frontiers | Research Topics

Current advances in genomics and gene editing tools for crop improvement in a changing climate scenario

Edited by
Vijay Rani Rajpal, Deepmala Sehgal, Ravi Valluru and Sukhwinder Singh

Published in
Frontiers in Genetics

FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-2811-2
DOI 10.3389/978-2-8325-2811-2

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Current advances in genomics and gene editing tools for crop improvement in a changing climate scenario

Topic editors
Vijay Rani Rajpal — University of Delhi, India
Deepmala Sehgal — Syngenta, United Kingdom
Ravi Valluru — University of Lincoln, United Kingdom
Sukhwinder Singh — The International Maize and Wheat Improvement Center (CIMMYT), India

Citation
Rajpal, V. R., Sehgal, D., Valluru, R., Singh, S., eds. (2023). *Current advances in genomics and gene editing tools for crop improvement in a changing climate scenario*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2811-2

Table of contents

06 Editorial: Current advances in genomics and gene editing tools for crop improvement in a changing climate scenario
Vijay Rani Rajpal, Deepmala Sehgal, Ravi Valluru and Sukhwinder Singh

10 Integrating CRISPR-Cas and Next Generation Sequencing in Plant Virology
Muntazir Mushtaq, Aejaz Ahmad Dar, Umer Basu, Basharat Ahmad Bhat, Rakeeb Ahmad Mir, Sanskriti Vats, M. S. Dar, Anshika Tyagi, Sajad Ali, Monika Bansal, Gyanendra Kumar Rai and Shabir Hussain Wani

25 CRISPR/Cas Genome Editing in Potato: Current Status and Future Perspectives
Jagesh Kumar Tiwari, Tanuja Buckseth, Clarissa Challam, Rasna Zinta, Nisha Bhatia, Dalamu Dalamu, Sharmistha Naik, Anuj K. Poonia, Rajesh K. Singh, Satish K. Luthra, Vinod Kumar and Manoj Kumar

31 Novel miRNA-SSRs for Improving Seed Hardness Trait of Pomegranate (*Punica granatum* L.)
Prakash Goudappa Patil, Nripendra Vikram Singh, Abhishek Bohra, Shivani Jamma, Manjunatha N, Venkatesh S. C, Dhinesh Babu Karuppannan, Jyotsana Sharma and Rajiv A. Marathe

45 Production of Conjoined Transgenic and Edited Barley and Wheat Plants for *Nud* Genes Using the CRISPR/SpCas9 System
Yiming Zang, Qiang Gong, Yanhao Xu, Huiyun Liu, Hao Bai, Na Li, Lipu Du, Xingguo Ye, Caixia Lan and Ke Wang

55 Genome-Wide Association Mapping Reveals Novel Putative Gene Candidates Governing Reproductive Stage Heat Stress Tolerance in Rice
K. T. Ravikiran, S. Gopala Krishnan, K. P. Abhijith, H. Bollinedi, M. Nagarajan, K. K. Vinod, P. K. Bhowmick, Madan Pal, R. K. Ellur and A. K. Singh

72 A General-Purpose Machine Learning R Library for Sparse Kernels Methods With an Application for Genome-Based Prediction
Osval Antonio Montesinos López, Brandon Alejandro Mosqueda González, Abel Palafox González, Abelardo Montesinos López and José Crossa

84 A Prospective Review on Selectable Marker-Free Genome Engineered Rice: Past, Present and Future Scientific Realm
Rajveer Singh, Navneet Kaur, Umesh Preethi Praba, Gurwinder Kaur, Mohammad Jafar Tanin, Pankaj Kumar, Kumari Neelam, Jagdeep Singh Sandhu and Yogesh Vikal

99 **BSA-seq Identifies a Major Locus on Chromosome 6 for Root-Knot Nematode (*Meloidogyne graminicola*) Resistance From *Oryza glaberrima***
Gurwinder Kaur, Inderjit Singh Yadav, Dharminder Bhatia, Yogesh Vikal, Kumari Neelam, Narpinderjeet Kaur Dhillon, Umesh Preethi Praba, Gurjit Singh Mangat and Kuldeep Singh

114 **Advances in Crop Breeding Through Precision Genome Editing**
Gauri Nerkar, Suman Devarumath, Madhavi Purankar, Atul Kumar, R. Valarmathi, Rachayya Devarumath and C. Appunu

128 **Comprehending the evolution of gene editing platforms for crop trait improvement**
Priyanka Dhakate, Deepmala Sehgal, Samantha Vaishnavi, Atika Chandra, Apekshita Singh, Soom Nath Raina and Vijay Rani Rajpal

154 **Recent advances in date palm genomics: A comprehensive review**
Hifzur Rahman, Prashant Vikram, Zied Hammami and Rakesh Kumar Singh

174 **CRISPR for accelerating genetic gains in under-utilized crops of the drylands: Progress and prospects**
Kiran K. Sharma, Sudhakar Reddy Palakolanu, Joorie Bhattacharya, Aishwarya R. Shankhapal and Pooja Bhatnagar-Mathur

194 **RNA Pol III promoters—key players in precisely targeted plant genome editing**
Sakshi Dharmendra Kor, Naimisha Chowdhury, Ajay Kumar Keot, Kalenahalli Yogendra, Channakeshavaiah Chikkaputtaiah and Palakolanu Sudhakar Reddy

207 **A CRISPR way for accelerating cereal crop improvement: Progress and challenges**
Umer Basu, Syed Riaz Ahmed, Basharat Ahmad Bhat, Zunaira Anwar, Ahmad Ali, Aqsa Ijaz, Addafar Gulzar, Amir Bibi, Anshika Tyagi, Suresh M. Nebapure, Chengeshpur Anjali Goud, Shafat Ahmad Ahanger, Sajad Ali and Muntazir Mushtaq

231 **Machine learning in the estimation of CRISPR-Cas9 cleavage sites for plant system**
Jutan Das, Sanjeev Kumar, Dwijesh Chandra Mishra, Krishna Kumar Chaturvedi, Ranjit Kumar Paul and Amit Kairi

244 **CRISPR/Cas genome editing system and its application in potato**
Xin Hou, Xiaomeng Guo, Yan Zhang and Qiang Zhang

254 **Development of bread wheat (*Triticum aestivum* L) variety HD3411 following marker-assisted backcross breeding for drought tolerance**
Prashanth K. C. Kumar, Amasiddha Bellundagi, Hari Krishna, Mallana Gowdra Mallikarjuna, Ramya K. Thimmappa, Neha Rai, P. Shashikumara, Nivedita Sinha, Neelu Jain, Pradeep K. Singh, Gyanendra Pratap Singh and Kumble Vinod Prabhu

266 **Sustaining yield and nutritional quality of peanuts in harsh environments: Physiological and molecular basis of drought and heat stress tolerance**
Naveen Puppala, Spurthi N. Nayak, Alvaro Sanz-Saez, Charles Chen, Mura Jyostna Devi, Nivedita Nivedita, Yin Bao, Guohao He, Sy M. Traore, David A. Wright, Manish K. Pandey and Vinay Sharma

290 **Unclasping potentials of genomics and gene editing in chickpea to fight climate change and global hunger threat**
Charul Singh, Ramesh Kumar, Hansa Sehgal, Sharmista Bhati, Tripti Singhal, Gayacharan, M. S. Nimmy, Renu Yadav, Santosh Kumar Gupta, Naglaa A. Abdallah, Aladdin Hamwieh and Rajendra Kumar

| Frontiers in Genetics

TYPE Editorial
PUBLISHED 12 June 2023
DOI 10.3389/fgene.2023.1214679

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Andrew H. Paterson,
University of Georgia, United States

*CORRESPONDENCE
Vijay Rani Rajpal,
vijayrani2@gmail.com,
vrrajpal@hrc.du.ac.in
Deepmala Sehgal,
Deepmala.Sehgal@syngenta.com

RECEIVED 30 April 2023
ACCEPTED 06 June 2023
PUBLISHED 12 June 2023

CITATION
Rajpal VR, Sehgal D, Valluru R and Singh S (2023), Editorial: Current advances in genomics and gene editing tools for crop improvement in a changing climate scenario.
Front. Genet. 14:1214679.
doi: 10.3389/fgene.2023.1214679

COPYRIGHT
© 2023 Rajpal, Sehgal, Valluru and Singh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Current advances in genomics and gene editing tools for crop improvement in a changing climate scenario

Vijay Rani Rajpal[1]*, Deepmala Sehgal[2]*, Ravi Valluru[3] and Sukhwinder Singh[4]

[1]Depatment of Botany, Hansraj College, University of Delhi, Delhi, India, [2]Syngenta, Jeolett's Hill International Research Center, Bracknell, United Kingdom, [3]University of Lincoln, Lincoln, United Kingdom, [4]USDA-ARS, SHRS, Coral Gables, FL, United States

KEYWORDS

genomics, gene editing, crop improvement, abiotic stress, disease resistance, climate change, crop productivity

Editorial on the Research Topic
Current advances in genomics and gene editing tools for crop improvement in a changing climate scenario

Introduction

The unprecedented global climate change has severely impacted our environment and engendered severe threats to agricultural productivity (Shahzad et al., 2021; Cinner et al., 2022; Ozdemir, 2022). This has led to the emergence of new races of plant pathogens and insect pests, accentuated abiotic stresses, depleted water resources and shrunken arable land, posing grave challenges to the food security of an ever-increasing global population (IPCC Sixth Assessment Report, 2022). The advantages offered by the Green Revolution of the mid-1960s are also fading away, resulting in a fragile food system (Davis et al., 2019; John and Babu 2021). Agriculture today faces newer challenges exacerbated by genetic erosion, the narrow genetic base of commercial crops and environmental degradation. There is an urgent need to make agriculture more resilient and sustainable while still continuing to develop high-yielding, stress-resistant and climate-smart crop varieties.

The advancements in genomics and gene editing technologies have offered immense opportunities and potential solutions for the genetic improvement of crops (Gao 2021). A plethora of novel avenues opened by genomics and genome editing approaches are attributed to the evolution of valuable tools like next-generation sequencing (NGS) methods, state-of-the-art genotyping arrays, genome mapping and genomic selection technologies that have helped to expedite the crop breeding processes. Likewise, new gene editing platforms have allowed precise editing of agronomically important genes to generate new varieties with desired characteristics (Zhang 2020; Ahmad 2023).

The deployment of these technologies has laid down the foundation of modern breeding for effectively channelizing the underutilized diversity hidden in the crop wild relatives into elite gene pools (Sehgal et al., 2015; 2017; 2020; Singh et al., 2018; Singh et al., 2021). The

advanced breeding programs are assisted by cutting-edge genomics, sequencing and genome editing technologies and have integrated artificial intelligence, machine learning bioinformatics and other related disciplines to meet the global food demands (Huang et al., 2022; Robert et al., 2022; Ahmad 2023).

In this Research Topic, we have collated a total of 19 articles, including original research and review articles, to get an essence of the spectrum of current efforts undertaken by applying modern tools for crop trait improvement. To develop a background on the theme, we have included three review articles entitled "Comprehending the evolution of gene-editing platforms for crop trait improvement" by Dhakate et al., "Advances in crop breeding through precision genome editing" by Nerkar et al. and "Integrating CRISPR-Cas and next-generation sequencing in plant virology" by Mushtaq et al. that have very well built up the narrative of the Research Topic. The other crop-specific comprehensive reviews entitled "CRISPR for accelerating genetic gains in under-utilized crops of the drylands: progress and prospects" by Sharma et al., "Recent advances in date palm genomics-a comprehensive review" by Rahman et al., "Unclasping potentials of genomics and gene editing in chickpea to fight climate change and global hunger threat" by Singh et al., "A CRISPR way for accelerated improvement of cereal crops" by Basu et al., "CRISPR/Cas genome editing system and its application in potato" by Zhang et al., "Physiological and molecular basis of drought and heat stress tolerance to enhance productivity and nutritional quality of peanuts in harsh environments" by Puppala et al., "A perspective on selectable marker-free genome engineered rice: past, present and future scientific realm" by Singh et al. and "CRISPR/Cas genome editing in potato: current status and future" by Tiwari et al. provide an up-to-date detailed compilation of the published research in date palm, dry-land crops, peanuts, chickpea, potato, and rice.

There has been continuous evolution in the development of gene editing-based technologies involving CRISPR/Cas platforms. Traditionally used CRISPR/Cas nucleases followed Sequence-Specific Nucleases (SSNs) such as Zinc-Finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs), and led to domains such as epigenome editing, base editing, and prime editing. Dhakate et al., Tiwari et al. and Basu et al. have comprehensively reviewed the evolution of CRISPR/Cas systems into new-age methods of genome engineering across various plant species and the impact that they have had on tweaking plant genomes and associated outcomes on crop improvement initiatives. Hou et al. reviewed the four types of CRISPR/Cas structures and mechanisms available today and the application of CRISPR/Cas9 systems in overcoming the challenges of self-incompatibility and improving the quality and resistance of potato. The review describes how precise knocking or targeted mutagenesis of *S-RNase* and *Sli* genes using CRISPR/Cas9 has helped to create self-compatible and self-incompatible potatoes, respectively. Additionally, the suitability of CRISPR/Cas9 and CRISPR/Cas13a systems in knocking out more than 22 genes has been detailed. Kor et al. have described the role of RNA Pol III promoters in precisely targeting genetic variants in genome editing. Nerkar et al. has focused on the advancements in crop breeding through precision genome editing. This review has included an overview of different breeding approaches for agronomic traits such as disease resistance, abiotic stress tolerance, herbicide tolerance, yield, and quality improvement, reduction of anti-nutrients, improved shelf life; genome editing tools and approaches used for crop improvement and an update on the regulatory approval of the genome-edited crops. Sharma et al. has described the opportunities of implementing genome editing technologies in under-utilized crops to increase genetic gains.

Zang et al. generated *Nud* mutants in wheat (*TaNud*) and barley (*HvNud*) using CRISPR/Cas9-based gene editing and investigated the heritability of the mutations in wheat. *Nud* gene is a transcription factor controlling the formation of the caryopsis, and loss-of-function mutation in the gene leads to the naked hull phenotype in barley. With the latest CRISPR/Cas9-based gene editing, combined with PCR-RE (polymerase chain reaction-restriction enzyme) approach, Zang et al. achieved a high editing efficiency (51.7%) of the three *Nud* alleles/homologs in wheat. This study has proven that with the improved vector system and CRISPR/Cas technology, it is not difficult to achieve precise genetic modification even in complex polyploid crops such as wheat which remained calcitrant to genetic modification technologies for a long time.

Heat and drought stresses cause substantial yield losses to crop production. According to the latest estimates, the heatwave episodes will particularly intensify in the Indo-Gangetic plains of India, which supports rice–wheat cropping system (Krishnan et al., 2020). Simulation models have predicted a reduction in rice yield by 10% for every 1°C rise in ambient temperature (Peng et al., 2004; Mendez et al., 2021). Breeding heat tolerant rice varieties is therefore a major objective of Indian breeders. Vast genetic variation has been reported to be present for reproductive stage heat stress (RSHT) tolerance in rice, however, modern genetics and genomics approaches have not been utilized fully to explore this variation. Ravikiran et al. comprehensively investigated genetic variation in RSHT tolerance with the GWAS approach using the cutting-edge genotyping arrays available in rice. They utilized three GWAS models to identify significant marker-trait associations (MTAs) for spikelet fertility and grain yield. Most significantly, the authors reported a set of stable MTAs for both traits, showing an advantage of 6–10 g for yield and up to 28% for spikelet fertility. Additionally, they identified promising tolerant genotypes that carried favorable alleles of 29, 28, 25, or 24 putative MTAs, which could serve as new donors in nurseries.

Peanuts exposed to drought stress at the reproductive stage are prone to aflatoxin contamination, which imposes a restriction on the use of peanuts as health food and also adversely impacts the peanut trade (Masaka et al., 2022). The review by Puppala et al. focuses on the significant progress that has been made towards the characterization of germplasm for drought and heat stresses tolerance and identifying MTAs and QTLs associated with drought tolerance. Advances in phenomics and artificial intelligence to accelerate the timely and cost-effective collection of phenotyping data in large germplasm/breeding populations are also reviewed and discussed. A holistic breeding approach that considers drought and heat-tolerant traits to simultaneously address both stresses is introduced as a successful strategy to produce climate-resilient peanut genotypes with improved nutritional quality.

Kumar et al. reported development of bread wheat variety HD3411 following marker-assisted backcross breeding for

drought tolerance. They reportedly transferred four drought tolerance quantitative trait loci (QTLs) controlling traits, *viz.* canopy temperature, normalized difference vegetative index, chlorophyll content, and grain yield from the drought-tolerant donor line, C306 into a popular high-yielding, drought-sensitive variety HD2733. Marker-assisted selection coupled with stringent phenotypic screening was used to develop a promising genotype, HD3411. The line HD3411 has shown higher yield over selected cultivars and has been identified for varietal release and testing in the northeastern plain zone of the wheat-growing region in India.

Besides abiotic stresses, plant disease outbreaks threaten global food security significantly. The crop pathogens are responsible for a substantial reduction in global crop yield and productivity. The global burden of viral, bacterial and fungal pathogens in farmers' fields ranges between 20% and 30% (Savary et al., 2019). It poses grave challenges, such as the emergence, spread and evolution of novel and more virulent races. Further, to address the challenges of crop loss due to various pathogens, improved disease surveillance and detection methods along with developing disease-resistant varieties are extremely important. Mushtaq et al. have shown that the NGS coupled with CRISPR-based genome editing have enabled rapid engineering of resistance by directly targeting specific genomic sites of plant viruses and viroids. They also discussed the emerging developments in NGS technologies and CRISPR/Cas-based DNA or RNA tests for the characterization of plant viruses along with their potential advantages and limitations. Kaur et al. employed BSA-seq approach in a wild species of rice *Oryza glaberrima* and identified a novel locus on chromosome 6 for resistance to root-knot nematode (*Meloidogyne graminicola*). They reported 3 potential candidate genes within QTLs qNR6.1, qNR6.2 and qNR2.1. This research has expanded the breadth of genes available for resistance to root-knot nematode and the possibilities of deploying new genes in rice breeding.

Machine learning (ML) and artificial intelligence (AI) algorithms have been increasingly used nowadays to improvise genomic prediction models to predict the phenotypes of newly developed breeding lines. Montesinos-López et al. in the article "A general-purpose machine learning R library for sparse kernels methods (SKM) with an application for genome-based prediction" demonstrated the usefulness of six machine learning models based on sparse kernel algorithm in two major crops, wheat and maize. Most importantly, this new package with six models allows user to use different formats of data, i.e., binary, categorical, count and continuous response variables. Recently, ML models are also being explored to resolve the off-target problems associated with CRISPR-Cas9. An article on 'Machine learning in the estimation of CRISPR-Cas9 cleavage sites for plant system' by Das et al. developed models on three ML-based techniques to estimate the cleavage sites of a given sgRNA. Out of the 11 trained models, the models based on the random forest technique, Artificial Neural Networks (ANN1-ReLU) and Support Vector Machine (SVM-Linear), performed better in the estimation of CRISPR-Cas9 cleavage sites showing an accuracy of 96.27%.

Breeding for soft-seeded varieties is an important objective in pomegranate (*Punica granatum* L.) breeding programs. The genetic architecture of seed-type trait has not been investigated much using modern approaches, and this has hampered the development of farmer-preferred and commercially viable pomegranate varieties. The recent advances in whole genome sequencing and transcriptome profiling in pomegranate have opened vistas for large-scale discovery of markers for trait discovery and improvement. A previous study in pomegranate had identified novel pre-miRNAs for seed-type trait in pomegranate (Luo et al., 2018). However, PCR-based markers were not generated in the study. The comprehensive study by Patil et al. utilized 761 potential novel pre-miRNAs to identify SSRs and generate breeder-friendly markers for use in breeding. They identified 227 and 79 SSR motifs specific to 60 pri-miRNA and 65 pre-miRNA sequences, respectively, and mapped them on the Tunisia genome. Most importantly, using target prediction and network analysis they reported the association of five miRNA-SSRs i.e., miRNA_SH_SSR35, miRNA_SH_SSR36, miRNA_SH_SSR53, miRNA_SH_SSR69 and miRNA_SH_SSR103 with seed type trait, which could be deployed by pomegranate breeders.

To summarize, a diverse collection of research and review articles included in this Research Topic has generated valuable information on the development of genetic and genomic resources in a variety of cereals (wheat, barley and rice), legumes (chickpea and peanut), fruit (pomegranate and date palm) and underutilized dryland crops. While, the review articles presented information on the evolution and refinement of new-age genomics, genome editing, and genome prediction models based on ML and AI algorithms for crop improvement, the research articles involved their application culminating into disease resistant, drought and heat resistant, high yielding crop varieties for instance line HD3411 in wheat (Kumar et al.). We believe that the articles compiled in this Research Topic will expand the existing knowledge on the strategies of crop improvement to mitigate climate change and ensure future food security.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahmad, M. (2023). Plant breeding advancements with "CRISPR-Cas" genome editing technologies will assist future food security. *Front. Plant Sci.* 14, 1133036. doi:10.3389/fpls.2023.1133036

Cinner, J. E., Caldwell, I. R., Thiault, L., Ben, J., Blanchard, J. L., Coll, M., et al. (2022). Potential impacts of climate change on agriculture and fisheries production in 72 tropical coastal communities. *Nat. Commun.* 13, 3530. doi:10.1038/s41467-022-30991-4

Davis, K. F., Chhatre, A., Rao, N. D., Singh, D., Ghosh-Jerath, S., Mridul, A., et al. (2019). Assessing the sustainability of post-Green Revolution cereals in India. *Proc. Natl. Acad. Sci. U. S. A.* 116 (50), 25034–25041. doi:10.1073/pnas.1910935116

Gao, C. (2021). Genome engineering for crop improvement and future agriculture. *Cell* 184 (6), 1621–1635. doi:10.1016/j.cell.2021.01.005

Huang, X., Huang, S., Han, B., and Li, J. (2022). The integrated genomics of crop domestication and breeding. *Cell* 185 (15), 2828–2839. doi:10.1016/j.cell.2022.04.036

IPCC Sixth Assessment Report (2022). *Climate change (2022): Impacts, adaptation and vulnerability*. Available at: https://www.ipcc.ch/report/ar6/wg2/.[

John, D. A., and Babu, G. R. (2021). Lessons from the aftermaths of green revolution on food system and health. *Front. Sustain. food Syst.* 5, 644559. doi:10.3389/fsufs.2021.644559

Krishnan, R., Sanjay, J., Gnanaseelan, C., Mujumdar, M., Kulkarni, A., and Chakraborty, S. (2020). *Assessment of climate change over the Indian region: A report of the ministry of earth sciences (MOES), government of India*. Springer Nature, 226.

Luo, X., Cao, D., Zhang, J., Chen, L., Xia, X., Li, H., et al. (2018). Integrated microRNA and mRNA expression profiling reveals a complex network regulating pomegranate (*Punica granatum* L) seed hardness. *Sci. Rep.* 8, 9292–9314. doi:10.1038/s41598-018-27664-y

Masaka, V. P., Ndlovu, N., Tshalibe, R. S., Mhande, T. C., and Jombo, T. Z. (2022). Prevalence of aflatoxin contamination in peanuts and peanut butter from an informal market, harare, Zimbabwe. *Int. J. Food Sci.* 2022, 3761078. doi:10.1155/2022/3761078

Mendez, K., Adviento-Borbe, M., Lorence, A., and Walia, H. (2021). Significant shift of ambient night-time air temperature during rice growing season in major US rice states. *Am. J. Clim. Change* 10, 134–151. doi:10.4236/ajcc.2021.101006

Ozdemir, D. (2022). The impact of climate change on agricultural productivity in asian countries: A heterogeneous panel data approach. *Environ. Sci. Pollut. Res. Int.* 29 (6), 8205–8217. doi:10.1007/s11356-021-16291-2

Peng, S., Huang, J., Sheehy, J. E., Laza, R. C., Visperas, R. M., Zhong, X., et al. (2004). Rice yields decline with higher night temperature from global warming. *Proc. Natl. Acad. Sci. U. S. A.* 101 (27), 9971–9975. doi:10.1073/pnas.0403720101

Robert, P., Brault, C., Rincent, R., and Segura, V. (2022). Phenomic selection: A new and efficient alternative to genomic selection. *Methods Mol. Biol. Clift. N.J.)* 2467, 397–420. doi:10.1007/978-1-0716-2205-6_14

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., and Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.* 3 (3), 430–439. doi:10.1038/s41559-018-0793-y

Sehgal, D., Autrique, E., Singh, R., Ellis, M., Singh, S., and Dreisigacker, S. (2017). Identification of genomic regions for grain yield and yield stability and their epistatic interactions. *Sci. Rep.* 7:41578. doi:10.1038/srep41578

Sehgal, D., Mondal, S., Crespo-Herrera, L., Velu, G., Juliana, P., Huerta-Espino, J., et al. (2020) Haplotype-based, genome-wide association study reveals stable genomic regions for grain yield in CIMMYT spring bread wheat. *Front. Genet.* 11:589490. doi:10.3389/fgene.2020.589490

Sehgal, D., Vikram, P., Sansaloni, C. P., Ortiz, C., Pierre, C. S., Payne, T., et al. (2015). Exploring and mobilizing the gene bank biodiversity for wheat improvement. *PLoS One* 10 (7), e0132112. doi:10.1371/journal.pone.0132112

Shahzad, U., Fareed, Z., Shahzad, F., and Shahzad, K. (2021). Investigating the nexus between economic complexity, energy consumption and ecological footprint for the United States: New insights from quantile methods. *J. Clean. Prod.* 279, 123806. doi:10.1016/j.jclepro.2020.123806

Singh, S., Jighly, A., Sehgal, D., Burgueño, J., Joukhadar, R., et al. (2021). Direct introgression of untapped diversity into elite wheat lines. *Nat. Food* 2, 819–827. doi:10.1038/s43016-021-00380-z

Singh, S., Vikram, P., Sehgal, D., Burgueño, J., Sharma, A., et al. (2018). Harnessing genetic potential of wheat germplasm banks through impact-oriented-prebreeding for future food and nutritional security. *Sci. Rep.* 8, 12527. doi:10.1038/s41598-018-30667-4

Zhang, B. (2020). CRISPR/Cas9: A robust genome-editing tool with versatile functions and endless application. *Int. J. Mol. Sci.* 21 (14), 5111. doi:10.3390/ijms21145111

frontiers in Genetics

REVIEW
published: 25 October 2021
doi: 10.3389/fgene.2021.735489

Integrating CRISPR-Cas and Next Generation Sequencing in Plant Virology

Muntazir Mushtaq[1†], Aejaz Ahmad Dar[2†], Umer Basu[3†], Basharat Ahmad Bhat[4], Rakeeb Ahmad Mir[5], Sanskriti Vats[6], M. S. Dar[7], Anshika Tyagi[8], Sajad Ali[8], Monika Bansal[9], Gyanendra Kumar Rai[2]* and Shabir Hussain Wani[10]**

[1]Division of Germplasm Evaluation, ICAR-National Bureau of Plant Genetic Resources, New Delhi, India, [2]School of Biotechnology, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Jammu, India, [3]Division of Plant Pathology, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Jammu, India, [4]Department of Bioresources, University of Kashmir, Srinagar, India, [5]Department of Biotechnology, School of Biosciences and Biotechnology, BGSB University, Rajouri, India, [6]Department of Agricultural Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali, India, [7]Division of Plant Pathology, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Srinagar, India, [8]Department of Biotechnology, Yeungnam University, Gyeongsan, South Korea, [9]School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, India, [10]Mountain Research Centre for Field Crops, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Srinagar, India

OPEN ACCESS

Edited by:
Ravi Valluru,
University of Lincoln, United Kingdom

Reviewed by:
Jaindra Nath Tripathi,
International Institute of Tropical Agriculture (IITA), Kenya
Chunjie Fan,
Chinese Academy of Forestry, China

***Correspondence:**
Sajad Ali
sajadali84@gmail.com
Gyanendra Kumar Rai
gkrai75@gmail.com
Shabir Hussain Wani
shabirhussainwani75@gmail.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 02 July 2021
Accepted: 16 September 2021
Published: 25 October 2021

Citation:
Mushtaq M, Dar AA, Basu U, Bhat BA, Mir RA, Vats S, Dar MS, Tyagi A, Ali S, Bansal M, Rai GK and Wani SH (2021) Integrating CRISPR-Cas and Next Generation Sequencing in Plant Virology. Front. Genet. 12:735489. doi: 10.3389/fgene.2021.735489

Plant pathology has been revolutionized by the emergence and intervention of next-generation sequencing technologies (NGS) which provide a fast, cost-effective, and reliable diagnostic for any class of pathogens. NGS has made tremendous advancements in the area of research and diagnostics of plant infecting viromes and has bridged plant virology with other advanced research fields like genome editing technologies. NGS in a broader perspective holds the potential for plant health improvement by diagnosing and mitigating the new or unusual symptoms caused by novel/unidentified viruses. CRISPR-based genome editing technologies can enable rapid engineering of efficient viral/viroid resistance by directly targeting specific nucleotide sites of plant viruses and viroids. Critical genes such as elf (iso) 4E or eIF4E have been targeted via the CRISPR platform to produce plants resistant to single-stranded RNA (ssRNA) viruses. CRISPR/Cas-based multi-target DNA or RNA tests can be used for rapid and accurate diagnostic assays for plant viruses and viroids. Integrating NGS with CRISPR-based genome editing technologies may lead to a paradigm shift in combating deadly disease-causing plant viruses/viroids at the genomic level. Furthermore, the newly discovered CRISPR/Cas13 system has unprecedented potential in plant viroid diagnostics and interference. In this review, we have highlighted the application and importance of sequencing technologies on covering the viral genomes for precise modulations. This review also provides a snapshot vision of emerging developments in NGS technologies for the characterization of plant viruses and their potential utilities, advantages, and limitations in plant viral diagnostics. Furthermore, some of the notable advances like novel virus-inducible CRISPR/Cas9 system that confers virus resistance with no off-target effects have been discussed.

Keywords: virus, viroid, NGS, genome editing, CRISPR

INTRODUCTION

Plant viral diseases present the most central challenge to twenty-first century agriculture systems on a global scale. Viruses are recognized to cause destructive plant diseases which lead to considerable losses in terms of yield as well as quality in the majority of crop plants worldwide (Mushtaq et al., 2020; Rubio et al., 2020). The projected cumulative crop damage caused due to pathogens is up to 15%, out of which viruses are instrumental and contribute 47% loss in the total yield (Boualem et al., 2016; Mushtaq et al., 2020). The global cost of controlling infections in cultivated crops due to viruses is anticipated to be higher than US $30 billion per year (Nicaise, 2014; Sastry and Zitter, 2014; Cao et al., 2020). Viral infection in crops has intensified at an unprecedented speed because of climate change, global warming, the increasing food demands of the human population, and the movement of insect vectors are causing dramatic changes in farming practices and cropping systems that encourage the spread of catastrophic viral disease outbreaks (Trębicki et al., 2016; Mushtaq et al., 2020). In food-insecure countries, such epidemics are especially visible in subsistence agriculture. (Jones and Naidu, 2019). Agricultural explosive growth and rapid international trade expansion of plants and plant produce has contributed to the movement of viral diseases and disseminated them to wide geographical regions to mediate unpredictable implications on the ecosystem and food production (Mushtaq et al., 2020).

Owing to the unpredictable epidemiological nature of various virus pathosystems, there is not any versatile method to mitigate the harmful effects of viral diseases on different agro-ecosystems. Advances in technical expertise focusing on virus pathosystems, accelerated scientific progress, ground-breaking connectivity plans, and global logical networks create an incentive to develop epidemiological virus suppression intelligence for agricultural development and overall food security. A paradigm shift towards the production of interconnected, smart, and sustainable solutions is required to advance the management of virus diseases in various cropping systems.

Plant viruses are obligate intracellular parasites, which have limited coding capacity and rely on the host plant for completing their life cycle. Unlike other plant diseases, there are hardly any successful remedies to cure harmful plant viruses without devising a specialized strategy. Consequently, plant molecular breeding is being considered as an indispensable tool to generate immunity, resistance, or tolerance to plant viruses in order to improve agricultural production.

An effective strategy to check viral attacks entails useful detection methods and thereafter deliberating the insights into the targeted viral genomes. The initial screening tools include PCR-based techniques such as RT-PCR and other variants. These diagnostic tools need prior knowledge of viral genomes and as a result, it ends up exhibiting poor detection of viruses with little information regarding their genomes (Shahid et al., 2021). So far, at least 1,500 plant virus species (26 families) have been known and characterized based on the genomic sequences of the viruses (Cao et al., 2020). Against these drawbacks, next-generation sequencing (NGS) may serve as an unbiased technology for the diagnosis of plant viral diseases, since no prior information about the pathogen is required. With this technology, plant virology is closely bridged with molecular biology through in-depth genomic information, leading to precise targeting of viral pathogens with significant improvement over existing technologies. Present-day NGS tools are capable of sequencing any type of nucleic acid, concomitantly. NGS technologies have emerged as the tool of choice to detect novel viral diseases from very few viral titers (Villamor et al., 2019). Therefore, with this technology, our understanding regarding phytoviromes has expanded horizons to facilitate future targeted approaches which will readily achieve their desired results.

Genome editing technologies have evolved to induce specific and targeted modifications into the plant genome to obtain desired results, such as the development of next-generation plant breeding through precision breeding systems.

The evolution of higher organisms is highly augmented by gene-editing technologies, such as Meganucleases, Zinc Finger Nuclease (ZFN), Transcription Activator like Effector Nuclease (TALEN), and CRISPR-Cas9 systems (Wiedenheft et al., 2011; Jinek et al., 2013; Zhang et al., 2013; Shahid et al., 2021). Amongst these technologies, the most recent CRISPR-Cas system imparts several advantages such as precise and flexible genome editing at the preferred genomic site to induce desirable mutations (Bortesi and Fischer, 2015). The CRISPR/Cas system has evolved as the leading and pioneering technology to edit genomes across all the kingdoms, although plant genome editing experiments were successfully carried out for the first time in 2013 (Li et al., 2013; Nekrasov et al., 2013; Shan et al., 2013). Since then, CRISPR/Cas mediated genome editing in plants has increased at a fast rate in contrast to the rest of new plant breeding technologies (NPBTs). CRISPR tool is based on RNA-programmed DNA cleavage systems, which were revealed in bacteria and archaea (Hsu et al., 2014). The last decade witnessed several reports regarding the diverse working principles of CRISPR-Cas based genome editing, especially the CRISPR/Cas9 system (Chen et al., 2019; Hanna and Doench, 2020).

In brief, CRISPR-Cas9 was found for the first time in *Streptococcus pyogenes* and reported as a type II immune system of prokaryotes against invading bacteriophages (Jinek et al., 2012). The later system relies on double-strand breaks (DSBs) induced at specific sites in the invading viral DNA. Consequently, DSBs trigger a DNA-repair mechanism in host cells through homology-directed repair (HDR) or non-homologous end-joining (NHEJ) (**Figure 1**) and induces insertions or deletions (indels) in the target viral DNA to make it non-functional against the host bacteria (Zaidi et al., 2020).

In an engineered CRISPR system, at the CRISPR locus a small CRISPR RNA (crRNA) is transcribed, which hybridizes with target genomic sequences through a complementary sequence present on the sequence flanking protospacer-associated motif site (PAM). The canonical 5′-NGG-3′ PAM is important in order to be recognized by the Cas9 protein for recognition and action in the target viral genome to induce immunity in *Streptococcus*

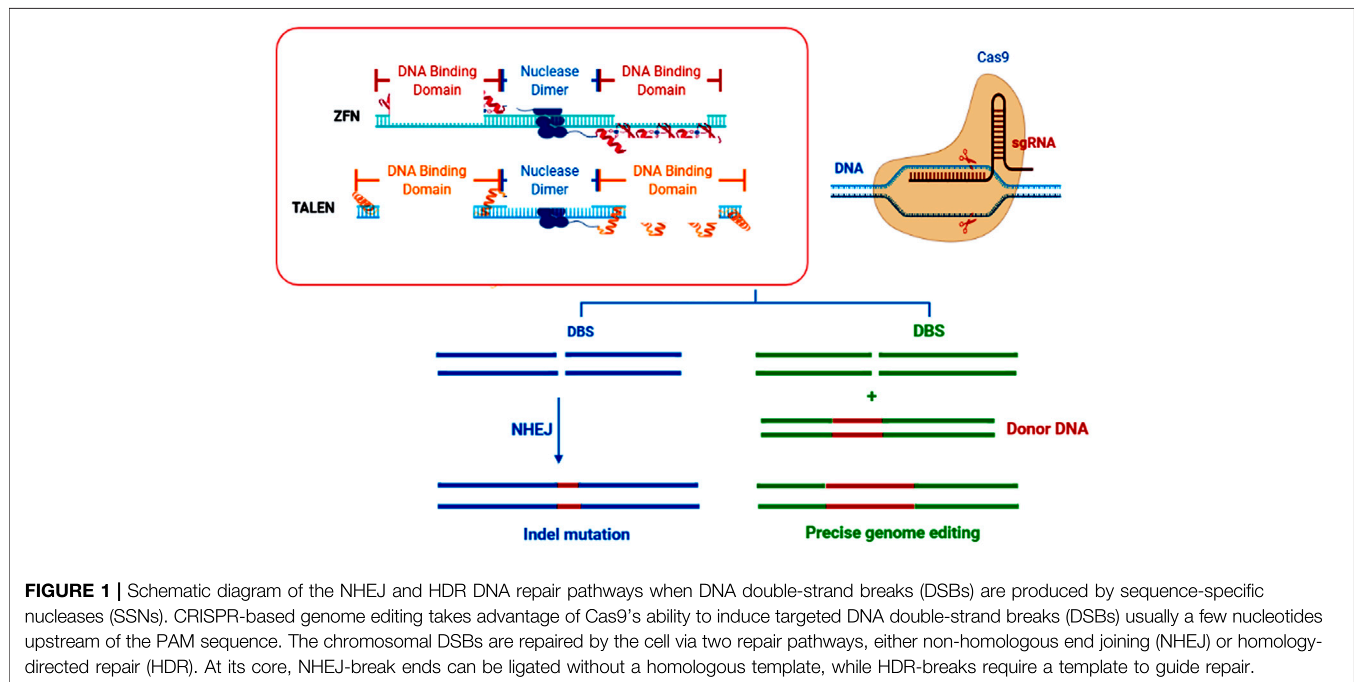


FIGURE 1 | Schematic diagram of the NHEJ and HDR DNA repair pathways when DNA double-strand breaks (DSBs) are produced by sequence-specific nucleases (SSNs). CRISPR-based genome editing takes advantage of Cas9's ability to induce targeted DNA double-strand breaks (DSBs) usually a few nucleotides upstream of the PAM sequence. The chromosomal DSBs are repaired by the cell via two repair pathways, either non-homologous end joining (NHEJ) or homology-directed repair (HDR). At its core, NHEJ-break ends can be ligated without a homologous template, while HDR-breaks require a template to guide repair.

pyogenes (Wright et al., 2016). Subsequently, a considerable portion of crRNA binds to trans-activating RNA (tracrRNA), and both bind to Cas9-gRNA complex to form a complete genome editing machinery. This complex now binds to complementary target sites in the target genome through gRNA and then Cas9 nuclease induces DSBs almost three nucleotides upstream of the PAM site (O'Connell et al., 2014). Accordingly, this system is therefore capable of generating precise, site-specific alterations in DNA via the synthetic single guide (sg) RNAs designed to direct Cas9-mediated cleavage at targeted sites (Hanna and Doench, 2020). The only criterion for CRIPSR to be used against the targeting of genes lies in the presence of a protospacer-adjacent pattern (PAM) sequence near the target site (Gleditzsch et al., 2019). Using CRISPR gene-editing for different targets requires only different spacer sequences; thus, it is quick, easy, effective, economical, and scalable (Zhang et al., 2020).

NGS is indispensable for genome editing experiments as well, especially clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) based gene editing. From validating (CRISPR) knockouts to examining off-target effects or other edits with targeted sequencing, NGS is employed at different steps of the genome editing workflow (**Figure 2**). Follow-up research can then be carried out using applications, for example, methylation and gene expression analysis with RNAseq to assess the functional impact of a certain gene edit (Bhat and Rao, 2020).

Contemporary advances in CRISPR/Cas based genome editing render it a desirable tool for developing or inducing plant defense. Two major pathways are employed by CRISPR/Cas systems to enhance the virus resistance in crop plants. The first way is the CRISPR/Cas-mediated targeted mutagenesis of specific genes in host plant responsible for contributing to the viral cycle, and second, CRISPR/Cas systems can be configured to work efficiently in plants for targeting viral genomes (Mushtaq et al., 2019; Kalinina et al., 2020; Zhao et al., 2020; Shahid et al., 2021). For instance, CRISPR/Cas9 systems could be used to directly target viruses with DNA as well as RNA genomes, while other CRISPR/Cas systems such as, Cas13a (Abudayyeh et al., 2017) and Cas9 from *Franciscella novicida* (FnCas9) (Price et al., 2015) could specifically target viruses which have RNA genomes (Green and Hu, 2017; Wolter and Puchta, 2018). In this review, we discuss the applications of CRISPR/Cas systems against diverse plant viruses by targeting the susceptible genes of the host or viral genomes (**Figure 3**), and additional advancements in this particular field. Further, we register certain possible recessive resistance genes which can be exploited in antiviral breeding programs and highlight the relevance of antiviral breeding based on recessive-resistance genes to produce virus-free plants. Finally, we address the problems and landscape for applications of CRISPR/Cas technology for the avoidance and management of plant viruses/viroids in the field. Overall, this review provides a snapshot vision of the role of NGS and spectacle applications of CRISPR-Cas editing technologies in plant virology.

NEXT-GENERATION SEQUENCING TECHNOLOGIES AS GO-TO TOOL FOR PLANT VIROLOGY

Several technological breakthroughs have been employed to overcome the precise detection of plant viromes. One among these modern technologies is Next-generation sequencing (NGS), a novel tool for viral detection in diseased plants. In 2009, NGS technologies were initiated in plant virology for genome

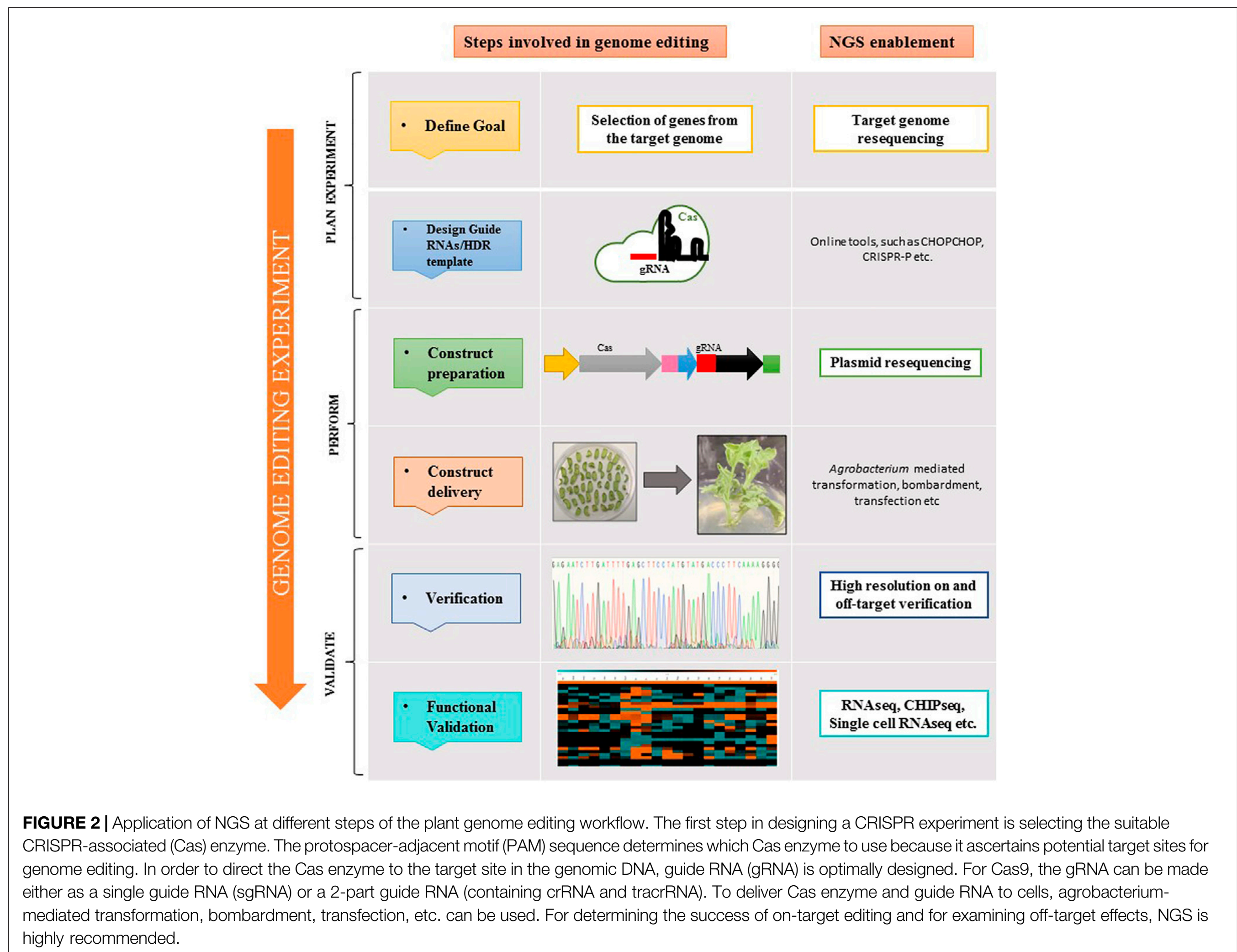


FIGURE 2 | Application of NGS at different steps of the plant genome editing workflow. The first step in designing a CRISPR experiment is selecting the suitable CRISPR-associated (Cas) enzyme. The protospacer-adjacent motif (PAM) sequence determines which Cas enzyme to use because it ascertains potential target sites for genome editing. In order to direct the Cas enzyme to the target site in the genomic DNA, guide RNA (gRNA) is optimally designed. For Cas9, the gRNA can be made either as a single guide RNA (sgRNA) or a 2-part guide RNA (containing crRNA and tracrRNA). To deliver Cas enzyme and guide RNA to cells, agrobacterium-mediated transformation, bombardment, transfection, etc. can be used. For determining the success of on-target editing and for examining off-target effects, NGS is highly recommended.

sequencing, discovery and identification, and epidemiology and ecology of viruses and viroids (Adams et al., 2009). The NGS pipeline involves the isolation of total RNA, DNA, or small RNA (sRNA) from the infected plant, the synthesis of cDNA and sequencing, then analyzing the sequencing data, contig development, and blasting the contigs sequence against a plant virus database assists in the recognition and characterization of target viruses. The detection and characterization of unknown and unidentified viruses/viroids from infected plants are probably the major promising application of such technologies (**Table 1**) (Barba et al., 2014). The RT-PCR helps to validate the NGS results using the complete RNA of the infected plant (Bhat and Rao, 2020). NGS provides a breakthrough to study viral diversity at taxonomic hierarchy levels. Its wide computational analyses by many programs and algorithms have created exciting opportunities for virus diagnostics and discovery. Unfolding evolutionary dynamics of viruses enhances the understanding of quasispecies diversity and the involvement of mutations in drug resistance and host switching, enabling the genotypic and phenotypic characterization of viruses (Kasibhatla et al., 2016).

Zhang et al. (2011), with the aid of deep and whole-genome sequencing, reported Grapevine vein clearing virus (DNA virus) in six grapevine cultivars linked with the vein-clearing symptom for the first time in Indiana, Missouri, and Illinois, indicating its widespread distribution in the Midwest of the United States. NGS coupled with metagenomic analysis was used to detect Pepino Mosaic Virus and Gayfeather Mild Mottle Virus infecting *Solanum lycopersicum* and *Gomphrena globosa* plant, respectively (Adams et al., 2009).

Ho and Tzanetakis (2014) have developed several barcodes for NGS through the degenerate oligonucleotide-primed RT-PCR method. Moreover, they created a universal bioinformatics tool, VirFind, exclusively for viral detection and discovery. They detected all the viruses in infected samples using this procedure. Such a method also mapped host reads, delivered files of virus reads, and hunted conserved domains for reads of unknown origin. Kehoe et al. (2014) demonstrated that on obtaining complete genomes of viruses through NGS, consideration should be given to sample preparation, efficient genome coverage, and assembly methods. Seguin et al. (2014) reported that deep siRNA sequencing is appropriate for universal

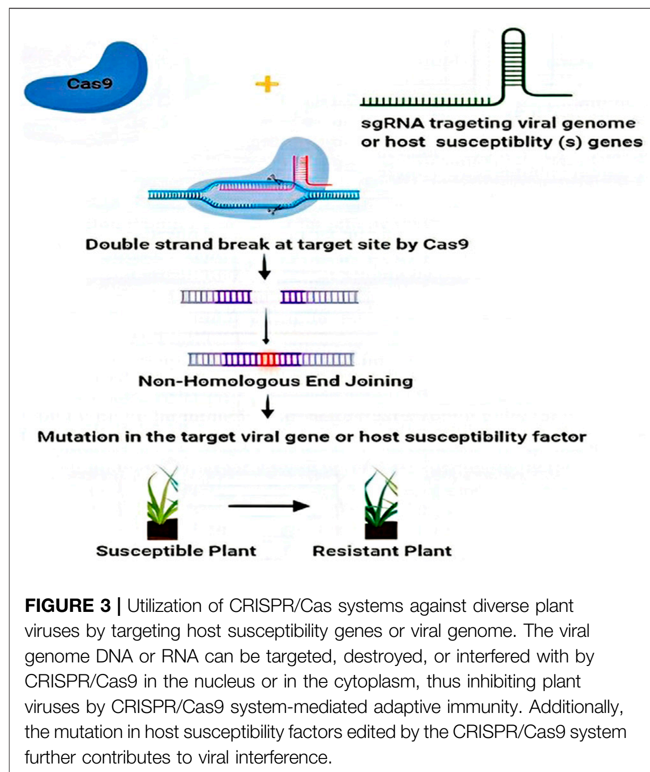


FIGURE 3 | Utilization of CRISPR/Cas systems against diverse plant viruses by targeting host susceptibility genes or viral genome. The viral genome DNA or RNA can be targeted, destroyed, or interfered with by CRISPR/Cas9 in the nucleus or in the cytoplasm, thus inhibiting plant viruses by CRISPR/Cas9 system-mediated adaptive immunity. Additionally, the mutation in host susceptibility factors edited by the CRISPR/Cas9 system further contributes to viral interference.

identification and classification of evolving viral quasispecies, and to understand fundamental mechanisms behind the biogenesis of siRNA and antiviral defense systems based on RNAi. The near-complete genome sequence of 22 isolates of several different virus species, namely, Potato virus Y, Tobacco vein banding mosaic virus, Cucumber mosaic virus, Tobacco mosaic virus, and *Brassica* yellow virus, were also identified and verified by NGS in infected tobacco plants of Anhui province of China (Akinyemi et al., 2016). Moreover, the discovery and identification of ssRNA viruses is accomplished by utilizing the ribo-depleted RNA in a *de novo* assembly-based method (Visser et al., 2016). They further suggested that a sequencing of one million reads can give adequate genome coverage, particularly for recognition of closterovirus, belonging to the family Closteroviridae and represented by 17 viral species, most of which cause necrosis and yellowing of phloem in plants (Fuchs et al., 2020).

The combinatorial approach based on NGS and automated viral surveillance and diagnosis (VSD) bioinformatics toolkit provided several workflows for distinct pathogenic viruses and viroids that facilitated the surveillance and diagnosis of viral pathogens in plants (Barrero et al., 2017). Multiple reports have proved the utility of NGS for viral detection and identification. Pecman et al. (2017) employed Illumina sequencing to identify and detect plant viruses by comparing RNA sequences of small RNAs with ribo-depleted RNA. The ribo-depleted RNA-based generated datasets were used to identify the putative novel Cytorhabdovirus, due to the reduced number of short reads in the latter. On the contrary, higher yields of viruses and viroid sequences were reported in sRNA pool with no RNA replicative intermediates. Rott et al. (2017) demonstrated the efficiency of NGS by comparing 178 infected tree fruit specimens by conventional as well as NGS methods. NGS was deemed an advanced tool for the identification of novel or poorly characterized viruses relative to traditional bioassays. Bomer et al. (2018) detected the genomes of novel isolates of genera *Badnavirus* and *Potyvirus* by NGS in *Dioscorea* spp. propagated by a robust tissue culture technique. The NGS confirmed its utility in diagnosing yam viruses, contributing towards the safe movement of germplasm between different countries. Liu et al. (2018) sequenced small RNAs by NGS in seven sunflower varieties imported from the United States and the Netherlands. After analysis, a novel endornavirus of double-stranded RNA molecule was detected in two sunflower varieties, X3939 and SH1108. High-throughput sequencing (HTS) goes hand in hand with bioinformatics algorithms for detecting viruses with a higher sensitivity rate. Different algorithms have been employed with HTS to detect twelve plant viruses through small RNA sequencing from three different infected plants (Massart et al., 2019). The virus detection sensitivity ranged from 35 to 100%, reproducibility was 91.6% and the false-positive detection rate was poor. High-throughput sequencing also significantly unveiled the presence of 03 isolates of *Potato virus M* in tomato plants in Slovakia (Glasa et al., 2019). One viroid and eight viruses have also been reported by sequencing of sRNA libraries from infected *Prunus persica* cv. *Nectarina* tree (Xu et al., 2019). In tomato plants, the 10 most abundant sequence variants of potato spindle tuber viroid RG1, differentially expressed with varying time periods, were identified by HTS and thereafter analyzed by employing *in silico* analysis for viroid derived small RNAs (vd-sRNA) (Adkar-Purushothama et al., 2020b). Other studies employing NGS for tomato crops have been reported. Mahmoudieh et al. (2020) surveyed tomato fields from different Peninsular Malaysian regions for the presence and distribution of begomoviruses, Tomato yellow leaf curl Kanchanaburi virus and Pepper yellow leaf curl Indonesia virus by an ORF-based study. They also recognized a novel virus, Ageratum yellow vein Malaysia virus, and its genome of single-stranded DNA and betasatellite component obtained by using NGS showed the maximum sequence similarity with Ageratum yellow vein virus (99%) and Pepper yellow leaf curl betasatellite (91%), respectively. Further, Uehara-Ichiki et al. (2020) detected broad bean wilt virus 2, asparagus virus 2, gaillardia latent virus and tomato spotted wilt orthotospovirus by NGS and RT-PCR analysis in *Valeriana fauriei* Briq. Chiapello et al. (2020) evaluated about 16 libraries from 150 grapevine cultivars infected with *Plasmopara viticola* to characterize the virome associated with the oomycete pathogen. Many plant virus variants including a new ilavirus were detected in grapevine.

NGS discovered four new viruses, namely Camellia yellow ringspot virus (CaYRSV), Camellia chlorotic ringspot viruses (CaCRSVs), Camellia-associated marafivirus (CaMaV), and Camellia-associated badnavirus (CaBaV) in Chinese *Camellia japonica* plants. These studies led to the validation of CaCRSVs as a novel genus belonging to family *Fimoviridae*. On the other hand, CaBaV, CaYRSV, and CaMaV have similar genome organizations and sequence homology with known viruses of the genera *Idaeovirus*, *Badnavirus,* and

TABLE 1 | List of viruses/viroids detected by NGS.

S.No.	Viruses/viroids identified	Host plant	Sequencing platform	Sequencing target	Reference
1.	Emaravirus	*Acer pseudoplatanus*	Illumina HiSeq2500	Total RNA	Rumbou et al. (2021)
2.	PSTVd	*Solanum lycopersicum*	Illumina MiSeq sequencer	vd-sRNA	Adkar-Purushothama et al. (2020a)
3.	GaP1V1 and Grapevine cogu-like virus	*Vitis* spp.	Illumina Novaseq technology	Total RNA	Bertazzon et al. (2020)
4.	GFLV, GaCLV-1, GaCLV-2, GaCLV-3, GLRaV-1, GLRaV-3, GalV-1, Entoleucaphenui-like virus 1, GaJV-1, GaJV-2 and PaPLV 1	*Vitis vinifera*	Illumina HiSeq	Total RNA and ribo-depleted RNA	Chiapello et al. (2020)
5.	Tomato yellow leaf curl Kanchanaburi virus, AYVMV	*Solanum lycopersicum* cv.MT1	Illumina Hiseq 2000	DNA	Mahmoudieh et al. (2020)
6.	23 viruses and viroids (including grapevine leafroll associated virus 4)	*Vitis vinifera*	Illumina NextSeq500 platform	mRNA and sRNA	Sidharthan et al. (2020)
7.	AV-2, BBWV-2, GalLV, CMV, TSWV and LNRSV	*Valeriana fauriei* Briq	-	Total RNA	Uehara-Ichiki et al. (2020)
8.	CaYRSV, CaCRSVs, CaBaV, and (CaMaV)	*Camellia japonica*	Illumina HiSeq X-ten platform	Total RNA	Zhang et al. (2020)
9.	PVM	*Solanum lycopersicum*	Illumina MiSeq platform along with High-throughput sequencing (HTS) and Sanger sequencing	Total RNA	Glasa et al. (2019)
10.	GLRaV-1, GRSPaV, GRGV and GYSVd1, GVA and Grapevine Syrah virus 1, HSVd	*Vitis berlandieri* × *V. riparia* 'Kober 125AA and	-	sRNA	Massart et al. (2019)
	PVX and PVB	*Solanum tuberosum*			
	ASGV	*Malus* × *domestica* 'Golden Delicious			
11.	CVT	*Prunus avium* L.	Illumina HiSeq 2500	Total RNA	Milusheva et al. (2019)
12.	PLMVd, APV1, PBNSPaV, CNRMV, CGRMV, CNRMV, NSPaV, ACLSV, and PeVD	*Prunus persica*cv. *Nectarina*	Illumina HiSeq. 2500 SE50	sRNA	Xu et al. (2019)
13.	DBRTV3-[2RT] and DBRTV3-[3RT]), YMV-NG	*Dioscorea* spp.	Illumina HiSeq4000	Total RNA	Bomer et al. (2018)
14.	CMV, ANSV and PYVV	*Solanum quitoense*	llumina HiSeq-2500	Total RNA and ribo-depleted RNA	Gallo et al. (2018)
15.	SMV	*Passifloraedulis*	llumina HiSeq-2500	Total RNA and ribo-depleted RNA	Jaramillo et al. (2018)
16.	HaEV	*Helianthus annuus*	Illumina HiSeq 2500	sRNA	Liu et al. (2018)
17.	GRSPaV, GVB, GFkV, GLRaV-3 and HSVd	*Vitis vinifera*	Illumina HiSeq2000	sRNAs	Barrero et al. (2017)
	PNRSV RNA1, PNRSV RNA2 and PNRSV RNA3	*Prunus persica*			
	(RBDV) RNA1, RBDV RNA2 and (RYNV)	*Rubusidaeus*			
	(RNA dependent RNA polymerase) [PCV (RdRp)] and Arisoteliachilensis virus (Reverse transcriptase) [ACV (RT)]	*Brassica sp.*			
	SPSMV-1	*Ipomoea batatas*			
	(SMoV) RNA1 and SMoV RNA2	*Fragaria ananassa*			
	SrMV	*Miscanthus sinensis*			
	CTV, CVd-VI and HSVd	*Citrus medica*			
	Citrus endogenous pararetrovirus	*Citrus sp.*			
	MCDV	*Pennisetum advena*			
	CTV	*Citrus latifolia*			
	AcVB	*Actinidia*			
	PVY and TVCV	*N. tabacum*			
18.	14 viral/viroid species detected by both sRNA and rRNA method. But a novel viral species from the CcyV1 genus was detected only by rRNA depleted totRNA approach	*Solanum tuberosum, Solanum lycopersicum, Brassica oleracea, N. tabacum, N. benthamiana, P. sativum* and *Prunu* ssp.	Illumina HiSeq 2500 (United States)	sRNA versus rRNA	Pecman et al. (2017)
19.	342 viruses/viroid sequences		Illumina HiSeq 2500	dsRNA	Rott et al. (2017)

(Continued on following page)

TABLE 1 | (*Continued*) List of viruses/viroids detected by NGS.

S.No.	Viruses/viroids identified	Host plant	Sequencing platform	Sequencing target	Reference
		P. domestica, *Prunusavium*, *P. persica*, *Pyrus communis and Malus domestica*			
20.	PVY,PeMV, BrYV, TMV, TVBM, CVMV, BBWV2, CMV	*N. tabacum*	Illumina HiSeq-2000	Total RNA	Akinyemi et al. (2016)
21.	PlAMV	*Viola grypoceras* and *Nandina domestica* Thunb	-	Total RNA	Komatsu et al. (2016)
22.	GLRaV-3 and HSVd CTV and CDVd	*Vitis vinifera* *Citrus paradisi*	Illumina HiSeq	sRNA and ribo-depleted RNA	Visser et al. (2016)
23.	PPSMV	*Cajanus cajan*	Illumina technology	Total RNA	Elbeaino et al. (2014)
24.	BYMV and ClYVV	*Lupinus angustifolius*	Illumina HiSeq2000	Total RNA	Kehoe et al. (2014)
25.	LChV1	Prunus *avium*	Roche 454 Pyrosequencing	dsRNA	Candresse et al. (2013)
26.	4 variants of PVY	Capsicum annuum L.	Roche 454	Total RNA	Fabre et al. (2012)
27.	GRSPaV,GSyV-1, HSVd and GYSVd1, GRVFV	Vitis vinifera	Illumina Genome Analyzer II	sRNAs	Giampetruzzi et al. (2012)
28.	Raspberry latent virus	Rubusidaeus	Illumina	dsRNA	Quito-Avila et al. (2011)
29.	Cotton leafroll dwarf virus	*Gossypium hirsutum*	Illumina Genome Analyzer	sRNAs	Silva et al. (2011)
30.	GVCV	*Vitis vinifera*	Solexa-Illumina platform	sRNA	Zhang et al. (2011)
31.	Rice stripe virus	*Oryza sativa*	Illumina Solexa Sequencer	sRNAs	Yan et al. (2010)
32.	PepMV GMMV	*Solanum lycopersicum*	GS FLX Genome Sequencer	cDNA	Adams et al. (2009)
33.	SPCSV and SPFMV	*Ipomoea batatas*	Illumina Genome Analyzer	sRNA	Kreuze et al. (2009)

Marafivirus, respectively. In addition, other known viruses such as geminivirus, bluner virus, and betaflexi viruses which existed in the form of heterogeneous mixtures were also detected (Zhang et al., 2020).

Moreover, it is reported that viral pathogens (variants) of green crinkle and apple russet ring are precisely identified by sequencing methods (Li et al., 2020). These studies led to the confirmation that one of the apple chlorotic leaf spot virus sequence variants infects apple to cause distinctive ring-shaped rust, and in addition, the apple stem pitting virus sequence variant was found to cause green crinkle on the fruits of infected apple plants (Li et al., 2020). Based on HTS technology results, Olmedo-Velarde et al. (2020) predicted that viroid-like RNAs (Vd-LRNAs) are correlated with fig trees. Molecular characterization showed that one of the RNAs was a circular RNA with a size ranging from 357 to 360 nucleotides. The biochemical and structural characteristics of this fig hammerhead viroid-like RNA (FHVd-LR) are noticeably identical to those earlier recorded for viroid-like satellite RNAs (Vd-LsatRNAs) and certain viroids. Further studies revealed that FHVd-LR is a unique viroid or Vd-LsatRNA. In accordance with the HTS results, the co-existence of FHVd-LR of dissimilar sizes inside the same host cannot be expelled. Bejerman et al. (2020) reported 70 new plant viral species belonging to negative-sense, single-stranded RNA virosphere by expertly reviewing the application of HTS approaches. It may further be noted that the viral families such as *Aspiviridae*, *Fimoviridae*, *Phenuiviridae*, *Rhabdoviridae*, and *Tospoviridae* include negative-sense and ambisense RNA (NSR) plant viral genomes. NGS-based techniques along with bioinformatic algorithms and (RT)-PCR-based assays have a large impact on viral discoveries by determining the viral genomic sequences, and thus authenticating its reliability in accurately detecting viruses infecting plants. Updated accounts pertaining to the potential of NGS-based high throughput sequencing provide a landmark in the deciphering of detailed information regarding the discovery of viromes to pave way for implementation of genome editing tools, especially CRISPR based tools to develop resistance against harmful viruses. The proceeding section will discuss a detailed and updated account on CRISPR mediated genome editing of desired plant species to develop resistance against economically important plant viruses.

CRISPR-BASED TECHNOLOGIES FOR PLANT VIRUS INTERFERENCE

A viral infection can cause up to 98% crop damage in most subtropical and tropical countries, which largely contributes to the global food crisis (Czosnek and Laterrot, 1997). Subsequently, to control the threat of viruses effectively, it is obligatory to boost immunity in crop plants to develop crop resistance to viruses. Over the last decade, limited success has been accomplished through conventional approaches to establish complete resistance against plant viruses. Molecular plant breeding could help in generating resilient plants, which are immune, resistant, or tolerant to viruses. Class II bacterial immune system-based CRISPR/Cas approaches have been extensively implemented and exploited for the modification and detection of nucleic acids (Garcia-Doval and Jinek, 2017; Ji et al., 2019; Cao et al., 2020). The editing of plant genomes based on CRISPR/Cas systems has advanced quickly in the direction of improving plants against devastating viruses. Viral resistance can be achieved in two ways, either by targeting host plant factors that are

TABLE 2 | Successful applications of CRISPR/Cas-mediated genome editing for enhancing plant resistance against viruses.

Plant species	Target region	Disease	Type of virus	Type of CRISPR variant	References
Targeting viral genomes using CRISPR/Cas-approaches					
Arabidopsis, *N. benthamiana*	CP, Rep, IR	BSCV	DNA	Cas9	Ji et al. (2015)
	RBS, 3x Rep, IR hairpin	Bean Yellow Dwarf Virus	DNA	Cas9	Baltes et al. (2015)
	IR, CP, RCRII of Rep	CMV/TMV (TYLCV)	DNA	Cas9	Ali et al. (2015)
	R, CP, RCRII of Re	Geminiviruses (CLCuKoV, MeMV)	DNA	Cas9	Ali et al. (2016)
	1A,CP, 3'UTR-A	CMV/TMV (TYLCV)	RNA	FnCas9 ˃ RNA	Zhang et al. (2018a)
	Hc-Pro, CP	TuMv	RNA	Cas13a	Aman et al. (2018a)
	IR, C1	CLCuMuV	DNA	Cas9	Yin et al. (2019)
Tomato	Rep, CP	TYLCV	DNA	Cas9	Tashkandi et al. (2018)
Potato	CI, Nib, CP, P3	PVY	DNA	Cas13a ˃ RNA	Zhan et al. (2019)
Barley	Rep/RepA, LIR, MP/CP	WDV	DNA	Cas9	Kis et al. (2019)
Rice	RNA	Southern rice black-streaked dwarf virus	RNA	LshCas13a	Zhang et al. (2019)
Targeting a host genome using CRISPR/Cas-approaches					
Arabidopsis	*AtEIF(iso)4E*	TMV	RNA	Cas9	Pyott et al. (2016)
Cucumber	*CseIF4E*	CVYV, ZYMV, PRSV-W	RNA	Cas9	Chandrasekaran et al. (2016)
Rice	*OseIF4G*	RTSV	RNA	Cas9	Macovei et al. (2018)
N. benthamiana	*CLC-Nb1a/b*	Reduced PVY intracellular replication	RNA	Cas9	Sun et al. (2018a)
Cassava (CBSV)	*MenCBP-1/2*	CBSV	RNA	Cas9	Gomez et al. (2019)
Glycine max (L.) Merr.	*GmF3H1/2, FNSII-1*	SMV	RNA	Cas9	Zhang et al. (2019)
M. balbisiana	ORFs of eBSV	Banana streak virus	DNA	Cas9	Tripathi et al. (2019)

responsible for replication of the viruses, or by destroying the viral genome and hence inhibiting viral replication (Borrelli et al., 2018; Mushtaq et al., 2020; **Table 2**).

Caulimoviridae and Geminiviridae are the most devastating DNA virus families infecting plants comprising single-stranded DNA, as well as those with double-stranded DNA genomes. Numerous independent studies have intended to specifically target and obliterate the genomic DNA of plant caulimoviruses or geminiviruses using genome-editing tools. Before the advent of the CRISPR/Cas systems, the zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) were grossly applied practical technologies to modify the plant host and viral DNAs. TALEN and ZFN mediated gene targeting in geminiviruses, including tobacco curly shoot virus and tomato yellow leaf curl China virus, resulted in viral resistant plants (Chen et al., 2014; Cheng et al., 2015). In contrast to ZFNs and TALENs, CRISPR/Cas systems are a more advantageous, easy, and promising tool for engineering plant resistance to viruses. The key to the susceptibility factor for plant-virus interactions lies in a versatile initiation factor 4E (eIF4E), a component of translation machinery in plants. Plant genomes harbor several recessive viral resistance genes, which encode up to 14 eukaryotic translational initiation factors (eIFs), such as, eIF4E, eIF4G, and other associated proteins. Cloning analysis of all these 14 viral resistant genes revealed that 4E (eIF4E) or its isoform eIF(iso)4E was coded by 12 genes (Wang and Krishnaswamy, 2012). Disruption of the eIF4E gene provides innate immunity to multiple potyviruses in different plant organisms. Following this information, CRISPR/Cas9 has been exercised to produce eIF4E-edited cucumber plants resistant to papaya ringspot mosaic virus-W and zucchini yellow mosaic virus (Chandrasekaran et al., 2016). Likewise, CRISPR/Cas9-based targeting of eIF4G in rice resulted in tungro spherical virus resistant rice plants (Macovei et al., 2018). In another study, Bastet et al. (2019) implemented the conversion of C >G (N176K) to the wild form eIF4E1 in *Arabidosis thaliana* with a cytidine base editor, subsequently resulting in resistance to clover yellow vein virus. Yoon et al. (2020) used CRISPR/Cas9 to produce targeted mutagenesis to determine whether eIF4E1 mutations in *Solanum lycopersicum* cv. Micro-Tom could impart resistance against potyviruses. Genotypic study of the eIF4E1-edited tomato plants in T_0, T_1, and T_2 generations illustrated that these mutations are transmissible to successive generations, and effectively confer resistance to PepMoV. Consequently, these investigations validated the applicability of CRISPR/Cas9 to augment the development of high-quality tomato crops for higher yield and biomass. Atarashi et al. (2020) demonstrated CRISPR/Cas9-mediated mutagenesis in the eIF4E1 gene of a commercial tomato cultivar. In addition to eIF4G, two deletions of three and nine nucleotides (3DEL and 9DEL) and a single nucleotide insertion (1INS) were found in close proximity to regions encoding amino acid residues essential for binding the 5′ mRNA cap structure. In agreement with earlier studies, inoculation tests with potato virus Y (genus Potyvirus) resulted in substantially reduced susceptibility to the N strain (PVYN), but not to the ordinary strain (PVYO), in 1INS plants. Results propose that genome editing could lead to additional resistance in contrast to mutation breeding. Editing of eIF4E alleles presents an alternative way to control CMV in tomato plants. They performed artificial mutagenesis in the eIF4E1 gene of a commercial tomato cultivar by utilizing CRISPR/Cas9. The recent successful recessive antiviral-type resistance approaches to potyviruses and associated plant viruses are largely based on eIF4s

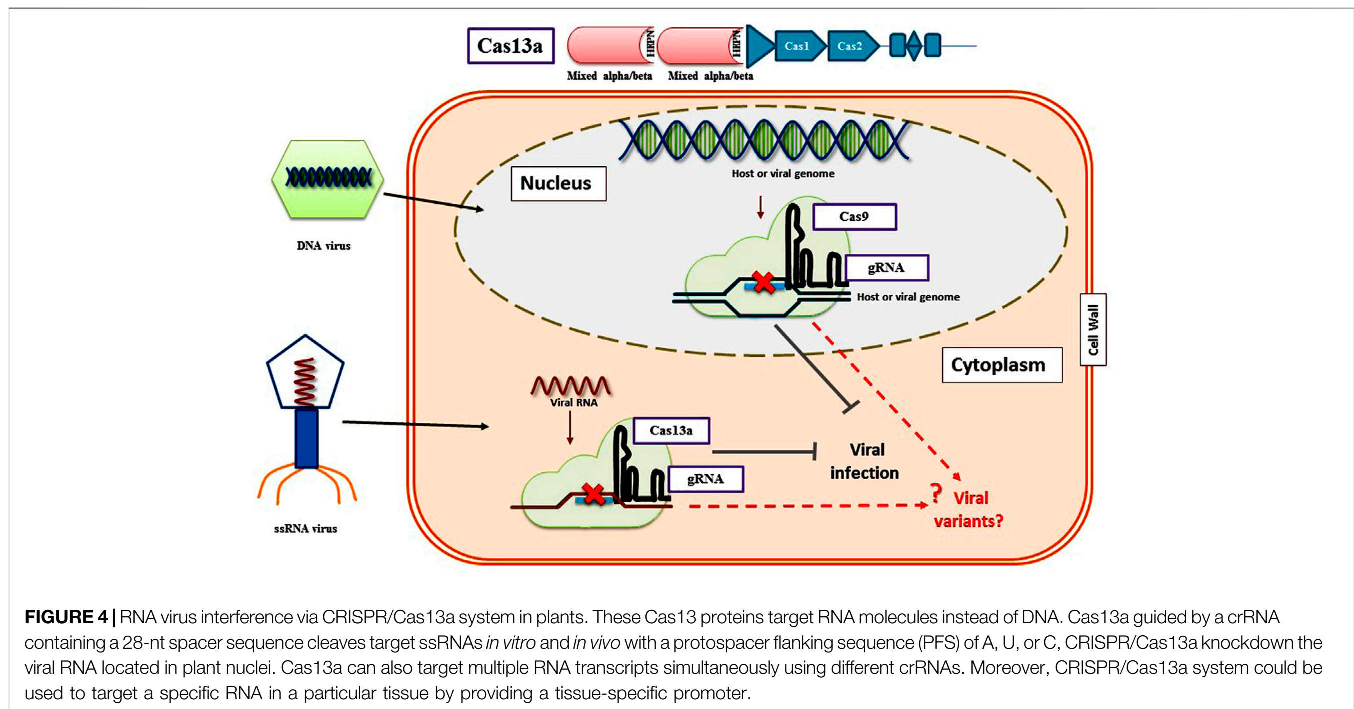


FIGURE 4 | RNA virus interference via CRISPR/Cas13a system in plants. These Cas13 proteins target RNA molecules instead of DNA. Cas13a guided by a crRNA containing a 28-nt spacer sequence cleaves target ssRNAs *in vitro* and *in vivo* with a protospacer flanking sequence (PFS) of A, U, or C, CRISPR/Cas13a knockdown the viral RNA located in plant nuclei. Cas13a can also target multiple RNA transcripts simultaneously using different crRNAs. Moreover, CRISPR/Cas13a system could be used to target a specific RNA in a particular tissue by providing a tissue-specific promoter.

and their homologs. Consequently, there is a need to identify more host susceptibility genes, which can be used as valuable genetic resources to combat economically vital plant viruses. CRISPR-mediated biomimicking mutations in Arabidopsis gene eIF4E1 led to the development of resistance against ClYVV (Bastet et al., 2019). Geminiviruses are known to cause significant losses to commercially valuable crops such as tomato, pepper, and sugarbeet (Langner et al., 2018). Several investigations were conducted for directly targeting the genomic DNA of geminiviruses via CRISPR/Cas9 approaches (Cao et al., 2020; Kalinina et al., 2020). Constructs containing sgRNAs that target the intergenic region (IR) and Rep (replication-associated protein) gene in beet severe curly top virus and bean yellow dwarf virus have been transformed into *Arabidopsis thaliana* and *N. benthamiana* respectively. The subsequent plants displayed a higher degree of resistance against the target viruses (Baltes et al., 2015; Ji et al., 2015). Undoubtedly, CRISPR-based genome editing tools can be effectively engineered to provide specific resistance towards geminiviruses in plants, but Ji et al. (2015) have shown in Arabidopsis plants that the importance of such an antiviral approach is compromised by the off-targeting observed by deep sequencing. They built two virus-inducible CRISPR/Cas9 vectors, which effectively inhibited the aggregation of Beet severely curly top virus in transient assays from transgenic lines of *Nicotiana benthamiana* and Arabidopsis. Deep sequencing did not detect any off-target mutations in the resulting transgenic Arabidopsis lines. Such types of virus-inducible gene-editing methods should be extensively used for designing virus-resistant plants exclusive of off-target costs. Similarly, Tomato and *N. benthamiana* plants expressing gRNA for coat protein or Rep sequences of tomato yellow leaf curl virus exhibit considerable resistance against viruses (Tashkandi et al., 2018). Further studies recorded that the use of CRISPR/Cas9 confers resistance against the wheat dwarf virus in barley plants and banana streak virus in banana plants by targeting conserved coding sequences present in the genome of the viruses (Kis et al., 2019; Tripathi et al., 2019).

Recently, RNA viruses have been targeted by Cas proteins that include Cas13a from *Leptotrichia shahii* and FnCas9 from *Francisella novicida*. These proteins target RNA molecules instead of DNA. The LshCas13a protein fused with the protospacers could be designed to knock down particular bacterial mRNAs (Abudayyeh et al., 2017). Researchers have engineered CRISPR/Cas13a machinery for *in planta* expression against different plant viruses (**Figure 4**). Aman et al. (2018a) successfully targeted the tobacco mosaic virus RNA in *N. benthamiana*. A later group of scientists also corroborated the applicability of CRISPR/LshCas13a to engineer *N. benthamiana* to develop resistance against Turnip mosaic virus (TuMV) (Aman et al., 2018a). These reports further paved the way to successfully engineered genomes of rice and *N. benthamiana* to develop resistance against viruses such as Southern rice black-streaked dwarf virus and TMV respectively (Zhang et al., 2019a). Moreover, in a study carried out by Zhang and co-workers, CRISPR/FnCas9 was used to degrade the cucumber mosaic virus and tobacco mosaic virus genome in transgenic lines of *N. benthamiana* and *Arabidosis thaliana* (Zhang et al., 2018b). The same strategies have been effectively used in rice to combat potato virus Y in tobacco and southern rice black-streaked draft virus (Zhang et al., 2019b). In the latest study conducted by Mahas et al. (2019), different Cas13 variants were characterized in order to identify the most specific interference against RNA viruses *in planta* in *N. benthamiana*. They demonstrated that

CRISPR-Cas13a system confers modest interference against RNA viruses. High interference activity of LwaCas13a, PspCas13b, and CasRx variants was reported against RNA viruses in transient assays.

In addition, the new Cas13 protein from *Ruminococcus flavefaciens* is classified as a type VI-D effector, called Cas13d (CasRx). Researchers established that Cas13d is advantageous over Cas13a, Cas13b, and Cas13 variants when used to target the CP, GFP, or HC-Pro region in the TuMV-GFP genome (Mahas et al., 2019). In a similar way, Cas13d has also been used in mammalian cells against novel coronavirus SARS-CoV-2 and influenza (Abbott et al., 2020).

Even though CRISPR/Cas-mediated genome editing is applied successfully to control the viral dissemination in transgenic plants, the probable risk of different virus escape events against the CRISPR/Cas9 cleavage results in resistance breakdown caused by the evolution of mutant viruses. Mehta et al. demonstrated that amid 33 and 48% of genome, edited viruses contain a conserved single base-pair mutation that imparted resistance against cleavage by CRISPR/CAS9 system, ensuing resistance breakdown to African cassava mosaic virus (ACMV) during inoculations in greenhouse conditions. While these novel variants of ACMV created by CRISPR/Cas9 mutagenesis might not multiply themselves, they depend on wild-type ACMV to proliferate in *N. benthamiana* (Mehta et al., 2019). The combination of two gRNAs, particularly at distance from each other, would significantly delay resistance breakdown in comparison to using only one sgRNA (Liu et al., 2018).

Thus, risks of virus escape from CRISPR-based antiviral immunity in plants predicts that this mechanism could be viewed as a double-edged sword for providing antiviral engineering. As it can destroy the genome of viruses to inhibit viral infection of crops, it poses a significant problem in that new variants of virus species might be created as by-products of genome editing, suggesting that it will increase the evolutionary process of viruses, or that evolved CRISPR-modified crops may lose their precise resistance to viral pathogens (Lassoued et al., 2019).

One essential aspect for the successful management of a disease is to detect the causal agent rapidly and with accuracy. Plant viruses are known for causing grave economic losses and pose a severe risk to agricultural sustainability. Therefore, optimization of the rapidity, sensitivity, practicability, portability, and precision of virus detection is urgently needed. Recent advances in genome editing technologies have shown that CRISPR-based systems, for example, Cas12a, Cas13a, and Cas14, encompass characteristics which can be used in the detection of nucleic acid (Gootenberg et al., 2017; Chen et al., 2018; Harrington et al., 2018). Cas12a possesses a DNase activity, which can randomly cut nonspecific ssDNA molecules into single/double nucleotides (Chen et al., 2018; Li et al., 2018a; Paul and Montoya, 2020). Researchers from the previous couple of years have used CRISPR-Cas9 protein variants, Cas12a and Cas13 to build easy, convenient, reliable, and economical platforms for nucleic acids detection at the molar level. The Zhang laboratory has exploited ribonuclease activity of the Cas13 protein to establish and refine the technique called Specific High Sensitivity Enzymatic Reporter UnLOCKING (SHERLOCK and SHERLOCKv2) (Gootenberg et al., 2017, Gootenberg et al.,2018). Whereas, Doudna's laboratory has exploited non-specific ssDNA degradation of Cas12a to establish a process referred to as DNA Endonuclease Targeted CRISPR *Trans* Reporter (DETECTR) (Chen et al., 2018). Both these nucleic acid detection tools exploited the promiscuous cleavage and degradation of adjacent ssRNA and ssDNA using Cas13 and Cas12a, to cleave and activate a reporter. Researchers demonstrated that the SHERLOCK and DETECTR showed the utmost sensitivity and accuracy for the detection of pathogenic viruses (Myhrvold et al., 2018; Chaijarasphong et al., 2019), transgenes (Zhang et al., 2020), and microorganisms (Zhang et al., 2020a).

Co-infection of apple trees with some viruses and viroids is widespread and declines the quality and yield of fruits. Rapid identification of viral pathogens with precision aids in the prevention of virus spread and reduces losses. Existing molecular tests used for the detection of apple viral pathogens involve specialized and costly apparatus. Jiao et al. (2020) optimized a CRISPR/Cas12a based detection approach for the identification of foremost prevailing RNA viruses or viroids in apple, namely apple necrotic mosaic virus, apple stem grooving virus, and stem-pitting virus. Each RNA virus was detected directly from raw leaf extracts following high specificity reverse transcription-recombinase polymerase amplification (RT-RPA). Nevertheless, this procedure was rapid and simple, requiring only about an hour to analyze the leaf samples. This innovative Cas12a-based approach is ideal for rapid and accurate identification of viruses in apple orchards, without sending samples to a specialist laboratory.

The above-mentioned CRISPR-based diagnostic tools involve the isothermal amplification of a target sequence, followed by detection of a target using Cas12 in DETECTR or Cas13 in SHERLOCK techniques and the collateral cleavage of a DNA or RNA reporter to specify the presence of the target (Kellner et al., 2019). Regardless of its extensive use for uncovering various pathogens in animals and humans (Li et al., 2019; Van Dongen et al., 2020). Mahas et al. (2020) reported the development and confirmation of a CRISPR-based nucleic acid diagnostic method exploiting the CRISPR–Cas12a system for detecting two geminiviruses, tomato yellow leaf curl virus (TYLCV) and tomato leaf curl New Delhi virus (ToLCNDV). They were successful in detecting TYLCV and ToLCNDV in infected plants with high sensitivity and specificity. This novel nucleic acid detection system can be used to perform an assay in approximately 1 h and makes available easy-to-infer visual readouts by the use of a simple, inexpensive fluorescence visualizer, thus providing an appropriate technique for point-of-use applications (**Figure 5**). Various reports have established the direct LAMP (loop-mediated isothermal amplification)-based amplification of viral sequences from crude extracts (Panno et al., 2020). This study demonstrates that the LAMP-coupled Cas12a technique is a valid rapid diagnostic tool for plant DNA viruses. Consequently, further upgrading of the LAMP-coupled Cas12a method could make possible the development of this assay as an in-field diagnostic test. The critical advantage of CRISPR-based

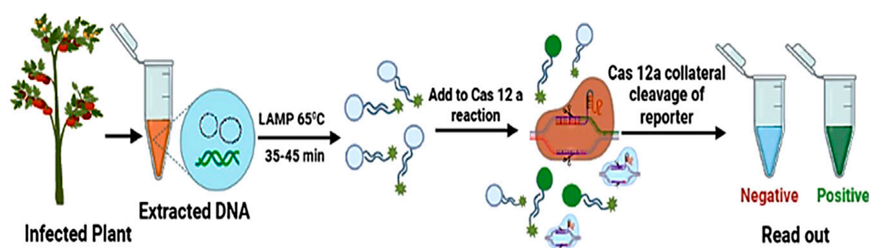


FIGURE 5 | LAMP-coupled Cas12-based assay for the detection of plant viruses. Viral DNA extracted from an infected tomato plant is amplified by loop-mediated isothermal amplification (LAMP), followed by CRISPR-mediated detection. Cas12a-based detection of the LAMP product triggers collateral cleavage of the reporter, thus producing a signal for visual detection.

genome editing lies in its lower off-targeting property to modulate crop genomes for antiviral detection against viruses. The aforementioned reports clearly demonstrate the potential of CRISPR-based genome-editing systems as versatile, efficient, and precise strategies to develop robust antiviral immune systems in crop plants.

CONCLUSION AND FUTURE DIRECTIONS

Recent approaches applied in virology are deeply influenced by the technical inputs of NGS technologies. Various sequencing platforms and sample preparation methods used worldwide in research laboratories have led to advancements in the detection and diagnosis of viroids and plant viruses. The last decade witnessed frequent involvement of NGS technologies, even though we also rely on alternative technical practices, preferably for characterization of different viruses. In the past, ELISA (1980s) and later PCR-based techniques (1990s) played the predominant role in detecting viral invasions and diagnosis of disease. However, NGS has facilitated the detection, investigation, and characterization of novel plant viruses that differentiates it from conventional diagnostic tools. The latest forms of NGS technologies, for instance, PacBio by Illumina, Oxford Nanopore, and ISS could be applied to considerably improve plant virology by offering rapid and more reliable viral detection with better precision. The use of HTS for viral diagnostics and the effect of this technique as a significant platform used in the detection of novel viruses have been thoroughly investigated. While different biotechnologies have their benefits and drawbacks, still we are in the infancy of utilizing the full capacity of RNAi and CRISPR/Cas in developing resistance against eukaryotic viruses. Despite the problem of GMO regulations, it can be seen that genome editing would be a powerful method for speeding resistance breeding, taking into account the ever-expanded CRISPR toolkit. Later tool kits can induce mutations to promote the generation of virus-resistant crop ideotypes in cases where resilience in natural variation and wild relatives is restricted.

Thus, the CRISPR/Cas method is widely used tool for selective genome engineering related to other editing approaches and has been developed and implemented in a vast range of plants which act as hosts, and in pathogens, to dissect molecular mechanisms responsible for plant-pathogenic interactions and to improve host resistance to both RNA and DNA viruses. Moreover, several reports suggest that the CRISPR/Cas method has the potential to develop genes with gain-of-function and loss-of-function mutants to decipher plant-virus interactions, and reduce the damage caused by harmful viruses in crops plants.

CRISPR-Cas13 could potentially be employed in disease management of plant viroids over transgenics. For instance, potato spindle tuber viroid (PSTVd) replicates in the nucleus of infected plants, and the mature PSTVd is resistant to RNA interference, hence a CRIPSR-Cas13 system could prove a potential genome editing tool in developing plants resistant to PSTVd. CRIPSR-Cas13 is advantageous over RNAi in terms of specificity, and the cleaved RNA may be further processed by RNAi to design plants with better disease resistance. To sum up, CRIPSR-Cas13 is a novel means to knock down RNA with improved specificity in contrast to RNAi, and it may bestow plants with stronger disease resistance because of the synergistic effect of RNAi.

CRISPR/Cas9 prime editors and base editors can be used to achieve correct genome editing of SNP and SNP typed QTLs effectively in plants, offering manifold resistance for viral pathogens. In a recent study the base-editing-mediated gene evolution (BEMGE) approach has been developed. This innovative crop breeding approach has the ability to artificially evolve every endogenous gene in a plant with a tiled sgRNA library related to the target locus in the genome. Therefore, BEMGE is a potential technique for the transformation of functional genes associated with a defense reaction in plants (Kuang et al., 2020).

In conclusion, CRISPR/Cas technology has the ability to investigate the dynamic spectrum of plant-pathogen interactions. Along with the recent transformation of agriculture and plant disease system, we look forward to CRISPR-based tools contributing to the deciphering of plant-virus interactions in the future and the development of plants with durable and broad-spectrum disease tolerance. NGS and CRISPR-Cas nexus have so far played a crucial role in controlling plant viral diseases. In the coming future, fundamental biological concerns for antiviral engineering could be intercepted using CRISPR technologies and the ongoing GMO-related concerns of plant biosafety regulators may be invalidated.

AUTHOR CONTRIBUTIONS

MM and AD contributed equally to the manuscript work. MM, AD, UB, MD have written the manuscript. MM and AD prepared the tables. AT, MM, SV, RM, and BB prepared the figures. MM, AD, UB, MD, AT, SV, MB, RM, SA, GR, BB, and SW have modified and edited the manuscript.

FUNDING

MM and AD are thankful to University Grants Commission, Maulana Azad National Fellowship (F1-17.1/2016-17/MANF-2015-17-JAM-72849), and the Department of Science and Technology-Science and Engineering Research Board (DST-SERB, PDF/2017/001488) India, respectively, for providing the research fellowship.

REFERENCES

Abbott, T. R., Dhamdhere, G., Liu, Y., Lin, X., Goudy, L., Zeng, L., et al. (2020). Development of CRISPR as an Antiviral Strategy to Combat SARS-CoV-2 and Influenza. *Cell* 181, 865–876. doi:10.1016/j.cell.2020.04.020

Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., et al. (2017). RNA Targeting with CRISPR-Cas13. *Nature* 550 (7675), 280–284. doi:10.1038/nature24049

Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., et al. (2009). Next-Generation Sequencing and Metagenomic Analysis: A Universal Diagnostic Tool in Plant Virology. *Mol. Plant Pathol.* 10, 537–545. doi:10.1111/j.1364-3703.2009.00545.x

Adkar-Purushothama, C. R., Bolduc, F., and Bru, P. (2020a). Tuber Viroid Quasi-Species from Infection to Disease. *Front. Microbiol.* 11, 1235. doi:10.3389/fmicb.2020.01235

Adkar-Purushothama, C. R., Bolduc, F., Bru, P., and Perreault, J-P. (2020b). *Insights into Potato Spindle*. Sherbrooke, QC, Canada.

Akinyemi, I. A., Wang, F., Zhou, B., Qi, S., and Wu, Q. (2016). Ecogenomic Survey of Plant Viruses Infecting Tobacco by Next Generation Sequencing. *Virol. J.* 13, 181. doi:10.1186/s12985-016-0639-7

Ali, Z., Abulfaraj, A., Idris, A., Ali, S., Tashkandi, M., and Mahfouz, M. M. (2015). CRISPR/Cas9-mediated Viral Interference in Plants. *Genome Biol.* 16, 238. doi:10.1186/s13059-015-0799-6

Ali, Z., Ali, S., Tashkandi, M., Zaidi, S. S.-e. -A., and Mahfouz, M. M. (2016). CRISPR/Cas9-Mediated Immunity to Geminiviruses: Differential Interference and Evasion. *Sci. Rep.* 6, 26912. doi:10.1038/srep26912

Aman, R., Ali, Z., Butt, H., Mahas, A., Aljedaani, F., Khan, M. Z., et al. (2018a). RNA Virus Interference via CRISPR/Cas13a System in Plants. *Genome Biol.* 19, 1. doi:10.1186/s13059-017-1381-1

Atarashi, H., Jayasinghe, W. H., Kwon, J., Kim, H., Taninaka, Y., Igarashi, M., et al. (2020). Artificially Edited Alleles of the Eukaryotic Translation Initiation Factor 4E1 Gene Differentially Reduce Susceptibility to Cucumber Mosaic Virus and Potato Virus Y in Tomato. *Front. Microbiol.* 11, 564310. doi:10.3389/fmicb.2020.564310

Baltes, N. J., Hummel, A. W., Konecna, E., Cegan, R., Bruns, A. N., Bisaro, D. M., et al. (2015). Conferring Resistance to Geminiviruses with the CRISPR-Cas Prokaryotic Immune System. *Nat. Plants* 1, 15145. doi:10.1038/nplants.2015.145

Barba, M., Czosnek, H., and Hadidi, A. (2014). Historical Perspective, Development and Applications of Next-Generation Sequencing in Plant Virology. *Viruses* 6, 106–136. doi:10.3390/v6010106

Barrero, R. A., Napier, K. R., Cunnington, J., Liefting, L., Keenan, S., Frampton, R. A., et al. (2017). An Internet-Based Bioinformatics Toolkit for Plant Biosecurity Diagnosis and Surveillance of Viruses and Viroids. *BMC Bioinformatics* 18, 26. doi:10.1186/s12859-016-1428-4

Bastet, A., Zafirov, D., Giovinazzo, N., Guyon-Debast, A., Nogué, F., Robaglia, C., et al. (2019). Mimicking Natural Polymorphism in eIF4E by CRISPR-Cas9 Base Editing Is Associated with Resistance to Potyviruses by CRISPR-Cas9 Base Editing Is Associated with Resistance to Potyviruses. *Plant Biotechnol. J.* 17 (9), 1736–1750. doi:10.1111/pbi.13096

Bejerman, N., Debat, H., and Dietzgen, R. G. (2020). The Plant Negative-Sense RNA Virosphere: Virus Discovery through New Eyes. *Front. Microbiol.* 11, 588427. doi:10.3389/fmicb.2020.588427

Bertazzon, N., Chitarra, W., Angelini, E., and Nerva, L. (2020). Two New Putative Plant Viruses from Wood Metagenomics Analysis of an Esca Diseased Vineyard. *Plants*. doi:10.3390/plants9070835

Bhat, A. I., and Rao, G. P. (2020). *Next-Generation Sequencing for Diagnosis of Viruses Characterization of Plant Viruses*. New York, NY: Springer Protocols Handbooks. Humana, 389–395. doi:10.1007/978-1-0716-0334-5_41

Bömer, M., Rathnayake, A. I., Visendi, P., Sewe, S. O., Sicat, J. P. A., Silva, G., et al. (2019). Tissue Culture and Next-Generation Sequencing: A Combined Approach for Detecting Yam (*Dioscorea* spp.) Viruses. *Physiol. Mol. Plant Pathol.*, 105, 54–66. doi:10.1016/j.pmpp.2018.06.003

Borrelli, V. M. G., Brambilla, V., Rogowsky, P., Marocco, A., and Lanubile, A. (2018). The Enhancement of Plant Disease Resistance Using CRISPR/Cas9 Technology. *Front. Plant Sci.* 9, 1245. doi:10.3389/fpls.2018.01245

Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 System for Plant Genome Editing and beyond. *Biotechnol. Adv.* 33, 41–52. doi:10.1016/j.biotechadv.2014.12.006

Boualem, A., Dogimont, C., and Bendahmane, A. (2016). The Battle for Survival between Viruses and Their Host Plants. *Curr. Opin. Virol.* 17, 32–38. doi:10.1016/j.coviro.2015.12.001

Candresse, T., Marais, A., Faure, C., and Gentit, P. (2013). Association of Little Cherry Virus 1 (LChV1) with the Shirofugen Stunt Disease and Characterization of the Genome of a Divergent LChV1 Isolate. *Phytopathology* 103, 293–298. doi:10.1094/phyto-10-12-0275-r

Cao, Y., Zhou, H., Zhou, X., and Li, F. (2020). Control of Plant Viruses by CRISPR/Cas System-Mediated Adaptive Immunity. *Front. Microbiol.* 11, 593700. doi:10.3389/fmicb.2020.593700

Chaijarasphong, T., Thammachai, T., Itsathitphaisarn, O., Sritunyalucksana, K., and Suebsing, R. (2019). Potential Application of CRISPR-Cas12a Fluorescence Assay Coupled with Rapid Nucleic Acid Amplication for Detection of white Spot Syndrome Virus in Shrimp. *Aquaculture* 512, 734340. doi:10.1016/j.aquaculture.2019.734340

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., et al. (2016). Development of Broad Virus Resistance in Non-transgenic Cucumber Using CRISPR/Cas9 Technology. *Mol. Plant Pathol.* 17 (7), 1140–1153. doi:10.1111/mpp.12375

Chen, J. S., Ma, E., Harrington, L. B., Da Costa, M., Tian, X., Palefsky, J. M., et al. (2018). CRISPR-Cas12a Target Binding Unleashes Indiscriminate Single-Stranded DNase Activity. *Science* 360 (6387), 436–439. doi:10.1126/science.aar6245

Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019). CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture. *Annu. Rev. Plant Biol.* 70, 667–697. doi:10.1146/annurev-arplant-050718-100049

Chen, W., Qian, Y., Wu, X., Sun, Y., Wu, X., and Cheng, X. (2014). Inhibiting Replication of Begomoviruses Using Artificial Zinc finger Nucleases that Target Viral-Conserved Nucleotide Motif. *Virus Genes* 48, 494–501. doi:10.1007/s11262-014-1041-4

Cheng, X., Li, F., Cai, J., Chen, W., Zhao, N., Sun, Y., et al. (2015). Artificial TALE as a Convenient Protein Platform for Engineering Broad-Spectrum Resistance to Begomoviruses. *Viruses* 7, 4772–4782. doi:10.3390/v7082843

Chiapello, M., Rodríguez-Romero, J., Nerva, L., Forgia, M., Chitarra, W., Ayllón, M. A., et al. (2020). Putative New Plant Viruses Associated withPlasmopara Viticola-infected grapevine Samples. *Ann. Appl. Biol.* 176, 180–191. doi:10.1111/aab.12563

Czosnek, H., and Laterrot, H. (1997). A Worldwide Survey of Tomato Yellow Leaf Curl Viruses. *Arch. Virol.* 142 (7), 1391–1406.

Elbeaino, T., Digiaro, M., Uppala, M., and Sudini, H. (2014). Deep Sequencing Of Pigeonpea Sterility Mosaic Virus Discloses Five RNA Segments Related To Emaraviruses. *Virus Res.*. doi:10.1016/j.virusres.2014.03.022

Fabre, F., Montarry, J., Coville, J., Senoussi, R., Simon, V., and Moury, B. (2012). Modelling the Evolutionary Dynamics of Viruses within Their Hosts: A Case

Study Using High-Throughput Sequencing. *Plos Pathog.* 8, e1002654. doi:10.1371/journal.ppat.1002654

Fuchs, M., Moshe Bar-Joseph, T. C., Maree, H. J., GiovanniMartelli, P. M. J. M., Wulf Angelantonio, M., and Sead, S. (2020). ICTV Virus Taxonomy Profile: Closteroviridae. *J. Gen. Virol.* 101, 364–365.

Gallo, Y., Toro, L. F., Jaramillo, H., Gutiérrez, P. A., and Marín, M. (2018). Identificación y caracterización molecular del genoma completo de tres virus en cultivos de lulo (*Solanum quitoense*) de Antioquia (Colombia). *Rev. Colomb. Cienc. Hortic.* 12, 281–292. doi:10.17584/rcch.2018v12i2.7692

Garcia-Doval, C., and Jinek, M. (2017). Molecular Architectures and Mechanisms of Class 2 CRISPR-Associated Nucleases. *Curr. Opin. Struct. Biol.* 47, 157–166. doi:10.1016/j.sbi.2017.10.015

Giampetruzzi, A., Roumi, V., Roberto, R., Malossini, U., Yoshikawa, N., La Notte, P., et al. (2012). A New grapevine Virus Discovered by Deep Sequencing of Virus- and Viroid-Derived Small RNAs in Cv Pinot Gris. *Virus. Res.* 163, 262–268. doi:10.1016/j.virusres.2011.10.010

Glasa, M., Šoltys, K., Predajňa, L., Sihelská, N., Budiš, J., Mrkvová, M., et al. (2019). High-throughput Sequencing of Potato Virus M from Tomato in Slovakia Reveals a Divergent Variant of the Virus. *Plant Protect. Sci.* 55, 159–166. doi:10.17221/144/2018-pps

Gleditzsch, D., Pausch, P., Müller-Esparza, H., Özcan, A., Guo, X., Bange, G., et al. (2019). PAM Identification by CRISPR-Cas Effector Complexes: Diversified Mechanisms and Structures. *RNA Biol.* 16 (4), 504–517. doi:10.1080/15476286.2018.1504546

Gomez, M. A., Lin, Z. D., Moll, T., Chauhan, R. D., Hayden, L., Renninger, K., et al. (2019). Simultaneous CRISPR/Cas9-Mediated Editing of Cassava eIF4E Isoforms nCBP-1 and nCBP-2 Reduces Cassava Brown Streak Disease Symptom Severity and Incidence. *Plant Biotechnol. J.* 17, 421–434. doi:10.1111/pbi.12987

Gootenberg, J. S., Abudayyeh, O. O., Kellner, M. J., Joung, J., Collins, J. J., and Zhang, F. (2018). Multiplexed and Portable Nucleic Acid Detection Platform with Cas13, Cas12a, and Csm6. *Science* 360 (6387), 439–444. doi:10.1126/science.aaq0179

Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., Joung, J., et al. (2017). Nucleic Acid Detection with CRISPR-Cas13a/C2c2. *Science* 356 (6336), 438–442. doi:10.1126/science.aam9321

Green, J. C., and Hu, J. S. (2017). Editing Plants for Virus Resistance Using CRISPR-Cas. *Acta Virol.* 61 (2), 138–142. doi:10.4149/av_2017_02_02

Hanna, R. E., and Doench, J. G. (2020). Design and Analysis of CRISPR-Cas Experiments. *Nat. Biotechnol.* 38 (7), 813–823. doi:10.1038/s41587-020-0490-7

Harrington, L. B., Burstein, D., Chen, J. S., Paez-Espino, D., Ma, E., Witte, I. P., et al. (2018). Programmed DNA Destruction by Miniature CRISPR-Cas14 Enzymes. *Science* 6416, 839–842. doi:10.1126/science.aav4294

Ho, T., and Tzanetakis, I. E. (2014). Development of a Virus Detection and Discovery Pipeline Using Next Generation Sequencing. *Virology* 471, 54–60. doi:10.1126/science.aav4294

Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* 157 (6), 1262–1278. doi:10.1016/j.cell.2014.05.010

Jaramillo Mesa, H., Marín Montoya, M., and Gutiérrez, P. A. (2018). Molecular Characterization of Soybean Mosaic Virus (SMV) Infecting Purple Passion Fruit (Passiflora edulis F. Edulis) in Antioquia, Colombia. *Arch. Phytopathol. Plant Prot.* 51, 617–636. doi:10.1080/03235408.2018.1505411

Ji, X., Wang, D., and Gao, C. (2019). CRISPR Editing-Mediated Antiviral Immunity: a Versatile Source of Resistance to Combat Plant Virus Infections. *Sci. China Life Sci.* 62, 1246–1249. doi:10.1007/s11427-019-9722-2

Ji, X., Zhang, H., Zhang, Y., Wang, Y., and Gao, C. (2015). Establishing a CRISPR-cas-like Immune System Conferring DNA Virus Resistance in Plants. *Nat. Plants* 1, 15144. doi:10.1038/nplants.2015.144

Jiao, J., Kong, K., Han, J., Song, S., Bai, T., Song, C., et al. (2020). Field Detection of Multiple RNA Viruses/viroids in Apple Using a CRISPR/Cas12a-Based Visual Assay. *Plant Biotechnol. J.*, 1–12. doi:10.1111/pbi.13474

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed Genome Editing in Human Cells. *eLife* 2, e00471. doi:10.7554/eLife.00471

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* 337, 816–821. doi:10.1126/science.1225829

Jones, R. A. C., and Naidu, R. A. (2019). Global Dimensions of Plant Virus Diseases: Current Status and Future Perspectives. *Annu. Rev. Virol.* 6, 387–409. doi:10.1146/annurev-virology-092818-015606

Kalinina, N. O., Khromov, A., Love, A. J., and Taliansky, M. E. (2020). CRISPR Applications in Plant Virology: Virus Resistance and beyond. *Phytopathology* 110, 18–28. doi:10.1094/phyto-07-19-0267-ia

Kasibhatla, S. M., Waman, V. P., Kale, M. M., and Kulkarni-Kale, U. (2016). "Analysis of Next-Generation Sequencing Data in Virology - Opportunities and Challenges," in *Next Gener Seq*. Editor J. K. Kulski (Rijeka: IntechOpen). doi:10.5772/61610

Kehoe, M. A., Coutts, B. A., Buirchell, B. J., and Jones, R. A. C. (2014). Plant Virology and Next Generation Sequencing: Experiences with a Potyvirus. *PLoS One* 9, e104580. doi:10.1371/journal.pone.0104580

Kellner, M. J., Koob, J. G., Gootenberg, J. S., Abudayyeh, O. O., and Zhang, F. (2019). Sherlock: Nucleic Acid Detection with CRISPR Nucleases. *Nat. Protoc.* 14, 2986–3012. doi:10.1038/s41596-019-0210-2

Kis, A., Hamar, É., Tholt, G., Bán, R., and Havelda, Z. (2019). Creating Highly Efficient Resistance against Wheat dwarf Virus in Barley by Employing CRISPR/Cas9 System. *Plant Biotechnol. J.* 17, 1004–1006. doi:10.1111/pbi.13077

Komatsu, K., Yamashita, A. M., Sugawara, K., Verbeek, M., Fujita, N., Hanada, K., et al. (2017). Complete Genome Sequences Of Two Highly Divergent Japanese Isolates Of Plantago Asiatica Mosaic Virus. *Arch. Virol.* 162, 581–584. doi:10.1007/s00705-016-3110-6

Konermann, S., Lotfy, P., Brideau, N. J., Oki, J., Shokhirev, M. N., and Hsu, P. D. (2018). Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* 173 (3), 665–676. doi:10.1016/j.cell.2018.02.033

Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., et al. (2009). Complete Viral Genome Sequence and Discovery of Novel Viruses by Deep Sequencing of Small RNAs: A Generic Method for Diagnosis, Discovery and Sequencing of Viruses. *Virology* 388, 1–7. doi:10.1016/j.virol.2009.03.024

Kuang, Y., Li, S., Ren, B., Yan, F., Spetz, C., Li, X., et al. (2020). Base-editing-mediated Artificial Evolution of *OsALS1 in Planta* to Develop Novel Herbicide-Tolerant rice Germplasms. *Mol. Plant* 13, 565–572. doi:10.1016/j.molp.2020.01.010

Langner, T., Kamoun, S., and Belhaj, K. (2018). CRISPR Crops: Plant Genome Editing toward Disease Resistance. *Annu. Rev. Phytopathol.* 56, 479–512. doi:10.1146/annurev-phyto-080417-050158

Lassoued, R., Macall, D. M., Smyth, S. J., Phillips, P. W., and Hesseln, H. (2019). Risk and Safety Consideration of Genome Edited Crops: Expert Opinion. *Curr. Res. Biotechnol.* 1, 11–21. doi:10.1016/j.crbiot.2019.08.001

Li, B., Yan, J., Zhang, Y., Li, W., Zeng, C., Zhao, W., et al. (2020). CRISPR-Cas12a Possesses Unconventional DNase Activity that Can Be Inactivated by Synthetic Oligonucleotides. *Mol. Ther. Nucleic Acids* 19, 1043–1052. doi:10.1016/j.omtn.2019.12.038

Li, J. F., Norville, J. E., Aach, J., McCormack, M., Zhang, D., Bush, J., et al. (2013). Multiplex and Homologous Recombination-Mediated Genome Editing in Arabidopsis and Nicotiana Benthamiana Using Guide RNA and Cas9. *Nat. Biotechnol.* 31, 688–691. doi:10.1038/nbt.2654

Li, S.-Y., Cheng, Q.-X., Wang, J.-M., Li, X.-Y., Zhang, Z.-L., Gao, S., et al. (2018a). CRISPR-Cas12a-assisted Nucleic Acid Detection. *Cell Discov.* 4, 20. doi:10.1038/s41421-018-0028-z

Li, Y., Li, S., Wang, J., and Liu, G. (2019). CRISPR/Cas Systems towards Next-Generation Biosensing. *Trends Biotechnol.* 37, 730–743. doi:10.1016/j.tibtech.2018.12.005

Liu, H., Soyars, C. L., Li, J., Fei, Q., He, G., Peterson, B. A., et al. (2018). CRISPR/Cas9-mediated Resistance to Cauliflower Mosaic Virus. *Plant Direct* 2, e00047. doi:10.1002/pld3.47

Liu, W.-w., Xin, M., Cao, M.-j., Qin, M., Liu, H., Zhao, S.-q., et al. (2018). Identification, Characterization and Full-Length Sequence Analysis of a Novel Endornavirus in Common sunflower (*Helianthus annuu*s L.). *J. Integr. Agric.* 17, 2281–2291. doi:10.1016/s2095-3119(18)61963-x

Macovei, A., Sevilla, N. R., Cantos, C., Jonson, G. B., Slamet-Loedin, I., Čermák, T., et al. (2018). Novel Alleles of rice eIF4G Generated by CRISPR/Cas9-targeted Mutagenesis Confer Resistance to Rice Tungro Spherical Virus. *Plant Biotechnol. J.* 16, 1918–1927. doi:10.1111/pbi.12927

Mahas, A., Hassan, N., Aman, R., Marsic, T., Wang, Q., Ali, Z., et al. (2020). LAMP-Coupled CRISPR-Cas12a Module for Rapid and Sensitive Detection of Plant DNA Viruses. *Viruses* 13, 466.

Mahas, A., Aman, R., and Mahfouz, M. (2019). CRISPR-Cas13d Mediates Robust RNA Virus Interference in Plants. *Genome Biol.* 20 (1), 263. doi:10.1186/s13059-019-1881-2

Mahmoudieh, M., Noor, M. R. M., Harikrishna, J. A., Othman, R. Y., and Othman, R. Y. (2020). Identification and Characterization of Ageratum Yellow Vein Malaysia Virus (AYVMV) and an Associated Betasatellite Among Begomoviruses Infecting *Solanum lycopersicum* in Malaysia. *J. Appl. Genet.* 61, 619–628. doi:10.1007/s13353-020-00574-4

Massart, S., Chiumenti, M., De Jonghe, K., Glover, R., Haegeman, A., Koloniuk, I., et al. (2019). Virus Detection by High-Throughput Sequencing of Small RNAs: Large-Scale Performance Testing of Sequence Analysis Strategies. *Phytopathology* 109, 488–497. doi:10.1094/phyto-02-18-0067-r

Mehta, D., Stürchler, A., Anjanappa, R. B., Zaidi, S. S.-e. -A., Hirsch-Hoffmann, M., Gruissem, W., et al. (2019). Linking CRISPR-Cas9 Interference in Cassava to the Evolution of Editing-Resistant Geminiviruses. *Genome Biol.* 20 (1), 80. doi:10.1186/s13059-019-1678-3

Mushtaq, M., Mukhtar, S., Sakina, A., Dar, A. A., Bhat, R., Deshmukh, R., et al. (2020). Tweaking Genome-Editing Approaches for Virus Interference in Crop Plants. *Plant Physiol. Biochem.* 147, 242–250. doi:10.1016/j.plaphy.2019.12.022

Mushtaq, M., Sakina, A., Wani, S. H., Shikari, A. B., Tripathi, P., Zaid, A., et al. (2019). Harnessing Genome Editing Techniques to Engineer Disease Resistance in Plants. *Front. Plant Sci.* 10, 550. doi:10.3389/fpls.2019.00550

Myhrvold, C., Freije, C. A., Gootenberg, J. S., Abudayyeh, O. O., Metsky, H. C., Durbin, A. F., et al. (2018). Field-deployable Viral Diagnostics Using CRISPR-Cas13. *Science* 360, 444–448. doi:10.1126/science.aas8836

Nicaise, V., Staskawicz, B., Weigel, D., Jones, J. D., and Kamoun, S. (2013). Targeted Mutagenesis in the Model Plant *Nicotiana benthamiana* Using Cas9 RNA-Guided Endonuclease. *Nat. Biotechnol.* 31, 691–693. doi:10.1038/nbt.2655

Nicaise, V. (2014). Crop Immunity against Viruses: Outcomes and Future Challenges. *Front. Plant Sci.* 5, 660. doi:10.3389/fpls.2014.00660

O'Connell, M. R., Oakes, B. L., Sternberg, S. H., East-Seletsky, A., Kaplan, M., and Doudna, J. A. (2014). Programmable RNA Recognition and Cleavage by CRISPR/Cas9. *Nature* 516 (7530), 263–266. doi:10.1038/nature13769

Olmedo-Velarde, A., Navarro, B., Hu, J. S., Melzer, M. J., and Di Serio, F. (2020). Novel Fig-Associated Viroid-like RNAs Containing Hammerhead Ribozymes in Both Polarity Strands Identified by High-Throughput Sequencing. *Front. Microbiol.* 11, 1903. doi:10.3389/fmicb.2020.01903

Panno, S., Matić, S., Tiberini, A., Caruso, A. G., Bella, P., Torta, L., et al. (2020). Loop Mediated Isothermal Amplification: Principles and Applications in Plant Virology. *Plants* 9, 461. doi:10.3390/plants9040461

Paul, B., and Montoya, G. (2020). CRISPR-Cas12a: Functional Overview and Applications. *Biomed. J.* 43, 8–17. doi:10.1016/j.bj.2019.10.005

Pecman, A., Kutnjak, D., Gutiérrez-Aguirre, I., Adams, I., Fox, A., Boonham, N., et al. (2017). Next Generation Sequencing for Detection and Discovery of Plant Viruses and Viroids: Comparison of Two Approaches. *Front. Microbiol.* 8, 1998. doi:10.3389/fmicb.2017.01998

Price, A. A., Sampson, T. R., Ratner, H. K., Grakoui, A., and Weiss, D. S. (2015). Cas9-mediated Targeting of Viral RNA in Eukaryotic Cells. *Proc. Natl. Acad. Sci. USA* 112 (19), 6164–6169. doi:10.1073/pnas.1422340112

Pyott, D. E., Sheehan, E., and Molnar, A. (2016). Engineering of CRISPR/Cas9-Mediated Potyvirus Resistance in Transgene-Free Arabidopsisplants. *Mol. Plant Pathol.* 17, 1276–1288. doi:10.1111/mpp.12417

Quito-Avila, D. F., Jelkmann, W., Tzanetakis, I. E., Keller, K., and Martin, R. R. (2011). Complete Sequence and Genetic Characterization of Raspberry Latent Virus, a Novel Member of the Family Reoviridae. *Virus. Res.* 155, 397–405. doi:10.1016/j.virusres.2010.11.008

Rott, M., Xiang, Y., Boyes, I., Belton, M., Saeed, H., Kesanakurti, P., et al. (2017). Application of Next Generation Sequencing for Diagnostic Testing of Tree Fruit Viruses and Viroids. *Plant Dis.* 101, 1489–1499. doi:10.1094/pdis-03-17-0306-re

Rubio, L., Galipienso, L., and Ferriol, I. (2020). Detection of Plant Viruses and Disease Management: Relevance of Genetic Diversity and Evolution. *Front. Plant Sci.* 11, 1092. doi:10.3389/fpls.2020.01092

Rumbou, A., Candresse, T., von Bargen, S., and Büttner, C. (2021). Next-Generation Sequencing Reveals a Novel Emaravirus in Diseased maple Trees from a German Urban forest. *Front. Microbiol.* 11, 621179. doi:10.3389/fmicb.2020.621179

Sastry, K. S., and Zitter, T. A. (2014). "Management of Virus and Viroid Diseases of Crops in the Tropics," in *Plant Virus and Viroid Diseases in the Tropic* Dordrecht: Epidemiology and Management Springer, 149–480. doi:10.1007/978-94-007-7820-7_2

Seguin, J., Rajeswaran, R., Malpica-López, N., Martin, R. R., Kasschau, K., Dolja, V. V., et al. (2014). De Novo Reconstruction of Consensus Master Genomes of Plant RNA and DNA Viruses from siRNAs. *PLoS ONE* 9, e88513. doi:10.1371/journal.pone.0088513

Shahid, M. S., Sattar, M. N., Iqbal, Z., Raza, A., and Al-Sadi, A. M. (2021). Next-Generation Sequencing and the CRISPR-Cas Nexus: A Molecular Plant Virology Perspective. *Front. Microbiol.* 11, 609376. doi:10.3389/fmicb.2020.609376

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., et al. (2013). Targeted Genome Modification of Crop Plants Using a CRISPR-Cas System. *Nat. Biotechnol.* 31, 686–688. doi:10.1038/nbt.2650

Sidharthan, V. K., Sevanthi, A. M., Jaiswal, S., and Baranwal, V. K. (2020). Robust Virome Profiling and Whole Genome Reconstruction of Viruses and Viroids Enabled by Use of Available mRNA and sRNA-Seq Datasets in Grapevine (Vitis vinifera L.). *Front. Microbiol.* 11, 1232. doi:10.3389/fmicb.2020.01232

Silva, T. F., Romanel, E. A., Andrade, R. R., Farinelli, L., Østerås, M., Deluen, C., et al. (2011). Profile of Small Interfering RNAs from Cotton Plants Infected with the Polerovirus Cotton Leafroll dwarf Virus. *BMC Mol. Biol* 12, 40. doi:10.1186/1471-2199-12-40

Sun, H., Shen, L., Qin, Y., Liu, X., Hao, K., Li, Y., et al. (2018a). CLC-Nt1affectsPotato Virus Yinfection via Regulation of Endoplasmic Reticulum Luminal Ph. *New Phytol.* 220, 539–552. doi:10.1111/nph.15310

Tashkandi, M., Ali, Z., Aljedaani, F., Shami, A., and Mahfouz, M. M. (2018). Engineering Resistance against Tomato Yellow Leaf Curl Virus via the CRISPR/Cas9 System in Tomato. *Plant Signal. Behav.* 13, e1525996. doi:10.1080/15592324.2018.1525996

Trębicki, P., Vandegeer, R. K., Bosque-Pérez, N. A., Powell, K. S., Dader, B., Freeman, A. J., et al. (2016). Virus Infection Mediates the Effects of Elevated CO2 on Plants and Vectors. *Sci. Rep.* 6, 22785. doi:10.1038/srep22785

Tripathi, J. N., Ntui, V. O., Ron, M., Muiruri, S. K., Britt, A., and Tripathi, L. (2019). CRISPR/Cas9 Editing of Endogenous Banana Streak Virus in the B Genome of *Musa Spp*. Overcomes a Major challenge in Banana Breeding. *Commun. Biol.* 2, 46. doi:10.1038/s42003-019-0288-7

Uehara-Ichiki, T., Ohashi, M., Hanada, K., Igarashi, M., and Hishida, A. (2020). Identification and Pathogenicity of Plant Viruses Infecting *Valeriana Fauriei* in Japan. *J. Gen. Plant Pathol.* 86, 310–315. doi:10.1007/s10327-020-00925-3

Van Dongen, J. E., Berendsen, J. T. W., Steenbergen, R. D. M., Wolthuis, R. M. F., Eijkel, J. C. T., and Segerink, L. I. (2020). Point-of-care CRISPR/Cas Nucleic Acid Detection: Recent Advances, Challenges and Opportunities. *Biosens. Bioelectron.* 166, 112445. doi:10.1016/j.bios.2020.112445

Villamor, D. E. V., Ho, T., Al Rwahnih, M., Martin, R. R., and Tzanetakis, I. E. (2019). High Throughput Sequencing for Plant Virus Detection and Discovery. *Phytopathology* 109, 716–725. doi:10.1094/phyto-07-18-0257-rvw

Visser, M., Bester, R., Burger, J. T., and Maree, H. J. (2016). Next-generation Sequencing for Virus Detection: Covering All the Bases. *Virol. J.* 13, 85. doi:10.1186/s12985-016-0539-x

Wang, A., and Krishnaswamy, S. (2012). Eukaryotic Translation Initiation Factor 4E-Mediated Recessive Resistance to Plant Viruses and its Utility in Crop Improvement. *Mol. Plant Pathol.* 13 (7), 795–803. doi:10.1111/j.1364-3703.2012.00791.x

Wiedenheft, B., van Duijn, E., Bultema, J. B., Waghmare, S. P., Zhou, K., Barendregt, A., et al. (2011). RNA-guided Complex from a Bacterial Immune System Enhances Target Recognition through Seed Sequence Interactions. *Proc. Natl. Acad. Sci.* 108, 10092–10097. doi:10.1073/pnas.1102716108

Wolter, F., and Puchta, H. (2018). The CRISPR/Cas Revolution Reaches the RNA World: Cas13, a New Swiss Army Knife for Plant Biologists. *Plant J.* 94, 767–775. doi:10.1111/tpj.13899

Wright, A. V., Nuñez, J. K., and Doudna, J. A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* 164 (1–2), 29–44. doi:10.1016/j.cell.2015.12.035

Xu, Y., Li, S., Na, C., Yang, L., and Lu, M. (2019). Analyses of Virus/viroid Communities in Nectarine Trees by Next-Generation Sequencing and Insight into Viral Synergisms Implication in Host Disease Symptoms. *Sci. Rep.* 9, 12261. doi:10.1038/s41598-019-48714-z

Yan, F., Zhang, H., Adams, M. J., Yang, J., Peng, J., Antoniw, J. F., et al. (2010). Characterization of siRNAs Derived from rice Stripe Virus in Infected rice Plants by Deep Sequencing. *Arch. Virol.* 155, 935–940. doi:10.1007/s00705-010-0670-8

Yin, K., Han, T., Xie, K., Zhao, J., Song, J., and Liu, Y. (2019). Engineer Complete Resistance to Cotton Leaf Curl Multan Virus by the CRISPR/Cas9 System in Nicotiana Benthamiana. *Phytopathol. Res.* 1, 9. doi:10.1186/s42483-019-0017-7

Yoon, Y., Venkatesh, J., Lee, J., Kim, J., Lee, H., Kim, D., et al. (2020). Genome Editing of eIF4E1 in Tomato Confers Resistance to Pepper Mottle Virus. Front. *Plant Sci.* 11, 1098. doi:10.3389/fpls.2020.01098

Zaidi, S. S.-e. -A., Mahas, A., Vanderschuren, H., and Mahfouz, M. M. (2020). Engineering Crops of the Future: CRISPR Approaches to Develop Climate-Resilient and Disease-Resistant Plants. *Genome Biol.* 21, 289. doi:10.1186/s13059-020-02204-y

Zhan, X., Zhang, F., Zhong, Z., Chen, R., Wang, Y., Chang, L., et al. (2019). Generation of Virus-resistant Potato Plants by RNA Genome Targeting. *Plant Biotechnol. J.* 17, 1814–1822. doi:10.1111/pbi.13102

Zhang, M., Liu, C., Shi, Y., Wu, J., Wu, J., and Chen, H. (2020a). Selective Endpoint Visualized Detection of Vibrio Parahaemolyticus with CRISPR/Cas12a Assisted PCR Using thermal Cycler for On-Site Application. *Talanta* 214, 120818. doi:10.1016/j.talanta.2020.120818

Zhang, P., Du, H., Wang, J., Pu, Y., Yang, C., Yan, R., et al. (2019). Multiplex CRISPR/Cas9-mediated Metabolic Engineering Increases Soya Bean Isoflavone Content and Resistance to Soya Bean Mosaic Virus. *Plant Biotechnol. J.* 18. 1384–1395. doi:10.1111/pbi.13302

Zhang, Q., Xing, H. L., Wang, Z. P., Zhang, H. Y., Yang, F., Wang, X. C., et al. (2018b). Potential High-Frequency Off-Target Mutagenesis Induced by CRISPR/Cas9 in Arabidopsis and its Prevention. *Plant Mol. Biol.* 96, 445–456. doi:10.1007/s11103-018-0709-x

Zhang, S., Yang, L., Ma, L., Tian, X., Li, R., Zhou, C., et al. (2020). Virome of Camellia Japonica: Discovery of and Molecular Characterization of New Viruses of Different Taxa in Camellias. *Front. Microbiol.* 11, 945. doi:10.3389/fmicb.2020.00945

Zhang, T., Zheng, Q., Yi, X., An, H., Zhao, Y., Ma, S., et al. (2018a). Establishing RNA Virus Resistance in Plants by Harnessing CRISPR Immune System. *Plant Biotechnol. J.* 16 (8), 1415–1423. doi:10.1111/pbi.12881

Zhang, Y., Malzahn, A. A., Sretenovic, S., and Qi, Y. (2019b). The Emerging and Uncultivated Potential of CRISPR Technology in Plant Science. *Nat. Plants*. doi:10.1038/s41477-019-0461-5

Zhang, Y., Singh, K., Kaur, R., and Qiu, W. P. (2011). Association of A Novel DNA Virus With The Grapevine Vein-Clearing And Vine Decline Syndrome. *Phytopathol.* 101, 1081–1090. doi:10.1094/phyto-02-11-0034

Zhang, Y., Zhang, F., Li, X., Baller, J. A., Qi, Y., Starker, C. G., et al. (2013). Transcription Activator-Like Effector Nucleases Enable Efficient Plant Genome Engineering. *Plant Physiol.* 161, 20–27. doi:10.1104/pp.112.205179

Zhang, Y.-m., Zhang, Y., and Xie, K. (2020b). Evaluation of CRISPR/Cas12a-based DNA Detection for Fast Pathogen Diagnosis and GMO Test in Rice. *Mol. Breed.* 40, 11. doi:10.1007/s11032-019-1092-2

Zhang, Y., Malzahn, A. A., Sretenovic, S., and Qi, Y. (2019a). The Emerging and Uncultivated Potential of CRISPR Technology in Plant Science. *Nat. Plants* 5, 778–794. doi:10.1038/s41477-019-0461-5

Zhao, Y., Yang, X., Zhou, G., and Zhang, T. (2020). Engineering Plant Virus Resistance: From RNA Silencing To Genome Editing Strategies. *Plant Biotechnol. J.* 18, 328–336. doi:10.1111/pbi.13278

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Mushtaq, Dar, Basu, Bhat, Mir, Vats, Dar, Tyagi, Ali, Bansal, Rai and Wani. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

OPINION
published: 02 February 2022
doi: 10.3389/fgene.2022.827808

CRISPR/Cas Genome Editing in Potato: Current Status and Future Perspectives

Jagesh Kumar Tiwari[1], Tanuja Buckseth[1], Clarissa Challam[2], Rasna Zinta[1,3], Nisha Bhatia[1,4], Dalamu Dalamu[1], Sharmistha Naik[1,5], Anuj K. Poonia[4], Rajesh K. Singh[1], Satish K. Luthra[6], Vinod Kumar[1] and Manoj Kumar[6]*

[1]ICAR-Central Potato Research Institute, Shimla, India, [2]ICAR-Central Potato Research Institute, Shillong, India, [3]School of Bioengineering and Biosciences, Lovely Professional University, Phagwara, India, [4]School of Biotechnology, Shoolini University, Solan, India, [5]ICAR-National Research Centre for Grapes, Pune, India, [6]ICAR-Central Potato Research Institute, Meerut, India

Keywords: abiotic, biotic, CRiSPR/Cas, genome editing, Potato, quality

OPEN ACCESS

Edited by:
*Deepmala Sehgal,
International Maize and Wheat Improvement Center, Mexico*

Reviewed by:
*Muntazir Mushtaq,
National Bureau of Plant Genetic Resources (ICAR), India*

****Correspondence:***
*Jagesh Kumar Tiwari
jageshtiwari@gmail.com*

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *02 December 2021*
Accepted: *11 January 2022*
Published: *02 February 2022*

Citation:
Tiwari JK, Buckseth T, Challam C, Zinta R, Bhatia N, Dalamu D, Naik S, Poonia AK, Singh RK, Luthra SK, Kumar V and Kumar M (2022) CRISPR/Cas Genome Editing in Potato: Current Status and Future Perspectives. Front. Genet. 13:827808. doi: 10.3389/fgene.2022.827808

INTRODUCTION

Potato (*Solanum tuberosum* L.) ($2n = 4x = 48$) is the third most important food crop after rice and wheat in terms of human consumption. Potato is considered as the staple food in Europe and parts of Americas. In 2018, the world total potato production was 368.17 million tonnes led by China (90.26 mt) followed by India (48.53 mt) (FAOSTAT, 2018). The increasing world population from the now 7.7 to the expected 9.7 billion by 2050 has posed a great challenge of food availability (United Nations, 2019). Potato suffers from various pathogens, insect pests, and environmental abiotic stresses. The condition is worsening under the climate change scenario. In India, the mean potato productivity in major potato-growing states, which together account for about 90% of the national potato production, is likely to decline by 2.0% in 2050s and 6.4% in 2080s (Rana et al., 2020). To address these issues, conventional breeding has shown key roles in varietal development programs combined with the deployment of marker-assisted selection mainly for late blight, viruses, and potato cyst nematode–resistant varieties the world over such as Kufri Karan in India (ICAR-CPRI Annual Report, 2018-19). Later, potato transgenics have also been developed for resistance to diseases (e.g., late blight and viruses), abiotic stresses (e.g., heat and drought), insect pest (e.g., potato cyst nematode and potato tuber moth), processing quality (e.g., reduced cold-induced sweetening), but none of them are being applied at the field level. Hence, with the advancement of sequencing technologies and availability of the potato genome sequence (Potato Genome Sequencing Consortium, 2011), it is possible to modulate the target genes applying genomics tools like genome editing.

Genome editing is an advanced genomics tool which can be deployed for crop improvement by gene knock-out and insertion/deletion mutagenesis (Hameed et al., 2018). It allows double-stranded breaks (DSBs) at specific sites in the genome and repairs via naturally occurring DNA repair mechanisms, namely, nonhomologous end joining (NHEJ) or homologous recombination (HR). In the past, this system was earlier facilitated by protein-guided nucleases such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). But now, attention has been driven on the new RNA-guided nuclease called clustered regularly interspaced short palindromic repeats (CRISPR)—CRISPR associated (Cas) (Nadakuduti et al., 2018). The TALENs and ZFNs require particular expertise, longer timelines, and higher costs than those needed for assembling CRISPR/Cas. Indeed, a tremendous progress has been reported on the utility of CRISPR/Cas in crops. In potato, CRISPR/Cas has been demonstrated for tuber quality, disease resistance (late blight and potato virus Y), phenotype, and other traits (Dangol et al., 2019; Hameed et al., 2020; Hofvander et al., 2021). This article provides the current status of CRISPR/Cas, future perspectives, and challenges in potato.

TABLE 1 | Successful examples of application of CRISPR/Cas genome editing technology for biotic and abiotic stress resistance/tolerance, tuber quality, and phenotype and other traits in potato.

Target gene	Trait	CRISPR system	Delivery/ transformation system	Genotype	Key findings	Reference
Biotic stress resistance						
P3, *CI*, *NIb*, or *CP* (RNA virus genes)	PVY, PVS, and PVA resistance	LshCas13a	*Agrobacterium*	Desiree	Multiple PVY strain–resistant mutants	Zhan et al. (2019)
StDND1, *StCHL1* and *StDMR6*-1 (S-genes: Susceptibility genes)	Late blight resistance	Cas9	*Agrobacterium*	Desiree	Tetra-allelic mutants by knockout of *StDMR6-1* and *StCHL1* genes	Kieu et al. (2021)
Caffeoyl-CoA O-methyltransferase (*StCCoAOMT*)	Late blight resistance	Cas9	*Agrobacterium*	Russet Burbank	Increase in late blight resistance than control	Hegde et al. (2021)
Abiotic stress tolerance						
StMYB44 (MYB transcription factor)	Phosphate transport (roots)	Cas9	*Agrobacterium*	Desiree	Mutants (84%), *StMYB44* negatively regulates Pi transport by suppressing *StPHO1* gene expression	Zhou et al. (2017)
Tuber quality traits						
GBBS	Starch quality	Cas9	Protoplasts (PEG)	Kuras	Multiple allele mutants (67%) and amylopectin-rich and waxy potato	Andersson et al. (2017)
GBBS	Starch quality	Cas9/RNP	Protoplasts	Kuras	Regenerants without transgenes (9%)	Andersson et al. (2018)
GBBS	Starch quality	Cas9	Protoplasts	Desiree and Wotan	Mutants (35%)	Johansen et al. (2019)
GBBS I	Starch quality	Cas9	*Agrobacterium*	Sayaka	Mutants with all four alleles (25%), low amylose starch	Kusano et al. (2018)
GBBS I	Starch quality	Cas9	*Agrobacterium*	Desiree	Tetra-allelic mutants by knockout of amylose-producing *StGBSSI* gene	Veillet et al. (2019a)
Starch synthase gene (*StSS6*)	Starch biosynthesis	Cas9	*Agrobacterium*	Desiree	Specific gRNA design and successful knock-out *SS6*	Sevestre et al. (2020)
Starch-branching enzymes (SBEs) genes *SBE1*, *SBE2*	Starch quality	Cas9	*Agrobacterium* and protoplasts (PEG)	Desiree	Mutants with valuable starch properties	Tuncel et al. (2019)
SBE1, *SBE2*	Starch quality	Cas9/RNP	Protoplasts	Desiree	Three to four allele mutants (72%) with amylase starch with no branching	Zhao et al. (2021)
PHYTOENE desaturase (*PDS*)	Carotenoid biosynthesis	Cas9	*Agrobacterium*	Desiree	Mutants (2–10%)	Bánfalvi et al. (2020)
StPDS	Carotenoid biosynthesis	Cas9	*Agrobacterium rhizogenes*	Diploid, self-compatible F_1 hybrid DMF1 (DM1-3 × M6)	Transgenic hairy roots mutants (64–98%)	Butler et al. (2020)
PDS and *coilin*	Carotenoid biosynthesis	Cas9	*In vitro* study without delivery	Chicago	Stimulated activity *in vitro*	Khromov et al. (2018)
St16DOX	Glycoalkaloids	Cas9	*A. rhizogenes* (electroporation)	May Queen	Full knockout of steroidal glycoalkaloids	Nakayasu et al. (2018)
Sterol side chain reductase 2 (*StSSR2*)	Steroidal glycoalkaloids (SGAs)	Cas9	*Agrobacterium*	Atlantic	Mutants (64%) with significantly reduced SGAs	Zheng et al. (2021)
Polyphenol oxidases (PPOs) gene (*StPPO2*)	Enzymatic browning	Cas9/RNP	Protoplasts	Desiree	Mutants (69% in four alleles) with 73% reduction in PPO activity than the control	González et al. (2020)
Other traits						
StDMR6-1 and *StGBSSI*	Phenotype	Cas9	*Agrobacterium*	Desiree	SpCas9-NG application in genome editing	Veillet et al. (2020a)
StIAA2	Phenotype	Cas9	*Agrobacterium*		Mono- and bi-allelic homozygous mutants (83%)	Wang et al. (2015)

(Continued on following page)

TABLE 1 | (*Continued*) Successful examples of application of CRISPR/Cas genome editing technology for biotic and abiotic stress resistance/tolerance, tuber quality, and phenotype and other traits in potato.

Target gene	Trait	CRISPR system	Delivery/ transformation system	Genotype	Key findings	Reference
				S. tuberosum Gp Phureja double monoploid		
Acetolactate synthase1 (*StALS1*)	Herbicide tolerance	Cas9	*Agrobacterium* and Geminivirus replicon (GVR)	Desiree, diploid (MSX914-10)	Targeted mutants (87–100%)	Butler et al. (2015)
StALS	Herbicide tolerance	Cas9	*Agrobacterium* and GVR	Desiree, diploid (MSX914-10)	Improved homozygous recombinants but no change in nonhomologous end joining	Butler et al. (2016)
StALS1 and *StALS2*	Herbicide tolerance	Cas9/CBE (cytidine base editing)	*Agrobacterium*	Desiree	Transgene-free mutants (10%)	Veillet et al. (2019b)
StALS1 and *StALS2*	Herbicide tolerance	Cas9/prime editing	*Agrobacterium*	Desiree	Successful prime editing in potato with nucleotide transition/transversion	Veillet et al. (2020b)
Stylar ribonuclease gene (*S-RNase*)	Self-incompatibility	Cas9	*Agrobacterium*	DRH-195 and DRH-310 F1	Stable self-compatible mutants through *S-RNase* gene knockout	Enciso-Rodriguez et al. (2019)
S-RNase	Self-incompatibility	Cas9	*Agrobacterium*	*S. tuberosum* Gp Phureja S15-65	Knock out of *S-RNase* gene resulted in self-compatibility	Ye et al. (2018)
NbFT, *NbPDS3*, and *NbXT2B*	Virus-induced genome editing (VIGE)	Cas9	*Agrobacterium*	Solanaceous plants	Heritable mutants expressing multiple sgRNAs in *Nicotiana benthamiana*/potato	Uranga et al., 2021

GBBS, Granule-bound starch synthase gene*; PEG, polyethylene glycol; RNP, Ribonucleo protein.*

CRISPR/CAS GENOME EDITING AND ITS NEED IN POTATO

CRISPR/Cas is the most powerful biological tool to create targeted modification in the genome, which allows easy designing and construction of gene-specific single guide RNA (sgRNA). The sgRNA vectors are easily reprogrammable to direct *Streptococcus pyogenes* Cas9 (SpCas9) to generate DSBs and are then repaired endogenously by the error-prone NHEJ or HR pathways. CRISPR/Cas is divided into two distinct classes based on the sequence, structure, and functions of the Cas proteins. Class 1 consists of types I, III, and IV andutilizes a multi-protein effector complex, whereas class 2 includes types II, V, and VI and uses a single effector protein; of which, type II and V target DNA, whereas type VI targets RNA. CRISPR/Cas9 (class 2, type-II) is the most commonly exploited machinery for DNA target. Remarkable innovations in CRISPR/Cas9 variant FnCas9 (*Francisella novicida*) (Price et al., 2015) and CRISPR/Cas13a (type VI, LshCas13a from *Leptotrichia shahii*) (Aman et al., 2018) have opened new avenues for RNA targets also. The SpCas9 and RNase III ribonucleases generate the Cas9/guide RNA complex that recognizes and cleaves DNA sequences adjacent to the 5′-NGG protospacer adjacent motif (PAM) and induces site-specific DSBs (Khatodia et al., 2016; Cao et al., 2020). Currently, CRISPAR/Cas9 has revolutionized plant research due to its simplicity, multiplexing, cost-effectiveness, high efficiency, and minimum off targets. Unlike genetically modified organisms, CRISPR/Cas creates alterations in the existing genome without the introduction of foreign genes, particularly site-directed nucleases (SDN1 and SDN2). Hence, CRISPR/Cas is expected to be transgene free, and biosafety regulations are under consideration in various countries (Schmidt et al., 2020).

Several complex traits of agronomic importance are considered in potato while breeding a new variety. The multigenic-controlled biotic/abiotic stresses are difficult to improve through conventional breeding in less time, which could be possible by using CRISPR/Cas9. The gene knockout mechanism has been applied in potato for late blight resistance using susceptibility (S) genes (*StDND1*, *StCHL1*, and *StDMR6*-1) (Kieu et al., 2021). A few successful examples are discussed later for biotic/abiotic stress resistance/tolerance, tuber quality, and phenotype traits improvement in potato (**Table 1**, and **Supplementary Figures 1 and 2**).

APPLICATION OF CRISPR/CAS IN POTATO

Biotic and Abiotic Stress Resistance/ Tolerance Traits

CRISPR/Cas has emerged as an alternative and efficient technology to accelerate potato breeding (**Table 1**). It has been demonstrated for potato virus Y (PVY) and late blight (*Phytophthora infestans*) resistance in potato. Cas13a protein was deployed to confer resistance to three PVY strains (RNA virus) by targeting *P3*, *CI*, *Nib*, and *CP* viral genes (Zhan et al., 2019). Host genes like the eukaryotic translation initiation factor

eIF4E and *coilin* have also been found very effective for PVY resistance (Makhotenko et al., 2019). Recently, late blight resistance was demonstrated in potato by the knockout of susceptibility genes *StDMR6-1* and *StCHL1* (Kieu et al., 2021) and *Caffeoyl-CoA O-methyltransferase* (*StCCoAOMT*) (Hegde et al., 2021).

Abiotic stresses such as heat, drought, salinity, and cold are very important in potato, but with meagre work that is available in potato so far. Zhou et al. (2017) developed mutants (84%) by manipulating potato MYB transcription factor gene *StMYB44*, which negatively regulates phosphate transport in potato by suppressing *StPHO1* gene expression (**Table 1**). Considerable research work on abiotic stress has been reported in cereals and other crops, but not in potato. Recently, we have proposed the use of CRISPR/Cas to manipulate N metabolism genes for improving nitrogen use efficiency in potato (Tiwari et al., 2020).

Tuber Quality, Phenotype, and Other Traits

CRISPR/Cas studies have been reported in potato for traits like improved tuber starch quality (Andersson et al., 2017, 2018; Kusano et al., 2018; Johansen et al., 2019; Tuncel et al., 2019; Veillet et al., 2019a; Sevestre et al., 2020; Zhao et al., 2021), carotenoid biosynthesis (Khromov et al., 2018; Bánfalvi et al., 2020; Butler et al., 2020), glycoalkaloids (Nakayasu et al., 2018; Zheng et al., 2021), and enzymatic browning (González et al., 2020) (**Table 1**). Functional mutants were developed for variations in phenotype (Wang et al., 2015; Veillet et al., 2020a) and herbicide tolerance (Butler et al., 2015, 2016). Self-compatible regenerants were also produced using Cas9 via *Agrobacterium* (Ye et al., 2018; Enciso-Rodriguez et al., 2019) or virus-induced genome editing (VIGE) (Uranga et al., 2021a; 2021b). Researchers have demonstrated the utility of Cas9 base editing and prime editing tools for herbicide tolerance in potato (Veillet et al., 2019b; 2020a; 2020b; 2020c).

CRISPR/CAS DELIVERY AND TRANSFORMATION SYSTEM AND CHALLENGES IN TETRAPLOID POTATO

Because potato is a highly amenable crop to tissue culture, transformation methods such as *Agrobacterium*, particle bombardment or biolistic, floral-dip, and protoplasts have been applied to it (Sandhya et al., 2020). The most common *Agrobacterium*-mediated transformation and protoplasts that have been successfully deployed in CRISPR/Cas in potato are sgRNA dicot-origin promoters like *Arabidopsis* (*AtUp*)/potato (*StU6p*)/U3p and plant promoters like *CaMV 35S* (Belhaj et al., 2013). However, the *Agrobacterium*-mediated method cannot be used to deliver ribonucleoprotein (RNP) complexes, and elimination of the Cas9 assembly from the plant genome via selfing or backcrossing is more complicated in genetically complex and vegetatively propagated tetraploid potato (Koltun et al., 2018). In potato, each botanical seed called True Potato Seed (TPS), which is a product of the meiosis process, is genetically different from another seed, hence the maintenance of the clonal identity is very crucial.

To address the above issues, the DNA-free delivery system is an ideal approach using somatic cells, i.e. protoplast. Polyethylene glycol (PEG)–mediated protoplast transformation has been found to be an excellent alternative for the efficient delivery of Cas9/gRNA-RNPs in potato (Andersson et al., 2017). DNA-free preassembled Cas9/gRNA-RNPs were directly delivered into the plant cells to induce mutations (Park and Choe, 2019) and were also demonstrated in lipofection-mediated DNA-free delivery (Liu et al., 2020). But with the establishment of suspension culture, protoplast isolation and regeneration into whole plants are the associated problems of the protoplast system (Sandhya et al., 2020).

VIGE is an emerging approach for CRISPR/Cas9 delivery. VIGE involving plant virus–derived vector such as geminivirus replicon has been demonstrated for fast and efficient delivery of sgRNAs in potato (Butler et al., 2015, 2016). This VIGE system bypasses the requirement of transformation and regeneration of plants which is a time-consuming and tedious process. But the large size of a Cas9 assembly challenges the use of the virus vector, as the length of a foreign insert negatively correlates with the stability of the vector.

Recently, base editing and prime editing are the upgraded and more efficient approaches of Cas9. The programmable base editing technology, like the adenine base editor that coverts A.T to G.C without DNA cleavage, has emerged as a boon for crop improvement (Gaudelli et al., 2017). Catalytically inactive Cas9 variant dCas9 or Cas9-nickase is fused with cytosine or adenosine deaminase domain to introduce the desired point mutations (C to T or A to G) in the target region (Mishra et al., 2020). Veillet et al. (2020c) deployed *Staphylococcus aureus*–cytosine base editor (CRISPR-SaCas9 CBE) to edit *StDMR6-1* in potato. Similarly, herbicide tolerance genes *Acetolactate synthase1* and *Acetolactate synthase2* (*StALS2*) were targeted through Cas9 cytidine base editing and Cas9 prime editing technologies, respectively (Veillet et al., 2019b; 2020b). Ariga et al. (2020) used the potato virus X vector to express a base editor consisting of modified Cas9 fused with cytidine deaminase to introduce the targeted nucleotide substitution in *Nicotiana benthamiana*. However, the size of the base editor is larger than Cas9 and this hindered the delivery into cells by the viral vectors.

Overall, high heterozygosity, tetrasomic inheritance, severe inbreeding depression, and vegetative propagation caused difficulties in the successful application of CRISPR/Cas in tetraploid potato. Furthermore, the selection of suitable sgRNA, robust CRISPR/Cas, and efficient transformation protocols and phenotypes without off targets are the main decisive factors in potato. Currently, gene knockout is a preferred mechanism in plants and even all four alleles were mutated through Cas9 in potato *StGBSS* gene (Andersson et al., 2017). PAM limitation (NGG) is one of the drawbacks of SpCas9, and therefore more diversity in CRISPR/Cas toolbox is necessary (Veillet et al., 2020a).

CONCLUSIONS

Desirable plant phenotypes, biotic/abiotic stress resistance/tolerance, and improved tuber quality traits play key roles in potato. The availability of robust CRISPR/Cas arrays, target genes selection, efficient plant transformation protocols, and minimum off-target mutants are the major issues in tetraploid potato. It is a fact that improvement of multigenic traits is difficult than that of the

monogenic traits, particularly in potato, due to polyploidy and clonal propagation. Despite this, considerable success has been achieved in potato for some traits and mostly through the gene knockout or insertion/deletion process. Studies have suggested that the use of multiplexing SpCas9 that can handle single or multiple sgRNA/RNPs via targeting conserved sequences combined with protoplast-mediated transformation is an ideal option in potato. Apart from this, awareness among people and policy makers/regulators would be necessary for the success of genome editing research. Collectively, CRISPR-Cas provides an effective next-generation toolbox for fast potato breeding to achieve sustainable crop yield.

AUTHOR CONTRIBUTIONS

JT conceived the idea and wrote the manuscript. All authors performed research, literature reviewing, and editing, and approved the manuscript for publication.

FUNDING

This work is funded by the Biotechnology programme by ICAR-CPRI, Shimla, CABin Scheme, and ICAR-LBS Outstanding Young Scientist Award project to JT.

ACKNOWLEDGMENTS

The authors thank ICAR-CPRI, CABin Scheme, and ICAR-LBS Outstanding Young Scientist Award project for support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.827808/full#supplementary-material

REFERENCES

Aman, R., Ali, Z., Butt, H., Mahas, A., Aljedaani, F., Khan, M. Z., et al. (2018). RNA Virus Interference via CRISPR/Cas13a System in Plants. *Genome Biol.* 19, 1. doi:10.1186/s13059-017-1381-1

Andersson, M., Turesson, H., Nicolia, A., Fält, A. S., Samuelsson, M., and Hofvander, P. (2017). Efficient Targeted Multiallelic Mutagenesis in Tetraploid Potato (*Solanum tuberosum*) by Transient CRISPR-Cas9 Expression in Protoplasts. *Plant Cel Rep* 36, 117–128. doi:10.1007/s00299-016-2062-3

Andersson, M., Turesson, H., Olsson, N., Fält, A. S., Ohlsson, P., Gonzalez, M. N., et al. (2018). Genome Editing in Potato via CRISPR-Cas9 Ribonucleoprotein Delivery. *Physiol. Plant* 164 (4), 378–384. doi:10.1111/ppl.12731

Ariga, H., Toki, S., and Ishibashi, K. (2020). Potato Virus X Vector-Mediated DNA-free Genome Editing in Plants. *Plant Cel Physiol* 61 (11), 1946–1953. doi:10.1093/pcp/pcaa123

Bánfalvi, Z., Csákvári, E., Villányi, V., and Kondrák, M. (2020). Generation of Transgene-free PDS Mutants in Potato by Agrobacterium-Mediated Transformation. *BMC Biotechnol.* 20 (1), 25. doi:10.1186/s12896-020-00621-2

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., and Nekrasov, V. (2013). Plant Genome Editing Made Easy: Targeted Mutagenesis in Model and Crop Plants Using the CRISPR/Cas System. *Plant Methods* 9, 39–47. doi:10.1186/1746-4811-9-39

Butler, N. M., Atkins, P. A., Voytas, D. F., and Douches, D. S. (2015). Generation and Inheritance of Targeted Mutations in Potato (*Solanum tuberosum* L.) Using the CRISPR/Cas System. *PLoS One* 10, e0144591. doi:10.1371/journal.pone.0144591

Butler, N. M., Baltes, N. J., Voytas, D. F., and Douches, D. S. (2016). Geminivirus-mediated Genome Editing in Potato (*Solanum tuberosum* L) Using Sequence-specific Nucleases. *Front. Plant Sci.* 7, 1045. doi:10.3389/fpls.2016.01045

Butler, N. M., Jansky, S. H., and Jiang, J. (2020). First-generation Genome Editing in Potato Using Hairy Root Transformation. *Plant Biotechnol. J.* 18 (11), 2201–2209. doi:10.1111/pbi.13376

Cao, Y., Zhou, H., Zhou, X., and Li, F. (2020). Control of Plant Viruses by CRISPR/Cas System-Mediated Adaptive Immunity. *Front. Microbiol.* 11, 593700. doi:10.3389/fmicb.2020.593700

Dangol, S. D., Barakate, A., Stephens, J., Çalıskan, M. E., and Bakhsh, A. (2019). Genome Editing of Potato Using CRISPR Technologies: Current Development and Future Prospective. *Plant Cel Tissue Organ. Cult* 139, 403–416. doi:10.1007/s11240-019-01662-y

Enciso-Rodriguez, F., Manrique-Carpintero, N. C., Nadakuduti, S. S., Buell, C. R., Zarka, D., and Douches, D. (2019). Overcoming Self-Incompatibility in Diploid Potato Using CRISPR-Cas9. *Front. Plant Sci.* 10, 376. doi:10.3389/fpls.2019.00376

FAOSTAT (2018). Available at: https://www.fao.org/faostat/en/ Accessed: November 20, 2021.

Gaudelli, N. M., Komor, A. C., Rees, H., Packer, M. S., Badran, A. H., Bryson, D. I., et al. (2017). Programmable Base Editing of A.T to G.C in Genomic DNA without DNA Cleavage. *Nature* 551, 464–471. doi:10.1038/nature24644

González, M. N., Massa, G. A., Andersson, M., Turesson, H., Olsson, N., Fält, A. S., et al. (2020). Reduced Enzymatic browning in Potato Tubers by Specific Editing of a Polyphenol Oxidase Gene via Ribonucleoprotein Complexes Delivery of the CRISPR/Cas9 System. *Front. Plant Sci.* 10, 1649. doi:10.3389/fpls.2019.01649

Hameed, A., Mehmood, M. A., Shahid, M., Fatma, S., Khan, A., and Ali, S. (2020). Prospects for Potato Genome Editing to Engineer Resistance against Viruses and Cold-Induced Sweetening. *GM Crops Food* 11 (4), 185–205. doi:10.1080/21645698.2019.1631115

Hameed, A., Zaidi, S. S., Shakir, S., and Mansoor, S. (2018). Applications of New Breeding Technologies for Potato Improvement. *Front. Plant Sci.* 9, 925. doi:10.3389/fpls.2018.00925

Hegde, N., Joshi, S., Soni, N., and Kushalappa, A. C. (2021). The Caffeoyl-CoA O-Methyltransferase Gene SNP Replacement in Russet Burbank Potato Variety Enhances Late Blight Resistance through Cell wall Reinforcement. *Plant Cel Rep* 40 (1), 237–254. doi:10.1007/s00299-020-02629-6

Hofvander, P., Andreasson, E., and Andersson, M. (2021). Potato Trait Development Going Fast-Forward with Genome Editing. *Trends Genet.* S0168-9525 (21), 00288–292. doi:10.1016/j.tig.2021.10.004

ICAR-CPRI Annual Report (2018). *Annual Report*. Shimla, Himachal Pradesh, India: ICAR-Central Potato Research Institute.

Johansen, I. E., Liu, Y., Jørgensen, B., Bennett, E. P., Andreasson, E., Nielsen, K. L., et al. (2019). High Efficacy Full Allelic CRISPR/Cas9 Gene Editing in Tetraploid Potato. *Sci. Rep.* 9, 17715. doi:10.1038/s41598-019-54126-w

Khatodia, S., Bhatotia, K., Passricha, N., Khurana, S. M., and Tuteja, N. (2016). The CRISPR/Cas Genome-Editing Tool: Application in Improvement of Crops. *Front. Plant Sci.* 7, 506. doi:10.3389/fpls.2016.00506

Khromov, A. V., Gushchin, V. A., Timberbaev, V. I., Kalinina, N. O., Taliansky, M. E., and Makarov, V. V. (2018). Guide RNA Design for CRISPR/Cas9-mediated Potato Genome Editing. *Dokl Biochem. Biophys.* 479 (1), 90–94. doi:10.1134/s1607672918020084

Kieu, N. P., Lenman, M., Wang, E. S., Petersen, B. L., and Andreasson, E. (2021). Mutations Introduced in Susceptibility Genes through CRISPR/Cas9 Genome Editing Confer Increased Late Blight Resistance in Potatoes. *Sci. Rep.* 11 (1), 4487. doi:10.1038/s41598-021-83972-w

Koltun, A., Corte, L. E. –D., Mertz-Henning, L. M., and Gonçalves, L. S. A. (2018). Genetic Improvement of Horticultural Crops Mediated by CRISPR/Cas: a New Horizon of Possibilities. *Hortic. Brasil.* 36, 290–298. doi:10.1590/s0102-053620180302

Kusano, H., Ohnuma, M., Mutsuro-Aoki, H., Asahi, T., Ichinosawa, D., Onodera, H., et al. (2018). Establishment of a Modified CRISPR/Cas9 System with Increased Mutagenesis Frequency Using the Translational Enhancer dMac3 and Multiple Guide RNAs in Potato. *Sci. Rep.* 8, 13753. doi:10.1038/s41598-018-32049-2

Liu, W., Rudis, M. R., Cheplick, M. H., Millwood, R. J., Yang, J.-P., Ondzighi-Assoume, C. A., et al. (2020). Lipofection-mediated Genome Editing Using DNA-free Delivery of the Cas9/gRNA Ribonucleoprotein into Plant Cells. *Plant Cel Rep* 39, 245–257. doi:10.1007/s00299-019-02488-w

Makhotenko, A. V., Khromov, A. V., Snigir, E. A., Makarova, S. S., Makarova, V. V., Suprunova, T. P., et al. (2019). Functional Analysis of Coilin in Virus Resistance and Stress Tolerance of Potato *Solanum tuberosum* Using CRISPR-Cas9 Editing. *Dokl Biochem. Biophys.* 484, 88–91. doi:10.1134/s1607672919010241

Mishra, R., Joshi, R. K., and Zhao, K. (2020). Base Editing in Crops: Current Advances, Limitations and Future Implications. *Plant Biotechnol. J.* 18, 20–31. doi:10.1111/pbi.13225

Nadakuduti, S. S., Buell, C. R., Voytas, D. F., Starker, C. G., and Douches, D. S. (2018). Genome Editing for Crop Improvement-Applications in Clonally Propagated Polyploids with a Focus on Potato (*Solanum tuberosum* L.). *Front. Plant Sci.* 9, 1607. doi:10.3389/fpls.2018.01607

Nakayasu, M., Akiyama, R., Lee, H. J., Osakabe, K., Osakabe, Y., Watanabe, B., et al. (2018). Generation of α-solanine-free Hairy Roots of Potato by CRISPR/Cas9 Mediated Genome Editing of the St16DOX Gene. *Plant Physiol. Biochem.* 131, 70–77. doi:10.1016/j.plaphy.2018.04.026

Park, J., and Choe, S. (2019). DNA-free Genome Editing with Preassembled CRISPR/Cas9 Ribonucleoproteins in Plants. *Transgenic Res.* 28 (Suppl. 2), 61–64. doi:10.1007/s11248-019-00136-3

Potato Genome Sequencing Consortium (2011). Genome Sequence and Analysis of the Tuber Crop Potato. *Nature* 475, 189–195. doi:10.1038/nature10158

Price, A. A., Sampson, T. R., Ratner, H. K., Grakoui, A., and Weiss, D. S. (2015). Cas9-mediated Targeting of Viral RNA in Eukaryotic Cells. *Procd. Natl. Acad. Sci. U. S. A.* 112, 6164–6169. doi:10.1073/pnas.1422340112

Rana, A., Dua, V. K., Chauhan, S., and Sharma, J. (2020). Climate Change and Potato Productivity in Punjab—Impacts and Adaptation. *Potato Res.* 63, 597–613. doi:10.1007/s11540-020-09460-2

Sandhya, D., Jogam, P., Allini, V. R., Abbagani, S., and Alok, A. (2020). The Present and Potential Future Methods for Delivering CRISPR/Cas9 Components in Plants. *J. Genet. Eng Biotechnol* 18, 25. doi:10.1186/s43141-020-00036-8

Schmidt, S. M., Belisle, M., and Frommer, W. B. (2020). The Evolving Landscape Around Genome Editing in Agriculture. *EMBO Rep.* 21, e50680. doi:10.15252/embr.202050680

Sevestre, F., Facon, M., Wattebled, F., and Szydlowski, N. (2020). Facilitating Gene Editing in Potato: a Single-Nucleotide Polymorphism (SNP) Map of the *Solanum tuberosum* L. Cv. Desiree Genome. *Sci. Rep.* 10 (1), 2045. doi:10.1038/s41598-020-58985-6

Tiwari, J. K., Buckseth, T., Singh, R. K., Kumar, M., and Kant, S. (2020). Prospects of Improving Nitrogen Use Efficiency in Potato: Lessons from Transgenics to Genome Editing Strategies in Plants. *Front. Plant Sci.* 11, 597481. doi:10.3389/fpls.2020.597481

Tuncel, A., Corbin, K. R., Ahn-Jarvis, J., Harris, S., Hawkins, E., Smedley, M. A., et al. (2019). Cas9-mediated Mutagenesis of Potato Starch-Branching Enzymes Generates a Range of Tuber Starch Phenotypes. *Plant Biotechnol. J.* 17 (12), 2259–2271. doi:10.1111/pbi.13137

United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019: Highlights (ST/ESA/SER.A/423).

Uranga, M., Aragonés, V., Selma, S., Vázquez-Vilar, M., Orzaez, D., and Daros, J.-A. (2021a). Efficient Cas9 Multiplex Editing Using Unspaced sgRNA Arrays Engineering in a Potato Virus X Vector. *Plant J.* 106, 555–565. doi:10.1111/tpj.15164

Uranga, M., Vazquez-Vilar, M., Orzáez, D., and Daròs, J. A. (2021b). CRISPR-Cas12a Genome Editing at the Whole-Plant Level Using Two Compatible RNA Virus Vectors. *CRISPR J.* 4 (5), 761–769. doi:10.1089/crispr.2021.0049

Veillet, F., Chauvin, L., Kermarrec, M. P., Sevestre, F., Merrer, M., Terret, Z., et al. (2019a). The *Solanum tuberosum GBSSI* Gene: a Target for Assessing Gene and Base Editing in Tetraploid Potato. *Plant Cel Rep* 38 (9), 1065–1080. doi:10.1007/s00299-019-02426-w

Veillet, F., Kermarrec, M. -P., Chauvin, L., Guyon-Debast, A., Chauvin, J. E., Gallois, J. –L., et al. (2020b). Prime Editing Is Achievable in the Tetraploid Potato, but Needs Improvement. *bioRxiv*. doi:10.1101/2020.06.18.159111

Veillet, F., Kermarrec, M. P., Chauvin, L., Chauvin, J. E., and Nogué, F. (2020a). CRISPR-induced Indels and Base Editing Using the *Staphylococcus aureus* Cas9 in Potato. *PLoS One* 15, e0235942. doi:10.1371/journal.pone.0235942

Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M. P., Guyon-Debast, A., Chauvin, J. E., et al. (2019b). Transgene-free Genome Editing in Tomato and Potato Plants Using Agrobacterium-Mediated Delivery of a CRISPR/Cas9 Cytidine Base Editor. *Inter. J. Mol. Sci.* 20 (2), 402. doi:10.3390/ijms20020402

Veillet, F., Perrot, L., Guyon-Debast, A., Kermarrec, M. P., Chauvin, L., Chauvin, J. E., et al. (2020c). Expanding the CRISPR Toolbox in *P. patens* Using SpCas9-NG Variant and Application for Gene and Base Editing in Solanaceae Crops. *Inter. J. Mol. Sci.* 21 (3), 1024. doi:10.3390/ijms21031024

Wang, S., Zhang, S., Wang, W., Xiong, X., Meng, F., and Cui, X. (2015). Efficient Targeted Mutagenesis in Potato by the CRISPR/Cas9 System. *Plant Cel Rep* 34, 1473–1476. doi:10.1007/s00299-015-1816-7

Ye, M., Peng, Z., Tang, D., Yang, Z., Li, D., Xu, Y., et al. (2018). Generation of Self-Compatible Diploid Potato by Knockout of. *S-RNaseNature Plant* 4 (9), 651–654. doi:10.1038/s41477-018-0218-6

Zhan, X., Zhang, F., Zhong, Z., Chen, R., Wang, Y., Chang, L., et al. (2019). Generation of Virus-Resistant Potato Plants by RNA Genome Targeting. *Plant Biotechnol. J.* 17, 1814–1822. doi:10.1111/pbi.13102

Zhao, X., Jayarathna, S., Turesson, H., Fält, A. S., Nestor, G., González, M. N., et al. (2021). Amylose Starch with No Detectable Branching Developed through DNA-free CRISPR-Cas9 Mediated Mutagenesis of Two Starch Branching Enzymes in Potato. *Sci. Rep.* 11 (1), 4311. doi:10.1038/s41598-021-83462-z

Zheng, Z., Ye, G., Zhou, Y., Pu, X., Su, W., and Wang, J. (2021). Editing Sterol Side Chain Reductase 2 Gene (*StSSR2*) via CRISPR/Cas9 Reduces the Total Steroidal Glycoalkaloids in Potato. *All Life* 14, 401–413. doi:10.1080/26895293.2021.1925358

Zhou, X., Zha, M., Huang, J., Li, L., Imran, M., and Zhang, C. (2017). StMYB44 Negatively Regulates Phosphate Transport by Suppressing Expression of PHOSPHATE1 in Potato. *J. Exp. Bot.* 68, 1265–1281. doi:10.1093/jxb/erx026

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Tiwari, Buckseth, Challam, Zinta, Bhatia, Dalamu, Naik, Poonia, Singh, Luthra, Kumar and Kumar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 12 April 2022
doi: 10.3389/fgene.2022.866504

Novel miRNA-SSRs for Improving Seed Hardness Trait of Pomegranate (*Punica granatum* L.)

Prakash Goudappa Patil[1]*, Nripendra Vikram Singh[1], Abhishek Bohra[2], Shivani Jamma[1], Manjunatha N[1], Venkatesh S. C[3], Dhinesh Babu Karuppannan[1], Jyotsana Sharma[1] and Rajiv A. Marathe[1]

[1]ICAR-National Research Centre on Pomegranate (NRCP), Solapur, India, [2]ICAR-Indian Institute of Pulses Research (IIPR), Kanpur, India, [3]Dept. of Biotechnology and Crop Improvement, University of Horticultural Sciences (UHS), Bagalkot, India

Present research discovered novel miRNA-SSRs for seed type trait from 761 potential precursor miRNA sequences of pomegranate. SSR mining and BLASTx of the unique sequences identified 69 non-coding pre-miRNA sequences, which were then searched for BLASTn homology against Dabenzi genome. Sixty three true pri-miRNA contigs encoding 213 pre-miRNAs were predicted. Analysis of the resulting sequences enabled discovery of SSRs within pri-miRNA (227) and pre-miRNA sequences (79). A total of 132 miRNA-SSRs were developed for seed type trait from 63 true pri-miRNAs, of which 46 were specific to pre-miRNAs. Through ePCR, 123 primers were validated and mapped on eight Tunisia chromosomes. Further, 80 SSRs producing specific amplicons were ePCR-confirmed on multiple genomes *i.e.* Dabenzi, Taishanhong, AG2017 and Tunisia, yielding a set of 63 polymorphic SSRs (polymorphism information content ≥0.5). Of these, 32 miRNA-SSRs revealed higher polymorphism level (89.29%) when assayed on six pomegranate genotypes. Furthermore, target prediction and network analysis suggested a possible association of miRNA-SSRs *i.e.* miRNA_SH_SSR69, miRNA_SH_SSR36, miRNA_SH_SSR103, miRNA_SH_SSR35 and miRNA_SH_SSR53 with seed type trait. These miRNA-SSRs would serve as important genomic resource for rapid and targeted improvement of seed type trait of pomegranate.

Keywords: functional markers, miRNA, pomegranate, SSR, seed type

OPEN ACCESS

Edited by:
Deepmala Sehgal, International Maize and Wheat Improvement Center, Mexico

Reviewed by:
Muntazir Mushtaq, National Bureau of Plant Genetic Resources (ICAR), India
Jagesh Kumar Tiwari, Central Potato Research Institute (ICAR), India

****Correspondence:***
Prakash Goudappa Patil
patilbt@gmail.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *31 January 2022*
Accepted: *01 March 2022*
Published: *12 April 2022*

Citation:
Patil PG, Singh NV, Bohra A, Jamma S, N M, C VS, Karuppannan DB, Sharma J and Marathe RA (2022) Novel miRNA-SSRs for Improving Seed Hardness Trait of Pomegranate (Punica granatum L.). Front. Genet. 13:866504. doi: 10.3389/fgene.2022.866504

INTRODUCTION

Pomegranate (*Punica granatum* L.) native to central Asia and is widely cultivated in tropical and subtropical regions. The most of the commercially cultivated pomegranate varieties in India are medium-to hard-seeded types. Therefore, breeding for the soft-seeded varieties is top priority for consumer point of view. Hard seeds of pomegranate are not preferred for consumption because they are too hard to chew and swallow (Xia et al., 2019). However, phytosterols and special fatty acid such as punicic acid found in pomegranate seeds offer a variety of health benefits. There is an urgent need

Abbreviations: miRNA, Micro RNA; ePCR, Electronic polymerase chain reaction; PIC, Polymorphism information content; QTL, Quantitative trait loci; UPGMA, Unweighted pair group method of arithmetic averages.

for breeding soft-seeded pomegranate cultivars that provide new products for the market and contribute to enhanced farmer incomes. However, the genetic architecture of seed type trait is largely unknown. Therefore, understanding the genetic mechanisms underlying seed type would facilitate the development of new commercially viable pomegranate varieties (Luo et al., 2018). There are three types of pomegranate varieties: soft-seeded (seed hardness<3.67 kg cm^{-2}), semi soft-seeded (seed hardness from 3.67 to 4.2 kg cm^{-2}), and hard-seeded (seed hardness>4.2 kg cm^{-2}) (Lu, 2006). Research has shown that the differences in the hard-seeded and soft-seeded cultivars could be attributed to the variation in expressions of genes or transcription factors *CCR, CAD, CelSy, SuSy, CCoA-OMT, MYB, WRKY and MYC* involved in lignin and cellulose biosynthesis (Xue et al., 2017). The divergence between hard- and soft-seeded pomegranates is the embodiment of germplasm diversity in pomegranate. Owing to the complexity of metabolic synthesis, underlying the degree of seed hardness of pomegranate is influenced by environmental factors and genetic background (Zhang et al., 2020). The softness of seeds is a desirable economic trait that enhances the consumptive qualities of fruits, but the complete soft-seeded pomegranate is restricted to a narrow ecological region and requires cold protection at low temperatures (Song et al., 2012). The quantitative nature of seed type may likely contain a regulator gene conferring tolerance to adverse environmental conditions (Lu, 2006). Deciphering ecological and evolutionary forces shaping the population structure of the soft-seeded pomegranate will help us understand the genetic makeup of the seed trait (Zhang et al., 2020). Investigations on the formation of the soft seed will elucidate the mechanism of lignin synthesis that contributes to plant growth and development and confers resistance to biotic and abiotic stresses (Liu et al., 2018).

Non-coding (nc) RNAs including miRNAs are known to regulate multiple aspects of plant growth and development and plant's response to a variety of stresses (Millar, 2020). High throughput sequencing has emerged as a promising tool to discover miRNAs and their gene targets at large scale in different plant species including pomegranate (Singh et al., 2017; Luo et al., 2018). Deep sequencing of small RNA libraries constructed from seeds at 60 and 120 DAF of soft-seeded and hard-seeded pomegranate enabled identification of miRNAs specific to seed type trait (Luo et al., 2018). The study suggested that a complex biological process mediated by miRNA-mRNA networks controls the seed type in pomegranate. The presence of genetic variations in pre-miRNAs and miRNAs are known to affect quantitative trait expression (Ferrao et al., 2015), hence DNA markers may be developed from the miRNA regions to assist the procedure of trait improvement (Kumar et al., 2017). Study by Joy et al. (2018) showed presence of SSRs in pre-miRNAs and the authors proposed a role for alternative splicing in creating mature RNA isoforms in response to stress. Therefore, earlier a large scale development of SSR markers from coding regions of miRNA sequences has been reported. For instance, Chen et al. (2010) performed a comprehensive analysis of the SSR prediction in 8,619 pre-miRNA sequences from 87 species from nine different taxonomic groups.

In pomegranate, recent advances such as whole genome sequencing (Akparov et al., 2017; Qin et al., 2017; Yuan et al., 2018; Luo et al., 2020), and transcriptome profiling (Ophir et al., 2014; Saminathan et al., 2016; Luo et al., 2018) have paved the way for large-scale discovery of functional molecular markers for use in genetic improvement programs (Mondal and Ganie, 2014). A previous research on genome wide discovery of miRNA-SSRs in pomegranate resulted in the identification of 897 and 168 SSR markers corresponding to pri-miRNAs and pre-miRNAs, respectively (Patil et al., 2020). In *Medicago truncatula* genome, Min et al. (2017) identified 189 miRNA-SSRs in pri-miRNA sequences extracted from 356 non-redundant (NR) miRNAs. A similar analysis in the *Arabidopsis* genome identified 147 miRNA-SSRs from 169 pre-miRNA transcripts (Kumar et al., 2017). Since, presence of SSRs in these miRNA coding regions creates enormous possibilities for the development of predictive DNA markers for important phenotypes regulated by miRNAs (Patil et al., 2020). Also, the availability of information on potentially novel miRNA candidates for seed type trait as reported by Luo et al. (2018). Here, we performed preliminary study to develop seed type specific miRNA-SSRs and target gene based EST-SSR markers for future genomic applications. Which can aid in discover of master miRNA through association analysis or candidate gene based genome editing application for genetic improvement of seed type trait in pomegranate.

MATERIALS AND METHODS

In silico Analysis and Identification of miRNA Coding Sequences

A total of 761 potential novel pre-miRNAs reported earlier by Luo et al. (2018) for seed type trait in pomegranate were retrieved. Firstly, all these sequences were searched for the presence of SSRs using MISA web tool (Beier et al., 2017; http://misaweb.ipk-gatersleben.de/misa/). We removed protein-coding sequences from pre-miRNA sequences through BLASTx against NR protein database (Altschul et al., 1997). Homology search was performed by using nc pre-miRNAs as query for BLASTn against Dabenzi genome (mismatch <1, with no gap and e-value ≤ 0.01) (Qin et al., 2017). The resulting contigs with flanking sequences of ~800 bp around each query sequences were retrieved. Contigs were then examined for pri-miRNA regions using sequence-structure motif base: pre-miRNA prediction webserver (http://www.regulatoryrna.org/webserver/SSMB/pre-miRNA/home.html). The contigs with no miRNA regions were excluded from the analysis. Finally, contigs with true pri-miRNAs and their pre-miRNA sequences were used for SSR survey.

Designing of SSRs Specific to Pri- and Pre-miRNAs

SSR motifs were searched by keeping a minimum repeat length of 12 bp and defining 12, 6, 4, 3, 3 and 2 for mono, di, tri, tetra, penta

and hexa nucleotide, respectively. The two SSRs interrupted within 100 bases were defined as compound SSRs. MISA statistics as obtained was used to draw frequency distribution graphs using Microsoft Excel. Batch Primer three v1.0 (https://wheat.pw.usda.gov/demos/BatchPrimer3) was used to design SSRs present in pri- and pre-miRNA sequences. All SSR primer pairs were designed to generate 100–400 bp amplicons with other specific parameters like: primer length (bp) 18–20 bp with 19 bp as optimum; GC content (%) 40–60, with the optimum value being 50% and Tm (°C) 52–60, with 55 as the optimum were used. For designing of EST-SSR primers Krait software was used with default parameters.

ePCR Validation and Localization of miRNA-SSRs on Chromosomes

To check the amplification efficiency and locate the newly designed miRNA-SSRs on Tunisia genome, ePCR was performed using GMATA (Genome-wide Microsatellite Analyzing Tool Package) software (Wang and Wang, 2016) using algorithm (Schuler, 1997). A physical map with the SSR loci was constructed using MapChart v 2.2 software (Voorrips, 2002). MiRNA-SSR primers producing one to two alleles in Tunisia genome were then evaluated in three draft genome assemblies (Dabenzi, Taishanhong and AG 2017) through ePCR. The amplicon sizes were recorded for 80 miRNA-SSRs across the four pomegranate genomes using GMATA. The marker parameters were computed using GenAlEx v.6.5 (Peakall and Smouse, 2012).

PCR Amplification and Diversity Analysis

Genomic DNA was extracted from the fresh leaf samples of 16 pomegranate genotypes (**Supplementary Table S1**) following the modified CTAB method (Ravishankar et al., 2000). For PCR, 32 pre-miRNA-SSR primers was initially synthesised and screened on a subset of six pomegranate genotypes *i.e.* Ganesh, Mridula, Jyoti, Yercaud, Kalpitiya and Co-white, following touch down PCR program with Prime-96™ Thermal Cycler (Himedia, India). Specific amplicons were confirmed by separating fragments on 3% metaphor gels. Subsequently, resulting subset of 10 informative miRNA-SSRs were selected for genetic diversity study in 16 pomegranate genotypes. For PCR experiments amplification was carried out in 10 µL reaction volume containing 1.0 µL of 10X PCR buffer, 1 µL (1 mM dNTP mix), 0.5 µL each of forward and reverse primers (10 pmol), 0.2 µL of *Taq* DNA polymerase 5U/µl (Himedia, India) and 1 µL (10 ng) of template DNA. Touchdown PCR was performed with the following conditions: 94°C for 5 min, followed by 16 cycles of 94°C for 30 s, decrease 0.2°C/cycle from 60°C for 30 s, 72°C for 45 s; followed by 25 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 45 s and a final extension at 72°C for 5 min. The amplified fragments were resolved on 3% metaphor agarose gels accompanied by visualization and documentation using a gel documentation system (Vilbert Dourmet, France). We computed genetic diversity parameters from the SSR genotyping data using GenAlEx v. 6.5 (Peakall and Smouse, 2012). The genetic cluster analysis was performed with the UPGMA (Unweighted pair group method with an arithmetic mean) method of NTSYS-pc v. 2.11 (Rohlf, 2000).

TABLE 1 | Characterization of SSRs in pri- and pre-miRNA sequences of pomegranate genome.

Parameters	Pri-miRNA	Pre-miRNA
Number of sequences examined	63	213
Examined sequences size (bp)	81,788	22,214
Total number of identified SSRs	227	79
Number of sequences with SSRs	60	65
Number of sequences with more than 1 SSRs	55	12
Number of compound SSRs	67	14

Note Pre-miRNAs: Precursor miRNAs; Pri-miRNAs: Primary miRNAs.*

Prediction of Potential Gene Targets for miRNAs

To elucidate the biological roles of selected pre-miRNAs, first homology search was performed against miRBase to determine their miRNA families. Then, target analysis for selected mature miRNAs was performed against 29,854 annotated gene models (mRNAs) reported for Tunisia genome (Luo et al., 2020). Prior to target analysis, CD-HIT tool with default parameters was used to reduce the redundancy (Li and Godzik, 2006), resulting in 21,877 unique sequences. Target analysis of mature miRNAs of 44 pre-miRNAs against 21,877 Tunisia gene models was performed using TAPIR (http://bioinformatics.psb.ugent.be/webtools/tapir; Bonnet et al., 2010). Blast2GO 6.0 (https://www.blast2go.com/blast2go-pro) was used to perform functional annotations of target genes for Gene Ontology (GO) and KEGG enrichment analysis, and Web Gene Ontology Annotation Plot was drawn using WEGO 2.07 (Ye et al., 2018). Further, based on the TAPIR alignment duplex target score (with cut off value ≤ 0.4) and duplex free energy ratio of hybridization (with cut off value ≥ 0.7) between miRNA/mRNA the regulatory networks were built using Gephi 0.9.2 software (Bastian et al., 2009). TAPIR target score considers the number of mismatches, gaps (introduced by bulges and loop structures) and number of G-U pairs located in position 2nd and 12th of 5' seed regions of miRNAs to the target mRNA. Whereas the duplex miRNA-mRNA free energy ratio (Allen et al., 2005), is the ratio of the free energy of duplex to the free energy of the same duplex having only perfect matches.

RESULTS

Characterization of miRNA-SSRs for Seed Type Trait

SSR survey of 761 pre-miRNA sequences facilitated detection of 199 SSR motifs corresponding to 144 pre-miRNAs. These sequences were then used for BLASTX search to remove protein-coding sequences. Resulting set of 69 non-coding pre-miRNA served as query sequences in BLASTn search against

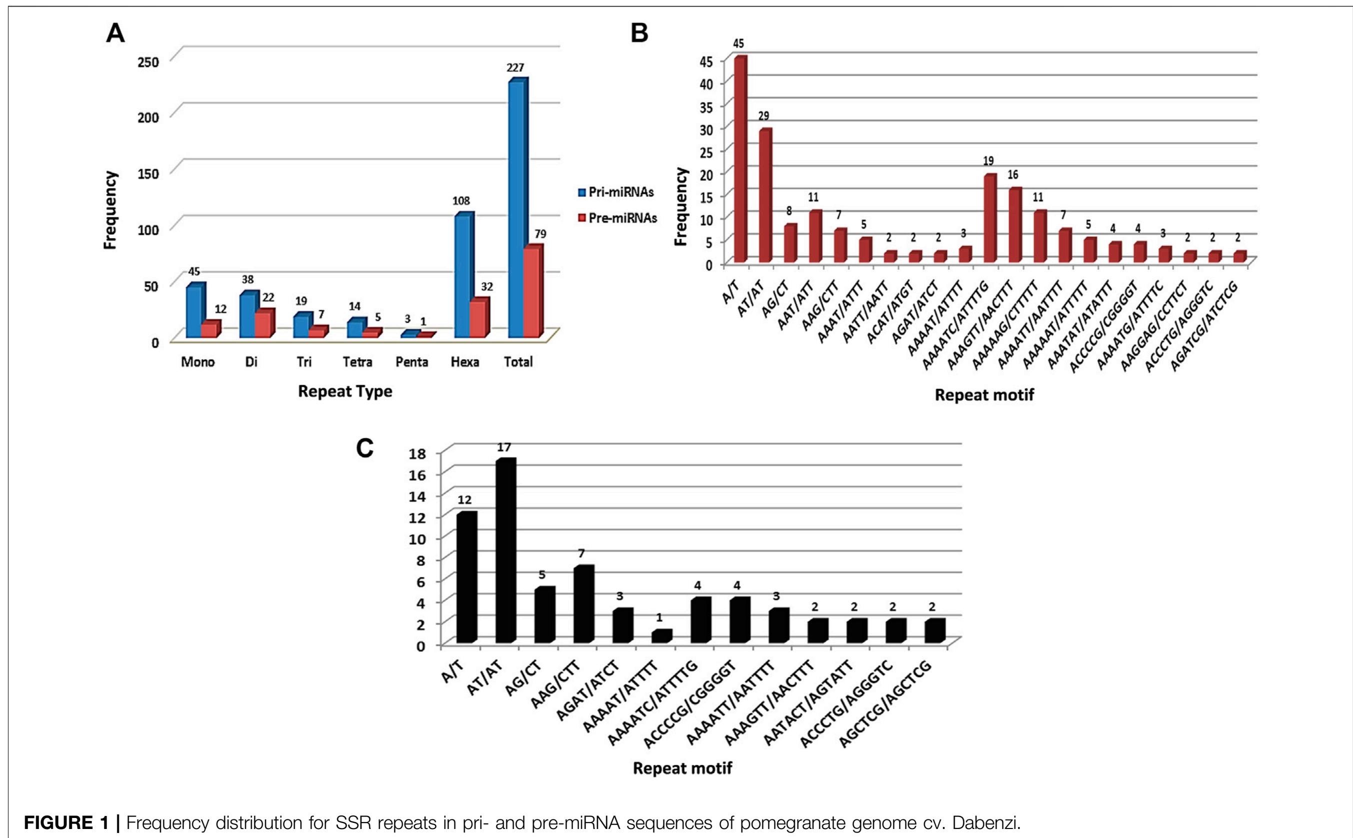


FIGURE 1 | Frequency distribution for SSR repeats in pri- and pre-miRNA sequences of pomegranate genome cv. Dabenzi.

pomegranate genome cv. Dabenzi. A total of 69 highly-homologous contigs with ~800bp flanking sequences around pre-miRNA complementarity regions were extracted. MiRNA prediction tool identified 63 contigs (~821) that harboured pri-miRNAs and encoding 213 pre-miRNAs. SSR survey of these sequences using MISA identified 227 SSR motifs specific to 60 (95.2%) pri-miRNA and 79 motifs to 65 (30.5%) pre-miRNA sequences (**Table 1**). The pri-miRNA and pre-miRNA sequences represented 81.7 Kb and 22.2 Kb of pomegranate genome, respectively. The distribution frequencies of one SSR locus per every 0.36 and 0.28 kb were observed for pri- and pre-miRNAs, respectively. Total 55 pri- (87.3%) and 12 pre-miRNAs (5.63%) showed more than one SSR motifs. Out of 227 SSR motifs specific to pri-miRNAs and 79 to pre-miRNAs, 67 (29.5%) and 14 (17.7%) respectively were of compound type. Concerning the abundance of SSR motifs, hexa-nucleotides were the most pronounced (47.58%) followed by mono- (19.82%) and di-nucleotides (16.74%) in pri-miRNAs (**Figure 1A**). Similar SSR distribution pattern was observed in pre-miRNAs. In addition to this, pri-miRNAs revealed more abundance of A/T (100%) repeats followed by AT/AT (76.31%), which is also witnessed in pre-miRNA sequences (**Figure 1B,C**).

Designing of miRNA-SSRs, ePCR Validation and Mapping on Chromosomes

We designed 132 primers for SSRs located within 60 pri-miRNA sequences. The details of primer pairs are given in (**Supplementary Table S2**). Out of 132 miRNA-SSR primers, 77 (58.33%) primers were designed targeting hexanucleotide repeats; followed by 30 (22.73%), 12 (9.09%), 10 (7.57%) and 3 (2.27%) for di, tri, tetra and penta nucleotide repeats, respectively. However, 46 (34.85%) primers were specific to pre-miRNA sequences.

To assess amplification efficiency, specificity and chromosome locations of the miRNA-SSRs, we performed ePCR or e-mapping of SSRs on the chromosomes of Tunisia genome. As a result, 123 of 132 SSRs were successfully got validated across eight chromosomes, producing alleles of single, two, three or more than three alleles in the Tunisia genome (**Table 2**). However, nine primers did not map to Tunisia chromosomes. Total 26 (21.14%) primers produced single amplicons, whereas 54 (43.90%) primers yielded two alleles and 43 primers had ≥3 alleles when assayed across the chromosomes.

Physical locations of miRNA-SSRs were visualized on eight chromosomes. A total of 123 markers were mapped onto individual chromosomes of Tunisia, of which Chm_2 (36 markers), Chm_6 (15) and Chm_7 (15) had the higher number of SSRs, followed by Chm_4 (13), Chm_1 (12), Chm_3 (11), Chm_8 (11) and Chm_5 (10) respectively. **Figure 2** illustrates the chromosome-wise locations (Mb) of all the miRNA-SSRs. The SSR map presented here would enable accurate choice of informative miRNA-SSRs based on their genomic positions (**Supplementary Table S3**).

TABLE 2 | Experimental validation of 123 miRNA_SSRs in Tunisia genome in comparison to other three draft genomes of pomegranate cultivars Dabenzi, Taishanhong and AG2017 through ePCR or eMapping.

	ePCR validation of 123 miRNA_ SSRprimers for genotyping applications																			
	Tunisia genome					Dabenzi genome					Taishanhong genome					AG2017 genome				
	Allele No					Allele No					Allele No					Allele No				
	one	two	three	>three	TNP	one	two	three	>three	TNP	One	two	three	>three	TNP	one	two	three	>three	TNP
Chm_1	1	6	5	—	12	1	7	4	—	12	1	6	5	—	12	1	4	6	—	11
Chm_2	7	16	13	—	36	5	15	16	—	36	3	19	13	—	35	4	12	14	2	32
Chm_3	2	4	5	—	11	1	5	4	1	11	0	6	5	—	11	1	4	2	3	10
Chm_4	2	8	3	—	13	1	7	5	—	13	4	6	3	—	13	1	6	4	1	12
Chm_5	2	5	3	—	10	1	5	4	—	10	1	6	3	—	10	0	4	5	—	9
Chm_6	5	5	3	2	15	3	7	5	—	15	3	4	5	3	15	4	7	1	—	12
Chm_7	5	6	4	—	15	0	9	6	—	15	4	7	4	—	15	3	6	4	—	13
Chm_8	2	4	5	—	11	2	3	5	1	11	1	5	4	—	10	1	5	3	2	11
Total	26	54	41	2	123	14	58	49	2	123	17	59	42	3	121	15	48	39	8	110

Note Chm: chromosomes; TNP: total number of primers.*

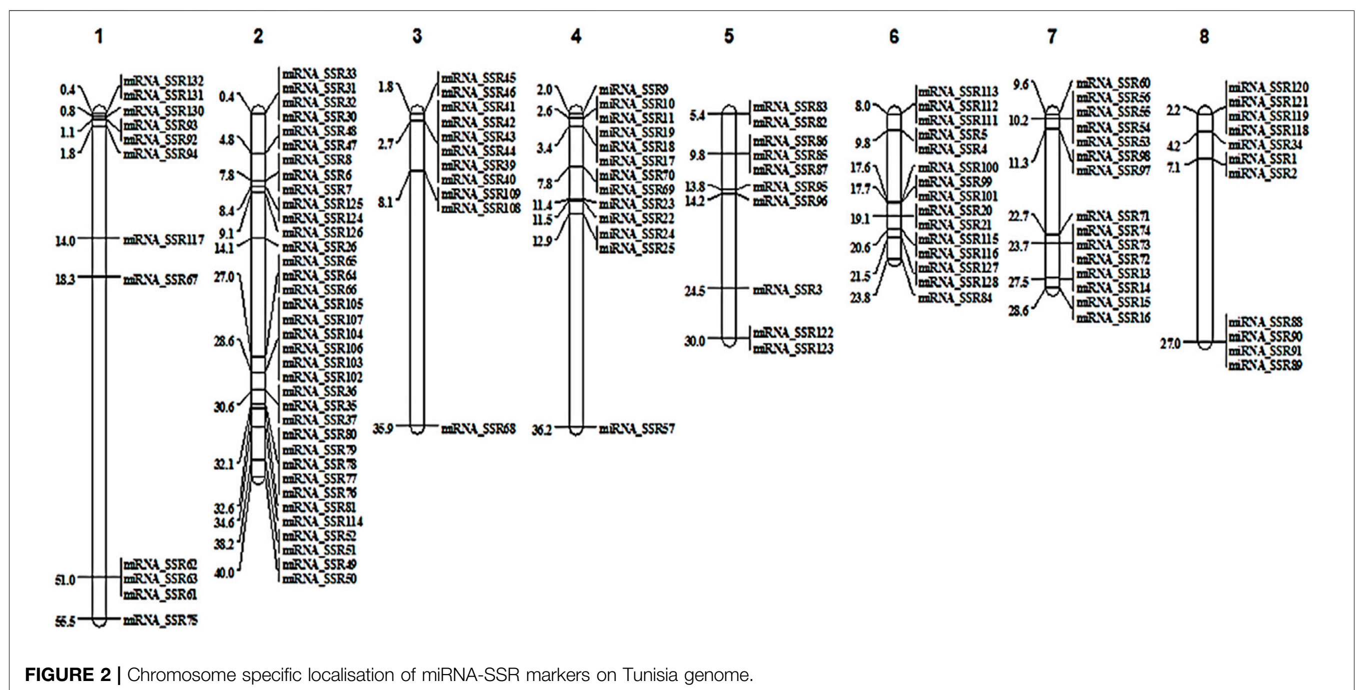


FIGURE 2 | Chromosome specific localisation of miRNA-SSR markers on Tunisia genome.

TABLE 3 | Chromosome specific marker statistics for 80 miRNA-SSR primers assayed through ePCR across the four pomegranate genotypes based on their genome sequences.

Chromosome	TNP	TPP	Na	MAF	Ne	*I*	*Ho*	*He*	PIC
Chm_1	7	7	19 (2.71)	0.55	2.27	0.87	0.89	0.54	0.63
Chm_2	23	23	65 (2.83)	0.55	2.30	0.89	0.78	0.54	0.62
Chm_3	6	6	15 (2.50)	0.59	2.07	0.79	0.76	0.51	0.58
Chm_4	10	9	24 (2.40)	0.60	2.08	0.74	0.80	0.47	0.54
Chm_5	7	7	18 (2.57)	0.57	2.20	0.84	0.79	0.53	0.61
Chm_6	10	8	22 (2.20)	0.66	1.82	0.62	0.63	0.40	0.45
Chm_7	11	11	30 (2.73)	0.61	2.03	0.80	0.73	0.49	0.56
Chm_8	6	6	20 (3.33)	0.52	2.83	1.03	0.83	0.58	0.67
Total/mean	80	77	213 (2.66)	0.58	2.20	0.82	0.78	0.51	0.58

Note Chm: Chromosome; TNP: total number of primers; TPP: total number of polymorphic primers; Na: Numbers of alleles; MAF: major allelic frequency; Ne: Number of Effective Alleles;* I*: Shannon's Information Index;* Ho*: Observed heterozygosity;* He*: Expected heterozygosity; PIC: polymorphic information content.*

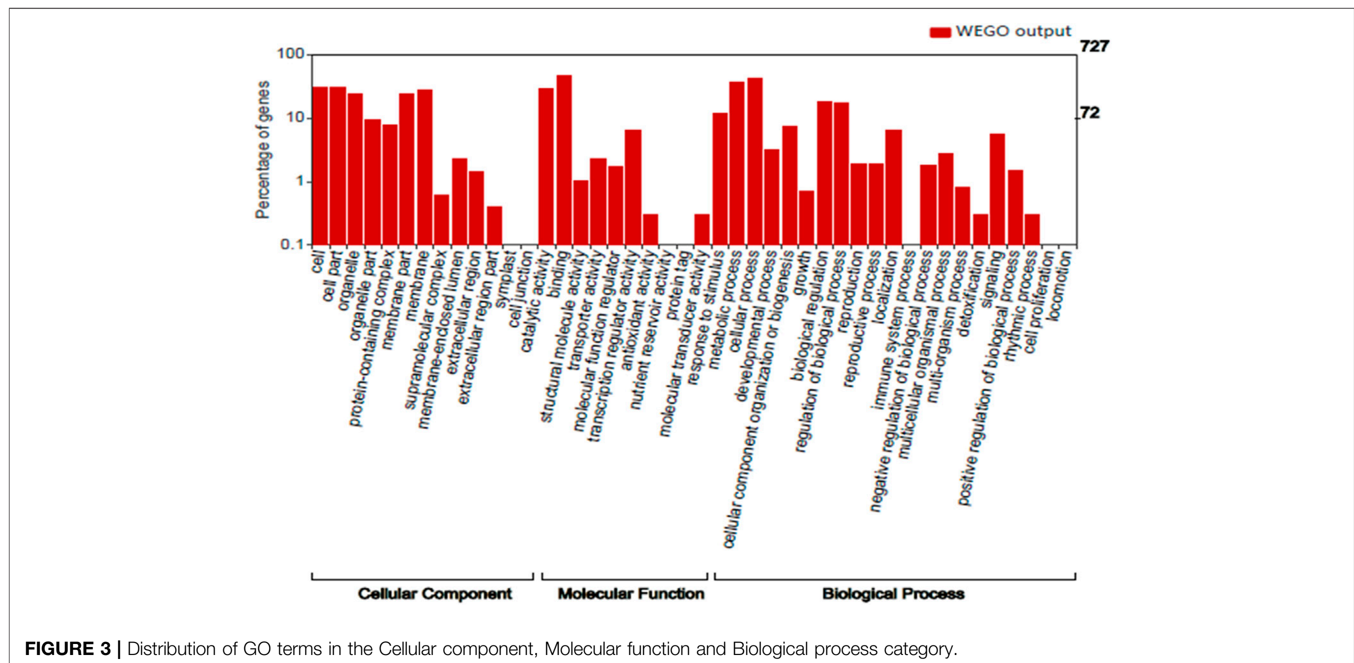


FIGURE 3 | Distribution of GO terms in the Cellular component, Molecular function and Biological process category.

Identification of Informative miRNA-SSRs Through ePCR Across the Pomegranate Genomes

We confirmed amplification of 123 miRNA-SSRs across three assemblies of pomegranate genome *i.e.* Dabenzi, Taishanhong and AG 2017. As a result, 80 SSRs were identified that produced one to two alleles across three genomes including Tunisia (**Table 3**). Further, the amplicons of these genotypes were recorded to compute various marker parameters (**Supplementary Table S4**). Of the total 80 SSRs, 77 (96.25%) showed polymorphism across four genotypes. The assay generated a total of 213 alleles spanning eight chromosomes. The Na per locus ranged from 2 to 6, with an average value of 2.66 alleles/loci. The MAF per locus varied between 0.25 and 0.88, with an average of 0.58. The PIC values ranged from 0.25 to 0.93, with an average of 0.58. In the present dataset, 63 SSRs had the PIC values ≥ 0.50, implying their highly informative nature. The average Shannon information index was 0.82 for the four genomes tested.

Functional Classification and Pathway Enrichment Analysis for miRNA Target Genes

To assign functional roles to the identified miRNA-SSRs, we carried out target analysis using 21,877 unique gene models from Tunisia genome. This resulted in identification of a total of 2,306 targets, of which 1935 were found unique targets and 371 as common targets (**Supplementary Table S5**). The predicted targets belonged to 24 miRNA families, with -MIR156b having the highest targets (655), followed by ath-MIR5655 (592), MIR5021_1 and 2 (522), ath-MIR5651_3 (394), ath-MIR157c (43) and ath-MIR5651_1 (39). Further, based on lowest target score (≤4) and highest minimum free energy (Mfe) ratio of hybridization (≥0.7), 2,306 targets were narrowed down to a set of 754 candidate genes, which were negatively regulated by 24 miRNA families. Target analysis led to the identification of five informative miRNA-SSRs *i.e.* miRNA_SH_SSR69, miRNA_SH_SSR36, miRNA_SH_SSR103, miRNA_SH_SSR35 and miRNA_SH_SSR53 influencing expression of multiple genes serving as transcription factors, enzymes and transporters involved in seed development and maturation to impart seed type (**Supplementary Table S6**).

Gene Ontology Analysis

We performed Gene Ontology for 727 top genes targeted by miRNAs to find the potential contributions of these genes during seed development and maturation. The target genes were grouped into three classes: biological process (20 GO terms), molecular function (11 GO terms), and cellular component (13 GO terms) (**Figure 3**). GO analysis showed most of the genes were found to be associated with biological process like metabolic process, cellular process, biological process regulations, cellular component organization or biogenesis, localization and signaling. With respect to cellular components, many genes are part of cell, organelle and parts, protein containing complex and membrane. However, with respect to molecular functions many genes have role of binding and catalytic activity, transcription regulator activity, transporter activity and response to stimulus. Therefore, the GO analysis clearly showed the role of miRNA-targeted genes in seed development and maturation.

The KEGG pathway enrichment analysis revealed 95 of 727 targets code for 101 enzymes involved in 58 pathways, with the highest representation from Starch and sucrose metabolism, Glycerophospholipid metabolism, Glycerolipid metabolism,

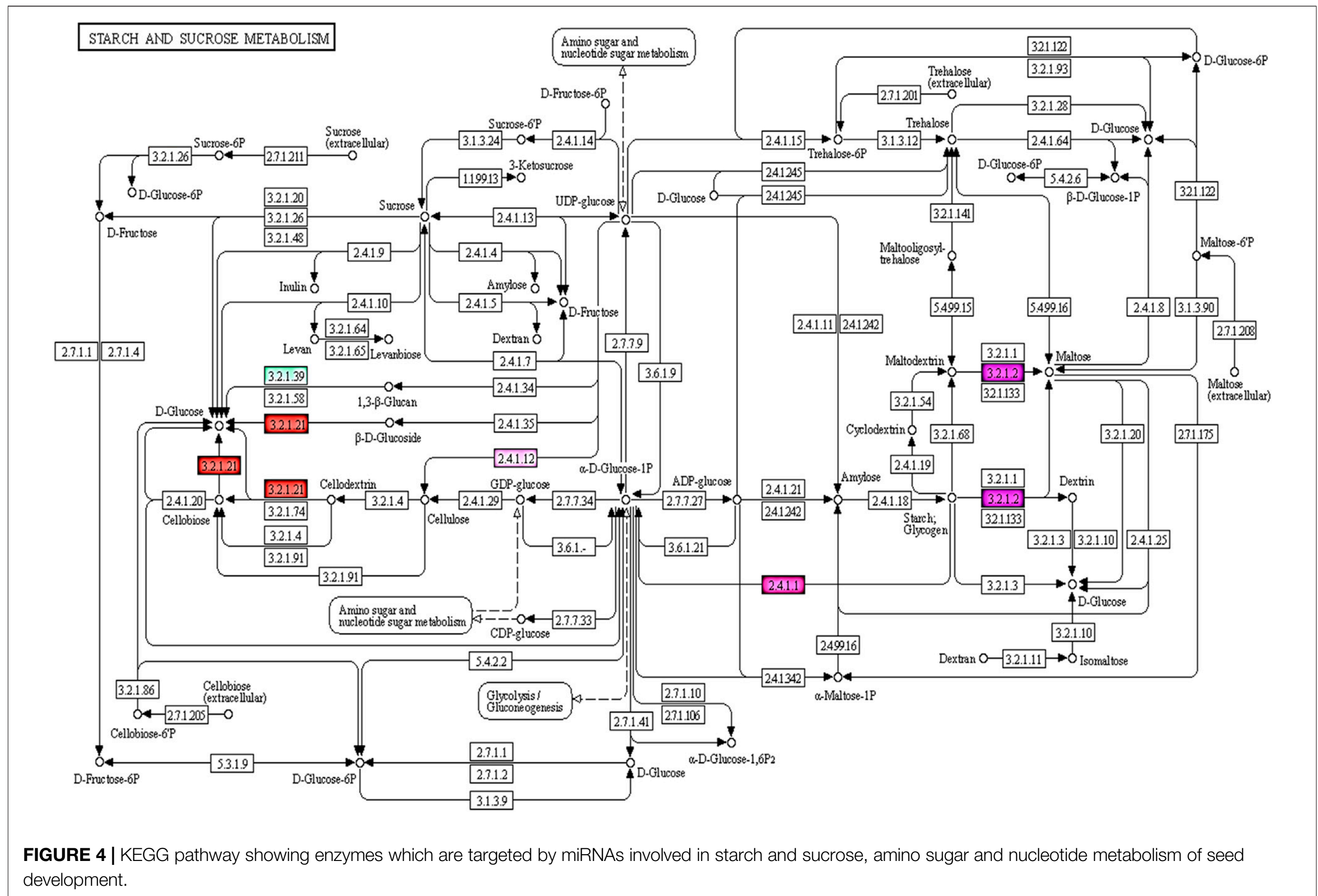


FIGURE 4 | KEGG pathway showing enzymes which are targeted by miRNAs involved in starch and sucrose, amino sugar and nucleotide metabolism of seed development.

Amino sugar and nucleotide sugar metabolism, Pentose and glucuronate interconversions and so forth (**Supplementary Table S7**). Among these pathways maximum of seven enzymes were part of starch and sucrose metabolism & four enzymes involved in Amino sugar and nucleotide sugar metabolism which is depicted above *i.e.* ec:3.2.1.39 -endo-1,3-beta-D-glucosidase, ec:2.7.7.27-adenylyltransferase, ec:3.2.1.26-invertase, ec:2.4.1.12 - synthase (UDP-forming), ec:3.2.1.21-gentiobiase, ec:3.2.1.2-saccharogen amylase, ec:2.4.1.1-phosphorylase, ec:1.1.1.22-6-dehydrogenase, ec:3.2.1.55-end alpha-L-arabinofuranosidase,ec:2.4.1.43-4-alpha-galacturonosyltransferase (**Figure 4**).

Construction of miRNA-Mediated Regulatory Networks

Ten independent networks were obtained for 24 miRNA families targeting 754 candidate genes with lowest target score and highest Mfe (**Figure 5**). The ath-MIR156b had maximum targets (403 genes), followed by ath-MIR5655 (248 genes), ath-MIR5651_3 (38 genes), ath-MIR5021_2 (19 genes) and ath-MIR157c (9 genes). Several other miRNAs had one to three targets. Network graphs depicted the candidate genes involved in seed development and their regulation by the different miRNA-families (**Supplementary Table S8**).

Developing EST-SSRs From Target Genes

To provide markers for selecting seed type trait, SSR survey of 1935 unique target genes facilitated in identification of 7,688 perfect SSR motifs, with highest frequency of hexa-nucleotide repeats (3,497), followed by di- (1,632) and tri-nucleotides (1,661). Penta-nucleotides (181) were the least frequent followed by mono-nucleotides (276) (**Supplementary Figure S1**). Finally, a total of 413 functional EST-SSR markers were designed, majority of which (204) targeted hexa-nucleotide motifs accompanied by di- (95), and tri-nucleotides (73). These EST-SSRs represent important genomic resources that would enable discovery of genes/QTLs for seed type traits in pomegranate (**Supplementary Table S9**).

PCR Based Validation and Diversity Analysis

We synthesized a set of 32 pre-miRNA-SSRs specific to seed type and assayed these selected SSRs on six pomegranate genotypes. As a result, 28 (87.5%) miRNA-SSRs yielded the amplicons of expected size, whereas no amplification was recorded for four miRNA-SSRs. Of these, 25 (89.29%) miRNA-SSRs revealed polymorphism across six pomegranate genotypes, while remaining three markers (MIR_SH_SSR41, 69 and 90) were monomorphic (**Supplementary Table S10**). Marker profiles of pomegranate genotypes using selected SSRs (**Supplementary Figure S2**).

FIGURE 5 | Genes regulatory network for 24 miRNA families. Darker lines with color codes for each family indicate stronger interactions of miRNAs with their target genes based on highest minimum free energy ratio of hybridization and lowest alignment scores.

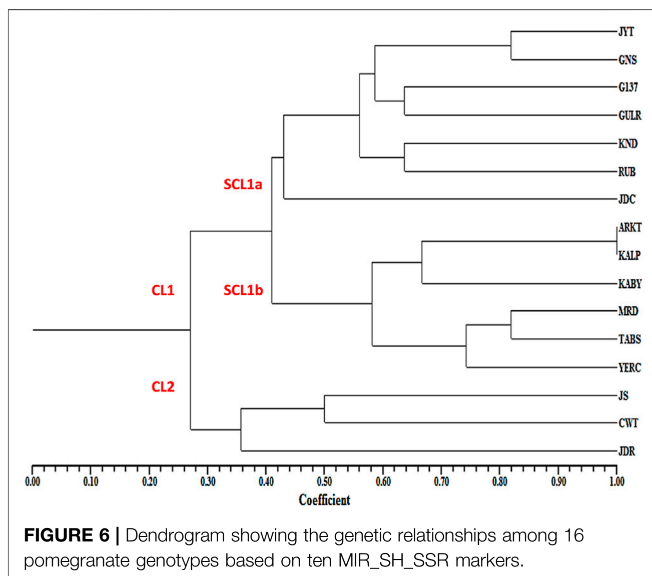


FIGURE 6 | Dendrogram showing the genetic relationships among 16 pomegranate genotypes based on ten MIR_SH_SSR markers.

The 28 primers produced one to two alleles, with the PIC values ranging between 0 and 0.54. Based on these results a subset of 10 miRNA-SSRs was selected and genotyped on 16 pomegranate genotypes to estimate genetic diversity. These markers amplified total 22 alleles with average of two alleles per locus. The MAF per locus ranged from 0.50 to 0.87, with average of 0.70. The *He* ranged from 0.23 to 0.50, with an average of 0.39. The PIC values ranged from 0.24 to 0.52, with an average of 0.40 (**Supplementary Table S11**). The mean Shannon's information index value of 0.57 was observed among the genotypes.

In the UPGMA tree, all the 16 pomegranate genotypes were grouped into two major clusters, with cluster one harboring 13 and cluster two containing three pomegranate genotypes. The cluster one was further divided into two sub clusters 1a and 1b (**Figure 6**).

DISCUSSION

Recent advances in DNA sequencing and computational analysis in pomegranate have facilitated large-scale mining of DNA markers including EST-SSRs (Ophir et al., 2014), EST-SNPs (Harel-Beja et al., 2015), genomic SSRs (Liu et al., 2020), and most recently miRNA-SSRs (Patil et al., 2020). This provided immense opportunities to integrate information from genome,

transcriptome and non-coding RNAs to understand genetic mechanisms that are regulating important traits in pomegranate. The DNA markers present in miRNAs have potential utility for the identification of master miRNAs that regulate the key genes for fruit quality traits (Patil et al., 2020). Here, we provided novel pre-miRNAs reported for seed type trait in pomegranate (Luo et al., 2018), to develop, validate and localization of miRNA-SSRs markers on Tunisia chromosomes. Also report development of 413 functional EST-SSRs, which may be of potential use for functional validation for seed type and breeding applications.

Survey of SSR Motifs in the miRNA-Coding Sequences

SSR survey in pri-miRNAs resulted in designing of 132 miRNA-SSR primers, of which 46 primers are specific to pre-miRNAs. Previously, we reported genome-wide survey for miRNA-SSRs for seedling to fruit development stages in pomegranate genome using *in silico* approaches. Homology-based search has enabled development of large-scale miRNA-SSRs in several plant species including *Medicago truncatula* and *Arabidopsis thaliana* (Kumar et al., 2017; Min et al., 2017). In the present study, frequency distribution analysis of SSR occurrence in both pri-and pre-miRNAs revealed the abundance of hexa-nucleotides repeats followed by mono-and di-nucleotides. With respect to motif types, we found A/T repeats were the most abundant, followed by AT/AT (76.31%) in pri- and pre-miRNA sequences. These results remain in good agreement with our previous finding as reported in pomegranate (Patil et al., 2020). By contrast, mono- and di-nucleotide repeats were pronounced in miRNAs coding sequences of *M. trucatula*, *A. thaliana* and other species (Chen et al., 2010; Kumar et al., 2017; Min et al., 2017).

Functional Classification and Pathway Enrichment for miRNA Target Genes

Mining of miRNA-SSRs and use of its mature miRNA for target prediction followed by functional annotations could lead to identification of candidate genes underlying traits of interest. Therefore, miRNA-SSRs (132) and EST-SSRs (413) derived from miRNA-coding sequences and their 1935 unique gene targets, respectively have greater implications for improving seed type trait in pomegranate. Our findings on target gene functions are congruent with the previous reports in pomegranate based on an integrated analysis of microRNA identification and mRNA expression profiling for seed type trait (Luo et al., 2018). The study reported 408 and 335 mRNA targets in the Tunisia and Sanbai genomes at 60 and 120 DAF, respectively (Luo et al., 2018). Xue et al. (2017) also reported 34,221 genes that had differential expression (DEGs) between soft and hard-seeded pomegranate using global gene expression profiling. As elucidated from the GO annotations, the functional roles of the miRNA gene targets remain in strong agreement with our present study.

In the present study, KEGG pathway enrichment analysis uncovered possible pathways associated with hard and soft seeded types in pomegranate. Cao et al. (2015), through comparative proteomics analysis identified key enzymes involved in tricarboxylic acid cycle (TCA) and mitochondrial metabolism during fruit maturation periods in soft seeded pomegranate cultivar Zhongnonghong. Recently, Luo et al. (2018) reported 14 and 8 DEG targets for miRNAs between soft and hard-seeded pomegranate at 60 and 120 DAF, which are part of 18 and eight pathways. Similarly, Xue et al. (2017) also reported DEGs between soft and hard-seeded pomegranate that participate in photosynthesis, benzene propane synthesis, phospholipid metabolism, ribosome metabolism and ubiquitin mediated proteolysis. Another study in pomegranate revealed genes targets of 41 miRNA families that were associated with in 107 major pathways controlling fruit development.

Here, we found 101 enzymes that are part of 58 pathways, seven enzymes participating in starch and sucrose metabolism, whereas four enzymes had roles in amino sugar and nucleotide sugar metabolism. Niu et al. (2018) found that ovules-to-seed transition in pomegranate was regulated by the co-expression of many proteins in the short term. The authors predicted protein-protein interactions among 505 and 549 proteins at 60 and 120 DAF, respectively. Of these, seven proteins were involved in phenylpropanoid biosynthesis whereas 15 had roles in starch and sucrose metabolism. Research suggests a close association between UDP-glucose pyrophosphorylase (UGP) and sucrose synthase (SUS) with cellulose biosynthesis in plants (Andersson-Guunneras et al., 2006). Similarly, differences in expression abundance at both gene and protein levels between Tunisia and Sanbai suggested that UGP2 and SUS3 were upregulated at 60 DAF and downregulated at 120 DAF in Tunisia (Niu et al., 2018). Evidence of lower lignin and higher cellulose during early fruit developmental stages supported the contradictory roles of lignin and cellulose in cell wall formation in soft-seeded pomegranate. Our findings are congruent with an earlier report of higher cellulose-related gene expression and cellulose content in soft-seeded pomegranate varieties in contrast with the hard-seeded varieties with higher lignin (Zarei et al., 2016). Its also may be due to fact that use of Tunisia gene models for target analysis in our study identified several targets engaged in cellulose synthesis.

miRNA Regulatory Networks

MiRNAs are usually known to negatively regulate their targets. The complementarities between the miRNA and their target genes serves as a basis for identification of the gene targets engaged in a variety of processes to plant growth and development (Singh et al., 2017; Mishra and Bohra, 2018). Therefore, here we established regulatory networks among miRNA and mRNA to understand their roles in seed type. Previously, it was reported that seed type is related to cell wall biosynthesis (Zarei et al., 2016). Numerous candidates like MYB, WRKY, AP2-like, MYC and NAC are known to play important roles in pomegranate and hawthorns in regulating seed type. These transcription factors are involved in brassinosteroid biosynthesis, cell division, lignin, cellulose flavonoid and xyloglucan biosynthesis (Dai et al., 2013; Xue et al., 2017; Luo et al., 2020). In our study target and network analysis revealed

ath-miR156b had maximum targets followed by ath-MIR5655, ath-MIR5651_3, ath-MIR5021_2 and ath-MIR157c suggesting these are the most abundant families in pomegranate. We propose a possible role of all these miRNA families in complex regulation of seed maturation since these showed strongest interactions with MYB, auxin response factors, WRKY and NAC, AP2/ERF and B3 domain-containing transcription factors and enzymes mainly involved in cellulose, lignin and sugar metabolism etc. However, Luo et al. (2018) reported mdm-miR164e- and mdm-miR172b-targets included WRKY, MYC and NAC1 mainly involving brassinosteroid biosynthesis, cell division and lignin biosynthesis. Xia et al. (2019) characterized the role of a NAC transcription factor (PgSND1-like) involved in the regulation of seed type in pomegranate. These findings suggest that a complex biological process mediated by miRNA–mRNA network controls pomegranate seed type (Luo et al., 2018).

Huang et al. (2013) found mir156 is the largest family while studying the miRNAs and their purgative targets in *Brassica napus* seed maturation. They found majority of the family members were primarily expressed in the embryo and they may also regulate the developmental transition to germination. In pomegranate, Luo et al. (2018) identified 40 miRNA families for seed type trait, with miR156 family showing the highest 31 members (14.98% of the total). Similar reports exist in pear (Wu et al., 2014) and apple (Xing et al., 2016). Saminathan et al. (2016) found miR157 as the most abundant family in pomegranate. Interestingly, we noticed fewer targets for ath-MIR157c in this study. The miR156 and miR157 are known to regulate ovule development by targeting SQUAMOSA-promoter binding protein or box transcription factors (SPL/SBP) in plants (Liu et al., 2017). Evidence supports the role of miR156/157 and miR172 in controlling flowering and the vegetative-to-reproductive transition in plants (Wu et al., 2009). Therefore, participation of miRNA156 family has been well documented during seed cycle *i.e.* during dormancy modulation, germination, development and maturation. Earlier research has shown a role of AP2-like TFs in regulating seed size and seed mass in *Arabdopisis* (Ohto et al., 2009).

In our previous report, we found ath-MIR5651 is strongly interacting with MYB-like transcription factors during fruit development in pomegranate (Patil et al., 2020). In brassica, Singh et al. (2017) predicted involvement of miR5021 in starch and sugar metabolism and the corresponding target glycosyl hydrolase family proteins that participate in pentose and gluconate inter-conversion pathway during abiotic stress. Borges et al. (2011) reported 25 potentially novel miRNAs including miR5021 processed in sperm cells and pollen. The authors found sub functionalization of these miRNAs in association with a putative germline-specific Argonaute complex. ARGONAUTE 5 (AGO5) localizing preferentially to the sperm cell cytoplasm in mature pollen, suggest possible role of these miRNAs in germline differentiates and formation of mature male gametophyte. Amiteye et al. (2013) discovered 109 *Boechera* small RNAs while unrevealing the role of miRNAs during switch from sexual to apomictic reproduction in *Boechera* species. They found these miRNAs showing significant homology to 407 *Arabidopsis thaliana* small RNAs including the *A. thaliana* pollen-specific ath-miR5021 indicating possible role in reproduction.

On the other hand, we have identified families that have only few targets *viz.,* ath-MIR167d, ath-MIR169c and ath-MIR401 had two targets each, *i.e.* zinc metalloprotease EGY2 and chloroplastic LOC116189522, kinesin-like protein NACK2 and protein transport protein SEC16A homolog, uncharacterized LOC116188785 and pentatricopeptide repeat-containing protein At5g08490, respectively. For ath-MIR395f we found one target *i.e.* pentatricopeptide repeat-containing protein At1g56570. The essential role for miRNA167 in maternal control of embryonic and seed development was confirmed by the deletion of four *microRNA167* (*MIR167*) genes in *Arabidopsis* using gene editing. Plants with *mir167a* mutant and the *ARF* overexpression were found defective in anther dehiscence and ovule development (Yao et al., 2019). Zhou et al. (2007) reported UV-B stress induced upregulation of miR401 in *Arabidopsis*. Kim et al. (2010) reported miR395c/e differentially affects seed germination of *Arabidopsis* under stress conditions.

Besides, the miRNA families that specifically targeted embryogenesis-associated protein, small heat shock protein and resistance proteins RPM1 and RPP13 were mtr-MIR5253, ptr-mir-564, ath-MIR8180 and ath-MIR5021_2. This implied towards that these miRNA families orchestrate the expression of important TFs and enzymes during the early stages of seed development and regulate proteins involved in storage compound synthesis and transport in the mature seeds. Heat shock proteins expressed during pomegranate fruit maturation periods implicate them in protecting the pomegranate seeds from adverse stress such as extreme temperatures (Cao et al., 2015). Similarly, development of miRNA regulatory networks has greatly helped expanding knowledge about the miRNA and their target gene interactions in various other crops *viz.,* bread wheat (Nigam et al., 2015), radish (Zhang et al., 2016), maize (Wu et al., 2016) and brassica (Singh et al., 2017). Finally, through target analysis we could identify few informative miRNA-SSRs that can be deployed after functional validation for genetic improvement of seed type.

Experimental Validation and Mapping of miRNA-SSRs Having Associations With Seed Type

Despite availability of a diverse array of DNA marker systems in pomegranate, their applications in improvement programs remain limited, possibly due to lack of trait-associated markers (Patil et al., 2020). Mondal and Ganie (2014) reported miR172b-SSR that could differentiate rice genotypes with respect to their salt stress response. Therefore, such functional DNA markers hold great potential in pomegranate for trait mapping and genome editing applications.

E-mapping of 132 miRNA-SSRs on Tunisia chromosomes led the validation of 123 markers across all the chromosomes. A physical genetic map was developed based on the information on start positions of all the markers. Information on start positions of markers had allowed

successful anchoring of the new SSR markers to the physical map of groundnut (Lu et al., 2019). Portis et al. (2016) also reported genomic distribution of SSRs and their relations with annotated genomic components (gene space) based on the information on assembled pseudomolecules of globe artichoke genome. The information on chromosome wise location of miRNA-SSR markers could help in precise validation and introgression of genes/QTLs for seed type in pomegranate. Several reports exist on fine mapping of genes/QTLs enabled by an SSR-based physical map (Zhao et al., 2017).

Identification of Informative miRNA-SSRs Through ePCR

Through ePCR we identified 80 primers producing one to two alleles across the four pomegranate genomes. Out of these, 63 primers had PIC values greater than 0.50, implying their highly informative nature. Therefore, we believe that 63 highly polymorphic miRNA-SSRs identified in this study would be a valuable genomic tool for understanding the genetic makeup of seed type trait of pomegranate. Recently, Uncu and Uncu (2020)) developed chromosome anchored SSR markers for carrot genome assembly, and identified 51,160 single-locus markers through e-PCR. Further, they experimentally evaluated 50 markers across 17 carrot accessions and found 46 markers produced expected product sizes suggesting accuracy rate of 90% in predicting the amplification profiles by e-mapping. The DNA markers designed through *in silico* mining have been succefully validated through experimental assays in wheat (Han et al., 2015), cucumber (Liu et al., 2015), bitter gourd (Cui et al., 2017) and tobacco (Wang et al., 2018). Therefore, the dataset generated here assist pomegranate research community in future genetic research and applied breeding.

Apart from miRNA-SSRs, we report large scale development of 1935 unique target gene based 413 EST-SSR markers for mapping genes/QTLs for seed type trait in pomegranate. In a previous study, we reported a subset of 58 functional EST-SSR markers from 128 target genes (Patil et al., 2020). Similarly, miRNA-mRNA complementarity has permitted the development of SSR from predicted target genes in several other crops such as 700 SSRs from 621 target genes in *Brassica* (Singh et al., 2017).

Wet Lab Validation and Diversity Analysis

Through wet lab experiments, we observed that 28 primers showed clear and reproducible amplifications on metaphor gels, with an average PIC value of 0.38. Earlier we found an average PIC value of 0.29 for 47 pre-miRNA-SSRs validated on eight pomegranate genotypes (Patil et al., 2020). A similar PIC value of 0.43 was obtained when 34 miRNA-SSRs were screened on six *Brassica* species (Fu et al., 2013). In our study, 13 miRNA-SSRs had higher PIC (≥0.48) on metaphor gels, implying towards their greater potential for future research on seed type trait in pomegranate. According to Bandelj et al. (2004), DNA markers having PIC values greater than 0.5 can be considered informative.

Twenty two alleles were obtained by analysing 10 miRNA-SSRs on 16 pomegranate genotypes. However, 87 alleles obtained for 15 miRNA-SSRs in our earlier study on 18 pomegranate genotypes using fragment analyser indicated that SSR fragments can be better separated using high resolution systems (Patil et al., 2020). The gel detection system used in the present study has allowed detection of a relatively less number of alleles and lower PIC values, possibly due to its poor resolving power. Hence we advocate that an allele detection system with a higher resolving power such as based on capillary electrophoresis would be more appropriate for future research using these new SSR markers. In our study, the higher diversity index (0.62–0.69) and PIC values (0.44–0.52) were observed for five MIR_SH_SSRs 11, 23, 26, 37 and 64 suggesting highly informative in nature. The mean Shannon's information index indicated moderate genetic diversity level among the 16 genotypes assayed. These results are in agreement with earlier finding as reported for miRNA-SSRs in pomegranate (Patil et al., 2020).

Cluster analysis separated 16 pomegranate genotypes into two major clusters. The cluster one constituted two sub clusters, where most of the soft seeded genotypes were clearly separated from hard seeded types with few exceptions. Sub-cluster 1a is composed of seven soft seeded cultivars (Jyoti, Ganesh, G137, Gulesha Red, Kandhari, Ruby and Jodhpur collection). Similarly, sub cluster 1b contained four hard-seeded types (Kalpitiya, Kabuli Yellow, Tabesta and Yercuad) with a slight deviation. Cluster two had three pomegranate cultivars belonging to soft seeded (Jallore Seedless) and hard seeded cultivars (Co-white and Jodhpur Red). These results indicated that the seed type specific miRNA-SSRs developed here could differentiate the genotypes according to their seed trait. Sequence variations in miRNA precursor molecules can have profound impact on the expressions of the genes associated with varietal seed type trait in pomegranate. For instance, Joy et al. (2018) highlighted the significance of SSRs in pre-miRNAs. The study suggested a role of SSRs in the formation of mature RNA isoforms via alternative splicing events during stress response. A similar association of miRNA-SSRs with the gene expressions and phenotypic manifestations has also been reported in rice (Mondal and Ganie, 2014; Ganie and Mondal, 2015).

CONCLUSION

Here we reported integrated analysis of sequence data on genome, transcriptome and nc RNAs to generate trait specific novel DNA markers like miRNA-SSRs and EST-SSRs for seed type trait. *In silico* analysis enabled the development of 132 miRNA-SSR primers for seed type trait from 761 pre-miRNA sequences. One hundred twenty three markers were successfully validated and mapped onto Tunisia chromosomes to generate physical maps. Further, identified 63 highly polymorphic markers out of 80 miRNA-SSR primers producing unique amplicons through ePCR validation across multiple genomes of pomegranate. PCR validation and genetic diversity analysis confirmed the utility of these novel markers for future research and breeding. Following target prediction and SSR mining in resulting 1935 unique genes, we developed 413 EST-SSR markers with possible association with seed type trait. Besides, we demonstrated the utility of these markers in identifying master miRNAs and their candidate genes thorough target analysis, functional annotations of target genes and their KEGG pathways and network analysis. We identified a subset of five

informative miRNA-SSRs (miRNA_SH_SSR69, miRNA_SH_SSR36, miRNA_SH_SSR103, miRNA_SH_SSR35 and miRNA_SH_SSR53) that influence functioning of several genes involved in seed hardness process. The potential miRNAs and their candidate targets identified here create new avenues for gene editing applications for improving seed type trait in pomegranate.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Materials**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

PP and RM designed the research experiments. PP and SJ performed the *in silico* analysis to design miRNA-SSR markers and functional annotations. DK and NS contributed in collection of test materials. PP, SJ and MN performed the wet lab experiments for SSR validation and diversity analysis. PP, JS, VC and AB wrote the original paper with RM.

FUNDING

This work was carried out with the financial support from Indian Council of Agricultural Research (ICAR), New Delhi, India extended through ICAR-National Research Centre on Pomegranate, Solapur (M.S) under institute-funded project.

ACKNOWLEDGMENTS

The authors are highly grateful to the Indian Council of Agricultural Research (ICAR), New Delhi for extended financial support through ICAR-National Research Centre on Pomegranate, Solapur.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.866504/full#supplementary-material

Supplementary Table 1 | Details of 16 pomegranate genotypes used in this study.

Supplementary Table 2 | Details of 132 miRNA_SH_SSR primers designed in relation to seed hardness trait in pomegranate.

Supplementary Table 3 | Details of physical locations of 123 miRNA-SSR markers which are distributed across the 8 chromosomes of Tunisia.

Supplementary Table 4 | Marker statistics as computed for 80 miRNA-SSR primers screened through ePCR on four pomegranate genomes.

Supplementary Table 5 | Details of 1935 unique gene targets with functional details which are targeted by 24 miRNA families in this study.

Supplementary Table 6 | Details of miRNA-SSRs, their families and the target candidate genes engaged in seed development and maturation.

Supplementary Table 7 | Details of 58 KEGG pathways for annotated miRNA-target genes with enzymes details.

Supplementary Table 8 | Details of top 754 target genes based on lowest alignment score and highest mfe which are targeted by 24 miRNA families.

Supplementary Table 9 | Details of 7688 SSR motifs and 413 EST-SSR primers designed based on 1935 unique targets of 24 miRNA families.

Supplementary Table 10 | List of 32 pre-miRNA-SSRs initially validated on six pomegranate genotypes.

Supplementary Table 11 | Genetic diversity statistics of ten MIR_SH_ SSRs obtained from 16 pomegranate genotypes.

Supplementary Figure 1 | EST-SSR marker statistics and frequency distribution graphs of different SSR motifs in 1935 unique target genes of 24 miRNA families.

Supplementary Figure 2 | Allelic variations among **(A)** Six pomegranate genotypes using MIR_SH_SSR markers 64, 65, 67, 69, 71, 81, 84 and 86, and **(b)** Sixteen pomegranate genotypes using MIR_SH_SSRs 26, 29 and 37. (Where, L-100 bp DNA ladder, Lane 1–6 and Lane 1–16 set of pomegranate genotypes as mentioned in material and methods and listed in **Supplementary Table S1**)

REFERENCES

Akparov, Z., Amiraslanov, A., Hajiyeva, S., Abbasov, M., Kaur, K., Hamwieh, A., et al. (2017). De-novo Sequencing of Pomegranate (*Punica Granatum* L.) Genome. Heidelberg: EMBL. database.

Allen, E., Xie, Z., Gustafson, A. M., and Carrington, J. C. (2005). MicroRNA-directed Phasing during Trans-acting siRNA Biogenesis in Plants. *Cell* 121, 207–221. doi:10.1016/j.cell.2005.04.004

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a New Generation of Protein Database Search Programs. *Nucleic Acids Res.* 25, 3389–3444. doi:10.1093/nar/25.17.3389

Amiteye, S., Corral, J. M., Vogel, H., Kuhlmann, M., Mette, M. F., and Sharbel, T. F. (2013). Novel microRNAs and Microsatellite-like Small RNAs in Sexual and Apomictic Boechera Species. *Mirna* 2, 46–63. doi:10.2174/2211536611302010006

Andersson-Gunnerås, S., Mellerowicz, E. J., Love, J., Segerman, B., Ohmiya, Y., Coutinho, P. M., et al. (2006). Biosynthesis of Cellulose-Enriched Tension wood in Populus: Global Analysis of Transcripts and Metabolites Identifies Biochemical and Developmental Regulators in Secondary wall Biosynthesis. *Plant J.* 45, 144–165. doi:10.1111/j.1365-313X.2005.02584.x

Bandelj, D., Jakše, J., and Javornik, B. (2004). Assessment of Genetic Variability of Olive Varieties by Microsatellite and AFLP Markers. *Euphytica* 136, 93–102. doi:10.1023/b:euph.0000019552.42066.10

Bastian, M., Heymann, S., and Jacomy, M. (2009). "Gephi: an Open Source Software for Exploring and Manipulating Networks," in The Third international AAAI conference on weblogs and social media.

Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a Web Server for Microsatellite Prediction. *Bioinformatics* 33, 2583–2585. doi:10.1093/bioinformatics/btx198

Bonnet, E., He, Y., Billiau, K., and Van de Peer, Y. (2010). TAPIR, a Web Server for the Prediction of Plant microRNA Targets, Including Target Mimics. *Bioinformatics* 26, 1566–1568. doi:10.1093/bioinformatics/btq233

Borges, F., Pereira, P. A., Slotkin, R. K., Martienssen, R. A., and Becker, J. D. (2011). MicroRNA Activity in the Arabidopsis Male Germline. *J. Exp. Bot.* 62, 1611–1620. doi:10.1093/jxb/erq452

Cao, S.-Y., Niu, J., Cao, D., Li, H.-X., Xue, H., Chen, L.-N., et al. (2015). Comparative Proteomics Analysis of Pomegranate Seeds on Fruit Maturation Period (*Punica Granatum* L.). *J. Integr. Agric.* 14, 2558–2564. doi:10.1016/s2095-3119(15)61029-2

Chen, M., Tan, Z., Zeng, G., and Peng, J. (2010). Comprehensive Analysis of Simple Sequence Repeats in Pre-miRNAs. *Mol. Biol. Evol.* 27, 2227–2232. doi:10.1093/molbev/msq100

Cui, J., Cheng, J., Nong, D., Peng, J., Hu, Y., He, W., et al. (2017). Genome-wide Analysis of Simple Sequence Repeats in Bitter Gourd (*Momordica charantia*). *Front. Plant Sci.* 8, 1103. doi:10.3389/fpls.2017.01103

Dai, H., Han, G., Yan, Y., Zhang, F., Liu, Z., Li, X., et al. (2013). Transcript Assembly and Quantification by RNA-Seq Reveals Differentially Expressed Genes between Soft-Endocarp and Hard-Endocarp Hawthorns. *PLoS One* 8, e72910. doi:10.1371/journal.pone.0072910

Ferrao, L. F. V., Caixeta, E. T., Pena, G., Zambolim, E. M., Cruz, C. D., Zambolim, L., et al. (2015). New EST–SSR Markers of *Coffea Arabica*: Transferability and Application to Studies of Molecular Characterization and Genetic Mapping. *Mol. Breed.* 35, 1–5. doi:10.1007/s11032-015-0247-z

Fu, D., Ma, B., Mason, A. S., Xiao, M., Wei, L., and An, Z. (2013). MicroRNA-based Molecular Markers: a Novel PCR-Based Genotyping Technique in Brassica species. *Plant Breed* 132, 375–381. doi:10.1111/pbr.12069

Ganie, S. A., and Mondal, T. K. (2015). Genome-wide Development of Novel miRNA-Based Microsatellite Markers of rice (*Oryza Sativa*) for Genotyping Applications. *Mol. Breed.* 35, 1–12. doi:10.1007/s11032-015-0207-7

Han, B., Wang, C., Tang, Z., Ren, Y., Li, Y., Zhang, D., et al. (2015). Genome-wide Analysis of Microsatellite Markers Based on Sequenced Database in Chinese spring Wheat (*Triticum aestivum* L.). *PLoS One* 10, e0141540. doi:10.1371/journal.pone.0141540

Harel-Beja, R., Sherman, A., Rubinstein, M., Eshed, R., Bar-Ya'akov, I., Trainin, T., et al. (2015). A Novel Genetic Map of Pomegranate Based on Transcript Markers Enriched with QTLs for Fruit Quality Traits. *Tree Genet. Genomes* 11, 1–18. doi:10.1007/s11295-015-0936-0

Huang, D., Koh, C., Feurtado, J. A., Tsang, E. W., and Cutler, A. J. (2013). MicroRNAs and Their Putative Targets in *Brassica Napus* Seed Maturation. *BMC Genomics* 14, 140. doi:10.1186/1471-2164-14-140

Joy, N., Maimoonath Beevi, Y. P., and Soniya, E. V. (2018). A Deeper View into the Significance of Simple Sequence Repeats in Pre-miRNAs Provides Clues for its Possible Roles in Determining the Function of microRNAs. *BMC Genet.* 19, 29. doi:10.1186/s12863-018-0615-x

Kim, J. Y., Lee, H. J., Jung, H. J., Maruyama, K., Suzuki, N., and Kang, H. (2010). Overexpression of microRNA395c or 395e affects differently the seed germination of *Arabidopsis thaliana* under stress conditions. *Planta* 232, 1447–1454. doi:10.1007/s00425-010-1267-x

Kumar, A., Chauhan, A., Kompelli, S. K., Gahlaut, V., Ijaq, J., Singh, K. P., et al. (2017). Genome-Wide Mining, Characterization and Development of miRNA-SSRs in *Arabidopsis thaliana*. *BioRxiv p*, e203851. doi:10.1101/203851

Li, W., and Godzik, A. (2006). Cd-hit: a Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. *Bioinformatics* 22, 1658–1659. doi:10.1093/bioinformatics/btl158

Liu, C., Li, J., and Qin, G. (2020). Genome-wide Distribution of Simple Sequence Repeats in Pomegranate and Their Application to the Analysis of Genetic Diversity. *Tree Genet. Genomes* 16, 1–9. doi:10.1007/s11295-020-1428-4

Liu, J., Qu, J., Hu, K., Zhang, L., Li, J., Wu, B., et al. (2015). Development of Genomewide Simple Sequence Repeat Fingerprints and Highly Polymorphic Markers in Cucumbers Based on Next-Generation Sequence Data. *Plant Breed* 134, 605–611. doi:10.1111/pbr.12304

Liu, N., Tu, L., Wang, L., Hu, H., Xu, J., and Zhang, X. (2017). MicroRNA 157-targeted SPL Genes Regulate floral Organ Size and Ovule Production in Cotton. *BMC Plant Biol.* 17, 7. doi:10.1186/s12870-016-0969-z

Liu, Q., Luo, L., and Zheng, L. (2018). Lignins: Biosynthesis and Biological Functions in Plants. *Int. J. Mol. Sci.* 19, 335. doi:10.3390/ijms19020335

Lu, L. J. (2006). *Cloning and Sequence Measuring of the Marker Linked to Seed Trait in Soft Seeded Pomegranates*. Msc thesis. China: Anhui Agric. Univ. (in Chinese).

Lu, Q., Hong, Y., Li, S., Liu, H., Li, H., Zhang, J., et al. (2019). Genome-wide Identification of Microsatellite Markers from Cultivated Peanut (*Arachis hypogaea* L.). *BMC Genomics* 20, 1–9. doi:10.1186/s12864-019-6148-5

Luo, X., Cao, D., Zhang, J., Chen, L., Xia, X., Li, H., et al. (2018). Integrated microRNA and mRNA Expression Profiling Reveals a Complex Network Regulating Pomegranate (*Punica Granatum* L.) Seed Hardness. *Sci. Rep.* 8, 9292. doi:10.1038/s41598-018-27664-y

Luo, X., Li, H., Wu, Z., Yao, W., Zhao, P., Cao, D., et al. (2020). The Pomegranate (*Punica granatum* L.) Draft Genome Dissects Genetic Divergence between Soft- and Hard-seeded Cultivars. *Plant Biotechnol. J.* 18, 955–968. doi:10.1111/pbi.13260

Millar, A. A. (2020). The Function of miRNAs in Plants. *Plants* 9, 198. doi:10.3390/plants9020198

Min, X., Zhang, Z., Liu, Y., Wei, X., Liu, Z., Wang, Y., et al. (2017). Genome-wide Development of microRNA-Based SSR Markers in *Medicago Truncatula* with Their Transferability Analysis and Utilization in Related Legume Species. *Int. J. Mol. Sci.* 18, 2440. doi:10.3390/ijms18112440

Mishra, A., and Bohra, A. (2018). Non-coding RNAs and Plant Male Sterility: Current Knowledge and Future Prospects. *Plant Cel Rep* 37, 177–191. doi:10.1007/s00299-018-2248-y

Mondal, T. K., and Ganie, S. A. (2014). Identification and Characterization of Salt Responsive miRNA-SSR Markers in rice (*Oryza sativa*). *Gene* 535, 204–209. doi:10.1016/j.gene.2013.11.033

Nigam, D., Kadimi, P. K., Kumar, S., Mishra, D. C., and Rai, A. (2015). Computational Analysis of miRNA-Target Community Network Reveals Cross Talk Among Different Metabolisms. *Genomics Data* 5, 292–296. doi:10.1016/j.gdata.2015.04.028

Niu, J., Cao, D., Li, H., Xue, H., Chen, L., Liu, B., et al. (2018). Quantitative Proteomics of Pomegranate Varieties with Contrasting Seed Hardness during Seed Development Stages. *Tree Genet. Genomes* 14, 1–12. doi:10.1007/s11295-018-1229-1

Ohto, M.-A., Floyd, S. K., Fischer, R. L., Goldberg, R. B., and Harada, J. J. (2009). Effects of APETALA2 on Embryo, Endosperm, and Seed Coat Development Determine Seed Size in Arabidopsis. *Sex. Plant Reprod.* 22, 277–289. doi:10.1007/s00497-009-0116-1

Ophir, R., Sherman, A., Rubinstein, M., Eshed, R., Sharabi Schwager, M., Harel-Beja, R., et al. (2014). Single-nucleotide Polymorphism Markers from De-novo Assembly of the Pomegranate Transcriptome Reveal Germplasm Genetic Diversity. *PLoS One* 9, e88998. doi:10.1371/journal.pone.0088998

Patil, P. G., Singh, N. V., Parashuram, S., Bohra, A., Mundewadikar, D. M., Sangnure, V. R., et al. (2020). Genome Wide Identification, Characterization and Validation of Novel miRNA-Based SSR Markers in Pomegranate (*Punica granatum* L.). *Physiol. Mol. Biol. Plants* 26, 683–696. doi:10.1007/s12298-020-00790-6

Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research-Aan Update. *Bioinformatics* 28, 2537–2539. doi:10.1093/bioinformatics/bts460

Portis, E., Portis, F., Valente, L., Moglia, A., Barchi, L., Lanteri, S., et al. (2016). A Genome-wide Survey of the Microsatellite Content of the globe Artichoke Genome and the Development of a Web-Based Database. *PLoS One* 11, e0162841. doi:10.1371/journal.pone.0162841

Qin, G., Xu, C., Ming, R., Tang, H., Guyot, R., Kramer, E. M., et al. (2017). The Pomegranate (*Punica granatum* L.) Genome and the Genomics of Punicalagin Biosynthesis. *Plant J.* 91, 1108–1128. doi:10.1111/tpj.13625

Ravishankar, K. V., Anand, L., and Dinesh, M. R. (2000). Assessment of Genetic Relatedness Among Mango Cultivars of India Using RAPD Markers. *J. Hortic. Sci. Biotechnol.* 75, 198–201. doi:10.1080/14620316.2000.11511223

Rohlf, J. (2000). *Numerical Taxonomy and Multivariate Analysis System NTSYS-Pc V 2. 11L*. New York, USA: Applied Biostatistics.

Saminathan, T., Bodunrin, A., Singh, N. V., Devarajan, R., Nimmakayala, P., Jeff, M., et al. (2016). Genome-wide Identification of microRNAs in Pomegranate (*Punica granatum* L.) by High-Throughput Sequencing. *BMC Plant Biol.* 16, 122. doi:10.1186/s12870-016-0807-3

Schuler, G. D. (1997). Sequence Mapping by Electronic PCR. *Genome Res.* 7, 541–550. doi:10.1101/gr.7.5.541

Singh, I., Smita, S., Mishra, D. C., Kumar, S., Singh, B. K., and Rai, A. (2017). Abiotic Stress Responsive miRNA-Target Network and Related Markers (SNP, SSR) in *Brassica Juncea*. *Front. Plant Sci.* 8, 1943. doi:10.3389/fpls.2017.01943

Song, S. W., Liu, C. H., Zhang, F. M., and Tao, Y. H. (2012). Study on Cold Resistance of Some Varieties of Pomegranate. *J. Henan Agric. Univ.* 46, 43–146. [in Chinese]. doi:10.1016/j.matlet.2012.07.077

Uncu, A. O., and Uncu, A. T. (2020). High-throughput Simple Sequence Repeat (SSR) Mining Saturates the Carrot (*Daucus carota* L.) Genome with Chromosome-Anchored Markers. *Biotechnol. Biotechnological Equipment* 34, 1–9. doi:10.1080/13102818.2019.1701551

Voorrips, R. E. (2002). MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *J. Heredit.* 93, 77–78. doi:10.1093/jhered/93.1.77

Wang, X., Yang, S., Chen, Y., Zhang, S., Zhao, Q., Li, M., et al. (2018). Comparative Genome-wide Characterization Leading to Simple Sequence Repeat Marker Development for *Nicotiana*. *BMC Genomics* 19, 500. doi:10.1186/s12864-018-4878-4

Wang, X., and Wang, L. (2016). GMATA: an Integrated Software Package for Genome-Scale SSR Mining, Marker Development and Viewing. *Front. Plant Sci.* 7, 1350. doi:10.3389/fpls.2016.01350

Wu, G., Park, M. Y., Conway, S. R., Wang, J.-W., Weigel, D., and Poethig, R. S. (2009). The Sequential Action of miR156 and miR172 Regulates Developmental Timing in Arabidopsis. *Cell* 138, 750–759. doi:10.1016/j.cell.2009.06.031

Wu, J., Wang, D., Liu, Y., Wang, L., Qiao, X., and Zhang, S. (2014). Identification of miRNAs Involved in Pear Fruit Development and Quality. *BMC Genomics* 15, 953. doi:10.1186/1471-2164-15-953

Wu, X., Ding, D., Shi, C., Xue, Y., Zhang, Z., Tang, G., et al. (2016). microRNA-Dependent Gene Regulatory Networks in maize Leaf Senescence. *BMC Plant Biol.* 16, 73. doi:10.1186/s12870-016-0755-y

Xia, X., Li, H., Cao, D., Luo, X., Yang, X., Chen, L., et al. (2019). Characterization of a NAC Transcription Factor Involved in the Regulation of Pomegranate Seed Hardness (*Punica granatum* L.). *Plant Physiol. Biochem.* 139, 379–388. doi:10.1016/j.plaphy.2019.01.033

Xing, L., Zhang, D., Zhao, C., Li, Y., Ma, J., An, N., et al. (2016). Shoot Bending Promotes Flower Bud Formation by Mi RNA -mediated Regulation in Apple (Malus domestica Borkh). *Plant Biotechnol. J.* 14, 749–770. doi:10.1111/pbi.12425

Xue, H., Cao, S., Li, H., Zhang, J., Niu, J., Chen, L., et al. (2017). De Novo transcriptome Assembly and Quantification Reveal Differentially Expressed Genes between Soft-Seed and Hard-Seed Pomegranate (*Punica granatum* L.). *PloS one* 12, e0178809. doi:10.1371/journal.pone.0178809

Yao, X., Chen, J., Zhou, J., Yu, H., Ge, C., Zhang, M., et al. (2019). An Essential Role for miRNA167 in Maternal Control of Embryonic and Seed Development. *Plant Physiol.* 180, 453–464. doi:10.1104/pp.19.00127

Ye, J., Zhang, Y., Cui, H., Liu, J., Wu, Y., Cheng, Y., et al. (2018). WEGO 2.0: a Web Tool for Analyzing and Plotting GO Annotations, 2018 Update. *Nucleic Acids Res.* 46, W71–W75. doi:10.1093/nar/gky400

Yuan, Z., Fang, Y., Zhang, T., Fei, Z., Han, F., Liu, C., et al. (2018). The Pomegranate (*Punica granatum* L.) Genome Provides Insights into Fruit Quality and Ovule Developmental Biology. *Plant Biotechnol. J.* 16, 1363–1374. doi:10.1111/pbi.12875

Zarei, A., Zamani, Z., Fatahi, R., Mousavi, A., Salami, S. A., Avila, C., et al. (2016). Differential Expression of Cell wall Related Genes in the Seeds of Soft- and Hard-Seeded Pomegranate Genotypes. *Scientia Horticulturae* 205, 7–16. doi:10.1016/j.scienta.2016.03.043

Zhang, W., Xie, Y., Xu, L., Wang, Y., Zhu, X., Wang, R., et al. (2016). Identification of microRNAs and Their Target Genes Explores miRNA-Mediated Regulatory Network of Cytoplasmic Male Sterility Occurrence during Anther Development in Radish (*Raphanus sativus* L.). *Front. Plant Sci.* 7, 1054. doi:10.3389/fpls.2016.01054

Zhang, X., Zhao, Y., Ren, Y., Wang, Y., and Yuan, Z. (2020). Fruit Breeding in Regard to Color and Seed Hardness: A Genomic View from Pomegranate. *Agronomy* 10, 991. doi:10.3390/agronomy10070991

Zhao, C., Qiu, J., Agarwal, G., Wang, J., Ren, X., Xia, H., et al. (2017). Genome-wide Discovery of Microsatellite Markers from Diploid Progenitor Species, *Arachis duranensis* and *A. ipaensis,* and Their Application in Cultivated Peanut (*A. hypogaea*). *Front. Plant Sci.* 8, 1209. doi:10.3389/fpls.2017.01209

Zhou, X., Wang, G., and Zhang, W. (2007). UV-B Responsive microRNA Genes in *Arabidopsis thaliana*. *Mol. Syst. Biol.* 3, 103. doi:10.1038/msb4100143

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Patil, Singh, Bohra, Jamma, N, C, Karuppannan, Sharma and Marathe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 05 May 2022
doi: 10.3389/fgene.2022.873850

Production of Conjoined Transgenic and Edited Barley and Wheat Plants for *Nud* Genes Using the CRISPR/SpCas9 System

Yiming Zang[1,2†], Qiang Gong[2†], Yanhao Xu[3†], Huiyun Liu[2], Hao Bai[2], Na Li[2], Lipu Du[2], Xingguo Ye[2], Caixia Lan[1] and Ke Wang[2]**

[1]College of Plant Science & Technology, Huazhong Agricultural University, Wuhan, China, [2]Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China, [3]Hubei Key Laboratory of Food Crop Germplasm and Genetic Improvement, Food Crops Institute, Hubei Academy of Agricultural Sciences, Wuhan, China

OPEN ACCESS

Edited by:
*Deepmala Sehgal,
International Maize and Wheat Improvement Center, Mexico*

Reviewed by:
*Muntazir Mushtaq,
National Bureau of Plant Genetic Resources (ICAR), India
Sajid Fiaz,
The University of Haripur, Pakistan*

***Correspondence:**
*Caixia Lan
cxlan@mail.hzau.edu.cn
Ke Wang
wangke03@caas.cn*

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *11 February 2022*
Accepted: *29 March 2022*
Published: *05 May 2022*

Citation:
Zang Y, Gong Q, Xu Y, Liu H, Bai H, Li N, Du L, Ye X, Lan C and Wang K (2022) Production of Conjoined Transgenic and Edited Barley and Wheat Plants for Nud Genes Using the CRISPR/SpCas9 System. Front. Genet. 13:873850. doi: 10.3389/fgene.2022.873850

The *Nudum* (*Nud*) gene controls the caryopsis type of cereal crops by regulating lipid biosynthetic pathways. Based on the *HvNud* sequence and its homologous gene sequences in wheat, a conserved sgRNA was designed to obtain the mutants from the barley variety “Vlamingh” and the wheat variety “Fielder” *via Agrobacterium*-mediated transformation. A total of 19 and 118 transgenic plants were obtained, and 11 and 61 mutant plants were identified in T_0 transgenic plants in barley and wheat after PCR-RE detection, and the editing efficiencies of the targeted gene were 57.9 and 51.7% in barley and wheat, respectively. The grain shape of the barley mutants was naked. Five different combinations of mutations for wheat *TaNud* genes were identified in the T_0 generation, and their homozygous-edited plants were obtained in the T_1 generation. Interestingly, the conjoined plants in which one plant has different genotypes were first identified. The different tillers in an individual T_0 plant showed independent transgenic or mutant events in both barley and wheat, and the different genotypes can stably inherit into T_1 generation, indicating that the T_0 transgenic plants were the conjoined type. In addition, we did not find any off-target mutations in both barley and wheat. A candidate method for detecting putative-edited wheat plants was suggested to avoid losing mutations in this investigation. This study provides not only materials for studying the function of the *Nud* gene in barley and wheat but also a system for detecting the mutants in wheat.

Keywords: nudum (nud) gene, genome editing, agrobacterium-mediated transformation, conjoined plants, naked grain

INTRODUCTION

Genome-editing technologies contain three types of sequence-specific nucleases (SSNs), viz., zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeat-associated endonucleases (CRISPR/Cas) (Filippo et al., 2008; Lieber, 2010; Chapman et al., 2012; Jiang and Doudna, 2017; Mushtaq et al., 2019; Mushtaq et al., 2020; and Mushtaq et al., 2021a). At present, CRISPR/Cas is the most widely used genome-editing system in both animals and plants due to its easy assembling, straightforward guide RNA designing, and high activity (Mushtaq et al., 2021b). The most commonly used type for

CRISPR/Cas is the type II system derived from *Streptococcus pyogenes* (SpCas9) that mainly recognizes the PAM (protospacer adjacent motif) sequence 5'-NGG-3'. To date, the genomes of many plants and crop species such as *Arabidopsis thaliana*, rice (*Oryza sativa*), tobacco (*Nicotiana tabacum* L.), tomato (*Lycopersicon esculentum*), maize (*Zea mays*), wheat (*Triticum aestivum* L.), barley (*Hordeum vulgare* L.), oilseed rape (*Brassica campestris* L), soybean (*Glycine max*), and chickpea (*Cicer arietinum* L.) have been edited using this technique (Li et al., 2013; Brooks et al., 2014; Zhang et al., 2016; Kelliher et al., 2017; Fiaz et al., 2019; Wu et al., 2020; Badhan et al., 2021; Fiaz et al., 2021).

Wheat and barley are two important cereal crops worldwide and are closely associated with social economic development, food production supply, food security, and human health and nutrition (Haas et al., 2019). Therefore, the improvement in the yield, quality, disease resistance, and stress tolerance of wheat and barley using CRISPR/SpCas9 technology is of significant value for the two crops (Abdelrahman et al., 2018). Recently, with the advancement of genetic transformation and the CRISPR/Cas system in wheat, several traits have been genetically modified by introducing mutations in the target genes using CRISPR/SpCas9. Wheat plants with mutations in *TaGW2* showed an increase in grain size (GS) and thousand-grain weight (TGW) (Wang et al., 2018; Zhang et al., 2018). The mutations in the wheat genes *TaMs2* led to the recovery of male sterility (Tang et al., 2021). Knocking-out wheat *TaPLA* or *TaMTL* genes induced haploid plant production (Liu et al., 2019a; Liu et al., 2020a). However, genome editing of barley is more straightforward than that of hexaploid wheat due to only one genome in the former. The editing case of barley *HvPM19* gene, which encodes an ABA-induced plasma membrane protein, was the first application of CRISPR/Cas9 in barley, and the mutants showed a dwarf phenotype (Lawrenson et al., 2015). Barley *Hvckx1* mutations led to reduced root growth and an increased number of tillers and grains (Gasparis et al., 2019). Mutations of *HvMORC1* induced with CRISPR/Cas9 resulted in plants with enhanced resistance against fungal pathogens in barley (Kumar et al., 2018).

Barley can be divided into two types based on the caryopsis: naked and hulled. Most domesticated barley cultivars have caryopses with adhering hulls that are known as hulled barley; some barley cultivars have a free-threshing feature and are called hulless (or naked) barley, especially the cultivars grown in the Tibetan Plateau of China. The caryopsis type in barley is controlled by the transcription factor gene *Nudum* (*Nud*), which encodes a protein in the ethylene response factor (ERF) family located on chromosome arm 7HL (Taketa et al., 2006). The barley *Nud* gene is homologous to the *Arabidopsis WIN1/SHN1* transcription factor gene, which is thought to function in lipid biosynthesis. It has been shown that the caryopsis surface in hulled barley is overlaid with lipid compounds, which penetrate to the inner side of the hull to form the adhesion organ (Taketa et al., 2008).

Studies on X-ray-induced naked mutation alleles (Taketa et al., 2008) and *Nud* locus re-sequencing in 162 barley cultivars (Yu et al., 2016) showed that amino acid substitutions and frame shifts in *Nud* led to loss-of-function and further resulted in the naked phenotype. Moreover, a single nucleotide polymorphism of *HvNud* (T643A), which generated an amino acid substitution of valine (Val, V) by aspartate (Asp, D) at position 148 (Val 148 Asp), could lead to the naked caryopsis type in barley (Yu et al., 2016). There are also three homologous genes for the *Nud* gene in wheat that are located on chromosomes 7AL, 7BL, and 7DL, but their functions are presently unknown. Additionally, *Nud* may also influence other traits in barley. The role of the *Nud* gene has not been fully investigated at present, especially its roles in controlling the naked hull phenotype in barley for efficient breeding of naked barley varieties. Therefore, it is necessary to characterize the function of *Nud* genes in detail in wheat and barley.

In this study, mutants in *HvNud* and *TaNud* were generated by CRISPR/SpCas9 in both barley and wheat. The efficiency and heritability of the mutations were investigated in wheat. Moreover, this is the first report in which we found that different tillers in an individual T_0 plant were independent transgenic or genome-editing events in both barley and wheat, called as conjoined plants. Conjoined plants mean that an individual plant contains different genotyping which is produced possibly from different contiguously transformed cells and developed like one transgenic plant/event. Additionally, potential off-target sites for *HvNud* and *TaNud* in barley and wheat, respectively, were detected and no off-target mutations were found. A method to test candidate-edited wheat plants was put forward to avoid losing mutations in this study. In the present study, the *Nud* gene mutants developed in barley and wheat will be valuable for further investigation of the functions of this gene in the two crops.

MATERIALS AND METHODS

Plant Materials

The barley cultivar 'Vlamingh' and the wheat cultivar 'Fielder' were kindly provided by the National Crop Germplasm Bank, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China. The plants of both species were cultured in an environmental growth chamber at 24°C–16 h light/18°C–8 h dark with a light intensity of 300 μmol m^{-2} s^{-1} at 45% relative humidity.

Designing the sgRNA for the *Nud* Genes

The sequence of the barley *HvNud* gene that determines hulled vs. naked caryopsis was obtained from NCBI (https://www.ncbi.nlm.nih.gov/, Gene Bank accession AP009567.1). The wheat *Nud* genes (*TaNud*) were identified using *HvNud* as a query in a BLAST search of the IWGSCv1 wheat genome (https://urgi.versailles.inra.fr/blast/blast.php). In order to simultaneously edit the *Nud* genes in barley and wheat, a conserved 20-bp sgRNA sequence (5'-CGGCTCCTTGTTGAGCTCGA-3') containing a *SacI* restriction site was selected as the target site for both *HvNud* and *TaNud*. Off-target sites related to the 20-bp sgRNA sequence were predicted by searching the target sequence

in the IBSCv2 barley genome (https://webblast.ipk-gatersleben.de/barley_ibsc/) and the IWGSCv1 wheat genome.

Vector Construction

The full DNA sequence encoding SpCas9 (Ma et al., 2015) was inserted into the expression vector pWMB110 to generate a new plasmid, pWMB110-SpCas9 (Liu et al., 2019a). The wheat *TaU3* promoter was cloned onto plasmid pUC18 as a template, and the sgRNA designed for the *Nud* gene was linked with the *TaU3* cassette by overlapping PCR (Ma et al., 2015). The *TaU3* promoter–sgRNA expression cassette was then amplified and inserted onto the vector pWMB110-SpCas9 at the *MluI* cloning site to generate the recombinant plasmid pWMB110-SpCas9-Nud (**Supplementary Figure S3**). The final vector pWMB110-SpCas9-Nud was transferred into *Agrobacterium* strain C58C1 for transformation of wheat and barley.

Agrobacterium-Mediated Plant Transformation

Immature barley and wheat grains were collected approximately 14 days post anthesis (DPA). The immature grains were sterilized with 75% ethanol for 1 min, followed by 5% sodium hypochlorite for 15 min, and washed five times with sterile water.

Fresh immature embryos of wheat were isolated and transformed by *Agrobacterium*-mediated transformation to generate transgenic plants following the protocol described by Ishida et al. (2015) with slight modifications. In brief, immature wheat embryos were incubated with *Agrobacterium* strain C58C1 harboring the vector for 5 min in a WLS-inf medium at room temperature and co-cultivated for 2 days on the WLS-AS medium with the scutellum facing upward at 25°C in darkness. After co-cultivation, embryonic axes were removed with a scalpel, and the scutella were transferred onto plates containing the WLS-Res medium for delay culture for 5 days under the same conditions. Afterward, tissues were cultured on the WLS-P5 medium with 5 mg L^{-1} phosphinothricin (PPT, Sigma, 45,520) for callus induction. After two weeks, the calli were placed on the WLS-P10 medium with 5 mg L^{-1} PPT for 3 weeks in darkness. The 1/2 MS medium containing 5 mg L^{-1} PPT without zeatin was used for differentiation of embryonic calli other than the LSZ-P5 medium in the methods of Ishida et al. (2015) at 25°C with 100 μmol m^{-2} s^{-1} light. Regenerated shoots were transferred into cups filled with 1/2 MS medium with 5 mg L^{-1} PPT for shoot elongation and root formation.

Barley transformation was performed following the previously published protocols with a slight modification (Bartlett et al., 2008). Immature embryos of barley were isolated after sterilization of the immature grains by the same methods as wheat, subsequently incubated with *Agrobacterium* for 10 min, and co-cultivated for 2 days on CM medium. Then, embryo axes were remove, and the scutella were cultured on the first selection medium with 5 mg L^{-1} PPT. After 2 weeks, tissues were transferred onto the second selection medium with 10 mg L^{-1} PPT. After three weeks, embryonic calli were cultured on the DM medium with 5 mg L^{-1} PPT at 25°C with 100 μmol m^{-2} s^{-1} light for differentiation. Shoots were timely moved into a plastic box containing the RT medium. Last, plants were transplanted into pots filled with soil.

Detection of Transgenic Plants and Edited Mutations

Mixed leaf samples from different tillers of T_0 transgenic plants at the jointing stage were collected for genomic DNA extraction using the CWBIO NuClean Plant Genomic DNA Kit (CWBIO Biotech Co., Ltd.). The *SpCas9* and target genes *HvNud* and *TaNud* in the T_0 transgenic plants were amplified with gene-specific primers (**Supplementary Table S1**) using 2X Taq Master Mix (Vazyme Biotech Co., Ltd.) for positive detection. Two types of primers were employed to amplify the *TaNud* gene: 1) gene-conserved primer pairs designed by the barley *HvNud* gene to simultaneously amplify the homologous wheat *TaNud* genes from the A, B, and D genomes and 2) gene-specific primer pairs designed to amplify the individual wheat gene from each of the three genomes (**Supplementary Table S1**).

Mutations in the *Nud* genes were detected using a polymerase chain reaction-restriction enzyme (PCR-RE) approach. For this test, the amplification reactions were performed in a volume of 20 μl consisting of 2× PCR Mix, 50 ng of genomic DNA, and 0.25 μM of each primer. PCR amplification was performed in a Veriti 96 PCR system (Applied Biosystems) using the following program: an initial denaturation step at 94°C for 5 min, followed by 34 cycles of 94°C for 1 min, 60°C for 45 s and 72°C for 1 min and a final elongation step at 10 min at 72°C. The restriction enzyme digestion of the PCR products was performed in a 20 μl reaction volume containing the appropriate restriction enzyme buffer and 0.2 μl *Sac*I enzyme for 4–6 h at 37°C. The digested products were separated in a 1.5% agarose gel and visualized using a GelDoc XR System (Bio-Rad, United States). To distinguish the different mutant types, the biggest band in the PCR-RE test was directly sequenced for homozygous mutations or indirectly sequenced after cloning into the pMD18-T vector for heterozygous mutations (TaKaRa, Dalian, China) at Sangon Biotech (Shanghai, China). The software BioEdit is used for sequence alignment and analysis (Hall, 2011). The mutations were identified by aligning the sequenced sequences with the referenced sequences of the targeted genes.

RESULTS

Analysis of the *HvNud* Gene in Barley and its Homologs in Wheat

The structure of the barley *HvNud* gene (*HORVU7Hr1G089930*) consists of two exons and one intron, and its ORF encodes a deduced protein of 227 aa (Taketa et al., 2008, **Figure 1**). By using *HvNud* as a query in BLAST searches of the wheat genome database, three orthologous *TaNud* genes (*TraesCS7A02G376300*, *TraesCS7B02G277800*, and *TraesCS7D02G372700*) located on chromosomes 7A, 7B, and 7D were found. The *TaNud* genes have the same gene structure as *HvNud*. All of the proteins predicted from the *Nud* gene sequences contain an AP2/ERF domain, a middle motif, and a C-terminal motif (**Figure 1**).

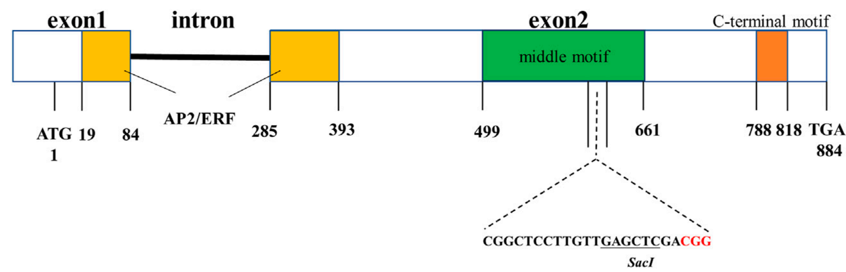


FIGURE 1 | Structure diagram of the *HvNud* gene. Boxes indicate exons, and the black bar between the boxes indicates intron.

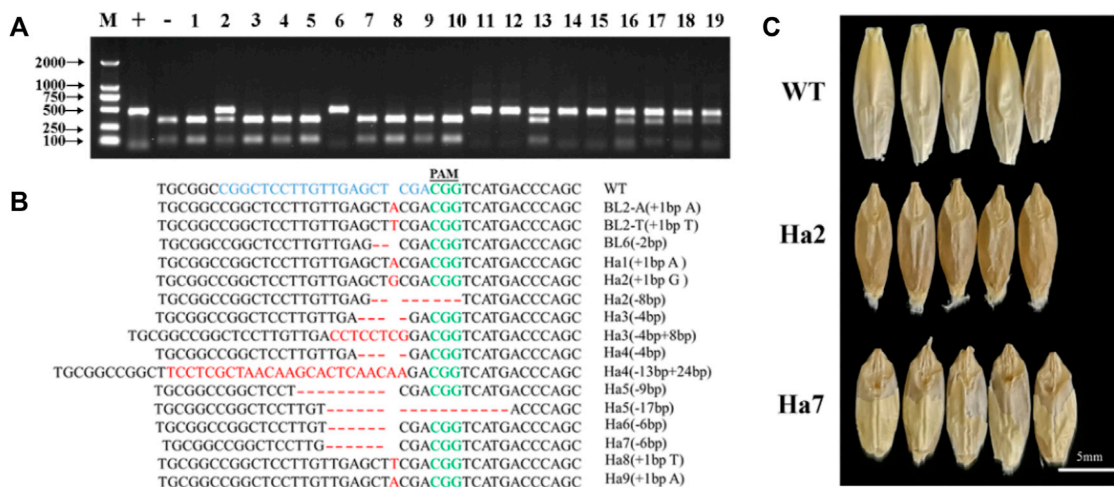


FIGURE 2 | Detection and phenotyping of *HvNud* mutations in transgenic barley plants. **(A)** PCR products of *HvNud* digested with *SacI* restriction enzyme, M: DNA marker; +: positive control; −: negative control; 1–10: B1-BL10; 11–19: Ha1–Ha9. **(B)** InDel mutations in the *HvNud* gene in edited T_0 transgenic barley plants. Nucleotide bases shown in red differ from the target sequence in the sgRNA. **(C)** Phenotype of T_1 grains harvested from mutant plants in comparison to WT grain and T_1 grains collected from *HvNud* knockout T_0 mutants. WT: wild type without threshing; Ha2: frame-shift mutation after threshing; Ha7: 6 bp deletion mutation after threshing.

Moreover, the DNA sequence similarity between the *HvNud* and *TaNud* genes is 87.7%, and their similarity in protein sequence is as high as 94%. A conserved sequence (5'-CGGCTCCTTGTT GAGCTCGA-3', containing a *SacI* restriction site) located in the second exon of both *HvNud* and *TaNud* was selected as the sgRNA for editing.

Detection of *HvNud* Gene-Edited Mutations in T_0 Transgenic Barley Plants

In total, 19 T_0 transgenic barley plants (BL1 to BL10 and Ha1 to Ha9) were generated from two experiments by *Agrobacterium*-mediated transformation using immature embryos. A pair of primers, HvNud-330F and HvNud-816R, was used to amplify the *HvNud* gene from mixed leaves in an individual plant, and the PCR products were then digested with *SacI*. Three types of band patterns (**Figure 2A**) were found in the PCR-RE experiment: heterozygous monoallelic mutants gave three bands, biallelic mutants only gave a band of 487 bp, and non-mutants as well as wild-type (WT) plants gave completely digested bands of 359 bp and 128 bp. A total of 11 mutant plants were obtained (**Figure 2A**), and the editing efficiency in the T_0 transgenic barley plants was 57.9%. A total of five and six mutant plants were confirmed to be biallelic and heterozygous monoallelic mutants in the experiment, respectively (**Supplementary Table S2**).

The undigested PCR products from the mutant plants were directly sequenced, and the results showed that the 1-bp insertion mutation type appeared in the five mutant plants, while there are nine different mutation types in T_0 generation (**Figure 2B**, **Supplementary Table S2**). It is interesting to find three peaks after the 5'-CGGCTCCTTGTTGAGCT-3' target sequence in BL2 (**Figure 3A**), which suggests that there might be three types of the *HvNud* sequence at this nucleotide position. The PCR-amplified DNA fragments from the targeted sequence in BL2 were subcloned onto vector pMD18-T and sequenced. Surprisingly, the sequencing results showed that there were three types of *HvNud* sequence at the target site in BL2, which confirmed the interpretation of the results: the first type has an A nucleotide

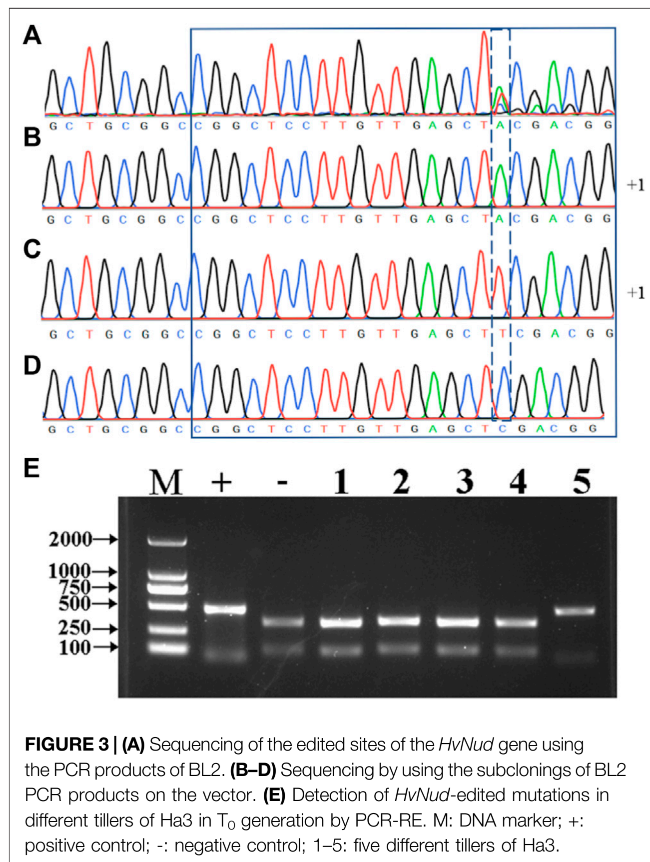


FIGURE 3 | (A) Sequencing of the edited sites of the *HvNud* gene using the PCR products of BL2. **(B–D)** Sequencing by using the subclonings of BL2 PCR products on the vector. **(E)** Detection of *HvNud*-edited mutations in different tillers of Ha3 in T_0 generation by PCR-RE. M: DNA marker; +: positive control; -: negative control; 1–5: five different tillers of Ha3.

insertion (**Figures 2B**, **3B**); the second type has a T nucleotide insertion (**Figures 2B**, **3C**); and the third type remains unchanged from the WT (**Figure 3D**). Unfortunately, we did not harvest seeds from plant BL2. The other mutant plants were randomly selected for mutation detection by tillers. Also, the similar results to BL2 were found in Ha3, in which only one tiller (Ha3-5) was a biallelic mutant and the other four tillers had no mutations after PCR-RE detection (**Figure 3E**). Sequencing results showed that one DNA strand in tiller Ha3-5 had a 4 bp deletion and the other DNA strand had a 4 bp deletion and an 8 bp insertion (−4 bp +8 bp) (**Figure 2B**). These results proved that the different tillers in Ha3 were different independent events.

Inheritance of *HvNud* Mutation Site in T_1 Generation

The T_0 transgenic barley plants numbered Ha2, Ha3, Ha4, Ha6, and Ha7 produced seeds normally. The seeds of Ha2 and Ha4 were naked (**Figure 2C**). But, the seeds of Ha6 and Ha7 were hulled and the same as their WT. In addition, the lemma covering the seeds was difficult to be removed due to 6 bp deletion in the two mutants (**Figures 2B,C**). As different tillers in Ha3 belonged to different genotypes (**Figure 3E**), the seeds in this plant were harvested by tillers. We found that the seeds from Ha3–5 were naked, and the seeds from other tillers were hulled. Moreover, the results by PCR-RE revealed that all the T_1 plants from Ha3–5 were biallelic mutants, while no mutations were found in the descendants of the other four tillers of Ha3 (Ha3-1, Ha3-2, Ha3-3, and Ha3-4) (**Supplementary Figure S1**). Sequencing results confirmed that Ha3-5-3 was a homozygous mutant with 4 bp deletion, while Ha3-5-1 and Ha3-5-2 were heterozygous biallelic mutants with −4 bp/−4 bp +8 bp. These findings were consistent with the detection results of Ha3 in T_0 generation (**Figure 2B**). The aforementioned results confirmed that the different tillers of Ha3 can inherit stably following the Mendelian rule.

Detection of Mutations in Different Tillers in T_0 Transgenic Wheat Plants

Totally, 118 transgenic wheat plants were generated and named from WL1 to WL118. At the outset of the testing, five individual T_0 plants (WL1, 5, 8, 15, and 16) were randomly selected to test the *bar* gene using QuickStix strips and *SpCas9* gene by PCR in different tillers. The results showed that all the tillers from plants WL1, 5, 15, and 16 were positive, while two of 11 tillers on plant WL8 were negative for the *SpCas9* (**Figure 4A**) and *bar* genes (**Figure 4B**). Moreover, both the negative and positive events from the tillers can be steadily detected in T_1 generation. The mutations in the *TaNud* genes were further detected in the four individual plants using the PCR-RE assay and positive detections for *SpCas9* and *bar* genes, and edited mutations for the *TaNud* gene were identified in plants WL5 and WL15. In particular, one tiller showed mutations in genomes A and B, another tiller showed mutations in genome A, and the other eight tillers showed no mutations (**Figure 5**) in plant WL15. These results were consistent with the findings achieved in plant Ha3 in barley, indicating that different tillers in an individual T_0 plant might belong to different independent transgenic or edited events.

Mutation Frequency and Type of *TaNud* Genes in T_0 Transgenic Wheat Plants

Genomic DNA was extracted from the mixed leaf samples of the T_0 transgenic plants and detected for mutations in the three *TaNud* genes on chromosomes 7A, 7B, and 7D by PCR-RE assay using the gene-specific primers (**Figures 6A–C**). Mutations with the *TaNud* genes were detected in 61 T_0 transgenic plants (**Table 1**), and the total editing efficiency of the target genes was 51.7%. In detail, the editing efficiencies for the three *TaNud* genes on chromosomes 7A, 7B, and 7D were 24.6, 33.1, and 8.5%, respectively. The efficiency for the simultaneous mutation of any two genes in a single plant was 14.4% (**Table 1**), and there was no plant identified with simultaneous mutations in the three *TaNud* genes. Theoretically, there were seven combinations (*aaBBDD*, *AAbbDD*, *AABBdd*, *aabbDD*, *AAbbdd*, *aaBBdd*, and *aabbdd*) of biallelic mutations in the three genes in the edited T_0 wheat plants, but only five biallelic mutation types were obtained in this study (**Table 1**).

Application Comparison of the Gene-Conserved and -Specific Primers Used to Detect Edited Plants

A total of nineteen T_0 transgenic wheat plants which carried the five combinations of biallelic mutations in the three *TaNud*

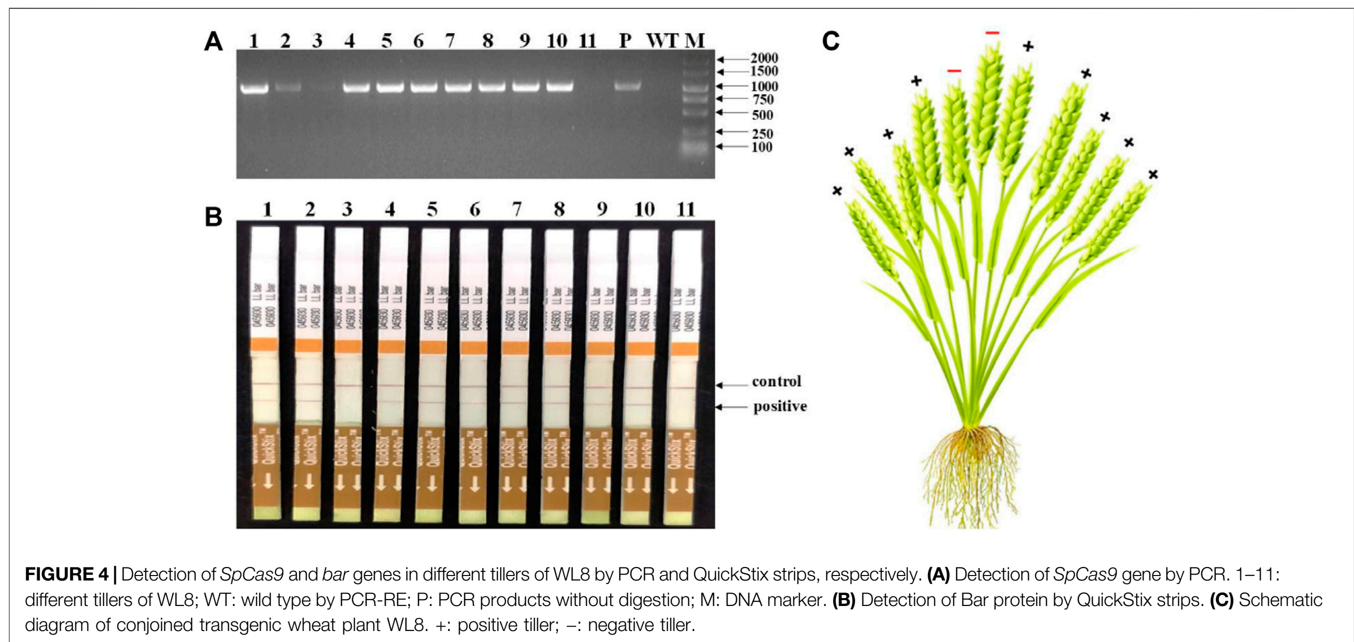


FIGURE 4 | Detection of *SpCas9* and *bar* genes in different tillers of WL8 by PCR and QuickStix strips, respectively. **(A)** Detection of *SpCas9* gene by PCR. 1–11: different tillers of WL8; WT: wild type by PCR-RE; P: PCR products without digestion; M: DNA marker. **(B)** Detection of Bar protein by QuickStix strips. **(C)** Schematic diagram of conjoined transgenic wheat plant WL8. +: positive tiller; −: negative tiller.

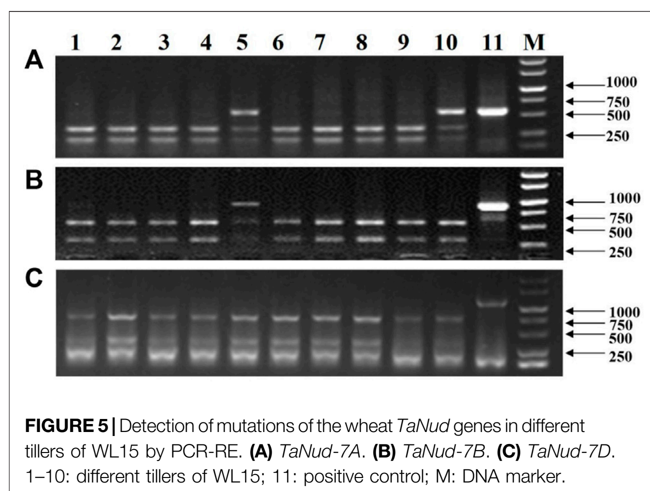


FIGURE 5 | Detection of mutations of the wheat *TaNud* genes in different tillers of WL15 by PCR-RE. **(A)** *TaNud-7A*. **(B)** *TaNud-7B*. **(C)** *TaNud-7D*. 1–10: different tillers of WL15; 11: positive control; M: DNA marker.

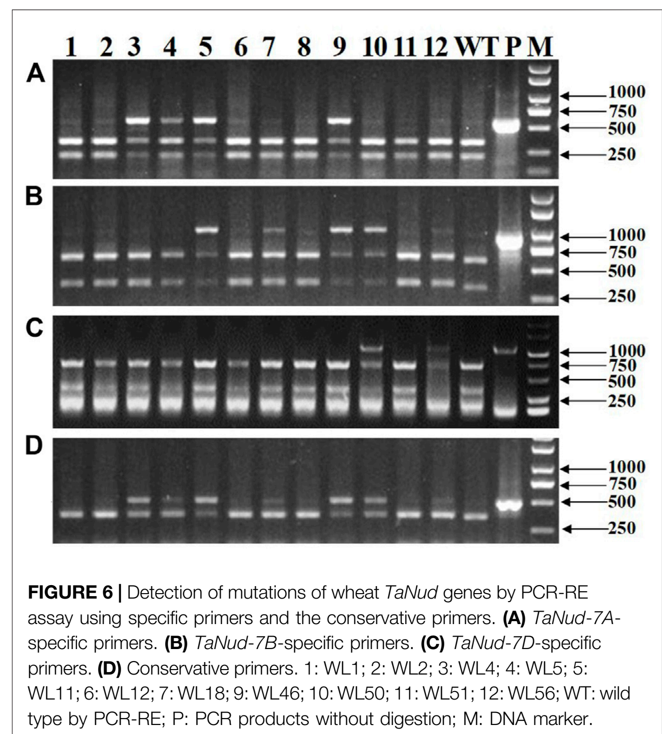


FIGURE 6 | Detection of mutations of wheat *TaNud* genes by PCR-RE assay using specific primers and the conservative primers. **(A)** *TaNud-7A*-specific primers. **(B)** *TaNud-7B*-specific primers. **(C)** *TaNud-7D*-specific primers. **(D)** Conservative primers. 1: WL1; 2: WL2; 3: WL4; 4: WL5; 5: WL11; 6: WL12; 7: WL18; 9: WL46; 10: WL50; 11: WL51; 12: WL56; WT: wild type by PCR-RE; P: PCR products without digestion; M: DNA marker.

genes were used to compare the gene-conserved primers (HvNud-330F and HvNud-816R) and gene-specific primers for the three alleles by PCR-RE assay. The results using the conserved primers (**Figure 6D**) showed the mutated DNA fragments as long as there was a mutation in any one of the three genes on chromosomes 7A, 7B, and 7D (**Figures 6A–C**). From this result, we can infer that the conserved primers are able to detect all of the mutations in the three different *TaNud* genes. Therefore, the conserved primers can be used to quickly screen the transgenic plants, and a detailed confirmation of the mutation types can then be performed using the specific primers from the plants which contained mutations. This approach is necessary to be adapted when large groups of transgenic plants are obtained, or the editing efficiency in an experiment is low.

Inheritance of *TaNud* Mutation Sites in T_1-Edited Wheat Plants

Genetic segregation of the mutations in the *TaNud* genes was detected in the T_1 generation by PCR-RE using gene-specific primers and DNA sequencing. For this purpose, five T_0 plants (WL4, WL11, WL18, WL50, and WL56), in which each plant belonged to a different mutation-type category, were selected for

TABLE 1 | Summary of the mutations of the *TaNud* genes in T_0-edited wheat plants.

Mutation allele	Mutant plant	Mutation rate (%)	Plant ID
7A	17	14.4	WL–4, 5, 10, 14, 20, 40, 48, 52, 75, 89, 92, 105, 106, 107, 110, 114, and 117
7B	22	18.6	WL–18, 21, 23, 32, 33, 36, 43, 53, 57, 59, 62, 63, 68, 76, 77, 80, 84, 90, 94, 96, 97, and 118
7D	5	4.2	WL–56, 67, 69, 95, and 109
7A and 7B	12	10.2	WL–11, 15, 17, 24, 27, 38, 46, 66, 74, 104, 113, and 115
7B and 7D	5	4.2	WL–50, 60, 82, 91, and 101

Lower case letters represent mutant types; capital letters represent wild type.

TABLE 2 | Summary of mutations in the *TaNud* genes in T_1-edited wheat plants.

T_0 line	T_0-edited allele	T_1-edited type		
		Homozygous	Heterozygous	WT
WL4	aaBBDD	3	3	4
WL18	AAbbDD	4	5	1
WL56	AABBdd	4	3	3
WL11	aabbDD	2	8	0
WL50	AAbbdd	1	9	0

```
                          PAM
 GCCGGCTCCTTGTTGAGCTCGACGGTCATGACCCA   TaNud-WT
GCCGGCTCCTTGTTGAGCTTCGACGGTCATGACCCA   WL4(TaNud-7A)  +1bp
GCCGGCTCCTTGTTGAGCTTCGACGGTCATGACCCA   WL11(TaNud-7A) +1bp
 GCCGGCTCCTTGTTGAGCT– GACGGTCATGACCCA  WL11(TaNud-7B) -1bp
 GCCGGCTCCTTGTTGAGCT– GACGGTCATGACCCA  WL18(TaNud-7B) -1bp
 GCCGGCTCCTTGTTGAGCT– GACGGTCATGACCCA  WL50(TaNud-7B) -1bp
  GCCGGCTCCTTGTTGAGC– ––––––––––GACCCA  WL50(TaNud-7D) -11bp
 GCCGGCTCCTTGTTGAGCTTGACGGTCATGACCCA   WL56(TaNud-7D) C→T
```

FIGURE 7 | InDel mutations in *TaNud* genes in the edited T_1 transgenic wheat plants.

further characterization in the next generation. A total of ten T_1 plants were detected in each of the edited lines. The frequencies of homozygous mutants were 30, 40, and 40% for the edited types *aaBBDD*, *AAbbDD*, and *AABBdd*, respectively (**Table 2**). The frequencies of simultaneous homozygous mutation plants at two loci were 20 and 10% for *aabbDD* and *aabbDD*, respectively (**Table 2**). The segregation ratios of the edited sites in the T_1 generation plants did not follow the Mendelian heritance pattern, which might be either due to the T_1 population being small or their T_0 generations being conjoined. In total, five types of homozygous mutations were identified in the T_1 plants, and DNA sequencing showed that the mutations in the *TaNud* genes included small nucleotide insertions, deletions, and substitutions (**Figure 7**, **Supplementary Table S2**).

Detection of Off-Target Mutations

The presence of potential off-target mutated sites in barley and wheat was further analyzed. Two putative off-target sites which had three or five SNPs compared with the target sequences were predicted for each of *HvNud* and *TaNud* genes from the IBSCv2 barley genome and IWGSCv1 wheat genome by a BLAST search (**Supplementary Table S3**). Specific PCR primers were designed to amplify the potential off-target regions (**Supplementary Table S3**) in two T_0 barley transgenic plants and 61 T_0 wheat transgenic plants. The PCR-RE results verified that there were no off-target mutations happened in our present experiments. Therefore, the CRISPR/Cas9 system precisely targeted the selected sites in the *Nud* gene sequences of the two crops in this study.

DISCUSSION

Compared to other techniques used to induce mutations, such as ethyl methyl sulfonate (EMS) treatment and ray irradiation, the CRISPR/Cas9 system is much precise and efficient in generating specific mutations. Therefore, CRISPR/Cas9 has been the common choice to induce mutations for target gene function analysis and crop improvement because the genetic transformation efficiency of wheat and barley has been significantly improved (Bartlett et al., 2008; Wang et al., 2016); the application of CRISPR/Cas9 will also be more widespread in these two important crops. However, since wheat is a hexaploid plant, most wheat genes are present in at least three copies, and the difficulty of editing multiple target genes simultaneously may limit the application of CRISPR/Cas9 in wheat.

Mutant Induction of the *Nud* Gene by CRISPR/Cas9 in Barley and Wheat

In this study, five biallelic mutant plants and six monoallelic heterozygous mutant plants were obtained for editing the barley *HvNud* gene with an editing efficiency of 57.9%. Biallelic mutation efficiency was up to 26.3%, and the naked grain phenotype was observed in three T_1 lines Ha2, Ha3, and Ha4. Wheat has always lagged behind other cereal crops with respect to the applications of genetic modification technologies due to its complex polyploid genome. Although there were many studies on the use of CRISPR/Cas9 in wheat, there were few reports describing the simultaneous mutation of target genes located on A, B, and D genomes (Wang et al., 2020). With the optimization on vector construction and the improvement on editing efficiency by the CRISPR/Cas9 system for wheat, the application of genome-editing technology will soon become routine in wheat. By using the optimized CRISPR/Cas9 system in wheat (Liu et al., 2019a), the editing efficiency of *TaWaxy* and *TaMTL* genes in T_0 plants reached 80.5 and 57.5%, respectively. In our current study, the editing efficiency of the three *TaNud* alleles also reached 51.7%, but the editing efficiency of *TaNud-7D* was only 8.5%. The low editing efficiency of *TaNud-7D* resulted in failure to obtain plants carrying mutations in all three *TaNud*

genes on the three wheat genomes. Therefore, the same target site might lead to different editing efficiencies for different homoeologous genes. It is also possible that the editing efficiency for genome D is lower than that for genomes A and B due to the special structure in the target region of genome D in wheat. In a word, CRISPR/Cas9 can be employed to accurately generate targeted mutations for crop improvement.

Potential Function of the *HvNud* Mutants by CRISPR/Cas9

Generally, the hulled barley is used for distilling, brewing, and animal feed, and the naked barley is an important food source in Asia and northern Africa. Compared with the hulled barley, the naked barley is easy to separate from its outer glume, which is convenient for processing and eating. Moreover, the naked barley has a high protein level, high β-glucan, some rare nutrients, and trace element contents (Östman et al., 2006). The barley β-glucan can reduce blood LDL-cholesterol and visceral fat obesity (Tiwari and Cummins, 2011; Aoe et al., 2017). Therefore, the mutants of *HvNud* gene could not only convert the hulled germplasm into naked barley for dietary purposes but also increase the nutrition value in barley grains.

Identification of Conjoined T_0 Transgenic Barley and Wheat Plants

In general, an individual T_0 transgenic plant has been thought to represent a single transgenic event. Previously, the transgenic seedlings arising from the same embryo were even considered to represent the same transgenic event. In a previous study, different phenotypes for editing the wheat *TaQ* gene were observed in different tillers of a single transgenic plant (Liu et al., 2020b). In our present study, different independent transgenic events were first identified in different tillers of an individual plant and could inherit into next generations in barley and wheat. In T_0 barley plants, Ha3 displayed two different genotypes in different tillers (**Figure 3E**), and the two genotypes could stably inherit in T_1 generation, respectively (**Supplementary Figure S1**). Theoretically, all of the transgenic plants should be positive because the transformants were rigorously screened by the selective agent. In fact, negative plants are often detected in T_0 populations. In this study, two of 11 tillers on plant WL8 were transgene negative (**Figure 4A**). It is very interesting that the middle two tillers were negative in this plant (**Figure 4C**). Moreover, the foreign integrating elements in positive plants can be stably inherited by tillers. In summary, Ha3 and WL8 were conjoined plants, not chimeric plants. Mosaic or chimeric plants are normally generated when some tissues are positive and other tissues are negative for the transgenes in a plant. Thereby, the transgenes or mutations in mosaic or chimeric plants cannot stably inherit. We speculated that a negative transformant and a positive transformant are tightly grown together like conjoined babies in humans, and the negative transformant can be survived by the resistance of the positive transformant to the selection pressure in the medium. Conjoined plants were originated from different cells, but the different cells are too close to separate during transformation, so the conjoined plants are generated. Just like a conjoined baby, the conjoined plants are independent individuals although they grow combined. Therefore, this is the reason that positive and negative tillers can be detected in a single plant. The conjoined plants such as WL15 from wheat and BL2 and Ha3 from barley in this study also led to different transgenic events.

When the T_0 transgenic plant in the genome-editing experiment was shown to be a conjoined plant that contained different independent transgenic events (**Figure 4C**), the T_0 transgenic plants should be screened for edited mutations by detecting the individual tillers, and this approach could be too labor-intensive. Thus, mixed leaf samples can be used when screening T_{0}-edited plants to avoid missing targeted mutations.

System for Detecting Mutation in Different Wheat Genomes

Wheat is a complex allopolyploid plant harboring three similar genomes, and the most homoeologous genes have very small sequence differences on the A, B, and D genomes. This fact makes genome editing and the subsequent mutation detection extremely challenging in wheat. Current methods for detecting DNA sequence mutations induced by genome editing include PCR-RE (Shan et al., 2014), PCR/RNP (Liang et al., 2018), the T7EI cleavage assay (Vouillot et al., 2015), next generation sequencing (NGS) (Liu et al., 2019b), high-resolution melting analysis (HRMA) (Dahlem et al., 2012), and fluorescent PCR-capillary gel electrophoresis (Ramlee et al., 2015). Even though each method has its shortcomings, PCR-RE and Sanger sequencing are the most direct, convenient, and widely used methods in many laboratories. Moreover, PCR-RE is the best, efficient, and most cost-effective method when a restriction enzyme site exists in the target sequences. PCR-RE can steadily identify heterozygous mutants, biallelic mutants, and un-mutated WT sequences. Based on our experiences detecting wheat mutations induced by CRISPR/Cas9, a detecting system was suggested as follows: the sgRNA was designed with an incorporated restriction enzyme recognition site. For most genes, a target site with a restriction enzyme site can be found. Mixing leaf samples from different tillers of single plants in the T_0 generation were collected for DNA extraction and PCR detection. Normally, individual conjoined plants in T_0 generation might be homozygous in T_1 generation. Mixing samples can help to avoid missing the edited mutants in T_0 generation. Generally, different primer pairs specific to the homologs on A, B, and D genomes should be used. However, when the mutation frequency is low and/or a large population needs to be screened, conserved primers that can amplify the target regions from the three homologous genes can be used for the initial screening, and then gene-specific primers can be used for mutation detection. When the mutation plants are heterozygous, the detecting results by PCR-RE will show three binds and the largest fragment which is of the same size to the undigested PCR product can be sequenced; when the mutation plants are biallelic, the PCR-RE results will show only one bind, in which the product can be directly sequenced, and the sequence can be further analyzed at http://skl.scau.edu.cn/dsdecode/; in

either way, the PCR-RE product can be ligated into a T-vector and then sequenced. When there is no available restriction enzyme site present in the sgRNA sequence, gene-specific primers can be used to first amplify the target regions, and DNA sequencing is followed to determine whether a mutation is present by examining the overlapping peaks in the sgRNA sequence. Finally, the sequencing results were analyzed at http://skl.scau.edu.cn/dsdecode/, or the PCR product is ligated into a vector to be sequenced.

CONCLUSION

In this study, we created barley and wheat mutations for the *Nud* gene by CRISPR/Cas9 and provided materials for studying the functions of the target gene in the two crops. The editing efficiencies for the *Nud* gene in barley and wheat were 57.9 and 51.7%, respectively. The biallelic mutant barley plants showed a naked phenotype. A total of five types of homozygous wheat mutation plants for *TaNud* genes were obtained in the T_1 generation, especially we identified conjoined plants in which different tillers in a T_0 individual plant were independent transgenic or genome-editing events in barley and wheat, and different genotypes in different tillers could inherit in the T_1 generation. Therefore, the transgenic or edited wheat and barley plants need to be detected by mixed leaf samples in the T_0 generation in case of missing some desired events, which might be a candidate method for detecting edited wheat plants to avoid the loss of possible mutations. The mutants of the *HvNud* gene could not only convert the hulled germplasm into naked barley for dietary purposes but also increase the nutrition value in barley grains.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

YZ, QG, and NL constructed the vectors and generated the transgenic lines. YZ, QG, and HB analyzed the mutations in the target genes in the transgenic lines. KW initiated the project and designed sgRNAs. HL generated some of the transgenic lines. YX, LD, CL, KW, and XY analyzed the data and revised the manuscript. All authors discussed the results and approved the manuscript.

FUNDING

This research was financially supported in part by grants from the National Natural Science Foundation of China (31971946 and 31971945) and the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (S2022ZD03) and the Chinese Academy of Agricultural Sciences in China (2060302-2-19).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.873850/full#supplementary-material

REFERENCES

Abdelrahman, M., Al-Sadi, A. M., Pour-Aboughadareh, A., Burritt, D. J., and Tran, L.-S. P. (2018). Genome Editing Using CRISPR/Cas9-targeted Mutagenesis: An Opportunity for Yield Improvements of Crop Plants Grown under Environmental Stresses. *Plant Physiol. Biochem.* 131, 31–36. doi:10.1016/j.plaphy.2018.03.012

Aoe, S., Ichinose, Y., Kohyama, N., Komae, K., Takahashi, A., Abe, D., et al. (2017). Effects of High β-glucan Barley on Visceral Fat Obesity in Japanese Individuals: A Randomized, Double-Blind Study. *Nutrition* 42, 1–6. doi:10.1016/j.nut.2017.05.002

Badhan, S., Ball, A. S., and Mantri, N. (2021). First Report of CRISPR/Cas9 Mediated DNA-free Editing of *4CL* and *RVE7* Genes in Chickpea Protoplasts. *Int. J. Mol. Sci.* 22, 396. doi:10.3390/ijms22010396

Bartlett, J. G., Alves, S. C., Smedley, M., Snape, J. W., and Harwood, W. A. (2008). High-throughput *Agrobacterium*-Mediated Barley Transformation. *Plant Methods* 4, 22. doi:10.1186/1746-4811-4-22

Brooks, C., Nekrasov, V., Lippman, Z. B., and Van Eck, J. (2014). Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic repeats/CRISPR-Associated9 System. *Plant Physiol.* 166, 1292–1297. doi:10.1104/pp.114.247577

Chapman, J. R., Taylor, M. R. G., and Boulton, S. J. (2012). Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Mol. Cel* 47, 497–510. doi:10.1016/j.molcel.2012.07.029

Dahlem, T. J., Hoshijima, K., Jurynec, M. J., Gunther, D., Starker, C. G., Locke, A. S., et al. (2012). Simple Methods for Generating and Detecting Locus-specific Mutations Induced with TALENs in the Zebrafish Genome. *Plos Genet.* 8, e1002861. doi:10.1371/journal.pgen.1002861

Fiaz, S., Ahmad, S., Noor, M., Wang, X., Younas, A., Riaz, A., et al. (2019). Applications of the CRISPR/Cas9 System for Rice Grain Quality Improvement: Perspectives and Opportunities. *Int. J. Mol. Sci.* 20 (4), 888. doi:10.3390/ijms20040888

Fiaz, S., Wang, X., Khan, S. A., Ahmar, S., Noor, M. A., Riaz, A., et al. (2021). Novel Plant Breeding Techniques to advance Nitrogen Use Efficiency in rice: A Review. *GM Crops Food* 12, 627–646. doi:10.1080/21645698.2021.1921545

Filippo, J. S., Sung, P., and Klein, H. (2008). Mechanism of Eukaryotic Homologous Recombination. *Annu. Rev. Biochem.* 77, 229–257. doi:10.1146/annurev.biochem.77.061306.125255

Gasparis, S., Przyborowski, M., Kała, M., and Nadolska-Orczyk, A. (2019). Knockout of the HvCKX1 or HvCKX3 Gene in Barley (*Hordeum Vulgare* L.) by RNA-Guided *Cas9* Nuclease Affects the Regulation of Cytokinin Metabolism and Root Morphology. *Cells* 8, 782. doi:10.3390/cells8080782

Haas, M., Schreiber, M., and Mascher, M. (2019). Domestication and Crop Evolution of Wheat and Barley: Genes, Genomics, and Future Directions. *J. Integr. Plant Biol.* 61, 204–225. doi:10.1111/jipb.12737

Hall, T. (2011). BioEdit: an Important Software for Molecular Biology. *GERF Bull. Biosci.* 2, 60–61.

Ishida, Y., Tsunashima, M., Hiei, Y., and Komari, T. (2015). Wheat (*Triticum aestivum* L.) Transformation Using Immature Embryos. *Methods Mol. Biol.* 1223, 189–198. doi:10.1007/978-1-4939-1695-5_15

Jiang, F., and Doudna, J. A. (2017). CRISPR-Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* 46, 505–529. doi:10.1146/annurev-biophys-062215-010822

Kelliher, T., Starr, D., Richbourg, L., Chintamanani, S., Delzer, B., Nuccio, M. L., et al. (2017). MATRILINEAL, a Sperm-specific Phospholipase, Triggers maize Haploid Induction. *Nature* 542, 105–109. doi:10.1038/nature20827

Kumar, N., Galli, M., Ordon, J., Stuttmann, J., Kogel, K.-H., and Imani, J. (2018). Further Analysis of Barley MORC1 Using a Highly Efficient RNA-Guided Cas9 Gene-Editing System. *Plant Biotechnol. J.* 16, 1892–1903. doi:10.1111/pbi.12924

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Østergaard, L., Patron, N., et al. (2015). Induction of Targeted, Heritable Mutations in Barley and *Brassica oleracea* Using RNA-Guided Cas9 Nuclease. *Genome Biol.* 16, 258. doi:10.1186/s13059-015-0826-7

Li, J.-F., Norville, J. E., Aach, J., McCormack, M., Zhang, D., Bush, J., et al. (2013). Multiplex and Homologous Recombination-Mediated Genome Editing in *Arabidopsis* and *Nicotiana Benthamiana* Using Guide RNA and Cas9. *Nat. Biotechnol.* 31, 688–691. doi:10.1038/nbt.2654

Liang, Z., Chen, K., Yan, Y., Zhang, Y., and Gao, C. (2018). Genotyping Genome-Edited Mutations in Plants Using CRISPR Ribonucleoprotein Complexes. *Plant Biotechnol. J.* 16, 2053–2062. doi:10.1111/pbi.12938

Lieber, M. R. (2010). The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annu. Rev. Biochem.* 79, 181–211. doi:10.1146/annurev.biochem.052308.093131

Liu, H., Wang, K., Jia, Z., Gong, Q., and Ye, X. (2019a). Editing *TaMTL* Gene Induces Haploid Plants Efficiently by Optimized Agrobacterium-Mediated CRISPR System in Wheat. *J. Exp. Bot.* 71, 4. doi:10.1093/jxb/erz529

Liu, Q., Wang, C., Jiao, X., Zhang, H., Song, L., Li, Y., et al. (2019b). Hi-TOM: a Platform for High-Throughput Tracking of Mutations Induced by CRISPR/Cas Systems. *Sci. China Life Sci.* 62, 1–7. doi:10.1007/s11427-018-9402-9

Liu, C., Zhong, Y., Qi, X., Chen, M., Liu, Z., Chen, C., et al. (2020a). Extension of the *In Vivo* Haploid Induction System from Diploid maize to Hexaploid Wheat. *Plant Biotechnol. J.* 18, 316–318. doi:10.1111/pbi.13218

Liu, H., Wang, K., Tang, H., Gong, Q., Du, L., Pei, X., et al. (2020b). CRISPR/Cas9 Editing of Wheat *TaQ* Genes Alters Spike Morphogenesis and Grain Threshability. *J. Genet. Genomics* 47, 563–575. doi:10.1016/j.jgg.2020.08.004

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., et al. (2015). A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants. *Mol. Plant* 8, 1274–1284. doi:10.1016/j.molp.2015.04.007

Mushtaq, M., Sakina, A., Wani, S. H., Shikari, A. B., Tripathi, P., Zaid, A., et al. (2019). Harnessing Genome Editing Techniques to Engineer Disease Resistance in Plants. *Front. Plant Sci.* 10, 550. doi:10.3389/fpls.2019.00550

Mushtaq, M., Mukhtar, S., Sakina, A., Dar, A. A., Bhat, R., Deshmukh, R., et al. (2020). Tweaking Genome-Editing Approaches for Virus Interference in Crop Plants. *Plant Physiol. Biochem.* 147, 242–250. doi:10.1016/j.plaphy.2019.12.022

Mushtaq, M., Bhat, J. A., Mir, Z. A., Sakina, A., Ali, S., Singh, A. K., et al. (2021a). CRISPR/Cas Approach: A New Way of Looking at Plant-Abiotic Interactions. *J. Plant Physiol.* 224-225, 156–162. doi:10.1016/j.jplph.2018.04.001

Mushtaq, M., Ahmad Dar, A., Skalicky, M., Tyagi, A., Bhagat, N., Basu, U., et al. (2021b). CRISPR-based Genome Editing Tools: Insights into Technological Breakthroughs and Future Challenges. *Genes* 12, 797. doi:10.3390/genes12060797

Östman, E., Rossi, E., Larsson, H., Brighenti, F., and Björck, I. (2006). Glucose and Insulin Responses in Healthy Men to Barley Bread with Different Levels of (1→3;1→4)-β-Glucans; Predictions Using Fluidity Measurements of *In Vitro* Enzyme Digests. *J. Cereal Sci.* 43, 230–235. doi:10.1016/j.jcs.2005.11.001

Ramlee, M. K., Yan, T., Cheung, A. M. S., Chuah, C. T. H., and Li, S. (2015). High-throughput Genotyping of CRISPR/Cas9-mediated Mutants Using Fluorescent PCR-Capillary Gel Electrophoresis. *Sci. Rep.* 5, 15587. doi:10.1038/srep15587

Shan, Q., Wang, Y., Li, J., and Gao, C. (2014). Genome Editing in Rice and Wheat Using the CRISPR/Cas System. *Nat. Protoc.* 9, 2395–2410. doi:10.1038/nprot.2014.157

Taketa, S., Awayama, T., Amano, S., Sakurai, Y., and Ichii, M. (2006). High-resolution Mapping of the *Nud* Locus Controlling the Naked Caryopsis in Barley. *Plant Breed.* 125, 337–342. doi:10.1111/j.1439-0523.2006.01207.x

Taketa, S., Amano, S., Tsujino, Y., Sato, T., Saisho, D., Kakeda, K., et al. (2008). Barley Grain with Adhering Hulls Is Controlled by an ERF Family Transcription Factor Gene Regulating a Lipid Biosynthesis Pathway. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4062–4067. doi:10.1073/pnas.0711034105

Tang, H., Liu, H., Zhou, Y., Liu, H., Du, L., Wang, K., et al. (2021). Fertility Recovery of Wheat Male Sterility Controlled by Ms2 Using CRISPR/Cas9. *Plant Biotechnol. J.* 19, 224–226. doi:10.1111/pbi.13482

Tiwari, U., and Cummins, E. (2011). Meta-analysis of the Effect of β-glucan Intake on Blood Cholesterol and Glucose Levels. *Nutrition* 27, 1008–1016. doi:10.1016/j.nut.2010.11.006

Vouillot, L., Thélie, A., and Pollet, N. (2015). Comparison of T7E1 and Surveyor Mismatch Cleavage Assays to Detect Mutations Triggered by Engineered Nucleases. *G3 (Bethesda)* 5, 407–415. doi:10.1534/g3.114.015834

Wang, K., Liu, H., Du, L., and Ye, X. (2016). Generation of Marker-free Transgenic Hexaploid Wheat via anAgrobacterium-Mediated Co-transformation Strategy in Commercial Chinese Wheat Varieties. *Plant Biotechnol. J.* 15, 614–623. doi:10.1111/pbi.12660

Wang, W., Pan, Q., He, F., Akhunova, A., Chao, S., Trick, H., et al. (2018). Transgenerational CRISPR-Cas9 Activity Facilitates Multiplex Gene Editing in Allopolyploid Wheat. *CRISPR J.* 1, 65–74. doi:10.1089/crispr.2017.0010

Wang, K., Gong, Q., and Ye, X. (2020). Recent Developments and Applications of Genetic Transformation and Genome Editing Technologies in Wheat. *Theor. Appl. Genet.* 133, 1603–1622. doi:10.1007/s00122-019-03464-4

Wu, J., Chen, C., Xian, G., Liu, D., Lin, L., Yin, S., et al. (2020). Engineering Herbicide-resistant Oilseed Rape by CRISPR/Cas9-mediated Cytosine Base-editing. *Plant Biotechnol. J.* 18, 1857–1859. doi:10.1111/pbi.13368

Yu, S., Long, H., Deng, G., Pan, Z., Liang, J., Zeng, X., et al. (2016). A Single Nucleotide Polymorphism of *Nud* Converts the Caryopsis Type of Barley (*Hordeum Vulgare* L.). *Plant Mol. Biol. Rep.* 34, 242–248. doi:10.1007/s11105-015-0911-9

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., et al. (2016). Efficient and Transgene-free Genome Editing in Wheat through Transient Expression of CRISPR/Cas9 DNA or RNA. *Nat. Commun.* 7, 12617. doi:10.1038/ncomms12617

Zhang, Y., Li, D., Zhang, D., Zhao, X., Cao, X., Dong, L., et al. (2018). Analysis of the Functions ofTaGW2homoeologs in Wheat Grain Weight and Protein Content Traits. *Plant J.* 94, 857–866. doi:10.1111/tpj.13903

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zang, Gong, Xu, Liu, Bai, Li, Du, Ye, Lan and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 10 May 2022
doi: 10.3389/fgene.2022.876522

Genome-Wide Association Mapping Reveals Novel Putative Gene Candidates Governing Reproductive Stage Heat Stress Tolerance in Rice

K. T. Ravikiran[1†], S. Gopala Krishnan[1]*, K. P. Abhijith[1], H. Bollinedi[1], M. Nagarajan[2], K. K. Vinod[1], P. K. Bhowmick[1], Madan Pal[3], R. K. Ellur[1] and A. K. Singh[1]

[1]Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi, India, [2]Rice Breeding and Genetics Research Centre, ICAR-IARI, Aduthurai, India, [3]Division of Plant Physiology, ICAR-Indian Agricultural Research Institute, New Delhi, India

OPEN ACCESS

Edited by:
Deepmala Sehgal,
International Maize and Wheat Improvement Center, Mexico

Reviewed by:
Longbiao Guo,
China National Rice Research Institute (CAAS), China
Vandana Jaiswal,
Institute of Himalayan Bioresource Technology (CSIR), India

***Correspondence:**
S. Gopala Krishnan
gopal_icar@yahoo.co.in

†Present address:
K. T. Ravikiran,
ICAR-Central Soil Salinity Research Institute, Regional Research Station, Lucknow, India

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 15 February 2022
Accepted: 25 March 2022
Published: 10 May 2022

Citation:
Ravikiran KT, Gopala Krishnan S, Abhijith KP, Bollinedi H, Nagarajan M, Vinod KK, Bhowmick PK, Pal M, Ellur RK and Singh AK (2022) Genome-Wide Association Mapping Reveals Novel Putative Gene Candidates Governing Reproductive Stage Heat Stress Tolerance in Rice. Front. Genet. 13:876522. doi: 10.3389/fgene.2022.876522

Temperature rise predicted for the future will severely affect rice productivity because the crop is highly sensitive to heat stress at the reproductive stage. Breeding tolerant varieties is an economically viable option to combat heat stress, for which the knowledge of target genomic regions associated with the reproductive stage heat stress tolerance (RSHT) is essential. A set of 192 rice genotypes of diverse origins were evaluated under natural field conditions through staggered sowings for RSHT using two surrogate traits, spikelet fertility and grain yield, which showed significant reduction under heat stress. These genotypes were genotyped using a 50 k SNP array, and the association analysis identified 10 quantitative trait nucleotides (QTNs) for grain yield, of which one QTN (*qHTGY8.1*) was consistent across the different models used. Only two out of 10 MTAs coincided with the previously reported QTLs, making the remaing eight novel. A total of 22 QTNs were observed for spikelet fertility, among which *qHTSF5.1* was consistently found across three models. Of the QTNs identified, seven coincided with previous reports, while the remaining QTNs were new. The genes near the QTNs were found associated with the protein–protein interaction, protein ubiquitination, stress signal transduction, and so forth, qualifying them to be putative for RSHT. An *in silico* expression analysis revealed the predominant expression of genes identified for spikelet fertility in reproductive organs. Further validation of the biological relevance of QTNs in conferring heat stress tolerance will enable their utilization in improving the reproductive stage heat stress tolerance in rice.

Keywords: rice, GWAS, marker-trait association, quantitative trait nucleotides, reproductive stage heat stress tolerance

INTRODUCTION

Climate change and global warming are seriously affecting agricultural productivity. The mean surface temperature of the earth today is 0.8°C higher than that in the pre-industrial era and is projected to increase up to 4.8°C by the end of this century (IPCC, 2014). In India also, there has been a concurrent increase of 0.63°C since 1986, triggering intermittent heatwaves, and this is predicted to increase to 4.7°C by the end of 2100. The heatwave episodes are expected to intensify particularly in the Indo-Gangetic plains of India, where rice–wheat is the most prevalent cropping system (Im et al., 2017; Krishnan et al., 2020). Rice is the major staple food crop of India with an area of 43.79 million

hectares (mha) and a production of 116.42 million tons (mt) during 2018–19 (GoI, 2019). It is highly sensitive to heat stress at the reproductive stage with the optimum temperature ranging from 22 to 28°C (Prasad et al., 2006). Temperatures beyond the tolerant threshold (35°C) at anthesis and booting will adversely affect rice yields (Satake and Yoshida, 1978; Yoshida et al., 1981). Simulation models have predicted that for every 1°C rise in ambient temperature during the sensitive stages, the rice yield will suffer by 2.5 up to 10% (Baker et al., 1992; Peng et al., 2004). The temperatures of many tropical rice-growing countries have already reached ~33°C, and any further increase will have severe taxing on the grain yield and quality (Wassmann et al., 2009). For instance, a heatwave (~36°C) for 2 consecutive days in early April 2021 coupled with low rainfall and humidity has devastated the 68,000 acres of a spring-grown rice crop of Bangladesh with an estimated economic loss of $39 million (Hossain, 2021).

Heat stress is one of the complex abiotic stresses, to which plants respond through an intricate network of signal transduction pathways (González-Schain et al., 2016; Wang et al., 2019). In rice, heat stress at the reproductive stage affects pollen viability and spikelet fertility, thereby reducing the grain yield. Pollen viability is reduced primarily due to pollen desiccation and denaturation of proteins. Spikelet sterility is attributed to poor anther dehiscence and reduced pollen production, reducing the number of viable pollens reaching the stigma (Matsui et al., 2000, 2001; Prasad et al., 2006). Additionally, tight closure of anther locules by cell layers could also hinder anther dehiscence (Matsui and Omasa, 2002). This is further exacerbated by impaired stigma receptivity due to its exertion out of the spikelet into hot ambience (Wu and Yang, 2019). However, the female reproductive organ is comparatively more resilient to heat stress compared to the male counterpart. Post-fertilization, grain filling, and maturation are equally sensitive to heat stress. High temperature reduces the grain filling period but hastens the grain maturation rate, leading to the impairment of grain filling and resulting in poorly filled chalky grains. These grains break easily on hulling and milling, affecting the head rice recovery in rice (Lyman et al., 2013).

Breeding rice for heat stress tolerance is one of the viable options for mitigating the ill effects of heat stress in rice. Genetic variability for tolerance to reproductive stage heat stress (RSHS) in rice has been well documented (Tenorio et al., 2013; Bheemanahalli et al., 2016; Pradhan et al., 2016; Cheabu et al., 2019; de Brito et al., 2019; Ravikiran et al., 2020). Various mechanisms conferring reproductive stage heat stress tolerance (RSHT) in rice have been reported, which include 1) escape through early morning flowering (Ishimaru et al., 2010; Julia and Dingkuhn, 2013; Hirabayashi et al., 2014), 2) avoidance through mainly evaporative cooling (Julia and Dingkuhn, 2013), and 3) tolerance (Jagadish et al., 2010). True tolerance is primarily adjudged through spikelet fertility, grain yield, and stress tolerance indices. Both spikelet fertility and grain yield under RSHS are quantitative traits, and a large number of QTLs governing RSHT have been documented (Ravikiran et al., 2020). Two major QTLs, *qHTSF1.1* and *qHTSF4.1*, for spikelet fertility were reported on chromosomes 1 and 4 from an upland *aus* cultivar, Nagina 22 (Ye et al., 2012). *qHTSF4.1* has been fine-mapped to around 1.2 Mb region using BC_5F_2 population, and its effect was validated in a set of 24 rice varieties and different genetic backgrounds (Ye et al., 2015a; b). However, in the majority of the studies for mapping RSHT, the phenotypic variance accounted for by the QTLs identified is very low. Furthermore, most of them are neither validated nor finely mapped and cannot be reliably utilized for marker assisted selection for RSHT in rice. Several putative candidate genes associated with heat stress tolerance such as *TOGRI* (*Thermotolerant Growth Required 1*) (Wang et al., 2016), *SLG1* (*Slender guy 1*) (Xu et al., 2020), and *psbA* (Chen et al., 2020) in addition to *OgTTI1* (*Thermo Tolerance 1*) (Li et al., 2015) encoding the α_2 subunit of 26s proteasome, heat shock proteins and heat shock transcription factors, have been proposed in rice.

A biparental mapping population generally limits the number of genes that can be detected as its genetic variation is restricted between the contrasting parents, both phenotypic and genotypic. This would essentially lead to a large number of key genes/alleles contributing to the variability of a particular trait going unaccounted. Furthermore, the number of false positives in linkage mapping is higher due to the existence of extensive genomic regions under disequilibrium. Genome-wide association studies (GWAS) provide a valuable alternative to linkage mapping since it is based on historic recombination, which considerably breaks the linkage disequilibrium (LD) blocks and also enhances the resolution of QTL detected. It is a particularly useful tool to dissect complex traits such as heat stress tolerance. GWAS is popular in rice as it is endowed with vast genetic variability conserved in gene banks and access to rich genomic resources. Recently, pan-genome data of 67 diverse rice accessions with 16.5 million SNPs, 5.5 million indels, and 0.9 million structural variants have been made available (Zhao et al., 2018). Several SNP arrays with varying densities already exist in rice enabling high-throughput genotyping, which include 44 K (Zhao et al., 2011), 6 K (Yu et al., 2014), 50 K (Singh et al., 2015), 700 K (McCouch et al., 2016), and 7 K (Thomson et al., 2017). GWAS is routinely utilized for mapping biotic (Li et al., 2019; Hada et al., 2020), abiotic (Rohilla et al., 2020; Yuan et al., 2020), and grain quality (Misra et al., 2019; Tang et al., 2019)-related attributes in rice. Even though vast genetic variation was reported for RSHT in rice, it has not been properly utilized for identifying MTAs through a GWAS. There is only one systematic GWAS effort which compares three strategies of association mapping (Lafarge et al., 2017). Hence, in the present study, GWAS for RSHT was carried out through the phenotypic characterization of a set of 192 rice accessions for RSHT in terms of spikelet fertility and grain yield and by genotyping through 50K SNP markers, which lead to the identification of significant marker trait associations (MTAs) through three models. The SNPs associated with the significant MTAs are identified as quantitative trait nucleotides (QTNs). Furthermore, the candidate genes governing RSHT in the vicinity of these MTAs were also identified and discussed.

MATERIALS AND METHODS

Phenotypic Characterization of Germplasm for Reproductive Stage Heat Stress

The present study was conducted on a set of 192 diverse rice genotypes assembled from a germplasm collection maintained at

the Division of Genetics, ICAR–Indian Agricultural Research Institute (ICAR-IARI). The experiment was conducted at the IARI–Rice Breeding and Genetics Research Center, Aduthurai, Tamil Nadu (11° 00′N; 79° 28′E, 19.5 m) during the late *Rabi* (December–April) season of 2018–19. The crop was raised under irrigated transplanted conditions. The genotypes were laid out in an augmented randomized complete block design with four blocks – 48 genotypes along with five checks were randomly allocated to 53 plots per block. In each plot, genotypes were planted in three rows with a spacing of 20 × 15 cm. Two staggered sowings with a gap of 30 days were completed to adjust the flowering time of the germplasm to the targeted seasonal temperatures. The first sowing was completed in the second fortnight of December, which served as the unstressed control since the peak anthesis of the genotypes coincided with optimum (max) temperatures, 33–35°C. The second staggering was taken up in the second fortnight of January, which served as reproductive heat stress treatment wherein peak anthesis of genotypes occurred at higher ambient temperatures, that is, 39–41°C. There were no differences in other agronomic practices in both staggered experiments, and all necessary care was taken to raise a healthy rice crop.

Observations Recorded and Data Analysis

The genotypes from two staggered sowings were closely monitored particularly for their flowering and anthesis to make sure that there are no escapes due to variation in flowering duration in these genotypes. At physiological maturity, five randomly selected plants from the middle row of every plot were harvested separately. One panicle from the main tiller of each genotype was sampled for spikelet fertility. The plants were then threshed separately and weighed to record data on a single plant yield. For spikelet fertility, the panicles were threshed individually, and the filled and unfilled grains were counted manually. The proportion of filled grains among the total number of grains per panicle was expressed as spikelet fertility percent. Additionally, the stress tolerance index (STI; Fernandez, 1992) was calculated for both grain yield per plant and spikelet fertility using the following formula:

$$\frac{(Ys)(Yp)}{\bar{Y}p^2},$$

where Ys and Yp are the average yield/spikelet fertility of genotypes under stressed and unstressed conditions, respectively, while Yp represents the mean yield/spikelet fertility of all genotypes under unstressed conditions. Statistical analysis of the phenotypic data was conducted using R statistical software by utilizing appropriate packages. The adjusted means from augmented RCB analysis were generated using the *agricolae* package run on the R studio (RStudio Team, 2016). The package *ggplot2* was utilized to draw frequency curves for different traits.

SNP Genotyping and Filtering

The SNP genotyping of the germplasm set carried out in an earlier study (Bollinedi et al., 2020) was utilized for this study as well. Briefly, 2-week-old seedlings from the nursery were sampled and processed in liquid nitrogen. DNA was extracted using the cetyl trimethyl ammonium bromide method (Murray and Thompson, 1980). DNA quality was first assessed on 0.8% agarose gel, which was further confirmed using a nanospectrophotometer (NanoDrop™ 2000/2000c, Thermo Fisher Scientific, DE, United States). DNA samples were then sent for SNP genotyping using 50k Affymetrix GeneChip (Thermo Fisher Scientific, United States). The technical details of this custom-made gene chip were explained in Singh et al. (2015). The array houses 50,051 SNPs selected from 18,980 genes covering 12 rice chromosomes with an interval of 1 kb between two adjacent SNPs. The genotyping data of 50,051 SNPs were first filtered for rare alleles with a minor allele frequency cutoff of 5%, and then for missing values, markers with >20% missing reads were dropped. The final number of markers utilized for downstream analysis was reduced to 32,712 SNPs.

Population Structure, Linkage Disequilibrium, and Association Analysis

The population structure of the germplasm and LD decay was worked out as explained in Bollinedi et al. (2020). However, principal component analysis (PCA), inbuilt in the R platform for association analysis, genome association, and prediction integrated tool (GAPIT), was conducted to cross-check the number of subpopulations reported (Lipka et al., 2012). The scree plot generated from PCA was utilized to decide the number of components explaining the optimum population structure and thereby the number of subpopulations. Linkage disequilibrium was estimated based on squared allele frequency correlations (r^2) with significant p values (<0.05) for each pair of loci. LD decay was depicted using bins of 200 kb, and the average r^2 value was plotted against the physical distance. The distance at which the r^2 value plummeted to half of its average maximum value was considered as the rate of LD decay. The association analysis was conducted in GAPIT by executing three different models—mixed linear model (MLM), fixed and random model circulating probability unification (FarmCPU), and Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK). Phenotypic data generated under both normal and stressed conditions were used for association analysis. The significant threshold for marker trait associations (MTAs) was fixed at $-\log_{10} p > 5.8$ (Bonferroni threshold) to avoid type 1 errors (false positives). However, to prevent type 2 errors (false negatives), the threshold was relaxed to $-\log10\ p > 5.0$, wherever appropriate (Melandri et al., 2020). For every significant MTA, quantitative trait nucleotide (QTN) was identified.

Co-Localized QTLs, Candidate Genes, and Their *In Silico* Expression Analysis

The physical positions of the MTAs in the rice genome were further analyzed for the presence of any reported QTLs for reproductive stage heat stress tolerance, and MTAs which did not co-localize with QTLs mapped in earlier studies were

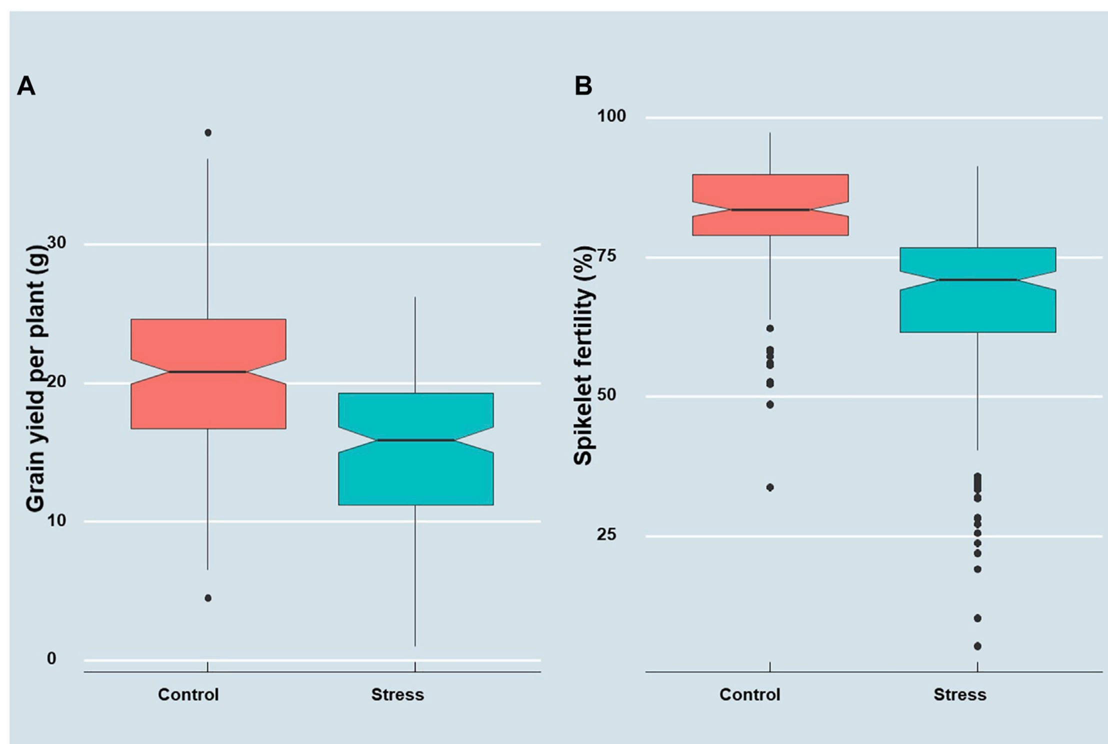


FIGURE 1 | Boxplots depicting the distribution of **(A)** grain yield plant^{-1} (g) and **(B)** spikelet fertility (%) under control and heat stress conditions.

considered novel. The candidate genes in and around these MTAs were identified using the genome browser of the Rice Genome Annotation Project (http://rice.uga.edu/cgi-bin/gbrowse/rice). The tissue-specific expression of the putative candidates identified was analyzed using the datasets available on the RiceXPro website (https://ricexpro.dna.affrc.go.jp). Furthermore, the candidate genes were compared with the results of previous transcriptomic studies and the expression dynamics of common genes.

RESULTS

Phenotypic Characterization of the Germplasm for Reproductive Stage Heat Stress Tolerance

A preliminary augmented ANOVA revealed significant test genotype effects, check effects, and checks versus test entry effects for both grain yield and spikelet fertility, particularly under heat stress (**Supplementary Table S1**). A significant difference was observed in the diurnal mean temperatures during the peak anthesis stage of the germplasm between the two staggered sowing windows. The temperature range during anthesis for the first staggered sown set was between 33 and 35°C, whereas it was 38–40°C for the late sown set (**Supplementary Figure S1** and **Figure 1** of Ravikiran et al., 2020). As a result, the performance of the genotype showed a significant reduction in the late sown set (**Figure 1**). A mean reduction of 26% was observed for grain yield, while it was 19% for spikelet fertility in the late sown set exposed to high-temperature stress at the reproductive stage. The grain yield and spikelet fertility ranged from 4.47 to 38.13 g and from 33.85 to 97.52%, respectively, under the timely sown unstressed situation. However, the grain yield under heat stress ranged from 1.04 to 26.23 g, while the spikelet fertility ranged from 5.30 to as high as 91.46% (**Table 1**). According to IRRI Standard Evaluation System (SES; IRRI, 2002), 27 genotypes are highly sterile with SF of <50%, 96 genotypes were partially sterile with SF ranging between 50 and 74%, 65 genotypes were fertile with SF varying between 75 and 89%, and the remaining four genotypes were highly fertile with spikelet fertility of ≥90% under heat stress conditions. For grain yield under heat stress, the genotypes Bhubana (26.24 g), Indravati (25.56 g), PRR127 (25.42 g), and PRR122 (24.93 g) were found to be superior, while for spikelet fertility, DV85 (91.46%), BJ1 (90.18%), and NDR359 (90.35%) were found to be the best. Grain yield under heat stress and spikelet fertility under the unstressed control showed high broad sense heritability. Both grain yield and spikelet fertility under heat stress followed near normal distributions (Shapiro–Wilk's p-value > 0.05). Furthermore, the range of stress tolerance index (STI) calculated for grain yield (STIGY) (0.01–2.14 with a mean of 0.81) was higher than that of spikelet fertility (STISF) (0.03–1.27 with a mean of 0.82).

TABLE 1 | Summary statistics of grain yield plant^{-1} and spikelet fertility under control and reproductive stage heat stress conditions.

Statistics	Grain Yield plant^{-1} (g)		Spikelet Fertility (%)		STI_{sf}	STI_{gy}
	Control	Heat Stress	Control	Heat Stress		
Mean	20.41	15.14	82.49	66.82	0.82	0.81
Min	4.47	1.04	33.85	5.30	0.03	0.01
Max	38.13	26.23	97.52	91.46	1.27	2.14
S.D.	6.60	5.77	9.87	16.85	0.25	0.47
S.E.	0.47	0.42	0.72	1.22	0.01	0.03
C.V. (%)	24.05	16.28	6.30	23.09	30.54	58.25
PCV	31.93	37.62	11.72	24.75		
GCV	20.89	33.94	9.89	16.54		
h^2 (broad sense)	42.83	81.36	71.19	65.91		
GA	5.78	9.58	14.21	10.08		

S.D., standard deviation; S.E., standard error of the mean; C.V., coefficient of variation; PCV, phenotypic coefficient of variation; GCV, genotypic coefficient of variation; h^2, heritability; GA, genetic advance; STI_{sf}, stress tolerance index calculated for spikelet fertility; STI_{gy}, stress tolerance index calculated for grain yield per plant.

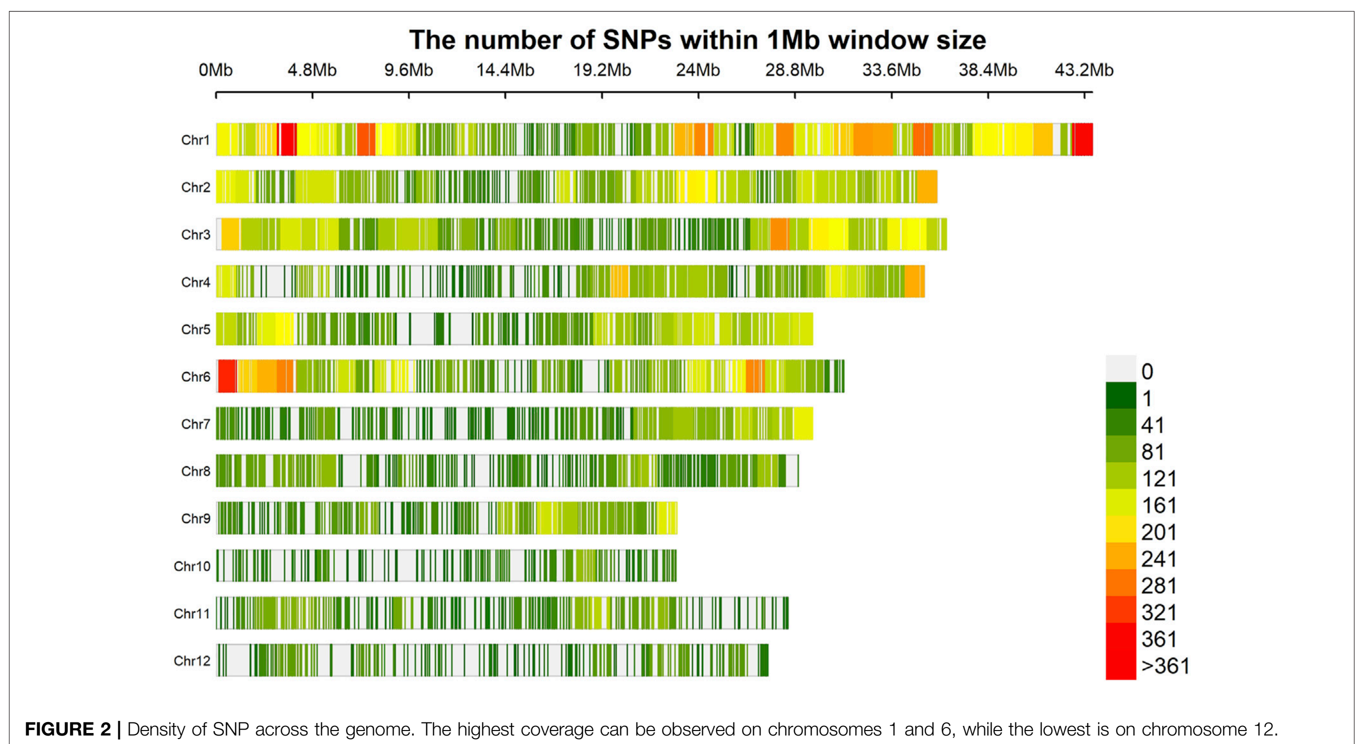


FIGURE 2 | Density of SNP across the genome. The highest coverage can be observed on chromosomes 1 and 6, while the lowest is on chromosome 12.

Population Structure and Linkage Disequilibrium

The marker density plot showed the coverage of markers with an average distance of around 30 kb with almost 90% of the markers spaced within a 5 kb distance (**Figure 2**). In addition, the majority of the genotypes (>150) and markers (>25,000) showed the least heterozygosity, reflecting the true breeding nature of genotypes. The highest r^2 (>0.8) was obtained with a genomic span of <5 kb, followed by a sudden dip (50%) at around 200 kb. Although there were many peaks and valleys, beyond this, there was a gradual general decline in LD which reached less than 0.1 at around 400 kb. Considering this, the marker coverage obtained in the present study is adequate and hence can be utilized for association analysis. Two covariates, population structure and kinship, were employed to cull out false positives. As described earlier (Bollinedi et al., 2020), the population structure analysis using STRUCTURE based on the graph drawn between ΔK and K values showed the existence of three subpopulations, denoted as POP1, POP2, and POP3. This was further predicted by principal component analysis conducted in the present study. The scree plot showed that a significant portion of variance was captured by PC1 itself (29%), followed by PC2 (7%) and PC3 (6%) (**Figure 3A**). Beyond PC3, the variation contributed by individual PCs was meager (<5%) and can be safely ignored. Hence, a PCA 3D plot between PC1, PC2, and PC3 was considered for interpreting the subpopulation composition of the association panel (**Figure 3B**). One of the subpopulations possessed the maximum number of genotypes (139 genotypes),

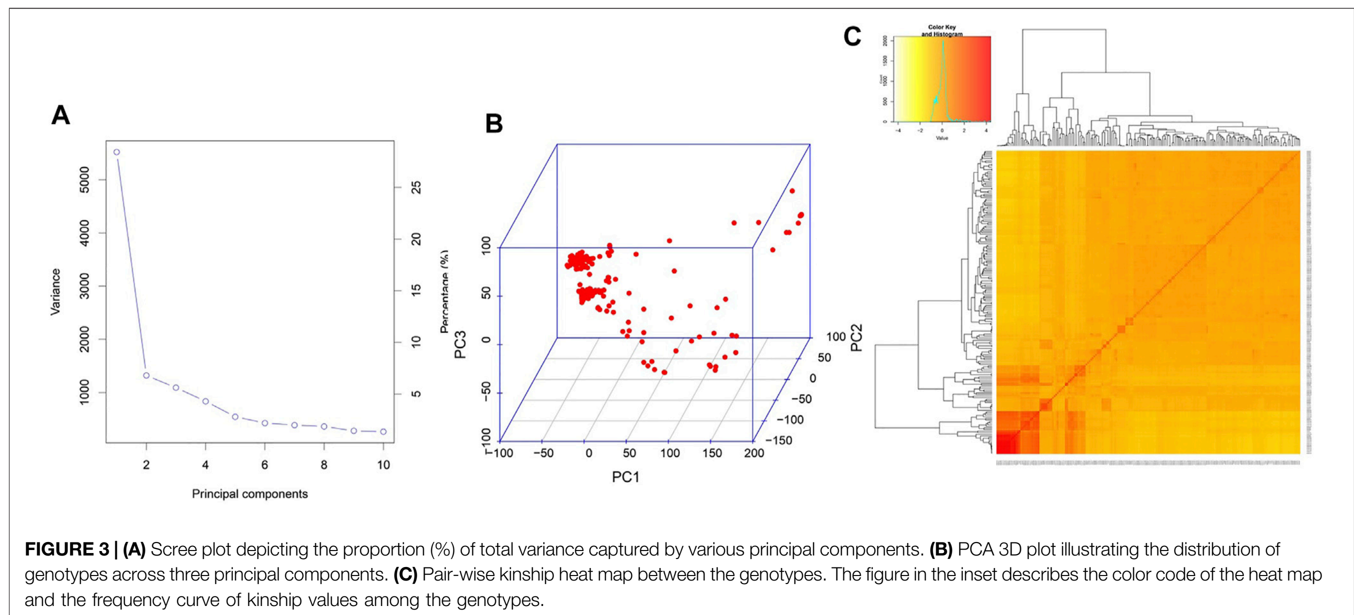


FIGURE 3 | (A) Scree plot depicting the proportion (%) of total variance captured by various principal components. **(B)** PCA 3D plot illustrating the distribution of genotypes across three principal components. **(C)** Pair-wise kinship heat map between the genotypes. The figure in the inset describes the color code of the heat map and the frequency curve of kinship values among the genotypes.

followed by the second subpopulation with 35 accessions and the third with 15 genotypes. The first subpopulation with the maximum membership is composed of most of the popular rice varieties of the country such as MTU1001, ADT 39, MAS946-1, Improved Sabarmati, and some advanced breeding lines. The second subpopulation is made of unique temperate rice landraces of Jammu and Kashmir. Furthermore, the heatmap of a kinship value revealed that the maximum number of kinship values populated around 0 to 0.5, indicating a very weak relatedness or maximum genetic diversity in the association panel utilized in the present study (**Figure 3C**).

Genome-Wide Association Analysis

Genome-wide association analysis was performed by executing three different models, namely, MLM, FarmCPU, and BLINK. The significant MTAs obtained from heat-stressed conditions are summarized in **Table 2** along with the corresponding Manhattan plots (**Figure 4**). However, no significant MTAs could be detected under normal unstressed conditions. The locations of the QTNs identified are depicted in **Figure 5**, and common QTNs across the models are shown in **Figure 6**. For grain yield, MLM failed to detect any significant association, while seven QTNs were detected through FarmCPU and three through BLINK. Only one MTA, *qHTGY8.1*, was common between these two models (**Figure 6A**). This MTA registered the highest probability through BLINK, while it was relatively lower through FarmCPU, but high R^2 values through both models. The MTA *qHTGY10.1* showed the highest probability under FarmCPU. MTAs with higher probability values, *qHTGY10.1* and *qHTGY11.1*, were also reported earlier but for other traits. For spikelet fertility, the highest number of MTAs was detected using the FarmCPU model (8), followed by MLM (7) and BLINK (7). Among these, one MTA was identified consistently across three models, *qHTSF5*.1 (**Figure 6B**), which also displayed the highest R^2 value (0.10). Furthermore, R^2 values of MTAs identified through MLM are slightly higher than those identified through the other two models. This reflects that MLM lays more emphasis on major QTLs and may miss some minor QTLs which play an equally important role in trait expression. The majority of these MTAs are novel in terms of RSHT except for five MTAs, which coincided with the positions of previously reported MTAs. The MTAs identified for STI are almost the same as that of original trait values, reflecting a high correlation between the two. Particularly under MLM, a group of MTAs clustered in the region 20.5 Mb. This region showed a significant hit with FarmCPU and BLINK as well.

Allelic Effects of Major Quantitative Trait Nucleotides Identified

Additionally, the allelic effects of a subset of QTNs identified through GWAS showing significant effects on the respective trait were inspected. For grain yield (**Figure 7**), three QTNs were selected, namely, *qHTGY10.1*, *qHTGY1.1*, and *qHTGY8.1*, with the linked SNP AX-95938448, AX-95918021, and AX-95937704, respectively. Among these three markers, the greatest significant difference for the grain yield per plant between the genotypes carrying alternate alleles was found for the marker AX-95918021, followed by AX-95937704 and AX-95938448. The genotypes carrying the ‘A’ allele for AX-95918021 showed a mean grain yield of 8 g, while those carrying its alternate allele, ‘G’, had a mean grain yield of 18 g. Similarly, genotypes with the ‘C’ allele of AX-95937704 exhibited a grain yield of 12 g, while those carrying the ‘T’ allele showed a grain yield of 18 g. Similarly, for spikelet fertility, seven major QTNs were selected, among which AX-95940947 with its linked QTN *qHTSF1.3* showed a highly significant difference in spikelet fertility values for the two alternate alleles closely followed by AX-95918542 linked to QTN *qHTSF1.1* (**Figure 8**). The ‘A’ allele of AX-95940947 showed an allelic effect in terms of spikelet fertility (%) of 47, while the ‘T’ allele showed an effect of 75. The ‘A’ allele of AX-95918542 exhibited an effect of 75%

TABLE 2 | Significant MTAs identified for grain yield plant^{-1} (GYPP) and spikelet fertility (SF) under heat stress and their respective stress tolerance indices.

Trait	SNP	Model	Fav Allele	Chr	Position	Probability	R^2	MTA	Previously Reported QTLs[a]
GYPP	AX-95938448	FarmCPU	C	10	12,691,143	9.11E-08	0.02	*qHTGY10.1*	*qhr3-1* Cao et al. (2003)
	AX-95938539		T	11	7,598,615	2.03E-07	0.05	*qHTGY11.1*	*qADL09-11* Tazib et al. (2015)
	AX-95918021		G	1	43,529,356	2.74E-07	0.03	*qHTGY1.1*	-
	AX-95959896		C	9	6,976,331	4.10E-06	0.02	*qHTGY9.1*	-
	AX-95957982		G	7	7,144,612	2.51E-05	0.02	*qHTGY7.1*	-
	AX-95938211		A	9	13,559,420	4.95E-05	0.02	*qHTGY9.2*	-
	AX-95937704		**C**	**8**	**3,014,640**	**8.57E-05**	**0.05**	***qHTGY8.1***	-
	AX-95937704	BLINK	**C**	**8**	**3,014,640**	**2.43E-10**	**0.05**	***qHTGY8.1***	-
	AX-95916113		C	1	2,051,081	8.53E-07	0.04	*qHTGY1.2*	-
	AX-95932,408		C	11	20,582,446	2.31E-05	0.03	*qHTGY11.2*	-
SF	AX-95926541	MLM	C	5	20,542,291	1.62E-06	0.10	*qHTSF5.1*	-
	AX-95961044		G	11	9,168,125	1.09E-05	0.08	*qHTSF11.1*	*qHTSF11.2* Ye et al. (2015a) *qLD10*-11 Tazib et al. (2015)
	AX-95926170		**G**	**5**	**20,547,603**	**1.70E-05**	**0.08**	***qHTSF5.1***	-
	AX-95927077		**T**	**5**	**20,540,979**	**1.70E-05**	**0.08**	***qHTSF5.1***	-
	AX-95962,696		C	4	16,049,320	1.73E-05	0.08	*qHTSF4.1*	*SSPF4* Xiao et al. (2011a) *qPF4* Xiao et al. (2011b)
	AX-95918542		A	1	28,440,408	4.95E-05	0.07	*qHTSF1.1*	-
	AX-95956527		T	6	28,019,302	5.42E-05	0.07	*qHTSF6.1*	-
	AX-95956527	FarmCPU	T	6	28,019,302	2.53E-10	0.07	*qHTSF6.1*	-
	AX-95926541		**C**	**5**	**20,542,291**	**3.57E-09**	**0.10**	***qHTSF5.1***	-
	AX-95930097		G	7	25,603,077	6.96E-08	0.03	*qHTSF7.1*	*qAL10-7* Tazib et al. (2015)
	AX-95924814		C	4	28,011,513	9.78E-08	0.04	*qHTSF4.2*	-
	AX-95944879		T	1	28,469,004	6.75E-06	0.06	*qHTSF1.1*	-
	AX-95923670		G	3	31,000,572	1.45E-05	0.03	*qHTSF3.1*	*qtl_3.4* Jagadish et al. (2010)
	AX-95918181		G	1	43,540,685	2.80E-05	0.04	*qHTSF1.2*	-
	AX-95949687		A	3	34,204,030	5.32E-05	0.04	*qHTSF3.2*	-
	AX-95926541	BLINK	**C**	**5**	**20,542,291**	**2.15E-10**	**0.10**	***qHTSF5.1***	-
	AX-95924814		C	4	28,011,513	7.73E-10	0.04	*qHTSF4.2*	-
	AX-95921100		T	2	30,694,562	2.67E-09	0.03	*qHTSF2.1*	*qtl_2.2* Jagadish et al. (2010)
	AX-95940947		T	1	10,799,807	5.56E-09	0.05	*qHTSF1.3*	-
	AX-95918181		G	1	43,540,685	3.66E-07	0.04	*qHTSF1.2*	-
	AX-95936404		C	5	16,852,857	4.32E-05	0.05	*qHTSF5.2*	-
	AX-95956527		T	6	28,019,302	6.21E-05	0.07	*qHTSF6.1*	-
STIGY	AX-95930775	FarmCPU	A	8	3,009,287	8.96E-05	0.03	*qSTIGY8.1*	-
	AX-95930775	BLINK	A	8	3,009,287	8.96E-05	0.03	*qSTIGY8.1*	-
STISF	AX-95926541	MLM	C	5	20,542,291	5.10E-07	0.11	*qSTISF5.1*	-
	AX-95927077		**T**	**5**	**20,540,979**	**5.11E-06**	**0.09**	***qSTISF5.1***	-
	AX-95926170		**G**	**5**	**20,547,603**	**5.11E-06**	**0.09**	***qSTISF5.1***	-
	AX-95952,837		A	5	20,642,828	8.34E-06	0.09	*qSTISF5.1*	-
	AX-95961044		G	11	9,168,125	1.53E-05	0.08	*qSTISF11.1*	*qHTSF11.2* Ye et al. (2015a) *qLD10-11* Tazib et al. (2015)
	AX-95927241		A	5	21,126,325	4.49E-05	0.07	*qSTISF5.2*	-
	AX-95962,696		C	4	16,049,320	4.54E-05	0.07	*qSTISF4.1*	*SSPF4* Xiao et al. (2011a) *qPF4* Xiao et al. (2011b)
	AX-95926541	FarmCPU	**C**	**5**	**20,542,291**	**3.77E-09**	**0.11**	***qSTISF5.1***	-
	AX-95924814		C	4	28,011,513	1.48E-07	0.03	*qSTISF4.2*	-
	AX-95940947		G	1	10,799,807	1.17E-06	0.04	*qSTISF1.1*	-
	AX-95919771		T	2	31,278,025	6.76E-06	0.03	*qSTISF2.1*	-
	AX-95926541	BLINK	**C**	**5**	**20,542,291**	**8.87E-15**	**0.11**	***qSTISF5.1***	-
	AX-95924814		C	4	28,011,513	1.50E-05	0.03	*qSTISF4.2*	-
	AX-95956527		T	6	28,019,302	3.67E-05	0.06	*qSTISF6.1*	-
	AX-95919771		T	2	31,278,025	5.25E-05	0.03	*qSTISF2.1*	-

[a]QTLs, reported exclusively for reproductive stage heat stress tolerance in rice; Chr, chromosome; GYPP, grain yield per plant; SF, spikelet fertility; STISF, stress tolerance index calculated for spikelet fertility; STIGY, stress tolerance index calculated for grain yield per plant; MTAs which are common across two to three models are highlighted in bold for various traits.

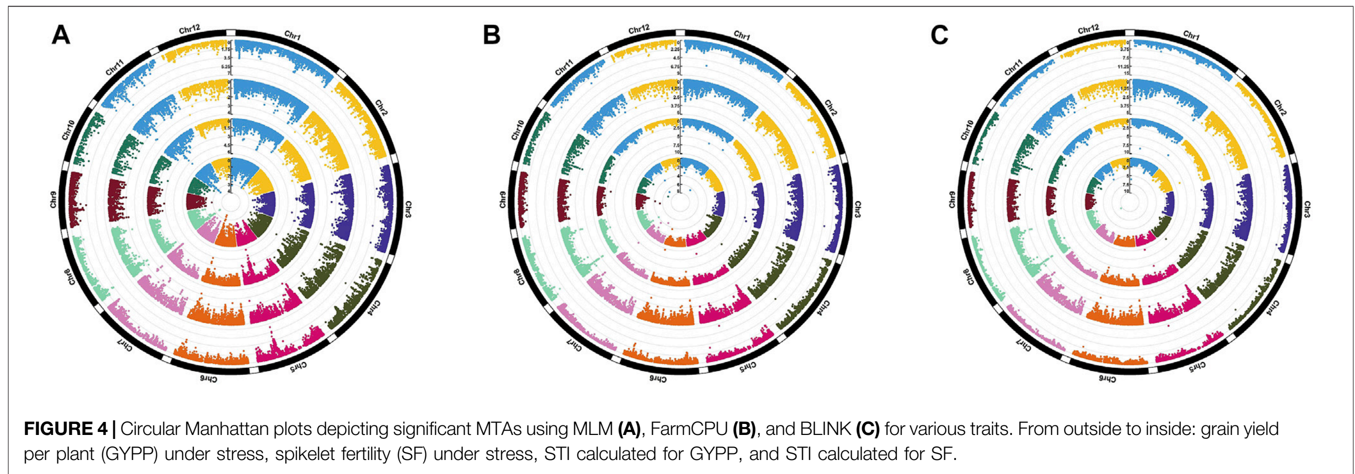


FIGURE 4 | Circular Manhattan plots depicting significant MTAs using MLM **(A)**, FarmCPU **(B)**, and BLINK **(C)** for various traits. From outside to inside: grain yield per plant (GYPP) under stress, spikelet fertility (SF) under stress, STI calculated for GYPP, and STI calculated for SF.

1: 0.0 TEL1; 2.0 *qHTGY1.2*; 10.8 *qHTSF1.3 qSTISF1.1*; 16.7 CEN; 28.4 *qHTSF1.1*; 43.5 *qHTGY1.1 qHTSF1.2*; 43.6 TEL2
2: 0.0 TEL1; 13.6 CEN; 30.7 *qHTSF2.1*; 31.3 *qSTISF2.1*; 35.9 TEL2
3: 0.0 TEL1; 19.4 CEN; 26.4 *qSTISF3.1*; 31.0 *qHTSF3.1*; 34.2 *qHTSF3.2*; 36.4 TEL2
4: 0.0 TEL1; 9.7 CEN; 16.1 *qHTSF4.1 qSTISF4.1*; 28.0 *qHTSF4.2 qSTISF4.2*; 35.5 TEL2
5: 0.0 TEL1; 12.4 CEN; 16.9 *qHTSF5.2*; 20.5 *qHTSF5.1 qSTISF5.1*; 21.1 *qSTISF5.2*; 31.3 TEL2
6: 0.0 TEL1; 15.3 CEN; 28.0 *qHTSF6.1 qSTISF6.1*; 31.3 TEL2
7: 0.0 TEL1; 7.1 *qHTGY7.1*; 12.1 CEN; 25.6 *qHTSF7.1*; 29.7 TEL2
8: 0.0 TEL1; 3.0 *qSTIGY8.1 qHTGY8.1*; 12.9 CEN; 28.4 TEL2
9: 0.0 TEL1; 2.8 CEN; 7.0 *qHTGY9.1*; 13.6 *qHTGY9.2*; 23.0 TEL2
10: 0.0 TEL1; 8.2 CEN; 12.7 *qHTGY10.1*; 23.2 TEL2
11: 0.0 TEL1; 7.6 *qHTGY11.1*; 9.2 *qHTSF11.1 qSTISF11.1*; 12.0 CEN; 20.6 *qHTGY11.2*; 29.0 TEL2

FIGURE 5 | Physical positions (Mb) of various MTAs identified across the rice genome through various models employed in GAPIT.

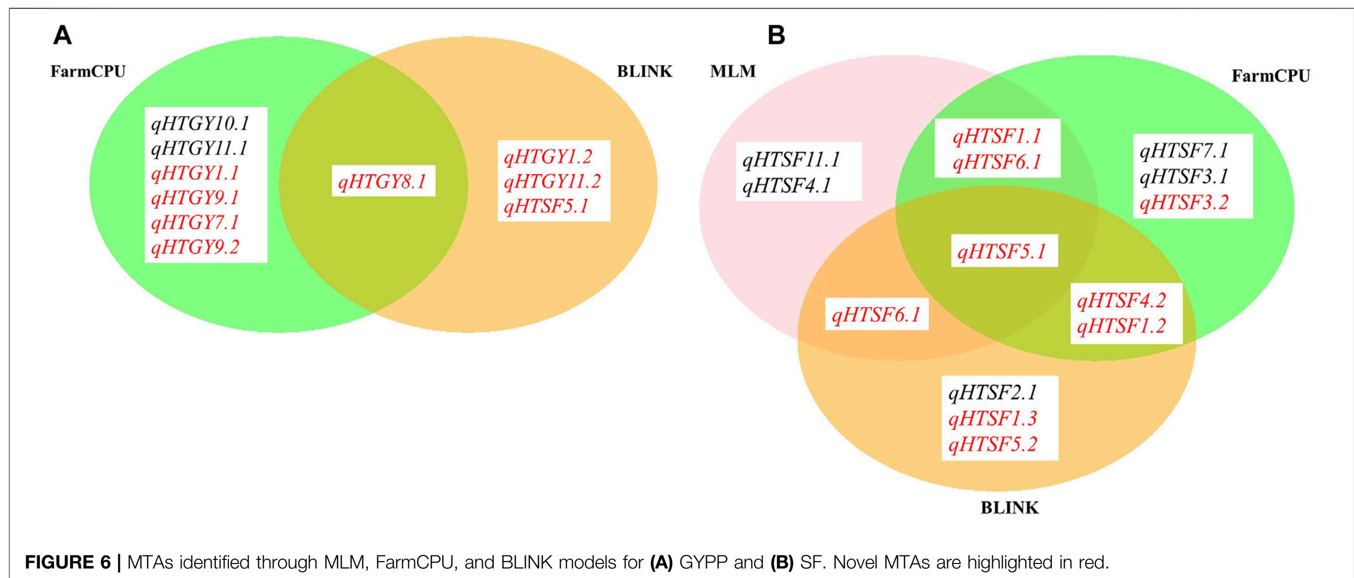


FIGURE 6 | MTAs identified through MLM, FarmCPU, and BLINK models for **(A)** GYPP and **(B)** SF. Novel MTAs are highlighted in red.

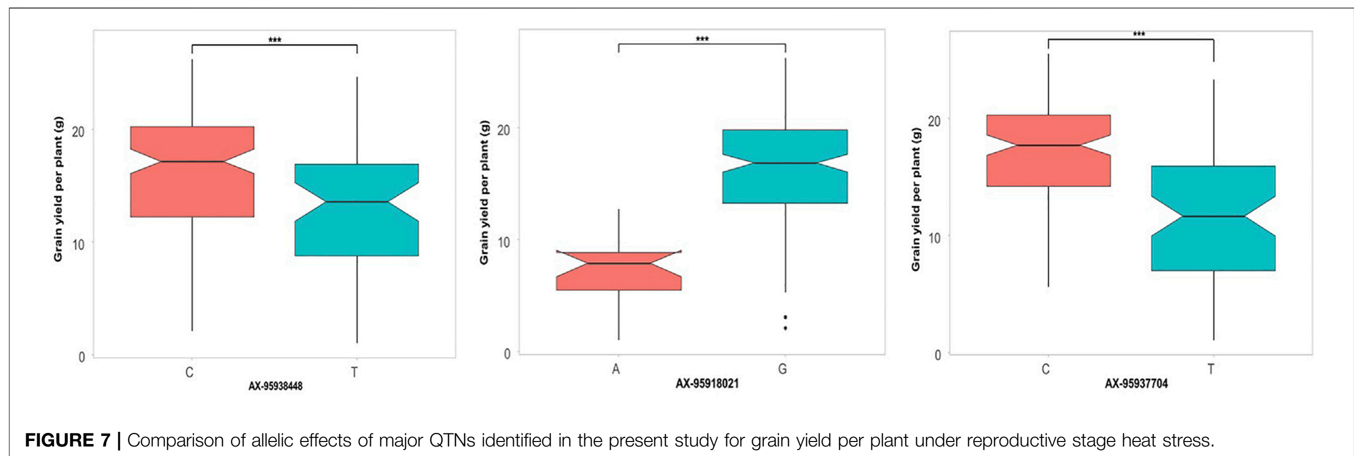


FIGURE 7 | Comparison of allelic effects of major QTNs identified in the present study for grain yield per plant under reproductive stage heat stress.

spikelet fertility, and the ‘T’ allele showed an effect of 50%. Interestingly, both these QTNs are located on chromosome 1 but 18 Mb apart. The remaining QTNs also showed statically significant ($p < 0.001$) differences between the trait values conferred by their respective alleles, except AX-95956527, with its linked QTN, *qHTSF6.1*, significant at only $p < 0.05$. This further attests to the robustness of major QTNs identified in the present study.

Identification of Putative Candidate Genes for RSHT and Their In Silico Expression Analysis

A total of 11 candidate genes were identified near 10 SNPs associated with grain yield under heat stress distributed on chromosomes 1 (2), 7 (1), 8 (3), 9 (2), 10 (1), and 11 (2) (**Table 3**). Of these genes, five of them are kinases–diacylglycerol kinase, Ser/Thr protein kinase, and receptor-like kinase involved in stress signaling cascades. Others include glucosylceramidase, zinc ion binding protein, DUF1336 domain-containing protein, cytochrome P450, and lipase with putative roles in plant defense responses. Similarly, 17 candidate genes were found in the vicinity of 10 SNPs associated with spikelet fertility under heat stress scattered across chromosomes 1 (2), 2 (3), 3 (5), 4 (1), 5 (3), 6 (1), 7 (1), and 11 (1). The majority of these genes are involved in protein processing and protein–protein interactions such as tetratricopeptide repeat domain-containing protein, an E3 ubiquitin ligase, BTB/POZ domain-containing protein, U-box domain-containing protein, ankyrin repeat-containing protein, and RING zinc finger protein. Others include genes involved in abiotic stress responses and other essential pathways of plant development. The survey of expression database (RiceXPro) revealed that the majority of the putative candidate genes identified for grain yield showed predominant vegetative stage-specific expression, whereas those identified for spikelet fertility expressed primarily in the reproductive organs (**Supplementary Figures S3A,B**). Additionally, the putative candidate genes identified in the present study

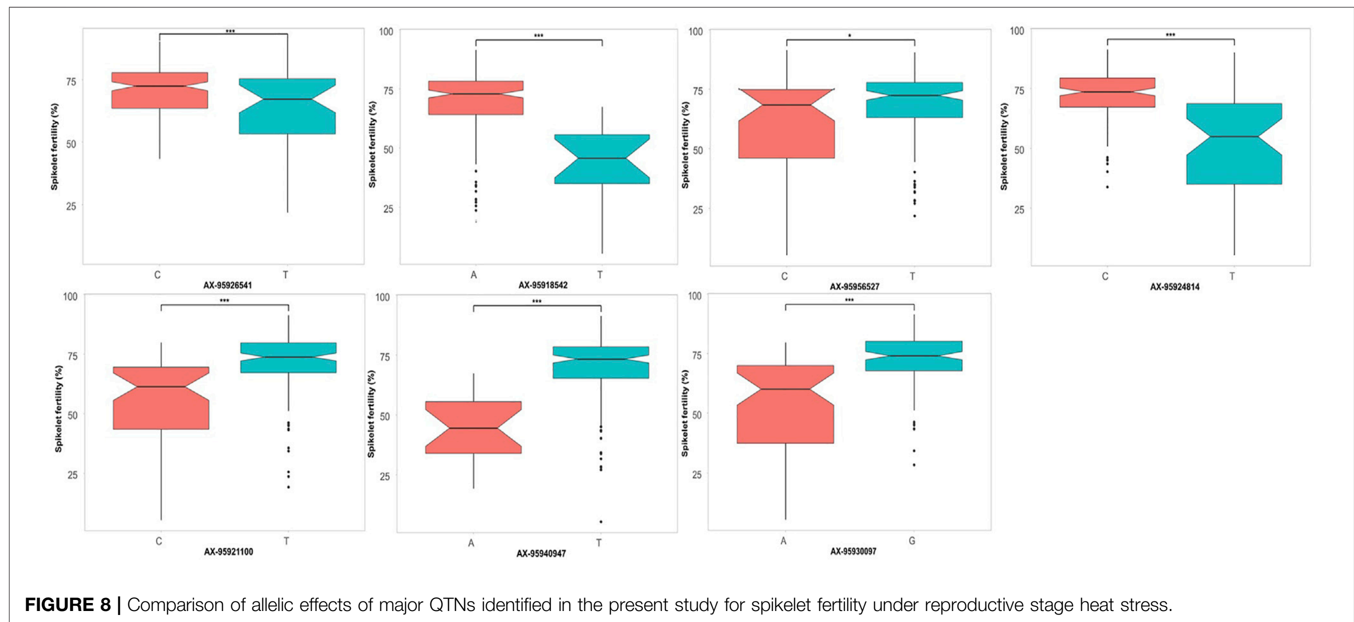


FIGURE 8 | Comparison of allelic effects of major QTNs identified in the present study for spikelet fertility under reproductive stage heat stress.

were compared with the differentially expressed genes (DEGs) identified in previous transcriptomic studies under heat stress in rice. Six genes were found common with the DEGs reported by Liu et al. (2020) in the thermo-tolerant genotype, SDWG005 (**Table 4**). Of these, three genes were upregulated, while three genes were downregulated under heat stress. There were four putative candidate genes that matched with the DEGs observed by Cai et al. (2020), of which two genes were upregulated and the other two were downregulated.

Selection of RSHT Genotypes With Superior Allelic Combination

The promising genotypes with the *per se* trait values of grain yield and spikelet fertility under stress and STI were shortlisted, and the number of positive alleles of 29 putative MTAs in these tolerant genotypes was investigated. A total of 18 genotypes were identified (**Table 5**). The genotype RIL 10 accumulated positive alleles for all the 29 SNPs coupled with superior grain yield and spikelet fertility under stress. This was followed closely by Selected Sabarmati, Sitwa Dhan, Bhadrakali, Samba Mashuri, Hema, and Haldamuri, which showed positive alleles for 28 SNPs. Similarly, UPRI 2003–45, B6144-MR-6-0-0, and Pant dhan 18 had positive alleles for 27 SNPs, while RNRM 7, Gouri, Indravati, and ADT 39 had positive alleles for 26 SNPs. The genotypes IR 77384–12-35-3-6-7-2-B and JR 75 showed positive alleles for 25 SNPs, while PRR117 showed positive alleles for 24 SNPs.

DISCUSSION

Crop yields and productivity are predicted to suffer acutely in the coming years due to climate change and global warming. Breeding heat-tolerant crop varieties is essential to address crop losses due to heat stress in rice. The present study aimed to map MTAs governing reproductive stage heat stress tolerance in rice using a diverse set of 192 rice germplasm lines. The two sowing windows utilized could adequately distinguish sensitive and tolerant genotypes, providing a unique situation in which two staggered sowings differed only for the heat stress at the reproductive stage. The screening was carried out during the late *rabi* season in Tamil Nadu, where winter is normally felt inconspicuous from normal weather conditions. The heat stress normally occurs at Aduthurai during the end of the rabi season, marking the beginning of summer. The weather change occurs within a span of 30 days, which is a significant shift from the normal temperature to a high temperature. We had two overlapping conditions: 1) one under a normal sowing which provided the most ideal normal weather conditions all throughout the crop period including the reproductive stage and 2) a late sowing that provided a heat stress only at the terminal stage. The highest temperature reached during the flowering of the second sowing was ~42°C. The overlap of both the sowings was almost 75%, meaning that both the normal and stressed conditions experienced the same environment most of the time, except for a 30-day window. In the first sowing, the window was during the initial stage, whereas in the second sowing, the window coincided with the reproductive stage. Therefore, both the sowings shared a common environmental influence most of the time. Moreover, weather-wise, the initial 30-day window for the first sowing was not different from that of the second sowing. This provides a unique situation, in which the environmental difference is maximized only during the reproductive stage between both the sowings, and hence, the data generated during this stage could be reliable for studying the RSHT among the genotypes. If found unique and reliable, single season data can be utilized for

TABLE 3 | Candidate genes in the QTN regions identified through three GWAS models for grain yield per plant, spikelet fertility, and their respective stress tolerance index.

Trait or gene ID	Closest SNP	Chr	Start	End	Gene Annotation	Putative Gene Function
Grain yield plant^{-1}						
LOC_Os10g24690	AX-95938448	10	12,690,416	12,689,328	Expressed protein	-
LOC_Os11g13810	AX-95938539	11	7,595,274	7,606,870	Non-lysosomal glucosylceramidase, putative, expressed	Catabolism of sphingolipids
LOC_Os09g12250	AX-95959896	9	6,975,494	6,978,083	Expressed protein	-
LOC_Os07g12520	AX-95957982	7	7,145,539	7,149,827	Zinc ion binding protein, putative, expressed	Reproductive development
LOC_Os09g22450	AX-95938211	9	13,559,548	13,563,071	Lipase, putative, expressed	Disease resistance
LOC_Os08g05640	AX-95937704	8	3,013,718	3,016,330	DUF1336 domain-containing protein	Disease resistance
LOC_Os08g05620	AX-95937704	8	3,007,241	3,009,195	Cytochrome P450, putative, expressed	Regulation of non-enzymatic antioxidant synthesis
LOC_Os08g05650	AX-95937704	8	3,017,356	3,022,047	Diacylglycerol kinase, putative, expressed	Plant stress response
LOC_Os01g04570	AX-95916113	1	2,048,717	2,052,521	Ser/Thr protein kinase, putative, expressed	Heat shock response
LOC_Os01g04580		1	2,053,583	2,057,638	Ser/Thr protein kinase, putative, expressed	
LOC_Os11g35120	AX-95932,408	11	20,589,041	20,591,080	OsWAK116 - OsWAK receptor-like cytoplasmic kinase OsWAK-RLCK, expressed	Signal transduction
Spikelet fertility						
LOC_Os11g16540	AX-95961044	11	9,156,589	9,173,338	Tetratricopeptide repeat domain-containing protein, expressed	Protein–protein interactions
LOC_Os01g49470	AX-95918700	1	28,447,887	28,458,977	E3 ubiquitin ligase, putative, expressed	Proteolysis
LOC_Os06g46240	AX-95956527	6	28,007,285	28,017,490	BTB/POZ domain-containing protein, putative, expressed	Transcriptional regulation and protein degradation
LOC_Os07g42750	AX-95930097	7	25,601,642	25,606,513	DDT domain-containing protein, putative, expressed	Chromatin remodeling
LOC_Os04g47170	AX-95924814	4	28,013,719	28,017,003	ATROPGEF7/ROPGEF7, putative, expressed	Secondary cell wall formation
LOC_Os01g49490	AX-95944879	1	28,467,690	28,471,172	Expressed protein	-
LOC_Os03g54084	AX-95923670	3	31,004,724	31,009,782	Phytochrome C, putative, expressed	Photoperiodic response
LOC_Os03g54091		3	31,009,978	31,013,940	OsTOP6A1-Topoisomerase 6 subunit A homolog 1, expressed	Meiotic recombination
LOC_Os03g60140	AX-95949687	3	34,198,836	34,201,827	U-box domain-containing protein, putative, expressed	Protein degradation
LOC_Os03g60130		3	34,194,882	34,197,992	Transcription elongation factor protein, putative, expressed	Regulation of flower induction
LOC_Os03g60150		3	34,202,415	34,206,308	Protein kinase domain-containing protein, expressed	Kinase activity
LOC_Os02g50270	AX-95921100	2	30,695,324	30,698,401	Ankyrin repeat-containing protein, putative, expressed	Protein–protein interactions and protein chaperoning
LOC_Os02g50280		2	30,698,990	30,701,615	Pentatricopeptide, putative, expressed	Abiotic stress responses
LOC_Os02g50290		2	30,701,962	30,703,636	RING zinc finger protein, putative, expressed	Protein ubiquitination
LOC_Os05g28730	AX-95936404	5	16,850,590	16,853,524	Zinc finger, C3HC4 type domain-containing protein, expressed	Signal transduction
LOC_Os05g28720		5	16,845,905	16,848,027	PPR repeat-containing protein, expressed	Abiotic stress responses
LOC_Os05g28740		5	16,859,683	16,860,977	Universal stress protein domain-containing protein, putative, expressed	Several abiotic stresses
LOC_Os03g46640	AX-95949747	3	26,396,243	26,397,459	Deoxyuridine 5-triphosphate nucleotidohydrolase, putative, expressed	Floral organ development
LOC_Os03g46650		3	26,398,523	26,406,544	WD domain, G-beta repeat domain-containing protein, expressed	Signal transduction

Chr, chromosome.

the GWAS study in rice (Melandri et al., 2020). Field-based phenotyping for RSHT was adopted in other studies (Bheemanahalli et al., 2016; Huang et al., 2016; Pradhan et al., 2016; Sukkeo et al., 2017; Cheabu et al., 2019). A similar sowing date (January last week) to expose the genotypes to heat stress was also chosen by Chidambaranathan et al. (2021) and Pradhan et al. (2016). Wide variation was observed for RSHT in terms of both grain yield and spikelet fertility. A similar variability at this scale (>150 genotypes) for RSHT was also reported in previous studies (Tenorio et al., 2013; Bheemanahalli et al., 2016; Pradhan et al., 2016; Cheabu et al., 2019; de Brito et al., 2019). From a set of 182 *indica*, *japonica*, and *indica*/*japonica* genotypes assessed for RSHT, two genotypes, LTB 14301 and BRS Pampa, were found tolerant (de Brito et al., 2019). Out of 198 genotypes exposed to heat stress, 15 genotypes showed spikelet sterility of <15% (Chidambaranathan et al., 2021). Similarly, from 169 rice accessions evaluated under very high temperatures (40–45°C), four genotypes, namely, AUS17, M9962, SONALEE, and AUS16, were found to be tolerant (with a seed-setting rate of >75%) along with N22. In another study, a total of 240 genotypes were assessed for RSHT at Cuttack, of which 59 genotypes including N22 showed spikelet fertility >60% (Pradhan et al., 2016). From a total of 511 rice genotypes evaluated in an open field, 200 genotypes showing high spikelet fertility (>60%) were further

TABLE 4 | Differential expression of some putative candidate genes identified in the present study from published datasets.

Gene ID	Log_2FC^a (6 h)	Log_2FC^a (12 h)	Log_2FC^b (36°C)	Log_2FC^b (38°C)	Log_2FC^c (36°C)	Log_2FC^c (38°C)
LOC_Os08g05620	−1.51	−1.40	-	-	-	-
LOC_Os01g04580	-4.87	−2.35	-	-	-	-
LOC_Os05g28740	-	−2.12	-	-	-	-
LOC_Os07g42750	2.45	1.65	-	-	-	-
LOC_Os01g04570	1.68	1.48	-	-	-	-
LOC_Os03g54091	1.43	1.12	-	-	-	-
LOC_Os10g24690	-	-	-	-	−1.89	−3.02
LOC_Os05g28730	-	-	1.67	2.18	2.08	2.45
LOC_Os08g05620	-	-	1.66	2.20	−0.17	0.28
LOC_Os03g46640	-	-	−1.36	−2.10	−0.69	−1.02

[a]*Logarithm of fold change values in thermo-tolerant genotype SDWG005 exposed to 6 and 12 h of heat stress during anthesis (Liu et al., 2020).*
[b]*Logarithm of fold change values in thermo-tolerant genotype SDWG005 exposed to heat stress (36 and 38°C) during the meiosis stage (Cai et al., 2020).*
[c]*Logarithm of fold change values in thermo-sensitive genotype MH101 exposed to heat stress (36 and 38°C) during the meiosis stage (Cai et al., 2020).*

TABLE 5 | List of best-performing genotypes in terms of grain yield, spikelet fertility, and their respective tolerance indices and number of positive alleles of 29 MTAs identified for RSHT.

S. No	Genotype	GY_C	GY_H	STI_GY	SF_C	SF_HT	STI_SF	NPA
1	RIL 10	29.87	23.5	1.68	90.59	86.23	1.14	29
2	Selected Sabarmati	21.98	20.38	1.07	86.8	86.79	1.1	28
3	Sitwa Dhan	26.15	24.66	1.54	77.49	75.02	0.85	28
4	Bhadrakali	25.5	21.54	1.31	90.73	89.81	1.19	28
5	Samba Mahsuri	24.23	22.32	1.29	88.64	78.24	1.02	28
6	Hema	24.88	22.22	1.32	86.79	86.01	1.09	28
7	Haldimuri	25.14	23.38	1.4	97.52	88.35	1.26	28
8	UPRI 2003–45	23.32	21.5	1.2	90.86	83.32	1.11	27
9	B6144-MR-6-0-0	16.81	15.33	0.62	83.59	82.6	1.01	27
10	Pant dhan 18	19.98	19.27	0.92	91.15	88.72	1.19	27
11	RNRM 7	24.62	20.25	1.19	90.63	84.11	1.12	26
12	Gouri	27.44	24.9	1.63	78.06	76.81	0.88	26
13	Indravati	27.88	25.56	1.7	83.64	82.81	1.02	26
14	ADT 39	29.21	23.57	1.64	89.85	79.38	1.05	26
15	IR 77384–12-35-3-6-7-2-B	17.23	15.2	0.81	87.58	83.88	1.08	25
16	JR 75	23.58	19.88	1.12	93.18	87.06	1.19	25
17	PRR 117	16.74	14.53	0.58	87.13	85.58	1.09	24
18	IR 70	25.04	22.48	1.34	81.31	80.29	0.96	24

GY_C, grain yield under control; GY_H, grain yield under heat stress; STI_GY, stress tolerance index for grain yield, SF_C, spikelet fertility under control; SF_H, spikelet fertility under heat stress; STI_SF, stress tolerance index for spikelet fertility; NPA, number of positive alleles.

assessed in a controlled environment (38°C), of which 28 genotypes with high fertility were finally selected as donors of RSHT (Tenorio et al., 2013). There were certain genotypes, such as Bhuban, PRR127, DV85, and Improved Samba Mashuri, which could maintain exceptionally higher yields and spikelet fertility under high temperature and can be valuable donors for RSHT after further validation. Grain yield suffered more than spikelet fertility, implying the role of other traits in addition to spikelet fertility which contributes to grain yield under heat stress. Both spikelet fertility and grain yield showed quantitative inheritance qualifying for a GWAS. Considering both phenotypes and genotypes (positive alleles of significant MTAs), RIL 10 was found to be the best performer of all the genotypes and can be a potential heat-tolerant donor after validation in further studies.

In the present study, a total of 32,712 SNPs were utilized for GWAS, which is greater than previous reports. PCA-based population structure analysis showed the presence of three subpopulations as in the previous report (Bollinedi et al., 2020). Assessing the population structure is indispensable, particularly in rice, since it is a self-pollinated crop with clearly differentiated ecotypes—*indica, temperate japonica, tropical japonica, aus/boro,* and *Basmati/Sadri* (Chen et al., 2019). The classification of genotypes through structure analysis in the present study is mainly based on the degree of their domestication and improvement through breeding to which they had been exposed with popular varieties subsumed in one group, with the landraces in the other. Another important parameter to be considered for GWAS is familial relatedness. The kinship values between different genotype pairs clustered around near-zero values, reflecting optimum genetic diversity between the genotypes. Several models are available to conduct the GWAS study. Broadly, they are classified into two categories—single locus models and multi-locus models. Single locus methods test MTAs one SNP at a time akin to single-marker analysis or simple interval mapping of the QTL study. These include the general linear model (GLM) (fits only population

structure as a covariate in the model), mixed linear model (MLM) (Zhang Y-M. et al., 2005; Yu et al., 2006) (uses both ancestry coefficient and kinship as covariates), and its improvements and modifications. To detect more MTAs with the least type I errors, multi-locus models have been proposed, which control background noise generated by other loci (termed pseudo-QTNs) which are in LD with the locus being tested. This is similar to the CIM and MIM strategies of QTL mapping. The multi-locus methods include MLMM (Segura et al., 2012), FarmCPU (Liu et al., 2016), and BLINK (Huang et al., 2019). Several other models were proposed after these recently (Zhang et al., 2020). Among them, in the present study, we have utilized these three models inbuilt in the GAPIT package for association analysis—MLM, FarmCPU, and BLINK (Lipka et al., 2012). In a GWAS study, it is important to keep the critical value very stringent to identify true MTAs by eliminating any false positives. Although the Bonferroni threshold is commonly used, it is likely to exclude some of the real positives if the probability of such MTAs lies too close to the threshold. Therefore, it is common to relax the threshold to bring such MTAs into selection. In this study too, the cutoff was relaxed to 5.0, but only those MTAs consistently appearing under multiple models were considered.

Although GWAS had been frequently adopted for mapping in rice, there were only two previous reports of GWAS for RSHT in rice. A set of 20 previously reported SSR markers was validated in one study using 62 rice genotypes (Pradhan et al., 2016). In another study, GBS-based genotyping of 167 rice accessions was performed to identify MTAs for 20 traits (Lafarge et al., 2017) using a set of 13,160, including 6667 SNPs and 6593 DArT markers. The authors compared three strategies of arriving at MTAs, single-marker-based regression, haplotype-based GWAS, and Bayesian Lasso-based analysis, which were utilized for identifying the MTAs, as in the present study. GWAS was also utilized for identifying genes for RSHT in other plant species such as Arabidopsis (Bac-Molenaar et al., 2015), *Brassica napus* (Rahaman et al., 2018), wheat (Tadesse et al., 2019; Guo et al., 2020; Sharma et al., 2020), and field pea (Tafesse et al., 2020).

Both *per se* trait values and the stress tolerance index extracted from them were utilized to establish marker-trait associations. GWAS for stress tolerance indices have been carried out in wheat (Gahlaut et al., 2019), cotton (Baytar et al., 2018), and *Brassica* (Khanzada et al., 2020). Several MTAs were identified in the present study for grain yield and spikelet fertility. The number of MTAs identified through FarmCPU and BLINK was higher than MLM, indicating that multi-locus models are effective in identifying reliable MTAs. False negatives were found to be more in the MLM model, implying its medium power of QTL detection. MLM weakens the associations in an effort to control the inflation of *p*-values *via* the population structure, thus missing out many minor effect MTAs. These minor effect MTAs are particularly relevant for complex traits such as heat stress tolerance. MTAs identified through the remaining two models cannot be deemed false positives given the appreciable phenotypic variances (R^2 values) observed for these MTAs. Hence, either FarmCPU or BLINK can be employed for GWAS (Merrick et al., 2022). Only nine previously reported QTLs were identified through the present study, indicating that most of the MTAs were novel. Two QTNs for grain yield per plant, five QTNs for spikelet fertility, and two QTNs for STISF were found to co-localize with previously reported QTLs. QTN for grain yield per plant identified through FarmCPU *qHTGY10.1* was co-localized with *qhr3-1* reported for heat tolerance by Cao et al. (2003) and *qHTGY11.1* overlapped with *qADL09-11* identified by Tazib et al. (2015) for anther dehiscence length. The QTN, *qHTST11.1* (or *qSTISF11.1* for STISF) for spikelet fertility identified in the present study through MLM coincided with two previously reported QTLs, *qHTSF11.2* (Ye et al., 2015a) for spikelet fertility and *qLD10*-11 (Tazib et al., 2015) for longitudinal dehiscence of anthers. Another QTN, *qHTSF4.1* (and *qSTISF4.1* for STISF), for spikelet fertility on chromosome 4 coincided with *SSPF4* (Xiao et al., 2011a) and *qPF4* (Xiao et al., 2011b) mapped for seed set percentage and pollen fertility under heat stress, respectively. Furthermore, *qHTSF7.1* identified through FarmCPU for spikelet fertility co-localized with *qAL10-7* reported for anther length (Tazib et al., 2015). The MTA *qHTST2.1*, identified through BLINK, coincided with *qtl_2.2* reported for absolute spikelet fertility (Jagadish et al., 2010).

The regions identified as significant MTAs were analyzed using the genome browser on the Rice Genome Annotation Project website in order to identify putative candidate genes. The majority of those genes identified for grain yield per plant are involved in stress signaling, such as the glucosylceramidase (GCD) gene, zinc ion binding protein, DUF1336 domain-containing protein, cytochrome P450, diacylglycerol kinase, Ser/Thr protein kinase, OsWAK116–OsWAK receptor-like cytoplasmic kinase OsWAK-RLCK, and so forth. A DUF1336 domain-containing protein was identified on chromosome 8. Several transcriptomic studies reported differential expression of DUF domain-containing proteins under heat stress at the anthesis stage in rice, indicating their possible role in heat stress responses (Endo et al., 2009; Zhang et al., 2012; González-Schain et al., 2016). The cytochrome P450 proteins (like the one found on chromosome 8) have varied roles in the biotic and abiotic stress responses of plants, particularly in the synthesis of secondary metabolites (Jun et al., 2015; Pandian et al., 2020). Their relevance under heat stress has already been demonstrated in rice (Endo et al., 2009; González-Schain et al., 2016; Wang et al., 2019), mustard (Rahaman et al., 2018), *Panicum virgatum* (Li et al., 2013), *Rhazya stricta* (Obaid et al., 2016), and *Panicum maximum* (Wedow et al., 2019). Four kinase-encoding genes were found—one on chromosome 8, two on chromosome 1, and one on chromosome 11. Protein kinases are central to abiotic stress signal transduction pathways. Among these, *OsWAK116*, receptor-like wall-associated kinase proteins can be presumed to be a transducer of heat stress signal between the cytoplasm and cell wall (Zhang S. et al., 2005).

Several genes identified for spikelet fertility are involved in protein chaperoning pathways central to plant heat stress responses. The proteins containing the tetratricopeptide repeat (TRP) motif (like the one found on chromosome 11) are, in general, involved in protein–protein interactions which unite into multi-protein complexes to assist plants in warding off external stresses (Paeng et al., 2020). An E3 ubiquitin ligase identified on chromosome 1 might be involved in the degradation of denatured proteins due to heat stress. Ubiquitin ligases are frequently implicated under heat stress in rice (Mittal et al., 2012; Zhang

et al., 2012; González-Schain et al., 2016). Broad-complex Tramtrack and the Bric-a-brac (BTB) domain/POX virus and Zinc finger (POZ) (BTB/POX) domain-containing proteins (chromosome 6) are implicated in transcriptional regulation and protein degradation (He et al., 2019). These BPM proteins are reported to negatively regulate the degradation of DREB2A and contribute to plant thermo-tolerance (Morimoto et al., 2017). The U-box domain-containing proteins (chromosome 3), called PUBs (plant U-box proteins), are a part of the ubiquitin–proteasome system (UPS) involved in the targeted ubiquitination and degradation of proteins when exposed to various environmental stresses (Hur et al., 2012; Byun et al., 2017; Ryu et al., 2019).

Ankyrin (ANK) repeat-containing proteins (chromosome 2) are involved in various protein–protein interactions (Michaely and Bennett, 1992; Li et al., 2006) and protein chaperoning (Shen et al., 2010) with putative roles in pollen germination and pollen tube growth in lily (Huang et al., 2006) and rice (Huang et al., 2009). These ANK repeat proteins also showed differential expression under heat stress at the anthesis stage in rice, in line with the present finding (González-Schain et al., 2016). RING finger proteins (chromosome 2) are a family of zinc finger proteins, with the majority being U3 ubiquitin ligases (Ciechanover 1998), which stem from the RING domain. There are reports of association of RING finger proteins in heat stress responses in rice. For instance, both *Oryza sativa* heat- and cold-induced 1(*OsHCI1*) and *Oryza sativa* heat-induced RING finger protein 1(*OsHIRP1*) act as E3 ligases and positively regulate heat stress responses (Lim et al., 2013; Kim et al., 2019). The rice *OsRZFP34* gene and *HEAT TOLERANCE AT SEEDLING STAGE* (*OsHTAS*) gene (an E3 ligase) improve high-temperature tolerance (Hsu et al., 2014; Liu et al., 2016). Corroborating this, the candidate gene of thermo-tolerance 1 (TT1) QTL identified from *O. glaberrima* encodes the α2 subunit of the 26S proteasome (Li et al., 2015). Several other genes such as DDT domain-containing protein, ATROPGEF7/ROPGEF7, phytochrome C, OsTOP6A1–Topoisomerase 6 subunit A homolog 1, transcription elongation factor protein, C3HC4-type domain-containing protein, PPR repeat-containing protein, universal stress protein domain-containing protein, deoxyuridine 5-triphosphate nucleotidohydrolase and WD domain, and G-beta repeat domain-containing protein were also found in the regions associated with spikelet fertility with putative functions in plant abiotic stress responses. An *in silico* expression analysis of these genes revealed an interesting pattern. The majority of the genes identified for spikelet fertility showed upregulation in reproductive and grain tissues, while the genes identified for grain yield showed higher expression levels in vegetative organs. Furthermore, nearly 10 genes matched with two heat stress-related DEGs identified in the previous transcriptomics studies (Cai et al., 2020; Liu et al., 2020). However, the logarithm of foldchange (Log_2FC) values of the genes are <2.5 except for LOC_Os01g04580, a Ser/Thr protein kinase which was significantly downregulated at 6 h of heat stress exposure. This is understandable since the genotypes (SDWG005 and MH101) used in these two reports were completely different from those in the current study.

Thus, in the present study, significant reductions in grain yield and spikelet fertility were observed among the rice genotypes characterized for RSHT. GWAS identified many novel MTAs, explaining high phenotypic variance for both these traits. There was a clear difference between the effects of alternate alleles, indicating their significance in governing RSHT in rice. The majority of the candidate genes identified around these MTAs were either directly or indirectly involved in heat stress and other abiotic stress responses, which are valuable candidates for marker-assisted selection for the improvement of heat stress tolerance at a reproductive stage after further validation in future. Some uncharacterized genes were also observed for both grain yield and spikelet fertility, whose function needs to be elucidated in future studies.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The name of the repository and link to the data can be found below: ICAR; https://krishi.icar.gov.in/jspui/handle/123456789/31947.

AUTHOR CONTRIBUTIONS

SG and AS conceived the idea and formulated the research plan. KR carried out the research work and prepared the manuscript. MN and PB assisted in the execution of field trials. KV, RE, and KA assisted in data analysis. BH shared seed materials and generated the genotypic data of the germplasm. SG, KV, and MP improved the manuscript. AS provided overall guidance in each of these activities.

FUNDING

This research was funded by the ICAR-funded network project on the National Innovations on Climate Resilient Agriculture (NICRA) and the National Agricultural Higher Education Project (NAHEP) – Center for Advanced Agricultural Science and Technology (CAAST) project on “Genomics Assisted Breeding for Crop Improvement” of the World Bank and the Indian Council of Agricultural Research, New Delhi.

ACKNOWLEDGMENTS

The study is part of the PhD research of the first author. RKT is thankful to the Director, ICAR-CSSRI, Karnal, and ICAR-IARI, New Delhi, for facilitating his PhD.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.876522/full#supplementary-material

REFERENCES

Bac-Molenaar, J. A., Fradin, E. F., Becker, F. F. M., Rienstra, J. A., van der Schoot, J., Vreugdenhil, D., et al. (2015). Genome-Wide Association Mapping of Fertility Reduction upon Heat Stress Reveals Developmental Stage-Specific QTLs in *Arabidopsis thaliana*. *Plant Cell* 27, 1857–1874. doi:10.1105/tpc.15.00248

Baker, J. T., Allen, L. H., Jr, and Boote, K. J. (1992). Temperature Effects on Rrice at Elevated CO_2 Concentration. *J. Exp. Bot.* 43, 959–964. doi:10.1093/jxb/43.7.959

Baytar, A. A., Peynircioğlu, C., Sezener, V., Basal, H., Frary, A., Frary, A., et al. (2018). Genome-Wide Association Mapping of Yield Components and Drought Tolerance-Related Traits in Cotton. *Mol. Breed.* 38, 1–16. doi:10.1007/s11032-018-0831-0

Bheemanahalli, R., Sathishraj, R., Tack, J., Nalley, L. L., Muthurajan, R., and Jagadish, K. S. V. (2016). Temperature Thresholds for Spikelet Sterility and Associated Warming Impacts for Sub-Tropical Rice. *Agric. For. Meteorol.* 221, 122–130. doi:10.1016/j.agrformet.2016.02.003

Bollinedi, H., Yadav, A. K., Vinod, K. K., Gopala Krishnan, S., Bhowmick, P. K., Nagarajan, M., et al. (2020). Genome-Wide Association Study Reveals Novel Marker-Trait Associations (MTAs) Governing the Localization of Fe and Zn in the Rice Grain. *Front. Genet.* 11, 213. doi:10.3389/fgene.2020.00213

Byun, M. Y., Cui, L. H., Oh, T. K., Jung, Y.-J., Lee, A., Park, K. Y., et al. (2017). Homologous U-Box E3 Ubiquitin Ligases OsPUB2 and OsPUB3 Are Involved in the Positive Regulation of Low Temperature Stress Response in Rice (*Oryza sativa* L.). *Front. Plant Sci.* 8, 16. doi:10.3389/fpls.2017.00016

Cai, Z., He, F., Feng, X., Liang, T., Wang, H., Ding, S., et al. (2020). Transcriptomic Analysis Reveals Important Roles of Lignin and Flavonoid Biosynthetic Pathways in Rice Thermotolerance During Reproductive Stage. *Front. Genet.* 11, 562937. doi:10.3389/fgene.2020.562937

Cao, L., Zhao, J., Zhan, X., Li, D., He, L., and Cheng, S. (2003). Mapping QTLs for Heat Tolerance and Correlation Between Heat Tolerance and Photosynthetic Rate in Rice. *Chin. J. Rice Sci.* 17, 223–227.

Cheabu, S., Panichawong, N., Rattanametta, P., Wasuri, B., Kasemsap, P., Arikit, S., et al. (2019). Screening for Spikelet Fertility and Validation of Heat Tolerance in a Large Rice Mutant Population. *Rice Sci.* 26, 229–238. doi:10.1016/j.rsci.2018.08.008

Chen, E., Huang, X., Tian, Z., Wing, R. A., and Han, B. (2019). The Genomics of *Oryza* Species Provides Insights into Rice Domestication and Heterosis. *Annu. Rev. Plant Biol.* 70, 639–665. doi:10.1146/annurev-arplant-050718-100320

Chen, J.-H., Chen, S.-T., He, N.-Y., Wang, Q.-L., Zhao, Y., Gao, W., et al. (2020). Nuclear-Encoded Synthesis of the D1 Subunit of Photosystem II Increases Photosynthetic Efficiency and Crop Yield. *Nat. Plants* 6, 570–580. doi:10.1038/s41477-020-0629-z

Chidambaranathan, P., Balasubramaniasai, C., Behura, N., Purty, M., Samantaray, S., Subudhi, H., et al. (2021). Effects of High Temperature on Spikelet Sterility in Rice (*Oryza sativa* L.): Association Between Molecular Markers and Allelic Phenotypic Effect in Field Condition. *Genet. Resour. Crop Evol.* 68, 1923–1935. doi:10.1007/s10722-021-01106-7

Ciechanover, A. (1998). The Ubiquitin-Proteasome Pathway: On Protein Death and Cell Life. *EMBO J.* 17, 7151–7160. doi:10.1093/emboj/17.24.7151

de Brito, G. G., Moraes, Í. L. d., Moura, D. S., Fagundes, P. R. R., Campos, A. D., Andres, A., et al. (2019). Non-Invasive Physiological Approaches for Plant Phenotyping: Rice Responses to Heat Stress. *Jas* 11, 453–464. doi:10.5539/jas.v11n3p453

Endo, M., Tsuchiya, T., Hamada, K., Kawamura, S., Yano, K., Ohshima, M., et al. (2009). High Temperatures Cause Male Sterility in Rice Plants with Transcriptional Alterations During Pollen Development. *Plant Cel. Physiol.* 50, 1911–1922. doi:10.1093/pcp/pcp135

Fernandez, G. C. (1992). "Effective Selection Criteria for Assessing Plant Stress Tolerance," in Proceeding of the International Symposium on Adaptation of Vegetables and Other Food Crops in Temperature and Water Stress, Shanhua, Taiwan, 257–270.

Gahlaut, V., Jaiswal, V., Singh, S., Balyan, H. S., and Gupta, P. K. (2019). Multi-Locus Genome Wide Association Mapping for Yield and its Contributing Traits in Hexaploid Wheat Under Different Water Regimes. *Sci. Rep.* 9, 1–15. doi:10.1038/s41598-019-55520-0

GoI (2019). *Agricultural Statistics at a Glance*. India: Directorate of Economics & Statistics.

González-Schain, N., Dreni, L., Lawas, L. M. F., Galbiati, M., Colombo, L., Heuer, S., et al. (2016). Genome-Wide Transcriptome Analysis During Anthesis Reveals New Insights into the Molecular Basis of Heat Stress Responses in Tolerant and Sensitive Rice Varieties. *Plant Cel. Physiol.* 57, 57–68. doi:10.1093/pcp/pcv174

Guo, J., Shi, W., Guo, J., Yue, L., Zhuang, L., Zhang, W., et al. (2020). Genome-Wide Association Studies on Heat Stress Tolerance During Grain Development in Wheat (*Triticum aestivum* L.). Research Square [Preprint]. doi:10.21203/rs.2.20460/v1

Hada, A., Dutta, T. K., Singh, N., Singh, B., Rai, V., Singh, N. K., et al. (2020). A Genome-Wide Association Study in Indian Wild Rice Accessions for Resistance to the Root-Knot Nematode *Meloidogyne graminicola*. *PLoS ONE* 15, e0239085. doi:10.1371/journal.pone.0239085

He, Y.-M., Liu, K.-K., Zhang, H.-X., Cheng, G.-X., Ali, M., Ul Haq, S., et al. (2019). Contribution of CaBPM4, a BTB Domain-Containing Gene, to the Response of Pepper to Phytophthora Capsici Infection and Abiotic Stresses. *Agronomy* 9, 417. doi:10.3390/agronomy9080417

Hirabayashi, H., Sasaki, K., Kambe, T., Gannaban, R. B., Miras, M. A., Mendioro, M. S., et al. (2014). *qEMF3*, a Novel QTL for the Early-Morning Flowering Trait from Wild Rice, *Oryza officinalis*, to Mitigate Heat Stress Damage at Flowering in Rice, *O. Sativa* L. *J. Exp. Bot.* 66, 1227–1236. doi:10.1093/jxb/eru474

Hossain, M. (2021). *As Rising 'heat Shocks' Ruin rice Crops, Bangladesh Faces Hunger Risk*. Thomson Reuters Foundation News, Thomson Reuters Foundation. Available at: https://news.trust.org/item/20210528065131-wnlhg/(Accessed May, 2021).

Hsu, K.-H., Liu, C.-C., Wu, S.-J., Kuo, Y.-Y., Lu, C.-A., Wu, C.-R., et al. (2014). Expression of a Gene Encoding a rice RING Zinc-finger Protein, OsRZFP34, Enhances Stomata Opening. *Plant Mol. Biol.* 86, 125–137. doi:10.1007/s11103-014-0217-6

Huang, J., Chen, F., Del Casino, C., Autino, A., Shen, M., Yuan, S., et al. (2006). An Ankyrin Repeat-Containing Protein, Characterized as a Ubiquitin Ligase, Is Closely Associated with Membrane-Enclosed Organelles and Required for Pollen Germination and Pollen Tube Growth in Lily. *Plant Physiol.* 140, 1374–1383. doi:10.1104/pp.105.074922

Huang, J., Zhao, X., Yu, H., Ouyang, Y., Wang, L., and Zhang, Q. (2009). The Ankyrin Repeat Gene Family in Rice: Genome-Wide Identification, Classification and Expression Profiling. *Plant Mol. Biol.* 71, 207–226. doi:10.1007/s11103-009-9518-6

Huang, L., Sun, Y., Peng, S., and Wang, F. (2016). Genotypic Differences of Japonica Rice Responding to High Temperature in China. *Agron.j.* 108, 626–636. doi:10.2134/agronj2015.0507

Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). BLINK: A Package for the Next Level of Genome-Wide Association Studies with Both Individuals and Markers in the Millions. *Gigascience* 8, giy154. doi:10.1093/gigascience/giy154

Hur, Y.-J., Yi, Y. B., Lee, J. H., Chung, Y. S., Jung, H. W., Yun, D. J., et al. (2012). Molecular Cloning and Characterization of OsUPS, a U-Box Containing E3 Ligase Gene that Respond to Phosphate Starvation in rice (*Oryza sativa*). *Mol. Biol. Rep.* 39, 5883–5888. doi:10.1007/s11033-011-1399-5

Im, E.-S., Pal, J. S., and Eltahir, E. A. B. (2017). Deadly Heat Waves Projected in the Densely Populated Agricultural Regions of South Asia. *Sci. Adv.* 3, e1603322. doi:10.1126/sciadv.1603322

IPCC (2014). "Summary for Policymakers," in *Climate Change 2013: The Physical Science Basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Editors T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, and Allen (Cambridge, United Kingdom and New York, NY: Cambridge University Press).

IRRI (2002). *Standard Evaluation System for rice (SES)*. Los Banos, Philippines: International Rice Research Institute.

Ishimaru, T., Hirabayashi, H., Ida, M., Takai, T., San-Oh, Y. A., Yoshinaga, S., et al. (2010). A Genetic Resource for Early-Morning Flowering Trait of Wild Rice *Oryza officinalis* to Mitigate High Temperature-Induced Spikelet Sterility at Anthesis. *Ann. Bot.* 106, 515–520. doi:10.1093/aob/mcq124

Jagadish, S. V. K., Cairns, J., Lafitte, R., Wheeler, T. R., Price, A. H., and Craufurd, P. Q. (2010). Genetic Analysis of Heat Tolerance at Anthesis in Rice. *Crop Sci.* 50, 1633–1641. doi:10.2135/cropsci2009.09.0516

Julia, C., and Dingkuhn, M. (2013). Predicting Temperature Induced Sterility of Rice Spikelets Requires Simulation of Crop-Generated Microclimate. *Eur. J. Agron.* 49, 50–60. doi:10.1016/j.eja.2013.03.006

Jun, X. U., Wang, X. Y., and Guo, W. Z. (2015). The Cytochrome P450 Superfamily: Key Players in Plant Development and Defense. *J. Integr. Agric.* 14, 1673–1686. doi:10.1016/S2095-3119(14)60980-1

Khanzada, H., Wassan, G. M., He, H., Mason, A. S., Keerio, A. A., Khanzada, S., et al. (2020). Differentially Evolved Drought Stress Indices Determine the Genetic Variation of Brassica Napus at Seedling Traits by Genome-Wide Association Mapping. *J. Adv. Res.* 24, 447–461. doi:10.1016/j.jare.2020.05.019

Kim, J. H., Lim, S. D., and Jang, C. S. (2019). *Oryza sativa* Heat-Induced RING Finger Protein 1 (*OsHIRP1*) Positively Regulates Plant Response to Heat Stress. *Plant Mol. Biol.* 99, 545–559. doi:10.1007/s11103-019-00835-9

Krishnan, R., Sanjay, J., Gnanaseelan, C., Mujumdar, M., Kulkarni, A., and Chakraborty, S. (2020). *Assessment of Climate Change Over the Indian Region*. Singapore: SpringerOpen. doi:10.1007/978-981-15-4327-2

Lafarge, T., Bueno, C., Frouin, J., Jacquin, L., Courtois, B., and Ahmadi, N. (2017). Genome-Wide Association Analysis for Heat Tolerance at Flowering Detected a Large Set of Genes Involved in Adaptation to thermal and Other Stresses. *PLoS ONE* 12, e0171254. doi:10.1371/journal.pone.0171254

Li, J., Mahajan, A., and Tsai, M.-D. (2006). Ankyrin Repeat: A Unique Motif Mediating Protein−Protein Interactions. *Biochemistry* 45, 15168–15178. doi:10.1021/bi062188q

Li, Y.-F., Wang, Y., Tang, Y., Kakani, V. G., and Mahalingam, R. (2013). Transcriptome Analysis of Heat Stress Response in Switchgrass (Panicum virgatumL.). *BMC Plant Biol.* 13, 153. doi:10.1186/1471-2229-13-153

Li, X.-M., Chao, D.-Y., Wu, Y., Huang, X., Chen, K., Cui, L.-G., et al. (2015). Natural Alleles of a Proteasome α2 Subunit Gene Contribute to Thermotolerance and Adaptation of African rice. *Nat. Genet.* 47, 827–833. doi:10.1038/ng.3305

Li, C., Wang, D., PengChen, S. Y., Chen, Y., Su, P., Chen, J., et al. (2019). Genome-wide Association Mapping of Resistance against rice Blast Strains in South China and Identification of a New *Pik* Allele. *Rice* 12, 47. doi:10.1186/s12284-019-0309-7

Lim, S. D., Cho, H. Y., Park, Y. C., Ham, D. J., Lee, J. K., and Jang, C. S. (2013). The rice RING finger E3 Ligase, *OsHCI1*, Drives Nuclear export of Multiple Substrate Proteins and its Heterogeneous Overexpression Enhances Acquired Thermotolerance. *J. Exp. Bot.* 64, 2899–2914. doi:10.1093/jxb/ert143

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: Genome Association and Prediction Integrated Tool. *Bioinform* 28, 2397–2399. doi:10.1093/bioinformatics/bts444

Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-wide Association Studies. *Plos Genet.* 12, e1005767. doi:10.1371/journal.pgen.1005767

Liu, G., Zha, Z., Cai, H., Qin, D., Jia, H., Liu, C., et al. (2020). Dynamic Transcriptome Analysis of Anther Response to Heat Stress during Anthesis in Thermotolerant Rice (*Oryza Sativa* L.). *Ijms* 21, 1155. doi:10.3390/ijms21031155

Lyman, N. B., Jagadish, K. S. V., Nalley, L. L., Dixon, B. L., and Siebenmorgen, T. (2013). Neglecting rice Milling Yield and Quality Underestimates Economic Losses from High-Temperature Stress. *PLoS ONE* 8, e72157. doi:10.1371/journal.pone.0072157

Matsui, T., and Omasa, K. (2002). Rice (Oryza Sativa L.) Cultivars Tolerant to High Temperature at Flowering: Anther Characteristics. *Ann. Bot.* 89, 683–687. doi:10.1093/aob/mcf112

Matsui, T., Omasa, K., and Horie, T. (2000). High Temperature at Flowering Inhibit Swelling of Pollen Grains, a Driving Force for Thecae Dehiscence in Rice (*Oryza sativa* L.). *Plant Prod. Sci.* 3, 430–434. doi:10.1626/pps.3.430

Matsui, T., Omasa, K., and Horie, T. (2001). The Difference in Sterility Due to High Temperatures During the Flowering Period Among Japonica-Rice Varieties. *Plant Prod. Sci.* 4, 90–93. doi:10.1626/pps.4.90

McCouch, S. R., Wright, M. H., Tung, C.-W., Maron, L. G., McNally, K. L., Fitzgerald, M., et al. (2016). Open Access Resources for Genome-Wide Association Mapping in Rice. *Nat. Commun.* 7, 10532. doi:10.1038/ncomms10532

Melandri, G., Prashar, A., McCouch, S. R., Van Der Linden, G., Jones, H. G., Kadam, N., et al. (2020). Association Mapping and Genetic Dissection of Drought-Induced Canopy Temperature Differences in Rice. *J. Exp. Bot.* 71, 1614–1627. doi:10.1093/jxb/erz527

Merrick, L. F., Burke, A. B., Zhang, Z., and Carter, A. H. (2022). Comparison of Single-Trait and Multi-Trait Genome-wide Association Models and Inclusion of Correlated Traits in the Dissection of the Genetic Architecture of a Complex Trait in a Breeding Program. *Front. Plant Sci.* 12, 772907. doi:10.3389/fpls.2021.772907

Michaely, P., and Bennett, V. (1992). The ANK Repeat: A Ubiquitous Motif Involved in Macromolecular Recognition. *Trends Cel. Biol.* 2, 127–129. doi:10.1016/0962-8924(92)90084-z

Misra, G., Anacleto, R., Badoni, S., Butardo, V., Jr, Molina, L., Graner, A., et al. (2019). Dissecting the Genome-Wide Genetic Variants of Milling and Appearance Quality Traits in Rice. *J. Exp. Bot.* 70, 5115–5130. doi:10.1093/jxb/erz256

Mittal, D., Madhyastha, D. A., and Grover, A. (2012). Genome-Wide Transcriptional Profiles During Temperature and Oxidative Stress Reveal Coordinated Expression Patterns and Overlapping Regulons in Rice. *PloS ONE* 7, e40899. doi:10.1371/journal.pone.0040899

Morimoto, K., Ohama, N., Kidokoro, S., Mizoi, J., Takahashi, F., Todaka, D., et al. (2017). BPM-CUL3 E3 Ligase Modulates Thermotolerance by Facilitating Negative Regulatory Domain-Mediated Degradation of DREB2A in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 114, E8528–E8536. doi:10.1073/pnas.1704189114

Murray, M. G., and Thompson, W. F. (1980). Rapid Isolation of High Molecular Weight Plant DNA. *Nucl. Acids Res.* 8, 4321–4326. doi:10.1093/nar/8.19.4321

Obaid, A. Y., Sabir, J. S. M., Atef, A., Liu, X., Edris, S., El-Domyati, F. M., et al. (2016). Analysis of Transcriptional Response to Heat Stress in *Rhazya stricta*. *BMC Plant Biol.* 16, 252. doi:10.1186/s12870-016-0938-6

Paeng, S. K., Kang, C. H., Chi, Y. H., Chae, H. B., Lee, E. S., Park, J. H., et al. (2020). AtTPR10 Containing Multiple ANK and TPR Domains Exhibits Chaperone Activity and Heat-Shock Dependent Structural Switching. *Appl. Sci.* 10, 1265. doi:10.3390/app10041265

Pandian, B. A., Sathishraj, R., Djanaguiraman, M., Prasad, P. V. V., and Jugulam, M. (2020). Role of Cytochrome P450 Enzymes in Plant Stress Response. *Antioxidants* 9, 454. doi:10.3390/antiox9050454

Peng, S., Huang, J., Sheehy, J. E., Laza, R. C., Visperas, R. M., Zhong, X., et al. (2004). Rice Yields Decline with Higher Night Temperature from Global Warming. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9971–9975. doi:10.1073/pnas.0403720101

Pradhan, S. K., Barik, S. R., Sahoo, A., Mohapatra, S., Nayak, D. K., Mahender, A., et al. (2016). Population Structure, Genetic Diversity and Molecular Marker-Trait Association Analysis for High Temperature Stress Tolerance in Rice. *PLoS ONE* 11, e0160027. doi:10.1371/journal.pone.0160027

Prasad, P. V. V., Boote, K. J., Allen, L. H., Sheehy, J. E., and Thomas, J. M. G. (2006). Species, Ecotype and Cultivar Differences in Spikelet Fertility and Harvest Index of Rice in Response to High Temperature Stress. *Field Crops Res.* 95, 398–411. doi:10.1016/j.fcr.2005.04.008

Rahaman, M., Mamidi, S., and Rahman, M. (2018). Genome-Wide Association Study of Heat Stress-Tolerance Traits in Spring-Type *Brassica napus* L. Under Controlled Conditions. *Crop J.* 6, 115–125. doi:10.1016/j.cj.2017.08.003

Ravikiran, K. T., Gopala Krishnan, S., Vinod, K. K., Dhawan, G., Dwivedi, P., Kumar, P., et al. (2020). A Trait Specific QTL Survey Identifies NL44, a NERICA Cultivar as a Novel Source for Reproductive Stage Heat Stress Tolerance in Rice. *Plant Physiol. Rep.* 25, 664–676. doi:10.1007/s40502-020-00547-z

Rohilla, M., Singh, N., Mazumder, A., Sen, P., Roy, P., Chowdhury, D., et al. (2020). Genome-Wide Association Studies Using 50 K Rice Genic SNP Chip Unveil Genetic Architecture for Anaerobic Germination of Deep-Water Rice Population of Assam, India. *Mol. Genet. Genomics* 295, 1211–1226. doi:10.1007/s00438-020-01690-w

Ryu, M. Y., Cho, S. K., Hong, Y., Kim, J., Kim, J. H., Kim, G. M., et al. (2019). Classification of Barley U-Box E3 Ligases and Their Expression Patterns in Response to Drought and Pathogen Stresses. *BMC Genomics* 20, 1–15. doi:10.1186/s12864-019-5696-z

Satake, T., and Yoshida, S. (1978). High Temperature-Induced Sterility in Indica Rices at Flowering. *Jpn. J. Crop Sci.* 47, 6–17. doi:10.1626/jcs.47.6

Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., et al. (2012). An Efficient Multi-Locus Mixed-Model Approach for Genome-Wide Association Studies in Structured Populations. *Nat. Genet.* 44, 825–830. doi:10.1038/ng.2314

Sharma, D., Jaiswal, J. P., Gahtyari, N. C., Chauhan, A., Chhabra, R., Saripalli, G., et al. (2020). Population Structure, Marker-Trait Association and Identification of Candidate Genes for Terminal Heat Stress Relevant Traits in Bread Wheat (*Triticum aestivum* L. Em Thell). *Plant Genet. Resour.* 18, 168–178. doi:10.1017/S1479262120000131

Shen, G., Kuppu, S., Venkataramani, S., Wang, J., Yan, J., Qiu, X., et al. (2010). ANKYRIN REPEAT-CONTAINING PROTEIN 2A is an Essential Molecular Chaperone for Peroxisomal Membrane-Bound ASCORBATE PEROXIDASE3 in *Arabidopsis*. *Plant Cell* 22, 811–831. doi:10.1105/tpc.109.065979

Singh, N., Jayaswal, P. K., Panda, K., Mandal, P., Kumar, V., Singh, B., et al. (2015). Single-Copy Gene Based 50 K SNP Chip for Genetic Studies and Molecular Breeding in Rice. *Sci. Rep.* 5, 11600. doi:10.1038/srep11600

Sukkeo, S., Rerkasem, B., and Jamjod, S. (2017). Heat Tolerance in Thai Rice Varieties. *ScienceAsia* 43, 61–69. doi:10.2306/scienceasia1513-1874.2017.43.061

Tadesse, W., Suleiman, S., Tahir, I., Sanchez-Garcia, M., Jighly, A., Hagras, A., et al. (2019). Heat-Tolerant QTLs Associated with Grain Yield and its Components in Spring Bread Wheat Under Heat-Stressed Environments of Sudan and Egypt. *Crop Sci.* 59, 199–211. doi:10.2135/cropsci2018.06.0389

Tafesse, E. G., Gali, K. K., Lachagari, V. B. R., Bueckert, R., and Warkentin, T. D. (2020). Genome-Wide Association Mapping for Heat Stress Responsive Traits in Field Pea. *Ijms* 21, 2043. doi:10.3390/ijms21062043

Tang, L., Zhang, F., LiuSun, A. J., Sun, J., Mei, S., Wang, X., et al. (2019). Genome-Wide Association Analysis Dissects the Genetic Basis of the Grain Carbon and Nitrogen Contents in Milled Rice. *Rice* 12, 101. doi:10.1186/s12284-019-0362-2

Tazib, T., Kobayashi, Y., Koyama, H., and Matsui, T. (2015). QTL Analyses for Anther Length and Dehiscence at Flowering as Traits for the Tolerance of Extreme Temperatures in Rice (*Oryza sativa* L.). *Euphytica* 203, 629–642. doi:10.1007/s10681-014-1291-1

Tenorio, F. A., Ye, C., Redona, E., Sierra, S., Laza, M., and Argayoso, M. A. (2013). Screening Rice Genetic Resources for Heat Tolerance. *SABRAO J. Breed. Genet.* 45, 371–381.

Thomson, M. J., Singh, N., Dwiyanti, M. S., Wang, D. R., Wright, M. H., Perez, F. A., et al. (2017). Large-Scale Deployment of a Rice 6 K SNP Array for Genetics and Breeding Applications. *Rice* 10, 40. doi:10.1186/s12284-017-0181-2

Wang, D., Qin, B., Li, X., Tang, D., Zhang, Y. e., Cheng, Z., et al. (2016). Nucleolar DEAD-Box RNA Helicase TOGR1 Regulates Thermotolerant Growth as a Pre-rRNA Chaperone in Rice. *Plos Genet.* 12, e1005844. doi:10.1371/journal.pgen.1005844

Wang, Y., Zhang, Y., Zhang, Q., Cui, Y., Xiang, J., Chen, H., et al. (2019). Comparative Transcriptome Analysis of Panicle Development Under Heat Stress in Two Rice (*Oryza sativa* L.) Cultivars Differing in Heat Tolerance. *PeerJ* 7, e7595. doi:10.7717/peerj.7595

Wassmann, R., Jagadish, S. V. K., Sumfleth, K., Pathak, H., Howell, G., Ismail, A., et al. (2009). Chapter 3 Regional Vulnerability of Climate Change Impacts on Asian Rice Production and Scope for Adaptation. *Adv. Agron.* 102, 91–133. doi:10.1016/S0065-2113(09)01003-7

Wedow, J. M., Yendrek, C. R., Mello, T. R., Creste, S., Martinez, C. A., and Ainsworth, E. A. (2019). Metabolite and Transcript Profiling of Guinea Grass (*Panicum maximum* Jacq) Response to Elevated [CO_2] and Temperature. *Metabolomics* 15, 51. doi:10.1007/s11306-019-1511-8

Wu, Y.-S., and Yang, C.-Y. (2019). Ethylene-Mediated Signaling Confers Thermotolerance and Regulates Transcript Levels of Heat Shock Factors in Rice Seedlings Under Heat Stress. *Bot. Stud.* 60, 1–12. doi:10.1186/s40529-019-0272-z

Xiao, Y. H., Pan, Y., Luo, L., Zhang, G., Deng, H., Dai, L., et al. (2011a). Quantitative Trait Loci Associated with Seed Set Under High Temperature Stress at the Flowering Stage in Rice (*Oryza sativa* L.). *Euphytica* 178, 331–338. doi:10.1007/s10681-010-0300-2

Xiao, Y. H., Pan, Y., Luo, L. H., Deng, H. B., Zhang, G. L., Tang, W. B., et al. (2011b). Quantitative Trait Loci Associated with Pollen Fertility Under High Temperature Stress at Flowering Stage in Rice (*Oryza sativa* L). *Rice Sci.* 18, 1–7. doi:10.1016/s1672-6308(11)60028-0

Xu, Y., Zhang, L., Ou, S., Wang, R., Wang, Y., Chu, C., et al. (2020). Natural Variations of *SLG1* Confer High-Temperature Tolerance in *Indica* Rice. *Nat. Commun.* 11, 1–13. doi:10.1038/s41467-020-19320-9

Ye, C., Argayoso, M. A., Redoña, E. D., Sierra, S. N., Laza, M. A., Dilla, C. J., et al. (2012). Mapping QTL for Heat Tolerance at Flowering Stage in Rice Using SNP Markers. *Plant Breed* 131, 33–41. doi:10.1111/j.1439-0523.2011.01924.x

Ye, C., Tenorio, F. A., Argayoso, M. A., Laza, M. A., Koh, H.-J., Redoña, E. D., et al. (2015a). Identifying and Confirming Quantitative Trait Loci Associated with Heat Tolerance at Flowering Stage in Different Rice Populations. *BMC Genet.* 16, 41. doi:10.1186/s12863-015-0199-7

Ye, C., Tenorio, F. A., Redoña, E. D., Morales–Cortezano, P. S., Cabrega, G. A., Jagadish, K. S. V., et al. (2015b). Fine-Mapping and Validating *qHTSF4.1* to Increase Spikelet Fertility Under Heat Stress at Flowering in Rice. *Theor. Appl. Genet.* 128, 1507–1517. doi:10.1007/s00122-015-2526-9

Yoshida, S., Satake, T., and Mackill, D. (1981). High Temperature Stress. *IRRI Res. Paper Ser.* 67, 1–15.

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., et al. (2006). A Unified Mixed-Model Method for Association Mapping that Accounts for Multiple Levels of Relatedness. *Nat. Genet.* 38, 203–208. doi:10.1038/ng1702

Yu, H., Xie, W., Li, J., Zhou, F., and Zhang, Q. (2014). A Whole-Genome SNP Array (RICE6K) for Genomic Breeding in Rice. *Plant Biotechnol. J.* 12, 28–37. doi:10.1111/pbi.12113

Yuan, J., Wang, X., ZhaoKhan, Y. N. U., Khan, N. U., Zhao, Z., Zhang, Y., et al. (2020). Genetic Basis and Identification of Candidate Genes for Salt Tolerance in Rice by GWAS. *Sci. Rep.* 10, 9958. doi:10.1038/s41598-020-66604-7

Zhang, S., Chen, C., Li, L., Meng, L., Singh, J., Jiang, N., et al. (2005). Evolutionary Expansion, Gene Structure, and Expression of the Rice Wall-Associated Kinase Gene Family. *Plant Physiol.* 139, 1107–1124. doi:10.1104/pp.105.069005

Zhang, Y.-M., Mao, Y., Xie, C., Smith, H., Luo, L., and Xu, S. (2005). Mapping Quantitative Trait Loci Using Naturally Occurring Genetic Variance Among Commercial Inbred Lines of Maize (*Zea mays* L.). *Genetics* 169, 2267–2275. doi:10.1534/genetics.104.033217

Zhang, X., Li, J., Liu, A., Zou, J., Zhou, X., Xiang, J., et al. (2012). Expression Profile in Rice Panicle: Insights into Heat Response Mechanism at Reproductive Stage. *Plos ONE* 7, e49652. doi:10.1371/journal.pone.0049652

Zhang, Y.-W., Tamba, C. L., Wen, Y.-J., Li, P., Ren, W.-L., Ni, Y.-L., et al. (2020). mrMLM v4.0.2: An R Platform for Multi-Locus Genome-Wide Association Studies. *Genomics, Proteomics Bioinformatics* 18, 481–487. doi:10.1016/j.gpb.2020.06.006

Zhao, K., Tung, C.-W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., et al. (2011). Genome-Wide Association Mapping Reveals a Rich Genetic Architecture of Complex Traits in *Oryza sativa*. *Nat. Commun.* 2, 467. doi:10.1038/ncomms1467

Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., et al. (2018). Pan-Genome Analysis Highlights the Extent of Genomic Variation in Cultivated and Wild Rice. *Nat. Genet.* 50, 278–284. doi:10.1038/s41588-018-0041-z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ravikiran, Gopala Krishnan, Abhijith, Bollinedi, Nagarajan, Vinod, Bhowmick, Pal, Ellur and Singh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 03 June 2022
doi: 10.3389/fgene.2022.887643

A General-Purpose Machine Learning R Library for Sparse Kernels Methods With an Application for Genome-Based Prediction

Osval Antonio Montesinos López[1], Brandon Alejandro Mosqueda González[2], Abel Palafox González[3], Abelardo Montesinos López[3] and José Crossa[4,5]*†*

[1]Facultad de Telemática, Universidad de Colima, Colima, Mexico, [2]Centro de Investigación en Computación (CIC), Instituto Politécnico Nacional (IPN), México City, Mexico, [3]Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI), Universidad de Guadalajara, Guadalajara, Mexico, [4]International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, [5]Colegio de Postgraduados, Montecillo, Mexico

OPEN ACCESS

Edited by:
Ravi Valluru,
University of Lincoln, United Kingdom

Reviewed by:
Alencar Xavier,
Corteva Agriscience™, United States
Moyses Nascimento,
Universidade Federal de Viçosa, Brazil

***Correspondence:**
Abelardo Montesinos López
aml_uach2004@hotmail.com
José Crossa
j.crossa@cgiar.org

†ORCID:
José Crossa
orcid.org/0000-0001-9429-5855

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 01 March 2022
Accepted: 02 May 2022
Published: 03 June 2022

Citation:
Montesinos López OA, Mosqueda González BA, Palafox González A, Montesinos López A and Crossa J (2022) A General-Purpose Machine Learning R Library for Sparse Kernels Methods With an Application for Genome-Based Prediction. Front. Genet. 13:887643. doi: 10.3389/fgene.2022.887643

The adoption of machine learning frameworks in areas beyond computer science have been facilitated by the development of user-friendly software tools that do not require an advanced understanding of computer programming. In this paper, we present a new package (sparse kernel methods, SKM) software developed in R language for implementing six (generalized boosted machines, generalized linear models, support vector machines, random forest, Bayesian regression models and deep neural networks) of the most popular supervised machine learning algorithms with the optional use of sparse kernels. The SKM focuses on user simplicity, as it does not try to include all the available machine learning algorithms, but rather the most important aspects of these six algorithms in an easy-to-understand format. Another relevant contribution of this package is a function for the computation of seven different kernels. These are Linear, Polynomial, Sigmoid, Gaussian, Exponential, Arc-Cosine 1 and Arc-Cosine L (with L = 2, 3, . . .) and their sparse versions, which allow users to create kernel machines without modifying the statistical machine learning algorithm. It is important to point out that the main contribution of our package resides in the functionality for the computation of the sparse version of seven basic kernels, which is indispensable for reducing computational resources to implement kernel machine learning methods without a significant loss in prediction performance. Performance of the SKM is evaluated in a genome-based prediction framework using both a maize and wheat data set. As such, the use of this package is not restricted to genome prediction problems, and can be used in many different applications.

Keywords: r package, machine learning, kernel, supervised learning, sparse kernels, genome-base prediction

INTRODUCTION

Machine learning has become the main approach for solving complex, data-based problems and it is being used everywhere from devices and digital services such as smartphones and websites, to scientific research in various fields (Wang et al., 2016; Ott et al., 2020; Shahin et al., 2020; Montesinos-López et al., 2021a). As machine learning research has progressed, so has the supply

and demand of software that facilitates its implementation. For this reason, numerous open-source packages for data related tasks and machine learning algorithms have become even more prevalent (Abadi et al., 2015; Wickham et al., 2015; Pandas development team, 2020).

One of the most used programming languages for data analysis is R (R Core Team, 2021) due to its statistical computing focus, free and open-source software and the thousands of packages that extend its power to all kind of analysis and related tasks of data science. In fact, it is difficult to find a machine learning algorithm not implemented within an R package. Likewise, it can even be said that some of the R packages contain more complete/specialized implementations (Ishwaran et al., 2008; Friedman et al., 2010; Meyer et al., 2019) than those available in other programming languages. As machine learning is strongly based on statistical models and R is the *de facto* language for statistics research, those who embark on machine learning will encounter R at some point.

Most R packages of machine learning algorithms include one type of model or a family of similar models. While using R packages have clear advantages, there are some challenges. For example, each package has been developed by different authors and there is no standardized code style guideline. This complicates the use of packages since it requires users to learn the expected data format, the name and expected parameters and the code convention (if any) in order to train a model or retrieve outputs. In addition, several complementing packages may be needed to perform cross validation of models, hyperparameter tuning, and compute accuracy metrics, among others. There are some libraries that seek to integrate a wide range of tools needed for machine learning in one place, such as scikit-learn (Pedregosa et al., 2011) in Python; H2O in Java (with both R and Python versions); and caret (Kuhn, 2016), mlr3 (Lang et al., 2019) and tidy models (Kuhn and Wickham, 2020) in R. All these options have their own philosophy, and they were designed using diverse approaches to implement machine learning models.

We consider the mlr3 as the most powerful R package for machine learning because of its potential scope. The mlr3 package is an object-oriented solution for machine learning focused on extensibility since it does not implement any model itself, but rather provide a unified interface for many existing packages in R. While this is a major advantage, such an approach does not completely solve the dependency of other packages, which require knowledge of both the package that implements the model and mlr3. It is worthwhile to learn how to use all the components in the mlr3 environment because it also provides efficient implementation of most data related tasks, parallelization, hyperparameter tuning and feature selection, among others. However, it takes times getting accustomed to the way mlr3 works and how things are defined in parts with the object-oriented paradigm, which is not so common in R programming. Nevertheless, this learning curve is relatively short.

Alternatively, we have caret and tidy models providing their own standardized interface, which is a very important factor in a good quality software. Like mlr3, these two packages use other third party packages of machine learning algorithms in tandem to train models as they provide with different options for the same algorithm. Caret is the oldest of these three packages, and as such, it still enjoys considerable popularity. Notwithstanding, the major advantage of tidy models is that they belong to the tidy verse, a collection of R packages tailored for data science that share an underlying design philosophy, grammar and data structures (Wickham et al., 2019); consequently, if users are familiar with tidy verse packages, they will naturally start using tidy models.

In the current paper, we present SKM (Sparse Kernels Methods), a new R package for machine learning that includes functions for model training, tuning, prediction, metrics evaluation and sparse kernels computation. The main goal of this package is to provide a stand-alone (or self-contained) R software, focused on the austere implementation of only six basic supervised learning models that are easy to understand from the user´s point-of- view. We will focus specifically on six types of supervised models, which are explained in the next section. The model functions in SKM were designed with simplicity in mind, and as such, the parameters, hyperparameters and tuning specifications are defined directly when calling the function; subsequently, users can understand how the package works by observing a handful of examples. Furthermore, we strive to provide clear documentation following a base convention in the functions. Likewise, all the parameters are validated with checkmate software (Lang, 2017) to inform the user when an error occurs through meaningful error messages—something that many other packages neglect. The most important hyperparameters of each model can be tuned with two different methods: grid search and Bayesian optimization (Osborne et al., 2009) based on the code of Bayesian Optimization package (Yan, 2016). Although Bayesian optimization is a very popular and effective method of tuning, the mlr3 and caret packages do not offer this option.

Kernels have proven to be useful in helping the conventional machine learning algorithms capture non-linear patterns in data (Montesinos-López et al., 2021b; Montesinos-López et al., 2022a). In addition to capturing complex non-linear patterns, the sparse kernel version of kernel methods can also save significant computational resources without a relevant loss in prediction accuracy (Montesinos-López et al., 2021b; Montesinos-López, et al., 2022a). In this paper by sparse kernels we define those kernels that are built with only a fraction of the total amount of inputs by assuming that the input matrix is a sparse matrix, that is, a matrix that contain many information with zeros. For this reason, the term level of compression, here is used, as one minus the proportion of the total lines (or rows) used to compute the sparse kernels thus representing the level of dimensionality reduction reached by using these sparse kernels. To the best of our knowledge, there is no existing R package for the computation of dense kernels and sparse kernels (that compress the dimension of the dense kernels), which is the added value of SKM and what gives it its name. The approach of sparse kernels implemented in the SKM library is based on the method proposed in Cuevas et al. (2020).

As software developers and consumers, we are aware of the importance of sharing our work with the community, and as such, SKM is a completely open-source software released under the GNU Lesser General Public License v3.0 (LGPLv3). As such,

TABLE 1 | Models that can be implemented in the SKM library.

Model	Name	Package of origin	Function in SKM	Response variables
M1	Generalized boosted machines	gbm (Greenwell et al., 2020)	generalized_ boosted_machine ()	Binary, categorical and continuous; only univariate
M2	Generalized linear models	Glmnet (Friedman et al., 2010)	Generalized _linear_model ( )	Binary, categorical, continuous, and count; univariate and multivariate only for continuous response variables
M3	Support vector machines	e1071 (Meyer, et al., 2019)	Support _vector_machine ()	Binary, categorical and continuous, only for univariate response variables
M4	Random forest	RandomForestSRC (Ishwaran, et al., 2008)	random_forest ()	Binary, categorical and continuous, univariate and multivariate
M5	Bayesian regression models	BGLR (Perez and de los Campos, 2014)	bayesian_model ()	Binary, categorical and continuous, univariate and multivariate only for continuous response variables
M6	Deep neural networks	keras (Allaire and Chollet, 2016)	deep_learning ()	Binary, categorical, continuous, and count; univariate and multivariate for all response variables

anyone can explore the source code, make modifications and build on it to develop other tools.

MACHINE LEARNING ALGORITHMS

The SKM package includes six different functions of supervised machine learning algorithms. **Table 1** shows the six models that can be implemented under the SKM package, and the package of origin that each of these models uses, in addition to the function to implement these models in the SKM library.

It is important to point out that all models that can be implemented in the SKM library will be able to implement the seven kernels methods and its sparse versions explained in the next section, whereas in the case of deep neural networks (M6), only fully connected networks can be implemented. Under the Bayesian methods, Bayesian Ridge regression (BRR), Bayes A (Bayes_A), Bayes B (Bayes_B), Bayes C (Bayes_C), Bayesian Lasso (Bayes_Lasso) and the best linear unbiased predictor (GBLUP) in its Bayesian version (BGBLUP) can be implanted. It should be highlighted that the six models that can be implemented in the SKM library, including all the Bayesian methods available in model M5, can also work with kernels. First, the matrix of inputs (X) is created; then the square root of the kernel is computed; next the design matrix of lines is post-multiplied by the square root of the kernel; and finally this design matrix is used as input in any of the six models when kernels are used. The exception is under the BGBLUP in model M5, where the computed kernels are directly used.

The additional layer of abstraction allows all functions to share the same data input format. Internally, data is adapted to the expected format of each package, where the result and prediction objects returned by these functions are also in the same format. Another benefit of these functions is that some parameters that can be inferred from data itself do not need to be supplied by the user, rather they are set automatically. For example, the family parameter of glmnet package which has to be "Gaussian" for continuous response variables, "binomial" for binary variables, "multinomial" for categorical response variables and "Poisson" for count variables, can be inferred from the response variable. In addition, the same functions permit hyperparameter tuning in an easy and user-friendly format without the need to call another function or initiate another object. In theory, as with all packages that internally call functions of other packages, ease of use and extended functionality is expected to improve with a slight increase in computational demand for the extra operations required. Furthermore, since these operations are of computationally low cost, there is no significant loss of power.

Supplementary Appendix SA included some comparative examples of the equivalent implementation of some machine learning models with mlr3, SKM and randomForestSRC, the original package.

SPARSE KERNELS

As Montesinos-López et al. (2021b) point out, kernel methods transform the independent variables (inputs) using a kernel function, followed by the application of conventional machine learning techniques to the transformed data to achieve better results, mainly when the inputs contain non-linear patterns. Kernel methods are excellent options in terms of computational efficiency when managing large, complex data that show non-linear patterns; likewise they can be used with any type of predictive machine. Consequently, we have included the kernelize function in SKM that can compute the same 7 kernels and their sparse versions as described in Montesinos-López et al. (2021b): Linear, Polynomial, Sigmoid, Gaussian, Exponential, Arc-Cosine 1 and Arc-Cosine L (with L = 2, 3, . . .). The kernel computation is independent from the model fitting process, which allows the kernelize function to be used with other packages or conversely, the machine learning algorithms implementation of SKM can be used without kernels.

Next the algorithm to approximate the kernels, here called sparse kernels is described in general terms. We assume that the response variable (y) is associated to the genomic effects (u) as:

$$\boldsymbol{y} = \mu\mathbf{1} + \boldsymbol{u} + \boldsymbol{e} \quad (1)$$

where μ is the overall mean, $\mathbf{1}$ is the vector of ones, and $\boldsymbol{y}$ is the vector of size n. Moreover, $\boldsymbol{u}$ is the vector of genomic effects $\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma_u^2\boldsymbol{K})$, where σ_u^2 is the genomic variance component and matrix $\boldsymbol{K}$ is the

TABLE 2 | Prediction performance of the Wheat data set for each environment and across environments (Global) of each of the six models.

Model	Metric	E1		E2		E3		E4		Global	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
M1	MAAPE	0.7307	0.0069	0.6852	0.0210	0.6993	0.0188	0.6922	0.0104	0.7082	0.0090
M1	MAE	0.6801	0.0133	0.6360	0.0272	0.6644	0.0314	0.5935	0.0162	0.5955	0.0123
M1	MSE	0.7359	0.0210	0.6931	0.0494	0.7908	0.0820	0.6007	0.0332	0.5951	0.0233
M1	NRMSE	0.8575	0.0174	0.8173	0.0263	0.8763	0.0271	0.7915	0.0157	0.8316	0.0081
M1	RMSE	0.8575	0.0123	0.8304	0.0298	0.8839	0.0490	0.7738	0.0221	0.7708	0.0155
M2	MAAPE	0.7134	0.0118	0.7506	0.0107	0.7460	0.0118	0.7635	0.0049	0.7500	0.0112
M2	MAE	0.7023	0.0303	0.7116	0.0292	0.7670	0.0283	0.7179	0.0131	0.6748	0.0163
M2	MSE	0.7845	0.0527	0.8980	0.0646	0.9876	0.0746	0.8581	0.0237	0.7645	0.0289
M2	NRMSE	0.8747	0.0193	0.9429	0.0200	0.9622	0.0145	0.9344	0.0123	0.9311	0.0116
M2	RMSE	0.8836	0.0305	0.9450	0.0349	0.9909	0.0380	0.9260	0.0129	0.8737	0.0167
M3	MAAPE	0.7857	0.0038	0.7835	0.0056	0.7877	0.0010	0.7848	0.0019	0.7856	0.0015
M3	MAE	0.7675	0.0186	0.7972	0.0267	0.7766	0.0082	0.7805	0.0228	0.7341	0.0133
M3	MSE	0.9014	0.0324	1.0875	0.0656	0.9583	0.0268	1.0724	0.0343	0.9035	0.0271
M3	NRMSE	0.9997	0.0013	1.0013	0.0021	1.0012	0.0045	1.0027	0.0017	1.0004	0.0004
M3	RMSE	0.9488	0.0171	1.0409	0.0320	0.9785	0.0138	1.0350	0.0166	0.9501	0.0142
M4	MAAPE	0.7161	0.0134	0.6835	0.0169	0.6902	0.0128	0.6898	0.0204	0.6965	0.0100
M4	MAE	0.6733	0.0273	0.6258	0.0081	0.7060	0.0196	0.5945	0.0094	0.5864	0.0083
M4	MSE	0.7063	0.0450	0.6793	0.0049	0.8221	0.0494	0.6291	0.0409	0.5769	0.0186
M4	NRMSE	0.8472	0.0163	0.8105	0.0159	0.8470	0.0178	0.7963	0.0151	0.8123	0.0080
M4	RMSE	0.8387	0.0264	0.8242	0.0030	0.9050	0.0275	0.7915	0.0252	0.7592	0.0123
M5	MAAPE	0.7133	0.0108	0.6956	0.0107	0.7233	0.0067	0.7455	0.0046	0.7211	0.0043
M5	MAE	0.7141	0.0183	0.6336	0.0116	0.6846	0.0272	0.6572	0.0291	0.6156	0.0056
M5	MSE	0.7987	0.0387	0.6587	0.0170	0.7696	0.0607	0.7021	0.0639	0.6183	0.0104
M5	NRMSE	0.8796	0.0230	0.8168	0.0220	0.8742	0.0121	0.8808	0.0148	0.8547	0.0081
M5	RMSE	0.8927	0.0212	0.8113	0.0104	0.8744	0.0355	0.8346	0.0369	0.7862	0.0066
M6	MAAPE	0.7056	0.0071	0.6991	0.0107	0.7149	0.0132	0.7058	0.0037	0.7075	0.0067
M6	MAE	0.6938	0.0144	0.6358	0.0204	0.6802	0.0280	0.6327	0.0103	0.6170	0.0122
M6	MSE	0.8160	0.0499	0.6978	0.0452	0.7807	0.0678	0.7183	0.0226	0.6645	0.0355
M6	NRMSE	0.8918	0.0067	0.8385	0.0188	0.8889	0.0230	0.8534	0.0167	0.8669	0.0119
M6	RMSE	0.9016	0.0279	0.8336	0.0267	0.8802	0.0386	0.8471	0.0133	0.8140	0.0217

Generalized boosted machines (M1), generalized linear models (M2), support vector machines (M3), random forest (M4), Bayesian regression models (M5) and deep neural networks (M6). The tuning process was done under the Bayesian optimization framework. Mean is the average of the five partitions for each metric and SE denotes the standard error for each metric. E1-E4 denotes location1, location2, location3 and location4.

dense kernel of order $n \times n$ constructed with any of the kernel methods explained above. The random residuals are assumed independent with normal distribution $\boldsymbol{e} \sim N(\boldsymbol{0}, \sigma_e^2\boldsymbol{I})$, where σ_e^2 is the error variance. The dense kernel, $\boldsymbol{K}$, can be approximated as $\boldsymbol{K} \approx \boldsymbol{Q} = \boldsymbol{K}_{n,m}\boldsymbol{K}_{m,m}^{-1}\boldsymbol{K}_{n,m}'$ (Williams and Seeger, 2001), where $\boldsymbol{Q}$ will have the rank of $\boldsymbol{K}_{m,m}$, that is, m. The computation of this kernel is facilitated since it is not necessary to compute and store the original matrix $\boldsymbol{K}$, since only $\boldsymbol{K}_{m,m}$ and $\boldsymbol{K}_{n,m}$ are required. This approximation of the dense kernel (which we call sparse kernel) use m out of n lines to compute $\boldsymbol{K}_{m,m}^{-1}$, then an eigen-value-decomposition of $\boldsymbol{K}_{m,m}^{-1} = \boldsymbol{U}\boldsymbol{S}^{-1/2}\boldsymbol{S}^{-1/2}\boldsymbol{U}'$ is used, where $\boldsymbol{U}$ are the eigen vectors of order $m \times m$ and $S_{m,m}$ is a diagonal matrix of order $m \times m$ with the eigen values ordered from largest to smallest. Next, these values are substituted in $\boldsymbol{Q} = \boldsymbol{K}_{n,m}\boldsymbol{U}\boldsymbol{S}^{-1/2}\boldsymbol{S}^{-1/2}\boldsymbol{U}'\boldsymbol{K}_{n,m}'$ resulting in $\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma_u^2\boldsymbol{K}_{n,m}\boldsymbol{U}\boldsymbol{S}^{-1/2}\boldsymbol{S}^{-1/2}\boldsymbol{U}'\boldsymbol{K}_{n,m}')$, and thus, model (1) can be expressed as:

$$\boldsymbol{y} = \mu 1_{\boldsymbol{n}} + \boldsymbol{P}\boldsymbol{f} + \boldsymbol{\varepsilon} \quad (2)$$

Model (2) is similar to model (1), except that $\boldsymbol{f}$ is a vector of order $m \times 1$ with a normal distribution of the form $\boldsymbol{f} \sim N(\boldsymbol{0}, \sigma_f^2\boldsymbol{I}_{m,m})$, where $\boldsymbol{P} = \boldsymbol{K}_{m,n}\boldsymbol{U}\boldsymbol{S}^{-1/2}$ is now the design matrix. This implies estimating only m effects that are projected into the n dimensional space in order to predict $\boldsymbol{u}$ and explain $\boldsymbol{y}$. Note that model (2) can be implemented under a conventional mixed model framework or under any statistical machine learning algorithm assuming that the $\boldsymbol{f}$ term of **Equation 2** is a fixed effect. For example, under a linear kernel the $\boldsymbol{K}_{m,n}$ and $\boldsymbol{K}_{m,m}$ can be computed as $\boldsymbol{K}_{m,m} = \frac{\boldsymbol{X}_{m,p}\boldsymbol{X}_{m,p}'}{p}$ and $\boldsymbol{K}_{n,m} = \frac{\boldsymbol{X}_{n,p}\boldsymbol{X}_{m,p}'}{p}$ respectively, where $\boldsymbol{X}_{m,p}$ is the centered and scaled matrix of markers with m lines and p markers, and $\boldsymbol{X}_{n,p}$ is the centered and scaled matrix of markers with n lines and p markers. In summary, according to Cuevas et al. (2020), the approximation described above consists of the following steps:

Step 1: Computing the following matrices, matrix $\boldsymbol{K}_{m,m}$ from m lines of the training set.
Step 2: Constructing matrix $\boldsymbol{K}_{n,m}$
Step 3: Eigen value decomposition of $\boldsymbol{K}_{m,m}$
Step 4: Computing matrix $\boldsymbol{P} = \boldsymbol{K}_{n,m}\boldsymbol{U}\boldsymbol{S}^{-1/2}$.
Step 5: Fitting the model under any of the above mentioned statistical machine learning using $\boldsymbol{P} = \boldsymbol{K}_{n,m}\boldsymbol{U}\boldsymbol{S}^{-1/2}$ as design matrix and $\boldsymbol{y}$ as response variable.

One of the major advantages of the sparse kernels is data dimensionality reduction since the number of parameters to be estimated is reduced significantly in comparison to the dense kernels. This is useful when working with high dimensional data where the number of columns is considerably greater than the number of rows, as there is few data, and the training process of the model is more efficient. More details about the kernels and the approximated kernels, here called sparse kernels, that were implemented in the SKM library can be found in detail in Montesinos-López et al. (2021b) and Montesinos-López, et al. (2022a).

In **Supplementary Appendix SB** we have included some examples of how to use the kernelize function of SKM to compute the different kernels and their sparse versions.

EVALUATION METRICS

Evaluating models' performance is an important task of all machine learning workflows. For this reason, in SKM we have included functions of the most popular metrics to evaluate models' performance for both regression and classification problems. The regression metrics included are: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Normalized Root Mean Squared Error (NRMSE, with four types of normalization: by standard deviation, mean, range and interquartile range), Mean Absolute Error (MAE) and Mean Arctangent Absolute Percentage Error (MAAPE). The classification metrics included are: accuracy, specificity, sensitivity, Kappa coefficient, Brier score, Matthews correlation coefficient, precision, recall, Area Under the ROC Curve (ROC-AUC), Precision-Recall Area Under the Curve (PR-AUC), F1 score and a function to compute the confusion matrix. In addition to the functions already mentioned, the wrapper functions numeric summary and categorical summary compute all the regression and classification metrics to obtain a complete summary of the model's performance in a simple function. More details about most of these metrics can be found in chapter 4 (Overfitting, model tuning and evaluation of prediction performance) of the book Multivariate statistical machine learning methods for genomic prediction (Montesinos-López et al., 2022b).

As expected, all these metric functions work in harmony with the machine learning algorithm functions since they use the same data format; no extra data processing is necessary when they are used correctly. This does not limit or complicate their use with other packages, as shown in the detailed documentation provided.

Supplementary Appendices SA, SB include examples of some metric functions that receive the observed and predicted values (or probabilities in classification) and return a numeric value.

INSTALLATION

SKM is a package built for the R ecosystem. As an open source project, the package has first been published in a GitHub repository at https://github.com/brandon-mosqueda/SKM where the full source code and another option of installing the development version (and most updated) can be found. This development version may include corrections of reported bugs and new functionalities, among others. Likewise, in the repository users can also find a place to report bugs or contribute to the project. In order to install the development version, the following commands must be executed in an R terminal.

```
devtools::install_github ("cran/randomForestSRC")
devtools::install_github ("gdlc/BGLR-R")
devtools::install_github ("rstudio/tensorflow")
```

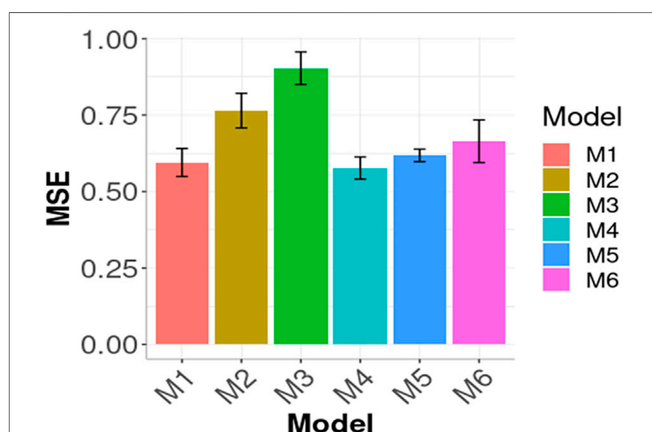


FIGURE 1 | Prediction performance in terms of Mean Squared Error of the six models (M1, M2, M3, M4, M5, M6) across environments (Global) in the wheat data. M1 denotes the generalized boosted machine model, M2 denotes the generalized linear model, M3 denotes the support vector machine model, M4 denotes the random forest model, M5 denotes the Bayesian regression model and M6 denotes the deep neural networks model.

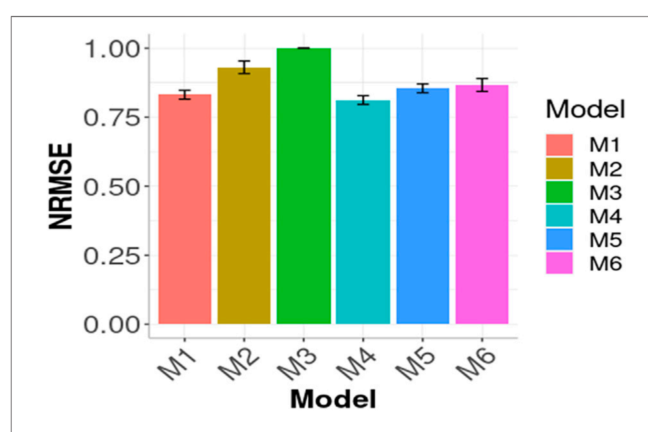


FIGURE 2 | Prediction performance in terms of Normalized Root Mean Squared Error of the six models (M1, M2, M3, M4, M5, M6) across environments (Global) in the wheat data. M1 denotes the generalized boosted machine model, M2 denotes the generalized linear model, M3 denotes the support vector machine model, M4 denotes the random forest model, M5 denotes the Bayesian regression model and M6 denotes the deep neural networks model.

TABLE 3 | Prediction performance of the Maize data set for each environment and across environments (Global) of each of the six models.

Model	Metric	E1		E2		E3		E4		Global	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
M1	MAE	0.2038	0.0024	0.4360	0.0122	0.2708	0.0035	0.5392	0.0088	0.3409	0.0047
M1	MSE	0.0700	0.0028	0.2991	0.0179	0.1125	0.0034	0.4670	0.0177	0.2059	0.0059
M1	NRMSE	0.8751	0.0259	0.9021	0.0077	0.9159	0.0135	0.9146	0.0196	0.8872	0.0147
M1	RMSE	0.2644	0.0051	0.5459	0.0166	0.3352	0.0050	0.6829	0.0129	0.4535	0.0066
M2	MAAPE	0.7672	0.0106	0.7787	0.0120	0.7734	0.0102	0.7580	0.0075	0.7565	0.0064
M2	MAE	0.2040	0.0067	0.4687	0.0144	0.2700	0.0060	0.5751	0.0166	0.3592	0.0072
M2	MSE	0.0713	0.0045	0.3460	0.0157	0.1131	0.0064	0.5281	0.0319	0.2336	0.0126
M2	NRMSE	0.9174	0.0119	0.9687	0.0057	0.9353	0.0163	0.9517	0.0070	0.9498	0.0040
M2	RMSE	0.2664	0.0083	0.5876	0.0134	0.3358	0.0095	0.7253	0.0225	0.4826	0.0127
M3	MAAPE	0.7861	0.0059	0.7829	0.0010	0.7871	0.0031	0.7870	0.0023	0.7852	0.0015
M3	MAE	0.2187	0.0054	0.4814	0.0130	0.2855	0.0049	0.6109	0.0151	0.3817	0.0069
M3	MSE	0.0847	0.0034	0.3701	0.0144	0.1287	0.0051	0.5861	0.0325	0.2603	0.0131
M3	NRMSE	1.0023	0.0027	1.0024	0.0031	0.9985	0.0010	1.0032	0.0013	1.0029	0.0014
M3	RMSE	0.2908	0.0059	0.6079	0.0119	0.3584	0.0070	0.7643	0.0215	0.5095	0.0126
M4	MAAPE	0.7450	0.0146	0.7615	0.0114	0.7432	0.0150	0.7418	0.0077	0.7444	0.0063
M4	MAE	0.2006	0.0053	0.4430	0.0100	0.2615	0.0069	0.5586	0.0119	0.3498	0.0034
M4	MSE	0.0678	0.0048	0.3073	0.0136	0.1070	0.0062	0.5041	0.0223	0.2215	0.0054
M4	NRMSE	0.8882	0.0082	0.9320	0.0076	0.9032	0.0173	0.9052	0.0076	0.8997	0.0042
M4	RMSE	0.2598	0.0091	0.5538	0.0122	0.3265	0.0096	0.7093	0.0157	0.4705	0.0058
M5	MAAPE	0.7853	0.0067	0.7601	0.0125	0.7600	0.0064	0.7275	0.0067	0.7483	0.0058
M5	MAE	0.2199	0.0033	0.4507	0.0074	0.2747	0.0086	0.5259	0.0089	0.3426	0.0060
M5	MSE	0.0796	0.0037	0.3269	0.0104	0.1166	0.0067	0.4500	0.0153	0.2099	0.0081
M5	NRMSE	0.9858	0.0087	0.9364	0.0111	0.9533	0.0223	0.8808	0.0063	0.9116	0.0065
M5	RMSE	0.2819	0.0065	0.5714	0.0091	0.3408	0.0100	0.6705	0.0113	0.4578	0.0089
M6	MAAPE	0.7980	0.0126	0.7792	0.0101	0.7819	0.0189	0.7681	0.0113	0.7747	0.0110
M6	MAE	0.2177	0.0075	0.4843	0.0157	0.2907	0.0094	0.5653	0.0200	0.3655	0.0112
M6	MSE	0.0798	0.0048	0.3775	0.0220	0.1398	0.0071	0.4992	0.0290	0.2396	0.0148
M6	NRMSE	0.9720	0.0214	1.0107	0.0114	1.0406	0.0199	0.9267	0.0249	0.9616	0.0215
M6	RMSE	0.2820	0.0084	0.6134	0.0180	0.3735	0.0096	0.7053	0.0212	0.4885	0.0155

Generalized boosted machines (M1), generalized linear models (M2), support vector machines (M3), random forest (M4), Bayesian regression models (M5) and deep neural networks (M6). The tuning process was done under the grid search framework. Mean is the average of the five partitions for each metric, SE denotes the standard error for each metric.

```
if (!require ("devtools")) {install.packages ("devtools")}
devtools::install_github ("brandon-mosqueda/SKM")
```

ILLUSTRATIVE EXAMPLES

Next, we will illustrate the use of the SKM library with two popular data sets in genomic selection using 5-random partitions to evaluate the prediction performance with the two available tuning options. The response variables in both datasets are numeric response variables, and as such, we present the prediction performance in terms of Mean Arctangent Absolute Percentage Error (MAAPE), Mean Absolute Error (MAE), Mean Squared Error (MSE) and Normalized Root Mean Squared Error (NRMSE). We have included a function in SKM to compute summaries for prediction performance with genomic selection data (summaries). This function requires a by data. frame with whole predictions in different folds, including genotype and environment information; this is used in all the examples described below.

Wheat Data

This data set was first used by Crossa et al. (2010) and Cuevas et al. (2016), Cuevas et al. (2017) and Cuevas et al. (2019) and is comprised of 599 wheat lines from the CIMMYT Global Wheat Program evaluated in four international environments representing four basic agroclimatic regions (mega-environments). The phenotypic trait considered for the 599 wheat lines evaluated in each of the four mega-environments was grain yield (GY). The 599 wheat lines were genotyped using 1447 Diversity Array Technology (DArT) markers generated by Triticarte Pty. Ltd.

In this example we evaluated the six models included in the package, each one using Bayesian optimization to tune its specific hyperparameters, with the exception of Bayesian methods (model M4), which do not require hyperparameter tuning. The cross-validation used to evaluate the predictions' accuracy was with five random (splits) partitions, where 80% of the data was used for training and 20% for the testing set, and the average of the five testing sets was reported as prediction performance. To tune the hyperparameters, an inner 5-fold cross validation was also used to evaluate each hyperparameter combination. It is important to point out that the inner 5-fold cross validation is implemented in each partition, which in this case, contains only 80% of the data. In this

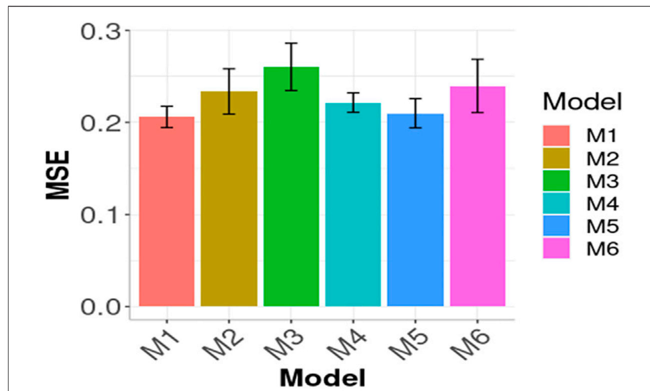


FIGURE 3 | Prediction performance in terms of Mean Squared Error of prediction of the six models (M1, M2, M3, M4, M5, M6) across environment (Global) in Maize data. M1 denotes the generalized boosted machine model, M2 denotes the generalized linear model, M3 denotes the support vector machine model, M4 denotes the random forest model, M5 denotes the Bayesian regression model and M6 denotes the deep neural networks model.

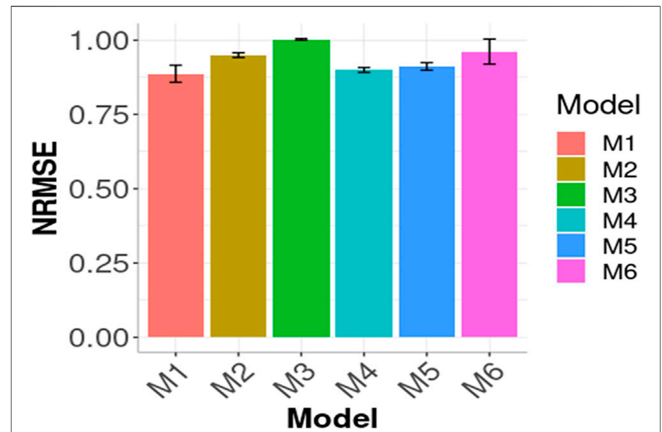


FIGURE 4 | Prediction performance in terms of Normalized Root Mean Squared Error of the six models (M1, M2, M3, M4, M5, M6) across environments (Global) in Maize data. M1 denotes the generalized boosted machine model, M2 denotes the generalized linear model, M3 denotes the support vector machine model, M4 denotes the random forest model, M5 denotes the Bayesian regression model and M6 denotes the deep neural networks model.

regard, each inner training contains only 64% of the data while the validation set contains only 16% of the data. In **Table 2**, the evaluation results are presented for the wheat data set, while the code for implementing the six models is given in **Supplementary Appendix SC**.

In **Figures 1**, **2** we compare the prediction performance of the six evaluated models across environments in terms of MSE and NRMSE, respectively. Both figures show a similar pattern in the prediction performance results. In terms of both metrics, M4, M1 and M5 produced the best prediction performance. In terms of MSE, the best model (M4) outperformed M1 by $(0.5951 - 0.5769) \times \frac{100}{0.5951} = 3.05\%$, M2 by $(0.7645 - 0.5769) \times \frac{100}{0.7645} = 32.51\%$, M3 (the worst) by $(0.9035 - 0.5769) \times \frac{100}{0.9035} = 36.14\%$, M5 by $(0.6183 - 0.5769) \times \frac{100}{0.6183} = 6.69\%$ and M6 by $(0.6645 - 0.5769) \times \frac{100}{0.6645} = 13.18\%$. Regarding NRMSE, the outperformance between models is not as large as in MSE terms. For example, the outperformance between the best (M4) and worst (M3) was $(1 - 0.8123) \times \frac{100}{1} = 18.77\%$, significantly different from the 36.14% in MSE terms. It should be noted that the model M5 was implemented in all the examples provided with Bayesian Ridge Regression (BRR; that works with the scaled matrix of markers Z), which is equivalent to BGBLUP [that works with the linear kernel computed as $ZZ^T/ncol(Z)$]. As mentioned before, the other Bayesian methods can be implemented by merely changing "BRR" in model to the other available options like: Bayes_A, Bayes_B, Bayes_C, Bayes_Lasso and BGLUP (See **Supplementary Appendix SB5**. Bayesian regression model).

Maize Data

This maize data set was included in Souza et al. (2017) and comes from USP (Universidad Sao Paulo). It consists of 722 (with 722 × 4 = 2888 observations) maize hybrids obtained by crossing 49 inbred lines. The hybrids were evaluated in four environments (E1-E4) in Piracicaba and Anhumas, São Paulo, Brazil, in 2016. The hybrids were evaluated using an augmented block design with two commercial hybrids as checks to correct for micro-environmental variation. At each site, two levels of nitrogen (N) fertilization were used. The experiment conducted under ideal N conditions received 100 kg ha-1 of N (30 kg ha-1 at sowing and 70 kg ha-1 in a coverage application) at the V8 plant stage, while the experiment with low N received 30 kg/ha at sowing. The parent lines were genotyped with an Affymetrix Axiom Maize Genotyping Array of 616 K SNPs. Markers with Minor Allele Frequency (MAF) of 0.05 were removed.

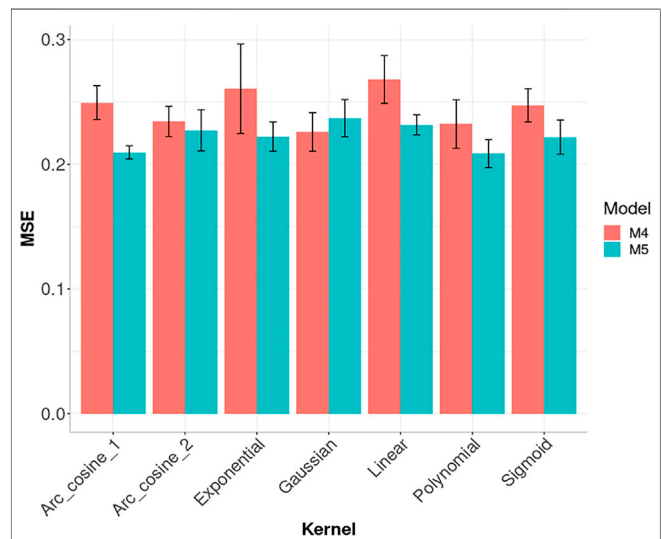


FIGURE 5 | Prediction performance across environments (Global) in Maize data in terms of Mean Square Error (MSE) of models M4 and M5 for seven kernel methods. M4 denotes the random forest model and M5 denotes the Bayesian regression model.

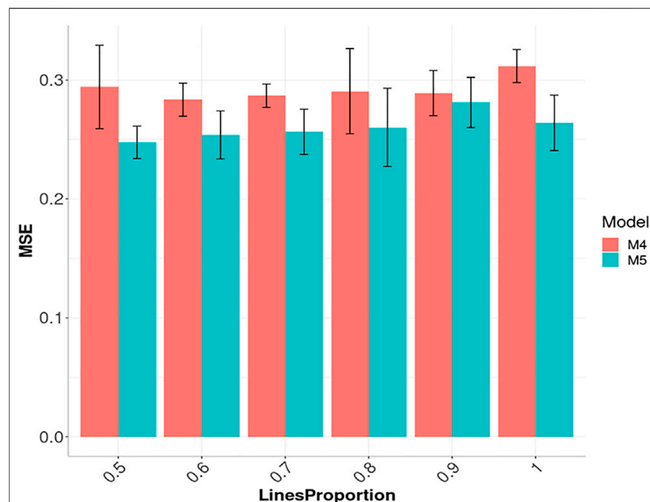


FIGURE 6 | Prediction performance across environments (Global) in Maize data in terms of Mean Square Error (MSE) of models M4 and M5 using the sparse Arc_cosine_1 kernel with six proportions of compression levels: 0.5, 0.4, 0.3, 0.2, 0.1 and 0, which correspond to using only the following proportions: 0.5, 0.6, 0.7, 0.8, 0.9 and 1 of the original lines (LinesProportion) for computing the kernels. M4 denotes the random forest model and M5 denotes the Bayesian regression model. The complement of level of compression level is equal to the proportion of lines used to compute the sparse kernel, that is, level of compression = 1 minus proportion of lines used to compute the sparse kernel.

After applying QC, 54,113 SNPs were available to make the predictions.

In this second example, we evaluated the same cases as the wheat data example using the grid search as a tuning strategy for the hyperparameters. Likewise, in this data set, the prediction performance was evaluated with five random partitions where 80% of the data was used for training and 20% for the testing set and the average of the five testing sets was reported as prediction performance. To tune the hyperparameters, an inner 5-fold cross validation was also used to evaluate each hyperparameter combination. In **Table 3** the evaluation results are shown for this data set (Maize). The complete R code for implementing the six models in the SKM library is provided in **Supplementary Appendix SD**.

In **Figures 3**, **4** the Global results of the maize data example are presented. **Figure 3** shows the prediction performance in terms of MSE and **Figure 4** the prediction performance in terms of NRMSE. According to **Figure 3**, the best Global results were observed in M1 with 0.2059 of MSE followed by M5 0.2099, that is $(0.2099 - 0.2059) \times \frac{100}{0.2099} = 1.9\%$ worst. M1 outperformed M4 by $(0.2215 - 0.2059) \times \frac{100}{0.2215} = 7.57\%$, M2 by $(0.2336 - 0.2059) \times \frac{100}{0.2336} = 11.85\%$, M6 by $(0.2396 - 0.2059) \times \frac{100}{0.2396} = 16.36\%$ and M3 (the worst) by $(0.2603 - 0.2059) \times \frac{100}{0.2603} = 20.89\%$. In **Figure 4** a similar pattern appears: M1 produced the best results since it has the lowest NRMSE. The only change in the order compared to that observed in **Figure 3** is that M4 outperformed M5 in terms of NRMSE. The remaining models' results agree with **Figure 3** given that the following best results in terms of NRMSE were obtained with M2, M3 and M6, respectively.

In **Figure 5**, we compared the performance of seven kernels for the maize data set: Linear, Polynomial, Sigmoid, Gaussian, Exponential, Arc-Cosine_1 and Arc-Cosine_2 for model M4 and M5. For model M5, the best prediction performance was observed under the Arc_cosine_1 and Polynomial kernel and the worst under the Gaussian kernel. While under model M4, the best performance in terms of MSE was observed under the Gaussian Kernel and the worst under the Linear kernel. The code used for implementing model M4 and M5 with the seven kernels are given in **Supplementary Appendix SE**. It is important to point out that in the SKM library it is possible to perform kernel and sparse kernels not only under the Bayesian BGBLUP method (a sub-model of model M5, that is implemented under a RKHS method in BGLR) but under the six models (M1 to M6) that can be implemented in this library. The kernels apart from one sub-model of model M5 (BGBLUP) are implemented not using as input directly the kernel, but with the square root of the kernel for this reason is possible to be implemented with all the six models. While the sparse kernels were implemented in a similar fashion but using the method explained above, proposed of Cuevas et al. (2020) and for this reason, also it is possible to be implemented with the six models here evaluated (M1, … , M6).

In **Figure 6**, we also provide the prediction accuracies in terms of MSE for models M4 and M5 with the Arc_cosine_1 kernel for six compression levels (0.5, 0.4, 0.3, 0.2, 0.1 and 0). It is important to point out that in **Figure 6**, the complement of the compression levels are given on the x-axis, which means the proportion of the columns (subsampling of lines without replacement; see Cuevas et al., 2020) of the complete (dense) kernel that are used as independent variables. We can observe in **Figure 6** that the best prediction performance for model M5 was obtained with the compression level at 50%, that is, when the model was trained with only half of the total columns of the complete kernel. However, the worst performance in model M5 was with a compression level of 10% (LinesProportion of 0.9). On the other hand, in model M4, the best and worst prediction performance in terms of MSE was observed under compression level of 0.4 (LinesProportion of 0.6) and 0 (LinesProportion of 1) respectively. The R code for reproducing the results given in **Figure 6** are provided in **Supplementary Appendix SF**.

Figure 6 for the Arc_cosine_1 sparse kernel, it is shown that even with the largest compression level, there is not a relevant loss in prediction accuracy. However, when the compression level is larger, less time (in hours) is required for the training process, and the reduction in time of execution is almost linear (**Figure 7A** for model M4 and **Figure 7B** for model M5 both for the Arc_cosine_1 sparse kernel). We can also observe in these Figures (**Figures 7A,B**) that the time required for the training process in model M5 is significantly less than the time required for model M4.

Figure 8 shows the prediction performance in terms of MSE for models M4 and M5 but now with the Gaussian kernel using the same six compression levels (0.5, 0.4, 0.3, 0.2, 0.1 and 0). For model M5, we did not observe any significant loss in terms of prediction performance with the six levels of compression levels evaluated. In model M5, we can observe that the best prediction performance was obtained with the largest compression level (0.5; LinesProportion of 0.5), but between the remaining compression levels we did not observed significant differences. The R code for reproducing the results given in **Figure 8** are provided in **Supplementary Appendix SF**.

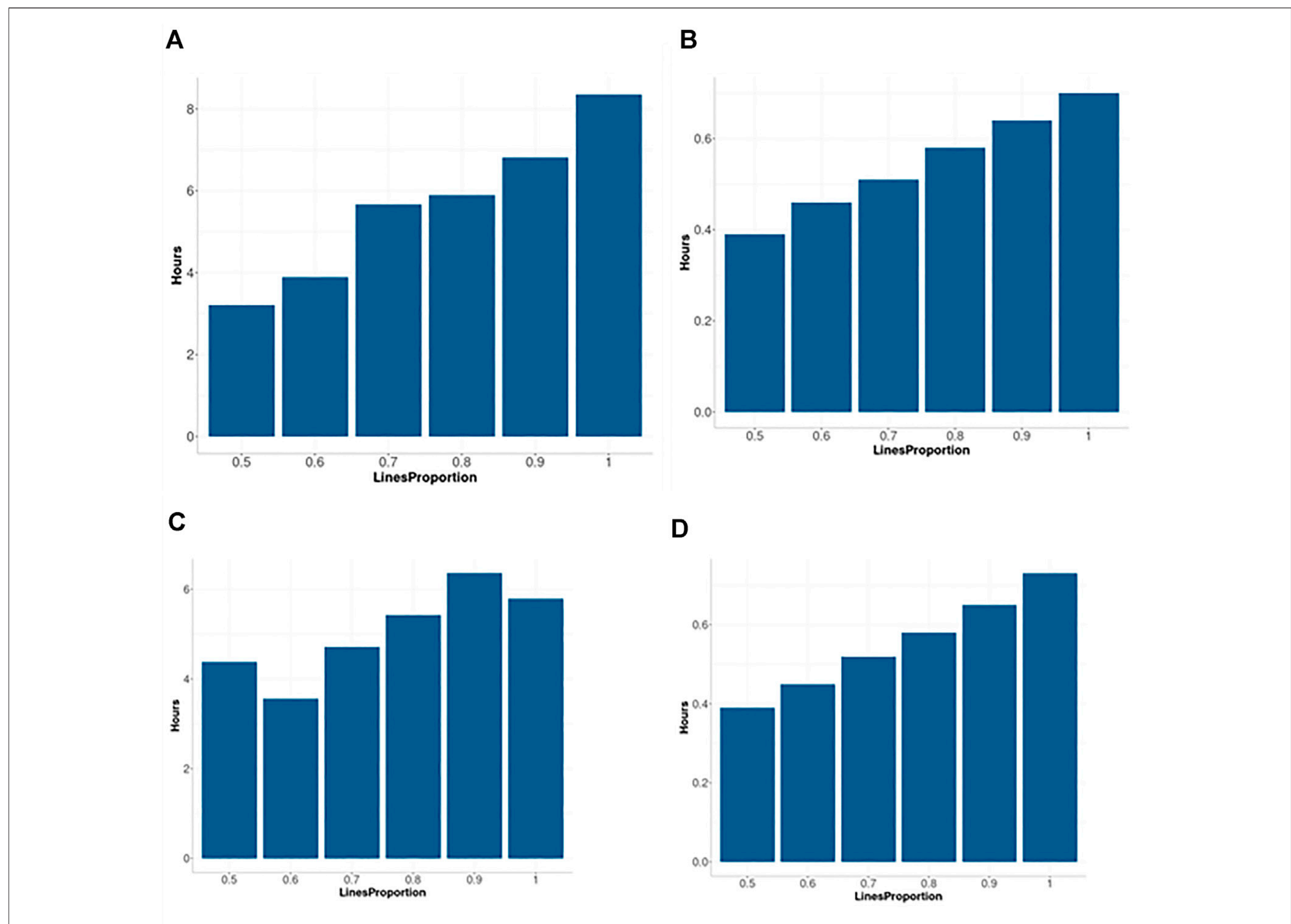


FIGURE 7 | Time in hours for implementing two sparse kernels (Arc_cosine_1 and Gaussian) with the maize data set as a function of the proportion of the compression level (0.5, 0.4, 0.3, 0.2, 0.1 and 0), which corresponds to using only the following proportions: 0.5, 0.6, 0.7, 0.8, 0.9 and 1 of the original lines (LinesProportion) for computing the sparse kernels since level of compression = 1 minus proportion of lines used to compute the sparse kernel. **(A)** corresponds to M4 and Arc_cosine_1 sparse kernel. **(B)** corresponds to M5 Arc_cosine_1 sparse kernel. **(C)** corresponds to M4 and Gaussian sparse kernel. **(D)** corresponds to M5 Gaussian sparse kernel. M4 denotes the random forest model and M5 denotes the Bayesian regression model.

It is observed in **Figure 8** that even with the large proportion of compression level, we did not experience a significant loss in prediction accuracy. When the compression level was larger, less time was required for the training process (**Figures 7C,D**). While the trend is not totally linear under model M4 and Gaussian sparse kernel, it is still clear that a significant reduction in time is achieved when the compression level increases. On the other hand, under model M5 with the Gaussian sparse kernel, a linear reduction is observed in the time required for training when the compression level is increased. This is particularly interesting since we can translate into significant savings of computational resources without a significant loss of prediction accuracy. Furthermore, **Figure 7D** also shows that model M5 requires considerably less time for the training process in comparison to the model M4 (**Figure 7C**).

The information provided in this **Figure 7**, illustrates that with the use of sparse kernels it is possible to gain a significant reduction in time for the implementation of the prediction models by means of dense kernels (without any level of compression). For example, **Figure 7** shows that the larger the level of compression the larger the reduction in computational resources. However, as observed in **Figures 6**, **8** caution must be exercised when determining the level of compression, because when this is large the level of accuracy could be negatively affected (will reduce the prediction performance). However, **Figures 6**, **8** depicted that even with level of compression of 50% genomic prediction accuracy is not dramatically affected. In general, M4 and M5 with sparse Gaussian kernel enhance the genome-based prediction accuracy of as compared with sparse kernel for all compression levels.

DEFAULT SETTINGS FOR THE ALGORITHMS

The default setting for those algorithms that require a tuning process (M1, M2, M3, M4 and M6) is the “Grid_search” strategy of tuning, but this only works when you specified

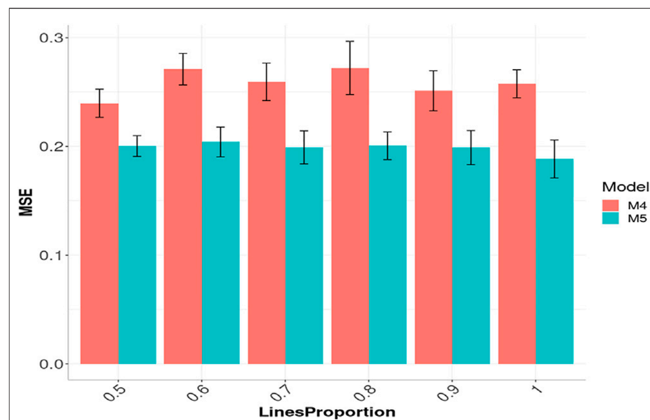


FIGURE 8 | Prediction performance across environments (Global) in Maize data in terms of Mean Square Error (MSE) of models M4 and M5 using the sparse Gaussian kernel with six proportions of compression levels: 0.5, 0.4, 0.3, 0.2, 0.1 and 0, which correspond to using only the following proportions: 0.5, 0.6, 0.7, 0.8, 0.9 and 1 of the original lines (LinesProportion) for computing the sparse kernels, since level of compression = 1 minus proportion of lines used to compute the sparse kernel. M4 denotes the random forest model and M5 denotes the Bayesian regression model.

at least for one of the hyper-parameters with more than two values to be evaluated, that is, a grid with a least two values for at least one hyperparameter. Also, for the tuning process by default is implemented an inner (nested) K-fold cross validation with K = 5 by default. When the Bayesian optimization is selected for the tuning process by default are explored 10 iterations. In **Table 4** are given the default hyperparameters for each or the six models.

DISCUSSION

Data is playing an unprecedented role in the twenty-first century. For this reason, many companies consider data science as a fundamental component to extract useful knowledge, make better decisions, reduce losses, analyze market trends and increase profits. Likewise, it is playing an essential role in increasing the rate of scientific and technological discoveries. For these reasons, the demand for Data Scientists continues increasing and is expected to grow by 27.9% by 2026, according to the US Bureau of Labor Statistics (Rieley, 2018). However, to satisfy this growing demand, people with different backgrounds need to be trained in this area rather rapidly. In this vein, more open source and user-friendly software such the SKM library are need to extract useful knowledge more efficiently from raw data. Even though there are many tools for implementing supervised machine learning methods in the R statistical software, they are still insufficient to cover the broad spectrum of needs, as there are many complex tasks that are not covered by existing tools.

For example, our library (SKM), in addition to grid search for hyperparameter tuning, also included the Bayesian optimization method, which is a sequential design strategy for global optimization of black-box functions that does not presume any functional forms. It is generally employed to optimize functions that are expensive to evaluate. Bayesian optimization, contrary to a grid search that performs an exhaustive evaluation over each point of the grid of values given for each hyperparameter, needs very few evaluations as starting points, and based on the knowledge at hand, it can indicate which point should be evaluated next. Bayesian optimization makes these decisions with something called acquisition functions, which are heuristics for how desirable it is to evaluate a point based on our present model. At every step, the Bayesian optimization method determines the best point to evaluate according to the acquisition function by optimizing it (Mockus, 2012). The model is then updated, and this process is repeated to determine the next point to evaluate.

In order for machine learning algorithms to be able to successfully perform a grid search, very large amount of values for each hyperparameter is required, and as such, this method is frequently rendered impractical since the required computational resources are substantial. For this reason, our library (SKM) is novel since it can be implemented for hyperparameter tuning with the Bayesian optimization algorithm, which is well suited when the function evaluations are expensive.

We do not expect the proposed SKM library to replace libraries like mlr3 and scikit-learn, since these libraries will continue to be suitable options for those who seek a complete solution for a particular machine learning implementation. Nonetheless, our library (SKM)

TABLE 4 | Default hyper-parameters for each of the models that can be implemented in the SKM library.

Model	Name	Default Hyper-parameter values
M1	Generalized boosted machines	trees_number = 500, max_depth = 1 node_size = 10, shrinkage = 0.1 sampled_records_proportion = 0.5
M2	Generalized linear models	alpha = 1; with alpha between 0 and 1 (for Elastic net Regression) and alpha = 0 for Ridge regression and alpha equal to 1 for Lasso Regression
M3	Support vector machines	kernel = "linear", degree = 3, gamma = 1/NCOL(x), coef0 = 0 and cost = 1
M4	Random forest	trees_number = 500, node_size = 5, node_depth = NULL and sampled_x_vars_number = NULL
M5	Bayesian regression models	Not applied since are not required hyperparameters since run with the default values of the BGLR library
M6	Deep neural networks	learning_rate = 0.001, epochs_number = 500, batch_size = 32, layers = list (list (neurons_number = 50, neurons_proportion = NULL, activation = "relu", dropout = 0, ridge_penalty = 0, lasso_penalty = 0)), output_penalties = list (ridge_penalty = 0, lasso_penalty = 0), optimizer = "adam", shuffle = TRUE, early_stop = FALSE early_stop_patience = 50

will be a great alternative for its simplicity, as it can be used with six conventional machine learning algorithms with some kernel methods, and thus, help to better capture non-linear patterns in the data.

Additionally, to the best of our knowledge, this is the first library that permits kernels to be implemented with six conventional machine learning methods in a very simple way, which can help increase the prediction performance when the input data contains non-linear patterns. Furthermore, the SKM package permits the implementation of approximate kernels (here called spare kernels), which can help reduce the computational resources for data sets of large dimensions, without a significant reduction in accuracy. In comparison to typical kernels that reduce the input size to the number of observations, sparse kernels can reduce the input size to even less than the number of observations and in this way, save more computational resources for its implementation. It must be noted that since the building process of the kernels is first done in an independent process, this computed kernel can be implemented with any machine leaning method.

While the proposed SKM library only allows multivariate responses for continuous outcomes to be trained under the Bayesian framework and generalized linear models, it also allows multivariate continuous, binary and categorical outcomes to be trained under by the random forest method. Nevertheless, only deep neural networks allows multivariate responses for continuous, binary, categorical and count to be trained. Contrarily, only univariate models can be trained under generalized boosted machines and support vector machines. As we previously stated, the six models can be implemented with seven kernels. These kernels are Linear, Polynomial, Sigmoid, Gaussian, Exponential, Arc-Cosine 1 and Arc-Cosine L (with L = 2, 3 ...), which is useful for when the dimensionality of the input is larger than the training samples, greater computational resources are needed; however, using any of these kernels reduces the number of training samples which, in turn, reduces the computational resources needed, thus permitting non-linear patterns to be captured more efficiently.

With the illustrative examples provided, the library can implement supervised machine learning methods for binary, categorical, count and continuous response variables, with the advantage that the user does not need to specify the type of response to be implemented; by providing the response variable as a factor, the library will understand whether it will implement a binary or categorical model depending on the number of categories of the response variable. On the other hand, if the response variable is converted to numeric values, the library will implement a count or continuous model.

CONCLUSION AND FUTURE WORK

This new package will benefit both machine learning practitioners and researchers who want to implement predictive models in a simple way with state-of-the art methods for tuning hyperparameters like Bayesian optimization. We also expect people from different disciplines who are not programming experts to be able to take advantage of the simplicity of SKM to enter into the machine learning world.

The kernelize function in SKM is of special interest since this is the first package that allows kernels to be used with different machine learning algorithms as a new approach of working with complex non-linear and high dimensional data.

This new package is not intended to provide a full data science solution, but rather, new machine learning algorithms can be included in future versions along with more metric functions, model benchmarking, data input and other data science related tools.

With the plant breeding examples provided, we illustrated how this library can implement six machine learning algorithms and seven types of kernel methods in the context of genomic prediction. Moreover, we illustrated that the implementation of sparse kernels can save significant computation resources without a significant loss in prediction accuracy. Finally, in the appendices, we provided all the codes so that users from different backgrounds and areas of interest can easily implement all the models and tools provided in the SKM library.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the first author and/or the corresponding authors and can be found in: https://github.com/osval78/SKM_Library_Examples.

AUTHOR CONTRIBUTIONS

OAML, AML, and JC had the original idea and BM and AP assisted in writing the R codes. All the authors participated writing the first version and reviewing several of improved versions.

FUNDING

We are thankful for the financial support provided by the Bill & Melinda Gates Foundation [INV-003439, BMGF/FCDO, Accelerating Genetic Gains in Maize and Wheat for Improved Livelihoods (AG2MW)], the USAID projects USAID Amend. No. 9 MTO 069033, USAID-CIMMYT Wheat/AGGMW, AGG-Maize Supplementary Project, AGG (Stress Tolerant Maize for Africa), and the CIMMYT CRP (maize and wheat).

ACKNOWLEDGMENTS

We acknowledge the financial support provided by the Foundation for Research Levy on Agricultural Products (FFL) and the Agricultural Agreement Research Fund (JA) in Norway through NFR grant 267806.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.887643/full#supplementary-material

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Available: https://www.tensorflow.org/.

Allaire, J., and Chollet, F. (2016). Keras: R Interface to Keras. https://keras.rstudio.com (Accessed 07 11, 2018).

Crossa, J., Campos, G. d. l., Pe´rez, P., Gianola, D., Burgueño, J., Araus, J. L., et al. (2010). Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. *Genetics* 186 (2), 713–724. doi:10.1534/genetics.110.118521

Cuevas, J., Crossa, J., Montesinos-López, O. A., Burgueño, J., Pérez-Rodríguez, P., and de Los Campos, G. (2017). Bayesian Genomic Prediction with Genotype × Environment Interaction Kernel Models Environment Interaction Kernel Models. *G3 Genes, Genomes, Genet.* 7 (1), 41–53. doi:10.1534/g3.116.035584

Cuevas, J., Crossa, J., Soberanis, V., Pérez-Elizalde, S., Pérez-Rodríguez, P., Campos, G. d. l., et al. (2016). Genomic Prediction of Genotype × Environment Interaction Kernel Regression Models. *Plant Genome* 9 (3), 1–20. doi:10.3835/plantgenome2016.03.0024

Cuevas, J., Montesinos-López, O. A., Martini, J. W. R., Pérez-Rodríguez, P., Lillemo, M., and Crossa, J. (2020). Approximate Genome-Based Kernel Models for Large Data Sets Including Main Effects and Interactions. *Front. Genet.* 11, 567757. doi:10.3389/fgene.2020.567757

Cuevas, J., Montesinos-López, O., Juliana, P., Guzmán, C., Pérez-Rodríguez, P., González-Bucio, J., et al. (2019). Deep Kernel for Genomic and Near Infrared Predictions in Multi-Environment Breeding Trials. *G3 Genes|Genomes|Genetics* 9 (9), 2913–2924. doi:10.1534/g3.119.400493

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* 33 (1), 1–22. doi:10.18637/jss.v033.i01

Greenwell, B., Boehmke, B., Cunningham, J., and Developers, G. (2020). Gbm: Generalized Boosted Regression Models. Available: https://CRAN.R-project.org/package=gbm.

Ishwaran, I., Kogalur, U. B., Blackstone, E. H., and Lauer, M. S. (2008). Random Survival Forests. *Ann. Appl. Stat.* 2 (3), 841–860. doi:10.1214/08-aoas169

Kuhn, M. (2016). Caret: Classification and Regression Training. R package version 6.0-71. Available: https://github.com/topepo/caret/.

Kuhn, M., and Wickham, H. (2020). Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles. Available: https://www.tidymodels.org.

Lang, M., Binder, M., Richter, J., Schratz, P., Pfisterer, F., Coors, S., et al. (2019). mlr3: A Modern Object-Oriented Machine Learning Framework in R. *Joss* 4 (44), 1903. doi:10.21105/joss.01903

Lang, M. (2017). Checkmate: Fast Argument Checks for Defensive R Programming. *R J.* 9 (1), 437–445. doi:10.32614/RJ-2017-028

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2019). Misc Functions of the Department of Statistics. Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7- 2 https://CRAN.R-project.org/package=e1071.

Mockus, J. (2012). *Bayesian Approach to Global Optimization: Theory and Applications*. Berlin, Germany: Springer Science & Business Media.

Montesinos-López, O. A., Montesinos-López, A., Hernandez-Suarez, C. M., Barrón-López, J. A., and Crossa, J. (2021a). Deep-learning Power and Perspectives for Genomic Selection. *Plant Genome* 14 (3), e20122. doi:10.1002/tpg2.20122

Montesinos-López, A., Montesinos-López, O. A., Montesinos-López, J. C., Flores-Cortes, C. A., de la Rosa, R., and Crossa, J. (2021b). A Guide for Kernel Generalized Regression Methods for Genomic-Enabled Prediction. *Heredity* 126 (4), 577–596. doi:10.1038/s41437-021-00412-1

Montesinos-López, O. A., Montesinos-López, A., and Crossa, J. (2022b). "Overfitting, Model Tuning and Evaluation of Prediction Performance," in *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Editors O.A. Montesinos López, A. Montesinos López, and J. Crossa (Cham, Switzerland: Springer International Publishing), 2022, 109–139.

Montesinos-López, O. A., Montesinos-López, A., and Crossa, J. (2022a). "Reproducing Kernel Hilbert Spaces Regression and Classification Methods," in *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Editors O.A. Montesinos López, A. Montesinos López, and J. Crossa (Cham, Switzerland: Springer International Publishing), 2022, 251–336.

Osborne, M. A., Garnett, R., and Roberts, S. J. (2009). *Gaussian Processes for Global Optimization*. Oxford, UK: Learning and Intelligent Optimization, 1–15.

Ott, T., Palm, C., Vogt, R., and Oberprieler, C. (2020). GinJinn: An Object-detection Pipeline for Automated Feature Extraction from Herbarium Specimens. *Appl. Plant Sci.* 8 (6), e11351. doi:10.1002/aps3.11351

Pandas development team (2020). *Pandas-dev/pandas*. Pandas. Zenodo. doi:10.5281/zenodo.3509134

Pedregosa, F., Gael Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Olivier Grisel, O., et al. (2011). Scikit-learn: Machine Learning in python. *Mach. Learn. PYTHON* 12, 2825–2830.

Pérez, P., and de los Campos, G. (2014). Genome-wide Regression and Prediction with the Bglr Statistical Package. *Genetics* 198 (2), 483–495. doi:10.1534/genetics.114.164442

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available: https://www.R-project.org/.

Rieley, M. (2018). Big Data Adds up to Opportunities in Math Careers. *Beyond Numbers Employ. Unempl.* 7 (8). https://www.bls.gov/opub/btn/volume-7/big-data-adds-up.htm.

Shahin, I., Nassif, A. B., and Hamsa, S. (2020). Novel Cascaded Gaussian Mixture Model-Deep Neural Network Classifier for Speaker Identification in Emotional Talking Environments. *Neural Comput. Applic* 32 (7), 2575–2587. doi:10.1007/s00521-018-3760-2

Souza, M., Cuevas, J., de Oliveira, C. E. G., Pérez-Rodríguez, P., Jarquín, D., Fritsche-Neto, R., et al. (2017). Genomic-Enabled Prediction in Maize Using Kernel Models with Genotype × Environment Interaction. *G3 (Bethesda)* 7 (6), 1995–2014. doi:10.1534/g3.117.042341

Wang, R., Aung, M. S. H., Abdullah, S., Brian, R., Campbell, A. T., Choudhury, T., et al. (2016). "CrossCheck," in UbiComp 2016 - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, September 12-16, 2016 (Association for Computing Machinery, Inc), 886–897. doi:10.1145/2971648.2971740

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the Tidyverse. *Joss* 4 (43), 1686. doi:10.21105/joss.01686

Wickham, H., François, R., Henry, L., and Müller, K. (2015). Dplyr: A Grammar of Data Manipulation. R Package Version 0.4.3 http://CRAN.R-project.org/package=dplyr.

Williams, C. K. I., and Seeger, M. (2001). "Using the Nyström Method to Speed up Kernel Machines," in *Advances in Neural Information Processing Systems*. Editors T.K. Leen, T.G. Diettrich, and V. Tresp (Cambridge, MA: MIT Press), 13, 682–688.

Yan, Y. (2016). RBayesianOptimization: Bayesian Optimization of Hyperparameters. R package version 1.1.0. [Online]. Available: https://CRAN.R-project.org/package=rBayesianOptimization.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Montesinos López, Mosqueda González, Palafox González, Montesinos López and Crossa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

REVIEW
published: 09 June 2022
doi: 10.3389/fgene.2022.882836

A Prospective Review on Selectable Marker-Free Genome Engineered Rice: Past, Present and Future Scientific Realm

Rajveer Singh[1†], Navneet Kaur[1†], Umesh Preethi Praba[1], Gurwinder Kaur[1], Mohammad Jafar Tanin[2], Pankaj Kumar[1]*, Kumari Neelam[1], Jagdeep Singh Sandhu[1] and Yogesh Vikal[1]*

[1]School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, India, [2]Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India

OPEN ACCESS

Edited by:
Sukhwinder Singh,
The International Maize and Wheat Improvement Center (CIMMYT), India

Reviewed by:
Jaindra Nath Tripathi,
International Institute of Tropical Agriculture (IITA), Kenya
Sajid Fiaz,
The University of Haripur, Pakistan

***Correspondence:**
Pankaj Kumar
ranapankajbhu@gmail.com
Yogesh Vikal
yvikal-soab@pau.edu

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 24 February 2022
Accepted: 29 April 2022
Published: 09 June 2022

Citation:
Singh R, Kaur N, Praba UP, Kaur G, Tanin MJ, Kumar P, Neelam K, Sandhu JS and Vikal Y (2022) A Prospective Review on Selectable Marker-Free Genome Engineered Rice: Past, Present and Future Scientific Realm.
Front. Genet. 13:882836.
doi: 10.3389/fgene.2022.882836

As a staple food crop, rice has gained mainstream attention in genome engineering for its genetic improvement. Genome engineering technologies such as transgenic and genome editing have enabled the significant improvement of target traits in relation to various biotic and abiotic aspects as well as nutrition, for which genetic diversity is lacking. In comparison to conventional breeding, genome engineering techniques are more precise and less time-consuming. However, one of the major issues with biotech rice commercialization is the utilization of selectable marker genes (SMGs) in the vector construct, which when incorporated into the genome are considered to pose risks to human health, the environment, and biodiversity, and thus become a matter of regulation. Various conventional strategies (co-transformation, transposon, recombinase systems, and MAT-vector) have been used in rice to avoid or remove the SMG from the developed events. However, the major limitations of these methods are; time-consuming, leftover cryptic sequences in the genome, and there is variable frequency. In contrast to these methods, CRISPR/Cas9-based marker excision, marker-free targeted gene insertion, programmed self-elimination, and RNP-based delivery enable us to generate marker-free engineered rice plants precisely and in less time. Although the CRISPR/Cas9-based SMG-free approaches are in their early stages, further research and their utilization in rice could help to break the regulatory barrier in its commercialization. In the current review, we have discussed the limitations of traditional methods followed by advanced techniques. We have also proposed a hypothesis, "DNA-free marker-less transformation" to overcome the regulatory barriers posed by SMGs.

Keywords: clustered regularly interspaced short palindromic repeats/crispr associated Cas9 (Crispr/Cas9), genetic engineering, genetically modified (GM) -regulation, rice, selectable marker genes (SMGs)

1 INTRODUCTION

The green revolution has led to remarkable progress through high-yielding crop varieties worldwide. Food security is the key mandate of agriculture systems to feed the ever-exceeding global human population (expected to be 10 billion by 2050). Rice (*Oryza sativa*) is one of the major staple food crops worldwide. Asian countries constitute approximately 91% of rice,

preceded by South America, North and Central America, Europe, and Oceania (Fraiture et al., 2016). However, its production has faced constant challenges due to the biotic and abiotic stresses that have emerged through climate change (Stallworth et al., 2020; Hernandez-Soto et al., 2021). Rice genetic improvement has been made through conventional breeding, molecular approaches, and genetic and genome engineering techniques to enhance yield potential and resistance to biotic and abiotic stresses (Das and Rao, 2015; Singh et al., 2020). Although molecular breeding is a leading method of crop improvement, including biotic and abiotic stresses (Waseem et al., 2021; Islam et al., 2022), during the continuous domestication and selection, significant genetic diversity has been lost (Singh et al., 2016). Moreover, breeding programs require ample time to transfer certain traits from wild relatives into elite cultivars, generally employing foreground and repeated background selections.

An alternative to these breeding strategies, genome engineering approaches represent a new way to tailor crop architecture in a comparably short time interval. At the beginning of the last decade (in the year 2013) the emergence of a new genome-editing tool, "Cluster Interspaced Short Palindromic Repeat" and its associated Cas9 nuclease (CRISPR/Cas9) has also enabled us to design the genetic architecture of rice for various traits including biotic stresses, abiotic stresses and other qualitative traits (Fiaz et al., 2021). For instance, transgenic rice expressing Dehydration-Responsive Element-Binding (DREB) genes for drought and salt tolerance (Lata and Prasad, 2011), *Cry* gene for insect resistance (Estiati, 2020) have been developed. Lectin genes such as Allium Sativum leaf lectin (ASAL) for sap-sucking insects (Yarasi et al., 2008) and *Cry1Ac::ASAL* hybrid fusion protein for multi-insect resistance (Boddupally et al., 2018) have been incorporated into different rice cultivars. Moreover, transgenes have been targeted for bacterial blight, blast, and sheath blight resistance (Sawada et al., 2004; Molla et al., 2020), nutritional traits like Golden rice enriched with beta-carotene (Paine et al., 2005), and many others, which have significantly improved its yield and quality.

Despite the great potential of genome engineering technologies, the journey of genetically engineered crops from labs to fields and finally to commercial release has been scrutinized substantially and blocked due to the socio-ethical concerns associated with their release process. Fraiture et al. (2016) have reported that the status of biotech rice is restricted to laboratory experiments or field evaluation. Garg et al. (2018) also inferred the maximum research in transgenics but minimum utilization at the commercial level. Apart from regulatory concerns of transgene expression (transgenic research) and off-target effects (genome editing research) in engineered rice, the main issue is the use of selectable marker genes (SMGs) placed next to the genetic construct in the transfer-DNA (T-DNA) region of the plasmid. *Neomycin phosphotransferase II* (*npt II*) and *hygromycin phosphotransferase* (*hpt*) are routinely used antibiotic resistance marker genes (ARMGs) (Hiei et al., 1997; Twyman et al., 2002; Breyer et al., 2014). The ARMGs present in transgenics is of no use but is of regulatory concern for the release and commercialization of transgenic crops (Breyer et al., 2014). The harness of ARMGs in transgenic plants has been questioned over the past few years as horizontal gene transfer from plant to soil bacteria or human intestinal microbes by plant products consumed as food. However, all these apprehensions are merely suppositional issues lacking scientific shreds of evidence (Ramessar et al., 2007; Breyer et al., 2014). The use of ARMGs in Genetically Modified (GM) plants is opposed strictly by many national governments, Non-Governmental Organizations (NGOs), industries, and regulators. The European Union (EU) raises concerns about the use of ARMGs and strictly opposes them in Genetically Modified Organisms (GMOs), as they may adversely affect human health and cause environmental risks (European Parliament Council of the European Union, 2001).

Alternative to selective antibiotics, second-generation non-antibiotic SMGs have also been employed in rice genetic transformation e.g., herbicide resistance gene for bialaphos (*bar*) (Rathore et al., 1993; Zhao et al., 2007). However, the use of herbicide resistant genes has several limitations related to the environment (Breyer et al., 2014). Additionally, *hpt* in Golden rice 1 (GR1) was opposed strictly due public perception of it, so new Golden rice 2 (GR2) events were developed by Syngenta. Instead of having an antibiotic marker, the *phosphomannose isomerase* (*pmi*) gene was used (Paine et al., 2005). More recently, *phosphite oxidoreductase* (*ptxd*) has been utilized as a selection marker in rice (Dormatey et al., 2021; Liu et al., 2021). A battery of scorable marker (positive selection) genes such as *gus* (*ß-glucuronidase*), *gfp* (*green fluorescent protein*), *luc* (firefly *luciferase*) and *manA* (*mannose A*) have been employed for screening transgenic rice to overcome the limitations posed by the use of antibiotics and herbicide resistant genes (Sah et al., 2014). A series of systems have been developed to avoid the use of SMGs and their removal from transgenic plants. The SMGs-free system includes co-transformation, site-specific recombinase, transposon-based, MAT (Multi Auto-Transformation) vector, DRB (Double Right Border)-binary vector, and marker-free transformation, which have been discussed in great detail in many reviews (Chong-Pérez and Angenon, 2013; Yau and Stewart, 2013; Breyer et al., 2014). The scope of the current review is not only to account in brief for these systems but also to discuss recently developed marker-free systems and their utility in developing rice free from selectable markers. Thus, it is imperative to study its current regulatory status to understand future visions for the commercialization of marker-free biotech rice.

2 ACCOUNT ON SELECTABLE MARKER GENES-FREE ENGINEERED RICE

Plant genetic engineering would not have become possible without selectable markers. The selectable markers allow the transformed cells to grow favorably where otherwise they face competition and being overgrown by non-transformed cells. The percent use of specific selectable markers in rice is

represented in **Figure 1A**. The study showed that the most widely used SMG is *hpt* (74.6%), followed by *npt II* (12.6%), *Bar* (4.7%), fluorescence, and *isopentyl transferase* (*ipt*) (3.1%), and *pmi* (1.5%) genes. The decline in the use of the *Bar* gene as the selectable marker is due to its positional effect and pleiotropic effect on the expression of plant genes (Miki et al., 2009). It is also worth accounting for the technique used in rice as a percentage, based on several publications (1996–2021) (**Figure 1B**). The co-transformation technique almost accounts for 62.2% of rice transformation, followed by site-specific recombination methods (20.5%), transposon (7.4%), and CRISPR (Clustered regularly interspaced short palindromic repeats)-based methods (7.5%). It is interesting to study the trend of various SMG-free technologies used so far from their beginning in rice. A timeline of diverse SMG-free techniques in rice has been retrieved from literature (1996–2021) and illustrated based on their year-wise use (**Figure 1C**). The most premier and prevalent technique used in rice is co-transformation was first reported in 1996 (Komari et al., 1996), with the most recent publication in 2018 (Rajadurai et al., 2018). It is anticipated that more publications on this subject will follow in the future. Besides co-transformation, site-specific recombination techniques including Reversible Recombinase system (R/Rs), Cyclic recombinase enzyme (Cre/lox), and Flippase/Flippase recognition target (FLP/FRT) are other methods of excising-out SMG using homologous recombination. These have been widely adopted in rice between 2001 and 2017, starting with R/Rs (2001–2002), but later on, the commonly used recombinase system was largely Cre/lox (2005–2017). However, only a single report on the FLP/FRT system use is available in rice (Woo et al., 2015). Another method of auto-excision used in rice is the transposon-based removal of SMG between 2002 and 2021. The majority of approaches used transposon system Ac/Ds (reported in five publications to date). "Piggyback" transposon from the cabbage lopper moth (discussed in the next section) was used in one study (Nishizawa-Yokoi et al., 2015).

With the dawn of CRISPR as a genome editing tool, its flexibility and versatility have allowed us to use it as a tool for removing SMG from engineered plants. Recently, there have been reports of the use of CRISPR/Cas9 to remove selectable markers using homology-directed repair (HDR) based marker excision (Dong et al., 2020), marker-free targeted gene insertion (Tan et al., 2022), and transgene-free Ribonucleoprotein (RNP) based genome editing in rice (Banakar et al., 2020). A comprehensive list of techniques used to produce SMG-free rice is shown in **Table 1**. The numerical data of these SMG-free techniques during their current and historical use in rice might assist with correlating their efficiency, ease, and even their regulatory aspects.

2.1 Traditional Methods to Make Selectable Marker Genes-Free Rice

The foremost concern of SMGs in engineered crops is socio-ethical issues and transgene expression. Even several copies of SMGs may result in the silencing of the essential genes of plants and affect plant metabolism (Rosellini, 2012). The batteries of

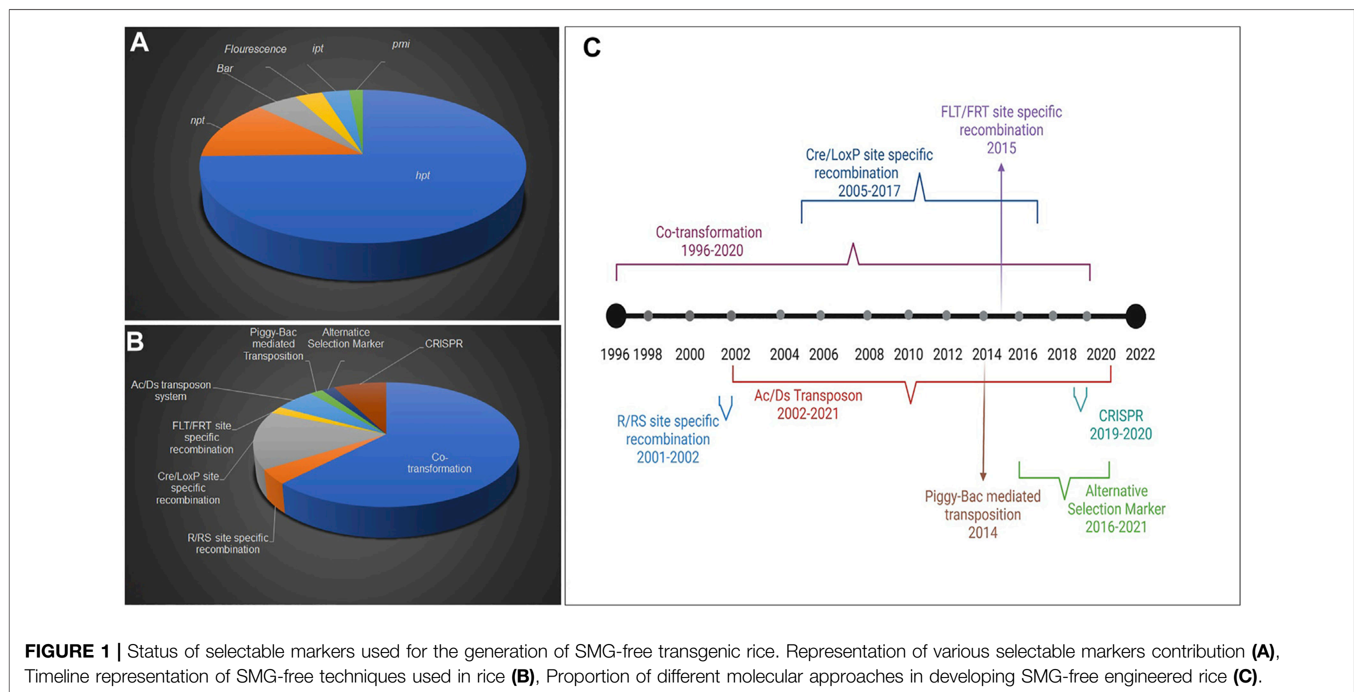


FIGURE 1 | Status of selectable markers used for the generation of SMG-free transgenic rice. Representation of various selectable markers contribution **(A)**, Timeline representation of SMG-free techniques used in rice **(B)**, Proportion of different molecular approaches in developing SMG-free engineered rice **(C)**.

TABLE 1 | Summary of selectable markers and techniques used to create SMG-free engineered rice.

S. no.	Method(s) used to generate SMG-free Plants	Selectable marker gene	Target gene(s)	Target Trait	References
1	Co-transformation	*HPT and NPT-II*	*GUS*	GUS activity in plant leaves	Komari et al. (1996)
2	Co-transformation	*HPT*	*uidA*	Gus activity in rice transgenic cells	Huang et al. (2001)
3	Co-transformation	*HPH, Bar*	*Rice ragged stunt virus (RRSV)*	Viral resistance	Lu et al. (2001)
4	R/RS site-specific recombination + Ac transposable elements	*HPT and NPT-II*	*R gene of zygosaccharomyces rouxii*	Generation of deletion in rice genome	Nakagawa et al.(2001)
5	Ac/Ds transposon system	*HPH*	*cry1B*	Insect resistance	Cotsaftis et al.(2002)
6	R/RS site-specific recombination	*IPT*	*Gus A, NPTII,* and *hpt*	Model genes of interest	Endo et al. (2002)
7	Co-transformation	*HPT*	*glutelin A (Antisense)*	Glutelin content in seeds	Maruta et al. (2002)
8	Co-transformation	*HPT and PMI*	*Phytoene synthase (psy), lycopene β-cyclase (lcy),* and *phytoene desaturase (crtl)*	Caroteneoid accumulation	Datta el al. (2003)
9	Co-transformation	*HPH*	*crylAb/crylAc*	Insect resistance (yellow stem borers and leaf-folders)	Tu et al. (2003)
10	Co-transformation	*HPT*	*bar*	Herbicide resistance	Breitler et al. (2004)
11	Co-transformation	*NPT-II and HPH*	*crtI , psy, and lyc*	Caroteneoid accumulation	Parkhi et al. (2005)
12	Cre/loxP site-specific recombination	*IPT*	*79 bp of XVE*	—	Sreekala et al. (2005)
13	Co-transformation	*HPH*	*psy, crtI, and lyc*	Accumulation of provitamin A in the endosperm tissue	Baisakh et al. (2006)
14	Cre/loxP site-specific recombination	*HPT*	*Vitreoscilla hemoglobin (VHb), trans-zeatin synthetase (tzs), and modified 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS)*	—	Cao et al. (2006)
15	Co-stranformation	*HPH*	*chip (pistil chitinase)*	Pistil-predominant chitinase (blast-disease resistance)	Hashizume et al. (2006)
16	Co-transformation	*HPT*	*Amphipathic protein (API)*	Enhanced disease resistance	Yu et al. (2006)
17	Co-transformation	*HPT*	*Xa21*	Bacterial blight (BB) resistance	Xia et al. (2006), Gao et al. (2011)
18	Co-transformation	*HPT*	*Human lactoferrin (hLF), a lysine-rich protein gene from potato (SB401), and a methionine-rich protein gene from rice (RZ10)*	—	Li et al. (2007)
19	Co-transformation	*Bar*	*CecropinB*	Resistance against a range of plant pathogenic bacteria (*Xanthomonas compestris pv oryzae*)	Zhao et al. (2007)
20	Cre/loxP site-specific recombination	*NPT-II*	*Gus controlled by OsMAD45*	Gus assay (Expression pattern of *OsMAD45* promotor)	Bai et al. (2008)
21	Co-transformation	*HPH*	*gluA-4XCII250–270*	Accumulating a type II-collagen tolerogenic peptide	Hashizume et al. (2008)
22	Co-transformation	*HPH*	*Rice chitinase (chi11)*	sheath blight resistance	Sripriya et al. (2008)
23	Co-transformation	*HPT*	*Cry1Ab*	Lepidopteran Pest Resistance	Qi et al. (2009)
24	Co-transformation	*HPT*	*crylA(c)*	resistance to chewing insects	Yu H. X. et al. (2009)
25	Co-transformation	*HPT*	*Waxy (Wx)*	high amylose content (AC)	Yu H. et al. (2009)
26	Co-transformation	*HPT*	*cry1B-1Aa*	Insect resistance (yellow stem borer)	Kumar et al. (2010)
27	Cre/lox site-specific recombination	*HPT*	*ASAL*	Resistance to sap-sucking planthoppers	Sengupta et al. (2010)
28	Co-transformation	*HPH*	*chi11*	Sheath blight disease resistance	Ramana Rao et al. (2011)
29	Cre/loxP site-specific recombination	*NPT*	*Gus A*	GUS assay	Khattri et al. (2011)
30	Cre/loxP site-specific recombination	*NPT and HPT*	*GUS driven by maize ubiquitin promoter*	GUS activity	Nandy and Srivastava, (2012)
31	Ac/Ds transposon system	*HPT*	*partial sequences of the first intron of rice epsps*	—	Li and Charng, (2012)
32	Co-transformation	*HPT*	*inverted-repeat (IR) structures targeting the rice stripe virus (RSV) coat protein (CP) and the special-disease protein (SP)*	Resistance to rice stripe virus (RSV)	Jiang et al. (2013)
33	Co-transformation	*HPT-II*	*High Molecular Weight Glutenin Subunits (HMW-GS) Gene- Glu-1Bx*	Increasing bread-making quality	Park et al. (2013)
34	Co-transformation	*HPT*	*cry1Ab*	Insect resistance (silkworm)	Qi et al. (2013)
35	Co-transformation	*HPT*	Phytoferritin	Increase iron content	Oliva et al. (2014)
36	Piggy bac mediated transposition	*HPT*	*ALS*	Herbicide bispyribac sodium (BS)-tolerant	Nishizawa-Yokoi et al. (2015)

(Continued on following page)

TABLE 1 | (*Continued*) Summary of selectable markers and techniques used to create SMG-free engineered rice.

S. no.	Method(s) used to generate SMG-free Plants	Selectable marker gene	Target gene(s)	Target Trait	References
37	Co- transformation	*HPT-II*	*Glu-1Dy10*	Increasing quality processing of bread and noodles	Park et al. (2014)
38	Co- transformation	*HPT*	*Bt*	Insect resistance	Gao et al. (2015)
39	FLP/FRT site-specific recombination	*HPT*	*NtTC*	Enhanced seed tocopherol content	Woo et al. (2015)
40	Alternative selection marker	*HPT*	*ptxD*	Weed control in rice	Manna et al. (2016)
41	Cre/loxPsite-specific recombination	*HPT-I*	*vip3BR*	Broad-spectrum insect resistance	Pradhan et al. (2016)
42	Co-transformation	*HPT*	RNAi targeting RBSDV (rice black-streaked dwarf virus)	Developing resistance	Ahmed et al. (2017)
43	Co-transformation	*HPT*	*NmDef02 antifungal defensin.*	Resistance against phytopathogenic fungus *Sarocladium oryzae*	Perez-Bernal et al. (2017)
44	Co-transformation	*HPT*	*AmA1*	Production of essential amino acids in rice seeds	Xu et al. (2017)
45	Cre/loxPsite-specific recombination	*HPT, NPT-II, BAR*	*OsB1, OsB2, OsDFR, OsC1*	Purple endosperm	Zhu et al. (2017)
46	Co-transformation	*HPT*	*cry2AX1*	Insect resistance	Rajadurai et al. (2018)
47	CRISPR	*DsRED fluorescence*	*IAA methyltransferase (IAMT)*	The difficulty for hypocotyl reorientation under gravistimulation increased growth rate of pollen tube	Aliaga-Franco et al. (2019)
48	CRISPR-Cas9 RNP	*Hygromycin*	*DROOPING LEAF (DL)*	Drooping leaf phenotype	Toda et al. (2019)
49	Co-transformation	*HPT*	*SSSII-2*	Soft kernels	Xu et al. (2020)
50	CRISPR-Cas9 RNP (co-delivered with plasmid)	*HPT*	*PDS*	Albino phenotype	Banakar et al. (2020)
51	CRISPR-Cas9	*Hygromycin*	*SSU-crtI and ZmPsy*	Enrichment of carotenoids in seeds	Dong et al. (2020)
52	Co-transformation	*HPT*	RNAi targeting RBSDV (rice black-streaked dwarf virus)	Developing resistance	Feng et al. (2021)
53	Ac/Ds transposon system	*Green and Red Fluorescence*	*Pi21*	Rice blast disease	Li et al. (2021)
54	Alternative selection marker	*HPT,NPT II*	*ptxDq*	Catalytic activity	Liu et al. (2021)

methods have been developed to make marker-free transgenic crops, including rice, as discussed below.

2.1.1 Co-Transformation

The maximum utilization of co-transformation is due to its simplicity and safety compared to other traditional methods. This method uses two T-DNAs containing the gene of interest (GOI) and the SMG, respectively. The chance of independent integration of GOI and SMG at different loci in the plant genome allows us to eliminate SMG by simple selection in subsequent generations (Breyer et al., 2014). The integration of SMG and GOI independently could be achieved in three ways: 1) using two strains of *Agrobacterium*, each with T-DNA, one with SMG, and the other with GOI. 2) Using a single *Agrobacterium* harboring two plasmids having independent SMG and GOI. 3) Using a single plasmid carrying two independent T-DNA regions in a single *Agrobacterium*. Co-transformation has been employed successfully in many monocots and dicots (Breyer et al., 2014). The best example is GR1, where the hygromycin resistance marker gene was eliminated (Al-Babili and Beyer, 2005). Later on, marker-free *Bt* transgenic rice was generated (Woo et al., 2015).

The efficiency of co-transformation utilizing a single vector containing two T-DNAs has been linked with a high frequency of (linked co-delivery of) the target gene and marker gene and interference with non-T-DNA sequences (McCormac et al., 2001). The co-transformation method is more efficient compared to other approaches and still it is under utilization in rice to date (Xu et al., 2017; Rajadurai et al., 2018). Another modification of the co-transformation vector system is the use of a DRB binary vector system. A DRB binary vector contains two copies of T-DNA right-border (RB) sequences adjoining a selectable marker followed by a GOI and behind with a copy of the left border (LB) sequence. Two different kinds of T-DNA could be inserted, the first RB contains the SMG and the GOI together, and the second RB contains only the GOI. Consequently, these could segregate away from each other, with the progeny resulting in GOI. Lu et al. (2001) followed this method and obtained positive progeny plants with only GOI for rice ragged stunt virus (RRSV)-derived synthetic resistance gene. Similarly, Xia et al. (2006) utilized the DRB-vector technique to make marker-free and vector backbone-free transgenic rice expressing *Xa21* gene for bacterial blight disease.

2.1.2 Site-Specific Recombination

Recombinase systems have also been used widely in various crops. Recombination is a well-known concept in biological systems. It occurs when two homologous sites in DNA molecules that contain a recombinase protein come together (Hirano et al., 2011). Site-specific fusion techniques in plants have been implemented to make marker-free foreign genes (Nanto and Ebinuma, 2008). The various recombinase systems (Cre-lox, FLP-FRT, and R/RS) classified under site-specific recombination are well described (Yau and Stewart, 2013). The Cre/lox system has been used to remove *hpt* and *NPT*-II in transgenic rice for the purple endosperm trait (Zhu et al., 2017). The chief limitations of recombinase systems include: 1) it is difficult to achieve 100% excision efficiency; 2) the prolonged presence of recombinase systems in the plant genome could lead to genetic and phenotypic changes making it less appealing than co-transformation; and 3) it has also been reported that chromosomal rearrangements use cryptic-target sites, and there are reports of leftover dispensable sequences of recombinase systems (Breyer et al., 2014; Nishizawa-Yokoi et al., 2015).

2.1.3 MAT-Vector System

MAT vectors use oncogenes (*ipt, iaaM/H, rol*) of *Agrobacterium* as selection markers, which control the endogenous levels of plant hormones and help to regenerate transgenic cells over non-transgenic cells (Ebinuma and Komamine, 2001). In this case, the oncogenes are combined with the site-specific recombination system (*R*/RS) for transformation. Later on, the oncogenes are removed by the R/RS system to generate marker-free transgenic plants (Ebinuma et al., 2005). This system has been used to eliminate the *ipt* marker gene from the transgenic rice (Endo et al., 2002).

2.1.4 Transposon-Based

Transposon-mediated transgene reintegration was used initially by Goldsbrough et al. (1993) to reposition a Dissociation (Ds) transposon-based GUS reporter gene in transgenic tomato (*Solanum lycopersicum*). The most characterized transposons belong to the *Ac/Ds* family. In this method, either GOI or SMG (present in T-DNA) is inserted between the Ds elements. Subsequently, an active transposase recognizes the Ds elements and cleaves either of them from their native position and reinserts them into another chromosomal location after the initial transformation. Later on, the SMG could be sorted out by subsequent selection (Yau and Stewart, 2013). In a few studies, this technique has been used in rice, and recently it has been used to remove selection markers in transgenic rice resistant to blast disease (Li et al., 2021). The major limitation of this technique is that it is labor-intensive to segregate out SMG from GOI, variable transposons efficiency, and they also cause mutations at an unknown site. Apart from the Ac/Ds system, another transposon named "piggyback" was used in excising the *hpt* gene from rice plants mutated for acetolactate synthase gene (ALS) using homologous recombination (HR)-mediated gene targeting (GT) (Nishizawa-Yokoi et al., 2015).

2.1.5 Marker-Less Transformation

Marker-free transformation refers to transforming without SMGs. It is an ideal way to obtain marker-free GM plants. Although the frequency of recovering transgenic events is lower (2 or 3-fold) than the use of SMGs, it could vary between 1%–25% (Breyer et al., 2014). The marker-free transformation has also been achieved *via* the pollen-tube pathway, in which exogenous DNA is taken up by egg cells or zygotes after fertilization. The pollen-tube channel has been used in certain crops like cotton, wheat, maize, and rice in China (Yang et al., 2009).

2.2 Recent Methods Adopted to Make Selectable Marker Genes-Free Rice-CRISPR Era

Recently, the most widely used genome editing tool known as CRISPR/Cas9 has also been brought into use to remove or avoid SMG in transgenic rice. Using the CRISPR/Cas9 tool, site-specific DSB is induced at the target site, followed by a repair mechanism either through homology-directed repair (HDR) or non-homologous end joining (NHEJ). Among both, the natural occurrence of HDR is rare and thus requires a donor template to repair DSB (Zafar et al., 2020). The delivery of donor templates is quite challenging due to the difficulties of its delivery and short-time stability in the cell. Therefore, recent efforts have aimed to increase HDR efficiencies, such as geminivirus-based donor template delivery (Wang et al., 2017) and Cas9-VirD2 chimeric protein (Ali et al., 2020). HDR-based SMG excision and marker-free gene insertion have been achieved (discussed next). It is imperative to mention that CRISPR is a more precise, efficient, and less time-consuming technology. Traditional methods, like co-transformation (using two independent T-DNA plasmids), transposon and recombinase systems (which leave cryptic sequences in the host genome) need a large screening population to segregate SMG. In contrast, the CRISPR/Cas9 based SMG-free approach utilizing HDR does not leave any foreign sequences in the genome. Moreover, RNP-based genome editing is considered DNA-free, and thus does not incorporate plasmid DNA sequences in the genome. It has now become possible to get rid of selectable markers as well as transgene cassettes that persisted in the plant genomes. The utilization of CRISPR/Cas9 as an SMG-free tool has been reported in the last few years and is in infancy. However, much is expected from this technology in terms of making SMG-free rice in the future. To date, only a few studies have reported the successful use of CRISPR/Cas9 as an SMG-free technique in rice, as discussed below.

2.2.1 Marker Excision

In addition to Cre/lox and Ac/Ds as auto-excision systems, CRISPR/Cas9-based HDR has been introduced as a marker excision system. Tan et al. (2022) used Pssi-driving CRISPR/Cas9-mediated HDR-based marker-free strategy (PssiCHMF) in rice. The "pssi" is a rice promoter that drives the high expression of the CRISPR/Cas9-HDR gene construct in shoot tip (containing

meristem) and inflorescence to enhance homology-directed marker excision in these tissues. The Cas9 induced double-strand break (DSB) repair pathway allows the deletion of large DNA fragments. The GUS marker gene was targeted for excision using the pYLPssi::Cas9 construct with a pair of 1027-bp homology arms to improve HDR efficiency. It resulted in a 55.6% homozygous excision of marker genes, 82.2% total excision rate, and 73% of the T_0 population showed marker excision. It is a more efficient marker excision strategy than the floral or pollen-specific promoter controlled Cre/lox systems.

2.2.2 Marker-Free Targeted-Gene Insertion

Dong et al. (2020) have demonstrated the targeted insertion of carotenoid gene cassette of GR2 (lacking selectable marker gene and T-DNA border sequences) at genomic safe harbors (GSHs) site. GSHs are the regions in the genome that can accommodate transgenes without producing detrimental effects on the host organism due to genome disruption. The GSHs were the five intergenic mutation sites identified by mutant screening, which do not exhibit visible morphological changes compared with parental phenotype. The CRISPR/Cas9-based DSB followed by donor templates assisted HDR at the target location was used to insert the gene cassette. T_0 plants were confirmed through polymerase chain reaction (PCR) for the presence of gene cassette and event (48-A7) with a golden color phenotype, which was characterized for the carotenoid using high-performance liquid chromatography (HPLC).

2.2.3 Ribonucleoprotein Based Transformation

Alternative to vector-mediated genome editing, a new method of DNA-free genome editing through RNP complex introduced by Svitashev et al. (2016) in maize by targeting four genes *viz*., (*liguleless1* (*LIG*), *acetolactate synthase* (*ALS2*), and two male fertility genes (*MS26* and *MS45*). Later on, this method was adopted in many plant species such as rice, wheat, pepper, brassica, tobacco, cabbage, apple, banana, etc. (Zhang et al., 2021). The delivery method of the RNP complex in protoplast and zygote utilized polyethylene glycol (PEG) followed by electroporation. However, particle bombardment has been used in rice, wheat, and maize embryos as well as calli (Zhang et al., 2021). In the case of rice, the premier work of RNP-based genome editing has been conducted by targeting the *phytoene desaturase* (PDS) gene to test the efficiency of different Cas9 variants using particle bombardment in scutellar derived embryos (Banakar et al., 2019; Banakar et al., 2020). In RNP-based genome editing, the RNP complex could be delivered into embryos or calli either alone (SMG-free) or co-delivered with a plasmid containing a selectable marker using standard particle delivery protocol. The detailed protocol for biolistic delivery of RNP complex is discussed in maize, wheat, and rice (Svitashev et al., 2016; Liang et al., 2017; Banakar et al., 2020). The main advantage of co-delivery of RNP complex and a plasmid containing SMG is that the transformed cells grow favorably on antibiotic selection media, and transformation efficiency increases in rice embryo-derived callus (Banakar et al., 2019). Apart from embryo and callus, the primarily and widely used explant for RNP-based genome editing is the protoplast using PEG and electroporation method. The lipofectamine reagent (TransIT-2020- water-soluble cationic lipid) has been used in a few studies to deliver RNP complex in immature embryos and calli (Svitashev et al., 2016; Banakar et al., 2020).

There are prospective reviews on the delivery methods and utilization of RNP-mediated transgene-free genome editing in various crops (Zhang et al., 2021). However, it is imperative to mention that RNP-based genome editing is challenging. It is in its starting phase, and its maximum utilization has only become possible in protoplasts, which are challenging to maintain and culture. Only a few labs have successfully utilized RNP-mediated editing versus vector-mediated genome editing (He et al., 2018). The basic workflow of RNP-based genome editing has been exhibited in various cells/tissues such as embryos, zygotes, protoplast, and callus utilizing different transformation methods (**Figure 2**). RNP-complex could be delivered through PEG or electroporation in protoplasts and zygotes, whereas in callus and embryo, RNP-complex could be bombarded by particle gun. It is noteworthy that T_0 embryo transformed plants will be chimeric, and mutation must be detected in the T_1 generation, while protoplasts, zygote, and callus-derived T_0 plants will be non-chimeric and screened through restriction digestion and targeted sequencing.

2.3 Ribonucleoprotein as a Key to Success for Marker-Free Engineered Plants

The RNP complex is constituted by nuclease and guide RNA is DNA- and SMG-free approach. Earlier, RNP-based edited rice plants have been generated for grain size and grain weight (Toda et al., 2019; Banakar et al., 2020). The fragrance is considered one of the essential grain quality traits in rice as it determines the market price. The aroma in rice is associated with an increased amount of 2-acetyl-1-pyrroline (2AP) controlled by the *betaine aldehyde dehydrogenase2* (*badh2*) gene (Buttery et al., 1983). The sequence alignment of the *OsBADH2* gene among non-fragrant and fragrant lines revealed few mutations i.e., 8-bp deletion and three SNPs in exon 7, 7-bp deletion in exon 2, and 803-bp (intronic) deletions between exon 4 and 5 (Shan et al., 2015). These mutations introduce a premature stop codon upstream of key coding regions, making this gene non-functional (*badh2*) (Hashemi et al., 2013; Shan et al., 2015). A few attempts have been made to introduce aroma in non-aromatic rice through RNAi (Niu et al., 2008) and genome-editing approaches. Recently, Ashokkumar et al. (2020) successfully created novel alleles in rice variety ASD16 by knocking out the *OsBADH2* gene through a vector-based CRISPR approach.

In our laboratory, we attempted the editing of the *OsBADH2* gene in non-aromatic rice. Basmati rice belongs to aromatic rice that has a pleasant and exquisite aroma with a low yield. However, elite cultivar PR114 lacks aroma in contrast to basmati rice. Its average yield is 6.9 tons per hectare, whereas, Basmati varieties have an average yield of 4.0 tons per hectare. The introduction of aroma in PR114 without disturbing its original genetic constitution will lead to premium quality aromatic high-yielding rice. It would lead to a major revolution for the stakeholders. A total of 1,100 embryos were

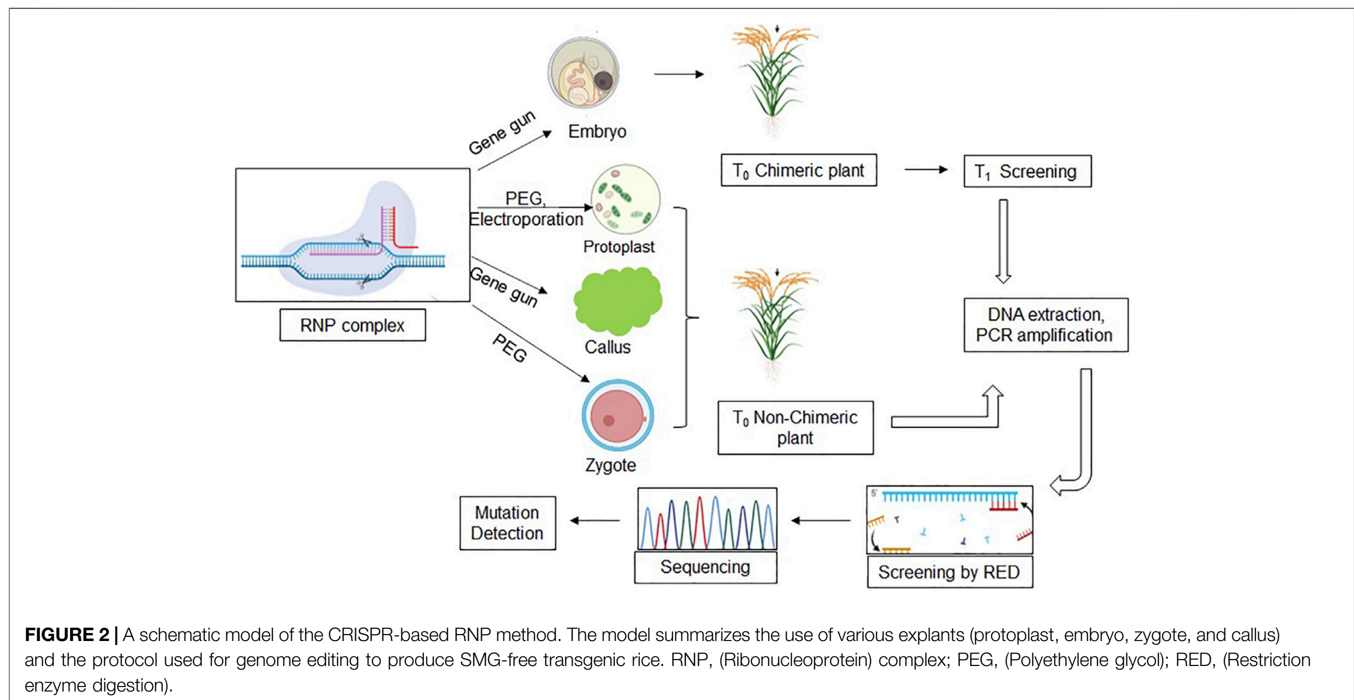


FIGURE 2 | A schematic model of the CRISPR-based RNP method. The model summarizes the use of various explants (protoplast, embryo, zygote, and callus) and the protocol used for genome editing to produce SMG-free transgenic rice. RNP, (Ribonucleoprotein) complex; PEG, (Polyethylene glycol); RED, (Restriction enzyme digestion).

bombarded by the RNP complex coated gold particles for exon 2 and exon 7 using the protocols outlined by Banakar et al. (2020). In total, 731 embryos were germinated under *in vitro* conditions on MS synthetic media, and 253 plantlets were transferred to soil. Only 35 plants survived in a glasshouse (**Figure 3A**), screened using the MSBSP-PCR (Mutation Site-Based Specific Primers-PCR) technique (Guo et al., 2018). Seven putative edited plants were obtained through the MSBSP-PCR (**Figure 3B**) and were subjected to Sanger sequencing (**Figure 3C**). The sequences of putative edited plants were aligned against the PR114 reference sequence using Clustal Omega software, which revealed the addition of a nucleotide "A" at 4-bp upstream of PAM sequence in the target site of the edited plant # 11–4 (Plant no. 11, tiller no. 4; **Figure 3D**). The alignment of the amino acid sequence of PR114 (**Figure 3E**) and plant 11–4 using the Expasy online tool showed the frameshift mutation in exon 7 (**Figure 3F**). The confirmed T_0 plant progeny will be raised and screened through molecular and biochemical analysis. To the best of out knowledge, this is the first report on RNP-based *OsBADH2* gene editing.

2.4 Regulatory Perspectives

The presence of SMGs, especially *hpt*, *npt II*, and *Bar* genes in transgenic rice is one of the major hurdles in their regulatory approval. The reasons behind their strict regulations are; the spread of their resistance in natural flora and fauna, and unintended changes in plant transcriptome and metabolome (pleiotropic effect) (Chong-Pérez and Angenon, 2013). Newly developed food that is genetically modified (GM) for a particular trait has to go through rigorous testing at molecular, biochemical, and metabolic levels for food and feed safety (including toxicity, allergenicity, and anti-nutrient). This process also makes sure the claims of substantial equivalence to non-GM wild type phenotypes are valid and that the genetically modified food is safe for environmental release (Giraldo et al., 2019). Regulatory concerns related to the presence of SMGs and the importance of their withdrawal from gene cassettes needed for further approval are apparent in a few examples of GM rice events produced in the past. The first best example of transgenic rice is "Xianyou 63", approved for release by China through co-transformation of two separate plasmids harboring *cry1Ab/Ac* and *hpt* selectable marker, respectively. The events developed were passed through the regulatory regime, and molecular characterization revealed the insertion of truncated *hpt* gene fragments (Lu, 2010). Another case is Golden rice 1 (GR1), harboring gene cassette for beta-carotene and *hpt* as a selectable marker. Event GR1 was unacceptable due to public concerns about the *hpt* marker gene. Thus, another event GR2, with a higher accumulation of beta-carotene than GR1, was produced by Syngenta using the *pmi* gene (Paine et al., 2005).

From its early development, the Golden rice trait (from GR2E event with single gene copy) has been successfully introgressed into elite rice cultivars *viz.* R64, PSBRc82, and BR29 using backcross breeding (MallikarjunaSwamy et al., 2021). After facing all the regulatory parameters, the GR2E event has been approved for consumption in different continental parts, including Australia, Canada, New Zealand, the Philippines, and the United States (https://www.goldenrice.org/). The regulation of newly developed GMOs comes under three categories. 1) process-based (for example, Europe) where the overall process or technique used to make GMO is regulated, 2) product-

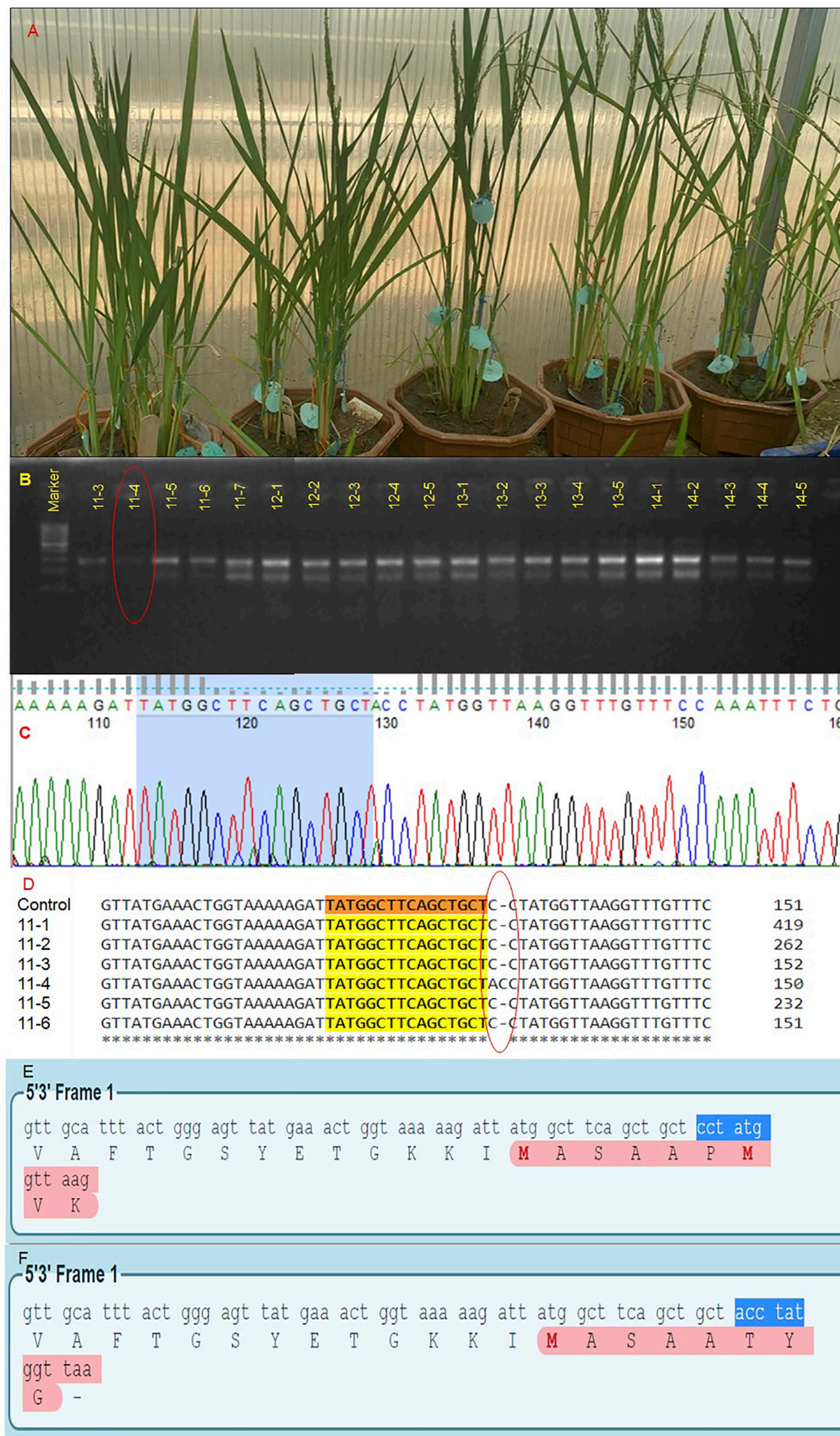


FIGURE 3 | Editing of *OsBADH2* gene for generation of aromatic rice using RNP approach. Acclimatized T_0 edited plants for *OsBADH2* gene grown under glasshouse conditions **(A)**, Detection of RNP-based editing in the T_0 generation through mutation site based specific primers technique (MSBSP). Encircled lane depicts the mutation **(B)**, A electropherogram showing the result of Sanger sequencing **(C)**, Multiple sequence alignment of putative T_0 plants showing the addition of a nucleotide "A" 4-bp upstream of the PAM site **(D)**, The ORF of *OsBADH2* exon seven in PR114 **(E)**, The ORF of *Osbadh2* exon seven in the edited plant, 11-4 showing change in the last four amino acid sequences indicating the disruption of protein chain **(F)**.

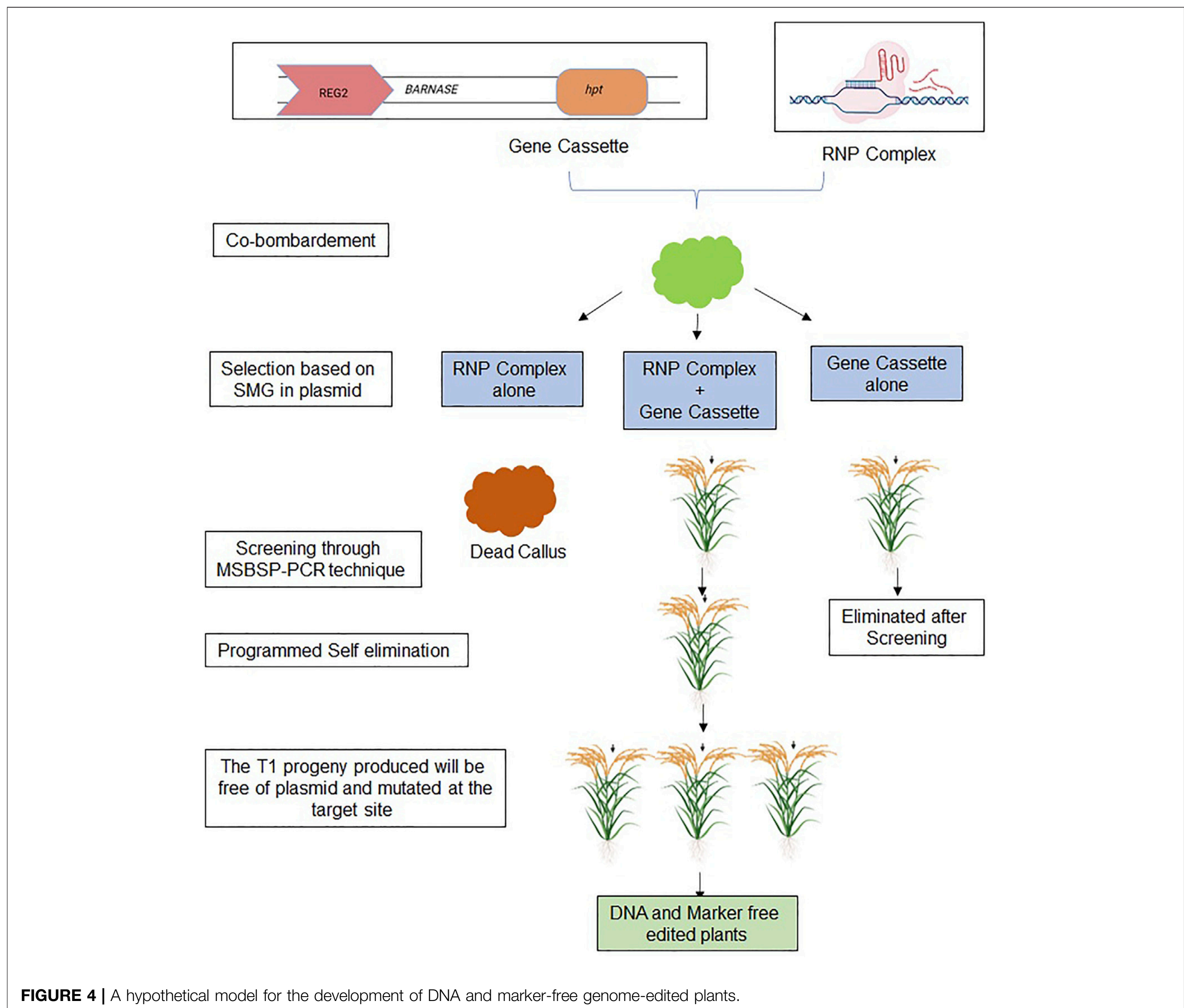


FIGURE 4 | A hypothetical model for the development of DNA and marker-free genome-edited plants.

based (for example, the United States) where the only final product is regulated, and 3) both at the process as well as product-level regulation (for example, India). The major opponents of Golden rice are the European Union (EU), where regulation is applied to food and feed products and is a process-based regulatory scheme (https://www.ncbi.nlm.nih.gov/books/NBK424533/). Even genome-edited crops using CRISPR/Cas9 were also included in the definition of GMO as per the European Court of Justice (ECJ) in 2018 (Turnbull et al., 2021). In contrast, North America and especially the United States do not have any specific federal laws for the process regulation through which GMOs are produced. The newly developed GM products are directed to specialized regulatory bodies to assess the health, safety, and environmental laws, which are the same as those used for conventional products. In Africa, the two main approaches for seed development include biotechnology and conventional, which contribute to food and nutritional security. The former is regulated under the Biosafety act and later through the Seed act and is often accompanied by National Performance Trials (NPTs) (Akinbo et al., 2021) to ensure harmony in decision making.

2.5 Future Prospects

Research that aims to create SMGs-free transgenic crops has always encouraged plant molecular biologists to adopt new ways to remove selectable markers from the GM plant background. The most widely used method is co-transformation. However, it is laborious to screen a segregating population for SMG-free plants and even could not be possible in vegetatively propagated crops (Breyer et al., 2014). Alternative to the traditional methods, CRISPR/Cas9-based genetic manipulations enable the development of SMG-free crops easily and precisely. The CRISPR/Cas9 method to make SMG-free rice is at the initial

stage and few attempts have been made to improve the technique. The CRISPR-based *npt* II marker degradation in transgenic tobacco has been reported by Rezaei et al. (2021) and programmed self-elimination in rice by He et al. (2018). These techniques could pave the way to making SMG-free engineered plants in the future.

There are a few limitations of the RNP-mediated genome-editing through CRISPR: 1) the low transformation efficiency of RNP; 2) the difficulty of screening plants; and 3) using embryos as an explant shows chimerism in the T_0 stage. A novel strategy has been proposed to overcome these limitations. Studies in rice have reported that over ten distinct plasmids could be delivered together into the plant genome by particle bombardment (Chen et al., 1998). The transformation of two plasmids using a biolistic gene gun exhibited a higher frequency (85%) in contrast to a single plasmid (Hilliou et al., 1999). The co-delivery of RNP and plasmid with selectable markers is a highly beneficial technique (Banakar et al., 2019). The RNP complex can edit the target gene without the integration of CRISPR elements into the genome and reduces the number of off-targets due to transient presence. The selectable marker in the plasmid facilitates the easy selection of transformed plants. This technique combines the benefits of the targeted mutation of RNP-mediated transformation and the easy selection process of a selectable marker in a plasmid. He et al. (2018) demonstrated the technique, TKC (Trangene Killer CRISPR), for the elimination of plasmids from the mutated plant using the suicidal gene (BARNASE) under the control of REG2 promoter (expressed during the early embryo development stage).

A combination of two approaches, the co-delivery of the RNP complex along with a gene cassette consisting of a suicidal gene and antibiotic-selectable markers (*hpt, npt II, etc.*) has been proposed as a new method of genome editing that is DNA and marker-free (**Figure 4**). In this approach, three scenarios are formed: 1) RNP transformed cells; 2) cells transformed with both RNP and cassette; and 3) cells transformed with only cassette. Transformed cells with RNP-cassette and cassette only would survive during the first screening step using selective media, while transformed plants with RNP only would be lost. In the second round of selection by MSBSP-PCR, the T_0 plants with cassette only would be eliminated. The plants with both RNP and cassette would be selected and advance to the next generation. When these screened plants reach the seed setting stage, embryos with cassette would be killed as per the Programmed Self elimination effect, whereas plants with mutated target site lacking cassette would survive. Hence, seeds obtained from the T_0 generation would be DNA and marker-free edited plants. We hypothesize that a straightforward and novel approach to making marker-free engineered crops for food security will support developing countries in introducing the product, thus contributing to the prologue of these products all over the world.

AUTHOR CONTRIBUTIONS

Conceptualization: YV and PK; Literature survey: RS, NK, MT, and UP; Methodology: NK, RS; Illustration preparation: NK, UP, JS, and GK; Analyzed the data: NK and YV; Original draft: RS, NK, UP, and GK; Finalized the manuscript: YV, JS, and KN; All authors approved the final version.

FUNDING

The guidance and support provided by Dr. Parveen Chhuneja, Director, School of Agricultural Biotechnology, and Dr. Navtej Singh Bains, Former Director of Research, Punjab Agricultural University, Ludhiana, is duly acknowledged.

ACKNOWLEDGMENTS

We are thankful to the Science and Engineering Research Board (SERB), India, Department of Science and Technology (DST), India and Confederation of Indian Industry (CII), India for granting the Prime Minister's Fellowship for Doctoral Research to RS. The guidance and support provided by Dr. Parveen Chhuneja, Director, School of Agricultural Biotechnology, and Dr. Navtej Singh Bains, Former Director of Research, Punjab Agricultural University, Ludhiana, is duly acknowledged.

REFERENCES

Ahmed, M. M. S., Bian, S., Wang, M., Zhao, J., Zhang, B., Liu, Q., et al. (2017). RNAi-Mediated Resistance to Rice Black-Streaked Dwarf Virus in Transgenic Rice. *Transgenic Res.* 26, 197–207. doi:10.1007/s11248-016-9999-4

Akinbo, O., Obukosia, S., Ouedraogo, J., Sinebo, W., Savadogo, M., Timpo, S., et al. (2021). Commercial Release of Genetically Modified Crops in Africa: Interface between Biosafety Regulatory Systems and Varietal Release Systems. *Front. Plant Sci.* 12, 314. doi:10.3389/fpls.2021.605937

Albabili, S., and Beyer, P. (2005). Golden Rice - Five Years on the Road - Five Years to Go? *Trends Plant Sci.* 10 (12), 565–573. doi:10.1016/j.tplants.2005.10.006

Ali, Z., Shami, A., Sedeek, K., Kamel, R., Alhabsi, A., Tehseen, M., et al. (2020). Fusion of the Cas9 Endonuclease and the VirD2 Relaxase Facilitates Homology-Directed Repair for Precise Genome Engineering in Rice. *Commun. Biol.* 3, 1–13. doi:10.1038/s42003-020-0768-9

Aliaga-Franco, N., Zhang, C., Presa, S., Srivastava, A. K., Granell, A., Alabadí, D., et al. (2019). Identification of Transgene-free CRISPR-Edited Plants of Rice, Tomato, and Arabidopsis by Monitoring DsRED Fluorescence in Dry Seeds. *Front. Plant Sci.* 10 (10), 1150. doi:10.3389/fpls.2019.01150

Ashokkumar, S., Jaganathan, D., Ramanathan, V., Rahman, H., Palaniswamy, R., Kambale, R., et al. (2020). Creation of Novel Alleles of Fragrance Gene OsBADH2 in Rice through CRISPR/Cas9 Mediated Gene Editing. *PloS One* 15 (8), e0237018. doi:10.1371/journal.pone.0237018

Bai, X., Wang, Q., and Chu, C. (2008). Excision of a Selective Marker in Transgenic Rice Using a Novel Cre/loxP System Controlled by a Floral Specific Promoter. *Transgenic Res.* 17 (6), 1035–1043. doi:10.1007/s11248-008-9182-7

Baisakh, N., Rehana, S., Rai, M., Oliva, N., Tan, J., Mackill, D. J., et al. (2006). Marker-free Transgenic (MFT) Near-Isogenic Introgression Lines (NIILs) of

'golden' Indica Rice (Cv. IR64) with Accumulation of Provitamin A in the Endosperm Tissue. *Plant Biotechnol. J.* 4 (4), 467–475. doi:10.1111/j.1467-7652.2006.00196.x

Banakar, R., Eggenberger, A. L., Lee, K., Wright, D. A., Murugan, K., Zarecor, S., et al. (2019). High-frequency Random DNA Insertions upon Co-delivery of CRISPR-Cas9 Ribonucleoprotein and Selectable Marker Plasmid in Rice. *Sci. Rep.* 9 (1), 19902. doi:10.1038/s41598-019-55681-y

Banakar, R., Schubert, M., Collingwood, M., Vakulskas, C., Eggenberger, A. L., and Wang, K. (2020). Comparison of CRISPR-Cas9/Cas12a Ribonucleoprotein Complexes for Genome Editing Efficiency in the Rice Phytoene Desaturase (OsPDS) Gene. *Rice* 13 (1), 4. doi:10.1186/s12284-019-0365-z

Boddupally, D., Tamirisa, S., Gundra, S. R., Vudem, D. R., and Khareedu, V. R. (2018). Expression of Hybrid Fusion Protein (Cry1Ac::ASAL) in Transgenic Rice Plants Imparts Resistance against Multiple Insect Pests. *Sci. Rep.* 8 (1), 8458. doi:10.1038/s41598-018-26881-9

Breitler, J.-C., Meynard, D., Boxtel, J. V., Royer, M., Bonnot, F., Cambillau, L., et al. (2004). A Novel Two T-DNA Binary Vector Allows Efficient Generation of Marker-free Transgenic Plants in Three Elite Cultivars of Rice (Oryza Sativa L.). *Transgenic Res.* 13 (3), 271–287. doi:10.1023/b:trag.0000034626.22918.0a

Breyer, D., Kopertekh, L., and Reheul, D. (2014). Alternatives to Antibiotic Resistance Marker Genes forIn VitroSelection of Genetically Modified Plants - Scientific Developments, Current Use, Operational Access and Biosafety Considerations. *Crit. Rev. Plant Sci.* 33 (4), 286–330. doi:10.1080/07352689.2013.870422

Buttery, R. G., Ling, L. C., Juliano, B. O., and Turnbaugh, J. G. (1983). Cooked Rice Aroma and 2-Acetyl-1-Pyrroline. *J. Agric. Food Chem.* 31 (4), 823–826. doi:10.1021/jf00118a036

Cao, M.-X., Huang, J.-Q., Yao, Q.-H., Liu, S.-J., Wang, C.-L., and Wei, Z.-M. (2006). Site-specific DNA Excision in Transgenic Rice with a Cell-Permeable Cre Recombinase. *Mb* 32 (1), 055–064. doi:10.1385/MB:32:1:055

Chen, L., Marmey, P., Taylor, N. J., Brizard, J.-P., Espinoza, C., D'Cruz, P., et al. (1998). Expression and Inheritance of Multiple Transgenes in Rice Plants. *Nat. Biotechnol.* 16 (11), 1060–1064. doi:10.1038/3511

Chong-Prez, B., and Angeno, G. (2013). "Strategies for Generating Marker-free Transgenic Plants," in *Genetic Engineering*. Editor S. N. Idah. Intechopen, 17–48. doi:10.5772/55573

Cotsaftis, O., Sallaud, C., Breitler, J. C., Meynard, D., Greco, R., Pereira, A., et al. (2002). Transposon-mediated Generation of T-DNA-And Marker-free Rice Plants Expressing a Bt Endotoxin Gene. *Mol. Breed.* 10 (3), 165–180. doi:10.1023/A:1020380305904

Das, G., and Rao, G. J. N. (2015). Molecular Marker Assisted Gene Stacking for Biotic and Abiotic Stress Resistance Genes in an Elite Rice Cultivar. *Front. Plant Sci.* 6, 698. doi:10.3389/fpls.2015.00698

Datta, K., Baisakh, N., Oliva, N., Torrizo, L., Abrigo, E., Tan, J., et al. (2003). Bioengineered 'golden' Indica Rice Cultivars with β-carotene Metabolism in the Endosperm with Hygromycin and Mannose Selection Systems. *Plant Biotechnol. J.* 1 (2), 81–90. doi:10.1046/j.1467-7652.2003.00015.x

Dong, O. X., Yu, S., Jain, R., Zhang, N., Duong, P. Q., Butler, C., et al. (2020). Marker-free Carotenoid-Enriched Rice Generated through Targeted Gene Insertion Using CRISPR-Cas9. *Nat. Commun.* 11 (1), 1–10. doi:10.1038/s41467-020-14981-y

Dormatey, R., Sun, C., Ali, K., Fiaz, S., Xu, D., Calderón-Urrea, A., et al. 2021). ptxD/Phi as Alternative Selectable Marker System for Genetic Transformation for Bio-Safety Concerns: a Review. *PeerJ.* 9, e11809. doi:10.7717/peerj.11809

Ebinuma, H., and Komamine, A. (2001). Mat (Multi-Auto-Transformation) Vector System. The Oncogenes of Agrobacterium as Positive Markers for Regeneration and Selection of Marker-free Transgenic plantsThe Oncogenes of Agrobacterium as Positive Markers for Regeneration and Selection of Marker-free Transgenic Plants. *In Vitro* Cell. *Vitro Cell.Dev.Biol.-Plant.* 37, 103–113. doi:10.1007/s11627-001-0021-2

Ebinuma, H., Sugita, K., Endo, S., Matsunaga, E., and Yamada, K. (2005). Elimination of Marker Genes from Transgenic Plants Using MAT Vector Systems. *Methods Mol. Biol.* 286, 237–254. doi:10.1385/1-59259-827-7:237

Endo, S., Sugita, K., Sakai, M., Tanaka, H., and Ebinuma, H. (2002). Single-step Transformation for Generating Marker-free Transgenic Rice Using the Ipt-type MAT Vector System. *Plant J.* 30 (1), 115–122. doi:10.1046/j.1365-313x.2002.01272.x

Estiati, A. (2020). Development of Bt Rice Potential for Yellow Stem Borer Control. *J. Crop Sci. Biotechnol.* 23, 395–403. doi:10.1007/s12892-020-00025-w

European Parliament Council of the European Union (2001). Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the Deliberate Release into the Environment of Genetically Modified Organism and Repealing Council Directive 90/220/EEC. Brussels: off. *J. Eur. Comm.* L106, 1–38.

Feng, Z., Yuan, M., Zou, J., Wu, L. B., Wei, L., Chen, T., et al. (2021). Development of Marker-free Rice with Stable and High Resistance to Rice Black-streaked Dwarf Virus Disease through RNA Interference. *Plant Biotechnol. J.* 19 (2), 212–214. doi:10.1111/pbi.13459

Fiaz, S., Ahmar, S., Saeed, S., Riaz, A., Mora-Poblete, F., and Jung, K.-H. (2021). Evolution and Application of Genome Editing Techniques for Achieving Food and Nutritional Security. *Ijms* 22 (11), 5585. doi:10.3390/2Fijms2211558510.3390/ijms22115585

Fraiture, M.-A., Roosens, N. H. C., Taverniers, I., De Loose, M., Deforce, D., and Herman, P. (2016). Biotech Rice: Current Developments and Future Detection Challenges in Food and Feed Chain. *Trends Food Sci. Technol.* 52, 66–79. doi:10.1016/j.tifs.2016.03.011

Gao, L., Xia, Z., Jiang, G., Peng, H., Zhao, X., and Zhai, W. (2011). Generation of Marker-free, Bacterial Blight-Resistant Transgenic Sterile Line and Hybrid Rice with Xa21. *Plant Breed.* 130 (4), 438–443. doi:10.1111/j.1439-0523.2011.01847.x

Gao, X., Zhou, J., Li, J., Zou, X., Zhao, J., Li, Q., et al. (2015). Efficient Generation of Marker-free Transgenic Rice Plants Using an Improved Transposon-Mediated Transgene Reintegration Strategy. *Plant Physiol.* 167 (1), 11–24. doi:10.1104/pp.114.246173

Garg, M., Sharma, N., Sharma, S., Kapoor, P., Kumar, A., and Ch, V. (2018). Biofortified Crops Generated by Breeding, Agronomy, and Transgenic Approaches Are Improving Lives of Millions of People Around the World. *Front. Nutr.* 5, 12. doi:10.3389/fnut.2018.00012

Giraldo, P. A., Shinozuka, H., Spangenberg, G. C., Cogan, N. O., and Smith, K. F. (2019). Safety Assessment of Genetically Modified Feed: Is There Any Difference from Food? *Front. Plant Sci.* 10, 1592. doi:10.3389/fpls.2019.01592

Goldsbrough, A., Lastrella, C., and Yoder, J. (1993). Transposition Mediated Repositioning and Subsequent Elimination of Marker Genes from Transgenic Tomato. *Nat. Biotechnol.* 11, 1286–1292. doi:10.1038/nbt1193-1286

Guo, J., Li, K., Jin, L., Xu, R., Miao, K., and Yang, F. (2018). A Simple and Cost-Effective Method for Screening of CRISPR/Cas9-induced Homozygous/biallelic Mutants. *Plant Methods* 14 (1), 1–10. doi:10.1186/s13007-018-0305-8

Hashemi, F. G., Rafii, M. Y., Ismail, M. R., Mahmud, T. M. M., Rahim, H. A., and Asfaliza, R. (2013). Biochemical, Genetic and Molecular Advances of Fragrance Characteristics in Rice. *Crit. Rev. Plant Sci.* 32 (6), 445–457. doi:10.1080/07352689.2013.807716

Hashizume, F., Hino, S., Kakehashi, M., Okajima, T., Nadano, D., and Aoki, N. (2008). Development and Evaluation of Transgenic Rice Seeds Accumulating a Type II-Collagen Tolerogenic Peptide. *Trans. Res.* 17 (6), 1117–1129. doi:10.1007/s11248-008-9187-2

Hashizume, F., Nakazaki, T., Tsuchiya, T., and Matsuda, T. (2006). Effectiveness of Genotype-Based Selection in the Production of Marker-free and Genetically Fixed Transgenic Lineages: Ectopic Expression of a Pistil Chitinase Gene Increases Leaf-Chitinase Activity in Transgenic Rice Plants without Hygromycin-Resistance Gene. *Plant Biotechnol.* 23 (4), 349–356. doi:10.5511/plantbiotechnology.23.349

He, Y., Zhu, M., Wang, L., Wu, J., Wang, Q., and Wang, R. (2018). Programmed Self-Elimination of the CRISPR/Cas9 Construct Greatly Accelerates the Isolation of Edited and Transgene-free Rice Plants. *Mol. plant* 11 (9), 1210–1213. doi:10.1016/j.molp.2018.05.005

Hernandez-Soto, A., Echeverria-Beirute, F., Abdelnour-Esquivel, Ana., Valdez, M., Boch, J., and Gatica-Arias, A. (2021). Rice Breeding in the New Era: Comparison of Useful Agronomic Traits. *Curr. Plant Biol.* 27 (7052), 100211. doi:10.1016/j.cpb.2021.100211

Hiei, Y., Komari, T., and Kubo, T. (1997). Transformation of Rice Mediated by Agrobacterium Tumefaciens. *Plant Mol. Biol.* 35, 205–218. doi:10.1023/A:100584761549310.1007/978-94-011-5794-0_20

Hilliou, F., Christou, P., and Leech, M. J. (1999). Development of an Efficient Transformation System for Catharanthus Roseus Cell Cultures Using Particle

Bombardment. *Plant Sci.* 140 (2), 179–188. doi:10.1016/S0168-9452(98)00225-8

Hirano, N., Muroi, T., Takahashi, H., and Haruki, M. (2011). Site-specific Recombinases as Tools for Heterologous Gene Integration. *Appl. Microbiol. Biotechnol.* 92, 227–239. doi:10.1007/s00253-011-3519-5

Huang, N., Wu, L., Nandi, S., Bowman, E., Huang, J., and Sutliff, T. (2001). The Tissue-specific Activity of a Rice Beta-Glucanase Promoter (Gns9) Is Used to Select Rice Transformants. *Plant Sci.* 161 (3), 589–595. doi:10.1016/S0168-9452(01)00447-2

Islam, M. R., Naveed, S. A., Zhang, Y., Li, Z., Zhao, X., and Fiaz, S. (2022). Identification of Candidate Genes for Salinity and Anaerobic Tolerance at the Germination Stage in Rice by Genome-wide Association Analyses. *Front. Genet.* 8. doi:10.3389/fgene.2022.822516

Jiang, Y., Sun, L., Jiang, M., Li, K., Song, Y., and Zhu, C. (2013). Production of Marker-free and RSV-Resistant Transgenic Rice Using a Twin T-DNA System and RNAi. *J. Biosci.* 38 (3), 573–581. doi:10.1007/s12038-013-9349-0

Khattri, A., Nandy, S., and Srivastava, V. (2011). Heat-inducible Cre-Lox System for Marker Excision in Transgenic Rice. *J. Biosci.* 36 (1), 37–42. doi:10.1007/s12038-011-9010-8

Komari, T., Hiei, Y., Saito, Y., Murai, N., and Kumashiro, T. (1996). Vectors Carrying Two Separate T-DNAs for Co-transformation of Higher Plants Mediated by Agrobacterium Tumefaciens and Segregation of Transformants Free from Selection Markers. *Plant J.* 10 (1), 165–174. doi:10.1046/j.1365-313x.1996.10010165.x

Kumar, S., Arul, L., and Talwar, D. (2010). Generation of Marker-free Bt Transgenic Indica Rice and Evaluation of its Yellow Stem Borer Resistance. *J. Appl. Genet.* 51 (3), 243–257. doi:10.1007/BF03208854

Lata, C., and Prasad, M. (2011). Role of DREBs in Regulation of Abiotic Stress Responses in Plants. *J. Exp. Bot.* 62 (14), 4731–4748. doi:10.1093/jxb/err210

Li, K. T., and Charng, Y. C. (2012). The Use of Hygromycin Phosphotransferase Gene (Hpt) with an Artificial Intron to Obtain Marker-Off Transgenic Plants. *Afr. J. Biotechnol.* 11 (6), 1330–1336. doi:10.5897/AJB11.2198

Li, X., Pan, L., Bi, D., Tian, X., Li, L., and Xu, Z. (2021). Generation of Marker-free Transgenic Rice Resistant to Rice Blast Disease Using Ac/Ds Transposon-Mediated Transgene Reintegration System. *Front. Plant Sci.* 12, 644437. doi:10.3389/fpls.2021.644437

Li, Z. H. U., Fu, Y. P., Liu, W. Z., Hu, G. C., Si, H. M., and Tang, K. X. (2007). Rapid Generation of Selectable Marker-free Transgenic Rice with Three Target Genes by Co-transformation and Anther Culture. *Rice Sci.* 14 (4), 239–246. doi:10.1016/S1672-6308(08)60001-3

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., and Zhao, Q. (2017). Efficient DNA-free Genome Editing of Bread Wheat Using CRISPR/Cas9 Ribonucleoprotein Complexes. *Nat. Commun.* 8, 14261. doi:10.1038/ncomms14261

Liu, T., Yuan, L., Deng, S., Zhang, X., Cai, H., and Ding, G. (2021). Improved the Activity of Phosphite Dehydrogenase and its Application in Plant Biotechnology. *Front. Bioeng. Biotechnol.* 9, 764188. doi:10.3389/fbioe.2021.764188

Lu, C. (2010). The First Approved Transgenic Rice in China. *Gm. Crops* 1 (3), 113–115. doi:10.4161/gmcr.1.3.12377

Lu, H. J., Zhou, X. R., Gong, Z. X., and Upadhyaya, N. M. (2001). Generation of Selectable Marker-free Transgenic Rice Using Double Right-Border (DRB) Binary Vectors. *Aust. J. Plant Physiol.* 28, 241–248. doi:10.1071/PP00129

MallikarjunaSwamy, B. P., Marundan, S., Samia, M., Ordonio, R. L., Rebong, D. B., and Miranda, R. (2021). Development and Characterization of GR2E Golden Rice Introgression Lines. *Sci. Rep.* 11 (1), 1–12. doi:10.1038/s41598-021-82001-0

Manna, M., Achary, V. M. M., Islam, T., Agrawal, P. K., and Reddy, M. K. (2016). The Development of a Phosphite-Mediated Fertilization and Weed Control System for Rice. *Sci. Rep.* 6 (1), 1–12. doi:10.1038/srep24941

Maruta, Y., Ueki, J., Saito, H., Nitta, N., and Imaseki, H. (2002). Transgenic Rice with Reduced Glutelin Content by Transformation with Glutelin A Antisense Gene. *Mol. Breed.* 8 (4), 273–284. doi:10.1023/A:1015208517669

McCormac, A. C., Fowler, M. R., Chen, D. F., and Elliott, M. C. (2001). Efficient Co-transformation of Nicotiana Tabacum by Two Independent T-DNAs, the Effect of T-DNA Size, and Implications for Genetic Separation. *Trans. Res.* 10, 143–155. doi:10.1023/A:1008909203852

Miki, B., Abdeen, A., Manabe, Y., and MacDonald, P. (2009). Selectable Marker Genes and Unintended Changes to the Plant Transcriptome. *Plant Biotechnol. J.* 7 (3), 211–218. doi:10.1111/j.1467-7652.2009.00400.x

Molla, K. A., Karmakar, S., Molla, J., Bajaj, P., Varshney, R. K., and Datta, S. K. (2020). Understanding Sheath Blight Resistance in Rice: the Road behind and the Road Ahead. *Plant Biotechnol. J.* 18 (4), 895–915. doi:10.1111/pbi.13312

Nakagawa, Y., Machida, C., Machida, Y., and Toriyama, K. (2001). A System to Induce the Deletion of Genomic Sequences Using R/RS Site-specific Recombination and the Ac Transposon in Transgenic Rice Plants. *Theor. Appl. Genet.* 102 (8), 1136–1141. doi:10.1007/s001220100580

Nandy, S., and Srivastava, V. (2012). Marker-free Site-specific Gene Integration in Rice Based on the Use of Two Recombination Systems. *Plant Biotech. J.* 10 (8), 904–912. doi:10.1111/j.1467-7652.2012.00715.x

Nanto, K., and Ebinuma, H. (2008). Marker-free Site-specific Integration Plants. *Trans. Res.* 17, 337–344. doi:10.1007/s11248-007-9106-y

Nishizawa-Yokoi, A., Endo, M., Ohtsuki, N., Saika, H., and Toki, S. (2015). Precision Genome Editing in Plants via Gene Targeting and piggyBac-Mediated Marker Excision. *Plant J.* 81 (1), 160–168. doi:10.1111/tpj.12693

Niu, X., Tang, W., Huang, W., Ren, G., Wang, Q., Luo, D., et al. (2008). RNAi-Directed Downregulation of OsBADH2 Results in Aroma (2-Acetyl-1-Pyrroline) Production in Rice (Oryza Sativa L.). *BMC Plant Biol.* 8, 100. doi:10.1186/1471-2229-8-100

Oliva, N., Chadha-Mohanty, P., Poletti, S., Abrigo, E., Atienza, G., and Torrizo, L. (2014). Large-scale Production and Evaluation of Marker-free Indica Rice IR64 Expressing Phytoferritin Genes. *Mol. Breed.* 33 (1), 23–37. doi:10.1007/s11032-013-9931-z

Paine, J., Shipton, C., Chaggar, S., Howells, R. M., Kennedy, M. J., and Vernon, J. (20052005). Improving the Nutritional Value of Golden Rice through Increased Pro-vitamin A Content. *Nat. Biotechnol.* 23, 482–487. doi:10.1038/nbt1082

Park, S. K., Shin, D., Hwang, W. H., Hur, Y. J., Kim, T. H., and Oh, S. Y. (2014). Development of Marker-free Transgenic Rice Expressing the Wheat Storage Protein, Glu-1Dy10, for Increasing Quality Processing of Bread and Noodles. *J. Life Sci.* 24 (6), 618–625. doi:10.5352/JLS.2014.24.6.618

Park, S. K., Shin, D., Hwang, W. H., Oh, S. Y., Cho, J. H., and Han, S. I. (2013). Development of Marker-free Transgenic Rice for Increasing Bread-Making Quality Using Wheat High Molecular Weight Glutenin Subunits (HMW-GS) Gene. *J. Life Sci.* 23 (11), 1317–1324. doi:10.5352/JLS.2013.23.11.1317

Parkhi, V., Rai, M., Tan, J., Oliva, N., Rehana, S., and Bandyopadhyay, A. (2005). Molecular Characterization of Marker-free Transgenic Lines of Indica Rice that Accumulate Carotenoids in Seed Endosperm. *Mol. Genet. Genom.* 274 (4), 325–336. doi:10.1007/s00438-005-0030-7

Pérez-Bernal, M., Delgado, M., Cruz, A., Abreu, D., Valdivia, O., and Armas, R. (2017). Marker-free Transgenic Rice Lines with a Defensin Gene Are Potentially Active against Phytopathogenic Fungus Sarocladium Oryzae. *Acta Phytopathol. Entomol. hung.* 52 (2), 135–144. doi:10.1556/038.52.2017.021

Pradhan, S., Chakraborty, A., Sikdar, N., Chakraborty, S., Bhattacharyya, J., and Mitra, J. (2016). Marker-free Transgenic Rice Expressing the Vegetative Insecticidal Protein (Vip) of Bacillus Thuringiensis Shows Broad Insecticidal Properties. *Planta* 244 (4), 789–804. doi:10.1007/s00425-016-2535-1

Qi, Y. B., Ye, S. H., Lu, Y. T., Jin, Q. S., and Zhang, X. M. (2009). Development of Marker-free Transgenic Cry1Ab Rice with Lepidopteran Pest Resistance by Agrobacterium Mixture-Mediated Co-transformation. *Rice Sci.* 16 (3), 181–186. doi:10.1016/S1672-6308(08)60077-3

Qi, Y., Chen, L., He, X., Jin, Q., Zhang, X., and He, Z. (2013). Marker-free, Tissue-specific Expression of Cry1Ab as a Safe Transgenic Strategy for Insect Resistance in Rice Plants. *Pest Manang. Sci.* 69 (1), 135–141. doi:10.1002/ps.3379

Rajadurai, G., Kalaivani, A., Varanavasiyappan, S., Balakrishnan, N., Udayasuriyan, V., and Sudhakar, D. (2018). Generation of Insect-Resistant Marker-free Transgenic Rice with a Novel cry2AX1 Gene. *Electron. J. Plant Breed.* 9 (2), 723–732. doi:10.5958/0975-928X.2018.00086.8

Ramana Rao, M. V., Parameswari, C., Sripriya, R., and Veluthambi, K. (2011). Transgene Stacking and Marker Elimination in Transgenic Rice by Sequential Agrobacterium-Mediated Co-transformation with the Same Selectable Marker Gene. *Plant Cell Rep*. 30 (7), 1241–1252. doi:10.1007/s00299-011-1033-y

Ramessar, K., Peremarti, A., Gomez-Galera, S., Naqvi, S., Moralejo, M., and Munoz, P. (2007). Biosafety and Risk Assessment Framework for Selectable Marker Genes in Transgenic Crop Plants: a Case of the Science Not Supporting the Politics. *Trans. Res.* 16, 261–280. doi:10.1007/s11248-007-9083-1

Rathore, K. S., Chowdhury, V. K., and Hodges, T. K. (1993). Use of Bar as a Selectable Marker Gene and for the Production of Herbicide-Resistant Rice Plants from Protoplasts. *Plant Mol. Biol.* 21, 871–884. doi:10.1007/BF00027118

Rezaei, A., Farsi, M., Malekzadeh-Shafaroudi, S., and Seifi, A. (2021). In Planta Removal of nptII Selectable Marker Gene from Transgenic Tobacco Plants Using CRISPR/Cas9 System. *Plant gene.* 26, 100288. doi:10.1016/j.plgene.2021.100288

Rosellini, D. (2012). Selectable Markers and Reporter Genes: A Well Furnished Toolbox for Plant Science and Genetic Engineering. *Crit. Rev. Plant Sci.* 5, 401–453. doi:10.1080/07352689.2012.683373

Sah, S. K., Kaur, A., Kaur, G., and Cheema, G. S. (2014). Genetic Transformation of Rice: Problems, Progress and Prospects. *J. Rice Res.* 3, 132. doi:10.4172/2375-4338.1000132

Sawada, K., Hasegawa, M., Tokuda, L., Kameyama, J., Kodama, O., and Kohchi, T. (2004). Enhanced Resistance to Blast Fungus and Bacterial Blight in Transgenic Rice Constitutively Expressing OsSBP, a Rice Homologue of Mammalian Selenium-Binding Proteins. *Biosci. Biotechnol. Biochem.* 68, 873–880. doi:10.1271/bbb.68.873

Sengupta, S., Chakraborti, D., Mondal, H. A., and Das, S. (2010). Selectable Antibiotic Resistance Marker Gene-free Transgenic Rice Harboring the Garlic Leaf Lectin Gene Exhibits Resistance to Sap-Sucking Planthoppers. *Plant Cell Rep.* 29 (3), 261–271. doi:10.1007/s00299-010-0819-7

Shan, Q., Zhang, Y., Chen, K., Zhang, K., and Gao, C. (2015). Creation of Fragrant Rice by Targeted Knockout of the OsBADH2 Gene Using TALEN Technology. *Plant Biotech. J.* 13 (6), 791–800. doi:10.1111/pbi.12312

Singh, N., Choudhury, D. R., Tiwari, G., Singh, A. K., Kumar, S., and Srinivasan, K. (2016). Genetic Diversity Trend in Indian Rice Varieties: an Analysis Using SSR Markers. *BMC Genet.* 17, 127. doi:10.1186/s12863-016-0437-7

Singh, P. K., Verma, R. L., Singh, R., Singh, R. K., Singh, H. B., and Arsode, P. (2020). "Biotic Stress Management in Rice (Oryza Sativa L.) through Conventional and Molecular Approaches," in *New Frontiers in Stress Management for Durable Agriculture*. Editors A. Rakshit, H. Singh, A. Singh, U. Singh, and L. Fraceto (Singapore: Springer), 609–644. doi:10.1007/978-981-15-1322-0-3010.1007/978-981-15-1322-0_30

Sreekala, C., Wu, L., Gu, K., Wang, D., Tian, D., and Yin, Z. (2005). Excision of a Selectable Marker in Transgenic Rice (Oryza Sativa L.) Using a Chemically Regulated Cre/loxP System. *Plant Cell Rep.* 24 (2), 86–94. doi:10.1007/s00299-004-0909-5

Sripriya, R., Raghupathy, V., and Veluthambi, K. (2008). Generation of Selectable Marker-free Sheath Blight Resistant Transgenic Rice Plants by Efficient Cotransformation of a Cointegrate Vector T-DNA and a Binary Vector T-DNA in One Agrobacterium Tumefaciens Strain. *Plant Cell Rep.* 27 (10), 1635–1644. doi:10.1007/s00299-008-0586-x

Stallworth, S., Schumaker, B., Fuller, M. G., and Tseng, T. M. (2020). "Consequences and Mitigation Strategies of Biotic and Abiotic Stress in Rice (Oryza Sativa L.)," in *Plant Stress Physiology*. Editor A. Hossain Intechopen, London. doi:10.5772/intechopen.91402

Svitashev, S., Schwartz, C., Lenderts, B., Young, J. K., and Cigan, A. M. (2016). Genome Editing in Maize Directed by CRISPR–Cas9 Ribonucleoprotein Complexes. *Nat. Commun.* 7, 13274. doi:10.1038/ncomms13274

Tan, J., Wang, Y., Chen, S., Lin, Z., Zhao, Y., and Xue, Y. (2022). An Efficient Marker Gene Excision Strategy Based on CRISPR/Cas9-mediated Homology-Directed Repair in Rice. *Int. J. Mol. Sci.* 23 (3), 1588. doi:10.3390/ijms23031588

Toda, E., Koiso, N., Takebayashi, A., Ichikawa, M., Kiba, T., and Osakabe, K. (2019). An Efficient DNA-And Selectable-marker-free Genome-Editing System Using Zygotes in Rice. *Nat. Plants* 5 (4), 363–368. doi:10.1038/s41477-019-0386-z

Tu, J., Datta, K., Oliva, N., Zhang, G., Xu, C., and Khush, G. S. (2003). Site-independently Integrated Transgenes in the Elite Restorer Rice Line Minghui 63 Allow Removal of a Selectable Marker from the Gene of Interest by Self-segregation. *Plant Biotechnol. J.* 1 (3), 155–165. doi:10.1046/j.1467-7652.2003.00012.x

Turnbull, C., Lillemo, M., and Hvoslef-Eide, T. A. K. (2021). Global Regulation of Genetically Modified Crops and the Gene Edited Crop Boom – A Review. *Front. Plant Sci.* 12, 630396. doi:10.3389/fpls.2021.630396

Twyman, R. M., Stoger, E., Kohli, A., Capell, T., and Christou, P. (2002). "Selectable and Screenable Markers for Rice Transformation," in *Testing for Genetic Manipulation in Plants. Molecular Methods of Plant Analysis*. Editors J. F. Jackson and H. F. Linskens (Berlin, Heidelberg: Springer), 22. doi:10.1007/978-3-662-04904-4_1

Wang, M., Lu, Y., Botella, J. R., Mao, Y., Hua, K., and Zhu, J.-,k. (2017). Gene Targeting by Homology-Directed Repair in Rice Using a Geminivirus-Based CRISPR/Cas9 System. *Mol. Plant.* 10, 1007–1010. doi:10.1016/j.molp.2017.03.002

Waseem, M., Huang, F., Wang, Q., Aslam, M. M., Abbas, F., Ahmad, F., et al. (2021). Identification, Methylation Profiling, and Expression Analysis of Stress-Responsive Cytochrome P450 Genes in Rice under Abiotic and Phytohormones Stresses. *GM Crops Food* 12 (1), 551–563. doi:10.1080/2F21645698.2021.190881310.1080/21645698.2021.1908813

Woo, H. J., Qin, Y., Park, S. Y., Park, S. K., Cho, Y. G., and Shin, K. S. (2015). Development of Selectable Marker-free Transgenic Rice Plants with Enhanced Seed Tocopherol Content through FLP/FRT-mediated Spontaneous Auto-Excision. *PLoS One* 10 (7), e0132667. doi:10.1371/journal.pone.0132667

Xia, Z. H., Li, X. B., Chen, C. Y., Fan, H. K., Jiang, G. H., Zhu, L. H., et al. (2006). Generation of Selectable Marker-free and Vector Backbone Sequence-free Xa21 Transgenic Rice. *Sheng Wu Gong Cheng Xue Bao* 22, 204–210.

Xu, H., Yu, M., Yin, Y., Zhu, C., Ji, W., and Zhang, C. (2020). Generation of Selectable Marker-free Soft Transgenic Rice with Transparent Kernels by Downregulation of SSSII-2. *Crop J* 8 (1), 53–61. doi:10.1016/j.cj.2019.05.006

Xu, M., Zhao, S., Zhang, Y., Yin, H., Peng, X., and Cheng, Z. (2017). Production of Marker-free Transgenic Rice (Oryza Sativa L.) with Improved Nutritive Quality Expressing AmA1. Iran. *J. Biotechnol.* 15 (2), 102. doi:10.15171/2Fijb.152710.15171/ijb.1527

Yang, A., Su, Q., An, L., Liu, J., Wu, W., and Qiu, Z. (2009). Detection of Vector- and Selectable Marker-free Transgenic Maize with a Linear GFP Cassette Transformation via the Pollen-Tube Pathway. *J. Biotechnol.* 139, 1–5. doi:10.1016/j.jbiotec10.1016/j.jbiotec.2008.08.012

Yarasi, B., Sadumpati, V., Immanni, C. P., Vudem, D. R., and Khareedu, V. R. (2008). Transgenic Rice Expressing Allium Sativum Leaf Agglutinin (ASAL) Exhibits High-Level Resistance against Major Sap-Sucking Pests. *BMC Plant Biol.* 8, 102. doi:10.1186/1471-2229-8-102

Yau, Y. Y., and Stewart, C. N. (2013). Less Is More: Strategies to Remove Marker Genes from Transgenic Plants. *BMC Biotechnol.* 13, 36. doi:10.1186/1472-6750-13-36

Yu, H. X., Liu, Q. Q., Ling, W. A. N. G., Zhao, Z. P., Li, X. U., and Huang, B. L. (2006). Breeding of Selectable Marker-free Transgenic Rice Lines Containing AP1 Gene with Enhanced Disease Resistance. *Agri. Sci. China* 5 (11), 805–811. doi:10.1016/s1671-2927(06)60128-4

Yu, H. X., Liu, Q. Q., Xu, L., Lu, M. F., Yang, X. J., and Gong, Z. Y. (2009). Quality Characteristics and Field Performance of Selectable Marker-free Transgenic Rice with Antisense Wx Gene and Improved Quality Derived from the Elite Parents of Hybrid Indica Rice. *J. Cereal Sci.* 50 (3), 370–375. doi:10.1016/j.jcs.2009.07.003

Yu, H., Yao, Q., Wang, L., Zhao, Z., Gong, Z., and Tang, S. (2009). Generation of Selectable Marker-free Transgenic Rice Resistant to Chewing Insects Using Two Co-transformation Systems. *Prog. Nat. Sci.* 19 (11), 1485–1492. doi:10.1016/j.pnsc.2009.04.005

Zafar, K., Sedeek, K. E. M., Rao, G. S., Khan, M. Z., Amin, I., and Kamel, R. (2020). Genome Editing Technologies for Rice Improvement: Progress, Prospects, and Safety Concerns. *Front. Genome Ed.* 2, 5. doi:10.3389/fgeed.2020.00005

Zhang, L., Zuris, J. A., Viswanathan, R., Edelstein, J. N., Turk, R., and Thommandru, B. (2021). AsCas12a Ultra Nuclease Facilitates the Rapid Generation of Therapeutic Cell Medicines. *Nat. Commun.* 2 (1), 3908. doi:10.1038/s41467-021-24017-8

Zhao, Y., Qian, Q., Wang, H. Z., and Huang, D. N. (2007). Co-transformation of Gene Expression Cassettes via Particle Bombardment to Generate Safe Transgenic Plant without Any Unwanted DNA. *In Vitro* Cell. *Dev. Biol.* 43 (4), 328–334. doi:10.1007/s11627-007-9051-8

Zhu, Q., Yu, S., Zeng, D., Liu, H., Wang, H., and Yang, Z. (2017). Development of "Purple Endosperm Rice" by Engineering Anthocyanin Biosynthesis in the Endosperm with a High-Efficiency Transgene Stacking System. *Mol. Plant* 10 (7), 918–929. doi:10.1016/j.molp.2017.05.008

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Singh, Kaur, Praba, Kaur, Tanin, Kumar, Neelam, Sandhu and Vikal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

ORIGINAL RESEARCH
published: 14 June 2022
doi: 10.3389/fgene.2022.871833

BSA-seq Identifies a Major Locus on Chromosome 6 for Root-Knot Nematode (*Meloidogyne graminicola*) Resistance From *Oryza glaberrima*

Gurwinder Kaur[1], Inderjit Singh Yadav[1], Dharminder Bhatia[2], Yogesh Vikal[1]*, Kumari Neelam[1], Narpinderjeet Kaur Dhillon[3], Umesh Preethi Praba[1], Gurjit Singh Mangat[2] and Kuldeep Singh[1,4]

[1]School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, India, [2]Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India, [3]Department of Plant Pathology, Punjab Agricultural University, Ludhiana, India, [4]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India

Root-knot nematode (*Meloidogyne graminicola*) is one of the emerging threats to rice production worldwide that causes substantial yield reductions. There is a progressive shift of the cropping system from traditional transplanting to direct-seeded water-saving rice production that favored the development of *M. graminicola*. Scouting and deploying new resistance genes is an economical approach to managing the root-knot nematodes. Here, we report that the inheritance of root-knot nematode resistance in *Oryza glaberrima* acc. IRGC102206 is governed by a single dominant gene. Traditional mapping coupled with BSA-seq is used to map nematode resistance gene(s) using the BC_1F_1 population derived from a cross of *O. sativa* cv. PR121 (S) and *O. glaberrima* acc. IRGC102206 (R). One major novel genomic region spanning a 3.0-Mb interval on chromosome 6 and two minor QTLs on chromosomes 2 and 4 are the potential genomic regions associated with rice root-knot nematode resistance. Within the QTL regions, 19 putative candidate genes contain 81 non-synonymous variants. The detected major candidate region could be fine mapped to accelerate marker-assisted breeding for root-knot nematode resistance in rice.

Keywords: BSA-QTLseq, candidate genes, *Oryza glaberrima*, SNPs, *Meloidogyne graminicola*

OPEN ACCESS

Edited by:
Sukhwinder Singh, The International Maize and Wheat Improvement Center (CIMMYT), India

Reviewed by:
Youlu Yuan, Cotton Research Institute (CAAS), China
Aamir W. Khan, University of Missouri, United States

Correspondence:
Yogesh Vikal
yvikal-soab@pau.edu

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *08 February 2022*
Accepted: *02 May 2022*
Published: *14 June 2022*

Citation:
Kaur G, Yadav IS, Bhatia D, Vikal Y, Neelam K, Dhillon NK, Praba UP, Mangat GS and Singh K (2022) BSA-seq Identifies a Major Locus on Chromosome 6 for Root-Knot Nematode (Meloidogyne graminicola) Resistance From Oryza glaberrima. Front. Genet. 13:871833. doi: 10.3389/fgene.2022.871833

INTRODUCTION

Rice (*Oryza sativa* L.) is one of the staple food crops that feed half of the world's population. Its production has been continuously increasing at a constant pace for the last 10 years. However, to meet the demand of the ever-growing population, there is still a need to increase rice production by 2050 (Ray et al., 2013). Over the changing agro-climatic conditions, various biotic and abiotic stresses have emerged that threaten rice production across the globe. Among biotic stresses, plant-parasitic nematodes pose the foremost warning to rice production worldwide (Mhatre et al., 2017). About 300 nematode species of 35 genera infect rice, and the *Meloidogyne* genus is the first among the top 10 plant-parasitic nematodes (Nicol et al., 2011). Within the *Meloidogyne* genus, root-knot nematode (*M. graminicola*) is the most widespread threat nowadays in almost all rice-growing systems—upland, lowland, deepwater, and irrigated rice—and causes significant yield losses ranging from 20% to 80% (Mantelin et al., 2017; Kumar, 2020).

The second stage juveniles (J2) of root-knot nematode penetrate behind the root cap due to the absence of differentiated endodermis near the root tip (Bridge et al., 2005). After penetration,

juveniles migrate intercellularly through the root cortex toward the apical meristem and the cellular differentiation region and establish a permanent feeding site in the vascular tissue to develop giant cells (Williamson, 1998). The metabolically active giant cells serve as a source of nutrients for nematodes to complete their life cycle between 19 and 27 days from juveniles to adults and release eggs around the root surface. These nematodes complete numerous generations in a single rice-growing season to build up a high damaging population in a shorter period (Shrestha et al., 2007). The mechanical disruption caused by the giant cells in metaxylem vessels interferes with the uptake of water and nutrients that strongly impair the root physiology and development (Singh, 2010). The disruption of water translocation and nutrient transport by the root vascular system leads to stunting, chlorosis, and loss of vigor, which eventually result in reduced growth and finally cause significant yield losses of the crop (Mantelin et al., 2017).

Some cultural practices are followed to limit the nematode population below a damaging threshold level. Flooding and crop rotation practices are partially effective and of limited use due to the broad host range of *M. graminicola* and an unacceptable cost of non-hosts, such as mung bean, mustard, and sesame, for small-scale farmers using rice as a staple food. The use of nematicides is uneconomic, unhealthy for the environment, and unsafe for human health. Alternatively, host resistance is effective and economical to manage root-knot nematode population densities below the threshold levels and gains significance in water-saving practices during the shifting of rice cultivation from irrigation to direct-seeded rice. Most Asian genotypes are susceptible to root-knot nematode, with only a few of them being resistant (Dimkpa et al., 2016). Natural resistance to *M. graminicola* has been reported in *O. glaberrima* (African rice) and *O. longistaminata* (Soriano et al., 1999). But due to the presence of sterility genes and the low yield potential of *O. glaberrima,* limited efforts have been made to introgress root-knot nematode resistance from *O. glaberrima* into *O. sativa*. However, fertility can be retained by recurring backcrossing for a few generations, but backcrossing for several generations increases the risk of losing desirable traits. Moreover, interspecific progenies do not show a similar type of resistance as *O. glaberrima* (Plowright et al., 1999). Several studies have reported the quantitative nature of resistance against *M. graminicola* as QTLs for root galling and the number of galls and eggs per root system have been identified using RIL populations (Shrestha et al., 2007; Jena et al., 2013; Galeng-Lawilao et al., 2018). In some crop systems, a single major gene confers resistance to different *Meloidogyne* species; for example, *rkn1* confers resistance to *M. incognita* in cotton (Wang et al., 2006), *Mi* from *Lycopersicon peruvianum* (Roberts, 1995) gives resistance to some of the root-knot nematode species in tomato (Abad et al., 2003), and *Hsa-1Og* provides resistance against cyst nematode (*Heterodera sacchari*) in rice (Lorieux et al., 2003). Recently, Mhatre et al. (2017) have reported hypersensitive response (HR) in rice cultivar Zhonghua11 (Asian rice) against *M. graminicola* and suggested that resistance is due to major genes rather than quantitative resistance. There is still a continuous need to explore and exploit related species of rice for resistance to *M. graminicola*.

Identification and characterization of genes/QTLs responsible for root-knot nematode resistance are important not only to unveil the molecular mechanisms of resistance but also to deploy the resistance genes for the development of nematode-resilient rice cultivars. Different molecular mapping strategies have been used to map genes/QTLs for several traits in rice. Bulked segregant analysis (BSA) is one of the effective methods to map genes or QTLs from a population having two extreme phenotypic traits (Michelmore et al., 1991; Venuprasad et al., 2009). Recent development in next-generation sequencing technologies has provided effective tools for genome-wide identification of SNPs and other structural variants, and also genotyping (Huang et al., 2009) has accelerated genetic mapping studies and marker development. A "BSA-seq" approach that couples whole-genome re-sequencing and BSA of extreme phenotypes is cost-effective and rapidly identifies genomic regions associated with a trait of interest (Takagi et al., 2013; Deokar et al., 2019). This approach has been used successfully in different crops such as rice, tomato, chickpea, and brassica to map QTLs of different genetic complexities from single genes to major QTLs from last few years (Takagi et al., 2013; Illa-Berenguer et al., 2015; Deokar et al., 2019; Zhang et al., 2020). There is only one report of QTL identification for rice root-knot nematode resistance through BSA-seq analysis using mapping population derived from indica and aus cultivar (Lahari et al., 2019). The present study aims to 1) determine the genetics of *M. graminicola* resistance in *O. glaberrima*, 2) identify major locus associated with root-knot nematode resistance through BSA-seq, and 3) explore candidate genes conferring *M. graminicola* resistance and identification of SNPs within the candidate genes.

MATERIALS AND METHODS

Plant Materials

The plant material consisted of *O. glaberrima* acc. IRGC102206, PR121, BC_1F_1, BC_2F_1, and BC_2F_2 populations derived from the cross of PR121 and *O. glaberrima* acc. IRGC102206 *O. glaberrima* acc. IRGC102206 were identified as highly resistant and susceptible to *M. graminicola*, respectively, in a previous study (Kaur, 2020). The resistance was confirmed over 2 years of screening both in the nematode-infested pots and sick plot under field conditions, as the number of galls per plant was significantly lower as compared to susceptible checks. *O. glaberrima* has characteristics of early maturity, moderate to tall height, seed shattering, lower yield, and more resistance to various diseases and pests. However, PR121 has short stature, bacterial blight resistance, and better lodging tolerance.

Phenotypic Evaluation of Backcross Generations for Nematode Infestation

The BC_1F_1, BC_2F_1, and BC_2F_2 populations were grown in the nursery during cropping seasons 2018–2020. The 25-day-old seedlings were transplanted in irrigated field conditions with plant-to-plant and row-to-row distances of 20 and 30 cm, respectively. Each plant of BC_1F_1, BC_2F_1, and BC_2F_2 after 15 days of transplanting was split into four plantlets, and three replicas of each BC_1F_1, BC_2F_1, and BC_2F_2 plant were transferred

to a nematode-infested sick plot with an initial population density of 1 J2/g of soil while one replica was raised under non-infested conditions. Standard agronomic practices were implemented during the raising of the crop, except that the soil was not flooded during screening in the nematode-infested sick plot. Each plant of all populations was uprooted from the nematode-infested sick plot after 60 days of transplanting. The roots were washed immediately under running tap water to count the galls per root system. Root gall index was calculated on a scale of 1–5 as given by Gaur et al. (2001). Rating was done as follows: 1 for 0–1 gall (highly resistant), 2 for 1–10 galls (resistant), 3 for 11–30 galls (moderately resistant), 4 for 30–100 galls (susceptible), and 5 for >100 galls (highly susceptible). The segregation pattern for nematode resistance was checked in each generation using standard chi-square analysis for the goodness of fit. Data for each BC_1F_1 plant on different morphological characters like plant height (cm), root length (cm), fresh shoot weight (g), and fresh root weight (g) were measured immediately after the uprooting of the plants. The shoots and roots of each plant were packed separately in brown paper bags for drying, to achieve constant weight for measuring dry shoot and root weight (g). Data were analyzed using a generalized linear model (GLM) of SAS software version 9.4 (SAS Institute, Cary, NC, United States).

Traditional Rough Mapping

Genomic DNA was isolated from the young leaves of parents and each BC_1F_1 plant using the CTAB method (Doyle and Doyle, 1987). The degradation and contamination of DNA were checked on 0.8% agarose gel while DNA was quantified using Thermo scientific NanoDropTM 1000 spectrophotometer. A total of 512 simple sequence repeat (SSR) markers from the universal core genetic map (Orjuela et al., 2010) spanning all 12 rice chromosomes were used for the parental polymorphic survey on PR121 and *O. glaberrima* acc. IRGC102206. Primer sequences were retrieved from the Gramene database (http://www.gramene.org/; IRGSP, 2005). *In vitro* amplification using polymerase chain reaction (PCR) was performed in a 96-well PCR plate in Eppendorf and Applied Biosystems master cyclers. The total PCR reaction of 20 µl was prepared using the following components: 100 ng template DNA, 0.50 µM each of forward and reverse primers and 2× Emerald Amp® GT PCR Master Mix containing an optimized buffer, PCR enzyme, dNTP mixture, gel-loading dye (green), and a density reagent. A negative control (without template DNA) was included in each amplification reaction. PCR profile of 95°C for 5 min, followed by 30 cycles of 1 min at 94°C, 1 min at 55–60°C, and 1 min at 72°C with a final extension of 10 min at 72°C was used for amplification. The amplified products were resolved in 3.0% agarose gel, and amplicons were scored by comparing them to parental alleles. The linkage map was constructed using the Kosambi mapping function of QTL IciMapping version 4.1 (Meng et al., 2015) through MAP functionality in a graphic form representing the position of markers within linkage groups by using a threshold LOD score of 3.0. Composite interval mapping (CIM) at a 95% threshold level was used for the identification of QTLs based on 1,000 permutation tests using Windows QTL cartographer version 2.5 (Wang et al., 2012). The position of putative QTLs corresponded to the location (in centiMorgans) of peak LOD scores in the scan of individual chromosomes and was designated according to the chromosome position. The proportion of observed phenotypic variance attributable to the QTL was estimated by the coefficient of determination (R^2) using the maximum likelihood of CIM.

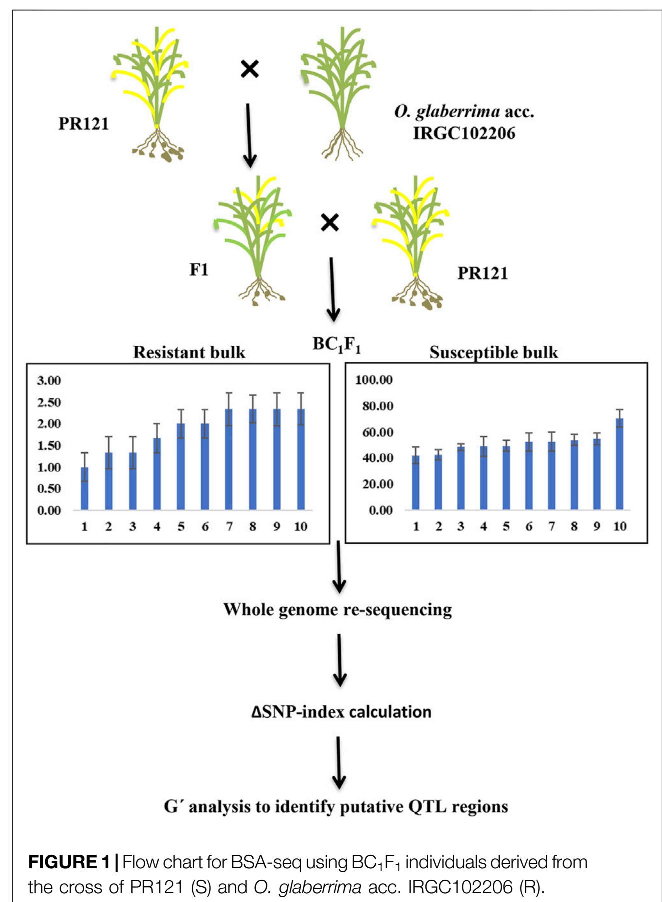


FIGURE 1 | Flow chart for BSA-seq using BC_1F_1 individuals derived from the cross of PR121 (S) and *O. glaberrima* acc. IRGC102206 (R).

Bulked Segregant Analysis Through Whole-Genome Re-Sequencing

The DNA concentration of each sample was normalized to 500 ng/µl, and 5 µg of the total DNA of the individual plant was used for making extreme bulks. An equal amount of DNA from 10 individual plants with few galls (1–2.33 galls) was mixed to generate the resistant DNA bulk. Similarly, an equal amount of DNA from 10 individual plants with higher gall numbers per plant (40–75 galls) was mixed to generate the susceptible DNA bulk (**Figure 1**). Paired-end sequencing libraries from the two extreme bulk and parents using 2 µg of DNA were prepared according to the Illumina manufacturer's instructions. The sequencing of libraries using the Illumina HiSeqTM 2500 platform was outsourced from NGB Diagnostics Pvt Ltd., India. Illumina sequencing of genomic libraries for each of the parents and the two bulks (2 × 150 bp) produced 33–35 million reads per sample for a total of 136 million raw reads. The quality

of raw reads was assessed using FASTQC (version 0.11.8; Andrews, 2010) with default parameters. Poor-quality sequences were filtered and removed while contaminated adapter sequences and any unwanted bias from their ends were trimmed using Trimmomatic (version 0.39; Bolger et al., 2014). A Phred score of 30 was kept as the overall quality threshold for raw reads. The filtered reads were further re-checked for quality using FASTQC.

BSA-seq Analysis

High-quality sequences were aligned and mapped to the *Oryza sativa* Indica Group ASM465v1 reference sequence of cultivar 93-11, available at Ensembl plants (https://plants.ensembl.org/Oryza_indica), using Bowtie 2 algorithm with default parameters (version v2.0.0; Langmead and Salzberg, 2012). The Bowtie 2 default mode is faster than all Burrows-Wheeler Aligner (BWA) modes and more than 2.5 times faster than the BWA default mode. All Bowtie 2 modes aligned a greater number of reads than either BWA or short oligonucleotide alignment program 2 (SOAP2). To keep only uniquely mapping reads, the output SAM files were converted into BAM files, then read groups were added, sorted, and indexed using SAMtools (version 0.1.19; Li et al., 2009). The output BAM files containing uniquely mapped reads were used for SNP Calling through the GATK (Genome Analysis Toolkit) pipeline. Subsequently, GATK's HaplotypCaller component (version 4.0; Mckenna et al., 2010) was used to perform a joint variant calling of all samples. The indels and missing data were filtered out using the variant filtration parameter of vcftools (https://github.com/vcftools/vcftools). The filtered VCF file in table format was used as an input file for QTLseqr (https://github.com/bmansfeld/QTLseqr) package developed by Mansfeld and Grumet (2018). SNPs with a reference allele frequency of 0.2 from both the bulks were filtered out as these might be due to sequencing or alignment error. In G' approach, run GprimeAnalysis first counted the number of SNPs within the sliding window, and then a tricube-smoothed ΔSNP index was calculated within a window size of 1.0-Mb genomic region. The ΔSNP index (>0.1) was used to calculate p-values (<0.05) and an FDR(q) of 0.01 to identify potential QTLs associated with root-knot nematode resistance. The G' determines the statistical significance of QTLs as background noise is less and also addresses the linkage disequilibrium (LD) between SNPs. One important advantage of this method is that p-values can be estimated for each SNP using non-parametric estimation of the null distribution of G' (Magwene et al., 2011). Comparison of the QTL-seq method (Delta-SNP index) and G' method by Mansfeld and Grumet (2018) showed that a confidence interval of 99% with the QTL-seq method was not as stringent as using an FDR of 0.01 in the G' method. All commands and codes for BSA-seq analysis are available at https://github.com/bmansfeld/QTLseqr.

Identification of Candidate Genes

The QTL regions harboring the candidate genes based on the annotation of *Oryza sativa* indica Group ASM465v1 (https://plants.ensembl.org/Oryza_indica) were identified. To identify non-synonymous SNPs among the candidate genes of the two parents (PR121 and *O. glaberrima*), nucleotide changes were investigated using ExPASy translate tool (http://web.expasy.org/translate/). Potential candidate genes and their corresponding Ensembl IDs were further subjected to the ShinyGO v0.74 database (Ge et al., 2019) to obtain gene ontology (GO) annotation against *O. sativa* subsp. indica. GO enrichment was calculated by a p-value cut-off (FDR) at 0.05 for the genes.

RESULTS

Inheritance of Root-Knot Nematode Resistance

The F_1s (with medium-sized ligule) generated from the cross of PR121 (S) and *O. glaberrima* acc. IRGC102206 (R) were partial to completely sterile; therefore these F_1s were backcrossed with PR121 to develop the BC_1F_1 population. A total of 10,800 spikelets of F_1 plants were cross-pollinated and a 0.95% seed setting of BC_1F_1 (103 seeds) was obtained. Out of the 103 seeds, 69 seeds (67%) were germinated in the nursery and transplanted in controlled (normal) conditions. After 25 days of transplanting, three replicas of each BC_1F_1 plant along with their parental genotypes were screened in the nematode-infested sick plot. Based on gall number, 39 and 30 BC_1F_1 plants were categorized as resistant and susceptible, respectively, for root-knot nematode resistance that corresponded to a single locus segregation ratio statistically (**Table 1**). The resistant parent IRGC102206 showed gall number from 0 to 1 while the susceptible parent PR121 exhibited gall number in the range of 64–82 (**Figure 2**). The average gall number and gall index among BC_1F_1 individuals ranged from 1.00 to 70.33 and 1.33 to 4.00, respectively (**Table 2**). The resistant parent *O. glaberrima* accession IRGC102206 had an overall mean gall index of 1.0 while the susceptible parent PR121 had a mean gall index of 4.0. The segregation pattern in successive generations, that is, BC_2F_1 and BC_2F_2, was authenticated, thereby confirming that nematode resistance is governed by a single dominant gene (**Table 1**). Overall, BC_1F_1 plants showed stunted growth under nematode infestation. Plant height and root length among BC_1F_1 plants ranged from 61 to 118 cm and 7.57 to 21.03 cm, respectively. Significant differences were observed among BC_1F_1 plants for all the traits (**Table 2**). There was a nominal decrease in growth parameters in the resistant individuals compared to the susceptible plants indicating that the plant growth parameters were affected by nematode infestation.

Traditional Mapping

We also carried out rough QTL mapping using 69 individuals of the BC_1F_1 population. A total of 512 microsatellite markers spanning 12 chromosomes of rice were used for the parental polymorphic survey and the markers per chromosome varied from 53 (chromosomes 1 and 2) to 31 (chromosome 9). Parental polymorphism among all chromosomes ranged from 7.6% to 50.9% with an average of 31.25%. A comprehensive list of polymorphic SSR markers is given in **Supplementary Table S1**. A total of 100 polymorphic SSR markers were genotyped and a genetic linkage map was generated with a total map length

TABLE 1 | Genetic analysis of root-knot nematode resistance in different generations derived from the cross of PR121 (S) and *O. glaberrima* acc. IRGC102206 (R).

Generation	Total no. of plants analyzed	No. of resistant plants	No. of susceptible plants	Expected segregation ratio	χ^2	*p*-value
BC_1F_1	69	39	30	1:1	1.16	0.28^{ns}
BC_2F_1	276	154	122	1:1	3.7	0.05^{ns}
BC_2F_2	231	175	56	3:1	0.07	0.79^{ns}

FIGURE 2 | Response of *O. glaberrima* acc. IRGC102206 (R), PR121 (S), and their derived BC_1F_1 progenies upon *M. graminicola* infestation.

TABLE 2 | Means of growth parameters, gall number, and gall index of parents and BC_1F_1 population derived from the cross of PR121 (S) and *O. glaberrima* acc. IRGC102206 (R) in nematode-infested conditions.

Parents/ population	Plant height (cm)	Root length (cm)	Fresh shoot weight (g)	Fresh root weight (g)	Dry shoot weight (g)	Dry root weight (g)	Gall number	Gall index
PR121	57.83 ± 0.1	11.67 ± 1.4	58.36 ± 1.9	30.95 ± 0.4	8.09 ± 1.5	4.70 ± 0.6	73.66 ± 5.23	4.00 ± 0.0
IRGC102206	101.16 ± 3.1	28.67 ± 0.5	112.93 ± 2.3	45.03 ± 0.4	20.73 ± 0.3	11.81 ± 0.3	0.67 ± 0.3	1.00 ± 0.0
BC_1F_1	87.00 ± 9.2 (61.0–118.0)[b]	14.30 ± 1.7 (7.57–21.03)	110.60 ± 23.1 (14.89–188.48)	41.50 ± 12.9 (6.01–84.42)	18.60 ± 5.8 (3.99–38.28)	8.40 ± 2.5 (2.37–17.81)	23.20 ± 10.8 (1.00–70.33)	3.00 ± 0.5 (1.33–4.00)
CV[a]	7.54	15.26	11.39	16.63	21.68	31.59	29.49	14.23
F value	15.99[c]	3.85[c]	28.65[c]	30.16[c]	16.81[c]	6.64[c]	20.86[c]	13.67[c]

[a]*CV, coefficient of variance.*
[b]*Value in parentheses indicates the range of a trait in BC_1F_1 population.*
[c]*Indicates significant level at* $p < 0.01$.

of 1,901.21 cM, with an average distance of 22.63 cM. One putative QTL associated with gall numbers was detected on chromosome 6 designated as *qGN6.1*, between the marker interval of RM3183 and RM27001 explaining the phenotypic variance of 41% at an LOD score of 3.95 (**Table 3**). Two QTLs for dry root weight and dry shoot weight (*qDRW3.1* and *qDRW3.2*) were co-localized on chromosome 3 whereas QTL for fresh root weight, *qFRW8.1*, was located on chromosome 8 with 17% of total phenotypic variance (**Table 3**).

Whole-Genome Re-Sequencing of Bulked Segregant Analysis Pools

Based on the gall number of BC_1F_1 population, 10 plants from each of the extreme values of frequency distribution were selected and pooled as resistant bulk (RB) and susceptible bulk (SB), respectively. The whole-genome re-sequencing data from PR121, IRGC102206, resistant, and susceptible bulks were aligned with the *O. sativa* cultivar 93-11 reference genome using Bowtie 2 algorithm with default parameters. A total of 35.0, 33.6, 34.3, and 34.0 million

TABLE 3 | Chromosomal locations and parameters associated with the quantitative trait loci (QTL) for resistance to rice root-knot nematode in BC_1F_1 population derived from the cross of PR121 (S) and *O. glaberrima* acc. IRGC102206 (R).

Trait	QTLs[a]	Flanking markers	Physical position (Mb)	LOD score[b]	PVE[c] (%)	AE[d]
Gall number	*qGN6.1*	RM3183-RM20071	12.29–16.36	3.95	41.90	–23.62
Fresh root weight	*qFRW8.1*	RM23174-RM210	21.07–22.46	3.4	17.43	35.70
Dry root weight	*qDRW3.1*	RM3204-RM15281	14.82–18.48	3.1	22.67	–5.60
	qDRW3.2	RM5626-RM168	24.67–27.89	4.8	25.40	6.29
Dry shoot weight	*qDSW3.1*	RM3204-RM15281	14.82–18.48	2.6	18.71	–11.46
	qDSW3.2	RM5626-RM168	24.67–27.89	4.1	21.72	11.98

[a]*Putative QTLs are designated by the corresponding chromosome in which they are found. The method described by McCouch et al. (1997) was followed for QTL nomenclature.*
[b]*The maximum LOD score associated with each QTL.*
[c]R^2 *estimates the proportion of phenotypic variance (%) explained by the detected QTL.*
[d]*The additive genetic effect of the putative QTL. A negative number indicates that the alleles for resistance are derived from the male donor parent (*O. glaberrima *acc IRGC102206) and a positive number means that the alleles are contributed by the female parent (PR121).*

TABLE 4 | Statistical summary of BSA-seq data of parental lines, resistant bulk, and susceptible bulk.

Parameters	PR121	*O. glaberrima* acc. IRGC102206	Resistant bulk	Susceptible bulk
Total sequenced reads	33,626,673	35,036,905	34,300,698	34,021,906
High-quality reads (Q30)	32,496,792 (96.63%)	33,982,343 (96.99%)	33,298,078 (97.07%)	33,082,740 (97.23%)
Low-quality reads	1,129,881 (3.37%)	1,054,562 (3.01%)	1,002,620 (2.93%)	939,166 (2.77%)
Uniquely mapped reads	31,233,034	29,789,156	31,077,973	31,052,534
Alignment rate (%)	96.12	87.71	93.35	93.89
Average depth	14.9	12.8	12.5	12.5

paired-end reads were generated from IRGC102206, PR121, RB, and SB, respectively (**Table 4**), and 32–34 million reads of paired-end sequences were retained which were equivalent to 12.8× to 14.9× coverage of the rice genome indicating the high quality of the sequencing data (**Supplementary Figure S1**). Approximately, 31.23 (96.11%), 29.78 (87.66%), 31.07 (93.33%), and 31.05 (93.86%) million read pairs of PR121, IRGC102206, RB, and SB were uniquely mapped to the reference genome, respectively (**Table 4**). The GC content of raw reads ranged from 42% to 43% for all the samples.

A total of 3,692,066 variants (SNPs/indels) were identified by a joint variant calling with the reference genome through the GATK pipeline. RB had the highest variant (1,641,906) followed by IRGC102206 (1,049,225), PR121 (695,856), and SB (305,079). Maximum SNPs (284,012) were specific to chromosome 1 while chromosome 11 had the least SNPs (123,492). Likewise, the variant rate (SNPs/Mb) also varied among chromosomes, and chromosome 6 possessed the highest variant rate (1 SNP/195 bases) while the lowest variant rate (1 SNP/136 bases) was detected for chromosome 7 (**Supplementary Table S2**). A total of 1,416,115 SNPs were detected between RB and SB. Based on reference allele frequency (0.20) and maximum total depth (300), 174,651 SNPs were used for QTL identification through BSA-seq analysis.

BSA-seq Identifies Root-Knot Nematode Resistance Major Locus on Chromosome 6

The BSA-seq analysis detected nine putative QTL regions on chromosomes 2, 3, 4, 6, 11, and 12, based on the calculation of G′ values of SNP within a window size of 1.0-Mb genomic region across the entire length of chromosomes. QTLs on chromosomes 3, 6, and 12 with large G′ peaks above FDR(q) of 0.01 were considered as major QTL regions responsible for root-knot nematode resistance in rice. Despite the major G′ peaks, five minor G′ peaks at the margin of significance have been identified on chromosomes 2, 3 (left to *qNR*3.1), 4, and 12 (on both sides of *qNR12.1*). Two adjacent QTLs on chromosome 3, namely, *qNR3.1* and *qNR3.2*, had peaked in genomic intervals of 22.00–23.3 Mb and 24.0–25.6 Mb, respectively (**Figure 3B**). Similarly, QTLs on chromosome 12, designated as *qNR12.1* and *qNR12.2*, were located at 5 Mb distance between intervals of 10.5–12.7 Mb and 17.5–22.9 Mb, respectively. *qNR12.2* was present at the proximal region on the long arm of chromosome 12 displaying two adjacent component peaks (**Figure 3F**). The region covered by significant QTLs varied from 0.2 Mb (*qNR4.1*) to 5.4 Mb (*qNR2.2*). The direction of the ΔSNP index value and the allele frequency difference (AFD) indicated that *qNR2.1*, *qNR4.1*, *qNR6.1*, and *qNR6.2* originated from the donor parent IRGC102206 (**Table 5**). However, QTLs on chromosomes 3, 11, and 12 originated from a susceptible parent, PR121. The number of SNPs present in *qNR3.1, qNR3.2, qNR11.1*, *qNR12.1*, and *qNR12.2* region was 192, 330, 328, 691, and 322, respectively.

The genomic interval between *qNR6.1 and qNR6.2* was 2.4 Mb (11.2–13.6 Mb) and 0.6 Mb (15.4–16.0 Mb), respectively. The main G′ peak exhibited a subpeak in the QTL, *qNR6.1*, region whereas *qNR6.2* exhibited a clear sharp peak region and fall near the centromere region (ranging

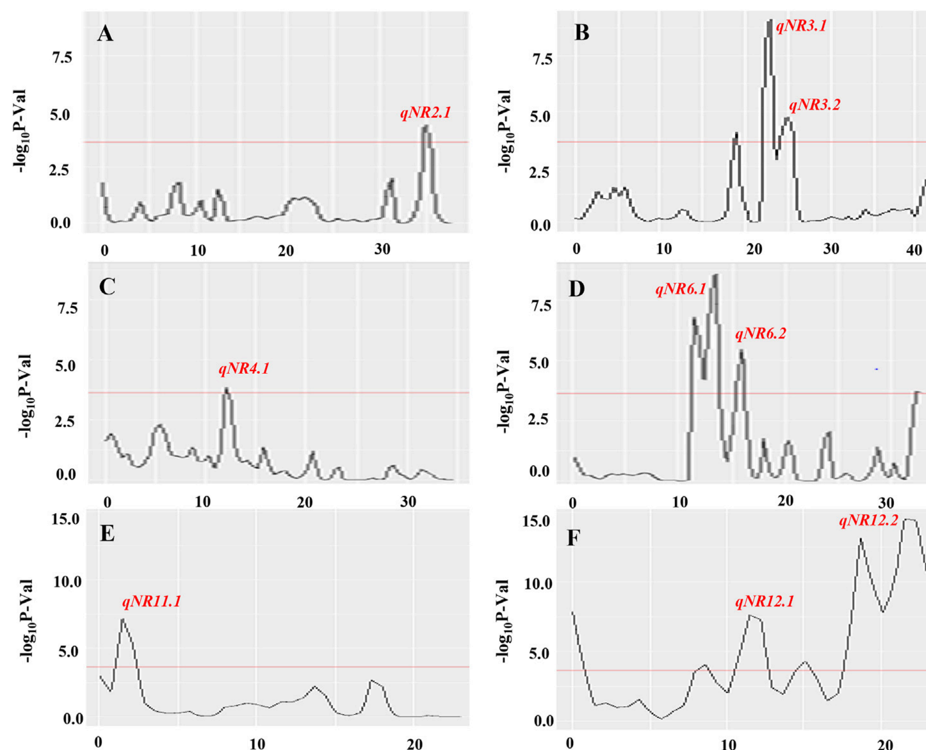


FIGURE 3 | Quantitative trait loci for root-knot nematode resistance identified on chromosomes 2 **(A)**, 3 **(B)**, 4 **(C)**, 6 **(D)**, 11 **(E)**, and 12 **(F)** using BSA-seq analysis. Distribution of –log10 *p*-value was calculated within a 1.0-Mb sliding window using tricube-smoothed kernel. The *Y*-axis represents – $\log_{10}$ *p*-values and the *X*-axis represents the position of chromosomes in Mb. The red line represents the significance threshold for FDR = 0.01, and the genomic region where –log10 *p*-value crosses the threshold was considered as significant QTL. Out of 12 chromosomes, significant QTLs identified on six chromosomes are shown.

TABLE 5 | List of QTLs identified through BSA-seq for rice root-knot nematode (*M. graminicola*) resistance.

Chromosome	QTL[a]	Start (Mb)	End (Mb)	Interval (Mb)	*p*-value	AFD[b]
2	*qNR2.1*	34.82	35.52	0.70	9.89E-05	0.29
3	*qNR3.1*	22.04	23.37	1.33	1.22E-05	–0.37
3	*qNR3.2*	24.05	25.61	1.56	6.11E-05	–0.29
4	*qNR4.1*	11.90	12.14	0.24	0.008847	0.33
6	*qNR6.1*	11.28	13.68	2.40	7.60E-06	0.18
6	*qNR6.2*	15.48	16.09	0.61	2.61E-05	0.74
11	*qNR11.1*	1.01	2.38	1.36	5.33E-06	–0.52
12	*qNR12.1*	10.52	12.73	2.20	1.92E-05	–0.20
12	*qNR12.2*	17.55	22.99	5.43	5.80E-06	–0.69

[a]*Putative QTLs are designated by the corresponding chromosome in which they are found. The method described by McCouch et al. (1997) was followed for QTL nomenclature.*
[b]*Allele frequency difference.*

approximately 17.00–22.59 Mb) of the short arm of chromosome 6. These two QTLs must be in continuity suggesting nematode resistance locus spanned within 3.0-Mb regions indicating a significant association for nematode resistance and hence considered as the major locus. The peaks of *qNR2*.1 and *qNR4.1* were at the margin of significance, that is, *qNR2.1* peak slightly exceeding and *qNR4.1* nearly reaching the threshold; and, therefore, referred to as minor QTLs. The number of SNPs present in *qNR2.1, qNR4.1, qNR6.1*, and *qNR6.2* regions was 345, 103, 264, and 27, respectively. Among all QTLs, *qNR6.1* and *qNR12.2* showed the highest G′ peaks that greatly exceeded the threshold level indicating the candidate regions for root-knot nematode resistance and susceptibility, respectively. Also, the genomic interval for root-knot nematode resistance on chromosome 6 overlapped with the identified gall number QTL (*qGN6.1*) through SSR markers. Therefore, *qNR6.1* and *qNR6*.2 were selected as promising QTL regions originating from the donor parent for the identification of candidate genes and their sequence variations allied with root-knot nematode resistance.

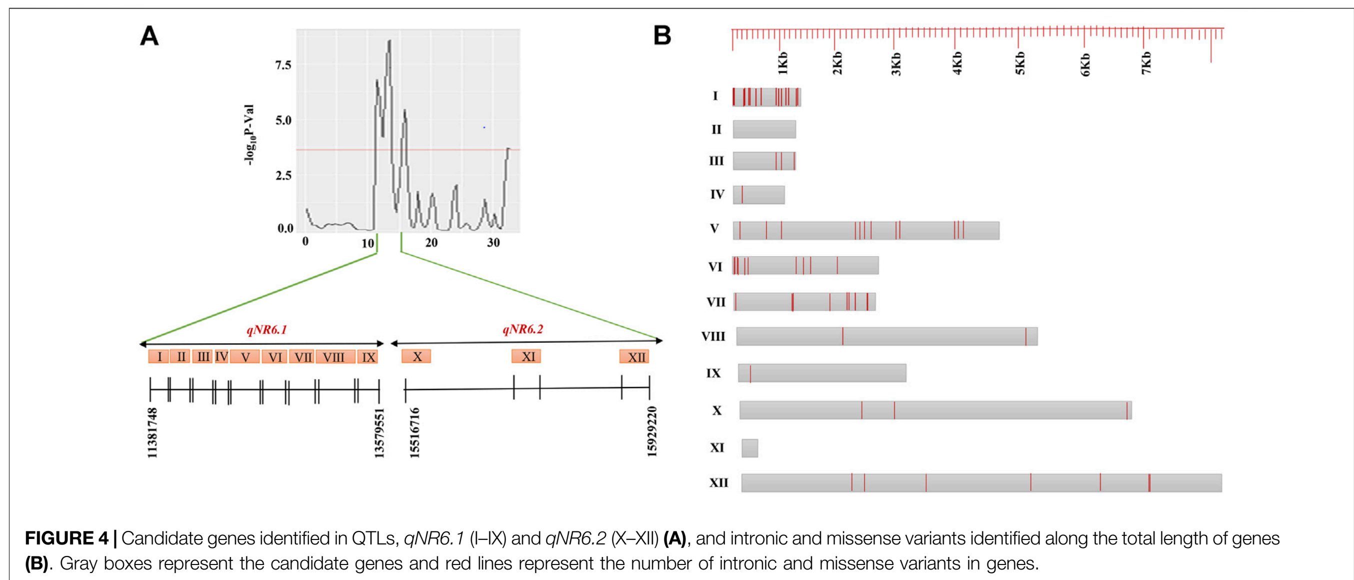


FIGURE 4 | Candidate genes identified in QTLs, *qNR6.1* (I–IX) and *qNR6.2* (X–XII) **(A)**, and intronic and missense variants identified along the total length of genes **(B)**. Gray boxes represent the candidate genes and red lines represent the number of intronic and missense variants in genes.

Candidate Genes Present Within the Identified QTLs

The candidate genes within the QTL regions were identified based on BGI annotation. The genomic interval of candidate genes and their annotation have been summarized in **Supplementary Table S3**. The total number of candidate genes annotated in QTLs, *qNR3.1*, *qNR3.2*, *qNR11.1*, *qNR12.1,* and *qNR12.2*, was 7, 10, 16, 10, and 66, respectively. A total of nine annotated genes were identified within a 2.4-Mb interval of *qNR6.1* and three genes within 0.6-Mb interval of *qNR6.2*. The less number of annotated genes within the QTL (*qNR6.1* and *qNR6.2*) interval is because of low recombination rate in the centromere region. In the *qNR2.1* region, seven genes were annotated whereas none of the candidate genes was annotated in the 240-Kb region of *qNR*4.1. We emphasized on those QTLs which originated from the donor parent to identify non-synonymous SNPs in the candidate genes. Among the nine annotated genes of *qNR6.1*, eight genes contained a total of 55 non-synonymous SNPs. In the candidate region of *qNR6.2*, a total of 10 non-synonymous SNPs were detected (**Figure 4**). The candidate gene encoding glycosyltransferase (BGIOSGA022720) had a maximum number of non-synonymous substitutions by SNPs (15), while peroxidase (BGIOSGA022766) and thioredoxin reductase (BGIOSGA021328) had the least number of non-synonymous SNPs (1). The examination of sequence reads of peroxidase and thioredoxin reductase on Integrative Genome Viewer (IGV) showed that non-synonymous SNPs were specific to the resistant parent; however, the susceptible parent also shared identical alleles to the reference genome (**Figure 5**). The majority of SNPs have been identified in the coding regions of candidate genes; however, mitogen-activated protein kinase (BGIOSGA022897), fructose-6-P-1-phosphotransferase subunit alpha (BGIOSGA022810), peroxidase (BGIOSGA022766), and thioredoxin reductase (BGIOSGA021328) had intronic SNPs (**Table 6**). Comparison of amino acid sequences of glycosyltransferase (BGIOSGA021421) and auxin-responsive protein (BGIOSGA022887) revealed that none of the SNPs caused non-synonymous substitutions.

In the *qNR2.1* region, a total of 85 SNPs were identified within seven candidate genes and only 16 of them had non-synonymous substitution. Among these genes, the miR164 gene (ENSRNA049493468) had no SNP, and none of the SNPs in growth-regulating factor 1 (BGIOSGA005497) and hexosyl transferase (BGIOSGA005471) caused non-synonymous substitutions. Pectin esterase (BGIOSGA005479) genes had 36 synonymous SNPs and 8 non-synonymous SNPs. The maximum non-synonymous SNPs were present in pectinesterase (BGIOSGA005479) followed by a UMP-CMP kinase gene (BGIOSGA009139) among the candidate genes of *qNR2.1* (**Table 6**). Therefore, candidate genes with non-synonymous SNPs could be the preferred genes for further studies. GO annotation showed the involvement of these genes in various biological processes and molecular functions. The candidate genes were predicted to be involved in different types of biosynthetic, metabolic, and catabolic processes (**Figure 6**). Transferase activity, protein metabolic process, and catabolic process appeared as the top enriched GO terms.

DISCUSSION

Genetics and Mapping of Root-Knot Nematode Resistance in *O. glaberrima*

M. graminicola is emerging as a devastating pest and affects both upland and irrigated rice causing up to 80% yield losses (Mantelin et al., 2017). It has been observed that intensely galled roots were shorter than the non-infested plants because terminal galls inhibited the roots from further elongation and prevented the plant's root system to absorb and translocate water (Kaur, 2020). Therefore, the severity of galling has a direct consequence on

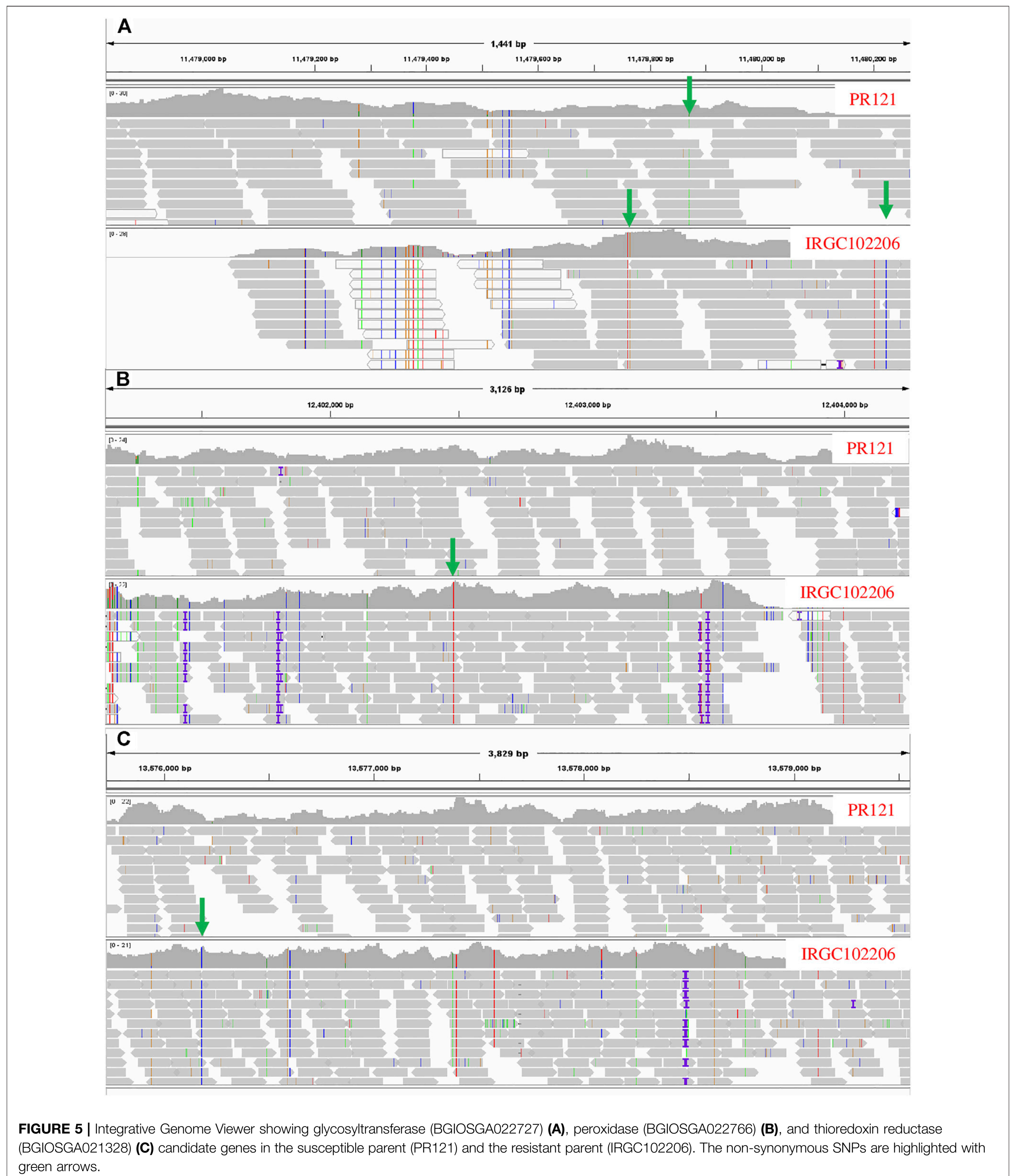


FIGURE 5 | Integrative Genome Viewer showing glycosyltransferase (BGIOSGA022727) **(A)**, peroxidase (BGIOSGA022766) **(B)**, and thioredoxin reductase (BGIOSGA021328) **(C)** candidate genes in the susceptible parent (PR121) and the resistant parent (IRGC102206). The non-synonymous SNPs are highlighted with green arrows.

plant yield. To alleviate the plant damage and yield losses, breeding for rice root-knot nematode resistance is one of the efficient and most economical strategies. With the advent of molecular markers, several nematode resistance genes and QTLs have been identified in different crop species such as tomato, potato, cotton, sugarbeet, and cowpea (Cai et al.,

TABLE 6 | Identification of non-synonymous SNPs in candidate genes present in putative genomic regions associated with rice root-knot nematode resistance.

QTL	Gene	SNP position (bp)	Reference allele	Alternate allele	Variant types	RB bulk variant rate	SB bulk variant rate
qNR2.1	BGIOSGA005507	34853274	G	C	Intronic	0.11	0.13
	BGIOSGA009139	35056763	G	A	Intronic	0.25	0.00
		35057030	T	C	Intronic	0.13	0.07
		35058165	C	A	Intronic	0.25	0.33
		35058435	T	C	Intronic	0.40	0.00
		35058636	T	C	Missense	0.25	0.25
		35058713	A	T	Intronic	0.42	0.41
		35058871	A	T	Intronic	0.30	0.07
	BGIOSGA005479	35345082	C	T	Missense	0.55	0.09
		35345096	G	T	Missense	0.47	0.10
		35345396	G	A	Intronic	0.47	0.40
		35345462	T	C	Missense	0.16	0.36
		35345484	C	G	Missense	0.27	0.47
		35346247	C	G	Missense	0.14	0.27
		35346258	T	G	Missense	0.30	0.65
		35346834	C	G	Missense	0.30	0.08
qNR6.1	BGIOSGA022720	11381794	C	G	Missense	1.00	0.87
		11381802	T	C	Missense	1.00	0.87
		11382081	T	C	Missense	0.37	0.80
		11382117	G	T	Missense	0.81	0.84
		11382178	A	G	Missense	0.75	0.86
		11382182	C	G	Missense	0.75	0.83
		11382328	C	T	Missense	0.41	1.00
		11382408	A	G	Missense	0.66	0.87
		11382674	A	G	Missense	0.50	1.00
		11382732	G	A	Missense	0.60	1.00
		11382765	C	A	Missense	0.18	0.00
		11382976	T	C	Missense	0.77	1.00
		11382982	T	C	Missense	0.77	1.00
		11383183	T	A	Missense	0.00	0.40
		11383201	T	C	Missense	0.88	0.66
	BGIOSGA022727	11479760	G	T	Missense	0.06	0.21
		11479871	G	A	Missense	0.75	0.80
		11480223	G	A	Missense	0.00	0.15
	BGIOSGA022766	12402483	G	T	Intronic	0.00	0.11
	BGIOSGA022770	12472994	A	C	Missense	0.20	0.37
		12473491	T	C	Missense	0.00	0.33
		12473749	C	T	Intronic	0.50	0.28
		12475176	C	G	Intronic	0.30	0.45
		12475239	G	T	Intronic	0.47	0.44
		12475312	G	A	Intronic	0.57	0.30
		12475415	G	A	Missense	0.52	0.70
		12475767	C	T	Missense	0.20	0.25
		12475832	G	C	Intronic	0.18	0.00
		12476695	C	T	Missense	0.33	0.50
		12476833	A	G	Missense	0.20	0.40
		12476962	A	G	Missense	0.15	0.20
	BGIOSGA022773	12525394	G	A	Missense	0.15	0.20
		12525408	A	G	Missense	0.20	0.23
		12525480	G	T	Missense	0.00	0.09
		12525513	C	A	Missense	0.10	0.08
		12525640	A	C	Intronic	0.10	0.11
		12525681	G	A	Intronic	0.00	0.25
		12526681	G	A	Missense	0.12	0.00
		12526824	C	T	Missense	0.20	0.00
		12526962	T	C	Missense	0.14	0.00
		12527407	C	T	Intronic	0.00	0.30
	BGIOSGA022777	12651941	C	T	Missense	0.14	0.00
		12653109	A	G	Intronic	0.00	0.07
		12653113	G	A	Intronic	0.00	0.07
		12653123	G	C	Intronic	0.00	0.07
		12653667	T	G	Intronic	0.12	0.50
		12653967	G	A	Missense	0.00	0.28
		12654024	A	G	Missense	0.12	0.20

(Continued on following page)

TABLE 6 | (*Continued*) Identification of non-synonymous SNPs in candidate genes present in putative genomic regions associated with rice root-knot nematode resistance.

QTL	Gene	SNP position (bp)	Reference allele	Alternate allele	Variant types	RB bulk variant rate	SB bulk variant rate
		12654207	G	T	Missense	0.11	0.07
		12654222	G	A	Missense	0.14	0.07
		12654401	T	C	Missense	0.75	0.93
		12654402	A	C	Missense	0.16	0.06
	BGIOSGA022810	13507912	C	G	Intronic	0.16	0.00
		13510886	G	A	Intronic	0.16	0.00
	BGIOSGA021328	13576180	A	C	Intronic	0.18	0.00
qNR6.2	BGIOSGA021280	15519303	T	A	Intronic	0.16	0.00
		15519833	A	G	Intronic	0.08	0.00
		15523461	T	G	Missense	0.14	0.00
	BGIOSGA022897	15923363	T	A	Intronic	0.12	0.00
		15923557	T	C	Intronic	0.12	0.00
		15924590	A	G	Intronic	0.12	0.00
		15924970	A	C	Intronic	0.11	0.00
		15927100	A	G	Intronic	0.10	0.00
		15928172	G	A	Intronic	0.09	0.00
		15928187	T	G	Intronic	0.18	0.00

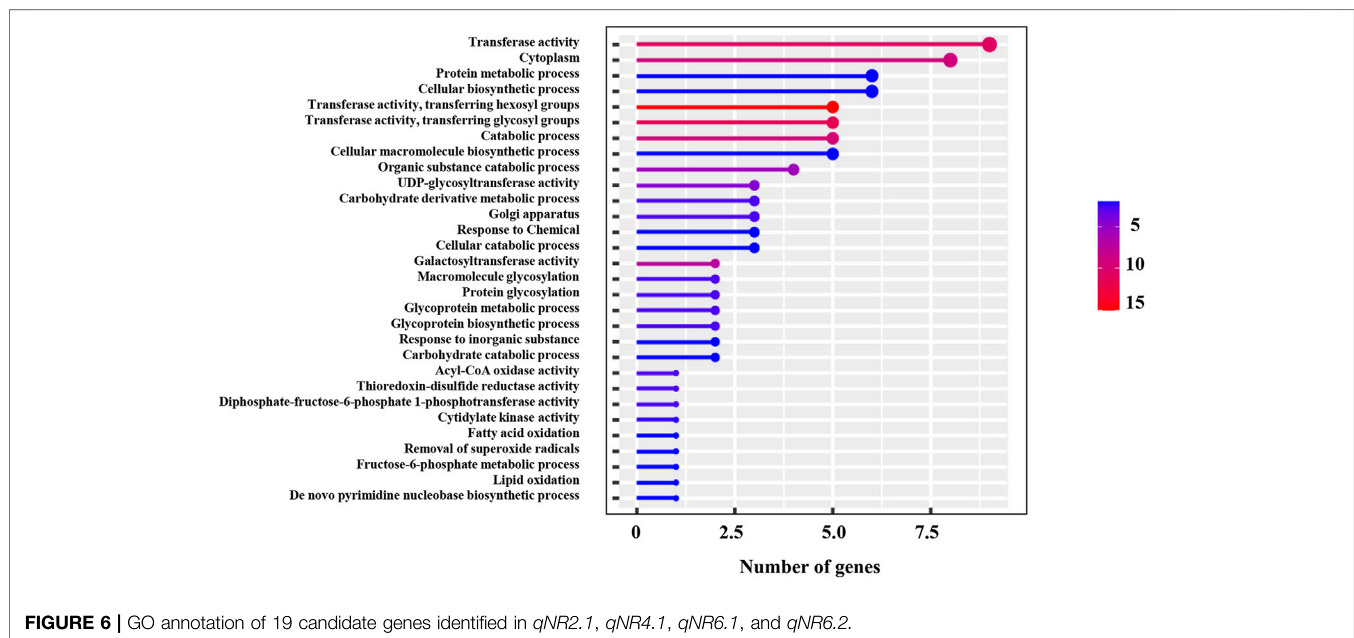


FIGURE 6 | GO annotation of 19 candidate genes identified in *qNR2.1*, *qNR4.1*, *qNR6.1*, and *qNR6.2*.

1997; Vander-Vossen et al., 2000; Wang et al., 2006; Wu et al., 2009; Ndeve et al., 2018).

In the past, most studies reported that rice root-knot nematode resistance was a quantitative trait (Amoussou et al., 2004; Prasad et al., 2006; Shrestha et al., 2007; Jena et al., 2013; Galeng-Lawilao et al., 2018). Here, we report that nematode resistance in *O. glaberrima* acc. IRGC102206 is under dominant monogenic control. Ideally, in BC_1F_1, 1:1 segregation is expected for heterozygous alleles and homozygous alleles of the recurrent parent. Due to the presence of heterozygous alleles, the trait under consideration should be moderately resistant. However, we got plants of varying resistances, from resistant to moderate resistant, suggesting that there might be non-allelic minor modifying genes that might act in an additive manner such that the phenotype is a resultant of the copy number of the alleles present in the genotype toward resistance. Similarly, the BC_2F_2 progenies showed wide variation for gall numbers indicating that there are a few modifying minor genes which affect the expression of a major gene to alter the resultant disease reaction. Previously, *O. glaberrima* variety CG14 has been identified as resistant to *M. graminicola* (Plowright et al., 1999; Soriano et al., 1999). But *O. glaberrima* has not been explored further to study the genetics of root-knot nematode resistance due to hybrid sterility in the interspecific crosses. However, introgression of *O. glaberrima* genes is possible by repeated backcrossing and doubled haploid breeding, even though there is a risk of losing the desirable traits from the parents (Jones et al., 1997). However, Neelam et al.

(2020) have successfully transferred the bacterial blight resistance gene *xa45(t)* into Pusa 44 background and mapped the gene on chromosome 8 from *O. glaberrima* acc. IRGC102600 by backcross breeding.

So far, no information is available for rice root-knot nematode resistance mapping in African rice cultivars. To the best of our knowledge, this is the first report on the identification of a major locus and a few minor QTLs for root-knot nematode resistance using interspecific BC_1F_1 population derived from a cross of *O. sativa* cv. PR121 and *O. glaberrima* acc. IRGC102206 by traditional QTL mapping and BSA-seq approach. Traditional QTL mapping identified one major large effect QTL for gall number, *qGN6.1*, explaining 41% of the phenotypic variance localized in the 12.29–16.36 Mb region with increasing alleles from the donor parent. This finding was consistent with the results of BSA-seq as the major locus for nematode resistance overlapped the genomic region between 11 and 16 Mb of chromosome 6, thereby validating the existence of a major QTL for root-knot nematode resistance in this region. We also detected QTLs on chromosomes 2 and 4 with minor effects using the BSA-seq approach. Based on the BC_1F_1 population, we anticipated that the SNP index of the RB had a mixture of PR121 and IRGC102206 alleles, and SB had PR121 alleles for the genes/QTLs associated with nematode resistance. The SNP index of the QTL regions on chromosomes 2, 4, and 6 agreed with this expectation, but the SNP index of the QTLs on chromosomes 1, 3, 11, and 12 displayed a contrasting blueprint. Thus, the present results inferred that QTLs on chromosomes 2, 4, and 6 probably conferred the nematode resistance, whereas QTLs on chromosomes 1, 3, 11, and 12 likely contributed to the susceptibility. Identification of the higher number of genomic regions contributed by PR121 is likely because we used a backcross population having a higher proportion of recurrent parent (PR121).

Previous studies identified QTLs for root galling, eggs per total root system and eggs per gram of roots, nematode reproduction, and nematode tolerance on chromosomes 1, 3, 4, 5, 7, 8, 9, 11, and 12 (Jena et al., 2013; Dimkpa et al., 2016; Galeng-Lawilao et al., 2018). To date, there is only one report inferring that a single dominant resistance gene (*Mg1(t)*) located on chromosome 10 in Asian rice cultivar Abhishek confers resistance against *M. graminicola* (Mhatre et al., 2017). So far, the *Hsa-1Og* nematode resistance gene has been identified from *O. glaberrima* TOG5681 on the long arm of chromosome 11 between markers RM206 and RM254 that confers resistance to the cyst nematode (Lorieux et al., 2003). Lahari et al. (2019) identified genomic region from 23 Mb to the bottom of rice chromosome 11 for root-knot nematode resistance through the QTL-seq approach, which might be co-localized with the results reported by Lorieux et al. (2003). Similarly, Dimkpa et al. (2016) also identified an SNP (id11008353) associated with gall number on the long arm of chromosome 11. Also, QTLs, *qNR12.1* and *qNR12.2*, on chromosome 12 have different genomic locations for gall number–associated SNPs (Dimkpa et al., 2016). However, the BSA-seq analysis in our study identified QTL for root-knot nematode resistance on the short arm of chromosome 11 (1.01–2.38 Mb) contributed by the susceptible parent. A root-knot nematode resistance locus from 23 Mbp to the bottom of rice chromosome 11 was identified by Dimkpa et al. (2016). This might be because different sources of root-knot nematode resistance were used to identify QTL regions, as the nature of gene actions varies with the background.

Shrestha et al. (2007) identified a QTL for the gall number on the long arm of chromosome 6 flanked by R2654 and RG778 markers, explaining 9.6% of the total phenotypic variation. The results indicated that QTL, *qGN6.1/qNR6.1-6.2*, and QTL reported by Shrestha et al. (2007) were present on either sides of the centromere of chromosome 6, suggesting that gall number QTL identified in our population was controlled by a different genetic locus. QTLs/linked markers for root-knot nematode resistance identified in previous studies have been summarized in **Supplementary Table S4**. In the present study, one QTL for fresh root weight and two QTLs for dry shoot weight and dry root weight were identified on chromosomes 8 and 3 (two QTLs), respectively. The QTLs for dry shoot weight and dry root weight were co-localized. Earlier, the QTL for fresh root weight was mapped on chromosomes 2 and 12 (Galeng-Lawilao et al., 2018). It has to be noteworthy that in the present study, genomic regions identified for different traits through either traditional mapping or BSA-seq are unique. The major QTL could be fine mapped to identify the putative candidate gene for nematode resistance.

Potential Roles of Candidate Genes in Root-Knot Nematode Resistance

Among the 19 candidate genes, 13 genes harbored non-synonymous mutations. Three glycosyltransferase genes were present in the *qNR*6.1 region, two of which (BGIOSGA022720 and BGIOSGA022727) had a total of 18 non-synonymous substitutions. Glycosyltransferases are involved in carbohydrate biosynthesis and cell-wall synthesis (Egelund et al., 2004). The upregulation of glycosyltransferase genes provided resistance against *Pseudomonas syringae* and *M. incognita and M. hapla* in *Arabidopsis* and tomato, respectively (Langlois-Meurinne et al., 2005; Schaff et al., 2007). However, its role in *M. graminicola* resistance is not understood currently in rice. In the present study, the peroxidase (BGIOSGA022766) gene located in *qNR6.1* plays an important role in ROS detoxification by regulating the H_2O_2 level and the oxidation of toxic reductants during pathogen attack. A higher level of peroxidase activity was reported in resistant genotypes of tomato during *M. incognita* infection (Chawla et al., 2013) and sweet potato (Yeon-Woo et al., 2019). Transcriptome profiling of *M. graminicola* infested rice roots showed the transcript abundance of peroxidases compared to uninfected roots (Kyndt et al., 2012). Rice root-knot nematode interactions revealed higher transcripts for nucleotide binding, catalytic, phosphatase, hydrolase, and ATPase activity after 3 and 7 days post inoculation (dpi) of *M. graminicola* (Haegeman et al., 2013).

Thioredoxin reductase has a role in ROS signaling and the protection of antioxidant enzymes to establish plant immunity during pathogen attacks. The ROS accumulation during pathogen attack results in oxidative modification of reactive free thiols in *S*-sulfenic acids (–SOH) of signaling proteins and antioxidant

enzymes such as catalase (CAT). Thioredoxin reductase reduces these thiol modifications to enable ROS signaling and protection of antioxidant enzyme activities (Mata-Perez and Spoel, 2019). Similar to thioredoxin reductase, two genes of the *qNR6.2* region, acyl-coenzyme A oxidase (BGIOSGA021280) and mitogen-activated protein kinase (BGIOSGA022897), might have a potential role in signaling pathways. Acyl-CoA oxidase metabolizes the pheromone (ascaroside) secreted by plant-parasitic nematode into chemical signals through the peroxisomal β-oxidation pathway that might act as a repellent for nematodes and thus reduces infection (Manohar et al., 2020). Mitogen-activated protein kinases (MAPKs) are intracellular signaling molecules that induce a defense response against root-knot nematode. The differential expression of the MAPK genes occurred after the activation of the effector-triggered immunity (ETI) pathway in rice upon *M. graminicola* infection as reported by Hatzade et al. (2020). It has been postulated that the genes involved in defense responses, phenylpropanoid, and hormone pathways were induced in response to *M. graminicola* infection in *O. glaberrima* line TOG5681, compared to Nipponbare (Petitot et al., 2017).

In the QTL region *qNR2.1*, the kinesin-like protein gene identified has diverse roles in biotic and abiotic stresses ranging from environmental to developmental processes like cell division, cell expansion, tropisms, and hormonal signaling. Kinesin modulates the cell wall structure and function by affecting the orientation and structure of cellulose microfibrils within the cell wall (Abdelkhalek et al., 2019). Therefore, kinesin-like protein and glycosyltransferase might be involved in the synthesis and maintenance of cell wall integrity during plant–nematode interactions. Growth-regulating factors are transcription factors regulated post-transcriptionally by miRNA396 (miR396) in different plant species (Omidbakhshfard et al., 2015). The growth-regulating factors, GRF1 and GRF3, in *Arabidopsis* target a plethora of genes having roles in programmed cell death, hormone signaling, and basal defense responses during biotic stresses (Piya et al., 2020). GRF1and GRF3 control the development and differentiation of syncytia during nematode (*Heterodera schachtii*) infection in *Arabidopsis*. In the present study, we also identified GRF1 (BGIOSGA009139) in the *qNR6.2* region. The presence of non-synonymous SNPs in these candidate genes suggests their potential role for root-knot nematode resistance in rice.

CONCLUSION

Understanding the genetic basis of root-knot nematode resistance and the development of functional markers to follow marker-assisted breeding is the prerequisite for nematode resistance breeding. In this study, we have identified a major QTL and two minor QTLs for *M. graminicola* resistance. Also, QTLs for fresh root weight, dry root weight, and dry shoot weight were detected on chromosomes 8 and 3, respectively. These results revealed the novel region of 11.03–16.25 Mb on chromosome 6 harboring *qNR6.1* and *qNR6.2* that may provide the basis for fine mapping and further exploration of novel genes associated with root-knot resistance in rice. A total of 10 and 3 candidate genes within *qNR6.1-qNR6.2* and *qNR2.1*, respectively, had non-synonymous SNPs that might be important for nematode resistance. The non-synonymous SNPs identified could be converted into KASP markers for effective deployment in rice breeding. The present results will increase our knowledge in understanding the molecular mechanism for root-knot nematode resistance.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

Conceptualization: YV, GM, and KS; methodology: GK, YV, and ND; performed the experiments: GK, KN and PP; analyzed the data: IY, GK, and DB; original draft: GK and YV; finalized the manuscript: GK, YV, and KN; funding: YV. All authors approved the final version.

FUNDING

This work was supported by the Department of Science and Technology, Ministry of Science and Technology, New Delhi, in the form of Extramural Research Grant (File No. EMR/2016/007539).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.871833/full#supplementary-material

Supplementary Figure S1 | Low quality reads **(A)** filtered through FastQC to obtain clean 856 reads **(B)** for reference alignment.

REFERENCES

Abad, P., Favery, B., Rosso, M.-N., and Castagnone-Sereno, P. (2003). Root-knot Nematode Parasitism and Host Response: Molecular Basis of a Sophisticated Interaction. *Mol. Plant Pathol.* 4 (4), 217–224. doi:10.1046/j.1364-3703.2003.00170.x

Abdelkhalek, A., Ismail, I. A., Dessoky, E. S., El-Hallous, E. I., and Hafez, E. (2019). A Tomato Kinesin-like Protein Is Associated with Tobacco Mosaic Virus Infection. *Biotechnol. Biotechnol. Equip.* 33, 1424–1433. doi:10.1080/13102818.2019.1673207

Amoussou, P. L., Ashurt, J., Bridge, J., Green, J., Jones, M., Koyama, M., et al. (2004). Broadly Based Resistance to Nematodes in the Rice and Potato Crops of Subsistence Farmers. *DFID Plant Sci. Res. Programme Annu. Rep.*, 9–14.

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30 (15), 2114–2120. doi:10.1093/bioinformatics/btu170

Bridge, J., Plowright, R. A., and Peng DeLiang, P. D. (2005). "Nematode Parasites of Rice," in *Plant–parasitic Nematodes in Subtropical and Tropical Agriculture*. Editors M. Luc, R. A. Sikora, and J. Bridge (Wallingford): CAB International), 87–130. doi:10.1079/9780851997278.0087

Cai, D., Kleine, M., Kifle, S., Harloff, H.-J., Sandal, N. N., and Marcker, K. A. (1997). Positional Cloning of a Gene for Nematode Resistance in Sugar Beet. *Science* 275, 832–834. doi:10.1126/science.275.5301.832

Chawla, N., Choudhary, K., Kaur, S., and Jindal, S. (2013). Changes in Antioxidative Enzymes in Resistant and Susceptible Genotypes of Tomato Infected with Root-Knot Nematode (*Meloidogyne incognita*). *Indian J. Nematol.* 43, 1–12.

Deokar, A., Sagi, M., Daba, K., and Taran, B. (2019). QTL Sequencing Strategy to Map Genomic Regions Associated with Resistance to Ascochyta Blight in Chickpea. *Plant Biotechnol. J.* 17 (1), 275–288. doi:10.1111/pbi.12964

Dimkpa, S. O., Lahari, Z., Shrestha, R., Douglas, A., Gheysen, G., and Price, A. H. (2016). A Genome-wide Association Study of a Global Rice Panel Reveals Resistance in *Oryza Sativa* to Root-Knot Nematodes. *J. Exp. Bot.* 67 (4), 1191–1200. doi:10.1093/jxb/erv470

Doyle, J. J., and Doyle, J. L. (1987). A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue. *Phytochem. Bull.* 19, 11–15.

Egelund, J., Skjot, M., Geshi, N., Ulvskov, P., and Petersen, B. L. (2004). A Complementary Bioinformatics Approach to Identify Potential Plant Cell Wall Glycosyltransferase-Encoding Genes. *Plant Physiol.* 136, 2609–2620. doi:10.1104/pp.104.042978

Galeng-Lawilao, J., Kumar, A., and De-Waele, D. (2018). QTL Mapping for Resistance to and Tolerance for the Rice Root-Knot Nematode, *Meloidogyne graminicola*. *BMC Genet.* 19 (1), 1–17. doi:10.1186/s12863-018-0656-1

Gaur, H. S., Singh, R. V., Kumar, S., Kumar, V., and Singh, J. V. (2001). *AICRP on Nematodes*. New Delhi: Division of Nematology. IARI.

Ge, X., Jung, D., and Yao, R. (2019). ShinyGO: a Graphical Enrichment Tool for Animals and Plants. *Bioinformatics* 36 (28), 2628–2629. doi:10.1093/bioinformatics/btz931

Haegeman, A., Bauters, L., Kyndt, T., Rahman, M. M., and Gheysen, G. (2013). Identification of Candidate Effector Genes in the Transcriptome of the Rice Root Knot Nematode *Meloidogyne graminicola*. *Mol. Plant Pathol.* 14 (4), 379–390. doi:10.1111/mpp.12014

Hatzade, B., Singh, D., Phani, V., Kumbhar, S., and Rao, U. (2020). Profiling of Defense Responsive Pathway Regulatory Genes in Asian Rice (*Oryza Sativa*) against Infection of *Meloidogyne graminicola* (Nematoda: Meloidogynidae). 3 *Biotech* 10 (2), 60. doi:10.1007/s13205-020-2055-3

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., et al. (2009). High-throughput Genotyping by Whole-Genome Resequencing. *Genome Res.* 19, 1068–1076. doi:10.1101/gr.089516.108

Illa-Berenguer, E., Van-Houten, J., Huang, Z., and Vander-Knaap, E. (2015). Rapid and Reliable Identification of Tomato Fruit Weight and Locule Number Loci by QTL-Seq. *Theor. Appl. Genet.* 128, 1329–1342. doi:10.1007/s00122-015-2509-x

Jena, M., Mohapatra, S. L., Panda, R. S., Mohanty, S. K., Thatoi, H. N., and Sahu, S. C. (2013). Genetic Loci Associated with Root-Knot Nematode Resistance in Rice Cv. Ramakrishna. *Oryza* 50 (2), 132–139.

Jones, M. P., Dingkuhn, M., Alukosnm, G. K., and Semon, M. (1997). Interspecific *Oryza Sativa* L. X *O. Glaberrima* Steud. Progenies in Upland Rice Improvement. *Euphytica* 94, 237–246. doi:10.1023/A:1002969932224

Kaur, G. (2020). *Mapping and Transfer of genes/QTLs for Nematode (*Meloidogyne graminicola*) Resistance from* Oryza Glaberrima *into* Oryza Sativa *L.* dissertation/PhD'sThesis (Ludhiana: Punjab Agricultural University).

Kumar, V. (2020). Data from: Social Innovation for Agricultural Development: A Study of System of Rice Intensification in Bihar, India. *Millenn. Asia* 11 (1), 99–118. doi:10.1177/0976399619900615

Kyndt, T., Nahar, K., Haegeman, A., De-Vleesschauwer, D., Hofte, M., and Gheysen, G. (2012). Comparing Systemic Defence-Related Gene Expression Changes upon Migratory and Sedentary Nematode Attack in Rice. *Plant Biol.* 1, 73–82. doi:10.1111/j.1438-8677.2011.00524.x

Lahari, Z., Ribeiro, A., Talukdar, P., Martin, B., Heidari, Z., Gheysen, G., et al. (2019). QTL-seq Reveals a Major Root-Knot Nematode Resistance Locus on Chromosome 11 in Rice (*Oryza Sativa* L.). *Euphytica* 215 (7), 117. doi:10.1007/s10681-019-2427-0

Langlois-Meurinne, M., Gachon, C. M., and Saindrenan, P. (2005). Pathogen-responsive Expression of Glycosyltransferase Genes UGT73B3 and UGT73B5 Is Necessary for Resistance to *Pseudomonas syringae* Pv. Tomato in Arabidopsis. *Plant Physiol.* 139 (4), 1890–1901. doi:10.1104/pp.105.067223

Langmead, B., and Salzberg, S. L. (2012). Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* 9 (4), 357–359. doi:10.1038/nmeth.1923

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Lorieux, M., Reversat, G., Garcia-Diaz, S. X., Denance, C., Jouvenet, N., Orieux, Y., et al. (2003). Linkage Mapping of *Hsa-1Og*, a Resistance Gene of African Rice to the Cyst Nematode, *Heterodera Sacchari*. *Theor. Appl. Genet.* 107, 691–696. doi:10.1007/s00122-003-1285-1

Magwene, P. M., Willis, J. H., and Kelly, J. K. (2011). The Statistics of Bulk Segregant Analysis Using Next Generation Sequencing. *PLoS Comput. Biol.* 7 (11), e1002255. doi:10.1371/journal.pcbi.1002255

Manohar, M., Tenjo-Castano, F., Chen, S., Zhang, Y. K., Kumari, A., Williamson, V. M., et al. (2020). Plant Metabolism of Nematode Pheromones Mediates Plant-Nematode Interactions. *Nat. Commun.* 11 (1), 208. doi:10.1038/s41467-019-14104-2

Mansfeld, B. N., and Grumet, R. (2018). QTLseqr: An R Package for Bulk Segregant Analysis with Next-Generation Sequencing. *Plant Genome* 11 (2). doi:10.3835/plantgenome2018.01.0006

Mantelin, S., Bellafiore, S., and Kyndt, T. (2017). *Meloidogyne graminicola*: a Major Threat to Rice Agriculture. *Mol. Plant Pathol.* 18 (1), 3–15. doi:10.1111/mpp.12394

Mata-Perez, C., and Spoel, S. (2019). Thioredoxin-mediated Redox Signaling in Plant Immunity. *Plant Sci.* 279, 27–33. doi:10.1016/j.plantsci.2018.05.001

McCouch, S. R., Chen, X., Panaud, O., Temnykh, S., Xu, Y., Cho, Y. G., et al. (1997). Microsatellite Marker Development, Mapping and Applications in Rice Genetics and Breeding. *Plant Mol. Biol.* 35, 89–99. doi:10.1007/978-94-011-5794-0_9

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20 (9), 1297–1303. doi:10.1101/gr.107524.110

Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: Integrated Software for Genetic Linkage Map Construction and Quantitative Trait Locus Mapping in Biparental Populations. *Crop J.* 3, 269–283. doi:10.1016/j.cj.2015.01.001

Mhatre, P. H., Pankaj, A. S., Singh, A. K., Ellur, R. K., and Kumar, P. (2017). Molecular Mapping of Rice Root-Knot Nematode (*Meloidogyne graminicola*) Resistance Gene in Asian Rice (*Oryza Sativa* L.) Using STMS Markers. *Indian J. Genet.* 77, 163–165. doi:10.5958/0975-6906.2017.00022.0

Michelmore, R. W., Paran, I., and Kesseli, R. V. (1991). Identification of Markers Linked to Disease-Resistance Genes by Bulked Segregant Analysis – a Rapid Method to Detect Markers in Specific Genomic Regions by Using Segregating Populations. *Proc. Natl. Acad. Sci. U.S.A.* 88, 9828–9832. doi:10.1073/pnas.88.21.9828

Ndeve, A. D., Matthews, W. C., Santos, J. R. P., Huynh, B. L., and Roberts, P. A. (2018). Broad-based Root-Knot Nematode Resistance Identified in Cowpea Gene-Pool Two. *J. Nematol.* 50, 545–558. doi:10.21307/jofnem-2018-046

Neelam, K., Mahajan, R., Gupta, V., Bhatia, D., Gill, B. K., Komal, R., et al. (2020). High-resolution Genetic Mapping of a Novel Bacterial Blight Resistance Gene Xa-45(t) Identified from *Oryza Glaberrima* and Transferred to *Oryza Sativa*. *Theor. Appl. Genet.* 133 (3), 689–705. doi:10.1007/s00122-019-03501-2

Nicol, J. M., Turner, S. J., Coyne, D. L., den-Nijs, L., Hockland, S., and Maafi, Z. T. (2011). "Current Nematode Threats to World Agriculture," in *Genomics and Molecular Genetics of Plant-Nematode Interactions*. Editors J. T. Jones, G. Gheysen, and C. Fenoll (Heidelberg): Springer), 21–43. doi:10.1007/978-94-007-0434-3_2

Omidbakhshfard, M. A., Proost, S., Fujikura, U., and Mueller-Roeber, B. (2015). Growth-Regulating Factors (GRFs): A Small Transcription Factor Family with Important Functions in Plant Biology. *Mol. Plant* 8 (7), 998–1010. doi:10.1016/j.molp.2015.01.013

Orjuela, J., Garavito, A., Bouniol, M., Moreno-Mayano, L. T., Kimball, J., Wilson, G., et al. (2010). A Universal Core Genetic Map for Rice. *Theor. Appl. Genet.* 120, 563–572. doi:10.1007/s00122-009-1176-1

Petitot, A. S., Kyndt, T., Haidar, R., Dereeper, A., Collin, M., de-Almeida, E. J., et al. (2017). Transcriptomic and Histological Responses of African Rice (*Oryza Glaberrima*) to *Meloidogyne graminicola* Provide New Insights into Root-Knot Nematode Resistance in Monocots. *Ann. Bot.* 119 (5), 885–899. doi:10.1093/aob/mcw256

Piya, S., Liu, J., Burch-Smith, T., Baum, T. J., and Hewezi, T. (2020). A Role for Arabidopsis Growth-Regulating Factors 1 and 3 in Growth-Stress Antagonism. *J. Exp. Bot.* 71 (4), 1402–1417. doi:10.1093/jxb/erz502

Plowright, R. A., Coyne, D. L., Nash, P., and Jones, M. P. (1999). Resistance to the Rice Nematodes *Heterodera Sacchari, Meloidogyne graminicola* and *M. incognita* in *Oryza Glaberrima* and *O. Glaberrima* X *O. Sativa* Interspecific Hybrids. *Nematology* 1 (7), 745–751. doi:10.1163/156854199508775

Prasad, P. V. V., Boote, K. J., Allen-, L. H., Jr, Sheehy, J. E., and Thomas, J. M. G. (2006). Species, Ecotype and Cultivar Differences in Spikelet Fertility and Harvest Index of Rice in Response to High-Temperature Stress. *Field crops Res.* 95 (2-3), 398–411. doi:10.1016/j.fcr.2005.04.008

Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS One* 8 (6), e66428. doi:10.1371/journal.pone.0066428

Roberts, P. A. (1995). Conceptual and Practical Aspects of Variability in Root-Knot Nematodes Related to Host Plant Resistance. *Annu. Rev. Phytopathol.* 33 (1), 199–221. doi:10.1146/annurev.py.33.090195.001215

Schaff, J. E., Nielsen, D. M., Smith, C. P., Scholl, E. H., and Bird, D. M. (2007). Comprehensive Transcriptome Profiling in Tomato Reveals a Role for Glycosyltransferase in Mi-Mediated Nematode Resistance. *Plant Physiol.* 144 (2), 1079–1092. doi:10.1104/pp.106.090241

Shrestha, R., Uzzo, F., Wilson, M. J., and Price, A. H. (2007). Physiological and Genetic Mapping Study of Tolerance to Root-knot Nematode in Rice. *New Phytol.* 176 (3), 665–672. doi:10.1111/j.1469-8137.2007.02185.x

Singh, P. (2010). The Rice Root-Knot Nematode, *Meloidogyne graminicola*: An Emerging Problem in Rice-Wheat Cropping System. *Indian J. Nematol.* 40, 1–11.

Soriano, I. R., Schmit, V., Brar, D. S., Prot, J. C., and Reversat, G. (1999). Resistance to Rice Root-Knot Nematode *Meloidogyne graminicola* Identified in *Oryza Longistaminata* and. *O. Glaberrima. Nematol.* 1 (4), 395–398. doi:10.1163/156854199508397

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013). QTL-seq: Rapid Mapping of Quantitative Trait Loci in Rice by Whole-Genome Resequencing of DNA from Two Bulked Populations. *Plant J.* 74 (1), 174–183. doi:10.1111/tpj.12105

Vander-Vossen, E. A. G., Rouppe vander Voort, J. N. A. M., Kanyuka, K., Bendahmane, A., Sandbrink, H., Baulcombe, D. C., et al. (2000). Homologues of a Single Resistance-Gene Cluster in Potato Confer Resistance to Distinct Pathogens: a Virus and a Nematode. *Plant J.* 23, 567–576. doi:10.1046/j.1365-313x.2000.00814.x

Venuprasad, R., Dalid, C. O., Del-Valle, M., Zhao, D., Espiritu, M., Sta-Cruz, M. T., et al. (2009). Identification and Characterization of Large-Effect Quantitative Trait Loci for Grain Yield under Lowland Drought Stress in Rice Using Bulk-Segregant Analysis. *Theor. Appl. Genet.* 120 (1), 177–190. doi:10.1007/s00122-009-1168-1

Wang, C., Ulloa, M., and Roberts, P. A. (2006). Identification and Mapping of Microsatellite Markers Linked to a Root-Knot Nematode Resistance Gene (*Rkn1*) in AcalaNem X Cotton (*Gossypium Hirsutum* L.). *Theor. Appl. Genet.* 112, 770–777. doi:10.1007/s00122-005-0183-0

Wang, S., Basten, C. J., and Zeng, Z. B. (2012). *Windows QTL Cartographer 2.5*. Raleigh, NC, USA: Department of Statistics, North Carolina State University. Available online at: http://statgen.ncsu.edu/qtlcart/WQTLCart.htm.

Williamson, V. M. (1998). Root-knot Nematode Resistance Genes in Tomato and Their Potential for Future Use. *Annu. Rev. Phytopathol.* 36, 277–293. doi:10.1146/annurev.phyto.36.1.277

Wu, W. W., Shen, H. L., and Yang, W. C. (2009). Sources for Heat Stable Resistance to Southern Root-Knot Nematode (*Meloidogyne incognita*) in *Solanum lycopersicum*. *Agric. Sci. China* 8, 697–702. doi:10.1016/S1671-2927(08)60267-9

Yeon-Woo, S., Lee, I., Shim, D., Lee, K. L., Nam, K., Yang, J. W., et al. (2019). Transcriptomic Changes in Sweet Potato Peroxidases in Response to Infection with the Root-Knot Nematode *Meloidogyne incognita*. *Mol. Biol. Rep.* 46 (4), 4555–4564. doi:10.1007/s11033-019-04911-7

Zhang, X., Zhang, K., Wu, J., Guo, N., Liang, J., Wang, X., et al. (2020). QTL-seq and Sequence Assembly Rapidly Mapped the Gene *BrMYBL2.1* for the Purple Trait in *Brassica Rapa*. *Sci. Rep.* 10, 2328. doi:10.1038/s41598-020-58916-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Kaur, Yadav, Bhatia, Vikal, Neelam, Dhillon, Praba, Mangat and Singh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

REVIEW
published: 14 July 2022
doi: 10.3389/fgene.2022.880195

Advances in Crop Breeding Through Precision Genome Editing

Gauri Nerkar[1], Suman Devarumath[2], Madhavi Purankar[1], Atul Kumar[1], R. Valarmathi[3], Rachayya Devarumath[1] and C. Appunu[3]**

[1]*Molecular Biology and Genetic Engineering Laboratory, Vasantdada Sugar Institute, Pune, India,* [2]*Vidya Pratishthan's College of Agricultural Biotechnology, Baramati, India,* [3]*ICAR-Sugarcane Breeding Institute, Coimbatore, India*

The global climate change and unfavourable abiotic and biotic factors are limiting agricultural productivity and therefore intensifying the challenges for crop scientists to meet the rising demand for global food supply. The introduction of applied genetics to agriculture through plant breeding facilitated the development of hybrid varieties with improved crop productivity. However, the development of new varieties with the existing gene pools poses a challenge for crop breeders. Genetic engineering holds the potential to broaden genetic diversity by the introduction of new genes into crops. But the random insertion of foreign DNA into the plant's nuclear genome often leads to transgene silencing. Recent advances in the field of plant breeding include the development of a new breeding technique called genome editing. Genome editing technologies have emerged as powerful tools to precisely modify the crop genomes at specific sites in the genome, which has been the longstanding goal of plant breeders. The precise modification of the target genome, the absence of foreign DNA in the genome-edited plants, and the faster and cheaper method of genome modification are the remarkable features of the genome-editing technology that have resulted in its widespread application in crop breeding in less than a decade. This review focuses on the advances in crop breeding through precision genome editing. This review includes: an overview of the different breeding approaches for crop improvement; genome editing tools and their mechanism of action and application of the most widely used genome editing technology, CRISPR/Cas9, for crop improvement especially for agronomic traits such as disease resistance, abiotic stress tolerance, herbicide tolerance, yield and quality improvement, reduction of anti-nutrients, and improved shelf life; and an update on the regulatory approval of the genome-edited crops. This review also throws a light on development of high-yielding climate-resilient crops through precision genome editing.

Keywords: Genome editing, crop breeding, new breeding techniques, CRISPR, disease resistance, abiotic stress tolerance, biofortification, climate-resilient crops

OPEN ACCESS

Edited by:
Ravi Valluru,
University of Lincoln, United Kingdom

Reviewed by:
Kalenahalli Yogendra,
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India
Sajid Fiaz,
The University of Haripur, Pakistan

****Correspondence:***
Rachayya Devarumath
rdevarumath@gmail.com
C. Appunu
cappunu@gmail.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *21 February 2022*
Accepted: *02 May 2022*
Published: *14 July 2022*

Citation:
Nerkar G, Devarumath S, Purankar M, Kumar A, Valarmathi R, Devarumath R and Appunu C (2022) Advances in Crop Breeding Through Precision Genome Editing. Front. Genet. 13:880195. doi: 10.3389/fgene.2022.880195

INTRODUCTION

The global climate change has a direct impact on the food security, agriculture, crop production and plant health (Tirado et al., 2010). According to the world population data sheet 2020, the world population is projected to increase from 7.8 billion in 2020 to 9.9 billion by 2050. Global cropland area per capita has decreased continuously from about 0.45 ha per capita in 1961 to 0.21 ha per capita in 2016 (https://www.fao.org). Further, the available area for cultivation is degraded due to various

TABLE 1 | Global production of top 10 crops essential for food security ranked based on their global harvested area.

Rank based on Harvested area	Crop	Harvested area (million hectares)		Production (million metric tons)		Yield (metric tons per hectare)	
		2016–17	2021–22	2016–17	2021–22	2016–17	2021–22
1	Wheat	222.11	222.11	753.09	778.6	3.39	3.51
2	Maize	183.06	203.89	1,065.11	1,210.45	5.82	5.94
3	Rice	161.48	166.47	481.54	513.03	4.45	4.60
4	Soybean	121.11	130.10	348.04	350.72	2.87	2.70
5	Barley	48.21	48.48	147.04	145.10	3.05	2.99
6	Sorghum	41.82	41.75	63.18	65.59	1.51	1.57
7	Rapeseed	33.65	37.73	68.86	71.18	2.05	1.89
8	Cottonseed	28.62	31.52	38.70	43.47	1.35	1.38
9	Cotton	29.44	32.07	105.88	120.20	783	816
10	Peanut	26.15	29.65	42.34	50.60	1.66	1.71

factors such as loss of agricultural land to non-agricultural uses, intensive use of land for cultivation through multiple cropping, reduction in fallow periods, excessive use of agrochemicals, and spread of monocultures (Alexandratos and Bruinsma, 2012). These factors undermine the long-term productive potential of the available cultivable land. Thus, on one hand when the global demand for food is increasing, there is a decline in the availability of the cultivable land for the production of food crops. **Table 1** lists the important crops for food security and comparison of their global production during 2016–17 and 2021–22 (data accessed from: http://www.fao.org/faostat/en/#home). This data show that the harvested area for the most grown crops of the world has remained fairly constant over the past 5 years and there is also a slight increase in the yield and production of these crops. Increasing the crop productivity in the face of the global climate change is therefore most important challenge in front of the crop breeders. The other important factors that limit the agricultural productivity are limited water availability and irrigation, declining soil fertility, untimely rainfall, high temperature, pests and pathogens (Singh et al., 2020).

The current efforts are focused on increasing the crop productivity without using pesticides and fertilizers. Conventional breeding programs are often laborious, time-consuming and difficult. Genetic engineering has greatly simplified the process of development of novel and improved varieties with better agronomic traits like disease resistance, abiotic stress tolerance, a better shelf life as well as improved crop productivity (Nerkar G. et al., 2018). The recent emergence of the novel plant breeding technologies like genome editing has opened up new doors for precise modification of the plant genomes without the introduction of foreign DNA (Altpeter et al., 2016). Their successful application in the development of elite germplasm, with high yield, quality, and resistance against biotic and abiotic stresses appears promising (Fiaz et al., 2020). Therefore, there is an urgent need to utilize these technologies for the development of novel and improved crop varieties to overcome the difficulties faced in the conventional breeding programs (Fiaz et al., 2021a).

Genome editing can facilitate genome modification by creating precise changes at specific sites of the genome and the reagents used in this process can be delivered into the cell without incorporating DNA into the genome (Zhang et al., 2016; Bandyopadhyay et al., 2020; Ma et al., 2020). The mutations resulting from genome editing are similar to those occurring in nature which potentially simplifies their regulation, unlike the traditional GMO crops. Another remarkable feature of this technology is that it creates inheritable mutations in the genome with a low probability of generating off-targets. Genome editing creates DNA modifications such as deletions, insertions, single nucleotide substitution (SNPs), and large fragment substitution. The site-specific nucleases (SSNs) that bring about nucleotide excision are: engineered homing endonucleases or mega-nucleases (MNs) (Smith et al., 2006), Zinc-Finger Nucleases (Kim et al., 1996), transcription activator like effector nucleases (TALENs) (Christian et al., 2010), and CRISPR-associated protein (Cas) (Jinek et al., 2012).

The site specific nucleases create double-stranded breaks (DSBs) in the genome. The era of precise genome editing in plants began with the discovery of I–SceI induced DSBs that enhanced homologous recombination in plants (Puchta et al., 1993). SSNs are programmed to recognize the preselected genomic sites and they make use of cellular DSB repair mechanisms such as non-homologous end joining (NHEJ) or homology-directed repair (HDR) (**Figure 1**). In NHEJ, a gene is rendered non-functional by random insertion or deletion of DNA at the cut site before reattaching of the free DNA ends (Puchta 2005). HDR involves addition of a Donor DNA of choice which is homologous to the site of the break. Cells use this as a patch to repair the DNA (Haber 2000) (**Figure 1**). Using the HDR pathway, scientists can introduce a new gene with vital function or correct a mutation by replacing the mutated sequence with a healthy sequence (Symington and Gautier 2011).

The global food security is largely affected by the changing climatic conditions, significant yield gaps between the actual yield and the potential yield, decrease in the number of farmers, lack of transportation infrastructure, post-harvest losses due to low shelf life of crops (Mackelprang and Lemaux, 2020; Fiaz and Wang, 2021c). Precision genome editing can help in addressing these problems by generating plants with sufficient yields in spite of changing climatic conditions. The crop varieties which remain underutilized due to low yields, high disease susceptibility can be made more resilient using genome editing. They can make specific plants a source of essential nutrients that are lacking in the diets of some populations.

Crop breeding has been greatly accelerated after the introduction of genome editing tools. Therefore, there is an urgent need to review the advances in crop breeding through precision genome editing. Here, we provide an overview of the different breeding approaches for crop improvement; genome editing tools and approaches used for crop improvement; agronomic traits such as disease resistance, abiotic stress tolerance, herbicide tolerance, yield and quality improvement, reduction of anti-nutrients, improved shelf life and an update on the regulatory approval of the genome-edited crops.

MECHANISM OF ACTION OF DIFFERENT GENOME EDITING TECHNOLOGIES USED FOR CROP IMPROVEMENT

Meganucleases have the target sites of up to 18 bp (Smith et al., 2006). ZFNs have a non-specific *Fok*I endonuclease domain combining with multiple zinc-finger DNA-binding domains that recognize a 3 bp module (Kim et al., 1996). TALENs consist of a *Fok*I endonuclease domain which pairs with multiple transcription activator-like effector domains that recognize single base pairs (Christian et al., 2010). TALENs have been widely applied in genome editing of crops, owing to their higher target binding specificity and generation of lesser number of off-targets compared with ZFNs (Liu et al., 2021).

In bacteria and archaea, Clustered Regularly Interspaced Palindromic Repeats (CRISPR) in combination with Cas (CRISPR associated) proteins form an adaptive immune system (Mohanraju et al., 2016). CRISPR-Cas immune systems consist of three distinct stages *viz.* adaptation, during which the short DNA fragments (spacers) from invading viruses and plasmids are recognized and acquired, processed and integrated into the CRISPR locus (Jackson et al., 2017); transcription during which the transcription of CRISPR locus to a long pre-CRISPR RNA (pre-crRNA) and the maturation of pre-crRNA to crRNA (guide RNA) occur (Charpentier et al., 2015); and finally interference in which the complementary target DNA sequences are recognized by Cas effector nucleases using the guide RNA. Consequent to the recognition of the target DNA, Cas effectors bind to the target DNA and generate a double-stranded DNA break (DSB) (Jinek et al., 2012).

There are six primary types of the CRISPR-Cas systems. Types I, III, and IV are characterized by multi-subunit effector complexes, while types II, V, and VI consist of single-subunit effector (Koonin et al., 2017; Shmakov et al., 2017). The class 2, type II clustered regularly interspaced short palindromic repeat (CRISPR)/Cas9 (CRISPR-associated) system from *Streptococcus pyogenes* is the ground-breaking technology for genome editing discovered around a decade ago, which is based on RNA-guided engineered nucleases (Jinek et al., 2012). While Meganucleases, ZFNs and TALENs recognize their sequence targets through protein/DNA interactions, CRISPR/Cas9 achieves targeting through a guide RNA (sgRNA). sgRNAs are short nucleotide sequences (~20 nt) with a specific sequence that can target the genomic sequence of interest. The Cas9 nuclease then cleaves the resulting RNA/DNA complex. Consequently, a DSB is created at the target site containing a conserved protospacer adjacent motif (PAM). The repair occurs by NHEJ which creating indels in the protein-coding regions causing frameshift or knock-down of the desired genes (Basso et al., 2020). The simplicity of DNA targeting through base-pairing has led to the quick and broad adoption of CRIPSR/Cas9 reagents for genome editing (Ahmar et al., 2021; Liu et al., 2021). The most widely used Cas9 is derived from *Streptococcus pyogenes* (SpCas9) which requires a protospacer adjacent motif (PAM) sequence of "NGG" in the target DNA sequence. Other Cas9 variants differing in their PAM requirements (SpCas9-VQR-"NGA", SpCas9-EQR- "NGAG", Cas9-NG-"NG", and xCas9 3.7-"NG/GAA/GAT") have also been used for plant genome editing (Anzalone et al., 2019; Zhang et al., 2019; Nadakuduti and Enciso-Rodríguez, 2021).

Cas12 nucleases belonging to class 2, type-V CRISPR systems were later added to the CRISPR toolbox (Zetsche et al., 2015). The major differences between Cas9 and Cas12 nucleases are that Cas12 nucleases are mostly guided by a single crRNA of ~42 nt compared to the Cas9 guide RNA of ~100 nt; unlike Cas9, Cas12 effectors lack HNH domain and possess only RuvC-like domain that cleaves both strands of the DNA target site resulting into a staggered cut with a 4–5 nt 5′ overhang (Zetsche et al., 2015). LbCas12a is the most widely used Cas12 variant for gene-editing in plants and it recognizes a T-rich PAM "TTTV" (Zhang et al., 2019). Engineered variants of Cas12a with increased activities and target ranges have also been developed (Kleinstiver et al., 2015). A distinct feature of Cas12a is that it does not require tracrRNA for the processing of mature crRNA making it advantageous for multiplex gene editing, transcription, epigenetic modulations and base editing (Safari et al., 2019).

The most recent addition to the CRISPR toolbox is the CRISPR-Cas8, which is a hypercompact type-V CRISPR system consisting of a single Cas8 protein of ~70-kDa that is about half the size of Cas9 or Cas12a (Nadakuduti and Enciso-Rodríguez, 2021). Like Cas12a, Cas8 also does not require a tracrRNA and generates a staggered cut with 5′-overhangs (Pausch et al., 2020) and requires the PAM of 5′-TBN-3′ (where B = G, T, or C). CRISPR-Cas8 mediated genome editing has been reported in Arabidopsis with an editing efficiency of 0.85% (Pausch et al., 2020).

Recently, precise modification of DNA and RNA has also been reported at single-base level through base editing which can convert one target DNA nucleotide to another. Base editors (BEs) can precisely modify nuclear and organellar genomes (DNA BEs) as well as transcriptomes (RNA BEs) of dividing as well as non-dividing cells (Molla et al., 2021). BEs consist of a catalytically impaired Cas nuclease (dCas9: D10A and H840A) that is fused to a nucleotide deaminase and DNA repair proteins. Unlike the SpCas9-generated DSBs that are repaired by error prone NHEJ, the BE-generated individual nicks are repaired by a more precise base excision repair pathway (BER), thus, minimizing the undesired by-products due to gene-editing (Ran et al., 2013). DNA BEs can be classified as: cytosine BEs (CBEs), adenine BEs (ABEs), C-to-G BEs (CGBEs), dual-base editors and organellar BEs. Genome editing using CBEs has been reported in some of the major crops like Arabidopsis (Chen et al., 2017), rice (Shimatani et al., 2017), wheat (Zhang et al., 2019),

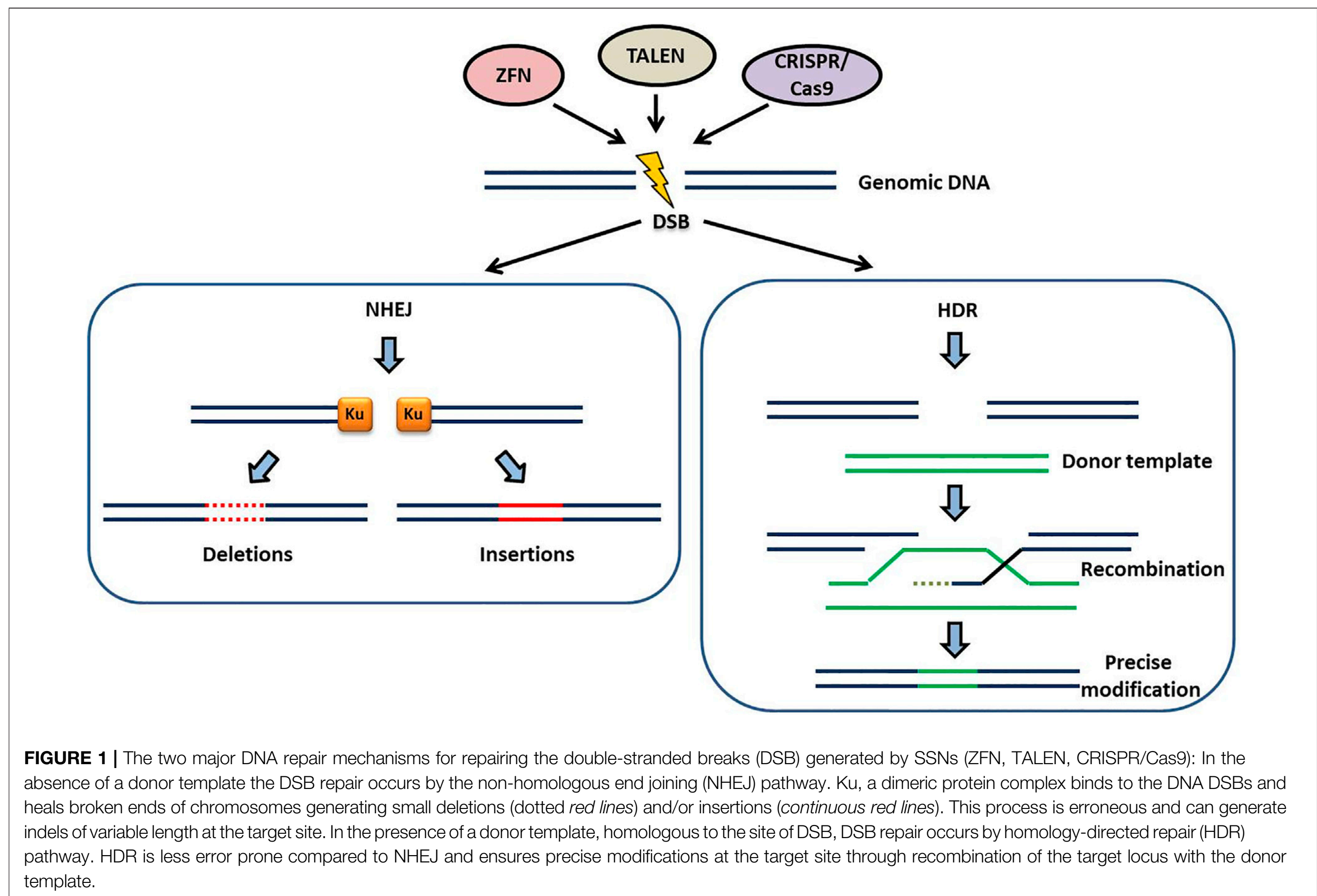


FIGURE 1 | The two major DNA repair mechanisms for repairing the double-stranded breaks (DSB) generated by SSNs (ZFN, TALEN, CRISPR/Cas9): In the absence of a donor template the DSB repair occurs by the non-homologous end joining (NHEJ) pathway. Ku, a dimeric protein complex binds to the DNA DSBs and heals broken ends of chromosomes generating small deletions (dotted *red lines*) and/or insertions (*continuous red lines*). This process is erroneous and can generate indels of variable length at the target site. In the presence of a donor template, homologous to the site of DSB, DSB repair occurs by homology-directed repair (HDR) pathway. HDR is less error prone compared to NHEJ and ensures precise modifications at the target site through recombination of the target locus with the donor template.

maize (Zong et al., 2017), tomato (Hunziker et al., 2020), potato (Veillet et al., 2019a), cotton (Qin et al., 2020), soybean (Cai et al., 2020), and rapeseed (Cheng et al., 2021). Although base editors can bring about base transitions without DNA donors, they cannot be used with other base transversions, insertions or deletions.

The 'search and replace' prime editing (PE) is the most recent and by far the most advanced tool for genome editing which can copy the desired edit incorporated within the guide RNA without using DSBs or donor repair template (Anzalone et al., 2019) and generate targeted insertions or deletions, or directly install precise transition and transversion mutations at targeted genomic loci.

Prime editing is mediated by a complex consisting of prime editing guide RNA (pegRNA) and a catalytically impaired Cas9 endonuclease [nCas9 (H840A)] that is fused with a reverse transcriptase from engineered moloney murine leukaemia virus (M-MLV RT). This complex binds to the regions of protospacer and PAM. nCas9 (H840A) nicks the edited sequence by RuvC-like domains at a position three base pairs upstream of PAM. Subsequently, as the primer binding site (PBS) matches to the exposed 3′ end of the edited sequence, reverse transcription is initiated whereby the editing information is transformed to the edited sequence from the RT template. A mismatch formed in the heteroduplex DNA which contains one edited and one unedited sequence is repaired by using the edited sequence as a template (Hao et al., 2021). Prime editing has been reported in cereal crops (Butt et al., 2020; Lin et al., 2020; Tang et al., 2020; Fiaz et al., 2021b) and has an advantage of fewer bystander mutations compared to base editing and also less restricted to PAM availability compared to the other genome editing methods (Anzalone et al., 2019). However, base editors still remain widely applicable due to their improved efficiency with superior on-target and off-target DNA editing profiles, product purity, and DNA specificity (Anzalone et al., 2019; Yu et al., 2020). Thus, the choice of suitable editing strategy largely depends on the specific application such as the desired edit, availability of PAMs, editing efficiency and generation of off-target/bystander mutations (Hao et al., 2021).

BREEDING APPROACHES FOR CROP IMPROVEMENT

The traditional breeding approaches have greatly contributed to genetic improvement of the present day elite crop varieties. The recent advances in the traditional breeding methods include wide crosses, introgression from wild-crop by hybrid breeding, mutation breeding, double haploid technology, tissue culture-based approaches like embryo and ovule rescue and protoplast fusion. Introgression through hybridization and back-crossing is

one of the most widely adopted methods for developing elite crop varieties with desirable traits. It is also an important breeding strategy for transferring desirable traits from wild species to cultivated varieties. Interspecific gene flow has contributed to the origin of crop plants, restoration of crop diversity after domestication and to the adaptation to challenging environments (Foria et al., 2022). Introgressive hybridization has been used to tap the secondary gene pool for accessing the genetic variation for crop improvement in wheat (Mujeeb-Kazi and Asiedu, 1990), rice (Jena, 2010), potato (Jansky and Oeloquin, 2006), tomato (Schouten et al., 2019), cassava (Wolfe et al., 2019). Wild-crop introgressive hybridization has been used for incorporating disease resistance traits into newly released crop varieties (Hajjar and Hodgkin., 2007). Another frequently used approach for crop breeding is inbreeding. Inbred varieties are produced by self-fertilization in order to preserve the original traits and to produce true breeding cultivars which can be as parents in the hybridization programmes. Inbreeding can be used to improve the results of selection when the heritability for a trait is small. In hybrid breeding, two different inbred varieties are crossed to produce an offspring with stable characteristics and hybrid vigor, where the offspring is much more productive than either parent (Caligari, 2001). Hybrid crop varieties perform better than their inbred progenitors in an array of crops like maize and oil palm (Labroo et al., 2021).

In mutation breeding, genotypes showing spontaneous mutations are selected for breeding or mutations are induced using physical or chemical mutagens to create mutant phenotypes with desired traits (Devarumath et al., 2015; Purankar et al., 2022). Marker assisted breeding makes use of molecular markers, in combination with linkage maps and genomics, to alter and improve the crop traits on the basis of genotypic assays (Jiang, 2013). The morphological (trait-specific), proteinaceous (isoenzyme), cytological (chromosome-specific), and DNA markers have been used in plant breeding. DNA markers have been extensively utilized for marker-assisted selection of crop plants (Madina et al., 2013; Kumawat et al., 2020). Recently, the advanced molecular breeding tools such as SSRs, Indels, SNPs, genome sequencing, genotype by sequencing, and microRNAs have been used for crop improvement (Devarumath et al., 2014; Platten et al., 2019; Bohar et al., 2020) to confer biotic and abiotic stress tolerance (Shriram et al., 2016; Devarumath et al., 2019).

Genetic engineering of plants commenced nearly three decades ago after the first successful regeneration of a transgenic plant was reported from transformed cells of tobacco (Barton and Brill, 1983) petunia, and tomato (Horsch et al., 1985). The invention of biolistic gene gun (Sanford et al., 1987; Klien et al., 1992) paved a way into transformation of recalcitrant crops which were not amenable to *Agrobacterium*-mediated transformation. Genetic engineering facilitated introduction of desired traits into crops as well as understanding novel gene functions (Anjanappa and Gruissem, 2021; Nerkar et al., 2018a, b; Devarumath et al., 2015). Till date, 525 different transgenic events across 32 crops have been approved for cultivation (Kumar et al., 2020; Anjanappa and Gruissem, 2021). However, the major challenges for plant transformation are the expensive, time-consuming, and recalcitrant crops. Genetic transformation involves random integration of transgenes in the nuclear genome often leading to transgene silencing which can be overcome by precision genome editing.

PRECISION GENOME EDITING FOR DEVELOPING CROPS WITH IMPORTANT AGRONOMIC TRAITS

Genome editing has revolutionized the crop improvement through the generation of precise changes in the plant genome that was a long-standing goal of the plant breeders across the globe. Since the first report on genome editing in rice (Liu et al., 2012), genome editing has been reported in and array of food crops such as vegetable crops (Cabbage, Carrot, Pumpkin Tomato Potato, Cucumber, Sweet potato, Basil, Cassava, Chilly, Kale, Lettuce, Lactuca sativa) fruit crops (Apple, Banana, Grapefruit, Coconut, Date Palm, Grapes, Lychee, Melon, Orange, Papaya, Pear, Strawberry, Watermelon, Kiwifruit, Blueberry, Citrus) cereal crops (Barley, Rice Wheat, Maize, Oats) legume crops (Chickpea, Cowpea), sugar producing crops (Sugarcane, Sugar beet) spice crops (Pepper, Saffron) as well as other industrial crops (Coffee, Dandelion, *Jatropha curcas*, Millet, Sorghum, Switchgrass) oil crops (canola, flax, oil palm, oilseed rape, soybean and sunflower) as reviewed by Liu et al. (2021). The application of CRISPR-Cas9 for development of crops with important agronomic traits is discussed in the further sections.

CRISPR-CAS9 MEDIATED GENOME EDITING OF CROPS FOR DISEASE RESISTANCE

CRISPR-mediated engineering of plants for disease resistance has been reported in major crops (as reviewed by Zaidi et al., 2020) such as rice, wheat, tomato, banana, citrus, grapes, cassava and cucumber (**Table 2**). Broad spectrum resistance is an effective strategy for disease management in crops as these loci confer resistance to diverse species or races of pathogen. Zhou et al. (2018) discovered the *bsr-k1* allele in rice and also developed the *bsr-k1* (broad spectrum resistance Kitaake-1) mutant, which confers broad-spectrum resistance against *Magnaporthe oryzae* and *Xanthomonas oryzae* pv *oryzae* without affecting the major agronomic traits. The bacterial blight caused by *Xanthomonas oryzae* pv. *Oryzae* causes significant yield losses in rice. The expression of sucrose transporter genes *SWEET1*, *SWEET3* and *SWEET14* causes disease susceptibility. Oliva et al. (2019) engineered broad-spectrum resistance into the rice line Kitaake and two mega varieties IR64 and Ciherang-Sub1.

Simultaneous mutation of the three homeoalleles of TaMLO conferred heritable broad-spectrum resistance to powdery mildew in hexaploid bread wheat (Wang et al., 2014). Similarly, Zhang et al. (2017) generated wheat *edr1*

TABLE 2 | Genome editing using CRISPR-Cas system in major crops for disease resistance and abiotic stress tolerance.

Plant	Gene modified	Function	Agronomic trait	Transformation method	References
Disease resistance					
Rice	*Bsr-k1*	Broad spectrum resistance	Broad spectrum resistance	Agrobacterium-mediated transformation	Zhou et al. (2018)
	OsSWEET11, OsSWEET13, and OsSWEET14	Susceptibility genes for bacterial blight	Resistance to bacterial blight	Agrobacterium-mediated transformation	Oliva et al. (2019)
Wheat	TaMlo-A1, TaMloB1, and TaMlo-D1	Mildew resistance locus proteins	Resistance to powdery mildew	Biolistic transformation	Wang et al. (2014)
	TaEdr1 (three homologs)	Negative role in powdery mildew resistance	Resistance to powdery mildew	Biolistic transformation	Zhang et al. (2017)
Tomato	Pmr4	Negatively controls the SA-associated defense pathway	Resistance to powdery mildew	Agrobacterium-mediated transformation	Santillan Martinez et al. (2020)
	Jaz2	Major COR/JA-Ile co-receptor in Arabidopsis controlling stomata dynamics during bacterial invasion	Resistance to bacterial speck disease	Agrobacterium-mediated transformation	Ortigosa et al. (2019)
Banana	RGA2, Ced9	Antiapoptosis gene, prevention of fungal-induced cell death and maintenance of organelle homeostasis	Resistance to *Fusarium* wilt	Agrobacterium-mediated transformation	Dale et al. (2017)
Citrus	CsLOB1	Citrus canker disease susceptibility gene	Resistance to citrus canker	Agrobacterium-mediated transformation	Jia et al. (2017); Peng et al. (2017)
Grapes	VvWRKY52	WRKY transcription factor playing a role in biotic stress	Resistance to *B. cinerea*	Agrobacterium-mediated transformation	Wang et al. (2017)
Cassava	nCBP-1, nCBP-2	Novel cap binding proteins from the eIF4E protein family playing an essential role in the initiation of cap-dependent mRNA translation	Resistance to brown streak disease	Agrobacterium-mediated transformation	Gomez et al. (2019)
Cucumber	eIF4E	Eukaryotic translation initiation factor 4E playing role in biotic stress	Resistance to Cucumber vein yellowing virus (Ipomovirus)	Agrobacterium-mediated transformation	Chandrasekaran et al. (2016)
Abiotic stress tolerance					
Rice	OsMYB30	Cold tolerance	Cold tolerance	Agrobacterium-mediated transformation	Zeng et al. (2020)
	OsNAC14	Transcription factor	Drought tolerance	Agrobacterium-mediated transformation	Shim et al. (2018)
	PQT3	Ubiquitin ligase	Salinity tolerance	Agrobacterium-mediated transformation	Alfatih et al. (2020)
	AOX1a, AOX1b, AOX1c, BEL	Breeding stress marker	Multiple stress tolerance	Agrobacterium-mediated transformation	Xu et al. (2015)
	ALS	Acetolactate synthase	Herbicide tolerance	Agrobacterium-mediated transformation	Endo et al. (2016)
Wheat	DREB2, DREB3, ERF3	Dehydration responsive element binding protein	Drought tolerance	PEG-mediated transformation	Kim et al. (2018)
	EPSPS	Synthesis of amino acids (aromatic)	Herbicide tolerance	Biolistic transformation	Arndell et al. (2019)
	INOX, PDS	Inositol oxygenase, Phytoene desaturase	Multiple stress tolerance	Agrobacterium-mediated transformation	Upadhyay et al. (2013)
Maize	ALS	Acetolactate synthase	Herbicide tolerance	Biolistic transformation	Yadava et al. (2017)
Sugarcane	ALS	Acetolactate synthase	Herbicide tolerance	Biolistic transformation	Oz et al. (2021)
Soybean	ALS1	Acetolactate synthase	Herbicide tolerance	Biolistic transformation	Li et al. (2015)
Tomato	BZR1	Brassinosteroid regulator	Heat stress	Agrobacterium-mediated transformation	Yin et al. (2018)
	NPR1	Drought tolerance	Drought tolerance	Agrobacterium-mediated transformation	Wang et al. (2015); Li et al. (2019)
	CLV3	Regulates shoot and Floral meristem development	Salinity stress tolerance	Agrobacterium-mediated transformation	Li et al. (2018); Van Eck et al. (2019)
	PDS	Carotenoid biosynthesis	Multiple stress tolerance	Agrobacterium-mediated transformation	Woo et al., 2015
	ALS	Acetolactate synthase	Herbicide tolerance	Biolistic transformation	Veillet et al. (2019a)
Brassica napus	BnaA6.RGA (DELLA Protein)	Transcription factor	Drought tolerance	Agrobacterium-mediated transformation	Wu et al. (2020)

plants by simultaneous modification of the three homoeologs of TaEDR1 confirming its negative role in powdery mildew resistance. In tomato, Powdery Mildew Resistance 4 *PMR4* knock-out lines showed enhanced resistance against powdery mildew pathogen *Oidium neolycopersici* (Santillan Martinez et al., 2020). Genetic manipulation of defence pathways is limited due to the antagonistic interactions between the SA and JA defence pathways. Ortigosa et al. (2019) reported spatial uncoupling the SA-JA antagonism at the stomata and generated a tomato resistant to the bacterial speck disease caused by the pathogen Pto DC3000, without compromising resistance to necrotrophic pathogens, by editing the SlJAZ2 gene (Santillan Martinez et al., 2020). Resistance to *Fusarium* wilt (Banana), citrus canker (Citrus), *B. cinerea* (grapes), brown streak disease (Cassava) and *Ipomovirus* (cucumber) has also been reported (Dale et al., 2017; Jia et al., 2017; Peng et al., 2017; Wang et al., 2017; Gomez et al., 2019; Chandrasekaran et al., 2016; **Table 2**).

DEVELOPMENT OF ABIOTIC STRESS TOLERANT AND HIGH-YIELDING CROPS USING CRISPR-CAS9

Abiotic stresses pose a major threat to the crop yield and productivity. CRISPR/Cas has been adopted rapidly for the manipulation of crop genomes to develop abiotic stress tolerant and high-yielding mutants (Bhat et al., 2021). Simultaneous editing of three genes, OsPIN5b (a panicle length gene), GS3 (a grain size gene) and OsMYB30 (a cold tolerance gene) with the CRISPR-Cas9 resulted in several new rice mutants with high yield and excellent cold tolerance (Zeng et al., 2020; **Table 2**) which was also stable in the T2 generation. Overexpression of the BZR (brassinosteroid regulator) in tomato conferred thermo-tolerance via regulation of the Feronia (Fer) homologs (Yin et al., 2018).

Improvement in drought tolerance by modulating the important transcription factors has been reported in the major crops like rice (Shim et al., 2018), wheat (Kim et al., 2018), tomato (Wang et al., 2015; Li et al., 2019) and *Brassica napus* (Wu et al., 2020). Shim et al. (2018) reported the functional characterization of the rice drought responsive transcription factor OsNAC14. Overexpression of OsNAC14 conferred drought tolerance in the rice mutants at the vegetative stage of growth. Field performance of OsNAC14 overexpressing transgenic rice lines revealed that these lines exhibited higher number of panicle and filling rate compared to non-transgenic plants under drought conditions. In wheat, CRISPR-Cas9 mediated genome editing of dehydration responsive element binding protein 2 (TaDREB2) and ethylene responsive factor 3 (TaERF3) resulted in improved drought tolerance (Kim et al., 2018). Li et al. (2019) isolated *SlNPR1* (non-expressor of pathogenesis-related gene 1) from tomato and generated *slnpr1* mutants using the CRISPR/Cas9 system and found that lines overexpressing *SlNPR1* showed reduced drought tolerance. This work throws a light on function of NPR1 in plant response. In rapeseed, CRISPR-Cas9 editing of *bnaa6. rga-D* and *bnarga* genes helped in understanding roles of DELLA proteins in drought tolerance in *B. napus* (Wu et al., 2020). The *bnaa6. rga-D* mutants displayed enhanced drought tolerance and *BnaRGA*s physically interacted with BnaA10. ABF2, an essential transcription factor in ABA signaling.

Major work on salinity tolerance has been done in rice (as reviewed by Bhat et al. (2021). The Arabidopsis PARAQUAT TOLERANCE 3 (*AtPQT3*) encoding an E3 ubiquitin ligase confers an off-switch mechanism which enable plants to balance the growth and stress responses (Alfatih et al., 2020). *OsPQT3*, a rice homologue of *AtPQT3* was knock-out using CRISPR-Cas9 (Alfatih et al., 2020). The resulting *OsPQT3* knockout mutants (*ospqt3*) displayed enhanced resistance to oxidative and salt stress significantly enhanced agronomic performance with higher yield compared with the wild type under salt stress in greenhouse and in field conditions. Li et al. (2018) introduced desirable traits into four stress-tolerant wild-tomato accessions by using multiplex CRISPR–Cas9 editing of genes associated with morphology, flower and fruit production, and ascorbic acid synthesis. The Cas9-free progeny of edited plants had domesticated phenotypes and also retained disease resistance and salt tolerance traits from the parents (Li et al., 2018).

Herbicide tolerance has been engineered in rice (Endo et al., 2016), maize (Yadava et al., 2017), sugarcane (Oz et al., 2021), soybean (Li et al., 2015) and tomato (Veillet et al., 2019b) by editing the Acetolactate synthase (*ALS*) gene; and in wheat (Arndell et al., 2019) by editing the 5-enolpyruvylshikimate 3-phosphate synthase (*EPSPS*) gene to develop crop varieties resistant to chlorsulfuron and glyphosate, respectively.

In rice, CRISPR-Cas9 mediated gene editing of GS3 and Gn1a genes responsible for grain size and grain number resulted into generation of 3 mutant genotypes (gs3-N9108, gs3-Z22, and gs3gn1a-Z22) which showed 3–7% increase in grain yields than the WT (Shen et al., 2018). Hao et al. (2019) reported larger grain size in the genome edited mutants generated by editing GL2/OsGRF4 and OsGRF3 genes responsible for grain size and grain yield, respectively. CRISPR-Cas9 mediated genome editing of Gn1a, DEP1, GS3, IPA1 led to enhanced grain number; dense erect panicles; larger grain size; and variation in the tiller number in T2 generation (Li et al., 2016).

BIOFORTIFICATION OF CROPS USING CRISPR-CAS9

Biofortification of grains is one of the main goals of breeders to enhance the nutritive value of grains for controlling the nutrient-deficiency related diseases. Lysine content has been improved by up to 25-fold in rice by editing the gene *AK* (*lysC*) and *DHPS* (*dapA*) responsible for key enzymes in lysine biosynthesis (Yang et al., 2016) (**Table 3**). In addition, these high-lysine lines showed improved physic-chemical properties without affecting the starch composition. The plants showed normal growth in field trials with slight difference in plant

TABLE 3 | Genome editing using CRISPR-Cas system in major crops for increased yield, improved nutritive value, reduction in anti-nutritional factors and improved shelf-life.

Plant	Gene modified	Function	Agronomic trait	Transformation method	References
Increased yield					
Rice	GS3 and Gn1a	GS3: QTL regulating grain size; Gn1a: QTL regulating grain number	Grain size and grain number	Agrobacterium-mediated transformation	Shen et al. (2018)
	GL2/OsGRF4 and OsGRF3	GL2 transcript negatively regulated grain size and yield	Grain size and yield	Agrobacterium-mediated transformation	Hao et al. (2019)
	Gn1a; DEP1; GS3; IPA1	Gn1a: regulates grain number; DEP1: regulates panicle size; GS3: regulates grain size; IPA1: regulates plant architecture	grain number; panicle architecture; grain size; plant architecture	Agrobacterium-mediated transformation	Li et al. (2016)
Improved nutritive value					
Rice	AK (lysC) and DHPS (dapA)	Key enzymes in lysine biosynthesis	Lysine content	Agrobacterium-mediated transformation	Yang et al. (2016)
	CrtI, PSY	Carotenoid biosynthesis	High β-carotene content	Agrobacterium-mediated transformation	Dong et al. (2020)
Tomato	SlGAD2, SlGAD3	Glutamate decarboxylase- key enzyme in GABA synthesis	High GABA content	Agrobacterium-mediated transformation	Nonaka et al. (2017)
Potato	StSBE1, StSBE2	Starch branching enzymes	High amylose content	Agrobacterium-mediated transformation/PEG -mediated transformation	Tuncel et al. (2019)
Rapeseed	BnTT8	Transcription factor regulator activating pro anthocyanidins-specific genes in seed coat development	High oil production and GPC	Agrobacterium-mediated transformation	Zhai et al. (2020)
Reduction in anti-nutritional factors					
Rice	OsNramp5	Cd transporter mediating root uptake of Cd	Cd accumulation	Agrobacterium-mediated transformation	Tang et al. (2017)
	OsPLDα1	Regulates abscicic acid signalling	Low phytic acid content	Agrobacterium-mediated transformation	Khan et al. (2019)
Wheat	α-gliadin genes	Gluten proteins	Low gluten content	Biolistic transformation	Sanchez-Leon et al. (2018)
Rapeseed	BnITPK	Key enzyme ITPK (inositol tetrakisphosphate kinase), catalysing the penultimate step for the synthesis of Phytic Acid in plants	Low phytic acid content	Agrobacterium-mediated transformation	Sashidhar et al. (2020)
Improved shelf-life					
Tomato	ALC	An allele of nor (non-ripening) gene	Extended shelf life	Agrobacterium-mediated transformation	Yu et al. (2017)
	PL, PG2a, TBG4	Tomato pectin degrading enzymes determining softening in fleshy fruits	Long shelf life	Agrobacterium-mediated transformation	Wang et al. (2019)
	RIN	MADS-box transcription factor regulating fruit ripening	Slower ripening	Agrobacterium-mediated transformation	Ito et al. (2015)
Banana	MaACO1	Encodes ACC oxidase playing a role in ripening	Long shelf life	Agrobacterium-mediated transformation	Hu et al. (2020)
Petunia	PhACO1	Encodes ACC oxidase and expressed during flower development	Increased shelf life	PEG-mediated transfection	Xu et al. (2020)

height and grain colour (Yang et al., 2016). Carotenoid biofortification has been achieved in rice by genome editing of CrtI and PSY genes resulting in marker-free gene-edited mutants containing high β-carotene content (Dong et al., 2020). Biofortified tomato has been produced with diverse nutrient like γ-aminobutyric acid (GABA). GABA is a neurotransmitter that control anxiety and blood pressure. By deleting the C-terminal autoinhibitory domain of glutamate decarboxylase, a key enzyme in GABA biosynthesis, mutant tomatoes have been created in which GABA accumulation increased by seven-fold (Nonaka et al., 2017). Yellow-seeded mutants in rapeseed have been created using the CRISPR-Cas9 mediated editing of BnTT8 homologs which increased the oil content in the T2 generation by 9.47%. These BnTT8 double mutants with high oil yield potential and modified FA composition as well as improved the nutritional quality could have potential application in rapeseed breeding (Zhai et al., 2020). Tuncel et al. (2019) investigated Cas9-mediated mutagenesis of starch-

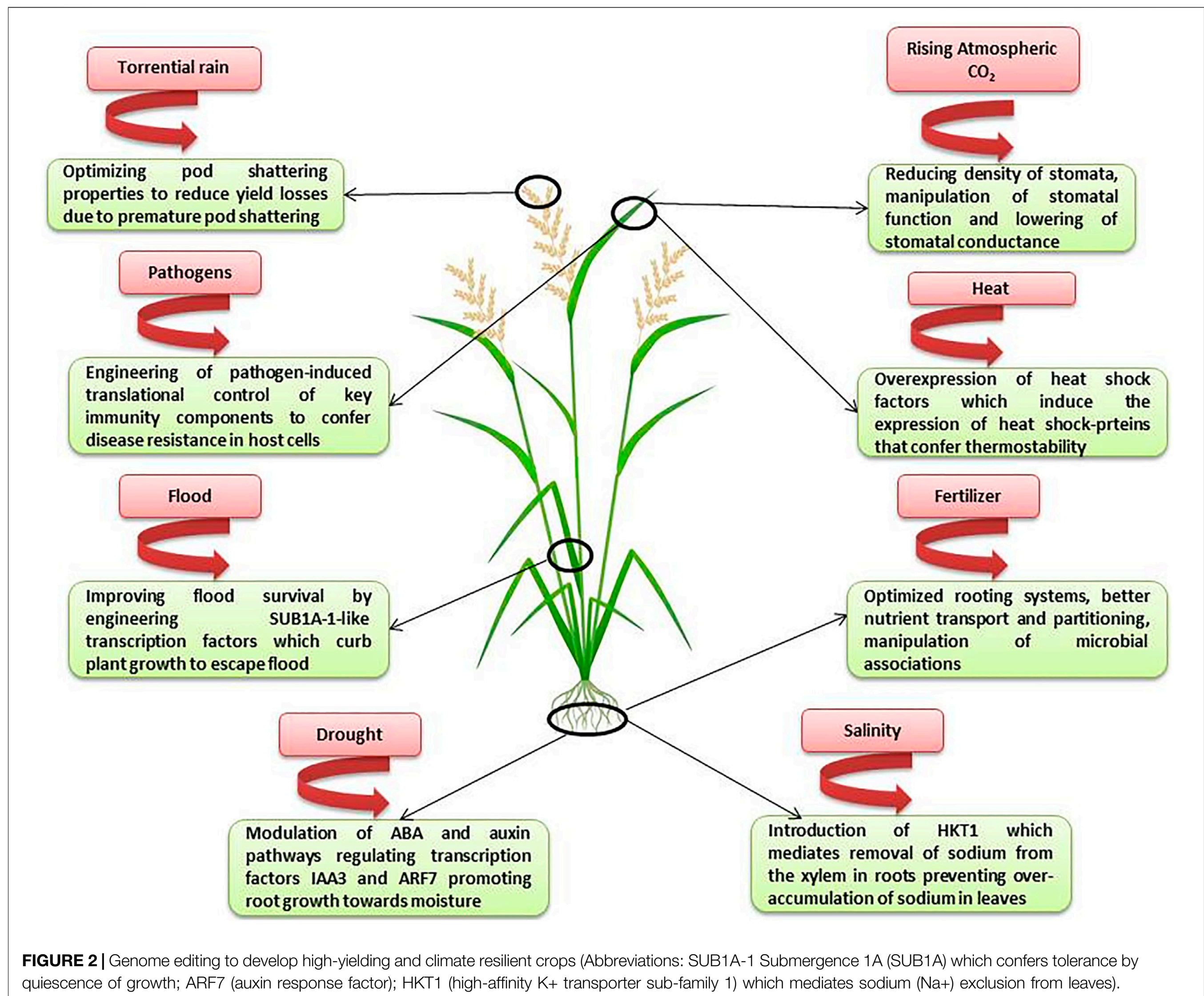


FIGURE 2 | Genome editing to develop high-yielding and climate resilient crops (Abbreviations: SUB1A-1 Submergence 1A (SUB1A) which confers tolerance by quiescence of growth; ARF7 (auxin response factor); HKT1 (high-affinity K+ transporter sub-family 1) which mediates sodium (Na+) exclusion from leaves).

branching enzymes (SBEs) in tetraploid potatoes and developed transgene-free mutant potato lines with elevated levels of resistant starch which can help in improving insulin control of blood sugar levels. Taken together, these results demonstrate that Cas9-mediated mutagenesis holds promise for development of commercially viable crops.

REDUCTION IN ANTI-NUTRITIONAL FACTORS IN CRISPR-CAS9 EDITED CROPS

In order to reduce the phytic acid content in rapeseeds, the *ITPK* gene encodes an enzyme that catalyzes the penultimate step of phytate synthesis. In rapeseed, the *ITPK* gene knock-out by CRISPR/Cas9 led to reduction in the phytic acid content by 35% (**Table 3**) without affecting the plant performance (Sashidhar et al., 2020). The gluten protein in wheat is another important anti-nutritional factor which can cause coeliac disease in gluten intolerant individuals. Reduction of the gluten content using the conventional breeding methods is difficult as this protein is encoded by more than 100 loci in the wheat genome. CRISPR/Cas9 mediated targeting of a conserved region of the α-gliadin genes has led to the production of low-gluten and transgene-free wheat lines (Sanchez-Leon et al., 2018). A remarkable application of the CRISPR/Cas9 technology in rice breeding is the generation of heavy metal pollution-safe rice cultivars. Cadmium (Cd) is a human carcinogen which can also lead to renal failure upon long-term consumption. Tang et al. (2017) developed novel Indica rice cultivars accumulating low Cd levels in the grain by mutating the *OsNramp5* gene, which mediates the root uptake of Cd. Field performance evaluation of osnramp5 mutants revealed that high Cd conditions did not affect the agronomic traits and the grain yield (Tang et al., 2017; Liu et al., 2021).

GENOME EDITING FOR DEVELOPING RESILIENT CROPS IN CHANGING CLIMATIC CONDITIONS

Genome editing tools have become the most widely used biotechnological tools in crop breeding. Presently, the genome editing of crops is at a stage of elucidating the genomic function and regulatory mechanisms (Liu et al., 2021) and there is a long way to go before the translation of research on genome edited crops from lab to the field. The climate change continues to be the major limiting factor in the crop improvement. Therefore, increasing crop yield in the sub-optimal environments is the most important goal for the breeders. Bailey-Serres et al. (2019) have enlisted the different factors affecting the crop productivity and suggested the breeding strategies for increasing crop yield in sub-optimal environments. Genome editing indeed has a crucial role to play in elucidating the gene functions during stress responses as well as the adaptive mechanisms that plants have evolved in response to the harsh environmental conditions. **Figure 2** depicts the areas where genome editing can find applications in breeding high-yielding and climate-resilient crops.

REGULATORY APPROVAL OF THE GENOME-EDITED CROPS

Genome editing is an innovative plant breeding technological advancement which creates targeted changes in the plant's own genome without the insertion of transgenic sequences. Genome editing is also referred to as a New Breeding Technique. Researchers argue that genome editing makes small genetic changes that could be found in nature. This is clearly different from introducing the DNA from other species into plant genomes. Unlike the older approaches, gene editing allows researchers to make more targeted changes in the genome.

Regulation of the genome-edited crops is crucial for its applicability for the betterment of crops which provide food, fibre and fuel for the growing population of the world in the face of a global climate change. On one hand when this technology has proved its versatile application in an array of important crops like rice, wheat, corn, soybean, tomato, potato, banana, cassava and oranges; international discussions are seeking legal clarity about the regulatory approval of genome editing and derived products (Lassoued et al., 2021). In 2019, soybean variety producing oil with a longer shelf life became the first commercialized gene edited food product to be launched in the United States by Calyxt of Roseville, Minnesota. Earlier this year, a gene-edited tomato with higher amounts of γ-aminobutyric acid (GABA) came in to the market in Japan. Recently, United Kingdom has planned to ease requirements for field research on gene-edited crops by allowing the researchers to conduct field trials of gene-edited plants without the need to submit risk assessments (Ledford 2021). Recognizing that the SDN1 (that introduce changes in the plant genome through small insertions/deletions) and SDN2 (that uses a small template to generate a desired change in the plant genome) categories of plants are free from any transgenes, the Ministry of Environment, Forest and Climate Change (Government of India) issued a notification on 30 March 2022 to exempt products of SDN1 and SDN2 (free from transgenes) from the provisions of Rules 7 and 11 (both inclusive) of Rules, 1989, whereas products of SDN3 (with transgenes) will be treated in the same way as GE organisms under Rules, 1989 (Ahuja, 2022). This decision will further boost the research and development of genome-edited products in India. However, genome-edited plants still need regulatory exemption from most of the countries in the world. Nevertheless, this will further boost the research and development of genome edited crops in India. However, as the many countries across the still await the exemption of genome-edited products from regulation, scientists believe that genome editing encompasses powerful tools for future food security that should be enabled and not delayed.

Any mutations leading to obviously deleterious phenotypes would be eliminated from breeding programmes (Carroll et al., 2016). Other hypothetical risks, such as a modified protein that turned out to be allergenic to humans, might equally well arise naturally in the absence of human intervention. The effects of genome editing are largely identical to those of the natural processes that continually create variation in the genomes of food animals. From this point of view, it is hard to see why the process of genome editing to introduce defined genetic changes should be regulated when the process of spontaneous mutation that introduces new random changes into every individual's genome, every generation, is not. Genome editing allows precise changes to be made in the genomes of agricultural organisms without the introduction of DNA from other species. The products of editing should be subject to the same oversight as other food products, based on the result rather than the process that yielded the result. This technology was developed largely with public funding, and the public should benefit from its intelligent and careful application.

Despite their promise, it is clear that not every issue can or should be solved with these technologies; many are societal problems that must be addressed by changing behavior and mindsets. Decisions to use, not to use, or how to use these tools should be made by informed stakeholders-including consumers and famers in collaboration with plant breeders. Using crops created through genetic engineering and genome editing cannot replace sustainable practices, such as cover cropping, crop rotation, or crop diversification. They can ideally be used in concert with these practices, serving as one tool of the many, that farmers at all production levels can use to adjust to local conditions and challenges.

CONCLUSION

Advances in the breeding strategies through the application of innovative technologies have the potential to furnish solutions to address the future challenges in global food security.

Combining genetic resources and innovative technologies like genome editing is important for developing crops with important agronomic traits that not only increase the global food security but also reduce the effects of agriculture on the environment. In less than a decade, CRISPR/Cas9 system has become the most widely used tool crop breeding. Considerable progress has been made in developing disease resistant and abiotic stress tolerant crops with improved yield, nutritive value and increased shelf life. Understanding novel gene functions and the regulatory mechanisms of genes controlling important agronomic traits in plants shall facilitate further progress in the application of the genome editing technologies for crop improvement. Through the identification and editing of genes involved in stress tolerance and yield improvement, it would be possible to develop robust crops that are resilient to the global climate change. Although, the translation of genome-edited crop research to the field is still a far way to go, the regulatory approval and consumer acceptance will play an essential role in commercializing the existing genome-edited crops.

AUTHOR CONTRIBUTIONS

RD and CA conceptualized the review. GN wrote the manuscript and prepared the tables and figures. MP, AK, and RV wrote the manuscript. SD, RD and CA corrected the manuscript.

ACKNOWLEDGMENTS

The authors are thankful to Shri. Shivajirao Deshmukh, Director General at Vasantdada Sugar Institute (Pune) and Dr. G. Hemaprabha, Director (Acting) at ICAR-Sugarcane Breeding Institute (Coimbatore); for their kind support and encouragement.

REFERENCES

Ahmar, S., Mahmood, T., Fiaz, S., Mora-Poblete, F., Shafique, M. S., Chattha, M. S., et al. (2021). Advantage of Nanotechnology-Based Genome Editing System and its Application in Crop Improvement. *Front. Plant Sci.* 12, 663849. doi:10.3389/fpls.2021.663849

Ahuja, V. (2022). Exemption of Genome Edited Plants Falling Under the Categories of SDN-1 and SDN-2 from the Provisions of the Biosafety Regulations in India. *South Asia Biosefety Program Newsletter* 19, 1-2.

Alexandratos, N., and Bruinsma, J. (2012). *World Agriculture towards 2030/2050: The 2012 Revision. ESA Working Paper No. 12-03.* Rome: FAO.

Alfatih, A., Wu, J., Jan, S. U., Zhang, Z. S., Xia, J. Q., and Xiang, C. B. (2020). Loss of Rice PARAQUAT TOLERANCE 3 Confers Enhanced Resistance to Abiotic Stresses and Increases Grain Yield in Field. *Plant Cell. Environ.* 43, 2743–2754. doi:10.1111/pce.13856

Altpeter, F., Springer, N. M., and Bartley, L. E. (20162016). Advancing Crop Transformation in the Era of Genome Editing. *Plant Cell.* 28, 1510–1520. doi:10.1105/tpc.16.00196

Anjanappa, R. B., and Gruissem, W. (2021). Current Progress and Challenges in Crop Genetic Transformation. *J. Plant Physiology* 261, 153411. doi:10.1016/j.jplph.2021.153411

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). Search-and-replace Genome Editing without Double-Strand Breaks or Donor DNA. *Nature* 576 (7785), 149–157. doi:10.1038/s41586-019-1711-4

Arndell, T., Sharma, N., Langridge, P., Baumann, U., Watson-Haigh, N. S., and Whitford, R. (2019). gRNA Validation for Wheat Genome Editing with the CRISPR-Cas9 System. *BMC Biotechnol.* 19 (1), 71. doi:10.1186/s12896-019-0565-z

Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G. E. D., and Schroeder, J. I. (2019). Genetic Strategies for Improving Crop Yields. *Nature* 575 (7781), 109–118. doi:10.1038/s41586-019-1679-0

Bandyopadhyay, A., Kancharla, N., Javalkote, V. S., Dasgupta, S., and Brutnell, T. P. (2020). CRISPR-Cas12a (Cpf1): A Versatile Tool in the Plant Genome Editing Tool Box for Agricultural Advancement. *Front. Plant Sci.* 11, 584151. doi:10.3389/fpls.2020.584151

Barton, K. A., and Brill, W. J. (1983). Prospects in Plant Genetic Engineering. *Science* 219 (4585), 671–682. doi:10.1126/science.6297007

Basso, M. F., Arraes, F. B. M., Grossi-de-Sa, M., Moreira, V. J. V., Alves-Ferreira, M., and Grossi-de-Sa, M. F. (2020). Insights into Genetic and Molecular Elements for Transgenic Crop Development. *Front. Plant Sci.* 11, 509. doi:10.3389/fpls.2020.00509

Bhat, M. A., Mir, R. A., Kumar, V., Shah, A. A., Zargar, S. M., Rahman, S., et al. (2021). Mechanistic Insights of CRISPR/Cas-mediated Genome Editing towards Enhancing Abiotic Stress Tolerance in Plants. *Physiol. Plant.* 172, 1255–1268. doi:10.1111/ppl.13359

Bohar, R., Chitkineni, A., and Varshney, R. K. (2020). Genetic Molecular Markers to Accelerate Genetic Gains in Crops. *Biotechniques* 69 (3), 158–160. doi:10.2144/btn-2020-0066

Butt, H., Rao, G. S., Sedeek, K., Aman, R., Kamel, R., and Mahfouz, M. (2020). Engineering Herbicide Resistance via Prime Editing in Rice. *Plant Biotechnol. J.* doi:10.1111/pbi.13399

Cai, Y., Chen, L., Zhang, Y., Yuan, S., Su, Q., Sun, S., et al. (2020). Target Base Editing in Soybean Using a Modified CRISPR/Cas9 System. *Plant Biotechnol. J.* 18, 1996–1998. doi:10.1111/pbi.13386

Caligari, P. D. (2001). Plant Breeding and Crop Improvement. *LS.* doi:10.1002/9780470015902.a0002024.pub3

Carroll, D., Van Eenennaam, A., Taylor, J., Seger, J., and Voytas, D. F. (2016). Regulate Genome-Edited Products, Not Genome Editing Itself. *Nat. Biotechnol.* 34, 477–479. doi:10.1038/nbt.3566

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., et al. (2016). Development of Broad Virus Resistance in Non-transgenic Cucumber Using CRISPR/Cas9 Technology. *Mol. Plant Pathol.* 17 (7), 1140–1153. doi:10.1111/mpp.12375

Charpentier, E., Richter, H., van der Oost, J., and White, M. F. (2015). Biogenesis Path- Ways of RNA Guides in Archaeal and Bacterial CRISPR-Cas Adaptive Immunity. *FEMS Microbiol. Rev.* 39 (3), 428–441. doi:10.1093/femsre/fuv023

Chen, Y., Wang, Z., Ni, H., Xu, Y., Chen, Q., Jiang, L., et al. (2017). CRISPR/Cas9-mediated Base-Editing System Efficiently Generates Gain-Of-Function Mutations in Arabidopsis. *Sci. China Life Sci.* 60, 520–523. doi:10.1007/s11427-017-9021-5

Cheng, H., Hao, M., Ding, B., Mei, D., Wang, W., Wang, H., et al. (2021). Base Editing with High Efficiency in Allotetraploid Oilseed Rape by A3A-PBE System. *Plant Biotechnol. J.* 19, 87–97. doi:10.1111/pbi.13444

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., et al. (2010). Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics* 186 (2), 757–761. doi:10.1534/genetics.110.120717

Dale, J., James, A., Paul, J-Y., Khanna, H., Smith, M., Peraza-Echeverria, S., et al. (2017). Transgenic Cavendish Bananas with Resistance to Fusarium Wilt Tropical Race 4. *Nat. Commun.* 8 (1), 1496. doi:10.1038/s41467-017-01670-6

Devarumath, R. M., Kalwade, S. B., Kulkarni, P. A., Sheelavanthmath, S. S., and Suprasanna, P. (2014). "Integrating Omics Approaches in Sugarcane Improvement," in *OMICS Applications in Crop Science.* Editor Debmalya Barh (NY: CRC Press, Taylor & Francis Group), 191–250.

Devarumath, R. M., Mirajkar, S. J., Thorat, A. S., Farsangi, F. J., and Suprasanna, P. (2019). "Genomic Landscapes of Abiotic Stress Responses in Sugarcane," in *Genomics Assisted Breeding of Crops for Abiotic Stress Tolerance.*

Sustainable Development and Biodiversity. Editors V. Rajpal, D. Sehgal, A. Kumar, and S. Raina (Cham: Springer), II21. doi:10.1007/978-3-319-99573-1_12

Devarumath, R. M., Nerkar, G. A., Farsangi, F. J., Nikam, A. A., and Babu, K. H. (2015). "Embracing Biotechnology Methods for Crop Improvement Research in Sugarcane," in *Current Status of Sugarcane Research in India*. Editors A. K. Tiwari, M. Lal, and A. K. Singh (USA: Nova Science Publishers).

Dong, O. X., Yu, S., Jain, R., Zhang, N., Duong, P. Q., Butler, C., et al. (2020). Marker-free Carotenoid Enriched Rice Generated through Targeted Gene Insertion Using CRISPR-Cas9. *Nat. Commun.* 11, 1–10. doi:10.1038/s41467-020-14981-y

Endo, M., Mikami, M., and Toki, S. (2016). Biallelic Gene Targeting in Rice. *Plant Physiol.* 170 (2), 667. doi:10.1104/pp.15.01663

Fiaz, S., Khan, S. A., Anis, G. B., Gaballah, M. M., and Riaz, A. (2021a). "CRISPR/Cas Techniques: a New Method for RNA Interference in Cereals," in *CRISPR and RNAi Systems: Nanobiotechnology Approaches to Plant Breeding and Protection*. Editor C. Cockle. 1st Edn (Amsterdam: Elsevier publisher), 233–252. doi:10.1016/b978-0-12-821910-2.00032-1

Fiaz, S., Khan, S. A., Noor, M. A., Ali, K., and Wang, X. (2021b). Novel Plant Breeding Techniques to Advance Nitrogen Use Efficiency in Rice: A Review. *GM Crop Food*. doi:10.1080/21645698.2021.1921545

Fiaz, S., and Wang, X. (2021c). Evolution and Application of Novel Plant Breeding Techniques for Achieving Food Security. *Int. J. Mol. Sci.* 22 (11), 5585. doi:10.3390/ijms22115585

Fiaz, S., Xiukang, W., Afifa, Y., Alharthi, B., and Ali, H. (2020). Apomixis and Strategies for Induce Apomixis to Preserve Hybrid Seed Vigor for Multiple Generations. *Gm. Foods Crop* 12, 57–70. doi:10.1080/21645698.2020.1808423

Foria, S., Magris, G., Jurman, I., Schwope, R., De Candido, M., De Luca, E., et al. (2022). Extent of Wild-To-Crop Interspecific Introgression in Grapevine (Vitis vinifera) as a Consequence of Resistance Breeding and Implications for the Crop Species Definition. *Hortic. Res.* 9, uhab010. doi:10.1093/hr/uhab010

Gomez, M. A., Lin, D. Z., Moll, T., Chauhan, R. D., Hayden, L., Renninger, K., et al. (2019). Simultaneous CRISPR/Cas9-Mediated Editing of Cassava eIF4E Isoforms nCBP-1 and nCBP-2 Reduces Cassava Brown Streak Disease Symptom Severity and Incidence. *Plant Biotechnol. J.* 17 (2), 421–434. doi:10.1111/pbi.12987

Haber, J. E. (2000). Partners and Pathways: Repairing a Double-Strand Break. *Trends Genet.* 16, 259–264. doi:10.1016/s0168-9525(00)02022-9

Hajjar, R., and Hodgkin, T. (2007). The Use of Wild Relatives in Crop Improvement: a Survey of Developments over the Last 20 Years. *Euphytica* 156, 1–13. doi:10.1007/s10681-007-9363-0

Hao, L., Pu, X., and Song, J. (2021). Introduction of Mutations in Plants with Prime Editing. *Methods* 194, 83–93. doi:10.1016/j.ymeth.2021.03.014

Hao, L., Ruiying, Q., Xiaoshuang, L., Shengxiang, L., Rongfang, X., Jianbo, Y., et al. (2019). CRISPR/Cas9-mediated Adenine Base Editing in Rice Genome. *Rice Sci.* 26, 125–128. doi:10.1016/j.rsci.2018.07.002

Horsch, R. B., Fry, J. E., Hoffmann, N. L., Wallroth, M., Eichholtz, D., Rogers, S. G., et al. (1985). A Simple and General Method for Transferring Genes into Plants. *Science* 227 (4691), 1229–1231. doi:10.1126/science.227.4691.1229

Hu, C., Sheng, O., Deng, G., He, W., Dong, T., Yang, Q., et al. (2020). CRISPR/Cas9-mediated Genome Editing of MaACO1 (Aminocyclopropane-1-carboxylate Oxidase1) Promotes the Shelf Life of Banana Fruit. *Plant Biotechnol. J.* 1-3, 654–656. doi:10.1111/pbi.13534

Hunziker, J., Nishida, K., Kondo, A., Kishimoto, S., Ariizumi, T., and Ezura, H. (2020). Multiple Gene Substitution by Target-AID Base-Editing Technology in Tomato. *Sci. Rep.* 10, 20471. doi:10.1038/s41598-020-77379-2

Ito, Y., Nishizawa-Yokoi, A., Endo, M., Mikami, M., and Toki, S. (2015). CRISPR/Cas9-mediated Mutagenesis of the RIN Locus that Regulates Tomato Fruit Ripening. *Biochem. Biophys. Res. Commun.* 467, 76–82. doi:10.1016/j.bbrc.2015.09.117

Jackson, S. A., McKenzie, R. E., Fagerlund, R. D., Kieper, S. N., Fineran, P. C., and Brouns, S. J. (2017). CRISPR-cas: Adapting to Change. *Science*. doi:10.1126/science.aal5056

Jansky, S. H., and Peloquin, S. J. (2006). Advantages of Wild Diploid Solanum Species over Cultivated Diploid Relatives in Potato Breeding Programs. *Genet. Resour. Crop. Evol.* 53, 669–674. doi:10.1007/s10722-004-2949-7

Jena, K. K. (2010). The Species of the Genus *Oryza* and Transfer of Useful Genes from Wild Species into Cultivated Rice. *O. Sativa. Breed. Sci.* 60, 518–523. doi:10.1270/jsbbs.60.518

Jia, H., Zhang, Y., Orbović, V., Xu, J., White, F. F., Jones, J. B., et al. (2017). Genome Editing of the Disease Susceptibility Gene CsLOB1 in Citrus Confers Resistance to Citrus Canker. *Plant Biotechnol. J.* 15 (7), 817–823. doi:10.1111/pbi.12677

Jiang, G. L. (2013). Molecular Markers and Marker-Assisted Breeding in Plants. *Plant Breed. laboratories fields* 3, 45–83. doi:10.5772/52583

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* 337 (6096), 816–821. doi:10.1126/science.1225829

Khan, M. S. S., Basnet, R., Islam, S. A., and Shu, Q. (2019). Mutational Analysis of OsPLDα1 Reveals its Involvement in Phytic Acid Biosynthesis in Rice Grains. *J. Agric. Food Chem.* 67, 11436–11443. doi:10.1021/acs.jafc.9b05052

Kim, D., Alptekin, B., and Budak, H. (2018). CRISPR/Cas9 Genome Editing in Wheat. *Funct. Integr. Genomics* 18 (1), 31–41. doi:10.1007/s10142-017-0572-x

Kim, Y.-G., Cha, J., and Chandrasegaran, S. (1996). Hybrid Restriction Enzymes: Zinc Finger Fusions to FokI Cleavage Domain. *Proc. Natl. Acad. Sci. U. S. A.* 93, 1156–1160. doi:10.1073/pnas.93.3.1156

Klein, T. M., Arentzen, R., Lewis, P. A., and Fitzpatrick-McElligott, S. (1992). Transformation of Microbes, Plants and Animals by Particle Bombardment. *Biotechnol. (N Y).* 10 (3), 286–291. doi:10.1038/nbt0392-286

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., et al. (2015). Engineered CRISPR-Cas9 Nucleases with Altered PAM Specificities. *Nature* 523 (7561), 481–485. doi:10.1038/nature14592

Koonin, E. V., Makarova, K. S., and Zhang, F. (2017). Diversity, Classifcation and Evolu- Tion of CRISPR-Cas Systems. *Curr. Opin. Microbiol.* 37, 67–78. doi:10.1016/j.mib.2017.05.008

Kumar, K., Gambhir, G., Dass, A., Tripathi, A. K., Singh, A., Jha, A. K., et al. (2020). Genetically Modified Crops: Current Status and Future Prospects. *Planta* 251, 91. doi:10.1007/s00425-020-03372-8

Kumawat, G., Kumawat, C. K., Chandra, K., Pandey, S., Chand, S., Mishra, U. N., et al. (2020). Insights into Marker Assisted Selection and its Applications in Plant Breeding. In (Ed.), *Plant Breeding - Current and Future Views. IntechOpen*. doi:10.5772/intechopen.95004

Labroo, M. R., Studer, A. J., and Rutkoski, J. E. (2021). Heterosis and Hybrid Crop Breeding: A Multidisciplinary Review. *Front. Genet.* 12, 643761. doi:10.3389/fgene.2021.643761

Lassoued, R., Phillips, P. W. B., Macall, D. M., Hesseln, H., and Smyth, S. J. (2021). Expert Opinions on the Regulation of Plant Genome Editing. *Plant Biotechnol. J.* 19 (6), 1104–1109. doi:10.1111/pbi.13597

Ledford, H. 2021. New Rules Will Make UK Gene-Edited Crop Research Easier. *Nature* (News). doi:10.1038/d41586-021-01572-0

Li, M., Li, X., Zhou, Z., Wu, P., Fang, M., Pan, X., et al. (2016). Reassessment of the Four Yield-Related Genes Gn1a, DEP1, GS3, and IPA1 in Rice Using a CRISPR/Cas9 System. *Front. Plant Sci.* 7, 377. doi:10.3389/fpls.2016.00377

Li, R., Liu, C., Zhao, R., Wang, L., Chen, L., Yu, W., et al. (2019). CRISPR/Cas9-mediated SlNPR1 Mutagenesis Reduces Tomato Plant Drought Tolerance. *BMC Plant Biol.* 19 (1), 38. doi:10.1186/s12870-018-1627-4

Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., et al. (2018). Domestication of Wild Tomato Is Accelerated by Genome Editing. *Nat. Biotechnol.* 36, 1160–1163. doi:10.1038/nbt.4273

Li, Z., Liu, Z. B., Xing, A., Moon, B. P., Koellhoffer, J. P., Huang, L., et al. (2015). Cas9-guide RNA Directed Genome Editing in Soybean. *Plant Physiol.* 169 (2), 960–970. doi:10.1104/pp.15.00783

Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., et al. (2020). Prime Genome Editing in Rice and Wheat. *Nat. Biotechnol.* 38 (5), 582–585. doi:10.1038/s41587-020-0455-x

Liu, D., Chen, X., Liu, J., Ye, J., and Guo, Z. (2012). The Rice ERF Transcription Factor OsERF922 Negatively Regulates Resistance to Magnaporthe Oryzae and Salt Tolerance. *J. Exp. Bot.* 63, 3899–3912. doi:10.1093/jxb/ers09710.1093/jxb/ers079

Liu, Q., Yang, F., Zhang, J., Liu, H., Rahman, S., Islam, S., et al. (2021). Application of CRISPR/Cas9 in Crop Quality Improvement. *Int. J. Mol. Sci.* 22 (8), 4206. doi:10.3390/ijms22084206

Ma, X., Zhang, X., Liu, H., and Li, Z. (2020). Highly Efficient DNA-free Plant Genome Editing Using Virally Delivered CRISPR–Cas9. *Nat. Plants* 6, 773–779. doi:10.1038/s41477-020-0704-5

Mackelprang, R., and Lemaux, P. G. (2020). Genetic Engineering and Editing of Plants: An Analysis of New and Persisting Questions. *Annu. Rev. Plant Biol.* 71, 659–687. doi:10.1146/annurev-arplant-081519-035916

Madina, M. H., Haque, M. E., Dutta, A. K., Islam, M. A., Deb, A. C., and Sikdar, B. (2013). Estimation of Genetic Diversity in Six Lentil (Lens Culinaris Medik.) Varieties Using Morphological and Biochemical Markers. *Int. J. Sci. Eng. Res.* 4, 819–825.

Mohanraju, P., Makarova, K. S., Zetsche, B., Zhang, F., Koonin, E. V., and van der Oost, J. (2016). Diverse Evolutionary Roots and Mechanistic Variations of the CRISPRCas Systems. *Science* 353 (6299), aad5147. doi:10.1126/science.aad5147

Molla, K. A., Sretenovic, S., Bansal, K. C., and Qi, Y. (2021). Precise Plant Genome Editing Using Base Editors and Prime Editors. *Nat. Plants* 7, 1166–1187. doi:10.1038/s41477-021-00991-1

Mujeeb-Kazi, A., and Asiedu, R. (1990). "Wide Hybridization- Potential of Alien Genetic Transfers for *Triticum aestivum* Improvement," in *Wheat. Biotechnology in Agriculture and Forestry*. Editor Y Bajaj (Berlin Heidelberg: Springer), 13, 111–127. doi:10.1007/978-3-662-10933-5_7

Nadakuduti, S. S., and Enciso-Rodríguez, F. (2021). Advances in Genome Editing with CRISPR Systems and Transformation Technologies for Plant DNA Manipulation. *Front. Plant Sci.* 11, 637159. doi:10.3389/fpls.2020.637159

Nerkar, G. A., Purankar, M. V., Sheelavantmath, S., and Devarumath, R. M. (2018b). "Biotechnological Approach: A New Dimension for Sugarcane Improvement," in *Biotechnology to Enhance Sugarcane Productivity and Stress Tolerance*. Editor Kalpana Sengar (NY: CRC Press, Taylor & Francis Group), 1–24. doi:10.1201/9781315152776-1

Nerkar, G., Thorat, A, Sheelavantmath, S., Babu, K. H., and Devarumath, R. (2018a). "Genetic Transformation of Sugarcane and Field Performance of Transgenic Sugarcane," in *Satbir Singh Gosal and Shabir Hussain Wani Biotechnologies of Crop Improvement*, (Switzerland AG: Springer Nature), Vol. 2. doi:10.1007/978-3-319-90650-8_9

Nonaka, S., Arai, C., Takayama, M., Matsukura, C., and Ezura, H. (2017). Efficient Increase of ɣ-Aminobutyric Acid (GABA) Content in Tomato Fruits by Targeted Mutagenesis. *Sci. Rep.* 7, 705. doi:10.1038/s41598-017-06400-y

Oliva, R., Ji, C., Atienza-Grande, G., Huguet-Tapia, J. C., Perez-Quintero, A., Li, T., et al. (2019). Broad-spectrum Resistance to Bacterial Blight in Rice Using Genome Editing. *Nat. Biotechnol.* 37 (11), 1344–1350. doi:10.1038/s41587-019-0267-z

Ortigosa, A., Gimenez-Ibanez, S., and Solano, R. (2019). Design of a Bacterial Speck Resistant Tomato by CRISPR/Cas9-mediated Editing of SlJAZ2. *Plant Biotechnol. J.* 17 (3), 665–673. doi:10.1111/pbi.13006

Oz, M. T., Altpeter, A., Karan, R., Merotto, A., and Altpeter, F. (2021). CRISPR/Cas9-Mediated Multi-Allelic Gene Targeting in Sugarcane Confers Herbicide Tolerance. *Front. Genome Ed.* 3, 673566. doi:10.3389/fgeed.2021.673566

Pausch, P., Al-Shayeb, B., Bisom-Rapp, E., Tsuchida, C. A., Li, Z., Cress, B. F., et al. (2020). CRISPR-cas *8* from Huge Phages Is a Hypercompact Genome Editor. *Science* 369, 333–337. doi:10.1126/science.abb1400

Peng, A., Chen, S., He, Y., Xu, L., He, Y., Wu, L., et al. (2017). Engineering Canker-Resistant Plants through CRISPR/Cas9-targeted Editing of the Susceptibility Gene CsLOB1 Promoter in Citrus. *Plant Biotechnol. J.* 15 (12), 1509–1519. doi:10.1111/pbi.12733

Platten, J. D., Cobb, J. N., and Zantua, R. E. (2019). Criteria for Evaluating Molecular Markers: Comprehensive Quality Metrics to Improve Marker-Assisted Selection. *PloS one* 14 (1), e0210529. doi:10.1371/journal.pone.0210529

Puchta, H., Dujon, B., and Hohn, B. (1993). Homologous Recombination in Plant Cells Is Enhanced by *In Vivo* Induction of Double Strand Breaks into DNA by a Site-specific Endonuclease. *Nucleic Acids Res.* 21 (22), 5034–5040. doi:10.1093/nar/21.22.5034

Puchta, H. (2005). The Repair of Double-Strand Breaks in Plants: Mechanisms and Consequences for Genome Evolution. *J. Exp. Bot.* 56, 1–14. doi:10.1093/jxb/eri025

Purankar, M. V., Nikam, A. A., Devarumath, R. M., and Suprasanna, P. (2022). Radiation Induced Mutagenesis, Physio-Biochemical Profiling and Field Evaluation of Mutants in Sugarcane Cv. CoM 0265. *Int. J. Radiat. Biol.* (in press). doi:10.1080/09553002.2022.2024291

Qin, L., Li, J., Wang, Q., Xu, Z., Sun, L., Alariqi, M., et al. (2020). High-efficient and Precise Base Editing of C•G to T•A in the Allotetraploid Cotton (Gossypium Hirsutum) Genome Using a Modified CRISPR/Cas9 System. *Plant Biotechnol. J.* 18, 45–56. doi:10.1111/pbi.13168

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., et al. (2013). Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell.* 154, 1380–1389. doi:10.1016/j.cell.2013.08.021

Safari, F., Zare, K., Negahdaripour, M., Barekati-Mowahed, M., and Ghasemi, Y. (2019). CRISPR Cpf1 Proteins: Structure, Function and Implications for Genome Editing. *Cell. Biosci.* 9, 1–21. doi:10.1186/s13578-019-0298-7

Sanchez-Leon, S., Gil-Humanes, J., Ozuna, C. V., Gimenez, M. J., Sousa, C., Voytas, D. F., et al. (2018). Low-gluten, Nontransgenic Wheat Engineered with CRISPR/Cas9. *Plant Biotechnol. J.* 16, 902–910. doi:10.1111/pbi.12837

Sanford, J. C., Klein, T. M., Wolf, E., and Allen, N. (1987). Delivery of Substances into Cells and Tissues Using a Particle Bombardment Process. *Part. Sci. Technol.* 5, 27–37. doi:10.1080/02726358708904533

Santillan Martinez, M. I., Bracuto, V., Koseoglou, E., Appiano, M., Jacobsen, E., Visser, R. G. F., et al. (2020). CRISPR/Cas9-targeted Mutagenesis of the Tomato Susceptibility Gene PMR4 for Resistance against Powdery Mildew. *BMC Plant Biol.* 20 (1), 284. doi:10.1186/s12870-020-02497-y

Sashidhar, N., Harloff, H. J., Potgieter, L., and Jung, C. (2020). Gene Editing of Three BnITPK Genes in Tetraploid Oilseed Rape Leads to Significant Reduction of Phytic Acid in Seeds. *Plant Biotechnol. J.* 18, 2241–2250. doi:10.1111/pbi.13380

Schouten, H. J., Tikunov, Y., Verkerke, W., Finkers, R., Bovy, A., Bai, Y., et al. (2019). Breeding Has Increased the Diversity of Cultivated Tomato in the Netherlands. *Front. Plant Sci.* 10, 1606. doi:10.3389/fpls.2019.01606

Shen, L., Wang, C., Fu, Y., Wang, J., Liu, Q., Zhang, X., et al. (2018). QTL Editing Confers Opposing Yield Performance in Different Rice Varieties. *J. Integr. Plant Biol.* 60, 89–93. doi:10.1111/jipb.12501

Shim, J. S., Oh, N., Chung, P. J., Kim, Y. S., Choi, Y. D., and Kim, J. K. (2018). Overexpression of OsNAC14 Improves Drought Tolerance in Rice. *Front. Plant Sci.* 9, 310. doi:10.3389/fpls.2018.00310

Shimatani, Z., Kashojiya, S., Takayama, M., Terada, R., Arazoe, T., Ishii, H., et al. (2017). Targeted Base Editing in Rice and Tomato Using a CRISPR-Cas9 Cytidine Deaminase Fusion. *Nat. Biotechnol.* 35, 441–443. doi:10.1038/nbt.3833

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., et al. (2017). Diversity and Evolution of Class 2 CRISPR-Cas Systems. *Nat. Rev. Microbiol.* 15 (3), 169–182. doi:10.1038/nrmicro.2016.184

Shriram, V., Kumar, V., Devarumath, R. M., Khare, T. S., and Wani, S. H. (2016). MicroRNAs as Potential Targets for Abiotic Stress Tolerance in Plants. *Front. Plant Sci.* 7, 817. doi:10.3389/fpls.2016.00817

Singh, R. K., Prasad, A., Muthamilarasan, M., Parida, S. K., and Prasad, M. (2020). Breeding and Biotechnological Interventions for Trait Improvement: Status and Prospects. *Planta* 252, 54. doi:10.1007/s00425-020-03465-4

Smith, J., Grizot, S., Arnould, S., Grizot, S., Epinat, J. C., Chames, P., et al. (2006). A Combinatorial Approach to Create Artificial Homing Endonucleases Cleaving Chosen Sequences. *Nucleic Acids Res.* 34 (22), e149. doi:10.1093/nar/gkl720

Symington, L. S., and Gautier, J. (2011). Double-Strand Break End Resection and Repair Pathway Choice. *Annu. Rev. Genet.* 45 (1), 247–271. doi:10.1146/annurev-genet-110410-132435

Tang, L., Mao, B., Li, Y., Lv, Q., Zhang, L., Chen, C., et al. (2017). Knockout of OsNramp5 Using the CRISPR/Cas9 System Produces Low Cd-Accumulating Indica Rice without Compromising Yield. *Sci. Rep.* 7, 14438. doi:10.1038/s41598-017-14832-9

Tang, X., Sretenovic, S., Ren, Q., Jia, X., Li, M., Fan, T., et al. (2020). Plant Prime Editors Enable Precise Gene Editing in Rice Cells. *Mol. Plant* 13, 667–670. doi:10.1016/j.molp.2020.03.010

Tirado, M. C., Clarke, R., Jaykus, L. A., Mcquatters -Gollop, A., and Franke, J. M. (2010). Climate Change and Food Safety: a Review. *Food Res. Int.* 43, 1745–1765. doi:10.1016/j.foodres.2010.07.003

Tuncel, A., Corbin, K. R., Ahn-Jarvis, J., Harris, S., Hawkins, E., Smedley, M. A., et al. (2019). Cas9-mediated Mutagenesis of Potato Starch-Branching Enzymes

Generates a Range of Tuber Starch Phenotypes. *Plant Biotechnol. J.* 17, 2259–2271. doi:10.1111/pbi.13137
Upadhyay, S. K., Kumar, J., Alok, A., and Tuli, R. (2013). RNA-guided Genome Editing for Target Gene Mutations in Wheat. *Genes., Genomes, Genet. (Bethesda)* 3 (12), 2233–2238. doi:10.1534/g3.113.008847
Van Eck, J., Keen, P., and Tjahjadi, M. (2019). *Agrobacterium Tumefaciens* Mediated Transformation of Tomato. *Methods Mol. Biol.* 1864, 225–234. doi:10.1007/978-1-4939-8778-8_16
Veillet, F., Chauvin, L., Kermarrec, M-P., Sevestre, F., Merrer, M., Terret, Z., et al. (2019a). The Solanum tuberosum GBSSI Gene: a Target for Assessing Gene and Base Editing in Tetraploid Potato. *Plant Cell. Rep.* 38, 1065–1080. doi:10.1007/s00299-019-02426-w
Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M. P., Guyon-Debast, A., Chauvin, J. E., et al. (2019b). Transgene-free Genome Editing in Tomato and Potato Plants Using Agrobacterium-Mediated Delivery of a CRISPR/Cas9 Cytidine Base Editor. *Int. J. Mol. Sci.* 20 (2), 402. doi:10.3390/ijms20020402
Wang, D., Samsulrizal, N. H., Yan, C., Allcock, N. S., Craigon, J., Blanco-Ulate, B., et al. (2019). Characterization of CRISPR Mutants Targeting Genes Modulating Pectin Degradation in Ripening Tomato. *Plant Physiol.* 179, 544–557. doi:10.1104/pp.18.01187
Wang, S., Zhang, S., Wang, W., Xiong, X., Meng, F., and Cui, X. (2015). Efficient Targeted Mutagenesis in Potato by the CRISPR/Cas9 System. *Plant Cell. Rep.* 34 (9), 1473–1476. doi:10.1007/s00299-015-1816-7
Wang, X., Tu, M., Wang, D., Liu, J., Li, Y., and Li, Z. (2017). CRISPR/Cas9-mediated Efficient Targeted Mutagenesis in Grape in the First Generation. *Plant Biotechnol. J.* 16 (4), 844–855. doi:10.1111/pbi.12832
Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous Editing of Three Homoeoalleles in Hexaploid Bread Wheat Confers Heritable Resistance to Powdery Mildew. *Nat. Biotechnol.* 32 (9), 947–951. doi:10.1038/nbt.2969
Wolfe, M. D., Bauchet, G. J., Chan, A. W., Lozano, R., Egesi, C., Ramu, P., et al. (2019). Historical Introgressions from a Wild Relative of Modern Cassava Improved Important Traits and May Be under Balancing Selection. *Genetics* 213, 1237–1253. doi:10.1534/genetics.119.302757
Woo, J. W., Kim, J., Il Kwon, S., Corvalan, C., Cho, S. W., Kim, H., et al. (2015). DNA-Free Genome Editing in Plants with Pre-Assembled CRISPR-Cas9 Ribonucleoproteins. *Nat. Biotechnol.* 33, 1162–1164. doi:10.1038/nbt.3389
Wu, J., Yan, G., Duan, Z., Wang, Z., Kang, C., Guo, L., et al. (2020). Roles of the Brassica Napus DELLA Protein BnaA6.RGA, in Modulating Drought Tolerance by Interacting with the ABA Signaling Component BnaA10. ABF2.*Front. Plant Sci.* 11, 577. doi:10.3389/fpls.2020.00577
Xu, J., Kang, B. C., Naing, A. H., Bae, S. J., Kim, C. K., and Kim, H. (2020). CRISPR/Cas9-mediated Editing of 1-Aminocyclopropane-1-Carboxylate Oxidase1 Enhances Petunia Flower Longevity. *Plant Biotechnol. J.* 18, 287–297. doi:10.1111/pbi.13197
Xu, R. F., Li, H., Qin, R. Y., Li, J., Qiu, C. H., Yang, Y. C., et al. (2015). Generation of Inheritable and "transgene Clean" Targeted Genome-Modified Rice in Later Generations Using the CRISPR/Cas9 System. *Sci. Rep.* 5, 11491. doi:10.1038/srep11491
Yadava, P., Abhishek, A., Singh, R., Singh, I., Kaul, T., Pattanayak, A., et al. (2017). Advances in Maize Transformation Technologies and Development of Transgenic Maize. *Front. Plant Sci.* 7, 1949. doi:10.3389/fpls.2016.01949
Yang, Q. Q., Zhang, C. Q., Chan, M. L., Zhao, D. S., Chen, J. Z., Wang, Q., et al. (2016). Biofortification of Rice with the Essential Amino Acid Lysine: Molecular Characterization, Nutritional Evaluation, and Field Performance. *J. Exp. Bot.* 67, 4285–4296. doi:10.1093/jxb/erw209
Yin, Y., Qin, K., Song, X., Zhang, Q., Zhou, Y., Xia, X., et al. (2018). BZR1 Transcription Factor Regulates Heat Stress Tolerance through FERONIA Receptor-like Kinase-Mediated Reactive Oxygen Species Signaling in Tomato. *Plant Cell. Physiology* 59 (11), 2239–2254. doi:10.1093/pcp/pcy146
Yu, Q.-h., Wang, B., Li, N., Tang, Y., Yang, S., Yang, T., et al. (2017). CRISPR/Cas9-induced Targeted Mutagenesis and Gene Replacement to Generate Long-Shelf Life Tomato Lines. *Sci. Rep.* 7, 11874. doi:10.1038/s41598-017-12262-1
Yu, Y., Leete, T. C., Born, D. A., Young, L., Barrera, L. A., Lee, S. J., et al. (2020). Cytosine Base Editors with Minimized Unguided DNA and RNA Off-Target Events and High On-Target Activity. *Nat. Commun.* 11, 2052. doi:10.1038/s41467-020-15887-5
Zaidi, S. Se. A., Mahas, A., Vanderschuren, H., and Mahfouz, M. M. (2020). Engineering Crops of the Future: CRISPR Approaches to Develop Climate-Resilient and Disease-Resistant Plants. *Genome Biol.* 21, 289. doi:10.1186/s13059-020-02204-y
Zeng, Y., Wen, J., Zhao, W., Wang, Q., and Huang, W. (2020). Rational Improvement of Rice Yield and Cold Tolerance by Editing the Three Genes OsPIN5b, GS3, and OsMYB30 with the CRISPR-Cas9 System. *Front. Plant Sci.* 10, 1663. doi:10.3389/fpls.2019.01663
Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell.* 163 (3), 759–771. doi:10.1016/j.cell.2015.09.038
Zhai, Y., Yu, K., Cai, S., Hu, L., Amoo, O., Xu, L., et al. (2020). Targeted Mutagenesis of BnTT8 Homologs Controls Yellow Seed Coat Development for Effective Oil Production in Brassica Napus L. *Plant Biotechnol. J.* 18 (5), 1153–1168. doi:10.1111/pbi.13281
Zhang, R., Liu, J., Chai, Z., Chen, S., Bai, Y., Zong, Y., et al. (2019). Generation of Herbicide Tolerance Traits and a New Selectable Marker in Wheat Using Base Editing. *Nat. Plants* 5, 480–485. doi:10.1038/s41477-019-0405-0
Zhang, Y., Bai, Y., Wu, G., Zou, S., Chen, Y., Gao, C., et al. (2017). Simultaneous Modification of Three Homoeologs of TaEDR1 by Genome Editing Enhances Powdery Mildew Resistance in Wheat. *Plant J.* 91 (4), 714–724. doi:10.1111/tpj.13599
Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., et al. (2016). Efficient and Transgene-free Genome Editing in Wheat through Transient Expression of CRISPR/Cas9 DNA or RNA. *Nat. Commun.* 7, 1–8. doi:10.1038/ncomms12617
Zhou, X., Liao, H., Chern, M., Yin, J., Chen, Y., Wang, J., et al. (2018). Loss of Function of a Rice TPR-Domain RNA-Binding Protein Confers Broad-Spectrum Disease Resistance. *Proc. Natl. Acad. Sci. U. S. A.* 115, 3174–3179. doi:10.1073/pnas.1705927115
Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., et al. (2017). Precise Base Editing in Rice, Wheat and Maize with a Cas9-Cytidine Deaminase Fusion. *Nat. Biotechnol.* 35, 438–440. doi:10.1038/nbt.3811

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Nerkar, Devarumath, Purankar, Kumar, Valarmathi, Devarumath and Appunu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Frontiers in Genetics

TYPE Review
PUBLISHED 23 August 2022
DOI 10.3389/fgene.2022.876987

Check for updates

OPEN ACCESS

EDITED BY
Jitendra Kumar,
Indian Institute of Pulses Research (ICAR), India

REVIEWED BY
Debjyoti Sen Gupta,
Indian Institute of Pulses Research (ICAR), India
Kumar Paritosh,
University of Delhi, India
Guoliang Yuan,
Oak Ridge National Laboratory (DOE), United States
Faiz Ahmad Joyia,
University of Agriculture, Pakistan

*CORRESPONDENCE
Vijay Rani Rajpal,
vijayrani2@gmail.com
Soom Nath Raina,
soomr@yahoo.com

SPECIALTY SECTION
This article was submitted to Plant Genomics,
a section of the journal Frontiers in Genetics

RECEIVED 16 February 2022
ACCEPTED 29 June 2022
PUBLISHED 23 August 2022

CITATION
Dhakate P, Sehgal D, Vaishnavi S, Chandra A, Singh A, Raina SN and Rajpal VR (2022), Comprehending the evolution of gene editing platforms for crop trait improvement.
Front. Genet. 13:876987.
doi: 10.3389/fgene.2022.876987

COPYRIGHT
© 2022 Dhakate, Sehgal, Vaishnavi, Chandra, Singh, Raina and Rajpal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Comprehending the evolution of gene editing platforms for crop trait improvement

Priyanka Dhakate[1], Deepmala Sehgal[2], Samantha Vaishnavi[3], Atika Chandra[4], Apekshita Singh[5], Soom Nath Raina[5]* and Vijay Rani Rajpal[6]*

[1]National Institute of Plant Genome Research, Aruna Asaf Ali Marg, New Delhi, India, [2]International Maize and Wheat Improvement Center (CIMMYT), México-Veracruz, Mexico, [3]Department of Botany, Central University of Jammu, Jammu, India, [4]Department of Botany, Maitreyi College, University of Delhi, New Delhi, India, [5]Amity Institute of Biotechnology, Amity Institute of Biotechnology, Amity University, Noida, India, [6]Department of Botany, Hansraj College, University of Delhi, New Delhi, India

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas (CRISPR-associated) system was initially discovered as an underlying mechanism for conferring adaptive immunity to bacteria and archaea against viruses. Over the past decade, this has been repurposed as a genome-editing tool. Numerous gene editing-based crop improvement technologies involving CRISPR/Cas platforms individually or in combination with next-generation sequencing methods have been developed that have revolutionized plant genome-editing methodologies. Initially, CRISPR/Cas nucleases replaced the earlier used sequence-specific nucleases (SSNs), such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), to address the problem of associated off-targets. The adaptation of this platform led to the development of concepts such as epigenome editing, base editing, and prime editing. Epigenome editing employed epi-effectors to manipulate chromatin structure, while base editing uses base editors to engineer precise changes for trait improvement. Newer technologies such as prime editing have now been developed as a "search-and-replace" tool to engineer all possible single-base changes. Owing to the availability of these, the field of genome editing has evolved rapidly to develop crop plants with improved traits. In this review, we present the evolution of the CRISPR/Cas system into new-age methods of genome engineering across various plant species and the impact they have had on tweaking plant genomes and associated outcomes on crop improvement initiatives.

KEYWORDS

CRISPR/Cas system, base editing, prime editing, epigenome editing, crop improvement

1 Introduction

Over the past decade, the gene-editing platforms have shown tremendous evolution to accommodate the dual concerns of biosafety of edited crops and the efficiency of the platform used. Efficient and rapid genomic sequencing platforms have facilitated a better understanding of plant genomes, particularly when used in conjunction with genome editing (GE). Restructuring genomes *via* introduction of heritable genomic changes for expressing desirable quality traits in crops has been the focus of research for decades. The primitive methods of genome restructuring involved the use of genotoxic agents to introduce random double-stranded breaks (DSB) that were subsequently repaired by inherent non-homologous end joining (NHEJ) pathways resulting in random mutations (Puchta, 2005). After decades of usage of these random mutations generating tools, GE platforms have gone through many phases of improvement over the years. For example, the discovery of sequence-specific nucleases (SSNs) such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) helped to engineer the genome at intended loci by mediating the cleavage of dsDNA. The use of these nucleases induced the native NHEJ pathway for DNA repair (Salomon and Puchta, 1998). This method of GE, however, is both cost- and labor-intensive as it requires the development of sequence-specific nucleases/proteins. In addition, GE using these nucleases was inefficient as unintended off-target edits were introduced by the induction of the error-prone NHEJ repair pathway.

Given the obvious limitations of ZFNs and TALENs, the vacuum was soon filled with the discovery of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas (CRISPR-associated) nucleases. In prokaryotes, the CRISPR/Cas system exists as a means of endogenous small RNA-based adaptive defense mechanism that protects the host bacterial cell *via* sequence-specific recognition and targeted cleavage of viral DNA (Jinek et al., 2012). With an approximate length of 32 bp, the length of CRISPR repeat sequences varies between 21 and 47 bp across prokaryotes. Every CRISPR repeat sequence harbors a unique sequence that is specific to the bacterial species processing it and has, therefore, been conserved over the course of evolution (Karginov and Hanon 2010). CRISPR was first discovered by a Japanese group in 1987 while studying the *iap* gene from the *E. coli* genome (Ishino et al., 1987). They identified CRISPR as homologous repeated sequences of only a few nucleotides interspersed by spacer sequences. Following this, CRISPRs were reported from the archaeal genome, *Haloferax mediterranei* (Mojica et al., 1993). However, the prodigious potential of the CRISPR/Cas9 as a GE platform was discovered just a decade ago (Jinek et al., 2012). To employ this tool, a customized small guide RNA (gRNA) is designed to identify the intended target and guide the associated Cas9 protein to introduce DSBs in the target genomic DNA. Indels are introduced at the target site as the repair pathway *via* NHEJ is triggered. Over the course of evolution of the platform, new variants of Cas proteins have been mobilized to increase the efficiency of the CRISPR/Cas9-mediated GE.

During the past decade, the term "CRISPR/Cas" has evolved into a synonym for GE following which off-targeting instances with the use of CRISPR/Cas systems have reduced manifold (Modrzejewski et al., 2020). However, the goal of achieving "no off-target" remains elusive. In addition, with the involvement of the NHEJ repair pathway, the efficiency of this platform has always been disputable. In the third phase of the evolution of GE platforms, the CRISPR/Cas platform evolved to target the epigenome of an organism which was termed epigenome editing (Konermann et al., 2013). In epigenome editing, chromatin modification at specific genomic loci involves the use of epi-effectors that are comprised of DNA recognition domains (ZFNs, TALENs, or CRISPR/Cas system) and catalytic domains from a chromatin-modifying enzyme. Epigenome editing has been slated to have promising results in numerous basic sciences to decipher functions of chromatin structure and associated modification in phenotypes.

In the fourth phase, the CRISPR/Cas system evolved into a new methodology called base editing, wherein RNA-guided endonucleases were employed to engineer all four possible transitions with increased precision (Komor et al., 2016). One of the major challenges that all of the aforesaid techniques still face is to simultaneously engineer the altered DNA at the intended target sites. These concerns were addressed with the introduction of prime editing, marking the fifth phase in the evolution of GE platforms. Prime editing is largely described as a "search-and-replace" technology that edits the intended genomic loci without generating DSBs (Anzalone et al., 2019). This platform efficiently addresses the concerns of frameshift mutations that arise with the introduction of indels, further reducing off-target mutations. In addition, prime editing can introduce all 12 possible nucleotide substitutions (including transversions and transitions) (Anzalone et al., 2019).

The availability of all new-age GE strategies has not stolen the thunder of the CRISPR/Cas platform owing to the ease of its use and relevance to editing genes in numerous crop plants. However, it is only a matter of time before rapidly changing GE methods will replace present-day CRISPR/Cas systems with more elegant and efficient platforms. With every refinement of the platform, we are getting only closer to generating precise introduced mutations/deletions with reduced off-target effects. In the present review, we evaluate the evolution of GE platforms, such as CRISPR/Cas, epigenome editing, base editing, and prime editing over the last decade to highlight the paradigm shift in our understanding of GE strategies and the relevance of these platforms in present-day agriculture.

2 Genome editing using zinc-finger nucleases and transcription activator-like effector nucleases

ZFNs and TALENs represent the first phase of the development of GE platforms. Essentially GE is achieved *via* the introduction of DSBs followed by a homologous repair

pathway or the NHEJ-DNA repair pathway. In the first phase of developing GE platforms, SSNs such as ZFNs and TALENs were employed to introduce heritable genomic changes. ZFNs are chimeric enzymes that work as a dimer. Each monomer has 3–5 zinc-finger repeats along with a *FokI* cleavage domain. Each of the zinc fingers is capable of recognizing 3 bp of genomic DNA. Therefore, a ZFN dimer can effectively identify an 18–30 bp DNA with a gap of 5–7 bp (Kim et al., 2007). In plants, the first study involving ZFNs was reported in *Arabidopsis*, wherein heat shock was found to augment ZFN expression. At least 10% of the transgenics obtained displayed the mutations induced by ZFNs in future generations (Lloyd et al., 2005). In maize, ZFNs were employed to introduce a DSB at ipk1, and following this, a herbicide tolerance gene was inserted that resulted in transgenics showing tolerance to herbicide (Shukla et al., 2009). One of the major disadvantages of ZFNs is that the zinc fingers could overlap and are largely dependent on the sequence context around them and the intended DNA segment. Therefore, employing ZFNs becomes both labor- and cost-intensive as for every edit, the zinc-finger array is designed, and the sites available for the edits are limited (Boch and Bonas 2010). Although many studies have reported ZFNs to edit genes, its use as a tool of choice for GE now stands outdated. Another type of nucleases, TALENs, with DNA binding domains, was also employed to engineer genomic changes (Boch and Bonas 2010). Thirty-four tandem repeats are typically present in the DNA binding domain along with repeat-variable di-residue (RVD) comprised of two amino acids at positions 12 and 13, providing the TALENs with the ability to identify the intended target DNA sequence (Cong et al., 2012; Streubel et al., 2012). Like ZFNs, TALENs also introduce DSBs in the intended genomic DNA sequences, completely disrupting the gene and (or) introducing mutations. In comparison to ZFNs, TALENs can be designed for more target sites in the genomic DNA (Boch and Bonas 2010). In rice, TALENs were used to mutate the *OsSWEET* gene to develop transgenic resistance to blight (Li et al., 2012). Similarly, in wheat, transgenic with increased resistance to powdery mildew was developed by employing TALENs induced mutations (Wang et al., 2014). In cabbage, early flowering plants were obtained by employing TALENs (Sun et al., 2013). Like ZFNs, using TALENs is cost- and labor-intensive with limited success, and therefore, their use has now been largely suspended for introducing genomic changes.

3 Clustered regularly interspaced short palindromic repeats/Cas system-mediated genetic modification

The CRISPR/Cas systems represent the second phase of evolution in the development of GE platforms. CRISPR/Cas systems are sequence-specific and, therefore, mediate targeted DNA cleavage with increased efficiency. Three major steps are involved in CRISPR/Cas mechanism. The first step is adaptation, wherein a small sequence from the mobile genetic elements (MGEs) is harbored into the host CRISPR resulting in a novel spacer sequence. This adaptive event helps the host bacterial cell evade the attack from the same virus in the future (Barrangou et al., 2007). The selection of the target sequence to be incorporated into the CRISPR array is sequence-specific. In type I, II, and V CRISPR/Cas systems, a small sequence, termed the protospacer adjacent motif (PAM), is found adjacent to the protospacer that is to be incorporated into the CRISPR array. Therefore, PAM is cardinal to both acquiring the protospacer and bringing about the subsequent interference (Datsenko et al., 2012; Zetsche et al., 2015; Fonfara et al., 2016). Although the acquisition mechanism of spacers is not yet fully deciphered, in almost all CRISPR/Cas systems, Cas1 and Cas2 proteins have been found to maneuver the acquisition of the spacer into the CRISPR array (Makarova et al., 2015; Shmakov et al., 2015). Both these proteins are found to be necessary for the acquisition of the spacer (Datsenko et al., 2012). The two proteins form a hetero-hexameric protein complex (Cas1–Cas2), which is central to both excision and incorporation of the protospacer DNA into the CRISPR array (Nuñez et al., 2014). Barring a few exceptions, invariably the spacers are chronologically added to the array (Shmakov et al., 2015). Cas1–Cas2 protein complex is central to protospacer acquisition across most type 1 and type II CRISPR/Cas systems. Therefore, this mode of spacer acquisition stands most well deciphered so far. In the second step, the CRISPR array is transcribed and processed. In addition, the associated *Cas* genes are also transcribed into crRNAs. This step is subtype-specific, and therefore, subtype-specific enzymes are employed. However, broadly across all CRISPR/Cas systems, the CRISPR array is first transcribed into a precursor crRNA (pre-crRNA). Different Cas proteins and ribonucleases cleave and process this in various types of CRISPR/Cas systems to yield a mature crRNA. In the third step, following infection, the mature crRNAs mediate subtype-specific machinery driven mostly *via* Cas proteins to ensure effective cleavage of the MGE. The mechanism of different Cas proteins employed in various CRISPR/Cas systems has been well documented in many studies (Liu L et al., 2020; Talakayala et al., 2022; Wada et al., 2022).

4 Classification of the clustered regularly interspaced short palindromic repeats/Cas system

The classification of the CRISPR/Cas systems identified so far is primarily based on the presence of the effector Cas proteins that cleave the invading foreign nucleic acids. The primary classification divides these systems into two classes: Class 1 and Class 2. Class 1 CRISPR/Cas systems employ a multi-

protein complex, and Class 2 CRISPR/Cas systems recruit a single effector protein. Further, classification of Class 1 and Class 2 CRISPR/Cas systems into subtypes (I through VI) is dependent on their mechanism of action. The effector module of the CRISPR/Cas system is divided into three stages: the adaptation stage, the expression and processing stage, and the interference stage. In class 1 CRISPR/Cas systems (with types I, III, and IV), type I and type III systems employ a multi-protein complex called the Cascade complex along with Cas3 nuclease-helicase and the Cmr complex for type I, type III-A, and type IIIB CRISPR/Cas systems, respectively (Koonin and Makarova, 2019; Chaudhuri et al., 2022). However, class 2 CRISPR/Cas systems (with types II, V, and VI) employ only one effector protein. In type II and type V CRISPR/Cas systems, the expression and processing of the crRNA are regulated by a single protein such as Cas9 and Cpf1, respectively (Makarova et al., 2015; Amitai and Sorek, 2016). Type VI systems have been recently discovered and are the only CRISPR/Cas systems to target RNA specifically (Chaudhuri et al., 2022). In Class 1 CRISPR Cas systems, type 1 and type III are more prevalent than type IV in diverse bacterial and archaeal populations. However, type II of the Class 2 CRISPR/Cas system is found across all bacterial species (Koonin and Makarova, 2019). Depending on their function, Cas proteins can be primarily classified into four categories; recombinases/nucleases that aid the acquisition of spacers, ribonucleases that regulate the processing of crRNAs, scanning complexes like the crRNP complex, and nucleases that mediate the cleavage of the intended target sequences (Van Der Oost et al., 2014).

Class 1 CRISPR systems, types I and III, bear structural similarities suggesting evolution *via* a common ancestor (Chaudhuri et al., 2022). In addition, they employ Cas9 endonuclease to process crRNA. Type I CRISPR/Cas systems are further divided into six subtypes, types I-A, I-B, I-C, I-D, I-E, and I-F, depending on the distinct PAMs that the subunits require to regulate recognition and acquisition. The type III systems are divided into four subtypes, type III-A, III-B, III-C, and III-D, based on variation in adaptation, recognition, and interference modules of the effector protein complex. Chaudhuri et al. (2022) discussed the further classification of type I and type III into subtypes at length. Class 2 CRISPR/Cas system is divided into three types, types II, V, and VI. Out of these, the type II system is the most dissected and well-understood system so far (Koonin and Makarova, 2019; Chaudhuri et al., 2022). This system employs the Cas9 endonuclease as the effector. Type V system uses a single effector protein, Cas12. However, Cas12 has six subtypes, types V-A, V-B, V-C, V-D, V-E, and V-U, that identify distinct PAM sequences (Chaudhuri et al., 2022). Owing to obvious advantages such as smaller size, no dependency on tracr for target recognition, and asymmetric cleavage sites, Cas12 has now been actively replacing the Cas9 system for GE in many animal and plant species. Type VI systems are characterized by the presence of higher eukaryotes and prokaryotes nucleotide-binding (HEPN) domains with RNase activity (Koonin and Makarova, 2019; Chaudhuri et al., 2022). Cas13a was the first protein identified for type VI CRISPR/Cas systems (Chaudhuri et al., 2022). The evolution of type VI-B, such as Cas13b, is thought to have occurred from transmembrane systems, making them unique from type VI systems into a new subtype type VI-B (Chaudhuri et al., 2022). Type VI systems only target RNAs, thus thought to have lower instances of off-targeting and, in turn, do not harm the host cell much. The extensive diversity of the CRISPR/Cas system, as evident by their classification, reflects the evolution of the CRISPR/Cas-based defense mechanism in both archaea and bacteria. In addition, this diversity of CRISPR/Cas systems presents researchers with varied tools of GE to introduce precise changes with efficacy. Table 1 summarizes the classification of the CRISPR/Cas systems identified so far.

5 Repurposing native clustered regularly interspaced short palindromic repeats/Cas9 for the development of genome-editing platforms

Class II CRISPR/Cas systems were found to be most suitable for development into a tool for genetic manipulation owing to the simplicity of their mechanism of action (Makarova et al., 2015). Type II CRISPR/Cas systems employ Cas9 protein that relies only on an RNA complex of crRNA:tracrRNA that is easy to engineer into a single guide DNA (gDNA) molecule (Jinek et al., 2012). These systems employ only two components: Cas9, a DNA endonuclease, and a customizable gRNA. A single gRNA is sufficient to direct the cleavage of the intended sequences. The gRNA molecules are customized to contain a sequence that Cas9 recognizes and a target sequence that guides the complex to the intended locus (Anders et al., 2014). To identify the intended target site, the Cas9-sgRNA complex scans the targeted DNA for a PAM site, following this 12 bases (seed region) of gRNAs proximal to PAM pair with the intended target sequence (Semenova et al., 2011). Mismatches in the seed region have been found to affect the activity of Cas9 adversely. However, mismatches in the 5’ PAM distal region are well-tolerated without affecting Cas9 nuclease activity (Liu et al., 2016). Catalytic domains of Cas9, HNH, and RuvC invariably result in a DSB in the DNA. Following this, DSB repair is initiated that is mediated either by homology direct repair (HDR) or the NHEJ pathway. The latter does not require a template for DNA repair and hence is error-prone. NHEJ is the active DNA repair mechanism in nature wherein Cas9-induced DSBs are repaired (Moore and Haber 1996). NHEJ can, therefore, lead to small insertions or deletions that could yield a host of mutations (Calvache et al., 2022; Wada et al., 2022). Such mutations are beneficial while knocking out a targeted gene

TABLE 1 Classification of the identified CRSIPR-Cas systems.

Class	Type	Effector module	Class	Type	Effector module
Class I	I-A	Cas8a2, Csa5	Class II	V-B	Cas12b
Class I	I-B	Cas8b	Class II	V-C	Cas12c
Class I	I-C	Cas8c	Class II	V-D	Cas12d
Class I	I-D	Cas10d	Class II	V-E	Cas12e
Class I	I-E	Cse1, Cse2	Class II	V-F	Cas14
Class I	I-F	Csy1, Csy2, Csy3, Cas6f	Class II	V-G	Cas12g
Class II	II-A	Csn2	Class II	V-H	Cas12h
Class II	II-B	Cas9 (Csx12 subfamily)	Class II	V-I	Cas12i
Class II	II-C	N/A	Class II	V-J	Cas12j
Class I	III-A	Csm2 (small subunit)	Class II	V-K	Cas12k
Class I	III-B	Cmr5 (small subunit)	Class II	VI-A	Cas13a
Class I	IV	DinG (Csf4)	Class II	VI-B	Cas13b, along with proteins, Csx27, and Csx28
Class II	V-A	Cas12a (previously known as Cpf1)	Class II	VI-C	Cas13c
			Class II	VI-D	Cas13d

using CRISPR/Cas9 systems. However, being random and unpredictable makes this mode of DNA repair unsuitable for precise editing of intended genes. To this effect, HDR is a more obvious choice of DSB repair mechanism for incorporation of desired sequences following cleavage by Cas9. In plants, GE HDR relies on a DNA template along with the gDNA and Cas9 for a successful DSB repair (Calvache et al., 2022; Wada et al., 2022). In plants, through genetic engineering, many outstanding repairs have been achieved *via* HDR, leading to gene replacement, DNA correction, and targeted knockouts. Figure 1 illustrates a diagrammatic representation of the adaptation to the CRISPR/Cas9 system in plants for gene editing.

6 Applications of clustered regularly interspaced short palindromic repeats/Cas9 system as a powerful tool in crop improvement

Present-day agriculture faces serious threats from both abiotic and biotic stresses. Rapidly changing climate and exponentially growing world population increase the pressure of ensuring food security for both present and future generations. To mitigate agricultural losses and to aid crops in realizing their full potential, the only sustainable solution is to develop climate-resilient crops. Since its discovery in 2012 as a potential tool for genetic engineering, CRISPR/Cas9 system and its derivatives have rapidly replaced genome engineering methods in crop improvement programs across the globe. In model crops such as maize, a CRISPR/Cas9 mediated knocking and replacement in the *liguleless-1* (*LIG1*) was reported (Svitashev et al., 2016). Similarly, in wheat, CRISPR/Cas9 GE system was employed to introduce targeted mutations in two wheat genes, *TaLox2* and *TaUbiL1*. This study also validated the efficiency of using the CRISPR/Cas9 system in combination with microspore technology in plants for both trait improvement and discovery (Bhowmik et al., 2018). In tomato, complete expression of the susceptibility gene SlyPMR4 was knocked down to generate tomato plants with resistance against powdery mildew (Martínez et al., 2020). CRISPR/Cas9-based GE systems are now employed to improve multigenic traits such as biotic and abiotic stresses in many crops. Table 2 summarizes studies wherein CRISPR/Cas has been used successfully for trait manipulation in crop plants. Figure 2 depicts schematic representation of the domains of crop sciences wherein CRISPR/Cas platforms have largely contributed.

One of the most important applications of CRISPR/Cas9 platforms across the globe has been to engineer disease resistance in crop plants. Plant pathogens such as bacteria, viruses, nematodes, insects, and fungi are the most potent biotic stress factors that impact the yield potential of crops across the globe. Continuously evolving new strains of lethal pests make the battle against the pathogens even more complicated and daunting (Razzaq et al., 2019). Therefore, to protect and aid crops, methodologies routed in concepts of genome engineering have been successfully developed (Jaganathan et al., 2018). Peng et al. (2017) reported the development of varieties of *Citrus sinensis* (Wanjincheng orange) with increased resistance to *Xanthomonas citri*, which is responsible for the citrus canker disease in oranges. In this study, the expression of the gene, *CsLOB1*, which is responsible for the development of the disease, was disrupted using the CRISPR/Cas9 system. Two alleles (*cslob1g* and *cslob1*) exist for the gene CsLOB1. The promoter region of both these alleles inhibits an effector binding site (EBE) that is recognized by the main effector PthA4 of Xcc to drive the expression of *cslob1* and results in the development of the disease. Five

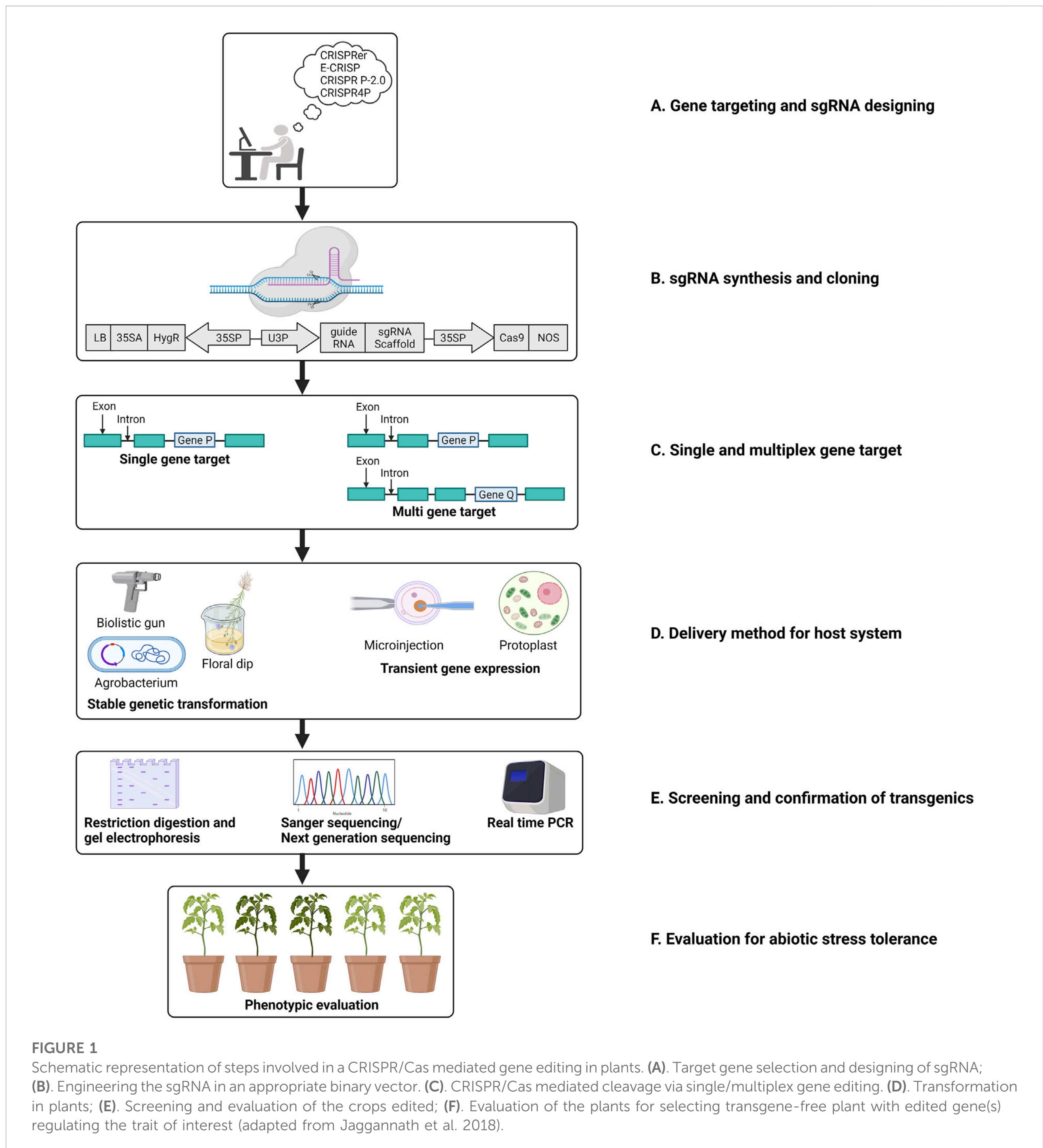


FIGURE 1
Schematic representation of steps involved in a CRISPR/Cas mediated gene editing in plants. **(A)**. Target gene selection and designing of sgRNA; **(B)**. Engineering the sgRNA in an appropriate binary vector. **(C)**. CRISPR/Cas mediated cleavage via single/multiplex gene editing. **(D)**. Transformation in plants; **(E)**. Screening and evaluation of the crops edited; **(F)**. Evaluation of the plants for selecting transgene-free plant with edited gene(s) regulating the trait of interest (adapted from Jaggannath et al. 2018).

independent constructs pCas9/CsLOB1sgRNA were employed to modify the effector binding site EBE in the promoter region of *CsLOB1* alleles. Homologous mutants wherein the EBE was completely disrupted were obtained, displaying no disease development following infection with *Xanthomonas citri* (Peng et al., 2017). In rice, an ethylene-responsive gene OsERF922 was knocked out using the CRISPR/Cas9 tool, which led to a marked reduction in the size and number of the blast lesions. This work led to the development of a rice cultivar with increased resistance against *Magnaporthe oryzae* (Wang et al., 2016). In another study, blight-resistant plants were produced using CRISPR/Cas9 system-mediated targeted mutagenesis of the *SWEET13* gene (Zhou et al., 2015).

Management of diseases in crop plants is dominated by the frequent use of insecticides to curb yield losses. The development of crops resistant to viruses is, therefore, an efficient strategy to

TABLE 2 CRISPR/Cas9-mediated improvement in major crop plants.

Plant species	Target gene	Trait of interest	References
Rice (*Oryza sativa*)	*OsAAP6*, *OsAAP10*	Reduced GPC	Wang M et al. (2020)
	OsBADH2	Fragrant rice	Kumar et al. (2021)
	eIF4G	Resistance to tungro spherical virus	Macovei et al. (2018)
	OsGAD3	Increased GABA content	Akama et al. (2020)
	CrtI, *PSY*	Increased β-carotene content	Dong et al. (2020)
	OsGS3, *OsGW2*, and *OsGn1a*	Increased grain length and width	Zhou et al. (2019)
	OsDST	Increased drought and salt tolerance	Kumar et al. (2021)
	OsPIN5b, *GS3*, and *OsMYB30*	Increased yield and cold tolerance	Zeng et al. (2020)
	OsPLDα1	Low phytic acid content	Khan et al., 2019
Wheat (*Triticum aestivum*)	*TaGW7*	Grain shape	Wang et al. (2019)
	EDR1	Resistant to powdery mildew	Zhang et al. (2017)
	TaGW2	Grain size	Wang X et al. (2018)
	α-Gliadin genes	Low gluten content	Sanchez et al. (2018)
	TaBAK1-2, *a-eIF4E*, *Ta-eIF(iso)4E*	Resistance to streak mosaic virus and yellow mosaic virus	Hahn et al., 2021
	TaSBEIIa	Grain quality	Li G et al. (2021)
	TaNP1	Male sterility	Li et al. (2020b)
Maize (*Zea mays*)	*SH2*, *GBSS*	Super sweet and waxy corn	Dong et al. (2019)
	Wx1	Waxy corn	Gao et al. (2020)
	ZmBADH2a, *ZmBADH2b*	Aromatic maize	Wang Z et al. (2021)
	CLE genes	Enhanced grain yield	Liu et al. (2021)
	GA20ox3	Semi-dwarf male plants	Zhang C et al. (2020)
Tomato (*Solanum lycopersicum*)	*ANT1*	Fruit color (purple)	Čermák et al. (2015)
	CLV3	Fruit size	Zsögön et al., 2020
	Psy1, *CrtR-b2*	Fruit color (yellow)	D'Ambrosio et al. (2018)
	OVATE, *Fas*, *Fw2.2*	Fruit size, oval fruit shape	Zsogon et al. (2018)
	ENO	Fruit size	Yuste-Lisbona et al. (2020)
	CRTISO	Fruit color (tangerine)	Ben Shlush et al. (2021)
	slyPDS	Increased lycopene content	Li J et al. (2018)
	SlNPR1	Increased drought tolerance	Li et al. (2019)
	SlCBF1	Increased cold tolerance	Li R et al. (2018)
	SlMAPK3	Increased drought tolerance	Wang et al. (2017)
	miR482b and *miR482c*	Resistance to *Phytophthora infestans*	Hong et al. (2021)
	SlyPMR4	Resistance against powdery mildew	Martínez et al. (2020)
	PL, *PG2a*, *TBG4*	Longer shelf life	Wang et al. (2019)
	SlLBD40	Enhanced drought tolerance	Liu et al. (2020)
Rapeseed (*Brassica napus*)	*BnaFAD2*	Improved fatty acid profile	Huang et al. (2020)
	BnaMAX1	Improved plant architecture and yield	Zeng et al. (2020)
	BnaA03.BP	Compact plant architecture	Fan et al. (2021)

yield a stable yet economically viable alternative (Wang W et al., 2021). To this effect, inducing deletions and introducing point mutations in the genes using the CRISPR/Cas9 system is one of the most organic adaptations of the platform. The eukaryotic translation initiation factor genes such as *eIF4E* and *eIF4G* are an absolute requirement for the translation of RNA viruses (Shopan et al., 2020). Therefore, CRISPR/Cas9 technology has been employed in numerous plant species to engineer induced mutations in these genes. In *Arabidopsis*, point mutations in *eIF(iso)4E* gene were found to impart complete resistance against the turnip mosaic virus (Pyott et al., 2016). Likewise, in cucumber, eukaryotic translation initiation factor *eIF(iso)4E* was engineered using the CRISPR/Cas9 system to generate heritable homozygous point mutations that conferred resistance to the mutants against zucchini yellow mosaic virus, papaya ringspot mosaic virus-W, and vein yellowing virus (Chandrasekaran et al., 2016). In *Nicotiana benthamiana*, sgRNA/Cas9-mediated broad-spectrum immunity was

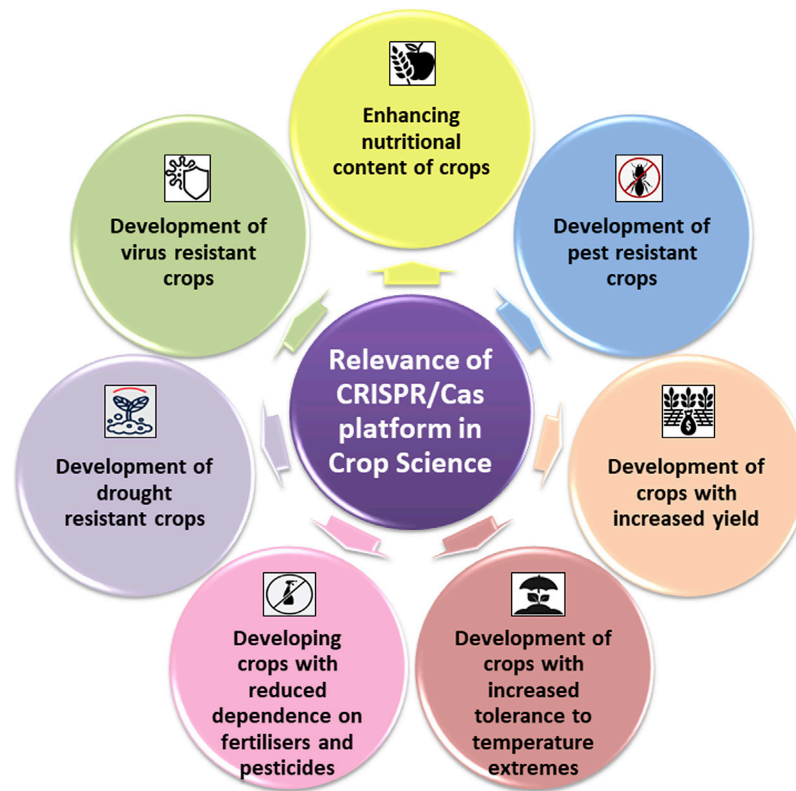


FIGURE 2
Schematic representations of the domains of crops sciences wherein CRISPR/Cas platforms have largely contributed.

achieved against viruses such as beet curly top virus, *Tomato leaf curl Sardinia virus*, *Tomato yellow leaf curl virus*, and *Cotton leaf curl Kokhran virus* (Ali et al., 2016). In rice, the CRISPR/Cas9 system was used to generate *eIF4G* alleles that conferred resistance against the R*ice tungro spherical virus* (Macovei et al., 2018). Recently, Wang et al. (2021) employed the CRISPR/Cas9 system to generate novel *eIF4G* alleles to yield transgenic plants displaying complete resistance to rice black-streaked dwarf virus. Engineering these mutations *via* the traditional backcrossing would have taken years, but using the CRISPR/Cas9 system expedited the process, and the goal was achieved in just a single generation.

The CRISPR/Cas9 system has also been used extensively over the past decade in generating climate-resistant cultivars in various crop species such as cotton, maize, rice, wheat, potato, soybean, and tomato (Khan et al., 2021; Wang et al., 2021; Rahman et al., 2022). In wheat, two regulatory genes (i.e., *TaDREB3* and *TaDREB2*) were mutated using the CRISPR/Cas9 system, which resulted in increased drought tolerance in the mutated plants in comparison to the wild cultivars (Kim et al., 2017). In maize, the *ZmARGOS8* gene that negatively regulates ethylene response was studied using the CRISPR/Cas9 system. The promoter of this gene was knocked out and replaced with maize GOS2 promoter in 5′-UTR of the target gene. The mutant plants were found to overexpress *ARGOS8*, which led to a stupendous increase in the yield in comparison to the wild type under drought conditions during the flowering stage without any yield penalty under irrigated environment (Shi et al., 2017). In rice, the CRISPR/Cas9 system was used to knock out gene *OsRR2*. The homozygous mutants obtained displayed increased tolerance to salinity stress (Zhang et al., 2019). In another study, three genes, *OsPIN5b*, *GS3*, and *OsMYB30*, that determine panicle length, grain size, and cold tolerance, respectively, were simultaneously edited using the CRISPR/Cas9 system (Zeng et al., 2020). T2 generations of the homozygous mutants of these genes displayed increased panicle length, enlarged grain size, and increased cold tolerance, respectively. The CRISPR/Cas9 tool has also been employed for the functional characterization of genes that regulate stress responses in plants. In *Arabidopsis*, three genes (*CBF1*, *CBF2*, and *CBF3*) have been identified to confer cold acclimatization and tolerance. However, the underlying mechanism remained undeciphered owing to the absence of any loss-of-function lines for these genes. Zhao et al. (2016) generated mutants of the cbf gene family, *cbf1*, *cbf2*, and *cbf3*. They generated cbf single, double, and triple mutants using the CRISPR/Cas9 platform. Interestingly, for the three genes, *cbf* triple mutants displayed compromised seedling development and

reduced salt tolerance. However, both triple and double (*cbf2cbf3*) mutants displayed increased sensitivity to feeding post-cold acclimatization in comparison to the wild-type control. The *cb1/cb3* double mutants displayed increased resistance, indicating that accumulation of *CBF2* is more important than *CBF1* and *CBF3* in regulating cold acclimation-dependent freezing tolerance. The functional role of many other genes with a potential role in stress tolerance was also investigated in the model system *Arabidopsis*. The expression of *UGT79-B2* and *B3* genes was induced by abiotic stresses such as salinity, drought, and cold. Overexpression of these genes was found to increase the resistance of the transgenics. However, gene *ugt79b2/b3* double mutants generated using the CRISPR/Cas9 system were found to be susceptible to abiotic stresses compared to the wild-type control. The overexpression mutants accumulated anthocyanins, but the *ugt79b2/b3* double mutants that displayed lower levels of anthocyanins were also found to be more susceptible to stresses than the wild-type control plants. These findings also suggested that an array of anthocyanins impart resistance against abiotic stresses (Li et al., 2017). In rice, knockout mutants for the *OsSAPK2* gene were developed for functional characterization of the gene. The mutants showed insensitivity to abscisic acid and increased sensitivity to drought and reactive oxygen species (ROS) during the germination/seedling stage compared to the wild-type control plants. These results suggested the active involvement of the *OsSAPK2* gene in mediating drought tolerance through increased stomatal closure (Lou et al., 2017). In another study, *OsAnn3*, a rice annexin gene, was knocked out in rice using the CRISPR/Cas9 system. The survival ratio of T1 mutant lines was found to be adversely affected, indicating that the expression of *OsAnn3* was central in imparting cold tolerance in rice (Shen et al., 2017).

Drought stress in plants is governed by mitogen-activated protein kinases (MAPKs). In tomato, functional characterization of MAPKs was achieved by knocking down *SlMAPK3* using the CRISPR/Cas9 system (Wang et al., 2017). The resulting *slmapk3* mutants displayed severe wilting symptoms along with lower antioxidant enzymes, increased hydrogen peroxide, and increased membrane damage in comparison to the wild-type control. In another study, a multiplex CRISPR/Cas9 system was used simultaneously to edit five tomato γ-aminobutyric acid (GABA) shunt genes (*CAT9*, *SSADH*, *GABA-TP1*, *TP2*, and *TP3*). These genes are repressors of GABA metabolism. Hence, targeted mutagenesis of these genes led to a 19-fold increase in the accumulation of GABA in fruits and leaves (Li R et al., 2017).

The multiplex CRISPR/Cas9 system has proven to be beneficial in improving yield substantially in various cereal crops. In rice, four genes [i.e., Grain Size 3 (*GS3*), Ideal Plant Architecture 1 (*IPA1*), Grain Number 1a (*Gn1a*), and DENSE AND ERECT PANICLE (*DEP1*)] were edited using the multiplex CRISPR/Cas9 technique. The mutant plants displayed marked improvement in all the aforesaid traits and resulted in better and improved yields concerning tiller number and grain yield (Li et al., 2016). Similarly, multiplex editing using the CRISPR/Cas9 system of four genes, that is, *GS3*, Grain Widths 2, 5, and 6 (*GW2*, *GW5*, and *GW6*), which are negative regulators of grain weight, was investigated in rice. A remarkable improvement was observed in grain weight and size (Xu et al., 2016). The CRISPR/Cas9 system was also employed in rice to knockout three heading date genes (i.e., *Hd2*, *Hd4*, and *Hd5*) (Li et al., 2017). The mutants displayed early heading and higher yield under drought stress conditions. Furthermore, a CRISPR/Cas9 mediated disruption of the *OsSWEET11* gene, known for grain filling and sucrose transportation in rice, led to reduced sucrose concentration and grain weight, which suggested that overexpression of these genes would be beneficial in obtaining a better grain quality (Ma et al., 2017). In wheat, *GASR7* was knocked out using the CRISPR/Cas9 tool, and the resulting mutants showed increased kernel weight (Zhang et al., 2016). In tomato, the use of CRISPR/Cas9 methods has also delivered seedless tomatoes (Ueta et al., 2017). In this study, a novel sgRNA/Cas9 was employed, resulting in additional somatic mutation in *SlIAA9*, a key parthenocarpy gene. The mutation rate was 100%, and there were no off-target mutations. The mutants hence obtained displayed parthenocarpic fruit along with an altered leaf shape.

7 Evolution of clustered regularly interspaced short palindromic repeats/Cas9 platform for precise gene manipulation

CRISPR/Cas9 systems have evolved over the years, and many other approaches have also been routed in this technology. As discussed earlier, CRISPR/Cas9-mediated gene editing necessarily introduces DSBs that are subsequently repaired by either NHEJ or HDR mechanisms (Kantor et al., 2020). This results in two major challenges in using CRISPR/Cas9 mechanisms. Firstly, although HDR promises insertion of only sequence-specific DNA, this pathway is synonymous with increased instances of indels and limited efficiency (Song et al., 2017). Secondly, reliance on the HDR mechanism of gene repair restricts gene editing to only dividing cells, adversely affecting the efficiency of this platform in manipulating the disease resistance in plants (Bollen et al., 2018). Many newer technologies that are primarily rooted in the CRISPR/Cas mechanism overcome some of these limitations and are more precise in achieving genome restructuring in plants. Some of these technologies are detailed in the following sections.

7.1 Multiplex genome editing

In plants, it is well documented that cellular processes are orchestrated *via* the interplay of several redundant genes. Therefore, editing a single gene from a gene family has not been found to confer the desired phenotype as the redundant genes from the same gene family compensate for the phenotype.

In polyploid crop species, this presents an additional layer of complication due to multiple gene dosages or homolog effects. Hence, a more efficient protocol for gene editing is required to aid multiplex gene editing. A single vector system has been used to design many sgRNA cassettes with single or multiple promoters in multiplex gene editing mediated *via* the CRISPR/Cas9 system (Liu et al., 2017). In *Arabidopsis thaliana*, two sgRNAs were successfully employed to disrupt two homologs of *CHLI* (magnesium-chelatase subunit I) to obtain an albino phenotype as both homologs have a function in the photosynthetic mechanism (Mao et al., 2013). In another study in *A. thaliana*, multiplex gene editing was successfully employed to obtain quadruple mutants displaying dwarf phenotype by deploying three gRNAs (Wang et al., 2017).

Further, Čermák et al.(2015) developed a tool kit wherein Csy-type (CRISPR system *yersinia*) ribonuclease 4 (Csy4) was employed along with tRNA-processing enzymes to simultaneously express multiple gRNAs. Using this method, they expressed 12 gRNAs from a single transcript to target deletions in six genes successfully. These Csy4 and tRNA expression systems have been found almost twice as effective in introducing mutations. The use of this platform has been validated in tobacco (*Nicotiana tabacum*), tomato (*Solanum lycopersicum*), wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), and *Medicago truncatula* (Čermák et al., 2015).

Xie et al. (2015) reported an endogenous tRNA-processing mediating gene editing by CRISPR/Cas9 in rice. Soon after, Tang et al. (2016) reportedly employed a single POL II promoter to drive the expression of a hammerhead ribozyme and multiple gRNAs. The ribozyme cleaved distinct sgRNAs, and post-transcription Cas9 processed functional Cas9 and gRNAs. In maize, the CRISPR/Cas9-based gene editing was successfully used to mutate the homologs that determine genic male sterility (Liu et al., 2022). Triple homozygous mutants were obtained that displayed complete male sterility. Over the course of CRISPR/Cas evolution, multiplex gene editing has emerged as an efficient tool to develop “multiple genes-knock-out-cultivars.” Concomitantly, this methodology has enhanced our understanding of gene functions of desired traits that are governed by multiple genes, gene families, or even pleiotropic genes. The technology has also opened vistas for investigating epistatic interactions/associations among genes or gene complexes, especially for complex traits, whose genetic architecture is largely influenced by epistasis.

7.2 New Cas variants to broaden the clustered regularly interspaced short palindromic repeats toolbox

Since the discovery of CRISPR/Cas9 mediated gene editing, numerous modifications have been incorporated into this technology to address the issue of incompatible off-target sequences due to gRNA mismatches. There have been many attempts to increase the efficiency of Cas9 enzymes and, at the same time, curb any off-target silencing with the use of enzymes such as dead cas9 (dcas9), SpCas9 Nickase (SpCas9n), and FokICas9 (fCas9) (Cong et al., 2013; Guilinger et al., 2014). Other studies have reported the extraction of Cas9 proteins with increased sequence specificity owing to their novel PAM sequences. Nmecas9 was extracted from *Neisseria meningitidis* specific for PAM sequence 5′-NNNNGATT (Lee et al., 2016). SpCas9 is most commonly used for gene editing with a PAM sequence 5′-NNGRRT (Ran et al., 2015). Modifications have been made for SpCas9 to identify shorter PAM sequences that not only increase the efficiency of the enzyme but also make the delivery of the system easier (Hu et al., 2018). In plants, CRISPR/Cas9 mediated gene editing has been employed in many plant species such as *A. thaliana*, rice, citrus, and tobacco (Jiang and Doudna, 2017). Furthermore, St1Cas9 and St3Cas9 extracted from *Streptococcus thermophilus* have also been employed in CRISPR-mediated gene editing (Jiang and Doudna, 2017). These Cas9 enzymes use different types of tracrRNA and crRNA for identifying PAM sequences (Steinert et al., 2015). Out of all these CRISPR systems employed so far, CRISPR/Cpf1, commonly known as Cas13, is the most popular (Zetsche et al., 2015). Unlike Cas9, Cas13 requires only a sgRNA with 4–5 nucleotide overhangs. In both animals and plants, the Cas13-mediated gene editing has been found to target the desired genes with none or very few off-targets (Endo et al., 2016). Due to their successes, type V CRISPR/Cpf1 has been popular in both plants and animals to engineer gene editing (Zhang et al., 2017). *Francisella novicida*-derived FnCpf1 was used to achieve targeted mutagenesis in both tobacco and rice. Similarly, Lachnospiraceae-derived LbCpf1 has also been used to achieve targeted mutagenesis (Yin et al., 2017).

7.3 Epigenome editing

Epigenome editing represents the third phase of plant GE, wherein changes are introduced to engineer the chromatin *via* modification of epigenome at specific sites. It involves targeted, locus-specific, reversible, and heritable alterations of the chromatin structure while bringing in no changes in the nucleotide sequences in the genomes by using epi-effectors. Epi-effectors are the epigenome engineering tools that represent a programmable DNA binding/DNA recognition domain in the genome. Additionally, the catalytic domains of chromatin-modifying enzymes (DNA methyltransferases and histone acetylases) represent components of an Epi-effector. Different epigenome editing tools are available for creating, erasing, and reading various epigenetic codes in plants (Jeltsch and Rots, 2018; Miglani and Singh, 2020; Miglani et al., 2020).

Currently, epigenome editing has been performed through three molecular platforms: zinc-finger proteins (ZFPs),

transcription activator-like effectors (TALEs), and CRISPR and dead CRISPR/Cas proteins. These act as DNA-binding domains (DBDs), and after interaction with epigenetic domains, they modify the epigenetic marks at targeted sites in the genome to bring about a restructuring of chromatin architecture and gene expression. The principle of epigenomic editing rests on the formation of fusion proteins between a designed DBD (ZFPs/TALEs/nuclease null or dead Cas9) that targets an attached enzymatic domain (chromatin modifiers; DNA methyltransferases (DNMTs) or histone acetyltransferases (HATs) to define genomic target sites. Hence, the DNA sequences of the target genomic site are presented to DNA-binding protein domains that affect DNA function in the presence of an enzymatic effector domain. This way, epigenome editing allows the precise modification of individual chromatin marks at selected genomic sites (Nakamura et al., 2021).

Besides modulating gene expression, epigenome editing is an appealing approach for understanding the mechanism of chromatin modification, cellular reprogramming, and regulatory functions. It has applications in both basic research involving gene expression studies and application-oriented epigenomic engineering of crop plants. The characterization of epialleles (i.e., alleles that are genetically alike but show variable genetic expression due to epigenomic modifications) is gradually picking up to be fully exploited in future crop improvement programs. Epigenome editing holds great promise in improving crops by creating novel epiallelic diversity that can be exploited for future precision and smart crop epi-breeding (Gahlaut S K et al., 2020; Giudice et al., 2021; Kakoulidou et al., 2021). For epigenome editing, a modified CRISPR/dCas9 known as dead, deactivated, null, or nuclease deficient Cas9 (dCas9) has been created by silencing two mutations of the RuvC1 (D10A) and HNH (H841A) nuclease domains (Qi et al., 2013). The CRISPR-dCas 9 approach is attractive as it helps overcome the limitation of the DBD approach, wherein for targeting a different sequence, a corresponding distinct protein is required, making it difficult to target a wide range of loci in the genomes. In this respect, CRISPR-dCas9 associated system offers flexibility as associated gRNAs help the Cas proteins achieve genomic specificity (Nakamura et al., 2021). A single dCas protein can be reoriented to target different loci simply by altering the sequence of its associated gRNA. This way, the technology offers a flexible platform for targeting almost any genomic sequence (Brocken et al., 2018). Epigenomic editing depends on inducing changes in chromatin architecture to influence gene transcription and relies on primarily inducing reversible and heritable changes in epigenetic marks such as DNA and histones' methylation, acetylation, and phosphorylation. This results in novel genetic variation in the form of epialleles and has tremendous potential for crop enhancement through epi-breeding. Although several publications have demonstrated the feasibility of epigenome editing in *A. thaliana* (Table 3), its modalities need to be standardized in crop plants for commercial application.

The first successful instance of epigenome editing was achieved in the model plant species *A. thaliana* (Johnson et al., 2014). A ZFN fused to RdDM (RNA-directed DNA methylase) component SU(VAR)3-9 HOMOLOG 9 (SUVH9) was involved in the recruitment of PolV during RdDM mediated *via* methyl-DNA binding SUVH2 and SUVH9 proteins at the FWA target to display DNA methylation induced gene silencing. Many other components of RdDM, such as SHH1, NRPD1, RDR2, DMS3, and RDM, when joined with ZFs, have also been shown to induce methylation at the FWA target in *A. thaliana* (Gallego-Bartolomé et al., 2019). A CRISPR dCas9-SunTag-based targeting system coupled with tobacco DRM methyltransferase (NtDRMcd) was used to target DNA methylation in *A. thaliana* (Zhong et al., 2014; Papikian et al., 2019). It resulted in the induction of DNA demethylation at FWA and SUPERMAN promoters affecting gene transcription and triggering a developmental phenotype. Further, a repressive effect of H3K9me2 and non-CG DNA methylation on both meiotic DSB and crossover formation in plant pericentromeric heterochromatin resulted in manipulation of the rate and positions of crossing over. Increase in meiotic recombination in proximity to the centromeres (pericentromeric recombination) and meiotic DNA double-strand breaks (DSBs) in Thale Cress (Papikian et al., 2019). Recently, Gallego-Bartolomé et al. (2018), Gallego-Bartolomé et al. (2019), and Gallego-Bartolomé (2020) used ZF and CRISPR-dcas9-SunTag systems fused with the catalytic domain of human demethylase TET1cd to test several RdDM components such as RNA-dependent RNA polymerase 2 (RDR2), Microchidia 1 and 6 (MORC1 and MORC6), RNA directed methylation 1 (RDM1), and defective in meristem silencing 3 (DMS3) to induce targeted DNA methylation/demethylation at *FWA* locus in *A. thaliana*. ZF fusion with catalytic domain human demethylase TET1cd and SunTag-TET1cd system resulted in demethylation of the promoter of *FWA* (Flowering Wageningen) gene and CACTA1 transposon and activation of gene expression. While the fusion of ZF-RdDM and ZF-MORC6 enhanced targeted FWA methylation, Microrchidia (MORC6) targeted DNA methylation and triggered AGO- and DRM2-dependent methylation and gene silencing in *A. thaliana* (Gallego-Bartolomé et al., 2019; Gallego-Bartolomé, 2020). These studies provide important experimental evidence to design and utilize a highly targeted and heritable DNA methylation/demethylation system to modulate gene expression in crop plants.

Fusion of CRISPR *dCas9-HAT1* gene resulted in hyperacetylation at *AREB1* (abscisic acid-responsive element-binding protein 1) locus leading to activation of endogenous promoter of *AREB1*. This improved transcription of the *AREB1* gene involved in ABA perception improved chlorophyll content and drought tolerance due to the activation of bZIP TF, which

TABLE 3 Epigenome editing in the model plant *Arabidopsis thaliana*.

DNB Domain/targeting system/ target gene	Epigenome editing/modification	Response	References
ZFN fused to SUVH9	Recruitment of PolV during RdDM through methyl-DNA binding SUVH2 and SUVH9 proteins	DNA methylation and gene silencing	Johnson et al. (2014)
CRISPR dCas9-SunTag based targeting system coupled with tobacco DRM methyltransferase (NtDRMcd)	Manipulation of DNA methylation at FWA promoter	Modification of gene expression, induction of DNA demethylation at FWA, and SUPERMAN promoter affecting gene transcription and triggering a developmental phenotype	Zhong et al. (2014), Papikian et al. (2019)
Mutation of the H3K9 methyl transferase genes *KYP/SUVH4 SUVH5*, *SUVH6*, or the CHG DNA methyl transferase gene *CMT3*	Disruption of histone 3 di-methylation on lysine 9 (H3K9me2) and non-CG DNA methylation *via* mutation of the H3K9 methyl transferase genes *KYP/ SUVH4 SUVH5*, *SUVH6*, or the CHG DNA methyl transferase gene *CMT3*	Manipulation of the rate and positions of crossing over (CO). Increase in meiotic recombination in proximity to the centromeres (pericentromeric recombination) and meiotic DNA double-strand breaks (DSBs). Repressive effect of H3K9me2 and non-CG DNA methylation on both meiotic DSB and crossover formation in plant pericentromeric heterochromatin	Underwood et al. (2018)
ZF fusion with catalytic domain human demethylase TET1cd and SunTag-TET1cd system	Demethylation of the promoter of *FWA* (Flowering Wageningen) gene and *CACTA1* transposon	Targeted, complete, highly specific, and heritable demethylation (removal of 5 mC at specific loci in the genome) at *FWA* promoter and activation of gene expression. Reactivation and upregulation of the FWA gene and a heritable late-flowering phenotype. Targeted demethylation and reactivation of heterochromatic TE-CACTA1, although demethylation was incomplete on this locus and remethylation and resilience occurred once the trigger construct was segregated out	Gallego-Bartolomé et al. (2018), Gallego-Bartolomé, (2020)
ZF-RNA directed DNA methylase (RdDM); ZF-MORC6	Co-targeting of both arms of the RdDM pathway, siRNA biogenesis, and co-targeting of Pol IV and Pol V synergistic recruitment	Enhanced targeted *FWA* methylation and silencing, microrchidia- (MORC6-) targeted DNA methylation. Trigger of AGO- and DRM2-dependent methylation	Gallego-Bartolomé et al. (2019), Gallego-Bartolomé, (2020)
CRISPR *dCas9-HAT1* gene	Hyperacetylation at *AREB1* (Abscisic acid-responsive element-binding protein 1) locus resulting in activation of endogenous promoter of *AREB1*	Improved transcription of *AREB1* gene involved in abscisic acid perception. Improved chlorophyll content and drought tolerance due to activation of bZIP TF that can activate several stress tolerance-related genes like RD29A	Paixão et al. (2019)
CRISPR dCas9-TET1	Essential requirement of methylated CG (mCG) and mCHG (where H can be A, C, or T) for targeting RdDM machinery to remethylable loci. RdDm target loci to form stable epialleles in the presence of specific histone and DNA methylation marks	Induction of alternation between two epi-allelic states at a specific locus	Li C et al. (2020)
CRISPR-bacterial methyltransferase MQ1v and CRISPR-SunTagMQ1v Systems	*De novo* induction of CG methylation at different loci with varying efficiency with CRISPR-MQ1v and CRISPR-SunTagMQ1v systems. CRISPR-SunTagMQ1v has shown to be more potent than CRISPR-MQ1v. Development of a CRISPR-based CG-specific targeted DNA methylation system	Improved heritability of induced target-specific CG methylation and high specificity of CRISPR-based MQ1v systems	Ghoshal et al. (2021)

can activate several stress tolerance-related genes such as *RD29A* (Paixão et al., 2019). Further, Li et al. (2020a) showed essential requirements of methylated CG (mCG) and mCHG by using CRISPR dCas9-TET1 fusion (where H can be A, C, or T) for targeting RdDM machinery to re-methylate loci. RdDm target loci were shown to form stable epialleles in the presence of specific histone and DNA methylation marks to induce alternation between two epiallelic states at a specific locus.

Recently, Ghoshal et al. (2021) used CRISPR-bacterial methyltransferase MQ1v and CRISPR-SunTagMQ1v and

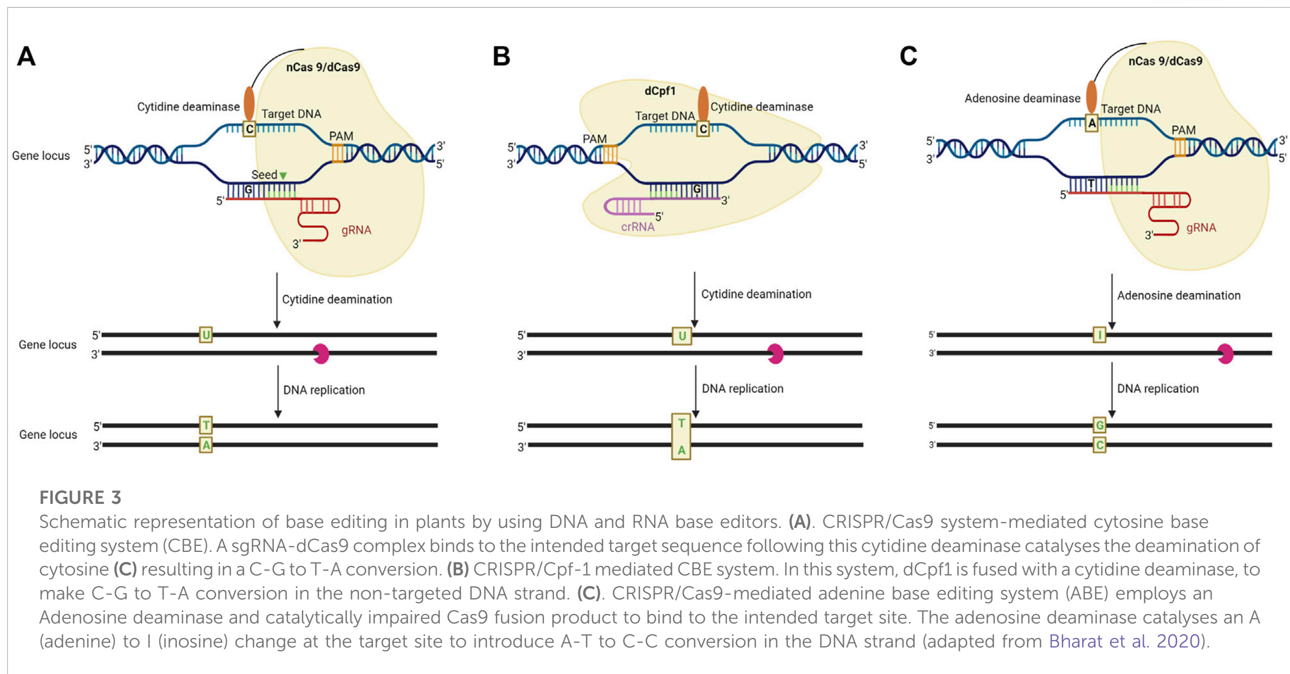


FIGURE 3
Schematic representation of base editing in plants by using DNA and RNA base editors. **(A)**. CRISPR/Cas9 system-mediated cytosine base editing system (CBE). A sgRNA-dCas9 complex binds to the intended target sequence following this cytidine deaminase catalyses the deamination of cytosine **(C)** resulting in a C-G to T-A conversion. **(B)** CRISPR/Cpf-1 mediated CBE system. In this system, dCpf1 is fused with a cytidine deaminase, to make C-G to T-A conversion in the non-targeted DNA strand. **(C)**. CRISPR/Cas9-mediated adenine base editing system (ABE) employs an Adenosine deaminase and catalytically impaired Cas9 fusion product to bind to the intended target site. The adenosine deaminase catalyses an A (adenine) to I (inosine) change at the target site to introduce A-T to C-C conversion in the DNA strand (adapted from Bharat et al. 2020).

developed a CRISPR-based CG-specific targeted DNA methylation system to achieve *de novo* induction of CG methylation at different loci with varying efficiency. CRISPR-SunTagMQ1v was shown to be more potent than CRISPR-MQ1v. These MQ1v-based tools appear to be attractive as they offer flexibility to induce methylation at different levels at different loci and show high specificity attributed to the Q147L mutation. Further, the study also demonstrated that for some loci, CG methylation alone was enough to silence gene expression, and for these loci, CRISPR-MQ1v and CRISPR-SunTagMQ1v systems were likely to be more efficient than the DRM2-based SunTag system developed by Papikian et al. (2019) described above.

The above examples show the potential of epigenome editing technology in modulating gene expression and showing observable changes in the phenotypes by altering the DNA methylation status at various genetic loci in *A. thaliana*. Similar studies need to be extended to crop species for exploiting the advantages of locus-specific modulation of DNA methylation through epigenome editing. The new tier of epigenetic variability generated by epigenome editing has significant potential in bringing about the genetic enhancement of crop species.

Epigenome editing, as discussed here and in many other reviews (Gahlaut V et al., 2020; Giudice et al., 2021; Kakoulidou et al., 2021), offers opportunities for editing epigenetic codes in plant genomes globally or at selected loci to create novel genetic variability. To harness the benefits of epigenomic editing, however, it is important to define the specific epimark(s) linked with specific phenotypes and agronomic traits of interest. In this context, genome-wide mapping of epigenomic marks and epigenetic target identification are among the current thrust research areas. A few genetic elements controlled by DNA methylation and linked to desired plant traits have been identified. For instance, naturally occurring epi-alleles that accumulate high levels of vitamin E in tomatoes are associated with differential methylation of a SINE retrotransposon located in the promoter region of gene *VTE3(1)* (Quadrana et al., 2014). In cotton, the *COL2* epi-allele is associated with DNA methylation changes and affects flowering time (Song et al., 2017). It is important to accumulate epigenomic data in various crop species to help identify the potential candidate editing targets. Information on genome-wide changes in DNA methylation in response to environmental stress has been gathered in crops such as rice (Guo et al., 2019; Rajkumar et al., 2020), wheat (Kumar et al., 2017), soybean (Song et al., 2012), and sesame (Komivi et al., 2018).

7.4 Base editing

Base editing (BE) is a novel GE technology representing the fourth phase of the evolution of GE platforms wherein a single nucleotide in a DNA or RNA can be substituted irreversibly. The process does not involve a double-stranded breaks (DSB) and hence bypasses the undesirable effects of NHEJ and HDR mechanisms. Of all the previous tinkering tools, BE is the most attractive for the simple reason that here the genome modification is "base-pointed" and precise. It does not involve additions or deletions in the genome (i.e., no change occurs in the DNA content of the organism). Neither does it involve the

TABLE 4 Base editing mediated proof of concept and improvement studies in major crop plants.

Aim	Editor	Plant	Genes targeted	References
Proof of concept/ demonstration of editing efficiency	CBE	Rice	*OsNRT1.1B*, *OzSLR1*, *OsCDC48*, *OsSPL14*, *OsSERK1*, *OsSERK2*, *OsPi-ta*, *OsSBEIIb*, *OsPDS*, *OsALS*, *OsAOS1*, *OsJAR1*, *OsJAR2*, *OsCOI2*, *OsSNB*, *OsSPL7*, *OsPMS3*, *OsSPL14*, *OsIPA1-T1, OsMKK6*, *OsEhd1*, *OsPi-d2*, *OsMPK3*, *OsROC*	Lu and Zhu (2017), Zong et al. (2017), Ren et al. (2017), Li P et al. (2017), Ren et al. (2018), Wang et al. (2019), Qin et al. (2019), Sretenovic et al. (2021)
		Wheat	*TaLOX2*	Zong et al. (2017)
		Maize	*ZmCENH3*	Zong et al. (2017)
		Arabidopsis	*LFY*	Choi et al. (2021)
		Tomato	*SlALS1*, *SlCYC-B*, *SlDET1*, *SlDDB1*, *SlETR1*, *SlETR2*, *SlHWS*, *SlDELLA*	Hunziker et al. (2020), Kashojiya et al. (2022)
		Rapeseed	*BnaCLV3*, *BnaRGA*, *BnaA3.IAA7*, *BnaDA1*, *BnaALS*	Hu et al. (2020), Cheng et al. (2021)
	ABE	Rice	*OsACC-T1*, *OsALS-T1*, *OsCDC48-T3*, *OsDEP1*, *OsNRT1.1B-T1*, *OsIPA1*, *OsSLR1*, *OsMPK6*, *OsMPK13*, *OsSERK2* and *OsWRKY45*, *OsSPL14*, *OsSPL17*, *OsSPL16*, *OsSPL18*, *OsIDS1*, *OsTOE1*, *OsSNB*, *OsPMS3*, *OsPMS1*, *OsSPL14*, *OsLF1*, *OsIAA13*, *OsSPL7*, *OsSPL4*, *OsMADS5*, *OsWx*, *OsPi-d3*, *OsGL2*, *OsGRF3*, *OsSLR1*, *OsWSL5*, *OsZEBRA3* (*Z3*), *OsROC*	Hua et al. (2018), Hua et al. (2019), Wang et al. (2019), Hua et al. (2020a), Sretenovic et al. (2021)
		Wheat	*TaDEP1*, *TaGW2*, *TaALS*, *TaTub*	Li J et al. (2018), Han et al. (2022)
		Tobacco	*NbPDS*	Wang W et al. (2021)
	CGBE	Rice	*OsALS*, *OsCGRS55*	Sretenovic et al. (2021)
		Tomato	*AGO7*	
		Poplar	*PtPDS1*, *PtPDS2*	
	DuBE	Rice	*OsAAT*, *OsACC*, *OsCDC48*, *OsDEP1*, *BADH2-2*, *FSD2-1*, *LAZY1-2*	Li et al. (2020a), Xu R et al. (2021)
Co-editing	CBE	Pear, apple	*PDS*, *ALS*	Malabarba et al. (2021)
Double CBE	CBE	Potato	*StDMR6-1*, *StGBSSI*	Veillet et al. (2020)
Simultaneous base editing	CBE and ABE	Rice	*OsSPL14*, *OsSPL17*, *OsSNB*	Hua et al. (2019)
To introduce premature stop codon		Poplar	*4CL1*, *PII*	Li R et al. (2021)
Resistance to biotic stress	CBE	Rice	*OsPi-d2, OsFLS2*	Ren et al. (2017)
Herbicide tolerance	CBE	Rice, wheat, watermelon, foxtail millet, *Arabidopsis*, potato, pear, tomato, rapeseed	*ALS1*, *ACC*, *GS1*, *TubA2*	Chen et al. (2017), Tian et al. (2018), Zhang A et al. (2019), Veillet et al. (2019), Veillet et al. (2020), Cheng et al. (2021), Kuang et al. (2020), Liu et al. (2020), Wu et al. (2020), Zhang J et al. (2020), Malabarba et al. (2021), Liang Y et al. (2022)
Improved grain/fruit/ seed quality	CBE	Rice	Waxy	Li et al. (2020a), Xu et al. (2020), Tra et al. (2021)

incorporation of DNA from another organism (i.e., the edited organism does not become a GMO). It minimizes the chances of unintended, unwarranted effects on the phenotype (Rees and Liu, 2018; Deb et al., 2022). With a perfect BE toolbox, one can envisage generating desirable alleles for a trait by simply making the required substitutions. All that is required is a base modifying enzyme linked to a modified endonuclease, such as dCas9, which can target a desired region in the genome but not cause a DSB. Since the advent of this technology in 2016, it has become possible to execute C to T and A to G transition and C to G transversion editing. Figure 3 presents a schematic representation of the working mechanism of the base editing methodology that has been employed for GE.

7.4.1 Cytosine base editors C to T

GE has been revolutionized by engineering the CRISPR/Cas9 to enable cytosine base editing (Komor et al., 2016). The first-generation cytosine base editors (BE1) comprised of catalytically dead dCas9 (D10A, H840A) fused with rat apolipoprotein B mRNA editing enzyme (rAPOBEC1), a

cytidine deaminase operating on ssDNA *via* a 16aa XTEN linker at its N-terminus (rAPOBEC1-XTEN-dCas9). Although BE1 was highly efficient in converting C:G to T:A *in vitro*, the same decreased considerably when assessed within cells because of the base excision repair mechanism (BER). To bypass the *in vivo* repair response and overcome decreased efficiency, second-generation cytosine base editors (BE2) were formed by fusion of Uracil DNA glycosylase inhibitor (UGI) to the C-terminal of BE1. This inhibited the action of Uracil DNA glycosylase (UDG), which would otherwise have catalyzed the removal of U, resulting in reversion to C:G through BER. The C:G to T:A conversion efficiency was sought to be further enhanced by generating a nick on the non-edited DNA strand, thereby stimulating the cellular mismatch repair mechanism (MMR), which would replace the G on the nicked strand opposite the U on the target strand by an A, resulting in a U:A, which gets repaired to result in the desired T:A substitution. This resulted in BE3, a BE2 with a dCas9 modified to enable nicking activity (nCas9-H840A), resulting in much more efficient C:G to T:A substitutions (Komor et al., 2016).

7.4.2 Adenine base editors A to G

Although CBEs use naturally occurring cytosine deaminases to convert cytosine to uracil or 5-methylcytosine to thymine, no known adenine deaminases could deaminate the adenosine in DNA. In a significant breakthrough, Gaudelli et al. (2017) used directed evolution to form a modified transfer RNA adenosine deaminase (TadA*), which could catalyze the deamination of deoxyadenosine in an ssDNA resulting in a deoxyinosine. TadA* was joined through the XTEN linked to the N-terminus of Cas9 nickase with a nuclear localization signal (NLS) at its C-terminus (TadA*–XTEN–nCas9–NLS). The group engineered seven generations of ABEs to arrive at ABE7.10, which had high efficiency in converting A:T to G:C (Gaudelli et al., 2017).

7.4.3 Cytosine to Guanosine base editor C to G

It had been observed that although the efficiency of C to T transitions increased considerably by fusing UGI to BE1, in absence of the glycosylase inhibitor, C to T conversions were not so clean and were accompanied by C to G and C to A transversions (Komor et al., 2016). This action of glycosylase, which sought to be inhibited in CBEs for improved recovery of clean C to T substitutions, was tapped for accomplishing C to G transversion in CGBEs. Uracil DNA N-glycosylase (ecUNG) from *Escherichia coli* (Kurt et al., 2021; Zhao et al., 2021) or rat XRCC1 (Chen et al., 2021) were linked to a nCas9 (D10A) and further fused with a rat cytidine deaminase rAPOBEC1 (Chen et al., 2021; Zhao et al., 2021) or its engineered variant rAPOBEC1 (R33A) (Kurt et al., 2021) or with human activation-induced cytidine deaminase (h-AID) (Zhao et al., 2021). The resultant CGBEs or GBEs (glycosylase base editors), UNG-nCas9-APOBEC1, XRCC1-nCas9-APOBEC1, UNG-APOBEC1-nCas9, and h-AID-nCas9-UNG, result in the conversion of C to U and subsequently to G *via* base excision repair (Chen et al., 2021) or by translesion polymerization (Liu et al., 2016). The nicking of the opposite strand triggers the repair machinery of the cell, which converts C:G to G:C.

7.4.4 Dual-base editors

Dual-base editors have recently been developed by merging the cytosine and adenine deaminases in a single editor termed variably as SPACE (synchronous programmable adenine and cytosine editor) (Grunewald et al., 2020), STEMEs (saturated targeted endogenous mutagenesis editors) (Li et al., 2020), ACBE (adenine and cytosine base editor) (Xie et al., 2020), and DuBEs (dual-base editors) (Xu et al., 2021). Grunewald et al. (2020) fused the monomeric TadA of miniABEmax-V82G6 and pmCDA1 of Target-AID5 with the adenine deaminase at the N-terminus and cytosine deaminase at the C-terminus of nCas9 (D10A). Sakata et al. (2020) and Xie et al. (2020) also used the same architecture. Zhang et al. (2020) developed DuBEs (A&C-BEmax) by fusing the two deaminases to the N-terminus and found that hAID-TadA-TadA*linked to nCas9 (D10A) along with two UGIs yielded higher editing efficiency compared to multiplexing with individual deaminase editors in human cells. Li et al. (2020) developed STEMEs by fusing both deaminases, APOBEC3A/ecTadA, to the N-terminus of nCas9 (D10A) and tested them in rice. They reported better C to T and A to G editing with the DuBE than that achieved using co-delivered deaminases and could generate herbicide resistance in rice. Overall, DuBEs were more efficient in C to T edits than A to G. However, the plant DuBE version 1 (pDuBE1) developed by Xu et al. (2021) using TadA-8e and LjCDA1L-4 (*Lethenteron japonicum* CDA1-like 4) fused to the opposite termini of nCas9 (D10A) displayed highly efficient simultaneous A to G/C to T edits (49.7%) in rice calli. Liang et al. (2022) furthered the scope of DuBEs by engineering an AGBE (fusing a CGBE with an ABE), which could render efficient C to G, C to T, C to A, and A to G editing possible in mammalian cells.

7.4.5 Base editing in plants

Base editing (C to T transitions) in plants was demonstrated for the first time in rice (Lu and Zhu, 2017; Ren et al., 2017; Zong et al., 2017; Li et al., 2017). Lu and Zhu (2017) formed a fusion protein, APOBEC1-XTEN-Cas9(D10A), as described by Komor et al. (2016), put it under the ubiquitin maize promoter, and used it for editing *OsNRT1.1B* and *OsSLR1* in rice. Sequencing confirmed C to T (1.4%–11.5%) and C to G (1.6%–3.9%) substitutions in both genes to be more in *SLR1* than *NRT1.1B*. Indels (10%) were much more than the <1% reported by Komor et al. (2016), probably because no uracil glycosylase inhibitor (UGI) was used. Zong et al. (2017) tailored the base editors by including UGI to form pnCas9-PBE (rAPOBEC1-nCas9-D10A-UGI) and pdCas9-PBE (rAPOBEC1-dCas9-UGI) and found that these bring about C to T substitutions in three rice (cell division cycle mutation 48 OsCDC48, nitrate transporter OsNRT1.1B,

and a plant architecture gene *OsSPL14*), one wheat (*TaLOX2*), and one maize (*ZmCENH3*) gene with hardly any indels. Cas9 nickase-based editor was more efficient than the one with dCas9. In the same year, Li et al. (2017), while reporting greater than 40% substitutions, proposed that editing efficiency could vary depending on the target locus amongst three targeted loci (one on *OsPDS* and two on *OsSBEIIb*) of rice.

One of the limitations that were obvious in the initial period of the use of this technology was the restriction imposed by the availability or otherwise the canonical PAM sites in a genome. To overcome this challenge, Cas variants/orthologues with relaxed PAM sites both naturally occurring and engineered have been employed. Further, since the first reported use of rAPOBEC cytidine deaminase from a rat in BE1, deaminases sourced from other organisms such as human apolipoprotein B mRNA editing enzyme (hAPOBEC3A) (Gehrke et al., 2018; Wang W et al., 2018), hAID (Hess et al., 2016), *Petromyzon marinus* cytidine deaminase 1 (PmCDA1) (Nishida et al., 2016), and their mutated forms with varying features *vis-a-vis* editing window, size, sequence preference, and so on have been reported (Cheng et al., 2019).

Various proof of concept studies conducted in plants for base editing using natural and engineered variants of Cas in combination with different cytidine/adenine deaminases have been listed in Table 4. A SpCas-9 variant, SpCas9-VQR (D1135V + R1335Q + T1337R), recognizes NGAN and NGNG PAM sites, broadening the reach within a genome (Kleinstiver et al., 2015). Ren et al. (2017) used this variant to develop two CBEs for rice, rBE3 (APOBEC1-XTEN-Cas9n-UGI-NLS) and rBE4 (APOBEC1-XTEN-Cas9nVQR-UGI-NLS), and successfully edited a blast susceptible protein and OsCERK1 (a receptor kinase) with an efficiency of 17%. Steinert et al. (2015) and Kaya et al. (2017) recommended the use of *Staphylococcus aureus* Cas9 (SaCas9) in plants because of its smaller size, longer target sequence, different PAM, and somewhat higher efficiency than spCas9. A variant with three mutations E782K/N968K/R105H (SaCas9-KKH SaKKH) has a relaxed PAM (NNNRRT) compared to the wild type (Kleinstiver et al., 2015). Qin et al. (2019) developed nSaCas9(D10A) and nSaKKH(D10A) nickase-based CBEs (Sa-BE3, SaKKH-BE3, Sa-eBE3, and SaKKH-eBE3) and ABEs (Sa-ABE and SaKKH-ABE/ABE-P5) reporting up to 71.9% cytosine edited (nSaCas9, SLR1 gene) and 63.2% adenine edited (nSaCas9, OsSPL17 gene) rice plants. Veillet et al. (2020) used the nickase SaCas9 (nSaCas9) with PmCDA1 to modify granule-bound starch synthase (StGBSS) and Downy Mildew Resistant 6 (StDMR6) in potato. It recognizes 5’--NNGGAT-3′ as a PAM site and has an editing window from −23 to −22. Nishimasu et al. (2018) engineered spCas9 to recognize NG (spCas9-NG), a relaxed PAM, and used the nickase version fused with activation-induced cytidine deaminase (nSpCas9-NG-AID/Target-AID-NG) to determine their editing efficiencies. Although Target-AID had a better efficiency at the canonical PAM, Target-AID-NG had a wider PAM repertoire and performed better than the former at other PAM sites, whereas xCas9-BE4 (Hu et al., 2018) was the least efficient in mammalian cells. Zhong et al. (2019) tested xCas9(D10A)-rAPOBEC1, xCas9(D10A)-PmCDA1-UGI, and Cas9(D10A)-NG-PmCDA1-UGI in rice and concluded that xCas9(D10A)-based editors were comparable in efficiency to those based on wtCas9(D10A). The former demonstrated better fidelity concerning the protospacer, and Cas9-NG-based editors were more efficient among all three tested at relaxed PAM sequences. Endo et al. (2019) used SpCas9-NGv1 nickase in rice. Veillet et al. (2020) used SpCas9NG-based CBE for editing granule-bound starch synthase (StGBSS) and Downy Mildew Resistant 6 (StDMR6-1) in potato. They also tested the performance of this editor in tomatoes by targeting two PAM sites in the acetolactate synthase (ALS) gene. GGT gave a lower efficiency (32%) than the canonical PAM NGN (64%).

Hua et al. (2018) adopted ABE7-10 (Gaudelli et al., 2017), developed adenine base editor plant version 1, ABE-P1 [TadA*7.10-SpCas9(D10A) nickase], and 2, ABE-P2 (TadA*7.10-SaCas9(D10A) nickase), and tested them on two rice genes: ideal plant architecture *OsIPA1* and slender plants *OsSLR1*. In 2019, they made several new versions, ABE-P3, P4, and P5, using SpCas9nVQR (D10A) and SpCas9-VRER (D10A) to increase target genome accessibility. They could successfully edit at four loci: *SPL14*, *SPL17*, *SPL16*, and *SPL18*. With the same set-up, they could demonstrate simultaneous cytosine and adenine editing using ABE-P2 and CBE-P1. Similar to reports in mammalian systems, there were no indels or off-target or any other unplanned base substitutions seen in rice. However, the editing windows were larger in the target genes. Hua et al. (2019) explored the use of SpCas9 and SaCas9 variants for widening the scope of the adenine base editing toolbox. They used nickases of VQR-, VRER-, and SAKKH-SpCas9 engineered variants to form three ABEs, ABE-P3 (pRABEspVQR), ABE-P4 (pRABEsp-VRER), and ABE-P5 (pRABEsa-SaKKH), and two CBEs with spCas9-VRER and saCas9-SAKKH, all of which were designed and tested in rice. The CBE and ABE formed with xCas9 were not efficient. Wang et al. (2021) compared the capabilities of ABE8e and ABE7.10 in *Nicotiana benthamiana* and established that ABE8e (60.87%) was more efficient than ABE7.10 (20.83%).

Sretenovic et al. (2021) studied the applicability of CGBEs, for affecting transversions in plants for the first time. They improvised the three CGBE platforms for successful use in humans (Chen et al., 2021; Zhao et al., 2021; Kurt et al., 2021) for use in three plant species: rice, tomato, and poplar. All three used the rat-derived rAPOBEC1 or its engineered variant rAPOBEC1 (R33A). rAPOBEC1 in combination with ecUNG or rXRCC1 was fused with nCas9 (D10A), whereas rAPOBEC1 (R33A) was linked to rescuing and nCas9 (D10A). Three, four, and two target sites were chosen for editing in rice, tomato, and poplar, respectively. As compared to BE3, all three CGBEs induced better C to G conversions, but the overall efficiency of conversion was less than that reported in

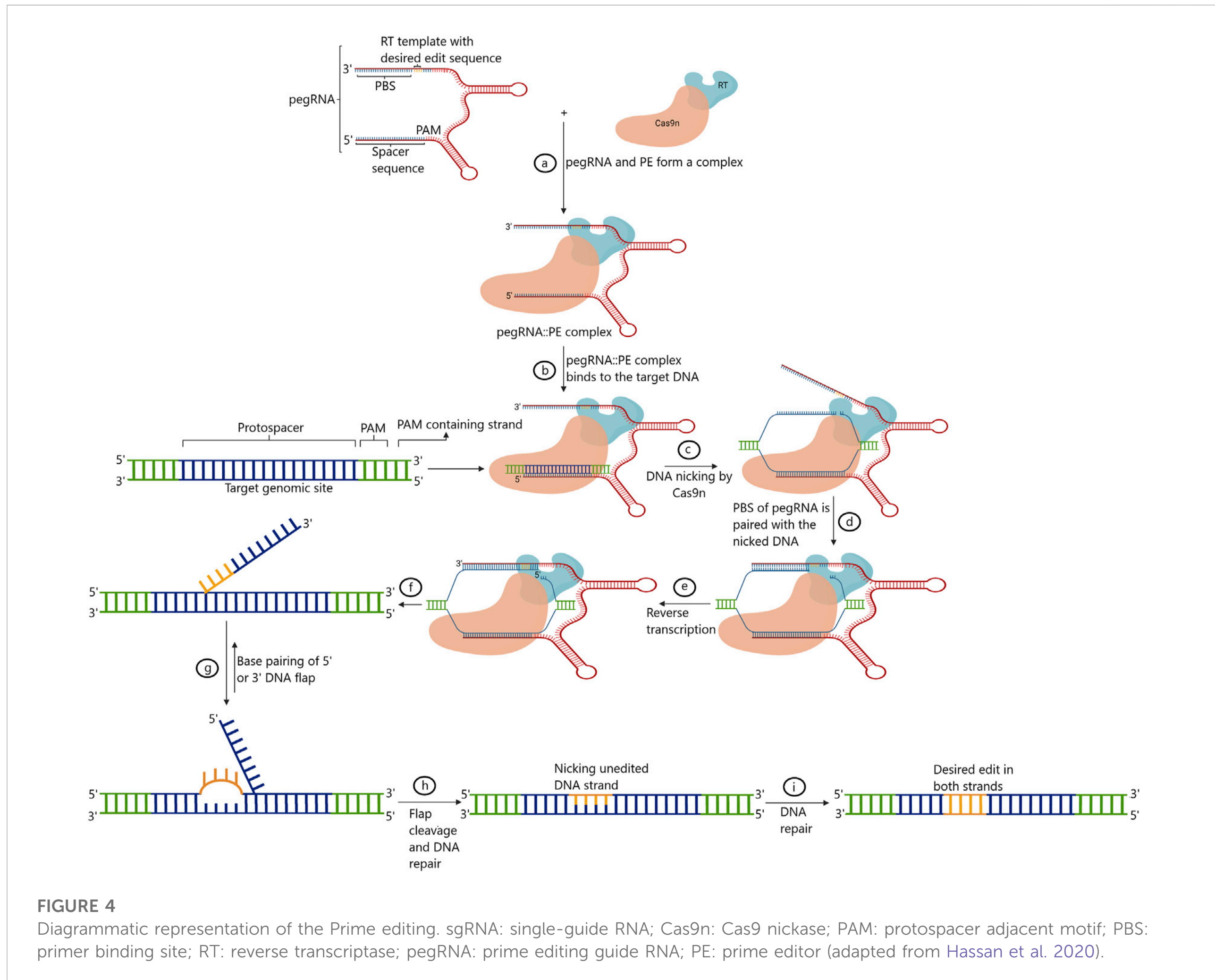


FIGURE 4
Diagrammatic representation of the Prime editing. sgRNA: single-guide RNA; Cas9n: Cas9 nickase; PAM: protospacer adjacent motif; PBS: primer binding site; RT: reverse transcriptase; pegRNA: prime editing guide RNA; PE: prime editor (adapted from Hassan et al. 2020).

humans. The efficiency of editing using SpRY, which is not PAM dependent, was also assessed. The authors achieved C to G editing, although the efficiency varied according to the system and target site. Because this was the first report, much needs to be done to improve the efficiency of plants.

Base editing is still an evolving technology, and many reports primarily demonstrate the successful use of a base-editing toolbox in different plants. This technology can create random variations within genomes, which can be screened and selected for advantageous traits. It also holds a great promise for improvement in traits affected by SNPs. Applications of the technology have been reported mainly as a gain of function for herbicide resistance and disease resistance and improvement in plant architecture, eating, and cooking quality (Table 4).

Base editing of acetyl-CoA carboxylase (*ACC*) and acetolactate synthase (*ALS1*) genes has been shown to confer herbicide resistance in rice (Li et al., 2020b; Liu et al., 2020; Zhang et al., 2020), tomato (Veillet et al., 2019; Veillet et al., 2020), potato (Veillet et al., 2019), watermelon (Tian et al., 2018), apple (Malabarba et al., 2021), pear (Malabarba et al., 2021), oilseed rape (Wu et al., 2020; Cheng et al., 2021), *Arabidopsis* (Chen et al., 2017), foxtail millet (Liang et al., 2022), and wheat (Zhang et al., 2019). The eating and cooking quality (ECQ) is of utmost importance for all cereals, and it is primarily determined by the amylose content in the grain, determined by the *Waxy (Wx)* gene-encoded granule-bound starch synthase I (GBSSI) (Li et al., 2016). Xu et al. (2021) used CBEs to develop rice lines expressing a range of amylose content (0%–12%), which improved its ECQ considerably by making several substitutions near the soft rice allele site in Wx. Similarly, Li et al. (2020a) lowered the amylose content in rice grains. Veillet et al. (2020) incorporated base substitutions in the *GBSSI* locus in potato, which could eventually be used for controlling amylose content in the tubers.

Traditional methods of inducing mutations become especially difficult in polyploid species because they possess more than two copies of a gene. Base editing has successfully generated heritable substitutions in polyploid species such as oilseed rape, wheat, and cotton. Hu et al. (2020) used

BnA3A1-PBE in rapeseed and demonstrated an editing efficiency of up to 50.5%, much higher than 23.6% reported by Cheng et al. (2021) and 1.8% by Wu et al. (2020). Li et al. (2018) demonstrated slight success (0.1%–1.1%) of *PABE* 1–7 in affecting A to G transitions in the *TaDEP1* and *TaGW2* wheat loci.

It is quite evident that this technology has immense potential, and once the challenges of discovering more efficient, PAM-independent DNA-binding proteins, better deaminases that can affect cleaner edits with zero off-targets, and engineering all possible substitutions are found, base editing can create a revolution in the field of plant sciences in general and crop improvement in particular.

7.5 Prime editing

Prime editing marks the fifth phase of evolution in GE platforms. The technique was first developed and standardized in human cells. Prime editing facilitates indels and all 12 possible base-to-base conversions, including transversions and transitions, without triggering the error-prone repair pathways by the DSB (Anzalone et al., 2019). Briefly, in this technique, paired/coupled prime editing guide RNA (pegRNA) is composed of single gRNA that is complementary to the one strand of the targeted DNA along with a primer-binding site (PBS), and the customized sequences to be replaced at the target site fused with Cas9 nickase are also present (Kumar et al., 2021). The PBS region primes to the second DNA strand to drive reverse transcriptase (RT) linked with the Cas9 nickase. RT transcribes and, in the process, copies the information straightaway from pegRNA into the intended target site. Following this, 5′ and 3’ are the single-stranded overhangs integrated into the genomic DNA *via* endogenous DNA repair mechanisms (Anzalone et al., 2019).

Research has successfully validated three generations of primer editors (PEs), PE1, PE2, and PE3, in humans so far. In PE1, the first-generation PEs, wild-type reverse transcriptase from commercial Moloney murine leukemia virus (M-MLV) fused to the C terminus of the Cas9 (H840A) nickase was used, triggered by the expression of pegRNA in a distinct plasmid. As mentioned earlier, pegRNA harbors a spacer sequence to recognize and bind to the intended target site. In addition, pegRNA carries an 8–15 nt of PBS and a template sequence to drive RT. However, the template sequence also contains a customized, altered DNA sequence to be incorporated at the intended site. The efficiency of this PE is largely determined by PBS length. Generally, 8–16 nt PBS length has been found to deliver results with increased efficiency (Anzalone et al., 2019). In an attempt to further increase the efficiency of this PE, numerous variants of M-MLV RT have been used. These variants were generated by inducing mutations in M-MLV RT. These mutations were found to alter processivity, thermostability, RNaseH activity, and DNA–RNA substrate affinity. In developing second-generation prime editors, PE2 an RT with five mutations (D200N, L603W, T330P, T306K, and W313F), when fused with the nickase, was found to increase the efficiency of the GE by 1.6–5.1 fold (Sretenovic and Qi 2022). The use of PE2 was found to hinder the efficiency primarily due to two factors. Firstly, the choice of single-stranded overhangs called “flaps” between unedited and edited to be paired with the native unmodified DNA strand. Secondly, choosing DNA strands as a template for DNA repair between unedited and edited was rather random (Gaudelli et al., 2017; Sretenovic and Qi, 2022). Many studies have shown that the introduction of nick in the unmodified strand enhanced the editing efficiency in both plants and animal cells (Komor et al., 2016; Gaudelli et al., 2017; Zong et al., 2017). Hence, to generate third-generation prime editors, PE3, nickase employed was used with an additional sgRNA to simultaneously nick the other complementary strand (Anzalone et al., 2019). This strategy enhanced the editing efficiency to introduce point mutations three-fold (Anzalone et al., 2019). With the use of the same protospacer, off-target instances were found much lower for PEs in comparison to the use of Cas9 (Jiang et al., 2021; Jiang et al., 2022). The increased efficiency of the prime editor is attributed to multiple DNA hybridization events that occur with the use of PEs. At first, the intended genomic DNA and spacer of the pegRNA hybridize. Next, hybridization occurs between the target sequence in the genomic DNA and the PBS of the pegRNA, adding to the sequence specificity of the system. Finally, the target DNA also hybridizes with the edited DNA, which further adds another layer of sequence specificity to the system (Jiang et al., 2021; Jiang et al., 2022). On the contrary, in a regular CRISPR/Cas9 system, only one step of hybridization occurs between the sgRNA and the target genomic DNA occurs (Jiang et al., 2022; Zhuang et al., 2022). Figure 4 presents a schematic representation of the working mechanism of the prime editing methodology that has been employed for GE.

The success of prime editing protocols hinges on optimizing critical parameters such as transformation system, selection of suitable vectors, design of prime editor cassettes (nuclease/nickase), structure/sequence of, for example, pegRNA, sgRNA, codon optimization of the vector constructs, promoters, use of novel/engineered endonuclease, ribozymes, reverse transcriptase, targeted genes, and method/s of detection. *Agrobacterium*-mediated transformation and floral dip agroinfiltration are the preferred modes of gene transfer as single copy inserts are efficiently achieved. However, other methods such as electroporation, PEG-mediated gene uptake, microinjection, and particle bombardment have been tested in different plants and are now expanding rapidly to include monocots (rice and maize), dicots (*Arabidopsis*, *Nicotiana benthamiana*, potato, and tomato), and even the bryophyte, *Physcomitrium patens* (Perroud et al., 2022) that is well known for incorporating DNA into specific genomic sites due to its innately high

frequencies of homologous recombination (Rensing et al., 2020).

Researchers have been experimenting extensively with the precise modeling of the molecular tool kit for high efficiency and specificity in several plants. As mentioned earlier, three versions of prime editors (PE1, PE2, and PE3) have been tested since 2019 in human and plant cells. The versions vary in the use of nickase, type of reverse transcriptase, position (C terminal or N-terminal fusion with nickase), length of the prime binding site, and types of editing predicted (Jiang et al., 2022). Promoters driving the expression of the prime editor apoprotein and the gRNAs play an important role in the overall scheme of prime editing in taxa and target gene of choice (Sretenovic and Qi 2022). Target sites have been categorized as type I and type II based on the position of the edit concerning the nicking site. If the edit is within 1–6 bp downstream of the pegRNA nicking site, then higher editing efficiencies are observed compared to the type II targets, where the targeted edit position(s) are 7–17 bp downstream of the pegRNA nicking site (Sretenovic and Qi, 2022). The editing efficiencies of the same vectors thus vary with the target genes. This was reported in rice, where the prime editor Sp-PE3 and gRNA were successful in introducing an S627N mutation in the endogenous *ALS* (acetolactate synthase) but were unsuccessful in editing the *APO1* (aberrant panicle organization) gene (Hua et al., 2020a). It was also successfully induced and present in regenerants. Three endogenous genes (*GAI*, *ALS2*, and *PDS1*) from tomato were tested for prime editing by PE3 strategy using an optimized prime editor. Prime editing frequencies of 0.025%–1.66% were observed in four pegRNAs out of seven tested, comparable to rice editing frequencies (Lu et al., 2020). Three genes (*OsPDS*, *OsACC1*, and *OsWx*) were used as targets to test the pPE2 system. Using the t-RNA processing strategy was also used to target a rice endogenous 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene (*OsEPSPS*) for prime editing to confer glyphosate resistance. A peg RNA with gRNA (59 bp RT, 13 nt PBS) and a second gRNA with the ability to nick at position 66 downstream were synthesized that could introduce triple mutations. For this gene-editing, the prime editing efficiency was 2.22% with both homozygous and heterozygous lines in rice (Li et al., 2020c). The pPPEM construct was tested in rice protoplasts, targeting gene *OsSULTR3*, six at two different edits for the bacterial leaf streak disease susceptibility. The editing efficiencies ranged from 0.7 to 2.2%. Besides editing endogenous genes, editing the transgenic reporter gene—fluorescent protein gene *EGFP* by SpPE2, SpPE3, and SaPE3—was tested in rice calli. The inactive insert was edited to active form successfully by SpPE3 at higher efficiencies than SpPE2, and none were observed with SaPE3, even though Sa compatible Cas9 and pegRNAs are required for efficient editing.

The prime-editing gRNAs of diverse structures with varied PBS and RT lengths and nicking position of gRNAs have also been reported to affect the prime editing efficiency (Xu et al., 2020; Hua et al., 2020a; Tang et al., 2020; Butt et al., 2020). Optimization of the melting temperature (Tm) of the PBS to around 30°C coupled with a dual-pegRNA strategy in plants (Lin et al., 2020) drastically increased the editing efficiencies by 17-fold in rice protoplasts, although stable expression and transmission of the edits remain to be seen. Inclusion of the t-RNA processing system (Xie et al., 2015) allows for the generation of multiple gRNAs that allow for "multiplex GE."

Detection of editing relies on the rates of transformation coupled with the rate of editing. Several studies have reported the co-transfection of T-DNA-containing vectors with the transgene and the PE vectors harboring the editor and the edit. The targeted sites are usually PCR amplified from the genomic DNA isolated from transformed plants and sequenced to identify the edits. Most researchers have done Sanger's sequencing, although the HRM-High Resolution Melting analysis has been included before sequencing by Perroud et al. (2022). Hi-TOM (high-throughput tracking of mutations) was used by Xu et al. (2022) in maize and rice.

Different selection and counter-selection strategies have been tested for the selection of transformed/edited cells. Perroud et al. (2022) have tested the use of APT/APRT (adenine phosphoribosyl transferase) enzyme that catalyzes the conversion of adenine to AMP in *Physcomitrium*. This enzyme can convert 2-fluoroadenine (2FA) supplemented in the culture medium into a toxic 2-fluoro AMP counter selective compound. Thus, if the editing vectors are successful, the APRT is mutated and the cells can grow and regenerate into plants on the 2FA medium. The DNA from these plants is further analyzed to detect edited sequences. In potato, the widely used acetolactate synthase (ALS) has been used for selection. ALS confers resistance to several herbicides, particularly chlorsulfuron, and the specific amino acid change in StALS Pro-187/186 to serine was targeted. In addition, the primary selection of transgenics was on kanamycin. A PE-PE2 system was designed by fusing hygromycin phosphotransferase (Hpt) to the C-terminus of the nSpCas9-M-MLV region with P2A, a self-cleaving 2A peptide, driven by Ubiquitin promoter of maize. PE-PE2 increased the editing efficiency by about threefold for three pegRNAs and gave improved editing frequencies (Perroud et al., 2022).

The ability to introduce both transversions and transitions is by far the most significant attribute of prime editing technology. In addition, PEs have been found to successfully introduce insertions, deletions, transitions, and transversions (Anzalone et al., 2019). Perroud et al. (2022) reported that 0.06% of transformed protoplasts of *Physcomitrium* were edited, which is less than the standard Cas9 mediated and base editing mutagenic strategies. However, the edit's specificity is higher than CRISPR/Cas systems, and off-targets are few or none. Substitutions, insertions, and deletions have been observed in the different taxa using the varied versions of prime editors.

The editing efficiency was similar in PE2- and PE3-based vectors in *Physcomitrium*, whereas in potato, same PE3 constructs failed to edit the *ALS* gene, which could be edited by PE2-based vectors albeit at low frequencies. In rice,

editing efficiencies were between 1.55% and 31.3% (Hua et al., 2020b; Butt et al., 2020; Li et al., 2020d; Lin et al., 2020; Tang et al., 2020; Xu et al., 2020). The editing efficiencies ranged from 0.7% to 2.2%. Overall, the PE3 strategies were less efficient in plant cells than animal cells. However, further modifications and adaptation of the technique would standardize prime editing for more crop systems. Wang et al. (2021) have reported insertion of up to 66 bases in *Arabidopsis* protoplasts, which is a four-fold increase over the 15-base insertion reported in rice. For prime editing in dicots and monocots, easy-use vectors on PE2 and PE3 strategies have been created, named pPPED and pPPEM (Wang et al., 2021). They have designed a pPEG cassette for insertion of peg RNA or sgRNA, and then pPEG is inserted in the vectors PPEM or PPED. The pPPED vector was targeted in *Arabidopsis*. Editing efficiency is thus influenced by the length of reverse transcriptase and primer-binding site in the designed pegRNAs and sgRNAs.

In addition to the biological parameters (plant taxa, molecular toolkit, transformation, and regeneration system), the physical temperature parameter has a profound impact on the editing frequencies. Because the efficiency of the M-MLV reverse transcriptase is enhanced at higher temperatures, 32°C and 37°C were tested, but no significant differences were reported. However, the temperature variations were also tried in prime editing (PPE) systems at 26°C and 37°C in rice, giving significantly higher editing activity at 37°C (Lin et al., 2020).

In summary, the modifications in the design of constructs, particularly to avoid by-products resulting from the scaffold of the pegRNAs and reduction of off-targets, have been found to increase the editing efficiencies. Gao (2015) suggested the shift from a knock-out strategy to a knock-in strategy by employing the homologous recombination process of DNA repair to increase targeted mutagenesis. This has been incorporated as a key attribute in the prime editing technology. Among the diverse strategies designed to achieve targeted mutagenesis, prime editing is a landmark advancement in methods achieving increased efficiency and reduced off-target effects. This method, for the first time, presented an efficient strategy to introduce all the 12-point mutations. With the availability of many diverse vectors (editors and pegRNAs) developed by the different research groups and web-based design algorithms available (Peg-finder, PE-Designer /PE-Analyzer, pegIT, PrimeDesign, and PlantPegDesigner), the deployment of this technique is at the threshold of revolutionizing precision breeding of crop plants. As most of the genes of importance rely on altering a few and specific nucleotide changes to confer traits rather than large-scale alteration of genes, prime editing presents an opportunity to drive the development of gene editing platforms that are precise, effective, and elegant.

8 Conclusion

Under the scenario of ever-rising food demands and climate change, there is tremendous pressure on scientists and breeders to speed up the development of climate-resilient-high-yielding cultivars. The application of molecular breeding approaches has achieved great success in accelerating performance gains in various crops in the past decade. However, the need of the hour is to integrate new biotechnological methods and technologies in the existing breeding programs to further realize genetic gains. The unprecedented advances made in GE technologies have shown great potential in genetic enhancement and boosting crop production. This review highlights how newly evolved CRISPR/Cas systems have successfully brought about a paradigm shift in crop improvement programs. There has been a significant advancement in understanding the functions of gene complexes underpinning complex traits, which was extremely daunting using the existing gene discovery approaches. The efficient use of GE tools in manipulating complex traits, especially in polyploid crops, has now become feasible, especially when used in combination with the next-generation sequencing platforms.

Despite the substantial deployment of the CRISPR/Cas platform in developing crops with desired traits, studies demonstrating the translation of the laboratory-based results into the field have been anecdotal. In addition to being relevant at the genome level, the improved traits must also be realized in the field without any trade-offs or counter effects on other traits of importance. Additionally, any genome strategy developed should pose no threat to the environment and should be able to reduce the application of pesticides and fertilizers. One of the major challenges in developing cultivars by the GE route is rooted in low transformation and regeneration efficiencies. Numerous agronomically important crops such as sunflower, cotton, and many others either have long transformation protocols with low efficiencies or are outrightly recalcitrant. In addition, in crops where transformation protocols have been established, regeneration efficiencies remain low, making the application of GE strategies challenging.

Furthermore, public acceptance of GE-modified crops has not come of age yet. A common misconception about these crops adversely affecting health and the environment has led many farmers to avoid reaping benefits from growing these crop cultivars. This bias automatically trickles down to the consumers and, in turn, results in limited acceptance of these crops for public consumption. Therefore, we believe, scientists across the globe need to ensure a healthy flow of information using present-day outreach tools, including social media, to educate the consumers about the differences between transgenic approaches and the risks and benefits of using modern GE-modified crops.

Although GE platforms are radically different, precise, and superior to traditional transgenic approaches, at the moment, these methods still go through governmental scrutiny and assessment in many countries. Nonetheless, in the foreseeable future, new-age GE platforms in plants are contemplated to be employed as a tool for efficiently engineering the majority of crop

plants. We expect and hope that these methods can be integrated into breeding programs globally with relatively lesser regulatory procedures compared to conventional transgenic approaches. The development of these measures will need comparable attention and consistent research efforts to continually assess developed crop varieties on various climatic and genomic parameters, especially in our present-day rapidly changing climate and pest pressure.

9 Future directions

The evolution of various GE platforms has made it possible for molecular biologists to precisely target gene(s) of interest. Primarily, only CRISPR/Cas has been used for gene editing. Only recently, techniques such as epigenome editing, prime editing, and base editing have been used for gene editing. These techniques are powerful alternative strategies that have been developed for gene editing in plants. However, glaring challenges still exist that continue to impede the goals of achieving sustainable crop production. These challenges stem from the complexity of both endogenous and exogenous cues in plant development, making it nearly impossible for any single GE platform to deliver efficiently. Present-day advances in GE protocols need to be primed toward generating platforms that are more precise, efficient, accurate, and, most importantly, feasible. At first, no off-target silencing should result from using these methods. Secondly, the delivery and results obtained in crop plants should not vary from species to species. In addition, the genomic changes should be traceable in future generations with precision and also remain feasible with respect to cost and labor. Lastly, at present, we need more dynamic regulatory measures in place to ease the development and use of these platforms in crop improvement programs.

Author contributions

VRR conceptualized and finalized the manuscript. PD, DS, SV, AS, AC and VRR participated in preparing and curating the manuscript and the revision. PD, SNR, and VRR helped in preparing and finalizing the draft of the MS. All authors read and approved the MS.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer KP declared a shared affiliation with the author(s) AC, VRR to the handling editor at the time of review.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Akama, K., Akter, N., Endo, H., Kanesaki, M., Endo, M., Toki, S., et al. (2020). An *in vivo* targeted deletion of the calmodulin-binding domain from rice glutamate decarboxylase 3 (Os GAD3) increases γ-aminobutyric acid content in grains. *Rice* 13 (1), 20. doi:10.1186/s12284-020-00380-w

Ali, Z., Ali, S., Tashkandi, M., Zaidi, S. S., and Mahfouz, M. M. (2016). CRISPR/Cas9- mediated immunity to geminiviruses: Differential interference and evasion. *Sci. Rep.* 6, 26912. doi:10.1038/srep26912

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513 (7519), 569–573.

Amitai, G., and Sorek, R. (2016). CRISPR–Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.* 14 (2), 67–76. doi:10.1038/nrmicro.2015.14

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157. doi:10.1038/s41586-019-1711-4

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712. doi:10.1126/science.1138140

Ben Shlush, I., Samach, A., Melamed-Bessudo, C., Ben-Tov, D., Dahan-Meir, T., Filler- Hayut, S., et al. (2021). CRISPR/Cas9 induced somatic recombination at the CRTISO locus in tomato. *Genes* 12 (1), 59. doi:10.3390/genes12010059

Bharat, S. S., Li, S., Li, J., Yan, L., and Xia, L. (2020). Base editing in plants: Current status and challenges. *Crop J.* 8 (3), 384–395.

Bhowmik, P., Ellison, E., Polley, B., Bollina, V., Kulkarni, M., Ghanbarnia, K., et al. (2018). Targeted mutagenesis in wheat microspores using CRISPR/Cas9. *Sci. Rep.* 8 (1), 6502. doi:10.1038/s41598-018-24690-8

Boch, J., and Bonas, U. (2010). Xanthomonas AvrBs3 family-type III effectors: Discovery and function. *Annu. Rev. Phytopathol.* 418, 419–436. doi:10.1146/annurev-phyto-080508-081936

Bollen, Y., Post, J., Koo, B. K., and Snippert, H. J. (2018). How to create state-of-the-art genetic model systems: strategies for optimal CRISPR-mediated genome editing. *Nucleic Acids Res.* 46 (13), 6435–6454. doi:10.1093/nar/gky571

Brocken, D. J. W., Tark-Dame, M., and Dame, R. T. (2018). dCas9: A versatile tool for epigenome editing. *Curr. Issues Mol. Biol.* 26, 15–32. doi:10.21775/cimb.026.015

Butt, H., Rao, G. S., Sedeek, K., Aman, R., Kamel, R., Mahfouz, M., et al. (2020). Engineering herbicide resistance via prime editing in rice. *Plant Biotechnol. J.* 18, 2370–2372. doi:10.1111/pbi.13399

Calvache, C., Vazquez-Vilar, M., Selma, S., Uranga, M., Fernández-del-Carmen, A., Daròs, J. A., et al. (2022). Strong and tunable anti-CRISPR/Cas activities in plants. *Plant Biotechnol. J.* 20 (2), 399–408. doi:10.1111/pbi.13723

Čermák, T., Baltes, N. J., Čegan, R., Zhang, Y., and Voytas, D. F. (2015). High-frequency, precise modification of the tomato genome. *Genome Biol.* 16 (1), 232. doi:10.1186/s13059-015-0796-9

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., et al. (2016). Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* 17, 1140–1153. doi:10.1111/mpp.12375

Chaudhuri, A., Halder, K., and Datta, A. (2022). Classification of CRISPR/Cas system and its application in tomato breeding. *Theor. Appl. Genet.* 135, 367. doi:10.1007/s00122-021-03984-y

Choi, M., Yun, J. Y., Kim, J. H., Kim, J. S., and Kim, S. T. (2021). The efficacy of CRISPR-mediated cytosine base editing with the RPS5a promoter in Arabidopsis thaliana. *Sci. Rep.* 11, 8087. doi:10.1038/s41598-021-87669-y

Chen, L., Park, J. E., Paa, P., Rajakumar, P. D., Prekop, H. T., Chew, Y. T., et al. (2021). Programmable C: G to G: C genome editing with CRISPR-cas9- directed base excision repair proteins. *Nat. Commun.* 12 (1), 1384. doi:10.1038/s41467-021-21559-9

Chen, Y., Wang, Z., Ni, H., Xu, Y., Chen, Q., Jiang, L., et al. (2017). CRISPR/Cas9-mediated base-editing system efficiently generates gain-of-function mutations in Arabidopsis. *Sci. China. Life Sci.* 60 (5), 520–523. doi:10.1007/s11427-017-9021-5

Cheng, H., Hao, M., Ding, B., Mei, D., Wang, W., Wang, H., et al. (2021). Base editing with high efficiency in allotetraploid oilseed rape by A3A-PBE system. *Plant Biotechnol. J.* 19 (1), 87–97. doi:10.1111/pbi.13444

Cheng, T. L., Li, S., Yuan, B., Wang, X., Zhou, W., and Qiu, Z. (2019). Expanding C-T base editing toolkit with diversified cytidine deaminases. *Nat. Commun.* 10, 3612. doi:10.1038/s41467-019-11562-6

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339 (6121), 819–823. doi:10.1126/science.1231143

Cong, L., Zhou, R., Kuo, Y. C., Cunniff, M., and Zhang, F. (2012). Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat. Commun.* 3, 968. doi:10.1038/ncomms1962

D'Ambrosio, C., Stigliani, A. L., and Giorio, G. (2018). CRISPR/Cas9 editing of carotenoid genes in tomato. *Transgenic Res.* 27 (4), 367–378. doi:10.1007/s11248-018-0079-9

Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K., Semenova, E., et al. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* 3, 945. doi:10.1038/ncomms1937

Deb, S., Choudhury, A., Kharbyngar, B., and Satyawada, R. R. (2022). Applications of CRISPR/Cas9 technology for modification of the plant genome. *Genetica*, 150. 1–12. doi:10.1007/s10709-021-00146-2

Dong, L., Qi, X., Zhu, J., Liu, C., Zhang, X., Cheng, B., et al. (2019). Supersweet and waxy: meeting the diverse demands for specialty maize by genome editing. *Plant Biotechnol. J.* 17 (10), 1853–1855. doi:10.1111/pbi.13144

Dong, O. X., Yu, S., Jain, R., Zhang, N., Duong, P. Q., Butler, C., et al. (2020). Marker-free carotenoid-enriched rice generated through targeted gene insertion using CRISPR-Cas9. *Nat. Commun.* 11 (1), 1178. doi:10.1038/s41467-020-14981-y

Endo, A., Masafumi, M., Kaya, H., and Toki, S. (2016). Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from *Francisella novicida*. *Sci. Rep.* 6, 38169. doi:10.1038/srep38169

Endo, M., Mikami, M., Endo, A., Kaya, H., Itoh, T., Nishimasu, H., et al. (2019). Genome editing in plants by engineered CRISPR–Cas9 recognizing NG PAM. *Nature Plants* 5 (1), 14–17.

Fan, S., Zhang, L., Tang, M., Cai, Y., Liu, J., Liu, H., et al. (2021). CRISPR/Cas9-targeted mutagenesis of the BnaA03. BP gene confers semi-dwarf and compact architecture to rapeseed (Brassica napus L.). *Plant Biotechnol. J.* 19 (12), 2383–2385. doi:10.1111/pbi.13703

Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., and Charpentier, E. (2016). The CRISPR- associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* 532, 517–521. doi:10.1038/nature17945

Gahlaut, S. K., Savargaonkar, D., Sharan, C., Yadav, S., Mishra, P., Singh, J. P., et al. (2020). SERS platform for dengue diagnosis from clinical samples employing a hand held Raman spectrometer. *Anal. Chem.* 92 (3), 2527–2534. doi:10.1021/acs.analchem.9b04129

Gahlaut, V., Zinta, G., Jaiswal, V., and Kumar, S. (2020). Quantitative epigenetics: A new avenue for crop improvement. *Epigenomes* 4, 25. doi:10.3390/epigenomes4040025

Gallego-Bartolomé, J. (2020). DNA methylation in plants: Mechanisms and tools for targeted manipulation. *New Phytol.* 227, 38–44. doi:10.1111/nph.16529

Gallego-Bartolomé, J., Gardiner, J., Liu, W., Papikian, A., Ghoshal, B., Kuo, H. Y., et al. (2018). Targeted DNA demethylation of the *Arabidopsis* genome using the human TET1 catalytic domain. *Proc. Natl. Acad. Sci. U. S. A.* 115, E2125–E2134. doi:10.1073/pnas.1716945115

Gallego-Bartolome, J., Liu, W., Kuo, P. H., Feng, S., Ghoshal, B., Gardiner, J., et al. (2019). Co-targeting RNA polymerases IV and V promotes efficient de novo DNA methylation in *Arabidopsis*. *Cell* 176, 1068–1082. doi:10.1016/j.cell.2019.01.029

Gao, C. (2015). Genome editing in crops: from bench to field. *Nat. Sci. Rev.* 2 (1), 13–15.

Gao, H., Gadlage, M. J., Lafitte, H. R., Lenderts, B., Yang, M., Schroder, M., et al. (2020). Superior field performance of waxy corn engineered using CRISPR–Cas9. *Nat. Biotechnol.* 38 (5), 579–581. doi:10.1038/s41587-020-0444-0

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., et al. (2017). Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage. *Nature* 551 (7681), 464–471. doi:10.1038/nature24644

Gehrke, J. M., Cervantes, O., Clement, M. K., Wu, Y., Zeng, J., Bauer, D. E., et al. (2018). An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* 36 (10), 977–982. doi:10.1038/nbt.4199

Ghoshal, B., Picard, C. L., Vong, B., Feng, S., and Jacobsen, S. E. (2021). CRISPR-based targeting of DNA methylation in *Arabidopsis thaliana* by a bacterial CG-specific DNA methyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2125016118. doi:10.1073/pnas.2125016118

Giudice, G., Moffa, L., Varotto, S., Cardone, M. F., Bergamini, C., De Lorenzis, G., et al. (2021). Novel and emerging biotechnological crop protection approaches. *Plant Biotechnol. J.* 19, 1495–1510. doi:10.1111/pbi.13605

Grunewald, J., Zhou, R. H., Lareau, C. A., Garcia, S. P., Iyer, S., Miller, B. R., et al. (2020). A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing. *Nat. Biotechnol.* 38 (7), 861–864. doi:10.1038/s41587-020-0535-y

Guilinger, J. P., Thompson, D. B., and Liu, D. R. (2014). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 32 (6), 577–582. doi:10.1038/nbt.2909

Guo, H., Wu, T., Li, S., He, Q., Yang, Z., Zhang, W., et al. (2019). The methylation patterns and transcriptional responses to chilling stress at the seedling stage in rice. *Int. J. Mol. Sci.* 20, 5089. doi:10.3390/ijms20205089

Han, H., Wu, Z., Zheng, L., Han, J., Zhang, Y., Li, J., et al. (2022). Generation of a high-efficiency adenine base editor with TadA8e for developing wheat dinitroaniline-resistant germplasm. *Crop J.* 10 (2), 368–374.

Hassan, M., Yuan, G., Chen, J. G., Tuskan, G. A., and Yang, X. (2020). Prime editing technology and its prospects for future applications in plant biology research. *BioDesign Res.* 2020.

Hess, G. T., Fresard, L., Han, K., Lee, C. H., Li, A., Cimprich, K. A., et al. (2016). Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* 13, 1036–1042. doi:10.1038/nmeth.4038

Hong, Y., Meng, J., He, X., Zhang, Y., Liu, Y., Zhang, C., et al. (2021). Editing miR482b and miR482c simultaneously by CRISPR/Cas9 enhanced tomato resistance to Phytophthora infestans. *Phytopathology* 111 (6), 1008–1016. doi:10.1094/PHYTO-08-20-0360-R

Hu, L., Amoo, O., Liu, Q., Cai, S., Zhu, M., Shen, X., et al. (2020). Precision genome engineering through cytidine base editing in rapeseed (Brassica napus. L). *Front. Genome Ed.* 15, 605768. doi:10.3389/fgeed.2020.605768

Hu, X., Meng, X., Liu, Q., Li, J., and Wang, K. (2018). Increasing the efficiency of CRISPR- Cas9-VQR precise genome editing in rice. *Plant Biotechnol. J.* 16 (1), 292–297. doi:10.1111/pbi.12771

Hua, K., Jiang, Y., Tao, X., and Zhu, J. K. (2020a). Precision genome engineering in rice using prime editing system. *Plant Biotechnol. J.* 18, 2167–2169. doi:10.1111/pbi.13395

Hua, K., Tao, X., Liang, W., Zhang, Z., Gou, R., Zhu, J. K., et al. (2020b). Simplified adenine base editors improve adenine base editing efficiency in rice. *Plant Biotechnol. J.* 18 (3), 770–778. doi:10.1111/pbi.13244

Hua, K., Tao, X., Yuan, F., Wang, D., and Zhu, J. K. (2018). Precise A·T to G·C base editing in the rice genome. *Mol. Plant* 11 (4), 627–630. doi:10.1016/j.molp.2018.02.007

Hua, K., Tao, X., and Zhu, J. K. (2019). Expanding the base editing scope in rice by using Cas9 variants. *Plant Biotechnol. J.* 17 (2), 499–504. doi:10.1111/pbi.12993

Hunziker, J., Nishida, K., Kondo, A., Kishimoto, S., Ariizumi, T., and Ezura, H. (2020). Multiple gene substitution by Target-AID base-editing technology in tomato. *Sci. Rep.* 10, 20471. doi:10.1038/s41598-020-77379-2

Huang, H., Cui, T., Zhang, L., Yang, Q., Yang, Y., Xie, K., et al. (2020). Modifications of fatty acid profile through targeted mutation at BnaFAD2 gene with CRISPR/Cas9-mediated gene editing in Brassica napus. *Theor. Appl. Genet.* 133 (8), 2401–2411. doi:10.1007/s00122-020-03607-y

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* 169 (12), 5429-5433. doi:10.1128/jb.169.12.5429-5433

Jaganathan, D., Ramasamy, K., Sellamuthu, G., Jayabalan, S., and Venkataraman, G. (2018). CRISPR for crop improvement: An update review. *Front. Plant Sci.* 9, 985. doi:10.3389/fpls.2018.00985

Jiang, F., and Doudna, J. A. (2017). CRISPR–Cas9 structures and mechanisms. *Annu. Rev. Biophys.* 46, 505–529. doi:10.1146/annurev-biophys-062215-010822

Jiang, T., Zhang, X. O., Weng, Z., and Xue, W. (2022). Deletion and replacement of long genomic sequences using prime editing. *Nat. Biotechnol.* 40 (2), 227–234. doi:10.1038/s41587-021-01026-y

Jiang, T., Zhang, X. O., Weng, Z., and Xue, W. (2021). *Programming large target genomic deletion and concurrent insertion via a prime editing-based method: Pedar*. USA: Cold Spring Harbor Laboratory. *bioRxiv*.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., Charpentier, E., et al. (2012). A programmable dual- RNA– guided DNA endonuclease in adaptive bacterial immunity. *Science* 337 (6096), 816–821. doi:10.1126/science.1225829

Johnson, L. M., Du, J., Hale, C. J., Bischof, S., Feng, S., Chodavarapu, R. K., et al. (2014). SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* 507, 124–128. doi:10.1038/nature12931

Kakoulidou, I., Avramidou, E. V., Baránek, M., Brunel-Muguet, S., Farrona, S., Johannes, F., et al. (2021). Epigenetics for crop improvement in times of global change. *Biology* 10, 766. doi:10.3390/biology10080766

Kantor, A., McClements, M. E., and MacLaren, R. E. (2020). CRISPR-Cas9 DNA base- editing and prime-editing. *Int. J. Mol. Sci.* 21 (17), 6240. doi:10.3390/ijms21176240

Karginov, F. V., and Hannon, G. J. (2010). The CRISPR system: Small RNA-guided defense in bacteria and archaea. *Mol. Cell* 37 (1), 7–19. doi:10.1016/j.molcel.2009.12.033

Kashojiya, S., Lu, Y., Takayama, M., Komatsu, H., Minh, L. H. T., Nishida, K., et al. (2022). Modification of tomato breeding traits and plant hormone signaling by target-AID, the genome-editing system inducing efficient nucleotide substitution. *Hortic. Res.* 9, uhab004. doi:10.1093/hr/uhab004

Kaya, H., Ishibashi, K., and Toki, S. (2017). A split *Staphylococcus aureus* Cas9 as a compact genome-editing tool in plants. *Plant Cell Physiol.* 58 (4), 643–649. doi:10.1093/pcp/pcx034

Khan, I., Khan, S., Zhang, Y., Zhou, J., Akhoundian, M., Jan, S. A., et al. (2021). CRISPR-Cas technology-based genome editing for modification of salinity stress tolerance responses in rice (Oryza sativa L.). *Mol. Biol. Rep.* 48 (4), 3605–3615. doi:10.1007/s11033-021-06375-0

Kim, H., Kim, S. T., Ryu, J., Kang, B. C., Kim, J. S., Kim, S. G., et al. (2017). CRISPR/Cpf1- mediated DNA-free plant genome editing. *Nat. Commun.* 8, 14406. doi:10.1038/ncomms14406

Kim, Y. G., Cha, J., and Chandrasegaran, S. (2007). Hybrid restriction enzymes: zinc finger fu- sions to fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.* 93, 1156–1160. doi:10.1073/pnas.93.3.1156

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Topkar, V. V., Zheng, Z., et al. (2015). Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* 33 (12), 1293–1298. doi:10.1038/nbt.3404

Komivi, D., Marie, A. M., Rong, Z., Qi, Z., Mei, Y., Ndiaga, C., et al. (2018). The contrasting response to drought and waterlogging is underpinned by divergent DNA methylation programs associated with transcript accumulation in sesame. *Plant Sci.* 277, 207–217. doi:10.1016/j.plantsci.2018.09.012

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533 (7603), 420–424. doi:10.1038/nature17946

Konermann, S., Brigham, M. D., Trevino, A. E., Hsu, P. D., Heidenreich, M., Cong, L., et al. (2013). Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 500 (7463), 472–476. doi:10.1038/nature12466

Koonin, E. V., and Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374 (1772), 20180087. doi:10.1098/rstb.2018.0087

Kuang, Y., Li, S., Ren, B., Yan, F., Spetz, C., Li, X., et al. (2020). Base-editing-mediated artificial evolution of OsALS1 in planta to develop novel herbicide-tolerant rice germplasms. *Mol. Plant* 13 (4), 565–572.

Kumar, A., Singh, P., Sharma, S., and Pandey, A. K. (2021). "Shifting paradigm towards the crops: From model plants to crops and employing the genome engineering to target traits," in *Agricultural Biotechnology: Latest research and trends* (Singapore: Springer), 511–535.

Kumar, S., Beena, A. S., Awana, M., and Singh, A. (2017). Salt-induced tissue-specific cytosine methylation downregulates expression of HKT genes in contrasting wheat (*Triticum aestivum* L.) genotypes. *DNA Cell Biol.* 36, 283–294. doi:10.1089/dna.2016.3505

Kurt, I. C., Zhou, R., Iyer, S., Garcia, S. P., Miller, B. R., Langner, L. M., et al. (2021). CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* 39, 41–46. doi:10.1038/s41587-020-0609-x

Lee, C. M., Cradick, T. J., and Bao, G. (2016). The Neisseria meningitidis CRISPR-Cas9 system enables specific genome editing in mammalian cells. *Mol. Ther.* 24 (3), 645–654. doi:10.1038/mt.2016.8

Li, C., Zong, Y., Wang, Y., Jin, S., Zhang, D., Song, Q., et al. (2018). Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol.* 19 (1), 1–9.

Li, C., Zhang, R., Meng, X., Chen, S., Zong, Y., Lu, C., et al. (2020). Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. *Nat. Biotechnol.* 38 (7), 875–882. doi:10.1038/s41587-019-0393-7

Li, G., Sretenovic, S., Eisenstein, E., Coleman, G., and Qi, Y. (2021). Highly efficient C-to-T and A-to-G base editing in a Populus hybrid. *Plant Biotechnol. J.* 19 (6), 1086–1088. doi:10.1111/pbi.13581

Li, H., Li, J., Chen, J., Yan, L., and Xia, L. (2020a). Precise modifications of both exogenous and endogenous genes in rice by prime editing. *Mol. Plant* 13, 671–674. doi:10.1016/j.molp.2020.03.011

Li, H., Li, J., Chen, J., Yan, L., and Xia, L. (2020b). Precise modifications of both exogenous and endogenous genes in rice by prime editing. *Mol. Plant* 13, 671–674. doi:10.1016/j.molp.2020.03.011

Li, H., Zhao, X., Dai, H., Wu, W., Mao, W., Zhang, Z., et al. (2012). Tissue culture responsive microRNAs in strawberry. *Plant Mol. Biol. Rep.* 30, 1047–1054. doi:10.1007/s11105-011-0406-2

Li, J., Wang, Z., He, G., Ma, L., and Deng, X. W. (2020c). CRISPR/Cas9-mediated disruption of TaNP1 genes results in complete male sterility in bread wheat. *J. Genet. Genomics* 47 (5), 263–272. doi:10.1016/j.jgg.2020.05.004

Li, J., Yang, D. L., Huang, H., Zhang, G., He, L., Pang, J., et al. (2020d). Epigenetic memory marks determine epiallele stability at loci targeted by de novo DNA methylation. *Nat. Plants* 6, 661–674. doi:10.1038/s41477-020-0671-x

Li, J., Zhang, X., Sun, Y., Zhang, J., Du, W., Guo, X., et al. (2018). Efficient allelic replacement in rice by gene editing: a case study of the *NRT1.1B* gene. *J. Integr. Plant Biol.* 60, 536–540. doi:10.1111/jipb.12650

Li, M., Li, X., Zhou, Z., Wu, P., Fang, M., Pan, X., et al. (2016). Reassessment of the four yield-related genes Gn1a, DEP1, GS3, and IPA1 in rice using a CRISPR/Cas9 system. *Front. Plant Sci.* 7, 377. doi:10.3389/fpls.2016.00377

Li, P., Li, Y. J., Zhang, F. J., Zhang, G. Z., Jiang, X. Y., Yu, H. M., et al. (2017). The Arabidopsis UDP-glycosyltransferases UGT79B2 and UGT79B3, contribute to cold, salt and drought stress tolerance via modulating anthocyanin accumulation. *Plant J.* 89, 85–103. doi:10.1111/tpj.13324

Li, R., Char, S. N., Liu, B., Liu, H., Li, X., Yang, B., et al. (2021). High-efficiency plastome base editing in rice with TAL cytosine deaminase. *Mol. Plant* 14 (9), 1412–1414. doi:10.1016/j.molp.2021.07.007

Li, R., Li, R., Li, X., Fu, D., Zhu, B., Tian, H., et al. (2017). Multiplexed CRISPR/Cas9-mediated metabolic engineering of γ-aminobutyric acid levels in Solanum lycopersicum. *Plant Biotechnol. J.* 16, 415–427. doi:10.1111/pbi.12781

Li, R., Liu, C., Zhao, R., Wang, L., Chen, L., Yu, W., et al. (2019). CRISPR/Cas9-Mediated SlNPR1 mutagenesis reduces tomato plant drought tolerance. *BMC Plant Biol.* 19 (1), 38. doi:10.1186/s12870-018-1627-4

Li, R., Zhang, L., Wang, L., Chen, L., Zhao, R., Sheng, J., et al. (2018). Reduction of tomato-plant chilling tolerance by CRISPR–Cas9-mediated SlCBF1 mutagenesis. *J. Agric. Food Chem.* 66 (34), 9042–9051. doi:10.1021/acs.jafc.8b02177

Liang, Y., Xie, J., Zhang, Q., Wang, X., Gou, S., Lin, L., et al. (2022). AGBE: a dual deaminase-mediated base editor by fusing CGBE with ABE for creating a saturated mutant population with multiple editing patterns. *Nucleic Acids Res.* 50 (9), 5384–5399. doi:10.1093/nar/gkac353

Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., et al. (2020). Prime genome editing in rice and wheat. *Nat. Biotechnol.* 38, 582–585. doi:10.1038/s41587-020-0455-x

Liu, H., Ding, Y., Zhou, Y., Jin, W., Xie, K., Chen, L. L., et al. (2017). CRISPR-P 2.0: an improved CRISPR-cas9 tool for genome editing in plants. *Mol. Plant* 10 (3), 530–532. Epub 2017 Jan 13. doi:10.1016/j.molp.2017.01.003

Liu, L., Gallagher, J., Arevalo, E. D., Chen, R., Skopelitis, T., Wu, Q., et al. (2021). Enhancing grain-yield-related traits by CRISPR–Cas9 promoter editing of maize CLE genes. *Nat. Plants* 7 (3), 287–294. doi:10.1038/s41477-021-00858-5

Liu, L., Zhang, J., Xu, J., Li, Y., Guo, L., Wang, Z., et al. (2020). CRISPR/Cas9 targeted mutagenesis of SlLBD40, a lateral organ boundaries domain transcription factor, enhances drought tolerance in tomato. *Plant Sci.* 301, 110683. doi:10.1016/j.plantsci.2020.110683

Liu, X., Homma, A., Sayadi, J., Yang, S., Ohashi, J., Takumi, T., et al. (2016). Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system. *Sci. Rep.* 6, 19675. doi:10.1038/srep19675

Liu, X., Zhang, S., Jiang, Y., Yan, T., Fang, C., Hou, Q., et al. (2022). Use of CRISPR/Cas9- based gene editing to simultaneously mutate multiple homologous genes required for pollen development and male fertility in maize. *Cells* 11 (3), 439. doi:10.3390/cells11030439

Lloyd, A., Plaisier, C. L., Carroll, D., and Drews, G. N. (2005). Targeted mutagenesis using zinc- finger nucleases in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2232–2237. doi:10.1073/pnas.0409339102

Lou, D., Wang, H., Liang, G., and Yu, D. (2017). OsSAPK2 confers abscisic acid sensitivity and tolerance to drought stress in rice. *Front. Plant Sci.* 8, 993. doi:10.3389/fpls.2017.00993

Lu, Y., and Zhu, J. K. (2017). Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. *Mol. Plant* 10 (3), 523–525.

Lu, Y., Tian, Y., Shen, R., Yao, Q., Zhong, D., Zhang, X., et al. (2020). Precise genome modification in tomato using an improved prime editing system. *Plant Biotechnol. J.* 19, 415–417. doi:10.1111/pbi.13497

Ma, L., Zhang, D., Miao, Q., Yang, J., Xuan, Y., Hu, Y., et al. (2017). Essential role of sugar transporter *OsSWEET11*during the early stage of rice grain filling. *Plant Cell Physiol.* 58, 863–873. doi:10.1093/pcp/pcx040

Macovei, A., Sevilla, N. R., Cantos, C., Jonson, G. B., Slamet-Loedin, I., Čermák, T., et al. (2018). Novel alleles of rice eIF4G generated by CRISPR/Cas9-targeted mutagenesis confer resistance to Rice tungro spherical virus. *Plant Biotechnol. J.* 16 (11), 1918–1927. doi:10.1111/pbi.12927

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* 13 (11), 722–736. doi:10.1038/nrmicro3569

Malabarba, J., Chevreau, E., Dousset, N., Veillet, F., Moizan, J., Vergne, E., et al. (2021). New strategies to overcome present CRISPR/Cas9 limitations in apple and pear: efficient dechimerization and base editing. *Int. J. Mol. Sci.* 22 (1), 319. doi:10.3390/ijms22010319

Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F., Zhu, J. K., et al. (2013). Application of the CRISPR–Cas system for efficient genome engineering in plants. *Mol. Plant* 6 (6), 2008–2011. doi:10.1093/mp/sst121

Martínez, M. I. S., Bracuto, V., Koseoglou, E., Appiano, M., Jacobsen, E., Visser, R. G., et al. (2020). CRISPR/Cas9-targeted mutagenesis of the tomato susceptibility gene PMR4 for resistance against powdery mildew. *BMC plant Biol.* 20 (1), 1–13.

Miglani, G. S., Kaur, A., and Lovepreet Kaur, L. (2020). Plant gene expression control using genome- and epigenome-editing technologies. *J. Crop Improv.* 34, 1–63. doi:10.1080/15427528.2019.1678541

Miglani, G. S., and Singh, R. (2020). "5 epigenome editing in crop improvement," in *Quantitative genetics, genomics and plant breeding*. 2nd edition (UK: CAB International), 44–70.

Modrzejewski, D., Hartung, F., Lehnert, H., Sprink, T., Kohl, C., Keilwagen, J., et al. (2020). Which factors affect the occurrence of off-target effects caused by the use of CRISPR/cas: a systematic review in plants. *Front. Plant Sci.* 11, 574959. doi:10.3389/fpls.2020.574959

Mojica, M. J., Juez, G., and Rodríguez-Valera, F. (1993). Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites. *Mol. Microbiol.* 9, 613–621. doi:10.1111/j.1365-2958.1993.tb01721.x

Moore, J. K., and Haber, J. E. (1996). Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 16, 2164–2173. doi:10.1128/mcb.16.5.2164

Nakamura, M., Gao, Y., Dominguez, A. A., and Qi, L. S. (2021). CRISPR technologies for precise epigenome editing. *Nat. Cell Biol.* 23, 11–22. doi:10.1038/s41556-020-00620-7

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353 (6305), aaf8729.

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., et al. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* 361 (6408), 1259–1262. doi:10.1126/science.aas9129

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., Doudna, J. A., et al. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat. Struct. Mol. Biol.* 21, 528–534. doi:10.1038/nsmb.2820

Paixão, J. F. R., Gillet, F. X., Ribeiro, T. P., Bournaud, C., Lourenço-Tessutti, I. T., Noriega, D. D., et al. (2019). Improved drought stress tolerance in Arabidopsis by CRISPR/dCas9 fusion with a histone acetyltransferase. *Sci. Rep.* 9, 8080. doi:10.1038/s41598-019-44571-y

Papikian, A., Liu, W., Gallego-Bartolomé, J., and Jacobsen, S. E. (2019). Site-specific manipulation of *Arabidopsis* loci using CRISPR-Cas9 SunTag systems. *Nat. Commun.* 10, 729. doi:10.1038/s41467-019-08736-7

Peng, A., Chen, S., Lei, T., Xu, L., He, Y., Wu, L., et al. (2017). Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene *CsLOB1* promoter in citrus. *Plant Biotechnol. J.* 15, 1509–1519. doi:10.1111/pbi.12733

Perroud, P. F., Guyon-Debast, A., Veillet, F., Kermarrec, M. P., Chauvin, L., Chauvin, J. E., et al. (2022). Prime Editing in the model plant Physcomitrium patens and its potential in the tetraploid potato. *Plant Sci.* 316, 111162. doi:10.1016/j.plantsci.2021.111162

Puchta, H. (2005). The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J. Exp. Bot.* 56 (409), 1–14. doi:10.1093/jxb/eri025

Pyott, D. E., Sheehan, E., and Molnar, A. (2016). Engineering of CRISPR/Cas9-mediated potyvirus resistance in transgene-free Arabidopsis plants. *Mol. Plant Pathol.* 17 (8), 1276–1288. doi:10.1111/mpp.12417

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., et al. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183. doi:10.1016/j.cell.2013.02.022

Qin, R., Li, J., Li, H., Zhang, Y., Liu, X., Miao, Y., et al. (2019). Developing a highly efficient and wildly adaptive CRISPR-SaCas9 toolset for plant genome editing. *Plant Biotechnol. J.* 17 (4), 706–708. doi:10.1111/pbi.13047

Quadrana, L., Almeida, J., Asís, R., Duffy, T., Dominguez, P. G., Bermúdez, L., et al. (2014). Natural occurring epialleles determine vitamin E accumulation in tomato fruits. *Nat. Commun.* 5, 3027. doi:10.1038/ncomms5027

Rahman, S. U., McCoy, E., Raza, G., Ali, Z., Mansoor, S., Amin, I., et al. (2022). Improvement of soybean; A way forward transition from genetic engineering to new plant breeding technologies. *Mol. Biotechnol.*, 1–19. doi:10.1007/s12033-022-00456-6

Rajkumar, M. S., Shankar, R., Garg, R., and Jain, M. (2020). Bisulphite sequencing reveals dynamic DNA methylation under desiccation and salinity stresses in rice cultivars. *Genomics* 112, 3537–3548. doi:10.1016/j.ygeno.2020.04.005

Ran, F. A. C. L., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., et al. (2015). *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature* 520 (7546), 186–191. doi:10.1038/nature14299

Razzaq, A., Saleem, F., Kanwal, M., Mustafa, G., Yousaf, S., Imran Arshad, H. M., et al. (2019). Modern trends in plant genome editing: an inclusive review of the CRISPR/Cas9 toolbox. *Int. J. Mol. Sci.* 20, 4045. doi:10.3390/ijms20164045

Rees, H. A., and Liu, D. R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* 19 (12), 770–788. doi:10.1038/s41576-018-0059-1

Ren, B., Yan, F., Kuang, Y., Li, N., Zhang, D., Lin, H., et al. (2018). A CRISPR/Cas9 toolkit for efficient targeted base editing to induce genetic variations in rice. *Sci. China. Life Sci.* 60 (5), 516–519. doi:10.1007/s11427-016-0406-x

Rensing, S. A., Goffinet, B., Meyberg, R., Wu, S. Z., and Bezanilla, M. (2020). The moss Physcomitrium (Physcomitrella) patens: A model organism for non-seed plants. *Plant Cell* 32 (5), 1361–1376.

Sanchez-Leon, S., Gil-Humanes, J., Ozuna, C. V., Gimenez, M. J., Sousa, C., Voytas, D. F., et al. (2018). Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. *Plant Biotechnol. J.* 16, 902–910.

Sakata, R. C., Ishiguro, S., Mori, H., Tanaka, M., Tatsuno, K., Ueda, H., et al. (2020). Base editors for simultaneous introduction of C-to-T and A-to-G mutations. *Nat. Biotechnol.* 38 (7), 865–869. doi:10.1038/s41587-020-0509-0

Salomon, S., and Puchta, H. (1998). Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO J.* 17 (20), 6086–6095. doi:10.1093/emboj/17.20.6086

Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., et al. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10098–10103. doi:10.1073/pnas.1104144108

Shen, C., Que, Z., Xia, Y., Tang, N., Li, D., He, R., et al. (2017). Knock out of the annexin gene OsAnn3 via CRISPR/Cas9-mediated genome editing decreased cold tolerance in rice. *J. Plant Biol.* 60, 539–547. doi:10.1007/s12374-016-0400-1

Shi, J., Gao, H., Wang, H., Lafitte, H. R., Archibald, R. L., Yang, M., et al. (2017). ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnol. J.* 15, 207–216. doi:10.1111/pbi.12603

Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., et al. (2015). Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol. Cell* 60 (3), 385–397. doi:10.1016/j.molcel.2015.10.008

Shopan, J., Lv, X., Hu, Z., Zhang, M., and Yang, J. (2020). Eukaryotic translation initiation factors shape RNA viruses resistance in plants. *Hortic. Plant J.* 6 (2), 81–88. doi:10.1016/j.hpj.2020.03.001

Shukla, V. K., Doyon, Y., Miller, J. C., DeKelver, R. C., Moehle, E. A., Worden, S. E., et al. (2009). Precise genome modification in the crop species Zea mays using zinc-finger nucleases. *Nature* 459, 437–441. doi:10.1038/nature07992

Song, Q., Zhang, T., Stelly, D. M., and Chen, Z. J. (2017). Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol.* 18, 99. doi:10.1186/s13059-017-1229-8

Song, Y., Ji, D., Li, S., Wang, P., Li, Q., Xiang, F., et al. (2012). The dynamic changes of DNA methylation and histone modifications of salt responsive transcription factor genes in soybean. *PLoS One* 7, e41274. doi:10.1371/journal.pone.0041274

Sretenovic, S., Liu, S., Li, G., Cheng, Y., Fan, T., Xu, Y., et al. (2021). Exploring C-To-G base editing in rice, tomato, and poplar. *Front. Genome Ed.* 24, 756766. doi:10.3389/fgeed.2021.756766

Sretenovic, S., and Qi, Y. (2022). Plant prime editing goes prime. *Nat. Plants* 8 (1), 20–22. doi:10.1038/s41477-021-01047-0

Steinert, J., Schiml, S., Fauser, F., and Puchta, H. (2015). Highly efficient heritable plant genome engineering using Cas9 orthologues from Streptococcus thermophilus and *Staphylococcus aureus*. *Plant J.* 84 (6), 1295–1305. doi:10.1111/tpj.13078

Streubel, J., Blucher, C., Landgraf, A., and Boch, J. (2012). TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.* 30, 593–595. doi:10.1038/nbt.2304

Sun, Z., Li, N., Huang, G., Xu, J., Pan, Y., and Wang, Z. (2013). Site-specific gene targeting using transcription activator-like effector (TALE)-based nuclease in *Brassica oleracea*. *J. Integr. Plant Biol.* 55, 1092–1103. doi:10.1111/jipb.12091

Svitashev, S., Schwartz, C., Lenderts, B., Young, J. K., and Cigan, A. M. (2016). Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. *Nat. Commun.* 7, 13274. doi:10.1038/ncomms13274

Talakayala, A., Ankanagari, S., and Garladinne, M. (2022). CRISPR-cas genome editing system: a versatile tool for developing disease resistant crops. *Plant Stress* 3, 100056. doi:10.1016/j.stress.2022.100056

Tang, X., Sretenovic, S., Ren, Q., Jia, X., Li, M., Fan, T., et al. (2020). Plant prime editors enable precise gene editing in rice cells. *Mol. Plant* 13, 667–670. doi:10.1016/j.molp.2020.03.010

Tang, X., Zheng, X., Qi, Y., Zhang, D., Cheng, Y., Tang, A., et al. (2016). A single transcript CRISPR-Cas9 system for efficient genome editing in plants. *Mol. Plant* 9 (7), 1088–1091. doi:10.1016/j.molp.2016.05.001

Tian, S., Jiang, L., Cui, X., Zhang, J., Guo, S., Li, M., et al. (2018). Engineering herbicide- resistant watermelon variety through CRISPR/Cas9-mediated base-editing. *Plant Cell Rep.* 37 (9), 1353–1356. doi:10.1007/s00299-018-2299-0

Tra, M. V. T., Yin, X., Bajal, I., Balahadia, C. P., Quick, W. P., and Bandyopadhyay, A. (2021). "Single base editing using cytidine deaminase to change grain size and seed coat color in rice," in *Rice genome engineering and gene editing*. New York, NY: Humana, pp. 135–143.

Ueta, R., Abe, C., Watanabe, T., Sugano, S. S., Ishihara, R., Ezura, H., et al. (2017). Rapid breeding of parthenocarpic tomato plants using CRISPR/Cas9. *Sci. Rep.* 7, 507. doi:10.1038/s41598-017-00501-4

Underwood, C. J., Choi, K., Lambing, C., Zhao, X., Serra, H., Borges, F., et al. (2018). Epigenetic activation of meiotic recombination near *Arabidopsis thaliana* centromeres via loss of H3K9me2 and non-CG DNA methylation. *Genome Res.* 28, 519–531. doi:10.1101/gr.227116.117

Van Der Oost, J., Westra, E. R., Jackson, R. N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat. Rev. Microbiol.* 12 (7), 479–492. doi:10.1038/nrmicro3279

Veillet, F., Chauvin, L., Kermarrec, M. P., Sevestre, F., Merrer, M., Terret, Z., et al. (2019). The Solanum tuberosum GBSSI gene: a target for assessing gene and base editing in tetraploid potato. *Plant Cell Rep.* 38 (9), 1065–1080. doi:10.1007/s00299-019-02426-w

Veillet, F., Kermarrec, M. P., Chauvin, L., Chauvin, J. E., and Nogué, F. (2020). CRISPR- induced indels and base editing using the *Staphylococcus aureus* Cas9 in potato. *PLoS One* 15 (8), e0235942. doi:10.1371/journal.pone.0235942

Wada, N., Osakabe, K., and Osakabe, Y. (2022). Expanding the plant genome editing toolbox with recently developed CRISPR-Cas systems. *Plant Physiol.* 188, 1825–1837. doi:10.1093/plphys/kiac027

Wang, D., Samsulrizal, N. H., Yan, C., Allcock, N. S., Craigon, J., Blanco-Ulate, B., et al. (2019). Characterization of CRISPR mutants targeting genes modulating pectin degradation in ripening tomato. *Plant Physiol.* 179 (2), 544–557. doi:10.1104/pp.18.01187

Wang, F., Wang, C., Liu, P., Lei, C., Hao, W., Gao, Y., et al. (2016). Enhanced rice blast resistance by CRISPR/Cas9-Targeted mutagenesis of the ERF transcription factor gene *OsERF922*. *PloS One* 11, e0154027. doi:10.1371/journal.pone.0154027

Wang, L., Chen, L., Li, R., Zhao, R., Yang, M., Sheng, J., et al. (2017). Reduced drought tolerance by CRISPR/Cas9-mediated SlMAPK3 mutagenesis in tomato plants. *J. Agric. Food Chem.* 65 (39), 8674–8682. doi:10.1021/acs.jafc.7b02745

Wang, M., Xu, Z., Gosavi, G., Ren, B., Cao, Y., Kuang, Y., et al. (2020). Targeted base editing in rice with CRISPR/ScCas9 system. *Plant Biotechnol. J.* 18 (8), 1645–1647. doi:10.1111/pbi.13330

Wang, W., Ma, S., Hu, P., Ji, Y., and Sun, F. (2021). Genome editing of rice eIF4G loci confers partial resistance to rice black-streaked dwarf virus. *Viruses* 13 (10), 2100. doi:10.3390/v13102100

Wang, W., Pan, Q., He, F., Akhunova, A., Chao, S., Trick, H., et al. (2018). Transgenerational CRISPR-Cas9 activity facilitates multiplex gene editing in allopolyploid wheat. *CRISPR J.* 1 (1), 65–74. doi:10.1089/crispr.2017.0010

Wang, X., Tu, M., Wang, D., Liu, J., Li, Y., Li, Z., et al. (2018). CRISPR/Cas9-mediated efficient targeted mutagenesis in grape in the first generation. *Plant Biotechnol. J.* 16 (4), 844–855. doi:10.1111/pbi.12832

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* 32, 947–951. doi:10.1038/nbt.2969

Wang, Z., Liu, X., Xie, X., Deng, L., Zheng, H., Pan, H., et al. (2021). ABE8e with polycistronic tRNA-gRNA expression cassette significantly improves adenine base editing efficiency in Nicotiana benthamiana. *Int. J. Mol. Sci.* 22 (11), 5663. doi:10.3390/ijms22115663

Wu, J., Chen, C., Xian, G., Liu, D., Lin, L., Yin, S., et al. (2020). Engineering herbicide- resistant oilseed rape by CRISPR/Cas9-mediated cytosine base-editing. *Plant Biotechnol. J.* 1857–1859. doi:10.1111/pbi.13368

Xie, J., Huang, X., Wang, X., Gou, S., Liang, Y., Chen, F., et al. (2020). ACBE, a new base editor for simultaneous C-to-T and A-to-G substitutions in mammalian systems. *BMC Biol.* 18, 131. doi:10.1186/s12915-020-00866-5

Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci. U. S. A.* 112 (11), 3570–3575. doi:10.1073/pnas.1420294112

Xu, R., Kong, F., Qin, R., Li, J., Liu, X., Wei, P., et al. (2021). Development of an efficient plant dual cytosine and adenine editor. *J. Integr. Plant Biol.* 63, 1600–1605. doi:10.1111/jipb.13146

Xu, R., Li, J., Liu, X., Shan, T., Qin, R., Wei, P., et al. (2020). Development of plant prime- editing systems for precise genome editing. *Plant Commun.* 1 (3), 100043. doi:10.1016/j.xplc.2020.100043

Xu, R., Yang, Y., Qin, R., Li, H., Qiu, C., Li, L., et al. (2016). Rapid improvement of grain weight via highly efficient CRISPR/Cas9-mediated multiplex genome editing in rice. *J. Genet. Genomics.* 43, 529–532. doi:10.1016/j.jgg.2016.07.003

Xu, Y., Lin, Q., Li, X., Wang, F., Chen, Z., Wang, J., et al. (2021). Fine-tuning the amylose content of rice by precise base editing of the Wx gene. *Plant Biotechnol. J.* 19 (1), 11–13. doi:10.1111/pbi.13433

Xu, W., Yang, Y., Yang, B., Krueger, C. J., Xiao, Q., Zhao, S., et al. (2022). A design optimized prime editor with expanded scope and capability in plants. *Nat. Plants* 8 (1), 45–52.

Yin, X., Biswal, A. K., Dionora, J., Perdigon, K. M., Balahadia, C. P., Mazumdar, S., et al. (2017). CRISPR-Cas9 and CRISPR-Cpf1 mediated targeting of a stomatal developmental gene EPFL9 in rice. *Plant Cell Rep.* 36 (5), 745–757. doi:10.1007/s00299-017-2118-z

Yuste-Lisbona, F. J., Fernández-Lozano, A., Pineda, B., Bretones, S., Ortíz-Atienza, A., García-Sogo, B., et al. (2020). ENO regulates tomato fruit size through the floral meristem development network. *Proc. Natl. Acad. Sci. U S A.* 117 (14), 8187–8195. doi:10.1073/pnas.1913688117

Zeng, Y., Wen, J., Zhao, W., Wang, Q., and Huang, W. (2020). Rational improvement of rice yield and cold tolerance by editing the three genes OsPIN5b, GS3, and OsMYB30 with the CRISPR–Cas9 system. *Front. Plant Sci.* 10, 1663. doi:10.3389/fpls.2019.01663

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163 (3), 759–771. doi:10.1016/j.cell.2015.09.038

Zhang, A., Liu, Y., Wang, F., Li, T., Chen, Z., Kong, D., et al. (2019). Enhanced rice salinity tolerance via CRISPR/Cas9-targeted mutagenesis of the OsRR22 gene. *Mol. Breed.* 39 (3), 47. doi:10.1007/s11032-019-0954-y

Zhang, C., Wang, F., Zhao, S., Kang, G., Song, J., Li, L., et al. (2020). Highly efficient CRISPR-SaKKH tools for plant multiplex cytosine base editing. *Crop J.* 8 (3), 418–423. doi:10.1016/j.cj.2020.03.002

Zhang, J., Zhang, X., Chen, R., Yang, L., Fan, K., Liu, Y., et al. (2020). Generation of transgene-free Semidwarf maize plants by gene editing of gibberellin-oxidase20-3 using CRISPR/Cas9. *Front. Plant Sci.* 104, 1048. doi:10.3389/fpls.2020.01048

Zhang, R., Liu, J., Chai, Z., Chen, S., Bai, Y., Zong, Y., et al. (2019). Generation of herbicide tolerance traits and a new selectable marker in wheat using base editing. *Nat. Plants* 5 (5), 480–485. doi:10.1038/s41477-019-0405-0

Zhang, X., Zhu, B., Chen, L., Xie, L., Yu, W., Wang, Y., et al. (2020). Dual base editor catalyzes both cytosine and adenine base conversions in human cells. *Nat. Biotechnol.* 38 (7), 856–860. doi:10.1038/s41587-020-0527-y

Zhang, Y., Bai, Y., Wu, G., Zou, S., Chen, Y., Gao, C., et al. (2017). Simultaneous modification of three homoeologs of *TaEDR1* by genome editing enhances powdery mildew resistance in wheat. *Plant J.* 91, 714–724. doi:10.1111/tpj.13599

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., et al. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat. Commun.* 7, 12617. doi:10.1038/ncomms12617

Zhao, C., Zhang, Z., Xie, S., Si, T., Li, Y., Zhu, J. K., et al. (2016). Mutational evidence for the critical role of CBF genes in cold acclimation in *Arabidopsis*. *Plant Physiol.* 171, 2744–2759. doi:10.1104/pp.16.00533

Zhao, D., Li, J., Li, S., Xin, X., Hu, M., Price, M. A., et al. (2021). Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat. Biotechnol.* 39 (1), 35–40. doi:10.1038/s41587-020-0592-2

Zhong, X., Du, J., Hale, C. J., Gallego-Bartolome, J., Feng, S., Vashisht, A. A., et al. (2014). Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell* 157, 1050–1060. doi:10.1016/j.cell.2014.03.056

Zhong, Z., Sretenovic, S., Ren, Q., Yang, L., Bao, Y., Qi, C., et al. (2019). Improving plant genome editing with high-fidelity xCas9 and non-canonical PAM-targeting Cas9-NG. *Mol. Plant* 12 (7), 1027–1036.

Zhou, J., Peng, Z., Long, J., Sosso, D., Liu, B., Eom, J. S., et al. (2015). Gene targeting by the TAL effector PthXo2 reveals cryptic resistance gene for bacterial blight of rice. *Plant J.* 82, 632–643. doi:10.1111/tpj.12838

Zhou, J., Xin, X., He, Y., Chen, H., Li, Q., Tang, X., et al. (2019). Multiplex QTL editing of grain-related genes improves yield in elite rice varieties. *Plant Cell Rep.* 38 (4), 475–485. doi:10.1007/s00299-018-2340-3

Zhuang, Y., Liu, J., Wu, H., Zhu, Q., Yan, Y., Meng, H., et al. (2022). Increasing the efficiency and precision of prime editing with guide RNA pairs. *Nat. Chem. Biol.* 18 (1), 29–37.

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., et al. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* 35 (5), 438–440. doi:10.1038/nbt.3811

Zsogon, A., Cermak, T., Naves, E. R., Notini, M. M., Edel, K. H., Weinl, S., et al. (2018). De novo domestication of wild tomato using genome editing. *Nat. Biotechnol.* 36, 1211–1216.

Frontiers in Genetics

TYPE Review
PUBLISHED 13 September 2022
DOI 10.3389/fgene.2022.959266

Check for updates

OPEN ACCESS

EDITED BY
Deepmala Sehgal,
International Maize and Wheat Improvement Center, Mexico

REVIEWED BY
Saurabh Badoni,
International Rice Research Institute (IRRI), Philippines
Mahmoud Yaish,
Sultan Qaboos University, Oman
Ulises Rosas,
Institute of Biology, National Autonomous University of Mexico, Mexico

*CORRESPONDENCE
Hifzur Rahman,
r.hifzurrahman@biosaline.org.ae,
rahman.biotech@gmail.com

SPECIALTY SECTION
This article was submitted to Plant Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 01 June 2022
ACCEPTED 11 August 2022
PUBLISHED 13 September 2022

CITATION
Rahman H, Vikram P, Hammami Z and Singh RK (2022), Recent advances in date palm genomics: A comprehensive review.
Front. Genet. 13:959266.
doi: 10.3389/fgene.2022.959266

COPYRIGHT
© 2022 Rahman, Vikram, Hammami and Singh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Recent advances in date palm genomics: A comprehensive review

Hifzur Rahman*, Prashant Vikram, Zied Hammami and Rakesh Kumar Singh

International Center for Biosaline Agriculture (ICBA), Dubai, United Arab Emirates

As one of the oldest fruit trees of the Arabian peninsula, other Middle-Eastern countries, and also North Africa, the date palm (*Phoenix dactylifera* L.), is highly significant for the economy of the region. Listed as part of UNESCO's Intangible Cultural Heritage of Humanity, the date palm is believed to be the first tree cultivated by human beings, and was probably first harvested for its fruit nearly 7,000 years ago. Initial research efforts in date palm genetics focused on understanding the genetic diversity of date palm germplasm collections and its phylogenetic history, both important prerequisites for plant improvement. Despite various efforts, the center of origin of the date palm is still unclear, although genomic studies suggest two probable domestication events: one in the Middle East and the other in North Africa, with two separate gene pools. The current review covers studies related to omics analyses that have sought to decipher the present genetic diversity of the date palm. With advances and cost reductions in sequencing technologies, rapid progress has been made in the past few years in date palm genomics research. Along with organellar genomes, several reference genomes of the date palm are now available. In addition, several genotypes have been re-sequenced, either to detect single nucleotide polymorphisms (SNPs), or to study domestication and identification of key genes/loci associated with important agronomic traits, such as sex, fruit color, and sugar composition. These genomics research progress has paved the way to perform fast-track and precise germplasm improvement processes in date palm. In this study, we review the advances made in the genetics and genomics of the date palm so as to strategize targeted crop improvement plans for marginal areas of the Middle Eastern peninsula, North Africa, and other parts of the world.

KEYWORDS

date palm, genomics, diversity, molecular markers, transcriptomics

Introduction

Date palm (*Phoenix dactylifera* L.) is one of the oldest fruit trees in the Arabian Peninsula (AP), other countries of the Middle East, and the arid regions of North Africa. Its origin is not yet known; however, evidence indicates this was possibly near Iraq. In the AP, the date palm holds cultural importance for the people, besides being a critically

TABLE 1 Cultivar diversity and production indices across major date palm producing countries.

Country	Cultivar evaluated[a]	Area harvested (ha)[b]	Yield (kg/ha)[b]	Production (tons)[b]
Algeria	1,000	170,500	6756.1	1,151,909
Egypt	52	50,834	33264.3	1,690,959
Iraq	400	245,033	3001.0	735,353
Iran	400	154,145	8326.6	1,283,499
Libya	95	32,868	5404.3	177,629
Morocco	453	61,332	2334.2	143,160
Oman	250	25,630	14380.7	368,577
Pakistan	NA	106,488	5101.7	543,269
Saudi Arabia	450	152,705	10096.4	1,541,769
Sudan	400	37,000	12576.3	465,323
Tunisia	250	72,205	4598.0	332,000
Yemen	321	15,038	4627.6	69,590
United Arab Emirates	120	38,422	8554.2	328,669

[a]Abul-Soad et al. (2017).
[b]Data from FAOSTAT (2020).

important staple food and a major source of income. It is known for multiple products and purposes, including fruit, fiber, fuel, and sheltering material. In addition to providing calories, dates serve as a source of vitamins and minerals (El Hadrami and Al-Khayri, 2012), rendering them a healthy and nutritious calorie option. The global trade in dates was valued at about USD $1.2 billion in 2016, thereby contributing to the livelihood and income of millions of rural smallholders in the AP and surrounding areas [FAO: Microsoft Word - Conference-side-event-Dates-Saudi-Arabia.docx (fao.org)]. Globally, date palms are cultivated on 1.1 million hectares of land with a production of about 8.7 million tons (FAOSTAT, 2018). Iran, Algeria, Iraq, Saudi Arabia, and Egypt share 59% of the total harvested area and 66.5% of the total production, with maximum yield coming from Egypt (Table 1) (FAOSTAT, 2020). The area under date palm cultivation has also increased continuously during recent decades because of the crop's adaptability to the harsh climate.

Despite being one of the most successful fruit crops in arid, semiarid, tropical, and subtropical regions, there has been relatively little research into the genetics and molecular genetics of the date palm compared to other commercial fruit trees. Genomics approaches are an exception here, and rapid advances have been made in the past decade. So far, the date palm genome, along with its organellar genomes, has been sequenced (Al-Dous et al., 2011; Fang et al., 2012; Khan et al., 2012; Al-Mssallem et al., 2013; Asaf et al., 2018). Several genotypes have been re-sequenced, either to detect single nucleotide polymorphisms (SNPs) (Thareja et al., 2018), or to study domestication and marker-trait association (Hazzouri et al., 2015, 2019; Gros-Balthazard et al., 2017). In addition to SNPs, other marker resources have been used in the past for diversity studies and the identification of cultivars, including random amplified polymorphic DNA (RAPD), inter simple sequence repeats (ISSR), simple sequence repeats (SSR), and amplified fragment length polymorphisms (AFLP), etc. Early sex deteremnation is an important trait in the date palm. Because the date palm is dioecious in nature, efforts have been made to develop specific markers for early detection of the female plant. Molecular markers have also been developed for brittle leaf disease (BLD) resistance in the tree. In the recent past, red palm weevil (*Rhynchophorus ferrugineus* Olivier) has had a devastating effect on date palm yields in the Arabian Peninsula (Kassem et al., 2020). At the International Center for Biosaline Agriculture (ICBA), UAE, efforts are underway to address this biotic stress with the help of advanced genomics tools (ICBA, unpublished).

Although there has been little progress in genomics applications in date palm improvement, this study represents an effort to review what progress has been made, as well as identify the future prospects for date palm genomics, given the importance of this crop for the livelihood of vast numbers of farmers in the AP.

Botanical description of date palm

The date palm (*Phoenix dactylifera* L.) is a perennial monocotyledonous plant belonging to the family Arecaceae (Palmae). Mature date palm plants are the tallest among *Phoenix* spp. and can attain heights as tall as 25–30 m, with a single main terminal shoot apex for linear growth. The date palm has a well-developed fibrous root system in which primary roots develop directly from the seeds/tree trunk with an average length of 4–6 m. The lateral roots originate from primary roots, which

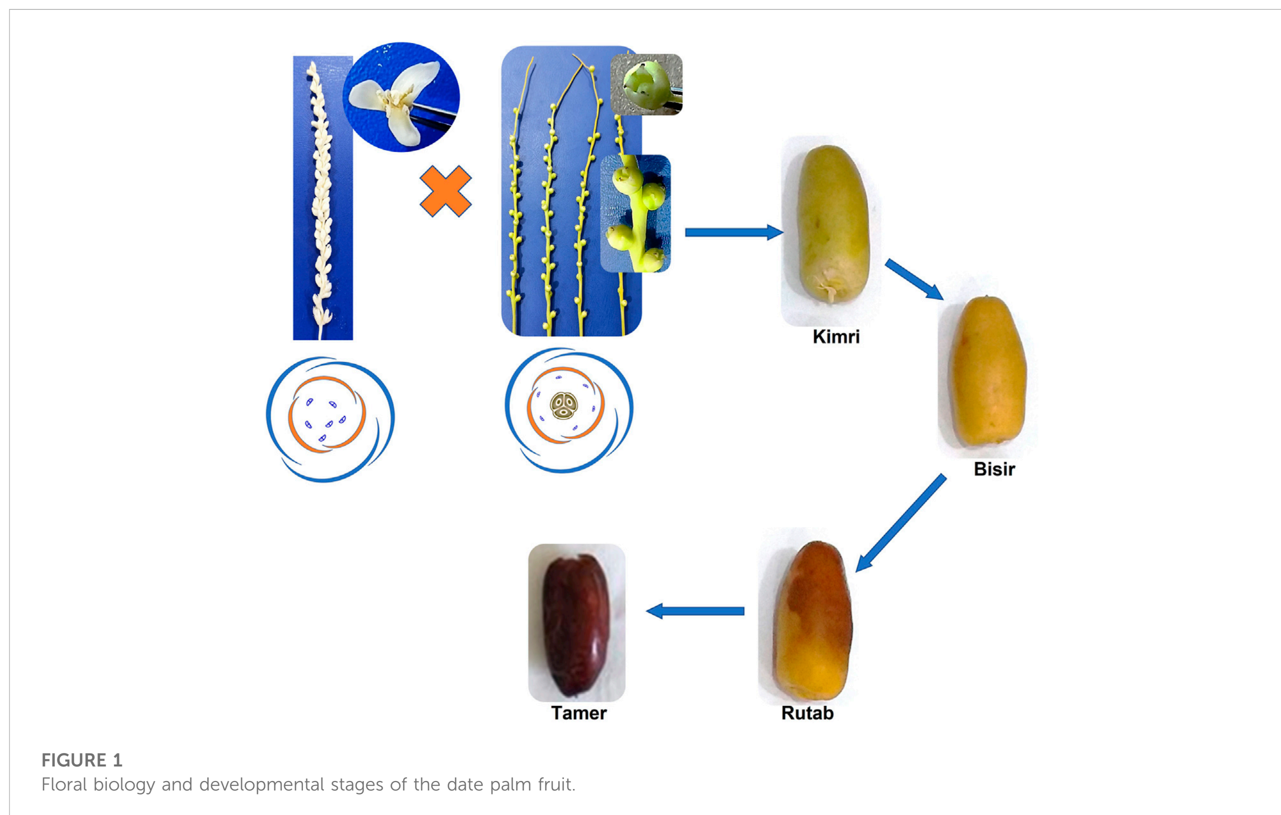


FIGURE 1
Floral biology and developmental stages of the date palm fruit.

further produce lateral roots throughout their length. All date palm roots contain pneumatics, which act as respiratory organs. The plant contains pinnate leaves arranged alternately along the trunk. An adult date palm plant contains 100–125 leaves, with 40% juvenile, 10% fast-growing, and 50% photosynthetically active (Zaid and De Wet, 1999). The date palm is dioecious in nature, with separate trees producing male and female flowers in clusters called spadixes or spikes, within axils of leaves of the growth of previous years. Rarely, both pistillate and staminate flowers are produced on the same spike, and hermaphrodite flowers have also been reported in the inflorescence (Mason, 1915; Milne, 1918), although in very few cases. The staminate flowers are sweet-scented and generally have six stamens, each composed of two little pollen sacs. The stamens are surrounded by three waxy sepals and petals. The female flowers contain rudimentary stamens and are tricarpellate, consisting of three carpels that are closely pressed together and surrounded by a short perianth with a superior ovary (Figure 1). Pollination occurs by wind, or artificially, by dusting pollen grains collected from male spikelets onto female inflorescences. The fruit normally develops after fertilization from one carpel, which develops faster, while the other two carpels degenerate and drop later. The development of seeded fruit follows a sigmoidal curve with four distinct ripening stages: kimri, khalal (also know as bisir/bisr), rutab, and tamer stages (the names being of Arabic origin), which represent immature green, mature full-colored, soft brown, and hard raisin-like stages, respectively, containing average moisture content of 80%, 60%, 40%, and 20%, respectively (Fayadh and Al-Showiman, 1990; Al-Shahib and Marshall, 2003). The date fruit varies in size and shape depending on the cultivar and environment. With advancement in the developmental stages of the fruit, antioxidant activity increases until bisir and then decreases (Awad et al., 2011; Mohamed Lemine et al., 2014), whereas sugar content increases with ripening in the date palm fruit (Al-Mssallem et al., 2013) (Figure 1).

Distribution and biodiversity

The date palm is one of the earliest cultivated fruit trees, with records of its cultivation in the areas of the Euphrates and Nile rivers going back to 3700 BC; while in Iran, Egypt, and Pakistan, the earliest records go back 7,000 years (Munier, 1973). The exact center of origin of the date palm is not certain; however, it is believed to have originated from the modern Iraqi region of Mesopotamia (Wrigley, 1995). Interestingly, the oldest seeds of palm, dating back to 5110 BC and 4670 BC, were reported from an island of Abu Dhabi known as Dalma (Rhouma et al., 2010). One of two wild species, *Phoenix reclinata* Jacq. from tropical Africa, or *P. sylvestris* (L.) Roxb. from India, or a hybrid of these two, is believed to be the progenitor of the date palm. However, many researchers believe that the tree originates from the Mesopotamian-Arabian Gulf area (Zohary and Hopf, 2000; Tengberg, 2012) and was later

introduced into North Africa. However, genomic studies suggest that the genotypes from North Africa and the Middle East are genetically distinct, with higher genetic diversity in the North African date palm population (Hazzouri et al., 2015). Along with archaeological records, the population structure suggests two probable domestication sources in the date palm, one from the Middle East and the other from North Africa, forming two separate gene pools that diverged before domestication (Hazzouri et al., 2015; Zehdi-Azouzi et al., 2015). Also, the presence of admixed genotypes suggests that gene flow occurred between populations of eastern and western origins, primarily from east to west, as a result of human-mediated dispersal of the species after domestication (Zehdi-Azouzi et al., 2015). Whole-genome sequencing of wild and cultivated date palms reveals a complex domestication history with the contribution of at least two wild sources to the African cultivated palms (Gros-Balthazard et al., 2017). The date palm is thought to have spread globally in two directions: one from Mesopotamia to Iran, India, and Pakistan, and the other from Egypt toward Libya and the countries of the Maghreb and Sahel (Racchi and Camussi, 2018). The date palm is now more abundant in the arid regions of the Old World than in the temperate regions of the New World, with the most date palm trees found in Middle Eastern countries (Iraq, Iran, Saudi Arabia, United Arab Emirates, Oman, Yemen, etc.), followed by Africa (Algeria, Egypt, Libya, Mali, Morocco, Mauritania, Niger, Somalia, Sudan, Chad, Tunisia, etc.) (FAOSTAT, 2019). Apart from the Middle East and Africa, with the expansion of Islam the date palm has also been introduced into the United States and Europe (Chao and Krueger, 2007; Rivera et al., 2013).

The genus *Phoenix* consists of 12 closely-related species, making them cross-compatible for natural hybridization (Moore, 1963; Munier, 1973). Several natural hybrids were obtained from different countries: *P. dactylifera* × *P. sylvestris* (India), *P. dactylifera* × *P. canariensis* (Morocco, Algeria, and Israel), and *P. dactylifera* × *P. reclinata* (Senegal).

The date palm is generally diploid in nature, with 2n = 36 chromosomes (Darlington and Wylie, 1956), although polyploidy has been reported in Iraqi varieties (n = 64) (Panga, 2014). Also, differences in chromosome numbers have been observed between varieties Sayer (2n = 32) and Khasab (2n = 36). Al Salih et al. (1987) reported 2n = 32, 34, 36, and 64 date palm chromosomes. It would be interesting to investigate the crossability among accessions with different ploidy levels for creating newer genetic variations.

Molecular marker-assisted genetic diversity in the date palm

Genetic diversity refers to the genetic variability present within species, subspecies, cultivars, or populations, and can be measured at the morphological, physiological, biochemical, or molecular levels. A total of 3,000–5,000 date palm cultivars exist globally. The cultivar diversity and production indices across major date palm growing countries are presented in Table 1 (Abul-Soad et al., 2015; Al-Yahyai and Al-Khanjari, 2008; Al-Yahyai and Khan, 2015; Ba-Angood, 2015; Bashah, 1996; Battaglia et al., 2015; Bouguedoura et al., 2015; Elshibli, 2009; Hajian and Hamidi-Esfahani, 2015; Mahar, 2007; Osman, 1984; Rabei et al., 2012; Sedra, 2015; Zabar and Borowy, 2012; Zaid and De Wet, 1999; FAOSTAT, 2020). Despite collections of traditionally cultivated genotypes, duplications have also been reported among them. Therefore, approximately only 10% of the total cultivars existing globally are believed to be unique and commercially important (Johnson, 2011). The commercialization of preferred date palm cultivars prompted an increase in area of cultivation, thereby enhancing the practice of monoculture, which ultimately led to a significant decline in genetic (or species) diversity in the date palm. Characterizing, conserving, and using date palm collections globally is one of the felt needs that can be effectively met using high-density genomics approaches. In recent years, initiatives aimed at characterizing large collections of gene banks have succeeded. As an example, 100,000 wheat and 30,000 maize accessions of the gene bank of the International Maize and Wheat Improvement Center (CIMMYT) were characterized (Seeds of Discovery: Unlocking the genetic potential of maize and wheat). This sort of approach could provide a suitable option for characterizing date palm genetic resources globally.

Research into developing biochemical and molecular markers for date palms began in the late 1970s, and was later used for genetic diversity analysis. Various isozyme markers were used to study the inheritance of multiple traits in date palm seedlings (Torres and Tisserat, 1980) and genetic diversity (Bennaceur et al., 1991), or to develop a cultivar identification system (Bendiab et al., 1998) for date palms. Advances in molecular biology, and the development of a PCR-based marker system during the 1990s led to studies using various molecular marker systems, either individually or in combination, to unravel the genetic diversity and phylogenetics of the date palm. Initially, DNA-based marker systems, such as RFLP and RAPD, were used to identify polymorphic markers associated with date palm cultivars (Corniquel and Mercier, 1994). Even though RAPD markers have low reproducibility, they have been used to study the genetic diversity in various date palm accessions in different countries (Table 2). Other types of dominant multi-locus markers, such as ISSR markers on their own or in combination with RAPD markers, have also been used in genetic diversity analysis of the date palm (Table 2). Genetic diversity analyses using RAPD or ISSR markers, or a combination of both, have shown that huge genetic similarity (more than 90%) exists among various date palm genotypes. A comparison between four female date palm trees and four unknown male trees of the Egyptian date palm, using RAPD markers, shows that genetic similarity existed not only between female cultivars (87.5%–98.9%), but also between unknown male trees (88.9%–95.3%) (Soliman et al., 2003). A more reproducible multi-locus marker, AFLP, has been used either alone, or in

TABLE 2 Application of biochemical and molecular markers in genetic diversity studies of the date palm.

Markers type	Markers used	Genotypes studied	Geographical location of studied genotypes	Study type	Citation
Isozyme	5	26 female and 20 male date palm cultivars and breeding populations	California	Inheritance in date palm seedlings	Torres and Tisserat (1980)
Isozyme	7	186 plants belonging to 31 cultivars	Algeria	Genetic diversity analysis	Bennaceur et al. (1991)
Isozyme	3	28 genotypes	Morocco	Varietal identification	Bendiab et al. (1998)
RFLP and RAPD	-	5	-	Cultivar identification	Corniquel and Mercier (1994)
RAPD	19	43	Morocco, Iraq, Tunisia	Genetic diversity	Sedra et al. (1998)
ISSR	12	18	Tunisia	Genetic diversity	Zehdi et al. (2002)
ISSR	15	8	Ethiopia	Genetic diversity	Takele et al. (2021)
RAPD		3 female and 4 male trees	Egypt	Genetic diversity	Soliman et al. (2003)
RAPD	37	13	Saudi Arabia	Genetic diversity	Al-Khalifah and Askari (2003)
RAPD	12	5	Saudi Arabia	Genetic diversity and cultivar identification	Al-Moshileh et al. (2004)
RAMPO	18	30 female and 10 male trees	Tunisia	Genetic diversity	Rhouma et al. (2008)
RAPD	3	10	Bahrain	Genetic diversity	Pathak and Hamzah (2008)
RAPD and ISSR	5 each	4	Saudi Arabia	Genetic diversity	Abdulla and Gamal (2010)
ISSR and AFLP	13 and 6	10	Saudi Arabia	Genetic diversity and cultivar identification	Sabir et al. (2014a)
SSR	22	24 female and 6 male trees	Iraq	Genetic diversity	Khierallah et al. (2011b)
AFLP	6	11 female and 7 male trees	Iraq	Genetic diversity	Khierallah et al. (2011a)
RAPD and ISSR	35 and 15	18 female and 5 male trees	Syria	Genetic diversity	Haider et al. (2012)
RAPD and ISSR	30 and 12	10 female and 7 male trees	Iraq	Genetic diversity	Khierallah et al. (2014)
RAMPO and AFLP	18 and 6	40	Tunisia	Genetic diversity	Soumaya et al. (2011)
ISSR and DAMD	5 and 8	9	UAE	Genetic diversity	Purayil et al. (2018)
ISSR	10	14	Iran	Genetic variability and population structure	Sharifi et al. (2018)
cpDNA sequences	-	47	Iran	Genetic diversity	Sharifi et al. (2018)
RAPD	3	10	Bahrain		Pathak and Hamzah (2008)
RAPD and ISSR	35 and 15	18	Syria	Genetic diversity	Haider et al. (2012)
RAPD	19	43	Morocco, Iraq, Tunisia	Genetic diversity and cultivar identification	Sedra et al. (1998)
RAPD	5	10	Nigeria	Genetic diversity	Emoghene et al. (2015)
RAPD and ISSR	27 and 21	20	Algeria	Genetic diversity	Guettouchi et al. (2017)
IR fluorescence- labeled AFLP markers	4	21	USDA germplasm collection	Genetic diversity	Cao and Chao (2002)
IR fluorescence-labeled AFLP markers	-	2	California (United States)	Genetic purity testing of Deglet Noor and Medjool	Devanand and Chao (2003a, 2003b)
Fluorescence-labeled AFLP	4	Various accessions of Medjool and Deglet Noor	Morocco, Egypt, and California (United States)	Genetic similarity/diversity within accession	Elhoumaizi et al. (2006)
AFLP	4	47 accessions of Medjool and Deglet Noor	Egypt	Genetic similarity/diversity within accession	El-Assar et al. (2005)
AFLP	5	5 individuals of 3 genotypes		Varietal identity among offshoots	Diaz et al. (2003)
AFLP	4	18	Iraq	Genetic relationship and varietal identification	Jubrael et al. (2005)

(Continued on following page)

TABLE 2 (*Continued*) Application of biochemical and molecular markers in genetic diversity studies of the date palm.

Markers type	Markers used	Genotypes studied	Geographical location of studied genotypes	Study type	Citation
AFLP	-	10	UAE	Genetic fidelity of tissue culture–raised plants	Al Kaabi et al. (2007)
AFLP	6	40	Tunisia	Genetic diversity	Rhouma et al. (2007)
SCoT (start codon targeted)	4	113 trees of 13 varieties	Iran	Genetic diversity	Saboori et al. (2020)
SSR	14	49	Tunisia	Genetic diversity	Zehdi et al. (2004)
SSR	16	37 female and 23 male trees	Sudan and Morocco	Genetic diversity	Elshibli and Korpelainen (2008)
SSR	10	200 individuals from 19 populations	Sudan	Genetic diversity	Elshibli and Korpelainen (2009)
SSR	5	26	Tunisia	Genetic diversity	Hammadi et al. (2011)
SSR	14	74 female and 27 male trees	Tunisia	Genetic diversity	Zehdi et al. (2012)
SSR	17	31	Algeria	Genetic diversity	Akkak et al. (2009)
SSR	37	18	-	Varietal identification	Johnson et al. (2009)
SSR	10	21	Oman, Bahrain, Iraq, and Morocco	Genetic diversity in tissue culture–raised plants	Al-Ruqaishi et al. (2008)
SSR	10	15	Qatar	Genetic diversity	Ahmed and Al-Qaradawi (2010)
SSR		11	Qatar	Genetic diversity	Elmeer et al. (2011)
SSR	14	59 plants from 12 cultivars	Qatar	Inter- and intra-varietal genetic diversity	Elmeer and Mattat (2015)
SSR	22	16	Iraq, Iran, and Africa	Genetic diversity	Arabnezhad et al. (2012)
SSR	16	377 female trees of 18 cultivars and 63 male trees	Libya	Molecular typing and diversity analysis	Racchi et al. (2014)
SSR	15	200 trees consisting of 191 females belonging to 26 cultivars, and 9 male trees	Morocco	Genetic diversity	Bodian et al. (2012)
SSR	16	377 trees belonging to 18 cultivars	Libya	Genetic diversity	Racchi et al. (2014)
SSR	10	89 female plants from 18 cultivars	Sudan	Genetic diversity	Elsafy et al. (2016)
SSR	46	45	Pakistan	Genetic diversity	Faqir et al. (2016)
SSR	22	32	Saudi Arabia	Genetic diversity	Al-Faifi et al. (2016)
Fluorescence-labeled SSRs	17	82	Australia	Genetic diversity	Al-Najm et al. (2017)
SSR	255	1,066 date palms from 411 cultivars	12 different countries	Genetic diversity	Salomon-Torres et al. (2017)
SSR	18	113 date palms involving 31 males and 82 females	Nigeria	Genetic diversity	Zango et al. (2017)
SSR and chloroplast mini-satellite	18 and 1	414 trees belonging to 114 varieties	Algeria	Genetic diversity	Moussouni et al. (2017)
SSR	12	38 wild date palm genotypes	Bangladesh	Genetic diversity	Huda et al. (2019)
SSR	8	24 male pollinizers	Tunisia	Genetic diversity	El Kadri et al. (2019)
Mitochondrial and plastid genome-based SNPs	-	9 varieties		Molecular characterization	Sabir et al. (2014b)
GBS-based SNPs	-	70 female genotypes and four other species			Mathew et al. (2015)
SNPs	-	62 cultivars		Genetic diversity and gene-trait relationship	Hazzouri et al. (2015)
GBS-based SNPs	-	191 cultivars		Genetic diversity	Thareja et al. (2018)
MatK, rbcl, atpB, and SNPs	-			SNP typing and varietal identification	Al-Dous et al. (2011)

(Continued on following page)

TABLE 2 (*Continued*) Application of biochemical and molecular markers in genetic diversity studies of the date palm.

Markers type	Markers used	Genotypes studied	Geographical location of studied genotypes	Study type	Citation
ITSs	-	15	Tunisia	Haplotype identification and diversity analysis	Maina et al. (2019)

combination with other marker systems, to assess genetic variations present in date palm cultivars across the globe. Initially, an attempt was made to develop a genetic map using AFLP markers with a population derived from Um-Assla and KL-96 (El-Kharbotly et al., 1998). Later, either AFLP or fluorescently-labeled AFLP primers were used to study genetic diversity among various accessions of the date palm, with genetic similarity ranging from approximately 10%–75% (Cao and Chao, 2002; Devanand and Chao, 2003a, 2003b; El-Khishin et al., 2003; El-Assar et al., 2005; Jubrael et al., 2005; Elhoumaizi et al., 2006; Rhouma et al., 2007; Khierallah H. et al., 2011, Khierallah et al., 2011 H. S.). The AFLP markers were also used to assess intra-varietal differences (El-Assar et al., 2005; Elhoumaizi et al., 2006), and to study the genetic fidelity of plants raised *in vitro* (Diaz et al., 2003; Al Kaabi et al., 2007). Apart from being used to study diversity, these multi-locus markers have also been used for cultivar identification (Table 2) (Corniquel and Mercier 1994; Al-Moshileh et al., 2004; Sabir J. S. et al., 2014).

Because of their abundance and dispersion throughout the genome, their co-dominance nature, ease of usage, and ability to automate, microsatellites or simple sequence repeats (SSR) have proven an ideal choice for cultivar identification and genetic diversity analysis, as well as for linkage and QTL mapping, and marker-assisted breeding. Billotte et al. (2004) made the first attempt to develop SSR markers for date palms using a $(GA)_n$-enriched library. They further designed 16 SSR primers, and tested their amplification on 40 DNA samples of *P. dactylifera* from various origins, as well as on 11 other species of *Phoenix*. Later, several attempts were made to identify and develop SSR markers for date palms using a genomic DNA library enriched for microsatellite sequences (Akkak et al., 2009; Arabnezhad et al., 2012; Al-Faifi et al., 2016). Apart from using microsatellite-enriched libraries, available sequence information, such as ESTs, have been used to develop EST-SSRs as well as gene-based SSRs, and they have been characterized with their functional annotations (Zhao et al., 2012, 2017). With the availability of various draft assemblies of date palm genomes, "genome sequence information has been used to develop 1090 SSRmarkers (Hamwieh et al., 2010). Using the draft genome sequence of the date palm, Mokhtar et al. (2016) identified 172,075 SSR motifs, with a frequency of 450.97 SSRs per Mb. A total of 130,014 SSRs (75.6%) were located within the intergenic regions, while 42,061 SSRs (24.4%) were located in the genic regions. Furthermore, 111,403 SSR primer pairs were designed, with a density of 291.9 SSR primers per Mb.

Numerous genetic diversity analyses have been conducted with the help of microsatellite markers in different countries: Tunisia, Qatar, Libya, Morocco, Sudan, Pakistan, Saudi Arabia, Niger, Algeria, Sudan, the United States (California), Australia, etc. (Zehdi et al., 2004, 2012; Al-Ruqaishi et al., 2008; Elshibli and Korpelainen, 2008, 2009; Akkak et al., 2009; Ahmed and Al-Qaradawi, 2010; Hamwieh et al., 2010; Hammadi et al., 2011; Bodian et al., 2014; Racchi et al., 2014; Elmeer and Mattat, 2015; Al-Faifi et al., 2016; Elsafy et al., 2016; Faqir et al., 2016; Al-Najm et al., 2017; Moussouni et al., 2017; Zango et al., 2017; El Kadri et al., 2019) (Table 2).

Salomon-Torres et al. (2017) reviewed the performance of 255 SSR markers for studying diversity among 1,066 date palm plants from 411 cultivars in 12 countries, and recommended a set of 19 SSR markers as useful for further genetic diversity analysis. Recently, studies have looked into the genetic diversity of worldwide date palm germplasm accessions, using SSRs (Chaluvadi et al., 2014; Zehdi-Azouzi et al., 2015; Salomon-Torres et al., 2017), SNPs (Hazzouri et al., 2015; Mathew et al., 2015), or comparisons of whole genomes (Hazzouri et al., 2015). Through sequencing of 62 varieties of date palms from 12 countries, Hazzouri et al. (2015) show that Middle Eastern genotypes form a separate group from North African genotypes, with North African genotypes having higher nucleotide diversity than Middle Eastern/South Asian genotypes. Similar results were obtained by Mathew et al. (2015), where the authors used the sequence data from 70 date palm accessions. Such accessions are mostly propagated through tissue culture; however, variations among accessions with the same name suggest that somaclonal mutation is ongoing during the process of subculturing during tissue culture. Under some circumstances, huge genetic variations within the same accession suggest that, since the date palm is dioecious in nature, there is a probability of random crossing events in addition to the controlled outcrossing, and the plants might have been raised from the seeds, resulting in an increase in the genetic distance between the genotypes.

Apart from their use in genetic diversity analysis, SSR markers have been used in developing a molecular identification key, as well as in molecular-typing for identification of the characterized cultivars (Zehdi et al., 2012; Racchi et al., 2014). These SSR

markers, developed and characterized across various date palm accessions, can further be used for identifying candidate genes and understanding the genetic basis of traits of interest, which may further help in molecular breeding for the genetic improvement of the date palm tree.

Single nucleotide polymorphisms

As the third generation of molecular markers, single nucleotide polymorphisms (SNPs) are more stable, and have higher conformity of inheritance than other marker systems (Gupta et al., 2001). With the whole-genome sequencing of the date palm, the first attempt to identify SNPs was carried out by Al-Dous et al. (2011), where researchers called 1,748,109 SNPs in 381 Mb of sequence, yielding a heterozygosity rate of 0.46%, or 1 SNP/217 bp, but the distribution of the SNPs was skewed, with 49% of the SNPs within every 50 bp. Sabir J. S. M. et al. (2014) used the mitochondrial and plastid genome sequences of nine date palm varieties to examine SNPs, but found a low level of variation, suggesting the preferred use of nuclear SNPs for molecular characterization of date palm cultivars. A genotyping-by-sequencing (GBS) approach was used to identify 13,000–65000 SNPs comparing the genomes of 70 female cultivars from different date palm growing regions and four other *Phoenix* species (Mathew et al., 2015). Based on the whole-genome re-sequencing of 62 cultivars, a catalog of approximately 7 million SNPs in date palms was developed (Hazzouri et al., 2015). Recently, the GBS approach was followed by using re-sequenced data of 191 date palm cultivars to identify SNPs and assess the genetic diversity among the date palm trees grown in Qatar. This study revealed that these trees in Qatar are of eastern origin and their genetic diversity does not associate with different regions (Thareja et al., 2018). Faqir et al. (2019) sequenced maturase K (matK), ribulose biphosphate carboxylase larger subunit (*rbcL*), the ATP synthase subunit b (atpB) gene of the chloroplast genome, and 12 DNA fragments from the nuclear genome of seven cultivars. Based on the sequenced data, the researchers identified unique SNP signatures and developed an SNP-typing system for varietal identification of date palm cultivars from Pakistan.

The internal transcribed spacer (ITS) sequences of 15 Tunisian date palm accessions were compared to identify four haplotypes, and the haplotypic and nucleotide diversities were found to be low among the studied genotypes (Maina et al., 2019). Further phylogenetic analysis revealed that the Tunisian populations of date palm evolved under a neutral model, and a demographic equilibrium seems to be maintained within the studied genotypes.

Trait-specific markers in the date palm

Most of the molecular studies of the date palm have been carried out for genetic diversity and phylogenetic analysis, as well as for cultivar identification, with limited progress made in developing trait-specific molecular markers. Most efforts have been aimed at identifying markers associated with Bayoud disease resistance, or for sex determination. *Fusarium oxysporum* f. sp. *albedensis* causes Bayoud disease, which is one of the most devastating of all diseases in date palm trees (Michielse and Rep, 2009; El Modafar, 2010).

Bendiab et al. (1992) carried out isozyme polymorphism analysis using esterase (EST), glutamate oxaloacetate transaminase (GOT), endopeptidase (ENP), and alcohol dehydrogenase (ADH) polymorphisms in different F_1 populations derived from seven female cultivars crossed with two males (Table 3). They found out three loci viz., Got2, Est 1, and Enp that could be used for hybrid screening. Benslimane et al. (1994, 1996) isolated two mitochondrial-like plasmid DNA (S and R plasmids) sharing 99% sequence similarity, except for 109 bp of sequence that was present in only the S plasmid. The S plasmid was found in Bayoud-susceptible genotypes, whereas the R plasmid was found in Bayoud-resistant Moroccan genotypes. Later, employing a PCR-based approach on 36 date palm varieties, Quenzar et al. (2001) confirmed the study of Benslimane et al. (1994, 1996), and reported that the simultaneous presence of the R plasmid and absence of the S plasmid can be considered a reliable marker for Bayoud resistance (Table 3). Salem et al. (2007) used this plasmid-based analysis system to check the susceptibility of Mauritanian date palm cultivars to Bayoud disease. Furthermore, using progenies of two controlled crosses, the authors showed that Bayoud strictly follows maternal transmission as controlled by the mitochondrial genome. The R and S mitochondrial plasmids have been used for molecular characterization of date palm cultivars from Algeria (Guettouchi et al., 2017), Syria (Haider and Nabulsi, 2012), and Saudi Arabia (Saleh et al., 2015).

Brittle leaf disease, known as *maladie des feuilles cassantes* in French, was first observed in southern Tunisia (Djerbi, 1983). It later spread to reach epidemic levels by 1986. The exact causal pathogen is not yet determined; however, the symptom of the disease is associated with manganese deficiency and the presence of a small double-stranded chloroplast RNA (Triki et al., 2003; Namsi et al., 2006, 2007; Marqués et al., 2008). Namsi et al. (2006) used chloroplast RNA, and developed a digoxigenin (DIG)-labeled probe for early diagnosis of BLD, which consistently gave positive hybridization signals, irrespective of cultivars, the severity of symptoms, or the geographic location (Table 3).

The date palm is a dioecious plant, and the sex of the plants can be determined only at the time of flowering, which takes 5–7 years (Shaheen, 1990). If the sex of the plants could be determined at the early seedling stage, this could save resources and time, as farmers need many female plants and only a few superior male plants for pollination. Hence, maintaining a proper male:female ratio is of the utmost importance for better production in the field. Sex determination at the early seedling stage is thus one of the major

TABLE 3 Trait-specific markers in date palms.

Trait	Marker	References
Bayoud disease	Biochemical: esterase (EST-1), glutamate oxaloacetate transaminase (GOT-2), endopeptidase (ENP)	Bendiab et al. (1992)
	R and S mitochondrial plasmid	Benslimane et al. (1994, 1996)
Brittle leaf disease	Double-stranded chloroplast RNA	Namsi et al. (2006, 2007); Triki et al. (2003)
Gender-specific	Biochemical: peroxidase and glutamate oxaloacetate activity higher in females	Bekheet et al. (2008); Qacif et al. (2007)
	RAPD: OPA10-490, OPA12-750, and OPD10-800 specific to females and OPA12-370 and OPD10-675 specific to males	Younis et al. (2008)
	ISSR: HB10-1010, HB9-340, HB12-375, 814-590, and 844A-920 specific for males	
	RAPD-derived SCAR marker	Dhawan et al. (2013)
	ISSR: IS_A02 (390) specific to female plants and IS_A71 (380bp) specific to male plants only	Al-Ameri et al. (2016b)
	SCoT-derived SCAR marker of size 253 bp specific to male trees	Al-Ameri et al. (2016a)
	RAPD-derived SCAR marker	Al-Qurainy et al. (2018)
	SRY gene-specific marker for identification of male plants	Mohei et al. (2019)
	SSRs: mPdIRDP80, mPdIRDP50, mPdIRDP52, mpdCIR48, and DP-168	Cherif et al. (2013); Elmeer and Mattat (2012); Maryam Jaskani et al. (2016)
	SNPs	Al-Dous et al. (2011)

requisites for establishing commercial date palm orchards. Therefore, the identification of markers linked to the sex of plants is of key importance for date palm cultivation. For the first time, Siljak-Yakovlev et al. (1996) developed a cytological method in which staining with chromomycin shows the presence of an extra heterochromatin region on both the arms of the male chromosome, which was considered sex determinant. Atia et al. (2017) describe cytological-based markers to distinguish date palm sex through localization of 45S and 5S rDNA markers on date palm chromosomes using the fluorescence *in situ* hybridization (FISH) technique.

A few biochemical markers, such as peroxidase and glutamate oxaloacetate, reportedly differentiate between male and female date palms, with a differential response of peroxidase and glutamate oxaloacetate activity observed in female plants versus male plants (Qacif et al., 2007; Bekheet et al., 2008). Over the past 2 decades, several attempts have been made to understand the genetic basis of sex determination in date palms using various types of DNA markers, such as RFLP, RAPD, ISSR, and SSRs. Using RAPD primers, several polymorphic markers have been identified with the potential to distinguish male from female plants among different cultivars (Ben et al., 2000; Soliman et al., 2003; Bekheet et al., 2008).

Younis et al. (2008) used a combination of RAPD and ISSR techniques to identify three fragments derived from RAPD markers specific to females (OPA10-490, OPA12-750, OPD10-800), and two for males in RAPD analysis (OPA12-370, OPD10-675), as well as five specific markers for males through ISSR analysis (HB10-1010, HB9-340, HB12-375, 814-590, 844A-920) (Table 3). However, in the past decade, attempts have been made to develop SCAR (sequence-characterized amplified region) markers for sex determination in date palms. The genomic DNA of 10 male genotypes of unknown origin and 10 female genotypes were pooled in equal quantities separately, and 100 RAPD primers and 104 ISSR primers were used to identify sex-specific markers. One of the RAPD primers, OPA-02, amplified an ≈1.0-kb fragment specifically in pooled as well as individual samples of male genotypes, and was later converted into a SCAR marker, which amplified a fragment of 406 bp in both female and male genotypes, and a unique fragment of 354 bp only in male genotypes (Dhawan et al., 2013). The developed SCAR marker was further validated in 25 female and 10 male date palms belonging to different varieties collected from different locations. Later, using an ISSR marker, Al-Ameri et al. (2016b) identified a 390-bp fragment from the amplicons of primer IS_A02, specifically in a female plant, and a 380-bp fragment from the amplicons of primer IS_A71, specifically in male plants only. These fragments were sequenced further to develop sequence-specific markers. Al-Ameri et al. (2016a) developed a SCAR marker of size 253 bp, specific to male trees based on cDNA fingerprinting of start codon targeted (SCoT) marker, and validated it independently on male and female trees. Al-Qurainy et al. (2018) developed a SCAR marker linked to sex-specific regions in the genome of the date palm using RAPD marker OPC-06, which was producing a band of 186 bp in male plants only. Recently, a gene, *SRY1*, involved in initiating sex determination, was identified on the Y chromosome of the date palm, and was tested with 100% efficiency for identifying male plants at the seedling stage (Mohei et al., 2019). Apart from RAPD and ISSR markers, a few microsatellite markers (e.g., mPdIRDP80, mPdIRDP50, mPdIRDP52, mpdCIR48, and DP-168) possessing the capacity

TABLE 4 Transcriptomic studies in date palm.

Study type	Focused trait studies	References
Transcriptome and metabolome	Carbon partitioning, sugars, and fatty acid metabolism	Bourgis et al. (2011)
cDNA sequencing	Cell division genes, ripening-related genes, and sugar/starch metabolism	Yin et al. (2012)
Transcriptome	30,854 annotated gene model and Gene Ontology and KEGG pathways assignment. Gene networks controlling organ development	Zhang et al. (2012)
Transcriptome	Differential expression of genes involved in energy metabolism in different tissues	Fang et al. (2012)
Transcriptome	Genes involved in fruit development and ripening	Al-Mssallem et al. (2013)
Transcriptome	Genes and small RNAs expressed in embryogenic calli	Naganeeswaran et al. (2020)
Transcriptome	Genes/pathways involved in imparting salinity tolerance	Radwan et al. (2015)
Transcriptome	Salinity-responsive small RNA libraries from roots and leaves	Yaish et al. (2015)
Transcriptome	Salinity-responsive genes in roots and leaves	Yaish et al. (2017)
Transcriptome and metabolome	Genes and metabolites in response to mild heat, drought, and combination of both stresses	Safronov et al. (2017)
Transcriptome	Genes involved in detoxifying cadmium toxicity	Rekik et al. (2019)
Suppression-subtractive hybridization	Genes involved in BLD tolerance	Saidi et al. (2010)
RT-PCR	Differential expression of genes in response to BLD in roots and leaves	Saidi et al. (2012)
Transcriptome	Differentially expressed genes in response to RPW infestation	Giovino et al. (2015)

for sex differentiation in the date palm have also been identified (Elmeer and Mattat, 2012; Cherif et al., 2013; Maryam Jaskani et al., 2016). Al-Dous et al. (2011) identified a region harboring 1,605 SNPs linked to sex through *de novo* genome sequencing, and proposed that the date palm follows an XY system of gender inheritance (Table 3). A 6-Mb region has been further mapped onto the distal end of chromosome 12, which has been found to be associated with sex determination (Hazzouri et al., 2019). Recently, Torres et al. (2021) identified 16-bp male-specific sequences in the date palm Y chromosome.

Date palms are facing a severe threat around the globe from red palm weevil (*Rhynchophorus ferrugineus* Olivier). So far, no molecular marker has been reported that deciphers resistance to this dreaded date palm pest. Using a historic long-term ongoing field trial with 18 date palm varieties, researchers at the International Center for Biosaline Agriculture have identified the pattern of preference/sensitivity and non-preference/ tolerance (anti-xenosis behavior) for red palm weevil of specific date palm varieties. They are further trying to understand if there could be a robust molecular/genetic basis of RPW resistance in the date palm, and further to identify the molecular markers linked to this RPW resistance. The developed markers will not only help in selecting resistant genotypes, but will also help in developing genotypes with RPW resistance through accelerated molecular breeding.

Date palm genomics

Genomics deals with the sequencing and analysis of the structure of the genome of an organism, predicting the genes, and their locations and functions in the genome. Initially, the date palm genome was considered to be relatively smaller than 250 Mb, with 41% of the region consisting of genes, and the remaining genome considered a non-coding region (Barakat et al., 1999). However, it was later found that the size of the date palm genome ranged from 550 to 650 Mbp (Malek, 2010). Initially, a random genomic library of Tunisian date palm varieties was constructed from total cellular DNA, and amplified using RAPD markers. The library consisted of inserts from 200 to 1,600 bp and was supposed to have potential application for generating probes for molecular characterization of date palm varieties through southern hybridization. Al-Faifi et al. (2017) generated 6,943 high-quality ESTs from a normalized cDNA library of the date palm cultivar, Sukkari. The generated ESTs were assembled into 6,362 unigenes and were further functionally annotated. The first genetic map of the date palm cv. Khalas was developed by Mathew et al. (2014), using ~4,000 SNPs spanning a total of 1,293 cM. Furthermore, the analysis suggested that the telomeric region on linkage group 12 may be the sex-determination region of the date palm. A total of 19% of the draft genome sequence scaffolds were placed onto the linkage groups, and the analysis results showed that approximately 1.9 cM represents 1 Mb on the map (Mathew et al., 2014). The chronological developments in genome sequencing of date palm is given in Figure 2.

Organellar genomes

With the advances in next-generation sequencing (NGS) technologies during the past decade, progress in the genomics

of the date palm has been made at an unprecedented pace. The complete chloroplast genome of the date palm cultivar Khalas was sequenced using pyrosequencing and was found to be of 158,462 bp in size, consisting of 112 unique genes and 19 duplicated fragments in the inverted repeat (IR) regions, and arranged in a typical quadripartite structure (Yang et al., 2010). Furthermore, 78 SNPs located in genes with vital functions were identified with potential for detecting intra-varietal polymorphisms within a date palm population. Using a combination of Sanger-based and next-generation sequencing strategies, Khan et al. (2012) sequenced the complete date palm chloroplast genome from the Pakistani cultivar Aseel. The size of the genome was found to be 158,458 bp, consisting of a large single-copy (LSC) region of 86,195 bp, and a small single-copy (SSC) region of 17,711 bp, separated by an IR region of 27,276 bp. The chloroplast genome consisted of 138 genes, of which 89 were protein-coding, 39 were tRNA, and 8 were rRNA genes. Furthermore, a comparison of the Khalas and Aseel chloroplast genome led to the identification of SNPs and mono-nucleotide SSRs. Recently, Khan et al. (2018) sequenced the chloroplast genome of two economically important date palm cultivars, Khanezi and Naghal, using the Illumina HiSeq4000 sequencing platform. The chloroplast genome sizes of Naghal and Khanezi were 158,210 bp and 158,211 bp, respectively, consisting of 138 genes. The phylogenetic analysis based on the whole chloroplast genome and 68 shared genes of four cultivars (Khanezi, Naghal, Khalas, and Aseel) yielded identical phylogenetic trees, with Khanezi and Naghal forming single clades with cultivars Khalas and Aseel, respectively.

Fang et al. (2012) published the first mitochondrial genome of the date palm cv. Khalas. The genome assembly consisted of 715,001 bp encoding 38 proteins, 30 tRNAs, and 3 ribosomal RNAs. The protein-coding sequence consists of only 6.5% (46,770 bp) of the mitochondrial genome, whereas the rest of the genome sequence (93.5%) was found to comprise chloroplast-derived (10.3%) and non-coding sequences. Recently, the mitochondrial genome of *P. dactylifera* var. Khanezi, consisting of 715,120 bp, was published (Asaf et al., 2018). The mitochondrial genome consisted of 67 genes encoding 24 transfer RNAs, 3 ribosomal RNAs, and 40 protein-coding genes. Apart from these two mitochondrial genomes, another unpublished assembly is available in GenBank from an unknown cultivar (MG257490.1), consisting of 585,493 bp (Figure 2).

Nuclear reference genomes

Whole-genome sequencing is fundamental for understanding the molecular basis of complex traits for crop improvement. With the rapid progress in NGS technology and the simultaneous availability of bioinformatics tools, the past decade has seen unprecedented developments in date palm genomics, leading to the development of two draft genome sequences and genetic maps. The first attempt to develop the draft genome sequence of the date palm cv. Khalas was made by Al-Dous et al. (2011). The genome was sequenced from tissue culture–raised plants using the Illumina platform. Unfortunately, it covered only ~60% of the genome and consisted of 380 Mb of sequence, spanning mainly gene-rich regions, including 25,059 gene models. However, this reference genome was found to be highly fragmented, with about 60,000 scaffolds showing a median length of ~30 kb. These authors further identified a genomic region linked to the sex of the plant, and provided evidence that the date palm follows an XY system of gender inheritance. Subsequently, using pyrosequencing, Al-Mssallem et al. (2013) reported another genome assembly of higher quality from the same date palm cultivar Khalas. This genome assembly has a total length of 605.4 Mb, covering more than 90% of the genome and 96% of the genes. They further built a larger pool of gene models, consisting of 41,660 models with a total of 42,957 isoforms in 10,363 scaffolds. The sequenced genome analysis demonstrated genome-wide duplication after either ancient whole-genome duplications or massive segmental duplications. Genetic diversity analysis showed that the stress resistance and sugar metabolism-related genes are enriched in the chromosomal regions where the density of SNPs is relatively low. Scrutiny of the late embryogenesis abundant (LEA) gene family revealed that group 2 LEA genes are specifically abundant in date palms, with 62 group 2 LEA members showing generally ubiquitous expression, whereas LEA1, LEA3, LEA4, LEA5, LEA6, seed maturation protein, and dehydrin were found to be either seed or male flower associated. This date palm draft genome assembly has also been included in the reference sequence (RefSeq) collection in the National Center for Biotechnology Information (NCBI), and gene models have been included in UniProtKB and the Kyoto Encyclopedia of Genes and Genomes (KEGG) databases for further exploration. In 2019, Hazzouri et al. (2019) released a new date palm draft genome (“BC4 male”). This draft genome spanned 772 Mb and was assembled into 2,390 scaffolds (Figure 2).

Functional genomics of the date palm

The first attempt to gain insight into carbon partitioning, comparative transcriptome, and metabolome analysis in oil palm and date palm mesocarp led to the identification of several sugars and fatty acid metabolism genes/transporters involved in fatty acid and sugar accumulation in date and oil palm, respectively (Bourgis et al., 2011). Yin et al. (2012) carried out cDNA sequencing of the date palm fruits of Khalas at seven different developmental stages and identified 10 core cell division genes, 18 ripening-related genes, and 7 starch metabolic enzymes, which are involved in nutrition storage and sugar/starch metabolism. To generate and annotate the gene model of the date palm, Zhang et al. (2012) carried out in-depth

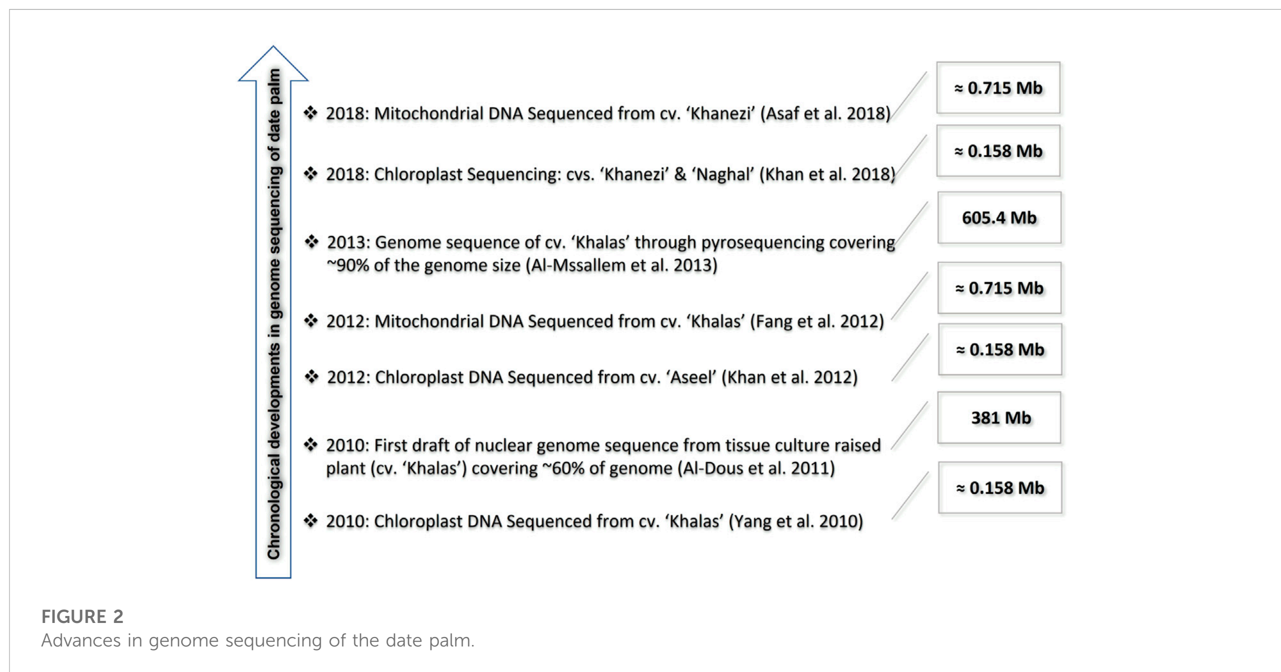


FIGURE 2
Advances in genome sequencing of the date palm.

transcriptomic sequencing from different tissues and at several developmental stages, and generated 30,854 annotated gene models from the cultivar Khalas. These were further assigned to Gene Ontology and KEGG pathways for future research aimed to unravel the genetic regulatory networks governing organ development and differentiation in the date palm (*P. dactylifera*). Whole-genome transcriptome analysis of eight tissues (root, seed, bud, fruit, green leaf, yellow leaf, female flower, and male flower), using the Roche/454 GS FLX platform, showed higher gene expression levels in developing tissues, such as male and female flower, root, and bud, than in the four other tissues, due to the need for more energy than in the relatively mature tissues (Fang et al., 2012). To identify the differentially expressed genes (DEGs) involved in fruit development and ripening, Al-Mssallem et al. (2013) carried out transcriptome analysis at seven distinct fruit developmental stages (0, 15, 30, 60, 90, 120, and 135 days post-pollination), and identified 4,134 DEGs whose expression varies significantly among the seven fruit developmental stages. The enrichment analysis of DEGs revealed that most of the molecular events involved in biological regulation, transcription, and regulation of RNA metabolic processes are down-regulated in the late stage of fruit development, whereas events involved in sugar accumulation, such as gluconeogenesis, cellular carbohydrate metabolism, and small molecule biosynthesis were up-regulated, resulting in unusually high sugar content in the dates. Hazzouri et al. (2019) carried out RNA-Seq analysis in date palm fruit at different developmental stages. The results indicated that the expression of alkaline/neutral invertase (*A/N-INV1*) was maximum at ≈105 days after pollination, whereas the expression of *cell wall invertase* (*CWINV1* and *CWINV3*) genes peaked at 120 days after pollination, showing their positive role in sugar accumulation during fruit development. Recently, Naganeeswaran et al. (2020) performed transcriptome assembly from the embryogenic calli of the date palm cultivar Khalas, and reported 63,888 Gene Ontology (GO) terms and 122 small RNAs that were annotated from the assembly (Table 4).

Date palms generally grow under adverse climatic conditions and have therefore developed stress tolerance during their evolution. The date palm can survive under extreme drought, heat, and relatively high soil salinity (Yaish and Kumar, 2015), thereby providing a valuable genome source for mining abiotic stress tolerance genes. However, limited research work has been carried out to identify and exploit the abiotic stress-responsive genes from the date palm. To understand the molecular mechanisms underlying salinity tolerance in the date palm, Radwan et al. (2015) undertook salinity-responsive transcriptome analysis in young roots of the date palm cv. Deglet Beida, which led to the identification of 1939 differentially expressed genes involved in tolerance of salt stress. RNA-Seq analysis further revealed that salinity stress activates abscisic acid signaling pathways through SNF1-related protein kinase 2, and several key genes involved in sodium uptake and transport were found to be down-regulated, thereby slowing down uptake and transportation in plant tissues under stress conditions. Yaish et al. (2015) generated salinity-responsive small RNA libraries from leaves and roots of date palm seedlings. Deep sequencing using Illumina Hiseq2000 led to the identification of 153 homologs of conserved miRNAs, 89 miRNA variants, and 180 putative novel miRNAs from the date palm plant. Differential expression analysis revealed that

TABLE 5 Whole-genome resequencing studies in date palms and their wild relatives.

References	Number of germplasm/accessions resequenced	Key findings
Al-Dous et al. (2011)	5 female and 3 male and one F_1 progeny	• First draft genome assembly of the date palm (cv. Khalas) • Identified 3,518,029 SNPs • Identified XY sex-determination model and region controlling sex on XY chromosomes
Al-Mssallem et al. (2013)	3 female and 1 male	• Improved genome assembly of the date palm (cv. Khalas) • Functional analysis of genes involved in abiotic stress tolerance and genes involved in sugar metabolism during fruit ripening
Hazzouri et al. (2015)	61 female and 1 male	• Genetic diversity analysis among the cultivars from North Africa and the Middle East • Candidate mutations for trait variation in genes involved in the pathways for key agronomic traits • *Virescens* (*VIR*) gene encoding R2R3 myb-like transcription factor was found to be associated with fruit color variation
Gros-Balthazard et al. (2017)	2 date palm cultivars 3 wild date palms 1 *Phoenix sylvestris* 1 *Phoenix atlantica*	• Discovered wild date palm populations in remote Oman • Studied population structure and diversity analysis in the date palm • Revealed complex domestication history of date palms
Torres et al. (2018)	15 female and 13 male trees representing all 14 *Phoenix* species	• Identified male-specific sequences • Further identified CYP703 and GPAT3 genes involved in male flower development in the date palm
Hazzouri et al. (2019)	145 female and 12 male	• Improved genome assembly for *P. dactylifera* • Genome-wide association studies of the sex-determining region and fruit traits • Confirmed previous finding that fruit color is controlled by *VIRESCENS* gene • Identified *invertase* genes controlling sugar composition in date palms

57 miRNAs in leaves and 27 miRNAs in roots were significantly regulated in response to salinity, whereas 12 miRNAs were commonly regulated in both leaves and roots. The targets of the identified miRNAs were the genes with known functions in plant salt tolerance, such as potassium channel AKT2-like proteins, vacuolar protein sorting-associated protein, calcium-dependent protein, and mitogen-activated proteins. Later, expression profiling in the leaves and roots of date palm seedlings revealed 194 differentially expressed transcripts in both leaf and root tissue in response to salinity stress (Yaish et al., 2017). Gene ontology analysis revealed that metabolic pathways, such as photosynthesis, sucrose and starch metabolism, and oxidative phosphorylation were enriched in leaves, whereas genes involved in membrane transport; phenylpropanoid biosynthesis; purine, thiamine, and tryptophan metabolism; and Casparian strip development, were enriched in roots in response to salinity stress. Salinity-responsive genes, such as putative potassium transporter 8, abscisic acid receptor PYR1 and 4, indole-3-acetic acid-amido synthetase GH3, along with a pyrophosphate-energized vacuolar membrane proton pump, were commonly induced in both roots and leaves. Using transcriptomic and metabolomic profiling, Safronov et al. (2017) studied the adaptation mechanism in the date palm toward mild heat, drought, and the combination of both. The results showed an increase in soluble carbohydrates, such as fructose and glucose derivatives, suggesting a switch to carbohydrate metabolism and cell wall biogenesis in response to these stresses. Increased transcriptional activation of genes involved in reactive oxygen species production occurred in response to all three treatments (drought, heat, and combined heat and drought). By contrast, under heat and combined heat and drought stress, genes enriched for circadian and diurnal rhythm motifs were differentially expressed, suggesting a stress avoidance mechanism in response to these stresses (Safronov et al., 2017). Another group of researchers employed salinity-responsive whole-genome bisulfite sequencing and mRNA sequencing in the roots of date palms (Al-Harrasi et al., 2018). The bisulfite sequencing revealed that the methylated regions increased in response to salinity, specifically at mCHG and mCHH sequences. However, when researchers correlated gene expression with DNA methylation, they observed that DNA methylation was not the primary agent that controls gene expression under salinity conditions (Al-Harrasi et al., 2018). Overexpressing the cDNA library of the date palm in *Saccharomyces cerevisiae*, and screening on a synthetic minimal medium containing 1.0 M of NaCl, resulted in the identification of genes such as aquaporins (*PIP*), serine/threonine protein kinases (*STKs*), ethylene-responsive transcription factor 1 (*ERF1*), and peroxidases (*PRX*) with potential salt-tolerance functions (Patankar et al., 2018). Rekik et al. (2019), through transcriptome analysis in leaves of *Phoenix dactylifera* cv. Deglet Nour, proposed a glutathione pathway involved in detoxifying cadmium under Cd

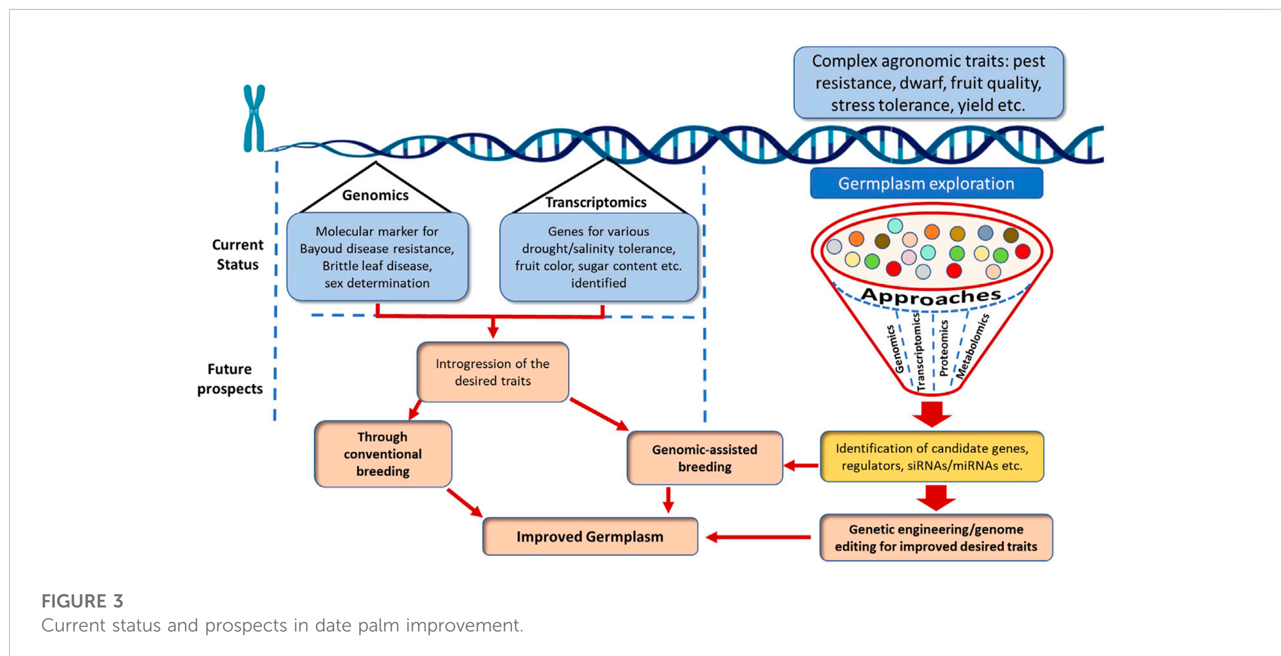


FIGURE 3
Current status and prospects in date palm improvement.

stress conditions, and further identified genes encoding heavy metal transporters and chelators in response to heavy metal stress. Patankar et al. (2019b) isolated aquaporin genes (*PdPIP1;2*) and characterized their role in response to drought and salinity tolerance by overexpressing them in yeast and *Arabidopsis*. The overexpression of an aquaporin gene in yeast resulted in improved oxidative stress tolerance, whereas overexpression in *Arabidopsis* resulted in increased salinity and drought tolerance with increased biomass, chlorophyll content, and root length in transgenic plants (Patankar et al., 2019a). Further, Patankar et al. (2019a) isolated metallothionein 2A (*PdMT2A*) and characterized its role in abiotic stress tolerance in yeast and *Arabidopsis*. The transformed yeast cells have shown tolerance against drought, salinity, and oxidative stresses. The *Arabidopsis* plants overexpressing the metallothionein 2A (*PdMT2A*) gene have shown tolerance against salinity by maintaining a high K^+/Na^+ ratio, and against drought and oxidative stress (Patankar et al., 2019b). Al-Harrasi et al. (2020) isolated a salt-inducible vascular highway 1-interacting kinase (*PdVIK*) and characterized its role in response to various abiotic stresses through heterologous overexpression in yeast and *Arabidopsis*. Jana and Yaish (2020, 2021) isolated and characterized the glyoxalase-I gene (*PdGLX1*) and glyoxalase III genes (*PdDJ-1*) for their roles in mitigation of abiotic stress tolerance through overexpression in bacterial and yeast systems. This study further suggested that *PdGLX1* and *PdDJ-1* genes play an important role in methylglyoxal detoxification and in maintaining reactive oxygen species balance under stress conditions in date palms.

Apart from understanding the transcriptional response of abiotic stress tolerance and fruit development, a couple of studies have been carried out to identify the genes involved in biotic stress tolerance. To understand the molecular mechanisms involved in the BLD of the date palm, Saidi et al. (2010) constructed suppression-subtractive cDNA libraries from BLD-affected and non-affected trees and identified the genes that were up-regulated in response to BLD. The genes associated with stress response, metabolism, protein synthesis, and signal transduction were found to be specifically up-regulated in BLD-affected trees. Later, through RT-PCR analysis, Saidi et al. (2012) showed that the transcripts of MnSOD decreased in affected leaves and roots, unlike the transcripts of FeSOD and Cu/Zn-SOD, whose expression increased in these tissues, revealing that BLD decreases the expression of manganese-related genes in date palm trees. To understand the molecular basis of red palm weevil (*Rhynchophorus ferrugineus* Olivier) resistance in *Phoenix canariensis*, Giovino et al. (2015) carried out deep transcriptome analysis in leaves of healthy and infested trees at two stages (middle and late infestation) and identified 54 genes that were differentially regulated during the middle stage in response to RPW infestation (Table 3). Further enrichment analysis showed that phenylpropanoid-related pathways were induced during the middle infestation period.

Resequencing of the date palm

With the availability of genetic maps, organellar and nuclear reference genomes of the date palm, several research groups carried out whole-genome resequencing of date palms to identify QTLs and SNP markers as well as to study date palm diversity and phylogenetic history. Hazzouri et al. (2015) resequenced 61 female date palm accessions and 1 male (cv. Fard4), and

detected 7,176,238 SNPs at a rate of ~12 SNPs per kb. Genome-wide scans for selection suggested that there were ~36 genomic regions in the genotypes of the Middle East, and 20 genomic regions in North African genotypes associated with positive selection which may underlie the geographic adaptation of these genotypes in these areas. They further characterized candidate mutations in the genes of the pathways associated with key agronomic traits, such as disease resistance, fruit ripening, fruit color, flowering time, and sugar metabolism. Hazzouri et al. (2015) further suggested that the R2R3 myb-like *virescens (VIR)* gene controls fruit color in the date palm. The varieties with red fruit color were found to have an intact *VIR* gene in the homozygous state, whereas the varieties with yellow fruit color had a *copia*-like retrotransposon insertion in the *VIR* gene in either the homozygous or heterozygous state. Using the GBS approach on 70 female cultivars from different date palm growing regions and four other *Phoenix* species, Mathew et al. (2015) showed that there are two centers of earliest cultivation and that the date palm is indigenous to North Africa. Whole-genome sequencing of several wild and cultivated date palms revealed a complex domestication history of date palm trees involving the contribution of a wild relative during the spread of cultivation from their original domestication center in the Arabian Gulf to North Africa (Gros-Balthazard et al., 2017). Sequence analysis of more than 200 mitochondrial and chloroplast genomes from a geographically diverse set of date palms showed that the most common cultivated date palms contain four haplotypes associated with the geographic region of cultivar origin (Mohamoud et al., 2019). Recently, Hazzouri et al. (2019) carried out genome-wide association studies of the sex-determining region, and of 21 fruit traits. GWAS analysis resulted in the identification of the R2R3-MYB transcription factor (*VIR* gene) associated with fruit color. The authors further identified an ≈1.1-Mb region consisting of invertase genes that were found to be associated with sugar composition in date palm fruit (Table 5).

Genomic databases for date palm

During the past decade, several attempts have been made to sequence and re-sequence the several date palm genotypes, leading to the accumulation of a huge amount of genomic data. Further, several SSR and SNP markers have been developed. However, this information is scattered across research publications. This necessitates the development of genomic databases for the date palm so that the developed genomic information can be used more efficiently. The first attempt at this was by Mokhtar et al. (2016), who established a Date Palm Molecular Markers Database (DPMMD) providing useful genomic information (http://dpmmd.easyomics.org/). This database contains information on more than 3,611,400 DNA markers involving SSRs and SNPs, genetic linkage maps, KEGG maps, DNA-barcode, as well as all previously published date palm articles in PubMed-indexed journals from 1976 to 2017. Apart from this, the DRDB (Date Palm Resequence Database) was developed by CAS Key Laboratory of Genome Sciences and Information and Joint Center of Excellence in Genomics, King Abdulaziz City for Science and Technology (He et al., 2017). This database consists of information about 6.3 million SNPs and 246,000 SSRs from 62 date palm cultivars. Apart from these two, there is no concise database for date palm genomics.

Summary and way forward

The date palm has immense regional relevance but requires global attention, as not many advanced research laboratories outside the Middle East and North Africa are giving due attention to date palm genomics. Although limited genomic studies of the date palm over the last decade have led to the identification of a couple of key genes associated with fruit color and sugar accumulation, this is still a long way from what is needed to unravel the hidden mysteries of this tree. Despite the huge existing diversity within the date palm genus, there is little understanding of the genetic factors underlying various biotic and abiotic stresses. The robustness and reliability of a marker are central to its usefulness in a genetic improvement program. Several breeder-friendly molecular markers, such as SSRs and SNPs, have been identified, but the extent to which these markers explain variation still needs to be validated on a large scale. Several abiotic stress responsive genes, and genes associated with fruit traits, have been identified. However, the identified genes/QTLs need to be introgressed in date palm improvement programs, either through breeding or genetic engineering. The use of genetic engineering tools for genome editing is the need of the moment, at least for game-changing traits such as the genetic mechanism of red palm weevil resistance, but this is still lagging because of limited concerted efforts with this crop. Further, studies on the role of small RNAs (siRNA and miRNA) are lacking. It is time to obtain feedback from stakeholders on desired traits in the different genetic backgrounds, and to generate foundational knowledge from diverse research disciplines, including genomics. An extensive germplasm exploration is required for the desired trait combinations ranging from plant architecture and stress tolerance to fruit yield and quality. A concerted effort is therefore needed, employing genomics, transcriptomics, proteomics and metabolomics for identification of candidate genes/genomic regions associated with complex agronomic traits, which can then be further introgressed in popular date palm cultivars/accessions, either through genetic engineering/editing or conventional breeding. An efficient ideotype breeding strategy for the desired date palm variety will be helpful for its improvement (Figure 3). In sum, there should be consortium- or mission-mode-based collaborative efforts to generate and use genomic information in breeding, genetic engineering, or

genome editing research for developing new farmer-friendly date palm varieties.

Author contributions

RS conceived the idea for the manuscript. HR drafted the manuscript, and PV, ZH, and RS revised it. All authors agreed to the final submitted version of the manuscript.

Funding

This work enjoyed funding support from the International Center for Biosaline Agriculture, Dubai.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdulla, M., and Gamal, O. (2010). Investigation on molecular phylogeny of some date palm (*Phoenix dactylifra* L.) cultivars by protein, RAPD and ISSR markers in Saudi Arabia. *Aust. J. Crop Sci.* 4, 23.

Abul-Soad, A. A., Mahdi, S. M., and Markhand, G. S. (2015). "Date palm status and perspective in Pakistan," in *Date palm genetic resources and utilization* (Springer), 153–205.

Ahmed, T. A., and Al-Qaradawi, A. Y. (2010). Genetic diversity of date palm genotypes in Qatar as determined by SSR and ISSR markers. *Acta Hortic.* 882, 279–286. doi:10.17660/actahortic.2010.882.30

Akkak, A., Scariot, V., Marinoni, D. T., Boccacci, P., Beltramo, C., and Botta, R. (2009). Development and evaluation of microsatellite markers in *Phoenix dactylifera* L. and their transferability to other *Phoenix* species. *Biol. plant.* 53, 164–166. doi:10.1007/s10535-009-0026-y

Al Kaabi, H. H., Zaid, A., Shephard, H., and Ainsworth, C. (2007). AFLP variation in tissue culture-derived date palm (*Phoenix dactylifera* L.) plants. *Acta Hortic.* 736, 135–160. doi:10.17660/actahortic.2007.736.12

Al Salih, A. A., Hussain, N., and Al Jarrah, A. (1987). Chromosomes number of a date palm male: Cultivar ghannami akhdar. *Date Palm J.*

Al-Ameri, A. A., Al-Qurainy, F., Gaafar, A.-R. Z., Khan, S., and Nadeem, M. (2016b). Molecular identification of sex in *Phoenix dactylifera* using inter simple sequence repeat markers. *Biomed. Res. Int.* 2016, 4530846. doi:10.1155/2016/4530846

Al-Ameri, A. A., Al-Qurainy, F., Gaafar, A. Z., Khan, S., and Nadeem, M. (2016a). Male specific gene expression in dioecious *Phoenix dactylifera* (Date palm) tree at flowering stage. *Pak. J. Bot.* 48, 131–135.

Al-Dous, E. K., George, B., Al-Mahmoud, M. E., Al-Jaber, M. Y., Wang, H., Salameh, Y. M., et al. (2011). De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nat. Biotechnol.* 29, 521–527. doi:10.1038/nbt.1860

Al-Faifi, S. A., Migdadi, H. M., Algamdi, S. S., Khan, M. A., Al-Obeed, R. S., Ammar, M. H., et al. (2017). "Analysis of expressed sequence tags (EST) in date palm," in *Date palm biotechnology protocols* (Springer), 283–313.

Al-Faifi, S. A., Migdadi, H. M., Algamdi, S. S., Khan, M. A., Ammar, M. H., Al-Obeed, R. S., et al. (2016). Development, characterization and use of genomic SSR markers for assessment of genetic diversity in some Saudi date palm (*Phoenix dactylifera* L.) cultivars. *Electron. J. Biotechnol.* 21, 18–25. doi:10.1016/j.ejbt.2016.01.006

Al-Harrasi, I., Al-Yahyai, R., and Yaish, M. W. (2018). Differential DNA methylation and transcription profiles in date palm roots exposed to salinity. *PloS one* 13, e0191492. doi:10.1371/journal.pone.0191492

Al-Harrasi, I., Patankar, H. V., Al-Yahyai, R., Sunkar, R., Krishnamurthy, P., Kumar, P. P., et al. (2020). Molecular characterization of a date palm vascular highway 1-interacting kinase (*PdVIK*) under abiotic stresses. *Genes* 11, 568. doi:10.3390/genes11050568

Al-Khalifah, N. S., and Askari, E. (2003). Molecular phylogeny of date palm (*Phoenix dactylifera* L.) cultivars from Saudi Arabia by DNA fingerprinting. *Theor. Appl. Genet.* 107, 1266–1270. doi:10.1007/s00122-003-1369-y

Al-Moshileh, A. M., Motawei, M. I., Ai-Wasel, A., and Abdel-Latif, T. (2004). Identification of some date palm (*Phoenix dactylifera* L.) cultivars in Saudi Arabia using RAPD fingerprints. *Jour. Agri. Mar. Scie.* 9, 1–3. doi:10.24200/jams.vol9iss1pp1-3

Al-Mssallem, I. S., Hu, S., Zhang, X., Lin, Q., Liu, W., Tan, J., et al. (2013). Genome sequence of the date palm *Phoenix dactylifera* L. *Nat. Commun.* 4, 2274. doi:10.1038/ncomms3274

Al-Najm, A., Luo, S., Ahmad, N. M., Pourkheirandish, M., and Trethowan, R. (2017). Molecular variability and population structure of a core collection of date palm (*Phoenix dactylifera* L.) cultivars from Australia and the Middle East. *Aust. J. Crop Sci.* 11, 1106–1115. doi:10.21475/ajcs.17.11.09.pne517

Al-Qurainy, F., Al-Ameri, A. A., Khan, S., Nadeem, M., Gaafar, A.-R. Z., and Tarroum, M. (2018). SCAR marker for gender identification in date palm (*Phoenix dactylifera* L.) at the seedling stage. *Int. J. Genomics* 2018, 3035406. doi:10.1155/2018/3035406

Al-Ruqaishi, I. A., Davey, M., Alderson, P., and Mayes, S. (2008). Genetic relationships and genotype tracing in date palms (*Phoenix dactylifera* L.) in Oman, based on microsatellite markers. *Plant Genet. Resour.* 6, 70–72. doi:10.1017/s1479262108923820

Al-Shahib, W., and Marshall, R. J. (2003). The fruit of the date palm: Its possible use as the best food for the future? *Int. J. Food Sci. Nutr.* 54, 247–259. doi:10.1080/09637480120091982

Al-Yahyai, R., and Al-Khanjari, S. (2008). Biodiversity of date palm in the Sultanate of Oman. *Afr. J. Agric. Res.* 3, 389–395.

Al-Yahyai, R., and Khan, M. M. (2015). "Date palm status and perspective in Oman," in *Date palm genetic resources and utilization* (Springer), 207–240.

Arabnezhad, H., Bahar, M., Mohammadi, H. R., and Latifian, M. (2012). Development, characterization and use of microsatellite markers for germplasm analysis in date palm (*Phoenix dactylifera* L.). *Sci. Hortic.* 134, 150–156. doi:10.1016/j.scienta.2011.11.032

Asaf, S., Khan, A. L., Al-Harrasi, A., and Al-Rawahi, A. (2018). First complete mitochondrial genome of *Phoenix dactylifera* var. Khanezi. *Mitochondrial DNA. B Resour.* 3, 778–779. doi:10.1080/23802359.2018.1491339

Atia, M. A., Adawy, S. S., and El-Itriby, H. A. (2017). "Date palm sex differentiation based on fluorescence *in situ* hybridization (FISH)," in *Date palm biotechnology protocols* (Springer), 245–256.

Awad, M. A., Al-Qurashi, A. D., and Mohamed, S. A. (2011). Antioxidant capacity, antioxidant compounds and antioxidant enzyme activities in five date cultivars during development and ripening. *Sci. Hortic.* 129, 688–693. doi:10.1016/j.scienta.2011.05.019

Ba-Angood, S. (2015). "Date palm status and perspective in Yemen," in *Date palm genetic resources and utilization* (Springer), 241–263.

Barakat, A., Han, D. T., Benslimane, A.-A., Rode, A., and Bernardi, G. (1999). The gene distribution in the genomes of pea, tomato and date palm. *FEBS Lett.* 463, 139–142. doi:10.1016/s0014-5793(99)01587-2

Bashah, M. A. (1996). *Date variety in the Kingdom of Saudi Arabia. Guidance booklet, palms and dates*. Riyadh: King Abdulaziz University Press.

Battaglia, M., Ghsera, B., Mancini, M., Bergesio, C., Camussi, A., and Racchi, M. L. (2015). "Date palm status and perspective in Libya," in *Date palm genetic resources and utilization* (Springer), 223–255.

Bekheet, S. A., Taha, H. S., Hanafy, M. S., and Solliman, M. E. (2008). Morphogenesis of sexual embryos of date palm cultured *in vitro* and early identification of sex type. *J. Appl. Sci. Res.* 4, 345–352.

Ben, A., Stiti, K., Lepoivre, P., and Jardin, P. D. (2000). Date palm (*Phoenix dactylifera* L.) cultivar identification using random amplified polymorphic DNA (RAPD). *Cah. Agric.* 9, 103–107.

Bendiab, K., Baaziz, M., Brakez, Z., and Sedra, My. H. (1992). Correlation of isoenzyme polymorphism and Bayoud-disease resistance in date palm cultivars and progeny. *Euphytica* 65, 23–32. doi:10.1007/BF00022196

Bendiab, K., Baaziz, M., and Majourhat, K. (1998). Preliminary date palm cultivar composition of Moroccan palm groves as revealed by leaf isoenzyme phenotypes. *Biochem. Syst. Ecol.* 26, 71–82. doi:10.1016/s0305-1978(97)00078-1

Bennaceur, M., Lanaud, C., Chevallier, M. H., and Bounaga, N. (1991). Genetic diversity of the date palm (*Phoenix dactylifera* L.) from Algeria revealed by enzyme markers. *Plant Breed.* 107, 56–69. doi:10.1111/j.1439-0523.1991.tb00528.x

Benslimane, A. A., Hartmann, C., Ouenzar, B., and Rode, A. (1996). Intramolecular recombination of a mitochondrial minicircular plasmid-like DNA of date-palm mediated by a set of short direct-repeat sequences. *Curr. Genet.* 29, 591–593. doi:10.1007/BF02426965

Benslimane, A. A., Rode, A., Quétier, F., and Hartmann, C. (1994). Characterization of two minicircular plasmid-like DNAs isolated from date-palm mitochondria. *Curr. Genet.* 26, 535–541. doi:10.1007/BF00309946

Billotte, N., Marseillac, N., Brottier, P., Noyer, J.-L., Jacquemoud-Collet, J.-P., Moreau, C., et al. (2004). Nuclear microsatellite markers for the date palm (*Phoenix dactylifera* L.): Characterization and utility across the genus *Phoenix* and in other palm genera. *Mol. Ecol. Notes* 4, 256–258. doi:10.1111/j.1471-8286.2004.00634.x

Bodian, A., El Houmaizi, M. A., Ndoye Ndir, K., Hasnaoui, A., Nachtigall, M., and Wehling, P. (2012). Genetic diversity analysis of date palm (*Phoenix dactylifera* L.) cultivars from Figuig oasis (Morocco) using SSR markers. *IJSAT* 2, 96–104.

Bodian, A., Nachtigall, M., Frese, L., Aziz Elhoumaizi, M., and Hasnaoui, A. (2014). Genetic diversity analysis of date palm (*Phoenix dactylifera* L.) cultivars from Morocco using SSR markers. *J. Biodivers. Biopros. Dev.* 1, 126. doi:10.4172/2376-0214.1000126

Bouguedoura, N., Bennaceur, M., Babahani, S., and Benziouche, S. E. (2015). "Date palm status and perspective in Algeria," in *Date palm genetic resources and utilization* (Springer), 125–168.

Bourgis, F., Kilaru, A., Cao, X., Ngando-Ebongue, G.-F., Drira, N., Ohlrogge, J. B., et al. (2011). Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *Proc. Natl. Acad. Sci. U. S. A.* 108, 12527–12532. doi:10.1073/pnas.1106502108

Cao, B. R., and Chao, C.-C. T. (2002). Identification of date cultivars in California using AFLP markers. *HortScience* 37, 966–968. doi:10.21273/hortsci.37.6.966

Chaluvadi, S. R., Khanam, S., Aly, M. A., and Bennetzen, J. L. (2014). Genetic diversity and population structure of native and introduced date palm (*Phoenix dactylifera*) germplasm in the United Arab Emirates. *Trop. Plant Biol.* 7, 30–41. doi:10.1007/s12042-014-9135-7

Chao, C. T., and Krueger, R. R. (2007). The date palm (*Phoenix dactylifera* L.): Overview of biology, uses, and cultivation. *HortScience* 42, 1077–1082. doi:10.21273/hortsci.42.5.1077

Cherif, E., Zehdi, S., Castillo, K., Chabrillange, N., Abdoulkader, S., Pintaud, J.-C., et al. (2013). Male-specific DNA markers provide genetic evidence of an XY chromosome system, a recombination arrest and allow the tracing of paternal lineages in date palm. *New Phytol.* 197, 409–415. doi:10.1111/nph.12069

Corniquel, B., and Mercier, L. (1994). Date palm (*Phoenix dactylifera* L.) cultivar identification by RFLP and RAPD. *Plant Sci.* 101, 163–172. doi:10.1016/0168-9452(94)90252-6

Darlington, C. D., and Wylie, A. P. (1956). *Chromosome atlas of flowering plants*. New York: The Macmillan Company, 519.

Devanand, P. S., and Chao, C. T. (2003a). Genetic variation within 'Medjool'and 'Deglet Noor' date (*Phoenix dactylifera* L.) cultivars in California detected by fluorescent-AFLP markers. *J. Hortic. Sci. Biotechnol.* 78, 405–409. doi:10.1080/14620316.2003.11511639

Devanand, P. S., and Chao, C. T. (2003b). Identification of genetic strains of 'Medjool' and 'Deglet Noor' date palm cultivars in California using amplified fragment length polymorpohism (AFLP) markers. *Acta Hortic.*

Dhawan, C., Kharb, P., Sharma, R., Uppal, S., and Aggarwal, R. K. (2013). Development of male-specific SCAR marker in date palm (*Phoenix dactylifera* L.). *Tree Genet. Genomes* 9, 1143–1150. doi:10.1007/s11295-013-0617-9

Diaz, S., Pire, C., Ferrer, J., and Bonete, M. J. (2003). Identification of *Phoenix dactylifera* L. varieties based on amplified fragment length polymorphism (AFLP) markers. *Cell. Mol. Biol. Lett.* 8, 891–899.

Djerbi, M. (1983). *Regional project for palm and dates research centre in near east and North Africa*.Diseases of the date palm (*Phoenix dactylifera*)

El Hadrami, A., and Al-Khayri, J. M. (2012). Socioeconomic and traditional importance of date palm. *Emir. J. Food Agric.* 24, 371.

El Kadri, N., Mimoun, M. B., and Hormaza, J. I. (2019). Genetic diversity of Tunisian male date palm (*Phoenix dactylifera* L.) genotypes using morphological descriptors and molecular markers. *Sci. Hortic.* 253, 24–34. doi:10.1016/j.scienta.2019.04.026

El Modafar, C. (2010). Mechanisms of date palm resistance to Bayoud disease: Current state of knowledge and research prospects. *Physiological Mol. Plant Pathology* 74, 287–294. doi:10.1016/j.pmpp.2010.06.008

El-Assar, A. M., Krueger, R. R., Devanand, P. S., and Chao, C.-C. T. (2005). Genetic analysis of Egyptian date (*Phoenix dactylifera* L.) accessions using AFLP markers. *Genet. Resour. Crop Evol.* 52, 601–607. doi:10.1007/s10722-004-0583-z

El-Kharbotly, A., El-Mardi, O., Al-Saidi, N., and Al-Mahruki, Y. (1998). "Towards the construction of genetic map of date palms using AFLP technique," in *Proceedings of the first international conference on date palm* (Al-Ain: United Arab Emirates University), 8–10.

El-Khishin, D. A., Adawy, S. S., Hussein, E. H., and El-Itriby, H. A. (2003). AFLP fingerprinting of some Egyptian date palm *(Phoenix dactylifera* L.*)* cultivars. *Arab J. Biotechnol.* 6, 223–234.

Elhoumaizi, M. A., Devanand, P. S., Fang, J., and Chao, C.-C. T. (2006). Confirmation of `Medjool' date as a landrace variety through genetic analysis of `Medjool' accessions in Morocco. *J. Am. Soc. Hortic. Sci.* 131, 403–407. doi:10.21273/jashs.131.3.403

Elmeer, K., and Mattat, I. (2015). Genetic diversity of Qatari date palm using SSR markers. *Genet. Mol. Res.* 14, 1624–1635. doi:10.4238/2015.March.6.9

Elmeer, K., and Mattat, I. (2012). Marker-assisted sex differentiation in date palm using simple sequence repeats. *3 Biotech.* 2, 241–247. doi:10.1007/s13205-012-0052-x

Elmeer, K., Sarwath, H., Malek, J., Baum, M., and Hamwieh, A. (2011). New microsatellite markers for assessment of genetic diversity in date palm (*Phoenix dactylifera* L.). *3 Biotech.* 1, 91–97. doi:10.1007/s13205-011-0010-z

Elsafy, M., Zborowska, A., Bryngelsson, T., Fatih, M., Mujaju, C., and Garkava-Gustavsson, L. (2016). Elucidating the genetic diversity of farmer cultivars of female date palms (*Phoenix dactylifera* L.) from Sudan by microsatellite markers. *Genet. Resour. Crop Evol.* 63, 975–986. doi:10.1007/s10722-015-0294-7

Elshibli, S. (2009). *Genetic diversity and adaptation of date palm (*Phoenix dactylifera *L.)*. Dissertation. Finland: University of Helsinki.

Elshibli, S., and Korpelainen, H. (2009). Excess heterozygosity and scarce genetic differentiation in the populations of *Phoenix dactylifera* L.: Human impact or ecological determinants. *Plant Genet. Resour.* 7, 95–104. doi:10.1017/s1479262108060905

Elshibli, S., and Korpelainen, H. (2008). Microsatellite markers reveal high genetic diversity in date palm (*Phoenix dactylifera* L.) germplasm from Sudan. *Genetica* 134, 251–260. doi:10.1007/s10709-007-9232-8

Emoghene, B. O., Asemota, O., Idu, M., and Eke, C. R. (2015). Molecular characterization of some date palms in Nigeria using RAPD markers. *J. Appl. Biol. Biotechnol.* 3, 026–030.

Fang, Y., Wu, H., Zhang, T., Yang, M., Yin, Y., Pan, L., et al. (2012). A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. *PloS one* 7, e37164. doi:10.1371/journal.pone.0037164

Faqir, N., Muhammad, A., Ali, G. M., Shehzad, A., Ali, K., Rahman, H. U., et al. (2019). Development of a single nucleotide polymorphism-based genetic identification key for date palm cultivars. *Crop Sci.* 59, 2509–2520.

Faqir, N., Muhammad, A., Shehzad, A., Hyder, M. Z., and Ali, G. M. (2016). Simple sequence repeat (SSR) markers show greater similarity among morphologically diverse date palm (*Phoenix dactylifera* L.) cultivars grown in Pakistan. *Pure Appl. Biol.* 5, 483. doi:10.19045/bspab.2016.50063

Fayadh, J. M., and Al-Showiman, S. S. (1990). Chemical composition of date palm (*Phoenix dactylifera* L.). *J. Chem. Soc. Pak.* 12, 84–103.

Giovino, A., Bertolini, E., Fileccia, V., Al Hassan, M., Labra, M., and Martinelli, F. (2015). Transcriptome analysis of *Phoenix canariensis* Chabaud in response to *Rhynchophorus ferrugineus* Olivier attacks. *Front. Plant Sci.* 6, 817. doi:10.3389/fpls.2015.00817

Gros-Balthazard, M., Galimberti, M., Kousathanas, A., Newton, C., Ivorra, S., Paradis, L., et al. (2017). The discovery of wild date palms in Oman reveals a complex domestication history involving centers in the Middle East and Africa. *Curr. Biol.* 27, 2211–2218. doi:10.1016/j.cub.2017.06.045

Guettouchi, A., Elshibli, S., Haider, N., Nabulsi, I., and Ykhlef, N. (2017). "Molecular diversity in date palm (*Phoenix dactylifera* L.) cultivars from Algeria indicated by RAPD and ISSR polymorphisms," in *Plant cell biotechnology and molecular biology*, 76–89.

Gupta, P. K., Roy, J. K., and Prasad, M. (2001). Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Curr. Sci.*, 524–535.

Haider, N., Nabulsi, I., and MirAli, N. (2012). Phylogenetic relationships among date palm (*Phoenix dactylifera* L.) cultivars in Syria using RAPD and ISSR markers. *J. Plant Biol. Crop Res.* 1, 12–24.

Haider, N., and Nabulsi, I. (2012). Molecular characterization of Syrian date palm cultivars using plasmid-like DNA markers. *Russ. J. Genet.* 48, 240–244. doi:10.1134/s1022795412010085

Hajian, S., and Hamidi-Esfahani, Z. (2015). "Date palm status and perspective in Iran," in *Date palm genetic resources and utilization* (Springer), 19–47.

Hammadi, H., Vendramin, G. G., and Ali, A. F. (2011). Microsatellite diversity among Tunisian date palm (*Phoenix dactylifera* L.) subpopulations. *Pak. J. Bot.* 43, 1257–1264.

Hamwieh, A., Farah, J., Moussally, S, Khaled El-Shamaa, Almer, K., Khierallah, H., et al. (2010). Development of 1000 microsatellite markers across the date palm (*Phoenix dactylifera* L.) genome. *Acta Hortic.* 882, 269–277. doi:10.17660/actahortic.2010.882.29

Hazzouri, K. M., Flowers, J. M., Visser, H. J., Khierallah, H. S., Rosas, U., Pham, G. M., et al. (2015). Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nat. Commun.* 6. doi:10.1038/ncomms9824

Hazzouri, K. M., Gros-Balthazard, M., Flowers, J. M., Copetti, D., Lemansour, A., Lebrun, M., et al. (2019). Genome-wide association mapping of date palm fruit traits. *Nat. Commun.* 10. doi:10.1038/s41467-019-12604-9

He, Z., Zhang, C., Liu, W., Lin, Q., Wei, T., Aljohi, H. A., et al. (2017). Drdb: An online date palm genomic resource database. *Front. Plant Sci.* 8, 1889. doi:10.3389/fpls.2017.01889

Huda, M. N., Hasan, M., Abdullah, H. M., and Sarker, U. (2019). Spatial distribution and genetic diversity of wild date palm (*Phoenix sylvestris*) growing in coastal Bangladesh. *Tree Genet. Genomes* 15, 3. doi:10.1007/s11295-018-1310-9

Jana, G. A., Krishnamurthy, P., Kumar, P. P., and Yaish, M. W. (2021). Functional characterization and expression profiling of glyoxalase III genes in date palm grown under abiotic stresses. *Physiol. Plant.* 172, 780–794. doi:10.1111/ppl.13239

Jana, G. A., and Yaish, M. W. (2020). Functional characterization of the Glyoxalase-I (*PdGLX1*) gene family in date palm under abiotic stresses. *Plant Signal. Behav.* 15, 1811527. doi:10.1080/15592324.2020.1811527

Johnson, C., Cullis, T. A., Cullis, M. A., and Cullis, C. A. (2009). DNA markers for variety identification in date palm (*Phoenix dactylifera* L.). *J. Hortic. Sci. Biotechnol.* 84, 591–594. doi:10.1080/14620316.2009.11512570

Johnson, D. V. (2011). "Introduction: Date palm biotechnology from theory to practice," in *Date palm biotechnology* (Springer), 1–11.

Jubrael, J. M., Udupa, S. M., and Baum, M. (2005). Assessment of AFLP-based genetic relationships among date palm (*Phoenix dactylifera* L.) varieties of Iraq. *J. Am. Soc. Hortic. Sci.* 130, 442–447. doi:10.21273/jashs.130.3.442

Kassem, H. S., Alotaibi, B. A., Ahmed, A., and Aldosri, F. O. (2020). Sustainable management of the red palm weevil: The nexus between farmers' adoption of integrated pest management and their knowledge of symptoms. *Sustainability* 12, 9647. doi:10.3390/su12229647

Khan, A., Khan, I. A., Heinze, B., and Azim, M. K. (2012). The chloroplast genome sequence of date palm (*Phoenix dactylifera* L. cv. 'Aseel'). *Plant Mol. Biol. Rep.* 30, 666–678. doi:10.1007/s11105-011-0373-7

Khan, A. L., Asaf, S., Lee, I.-J., Al-Harrasi, A., and Al-Rawahi, A. (2018). First chloroplast genomics study of *Phoenix dactylifera* (var. Naghal and Khanezi): A comparative analysis. *PloS one* 13, e0200104. doi:10.1371/journal.pone.0200104

Khierallah, H., Bader, S., Baum, M., and Hamwieh, A. (2011a). Assessment of genetic diversity for some Iraqi date palms (*Phoenix dactylifera* L.) using amplified fragment length polymorphism (AFLP) markers. *Afr. J. Biotechnol.* 10, 9570–9576. doi:10.5897/ajb11.055

Khierallah, H. S., Al-Sammarraie, S. K., and Mohammed, H. I. (2014). Molecular characterization of some Iraqi date palm cultivars using RAPD and ISSR markers. *J. Asian Sci. Res.* 4, 490.

Khierallah, H. S., Bader, S. M., Baum, M., and Hamwieh, A. (2011b). Genetic diversity of Iraqi date palms revealed by microsatellite polymorphism. *J. Am. Soc. Hortic. Sci.* 136, 282–287. doi:10.21273/jashs.136.4.282

Mahar, A. Q. (2007). *Post-harvest studies of different cultivars of date palm (*Phoenix dactylifera *L.) fruits, their protection, identification, processing and preservation at district Khairpur, Sindh, Pakistan*. Khairpur, Sindh, Pakistan: PhD thesis, Date Palm Research Institute, Shah Abdul Latif University.

Maina, N., Baraket, G., Salhi-Hannachi, A., and Sakka, H. (2019). Sequence analysis and molecular evolution of Tunisian date palm cultivars (*Phoenix dactylifera* L.) based on the internal transcribed spacers (ITSs) region of the nuclear ribosomal DNA. *Sci. Hortic.* 247, 373–379. doi:10.1016/j.scienta.2018.12.045

Malek, J. A. (2010). Next generation DNA sequencing applied to the Date palm tree (*Phoenix dactylifera*). *Acta Hortic.* 882, 249–252. doi:10.17660/actahortic.2010.882.27

Marqués, J., Fadda, Z. G. N., Duran-Vila, N., Flores, R., Bové, J. M., and Daròs, J. A. (2008). A set of novel RNAs transcribed from the chloroplast genome accumulates in date palm leaflets affected by brittle leaf disease. *Phytopathology* 98, 337–344. doi:10.1094/PHYTO-98-3-0337

Maryam Jaskani, M. J., Awan, F. S., Ahmad, S., and Khan, I. A. (2016). Development of molecular method for sex identification in date palm (*Phoenix dactylifera* L.) plantlets using novel sex-linked microsatellite markers. *3 Biotech.* 6, 22. doi:10.1007/s13205-015-0321-6

Mason, S. C. (1915). *Botanical characters of the leaves of the date palm used in distinguishing cultivated varieties*. Washington, D.C: US Department of Agriculture.

Mathew, L. S., Seidel, M. A., George, B., Mathew, S., Spannagl, M., Haberer, G., et al. (2015). A genome-wide survey of date palm cultivars supports two major subpopulations in *Phoenix dactylifera*. *G3* 5, 1429–1438. doi:10.1534/g3.115.018341

Mathew, L. S., Spannagl, M., Al-Malki, A., George, B., Torres, M. F., Al-Dous, E. K., et al. (2014). A first genetic map of date palm (*Phoenix dactylifera*) reveals long-range genome structure conservation in the palms. *BMC Genomics* 15, 285. doi:10.1186/1471-2164-15-285

Michielse, C. B., and Rep, M. (2009). Pathogen profile update: *Fusarium oxysporum*. *Mol. Plant Pathol.* 10, 311–324. doi:10.1111/j.1364-3703.2009.00538.x

Milne, D. (1918). *The date palm and its cultivation in the Punjab*. Punjab Government.

Mohamed Lemine, F. M., Mohamed Ahmed, M. V. O., Ben Mohamed Maoulainine, L., Bouna, Z. el A. O., Samb, A., and Boukhary, O. (2014). Antioxidant activity of various Mauritanian date palm (*Phoenix dactylifera*) fruits at two edible ripening stages. *Food Sci. Nutr.* 2 (6), 700–705. doi:10.1002/fsn3.167

Mohamoud, Y. A., Mathew, L. S., Torres, M. F., Younuskunju, S., Krueger, R., Suhre, K., et al. (2019). Novel subpopulations in date palm (*Phoenix dactylifera*) identified by population-wide organellar genome sequencing. *BMC Genomics* 20, 498. doi:10.1186/s12864-019-5834-7

Mohei, E. L., Mohasseb, H. A. A., Al-Khateeb, A. A., Al-Khateeb, S. A., Chowdhury, K., El-Shemy, H. A., et al. (2019). Identification and sequencing of date-SRY gene: A novel tool for sex determination of date palm (*Phoenix dactylifera* L.). *Saudi J. Biol. Sci.* 26, 514–523. doi:10.1016/j.sjbs.2017.08.002

Mokhtar, M. M., Adawy, S. S., El-Assal, S. E.-D. S., and Hussein, E. H. (2016). Genic and intergenic SSR database generation, SNPs determination and pathway annotations in date palm (*Phoenix dactylifera* L.). *PLOS one* 11, e0159268. doi:10.1371/journal.pone.0159268

Moore, H. E. (1963). *An annotated checklist of cultivated palms*. Palm Society.

Moussouni, S., Pintaud, J.-C., Vigouroux, Y., and Bouguedoura, N. (2017). Diversity of Algerian oases date palm (*Phoenix dactylifera* L., Arecaceae): Heterozygote excess and cryptic structure suggest farmer management had a major impact on diversity. *PloS one* 12, e0175232. doi:10.1371/journal.pone.0175232

Munier, P. (1973). *Le palmier dattier-techniques agricoles et productions tropicales*. France: Maisonneuve et Larousse.

Naganeeswaran, S., Fayas, T. P., and Rajesh, M. K. (2020). Dataset of transcriptome assembly of date palm embryogenic calli and functional annotation. *Data in Brief* 20.105760. doi:10.1016/j.dib.2020.105760

Namsi, A., Marqués, J., Fadda, Z., Takrouni, M. L., Mahamoud, O. B., Zouba, A., et al. (2006). Diagnosis of "maladie des feuilles cassantes" or brittle leaf disease of date palms by detection of associated chloroplast encoded double stranded RNAs. *Mol. Cell. Probes* 20, 366–370. doi:10.1016/j.mcp.2006.05.001

Namsi, A., Montarone, M., Serra, P., Mahamoud, O. B., Takrouni, M. L., Zouba, A., et al. (2007). Manganese and brittle leaf disease of date palm trees. *J. Plant Pathology*, 125–136.

Osman, A. M. A. (1984). "The performance of date palms in the Sudan," in *Acta horticulturae* (Leuven, Belgium): International Society for Horticultural Science ISHS), 231–238. doi:10.17660/ActaHortic.1984.143.25

Panga, J. A. (2014). *Plants of AMS garden: A garden in the arabian deserts of Dubai*. Xlibris AU.

Patankar, H. V., Al-Harrasi, I., Al Kharusi, L., Jana, G. A., Al-Yahyai, R., Sunkar, R., et al. (2019a). Overexpression of a metallothionein 2A gene from date palm confers abiotic stress tolerance to yeast and *Arabidopsis thaliana*. *Int. J. Mol. Sci.* 20, 2871. doi:10.3390/ijms20122871

Patankar, H. V., Al-Harrasi, I., Al-Yahyai, R., and Yaish, M. W. (2019b). Functional characterization of date palm aquaporin gene *PdPIP1;2* confers drought and salinity tolerance to yeast and *Arabidopsis*. *Genes* 10, 390. doi:10.3390/genes10050390

Patankar, H. V., Al-Harrasi, I., Al-Yahyai, R., and Yaish, M. W. (2018). Identification of candidate genes involved in the salt tolerance of Date Palm (*Phoenix dactylifera* L.) based on a Yeast functional bioassay. *DNA Cell Biol.* 37, 524–534. doi:10.1089/dna.2018.4159

Pathak, M. R., and Hamzah, R. Y. (2008). RAPD analysis of date palm cultivars of Bahrain. *Group* 1, 00.

Purayil, F. T., Robert, G. A., Gothandam, K. M., Kurup, S. S., Subramaniam, S., and Cheruth, A. J. (2018). Genetic variability in selected date palm (*Phoenix dactylifera* L.) cultivars of United Arab Emirates using ISSR and DAMD markers. *3 Biotech.* 8, 109. doi:10.1007/s13205-018-1108-3

Qacif, N., Baaziz, M., and Bendiab, K. (2007). Biochemical investigations on peroxidase contents of male and female inflorescences of date palm (*Phoenix dactylifera* L.). *Sci. Hortic.* 114, 298–301. doi:10.1016/j.scienta.2007.07.015

Quenzar, B., Trifi, M., Bouachrine, B., Hartmann, C., Marrakchi, M., Benslimane, A. A., et al. (2001). A mitochondrial molecular marker of resistance to Bayoud disease in date palm. *Theor. Appl. Genet.* 103, 366–370. doi:10.1007/s001220100572

Rabei, S., Said, W. M., Rizk, R. M., and El, S. F. (2012). Morphometric taxonomy of date palm diversity growing in Egypt. *J. Bot.* 189, 175–189.

Racchi, M. L., Bove, A., Turchi, A., Bashir, G., Battaglia, M., and Camussi, A. (2014). Genetic characterization of Libyan date palm resources by microsatellite markers. *3 Biotech.* 4, 21–32. doi:10.1007/s13205-013-0116-6

Racchi, M. L., and Camussi, A. (2018). The date palms of Al jufrah-Libya: A survey on genetic diversity of local varieties. *J. Agric. Environ. Int. Dev. (JAEID)* 112, 161–184.

Radwan, O., Arro, J., Keller, C., and Korban, S. S. (2015). RNA-Seq transcriptome analysis in date palm suggests multi-dimensional responses to salinity stress. *Trop. Plant Biol.* 8, 74–86. doi:10.1007/s12042-015-9155-y

Rekik, I., Chaâbene, Z., Kriaa, W., Rorat, A., Franck, V., Hafedh, M., et al. (2019). Transcriptome assembly and abiotic related gene expression analysis of date palm reveal candidate genes involved in response to cadmium stress. *Comp. Biochem. Physiol. C. Toxicol. Pharmacol.* 225, 108569. doi:10.1016/j.cbpc.2019.108569

Rhouma, S., Dakhlaoui-Dkhil, S., Salem, A. O. M., Zehdi-Azouzi, S., Rhouma, A., Marrakchi, M., et al. (2008). Genetic diversity and phylogenic relationships in date-palms (*Phoenix dactylifera* L.) as assessed by random amplified microsatellite polymorphism markers (RAMPOs). *Sci. Hortic.* 117, 53–57. doi:10.1016/j.scienta.2008.03.013

Rhouma, S., Zehdi-Azouzi, S., Dakhlaoui-Dkhil, S., Salem, A. O. M., Othmani, A., Cherif, E., et al. (2010). Genetic variation in the Tunisian date palm (*Phoenix dactylifera* L.). *Desert Plants*, 355–370. doi:10.1007/978-3-642-02550-1_17

Rhouma, S., Zehdi-Azouzi, S., Ould Mohamed Salem, A., Rhouma, A., Marrakchi, M., and Trifi, M. (2007). Genetic diversity in ecotypes of Tunisian date-palm (*Phoenix dactylifera* L.) assessed by AFLP markers. *J. Hortic. Sci. Biotechnol.* 82, 929–933. doi:10.1080/14620316.2007.11512328

Rivera, D., Johnson, D., Delgadillo, J., Carrillo, M. H., Obón, C., Krueger, R., et al. (2013). Historical evidence of the Spanish introduction of date palm (*Phoenix dactylifera* L., Arecaceae) into the Americas. *Genet. Resour. Crop Evol.* 60, 1433–1452. doi:10.1007/s10722-012-9932-5

Sabir, J. S., Abo-Aba, S., Bafeel, S., Zari, T. A., Edris, S., Shokry, A. M., et al. (2014a). Characterization of ten date palm (*Phoenix dactylifera* L.) cultivars from Saudi Arabia using AFLP and ISSR markers. *C. R. Biol.* 337, 6–18. doi:10.1016/j.crvi.2013.11.003

Sabir, J. S. M., Arasappan, D., Bahieldin, A., Abo-Aba, S., Bafeel, S., Zari, T. A., et al. (2014b). Whole mitochondrial and plastid genome SNP analysis of nine date palm cultivars reveals plastid heteroplasmy and close phylogenetic relationships among cultivars. *PloS one* 9, e94158. doi:10.1371/journal.pone.0094158

Saboori, S., Noormohammadi, Z., Sheidai, M., and Marashi, S. (2020). SCoT molecular markers and genetic fingerprinting of date palm (*Phoenix dactylifera* L.) cultivars. *Genet. Resour. Crop Evol.* 67, 73–82. doi:10.1007/s10722-019-00854-x

Safronov, O., Kreuzwieser, J., Haberer, G., Alyousif, M. S., Schulze, W., Al-Harbi, N., et al. (2017). Detecting early signs of heat and drought stress in *Phoenix dactylifera* (date palm). *PloS one* 12, e0177883. doi:10.1371/journal.pone.0177883

Saidi, M. N., Jbir, R., Ghorbel, I., Namsi, A., Drira, N., and Gargouri-Bouzid, R. (2012). Brittle leaf disease induces an oxidative stress and decreases the expression of manganese-related genes in date palm (*Phoenix dactylifera* L.). *Plant Physiol. biochem.* 50, 1–7. doi:10.1016/j.plaphy.2011.09.016

Saidi, M. N., Ladouce, N., Hadhri, R., Grima-Pettenati, J., Drira, N., and Gargouri-Bouzid, R. (2010). Identification and characterization of differentially expressed ESTs in date palm leaves affected by brittle leaf disease. *Plant Sci.* 179, 325–332. doi:10.1016/j.plantsci.2010.06.002

Saleh, A. A., El_Komy, M. H., Eranthodi, A., Hamoud, A. S., and Molan, Y. Y. (2015). Variation in a molecular marker for resistance of Saudi date palm germplasm to *Fusarium oxysporum* f. sp. *albedinis*, the causal agent of Bayoud disease. *Eur. J. Plant Pathol.* 143, 507–514. doi:10.1007/s10658-015-0703-y

Salem, A. O. M., Rhouma, S., Zehdi, S., Marrakchi, M., and Trifi, M. (2007). Molecular characterization of Mauritanian date palm cultivars using plasmid-like DNA markers. *Biol. Plant.* 51, 169–172.

Salomon-Torres, R., Ortiz-Uribe, N., Villa-Angulo, C., Villa-Angulo, R., and Yaurima-Basaldúa, V. H. (2017). Assessment of SSR markers used in analysis of genetic diversity of date palm (*Phoenix dactylifera* L.). *Plant Cell Biotechnol. Mol. Biol.* 18, 269–280.

Sedra, M. H. (2015). "Date palm status and perspective in Morocco," in *Date palm genetic resources and utilization* (Springer), 257–323.

Sedra, M. H., Lashermes, P., Trouslot, P., Combes, M.-C., and Hamon, S. (1998). Identification and genetic diversity analysis of date palm (*Phoenix dactylifera* L.) varieties from Morocco using RAPD markers. *Euphytica* 103, 75–82. doi:10.1023/a:1018377827903

Shaheen, M. A. (1990). Propagation of date palm through tissue culture: A review and an interpretation. *Ann. Agric. Sci.* 35, 895–909.

Sharifi, M., Sheidai, M., and Koohdar, F. (2018). Genetic fingerprinting of date palm (*Phoenix dactylifera* L.) by using ISSR and cpDNA sequences. *Indian J. Genet.* 78, 507–514.

Siljak-Yakovlev, S., Cerbah, M., Sarr, A., Benmalek, S., Bounaga, N., de la Pena, T. C., et al. (1996). Chromosomal sex determination and heterochromatin structure in date palm. *Sex. Plant Reprod.* 9, 127–132. doi:10.1007/bf02221391

Soliman, S. S., Ali, B. A., and Ahmed, M. M. M. (2003). Genetic comparisons of Egyptian date palm cultivars (*Phoenix dactylifera* L.) by RAPD-PCR. *Afr. J. Biotechnol.* 2, 86–87. doi:10.5897/ajb2003.000-1017

Soumaya, R. C., Ghada, B., Sonia, D. D., Salwa, Z.-A., and Mokhtar, T. (2011). Molecular research on the genetic diversity of Tunisian date palm (*Phoenix dactylifera* L.) using the random amplified microsatellite polymorphism (RAMPO) and amplified fragment length polymorphism (AFLP) methods. *Afr. J. Biotechnol.* 10, 10352–10365. doi:10.5897/ajb10.2242

Takele, D., Tsegaw, M., and Indracanti, M. (2021). Genetic diversity assessment in some landraces and varieties of date palm (*Phoenix dactylifera* L.) from Afar Region, Ethiopia, using ISSR markers. *Ecol. Genet. Genomics* 19, 100085. doi:10.1016/j.egg.2021.100085

Tengberg, M. (2012). Beginnings and early history of date palm garden cultivation in the Middle East. *J. Arid Environ.* 86, 139–147. doi:10.1016/j.jaridenv.2011.11.022

Thareja, G., Mathew, S., Mathew, L. S., Mohamoud, Y. A., Suhre, K., and Malek, J. A. (2018). Genotyping-by-sequencing identifies date palm clone preference in agronomics of the State of Qatar. *PLoS one* 13, e0207299. doi:10.1371/journal.pone.0207299

Torres, A. M., and Tisserat, B. (1980). Leaf isozymes as genetic markers in date palms. *Am. J. Bot.* 67, 162–167. doi:10.1002/j.1537-2197.1980.tb07637.x

Torres, M. F., Mathew, L. S., Ahmed, I., Al-Azwani, I. K., Krueger, R., Rivera-Nuñez, D., et al. (2018). Genus-wide sequencing supports a two-locus model for sex-determination in *Phoenix*. *Nat. Commun.* 9, 3969–9. doi:10.1038/s41467-018-06375-y

Torres, M. F., Mohamoud, Y. A., Younuskunju, S., Suhre, K., and Malek, J. A. (2021). Evidence of recombination suppression blocks on the Y chromosome of date palm (*Phoenix dactylifera*). *Front. Plant Sci.* 12, 697. doi:10.3389/fpls.2021.634901

Triki, M. A., Zouba, A., Khoualdia, O., Mahamoud, O. B., Takrouni, M. L., Garnier, M., et al. (2003). Maladie des feuilles cassantes" or brittle leaf disease of date palms in Tunisia: Biotic or abiotic disease? *J. Plant Pathology*, 71–79.

Wrigley, G. (1995). Date palm (*Phoenix dactylifera* L.). *Evol. Crop Plants* 2, 399–403.

Yaish, M. W., and Kumar, P. P. (2015). Salt tolerance research in date palm tree (*Phoenix dactylifera* L.), past, present, and future perspectives. *Front. Plant Sci.* 6, 348. doi:10.3389/fpls.2015.00348

Yaish, M. W., Patankar, H. V., Assaha, D. V., Zheng, Y., Al-Yahyai, R., and Sunkar, R. (2017). Genome-wide expression profiling in leaves and roots of date palm (*Phoenix dactylifera* L.) exposed to salinity. *BMC Genomics* 18, 246. doi:10.1186/s12864-017-3633-6

Yaish, M. W., Sunkar, R., Zheng, Y., Ji, B., Al-Yahyai, R., and Farooq, S. A. (2015). A genome-wide identification of the miRNAome in response to salinity stress in date palm (*Phoenix dactylifera* L.). *Front. Plant Sci.* 6, 946. doi:10.3389/fpls.2015.00946

Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., Yun, Q., et al. (2010). The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PloS one* 5, e12762. doi:10.1371/journal.pone.0012762

Yin, Y., Zhang, X., Fang, Y., Pan, L., Sun, G., Xin, C., et al. (2012). High-throughput sequencing-based gene profiling on multi-staged fruit development of date palm (*Phoenix dactylifera*, L.). *Plant Mol. Biol.* 78, 617–626. doi:10.1007/s11103-012-9890-5

Younis, R. A., Ismail, O. M., and Soliman, S. S. (2008). Identification of sex-specific DNA markers for date palm (*Phoenix dactylifera* L.) using RAPD and ISSR techniques. *Res. J. Agric. Biol. Sci.* 4, 278–284.

Zabar, A. F., and Borowy, A. (2012). in *Cultivation of date palm in Iraq*, 22, 39–54.*Annales Universitatis Mariae Curie-SkłodowskaSect. EEE, Hortic.*

Zaid, A., and De Wet, P. F. (1999). "Chapter I botanical and systematic description of date palm," in *FAO plant production and protection papers*, 1–28.

Zango, O., Cherif, E., Chabrillange, N., Zehdi-Azouzi, S., Gros-Balthazard, M., Naqvi, S. A., et al. (2017). Genetic diversity of Southeastern Nigerien date palms reveals a secondary structure within Western populations. *Tree Genet. Genomes* 13, 75. doi:10.1007/s11295-017-1150-z

Zehdi, S., Cherif, E., Rhouma, S., Santoni, S., Hannachi, A. S., and Pintaud, J. C. (2012). Molecular polymorphism and genetic relationships in date palm (*Phoenix dactylifera* L.): The utility of nuclear microsatellite markers. *Sci. Hortic.* 148, 255–263. doi:10.1016/j.scienta.2012.10.011

Zehdi, S., Trifi, M., Billotte, N., Marrakchi, M., and Christophe Pintaud, J. (2004). Genetic diversity of Tunisian date palms (*Phoenix dactylifera* L.) revealed by nuclear microsatellite polymorphism. *Hereditas* 141, 278–287. doi:10.1111/j.1601-5223.2004.01855.x

Zehdi, S., Trifi, M., Ould Mohamed Salem, A., Marrakchi, M., and Rhouma, A. (2002). Survey of inter simple sequence repeat polymorphisms in Tunisian date palms (*Phoenix dactylifera* L.). *J. Genet. Breed. (Italy).*

Zehdi-Azouzi, S., Cherif, E., Moussouni, S., Gros-Balthazard, M., Abbas Naqvi, S., Ludena, B., et al. (2015). Genetic structure of the date palm (*Phoenix dactylifera*) in the Old World reveals a strong differentiation between eastern and western populations. *Ann. Bot.* 116, 847–112. doi:10.1093/aob/mcv132

Zhang, G., Pan, L., Yin, Y., Liu, W., Huang, D., Zhang, T., et al. (2012). Large-scale collection and annotation of gene models for date palm (*Phoenix dactylifera*, L.). *Plant Mol. Biol.* 79, 521–536. doi:10.1007/s11103-012-9924-z

Zhao, Y., Keremane, M., Prakash, C. S., and He, G. (2017). Characterization and amplification of gene-based simple sequence repeat (SSR) markers in date palm. *Methods Mol. Biol.* 1638, 259–271. doi:10.1007/978-1-4939-7159-6_21

Zhao, Y., Williams, R., Prakash, C. S., and He, G. (2012). Identification and characterization of gene-based SSR markers in date palm (*Phoenix dactylifera* L.). *BMC Plant Biol.* 12, 237. doi:10.1186/1471-2229-12-237

| Frontiers in Genetics

TYPE Review
PUBLISHED 06 October 2022
DOI 10.3389/fgene.2022.999207

Check for updates

OPEN ACCESS

EDITED BY
Ravi Valluru,
University of Lincoln, United Kingdom

REVIEWED BY
Mehboob-ur- Rahman,
National Institute for Biotechnology and Genetic Engineering, Pakistan
Rebecca Ford,
Griffith University, Australia

*CORRESPONDENCE
Pooja Bhatnagar-Mathur,
p.b.mathur@cgiar.org
Kiran K. Sharma,
kksharma@teri.res.in

SPECIALTY SECTION
This article was submitted to Plant Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 20 July 2022
ACCEPTED 09 September 2022
PUBLISHED 06 October 2022

CITATION
Sharma KK, Palakolanu SR, Bhattacharya J, Shankhapal AR and Bhatnagar-Mathur P (2022), CRISPR for accelerating genetic gains in under-utilized crops of the drylands: Progress and prospects.
Front. Genet. 13:999207.
doi: 10.3389/fgene.2022.999207

COPYRIGHT
© 2022 Sharma, Palakolanu, Bhattacharya, Shankhapal and Bhatnagar-Mathur. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

CRISPR for accelerating genetic gains in under-utilized crops of the drylands: Progress and prospects

Kiran K. Sharma[1,2]*, Sudhakar Reddy Palakolanu[2], Joorie Bhattacharya[2,3], Aishwarya R. Shankhapal[4,5] and Pooja Bhatnagar-Mathur[2,6]*

[1]Sustainable Agriculture Programme, The Energy and Resources Institute (TERI), India Habitat Center, New Delhi, India, [2]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad, India, [3]Department of Genetics, Osmania University, Hyderabad, Telangana, India, [4]Division of Plant and Crop Sciences, School of Biosciences, University of Nottingham, Nottingham, United Kingdom, [5]Plant Sciences and the Bioeconomy, Rothamsted Research, Harpenden, Hertfordshire, United Kingdom, [6]International Maize and Wheat Improvement Center (CIMMYT), México, United Kingdom

Technologies and innovations are critical for addressing the future food system needs where genetic resources are an essential component of the change process. Advanced breeding tools like "genome editing" are vital for modernizing crop breeding to provide game-changing solutions to some of the "must needed" traits in agriculture. CRISPR/Cas-based tools have been rapidly repurposed for editing applications based on their improved efficiency, specificity and reduced off-target effects. Additionally, precise gene-editing tools such as base editing, prime editing, and multiplexing provide precision in stacking of multiple traits in an elite variety, and facilitating specific and targeted crop improvement. This has helped in advancing research and delivery of products in a short time span, thereby enhancing the rate of genetic gains. A special focus has been on food security in the drylands through crops including millets, teff, fonio, quinoa, Bambara groundnut, pigeonpea and cassava. While these crops contribute significantly to the agricultural economy and resilience of the dryland, improvement of several traits including increased stress tolerance, nutritional value, and yields are urgently required. Although CRISPR has potential to deliver disruptive innovations, prioritization of traits should consider breeding product profiles and market segments for designing and accelerating delivery of locally adapted and preferred crop varieties for the drylands. In this context, the scope of regulatory environment has been stated, implying the dire impacts of unreasonable scrutiny of genome-edited plants on the evolution and progress of much-needed technological advances.

KEYWORDS

CRiSPR/Cas, food security, gene editing, green revolution, new breeding technologies, speed breeding, underutilized crop, genetic gains

Introduction

Climate change, population growth, pandemic, conflicts and rising socio-economic disparities are putting considerable pressure on the already stressed food systems. While the precise impacts of climate change are still unclear, unpredictable climate changes are expected to impact several vulnerable regions more adversely. This requires informed approaches to address sustainability issues to meet the future food needs. While staple crops have limited resilience to changing climate in the drylands, locally grown underutilized crops despite being vital for diverse nutrient and local adaptations are grown in low-input conditions. Underutilized crops (under-researched compared to staple crops; Chapman et al., 2022) have significant potential to aid food security through increased food production in challenging environments where major crops are severely limited (Mayes et al., 2012). More recently, the discovery of genomes and candidate genes have aided the study of underutilized cereal and legumes and provided syntenic comparisons for the evolution of C4 photosynthesis, with the potential to improve the photorespiration efficiency, drought tolerance, and nutritional traits. However, significant efforts are still needed to identify and understand the underlying allelic variation for breeding applications (Chapman et al., 2022).

Advanced breeding technologies such as genome editing hold immense potential for improving crop yields and quality by inducing precise genetic alterations in the targeted genomes (Wolt et al., 2016; Miladinovic et al., 2021). Emergence of programmable site directed nucleases (SDNs) such as zinc finger nucleases or ZFNs (Carroll, 2011), transcription activator-like effector nucleases or TALENs (Zhang et al., 2013), clustered regularly interspersed short palindromic repeats or CRISPRs (Jiang et al., 2013a), and more recent base editing and prime editing tools have provided technological breakthrough for inducing precise and rapid genetic variations in organisms including plants (Komor et al., 2016; Mishra et al., 2020; Xu et al., 2020; Miladinovic et al., 2021). SDNs cleave the DNA sequence at specific sites and repair the double strand breaks (DSBs) through homologous recombination (HR) or non-homologous end joining (NHEJ) repair pathways, resulting in sequence replacements or creating insertions or deletions (INDELs) at predefined sites. Several successful applications have been reported for trait improvements in plants such as poplar (Fan et al., 2015), soybean (Li et al., 2015; Bao et al., 2019), wheat (Cui et al., 2019), tomato (Nekrasov et al., 2017), sorghum (Li A. et al., 2018), cassava (Hummel et al., 2018) and rice (Endo et al., 2016) for addressing complex traits such as heterosis (Wang et al., 2019), nutrition (Ku and Ha, 2020), stress tolerance (Jain, 2015) and yields (Huang et al., 2018). A detailed analysis of the applications of CRISPR in crops of tropical origin for better adaptation to current environmental conditions and market needs has recently been made (Rojas-Vasquez and Gatica-Arias, 2020) including regulatory environment in Africa (Tripathi et al., 2022). This article reviews the significance of genome editing tools in general, and the evolving CRISPR system and its applications for creating new precision breeding opportunities for important subsistence crops of the drylands.

Genome editing tools use different mechanisms for the recognition of target DNA. For example, while ZFNs and TALENS use the DNA-protein interactions, the CRISPR/Cas relies on DNA-RNA interactions (Gaj et al., 2013). The first generation SDNs such as ZFNs were used for editing a range of plant genomes (Lloyd et al., 2005; Kumar et al., 2015; Petolino, 2015), high skills and cost hindered its widespread applications (Sanjana et al., 2012). TALENs emerged as a relatively easy tool that still required sound molecular biology skills for construct preparation (Boch et al., 2009; Boch and Bonas, 2010; Li et al., 2011). However, in the last decade, CRISPR/Cas9 tools have been most efficient and successfully used for vast range of applications due to their low cost, effectiveness, and user friendliness, thereby providing attractive options for precision plant breeding (Hilscher et al., 2017; Ma et al., 2018; Oliva et al., 2019).

CRISPR/Cas has not only advanced at a very fast pace but has been efficient in simultaneous editing of several gene sequences (Zhang et al., 2016; Gao et al., 2017; Zaman et al., 2019; Li et al., 2021). The type II CRISPR/Cas systems hold an edge over other systems due to rapid continuous advancements with increased precision through numerous Cas protein variants including *dCas9*, *cas12a nickase79*, *fCas980*, *Cpf181* and other comparable nuclease systems While most of the Cas systems rely on the NHEJ DNA repair mechanism, the newly added tools such as base editing (Li G. et al., 2019) and prime editing (Anzalone et al., 2019) provide precise DNA base modification without induction of double strand breaks (DSBs). Although these have been largely exploited in mammals so far, they offer immense opportunities in agriculture and allied fields as well. Another factor influencing the editing efficiency of these tools is the delivery of editing components (Kang et al., 2020). While *Agrobacterium* transformation is the most stable delivery method for the development of edited plants (Sardesai and Subramanyam, 2018), several non-tissue culture-based delivery mechanisms have also been developed that can overcome the limitations of recalcitrancy in crops (Mahas et al., 2019). Moreover, CRISPR/Cas technologies together with rapid generation turnover (RGT)/speed breeding or double haploids are increasingly emerging to be more efficient in developing elite cultivars with safety, precision and speed (Hickey et al., 2019).

In the drylands, particularly in South Asian and Sub-Saharan African countries that have very limited cultivable arable land, little space is left for further crop expansion (Zdruli, 2014). This necessitates the broadening of crop diversity and reducing the burden on certain crops. Several neglected or underutilized crops are grown traditionally in their native environments and are

TABLE 1 Characteristic features of Cas variants and new genome editing tools.

Type of crop	Crop	Region grown	Area of production (m ha)	References
Cereals	Pearl millet (*Pennisetum glaucum*)	South Asia and Africa	34.79	Kumara Charyulu et al. (2014)
	Kodo millet (*Paspalum scrobiculatum*)	South-east and South Asia, Western Africa	0.2	Vetriventhan et al. (2020)
	Finger millet (*Eleusine coracana*)	Africa and South Asia	4–4.5	Vetriventhan et al. (2020)
	Foxtail millet (*Setaria italica*)	North America, Africa, South east and South Asia	0.72	Vetriventhan et al. (2020)
	Proso millet (*Panicum miliaceum*)	North America, Africa, South east and South Asia	1.37	Vetriventhan et al. (2020)
	Teff (Eragrostis tef)	South Africa, Australia, South East Asia and South Asia	NA	Vetriventhan et al. (2020)
	Fonio	Africa, South east Asia	0.96	Vetriventhan et al. (2020)
Legumes	Cowpea (*Vigna unguiculata*)	South Africa, West and central Africa	12.5	Ngalamu et al. (2015)
	Pigeonpea (*Cajanus cajan*)	South Asia, South Africa, West and central Africa, South East Asia	4.6	Saxena et al. (2015)
	Tepary Bean (*Phaseolus acutifolius*)	Sub-Saharan Africa, North America	NA	Jiri et al. (2017)
	Bambara groundnut (*Vigna subterranea* L. Verdc)	Sub-Saharan Africa and South Asia	0.25	Majola et al. (2021)

Table: Major underutilized crops and their area covered under dry regions in the world.

more adapted to marginal farming, that often have high nutritional value with rich genetic diversity (Jacob et al., 2018). The potential of underutilized crop is increasingly being recognized due to their superior trait qualities such as tolerance to biotic and abiotic stress for sustainable agriculture.

Extending the CRISPR/Cas toolbox for genome editing applications

The CRISPR/Cas protein endonuclease originates from several bacterial species including *Staphylococcus aureus*, *Streptococcus thermophilus*, *Francisella novocida* out of which *Streptococcus pyogenes* is the most widely used source. The SpCRISPR/Cas9 system predominantly recognizes protospacer adjacent motif (PAM) (5′-NGG-3′) and cleaves target DNA just three to four bases upstream of a PAM sequence to create blunt-end DSBs (Cong et al., 2013; Zhang et al., 2021). Several alternative Cas variants such as *dCas9*, *CRISPRi*, *iCas9*, *nickase79*, *fCas980*, *Cpf181*, *C2C2*, *13B*, *Cpf1*, etc. and other comparable nuclease systems have also been developed (Shmakov et al., 2017; Ghorbani et al., 2021). These Cas variants not only offer reduced off-target effects, but also provide higher precision in genome editing applications (Konermann et al., 2018). Among all Cas proteins, the type VI system has relatively simple and exclusive targets for RNA editing (Abudayyeh et al., 2017; Cox et al., 2017; Yan et al., 2018; Burmistrz et al., 2020). Faster customization of Cas variants has increased the target recognition capabilities resulting in multi-fold increase in precision and significantly lowered the off-target effects (Table 1).

Based on the constitution of effector protein, the CRISPR/Cas system is broadly classified into two major classes that have been further divided into six types (I–VI) and 33 sub-types (Makarova et al., 2020). While in the class I system (types I, III, and IV) the effector consists of multiple proteins, the class II system (II, V, and VI) compromises of a single effector with CRISPR RNAs (crRNAs) (Koonin et al., 2017; Chen et al., 2019). The class II system have been shown to have more flexible applications in inducing the sequence variations such as knock-ins, knockouts, exchange, genetic screening, imaging etc. (Tang and Fu, 2018). Within the class II, the CRISPR/Cas9 system has shown tremendous practical applications over others in a range of plant species from model systems to crop plants for efficient introduction of various traits such as disease resistance (Oliva et al., 2019), nutrition (Ku and Ha, 2020) and climate-resilience (Tripathi et al., 2019).

Base editing

While the evolution of CRISPR as a tool is remarkable and each shortcoming has been overcome with even more novel editing technologies, achieving precise single base DNA editing is an arduous task. Recently developed editing tools such as base editing (BE) ensure precise single base changes without the involvement of DSBs, HDR and donor DNA templates for selected irreversible nucleotide base substitutions at target

sites. Technically, the base editing system mainly comprises of catalytically impaired Cas protein, guide RNA, and a nucleobase deaminase domain.

Continuous advances in base editing tool offers improved editing efficiency or specificity or both by adding the base-edit repair inhibitor, a glycosylase inhibitor, to the fusion protein and modifying the Cas proteins (Marx, 2018). Base editing has several advantages over the existing CRISPR/Cas technologies and has been successfully carried out in several plant species. These include rice, wheat, maize, potato, watermelon, cotton, tomato, and *Arabidopsis* genomes (Chen et al., 2017; Hess et al., 2017; Yang et al., 2017; Zong et al., 2017; Tian et al., 2018; Monsur et al., 2020; Tra et al., 2021) for various traits including nitrogen use efficiency (Lu and Zhu, 2017) and herbicide resistance (Shimatani et al., 2017; Li C. et al., 2018; Monsur et al., 2020). Base editors also offer the disruption of genes by creating early stop codons or inducing transcript mis-splicing in plants (Veillet et al., 2019). While the ongoing base editor endeavors are constantly being improved to adjust to a wide range of crops (Mishra et al., 2020; Sretenovic et al., 2021), using appropriate Cas variants along with CBE and ABE base editors could broaden the horizon for crop improvement besides lowering the off-target effects.

Prime editing

Several genome editing tools encounter limitations with respect to the precision and utilization of the modified customized sequence simultaneously at the target site and perform single/few base substitutions. To overcome this, "prime editing" (PE) method is a "search-and-replace" system can alter the new genetic information directly at the targeted site without any DSBs or template DNA (Anzalone et al., 2019). Prime editors could efficiently develop all possible base conversions and small indels in a wider targeting range with limited off-target efficiency (Anzalone et al., 2019) and hold great promise for precision crop breeding. The PE components have been optimized to increase their efficiency and deployed in wheat and rice to generate several types of single base substitutions, multiple base substitutions and indels (Lin et al., 2020).

While the prime editing is less efficient than base editing for generating transition point mutations in plants, it generates transversion changes and all single base substitutions that cannot be made with other genome editing tools (Marzec and Hensel, 2020) that are important for applications in a range of crops. So far, most of the success has been achieved in monocots (Li and Xia, 2020a; Hua et al., 2020; Xu et al., 2020; Ren et al., 2021), with tomato as an exception among the dicots where editing of three genes *ALS2*, and *PDS1* was achieved at a frequency of 6.7% and 3.4%, respectively (Lu et al., 2021).

Epigenomic editing

Since the molecular basis of crop improvements is governed by both genome and epigenome of the plant (Kakoulodou et al., 2021), it is important to integrate them for realizing incremental genetic gains for improved adaptations and sustainable agriculture. CRISPR-based technologies are facilitating accelerated precision breeding, and epigenome editing is the next step in this direction to fast track the breeding process without the risk of genome instability and off-target effects. Since epigenetic modifications such as DNA methylation and histone modifications affect the expression of genes, with emerging knowledge on the functioning of epigenetics in plants, several efforts are ongoing for developing tools and technologies, thereby targeting epigenetic modifications that cause heritable changes. Epigenetic changes like modulation of chromatin, histone, cell differentiation, development and senescence have been shown to be involved in ensuring the survival of plants under stressful environments by enabling the plans to remember past stress events and dealing with these in the future, often referred to as "plant stress memory" (Singh et al., 2021; Sun et al., 2021; Shin et al., 2022). For example, methylation of histone H3 lysine 4 (H3K4) is involved in the persistent expression not only of high temperature-responsive genes, but also as hyper-induction of such genes during repeated heat stress treatments (Lämke et al., 2016). An inducible system for epigenome editing has recently been reported in *Arabidopsis* that uses a heat-inducible dCas9 to target a JUMONJI (JMJ) histone H3 lysine 4 (H3K4) demethylase domain to a locus of interest, the *APX2* gene in this case that showed transcriptional memory after heat stress (Oberkofler and Bäurle, 2022). Such newer tools enable targeted manipulation of epigenetic characters that could be used to specifically modify plant phenotype or to elucidate the relationship between the epigenome and transcriptional control (Hilton et al., 2015; Moradpour and Abdulah, 2020).

Emerging tools such as epigenetic QTLs or epigenetic single nucleotide polymorphisms tools also offer opportunities for activating or repressing candidate gene(s) or pathway(s) for trait improvement in crops, which could lead to the development of a new, efficient, and transgene-free breeding methods (Bilichak and Kovalchuk, 2016). While these new technological advances have shown the possibility of exploitation of epigenetic variation in crop breeding and acceleration and more efficient creation of climate-smart crop varieties, more work is needed in species beyond the model plant systems to gain a more comprehensive understanding of the mechanisms inducing and stabilizing epigenetic variation. In context of the underutilized crop plants, further studies are needed for identifying the specific traits and the association of stress-induced gene expression changes with alterations in DNA methylation and histone modifications, the mode of inheritance of these modifications, and their adaptive value (Chinnusamy and Zhu, 2009). In one such effort (Veley et al., 2021)

demonstrated that methylation to the TAL20 effector binding element within the *MeSWEET10a* promoter in cassava using synthetic zinc-finger DNA binding domain prevented TAL20 binding, blocking the transcriptional activation of *MeSWEET10a* displaying increased resistance to cassava bacterial blight (CBB). This offers potential opportunities for editing crop epialleles for adaptation traits. Nonetheless, this will require combined and multidisciplinary efforts in different areas of plant science and better integration of epigenomic data obtained in different crops.

Delivering genome editing components

The indirect and direct methods of delivery of genome editing reagents have been extensively and successfully used in several crops. Direct delivery methods such as *Agrobacterium* or particle bombardment possess persistent challenges in crops where efficient transformation systems are not available. While these plant transformation methods are cost-effective, convenient, and easily available in laboratories, their delivery efficiencies remain highly dependent on several factors such as type of explant, *Agrobacterium* strain, genotype, construction of independent editing reagents in multiple binary vectors etc. To overcome these drawbacks, direct methods of gene delivery, such as protoplast transfection, virus-mediated, RNP-based, meristem induction, lipofection-and PEG-mediated protoplasts and usage of aiding elements such as special peptides and nanoparticles have been developed and adopted (Ran et al., 2017; Nishizawa-Yokoi and Toki, 2021).

ZFNs, TALENs, and CRISPR have been successfully employed for gene knockout studies using protoplast transfection in different crops by using polyethylene glycol (PEG). However, the primary disadvantage of this technique lies in its inability to transform all plant species, especially monocots. While DNA-free genome editing methods have been obtained by delivering CRISPR/Cas reagents as *in vitro* transcripts or ribonucleoproteins (RNPs), these can give rise to multiple copies of the same gene causing undesirable altered expression (Liang et al., 2017).

More recently, nanoparticles (NPs) have gained pace as delivery vehicles since they can be designed according to the type of tissue and organism of interest (Ahmar et al., 2021). For example, mesoporous silica nanoparticles (MSNs) have been used to deliver the Cre recombinase into maize cells, leading to recombination of *lox* P sites and *DsRed2* expression (Martin-Ortigosa et al., 2014). NPs are highly stable and are flexible in terms of size, shape and distribution as carriers. Like NPs, polymers have been exploited as carriers due to their wide availability. Encapsulation or complexation with polymers, both synthetic and natural, can protect the components from enzymatic degradation and functionally activate them to bind to specific receptors for enhanced targeting. Additionally, lipid molecules have been effectively employed as delivery vehicles. Lipofectamine, a popular, commercial lipid reagent, has been utilized to deliver gene-editing proteins (Zuris et al., 2015).

Cas9 RNPs, containing negatively charged gRNA molecules quickly form a complex with cationic lipids. Nucleic acids have been exploited to function as polymeric substrates for Cas9 RNP delivery. However, further modification in the polymeric coating is vital to ensure that degradation through cellular pathways and enzymes does not occur. An alternative to encapsulation is the modification of the protein and nucleic acid. Cell-penetrating peptides (CPPs) are short peptide sequences that can penetrate the cell membrane easily. They can be conjugated with Cas9 protein and gRNA for enhanced delivery. However, these peptides do not protect the protein from protease degradation within the cell and can be complexed with other delivery methods. Further, nuclear localization sequences (NLSs) which are sequences synthesized in the cytoplasm for tagging proteins and transported into the nucleus are also being explored. They are poly-arginine/lysine and behave as signal molecules attached to proteins for nuclear transport. As Cas9 needs to be transported into the nucleus, NLS are excellent agents for delivery by synthesizing proteins containing NLS or encoding into the Cas9 construct (Glass et al., 2018). While most of these methods are still prevalent, they possess shortcomings, making them inadequate for efficient editing.

Tissue-culture-based techniques require plants to regenerate from transformed cells/explants which makes the procedure highly time consuming. Also, transformation protocols are genotype-dependent and effective protocols are not established for recalcitrant crop species. Therefore, new techniques have emerged that eliminate the need for traditional tissue culture techniques and *in planta* methods like floral dipping (Ji et al., 2020) and anther culture (Han et al., 2021). The gene-edited somatic cells are re-programmed into meristematic cells by expressions of developmental regulator (DR) genes, such as *WUSCHEL2(Wus2)* and *BABY BOOM* by using the genome-editing machinery. It has been demonstrated that the ectopic expression of DRs like *Wus2*, *SHOOT MERISTEMLESS(STM)* or *MONOPTEROS (MP)* induces the development of meristem-like structures in *Arabidopsis.* Additionally, the co-expression of Wus/STM and CRISPR/Cas9 cassette in *Nicotiana benthamiana* to target *phytoene desaturase* (*PDS*) gene was also carried out. The meristems later develop into shoots (Maher et al., 2020) demonstrating comparable mutation frequencies.

In another study, abundant shoots were successfully obtained through *in planta* transformation protocol in tobacco, where CRISPR/Cas9 expressing plants growing in soil were injected with *Agrobacterium* cultures carrying appropriate DR and sgRNA in the sites where meristems were removed. This study demonstrated altered development of edited somatic cells, induction of meristems and their growth in fertile plants

through co-expression of DRs and CRISPR/Cas9 system. Similar efforts have been made with editing the commercial varieties of wheat (Liu et al., 2021) that completely circumvented the need for tissue culture procedures to obtain genome-edited plants. The exclusion of tissue culture-based genome editing reduces cost, labor and amplifies efficiency. *De novo* meristem induction-mediated genome editing is still quite novel in terms of research. Additionally, abnormal growth has also been observed due to the constitutive expression of DRs. This can be overcome by inducible expression of the DRs. While the *de novo* shoot meristem induction has been exploited in grape, potato, and tomato crops, their feasibility in crops with heritable mutations is yet to be explored. This utility of the technique also needs to be extrapolated to staple food crops (Ji et al., 2020; Chennakesavulu et al., 2021).

Tissue culture-free genome editing technique has also been studied in meristem tissues developed from imbibed embryos of wheat seed (Hamada et al., 2018). The infection of meristematic tissues by virus (TRV) expressing SpCas9 protein was reported in tobacco wherein, the gRNA was fused with *Arabidopsis Flowering Locus T (FT)* mRNA, resulting in mutation of up to 65%–100% in the edited plants. Further, TRV was not detected in the progeny, subsequently protecting the progenies from any viral effects. This *in planta* gene-editing technique showed tremendous promise as it successfully generated small mutations in the gene. However, the method exhibited certain shortcomings in terms of identifying species-specific effective viral vectors and the gRNA-FT translocation abilities, highlighting the need for identification and characterization of viruses that infect meristematic tissues (Ellison et al., 2020; Kim et al., 2020).

Genome editing for underutilized dryland crops: Progress and prospects

The drylands make up to 41% of the total global land area and are characterized by low precipitation and drought. Comprising of mainly parts of western, central and southern United States, East and west Africa, the middle east, parts of Indian subcontinent and the central deserts of Australia, these areas are prone to environmental factors such as wind erosion, mineral weathering, and low fertility (Hyman et al., 2016). The climate change has made imminent the need for adapting climate-smart crops as well as growing more resilient underutilized having low water requirements. A comprehensive overview of the major crops grown in drylands and their area of production is given in Table 1.

To mitigate the challenges of agricultural productivity in the underdeveloped and developing countries and to make agriculture sustainable under diverse climatic conditions, it is critical to develop transformative strategies for breeding pipelines by using the new breeding acceleration techniques. While tweaking selection accuracy and intensity can lead to minor improvements in a breeding context, generating novel and useful genetic variations and rapidly fixing the traits facilitate crop genetic gains can allow faster turnover of improved cultivars for accelerated delivery of improved varieties to farmers. The availability of reference genomes and ever-increasing re-sequencing data has significantly advanced breeding applications and allowed to capture the genomic diversity and its effective mining. This revelation has helped in understanding genes and the mechanisms underlying various biotic and abiotic stress responsiveness, quality, besides nutrition and plant architecture parameters, thereby aiding considerably in developing crop species with adaptive and resilient traits.

Genetic resource collections that are deemed to harbor a wealth of undisclosed allelic variants are being unlocked by identifying allelic variation of relevant traits within these collections. The enormous genetic diversity present in wild species or landraces of crops as a source of allele-mining could very well be utilized and translated to elite backgrounds using genome editing tools, thereby potentially expanding the crop germplasm pool. Optimization of these tools in the underutilized food crops like sorghum, millets, groundnut, beans, cowpea, teff, banana, cassava etc. that are primarily cultivated by the poor and resource-poor farmers of the drylands would lead to huge impact in achieving the global food and nutritional security goals. This will not only accelerate the pace of the ongoing research but will potentially enable a disruptive reduction in cost for development of both farmer- and consumer-centric traits/products in these important crops. A comprehensive overview of the potential traits which can be explored in underutilized crops using genome editing has been given in Figure 1. Some of the important subsistence crops and traits of the drylands that could potentially be addressed through genome editing methods are discussed below.

For success with precision breeding, successful genetic transformation of underutilized crops is one of the prerequisites for delivery of recombinant DNAs as well as genome editing components into the plant cells that regenerate into whole plants. While *Agrobacterium* transformation has been successfully developed for almost all staples, there has not been a great deal of progress in improving the transformation frequencies for a many underutilized crops. Currently, transformation competent methods have been developed for crops such as finger millet (Ceasar and Ignacimuthu, 2011; Hema et al., 2014); foxtail millet (Ceasar et al., 2017; Santos et al., 2020) and pigeonpea (Sharma et al., 2006; Ghosh et al., 2017). While genetic transformation of several underutilized legumes is still in its infancy, stable and reproducible transformation system based on callus derived from floral buds and cotyledonary node region is available for tepary bean (Dillen et al., 1997; De Clercq et al., 2002; Zambre et al., 2005). In bambara groundnut, an efficient system for *in vitro* shoot induction from cotyledons derived from mature

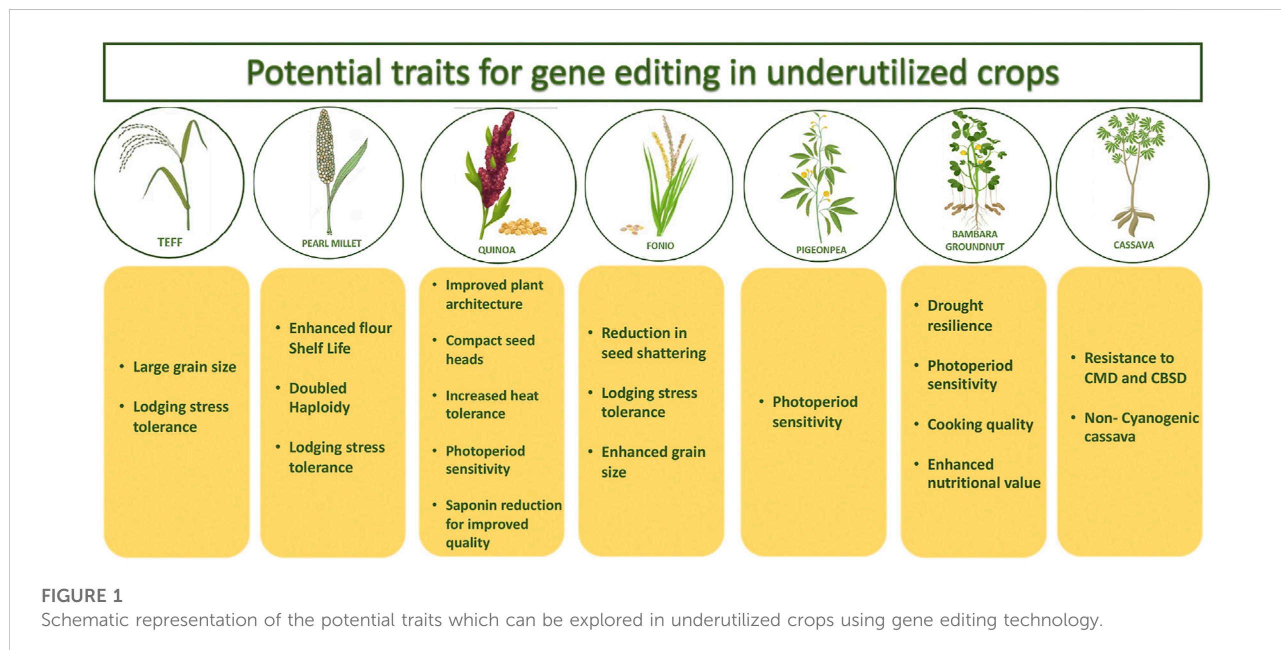


FIGURE 1
Schematic representation of the potential traits which can be explored in underutilized crops using gene editing technology.

seeds has been established to subsequently exploit transformation technologies in this important legume (Koné et al., 2011). In cassava, transformation systems have been developed and much progress has been made in the development of *Agrobacterium*-based transformation protocols (reviewed by Liu et al., 2011). Similarly, in quinoa, a rapid transformation system was established using hairy roots obtained from cotyledon-nod with hypocotyl, cotyledons and hypocotyl pieces at a transformation efficiency of 32%–68% (Wang et al., 2021). Concerning pearl millet, while several reports showed transient expression of the reporter genes in transformed calli (Ramineni et al., 2014), barring a report of Ignacimuthu and Kannan (2013), not many stable transformation methods have been reported. Owing to the inefficiencies and inconsistencies in the published protocols for several crops, several non-tissue culture-based approaches are being optimized for transformation that do not depend on the regeneration of adventitious shoot buds (Martins et al., 2015).

Green revolution traits-millets

The green revolution (GR) evolved from specific requirements in nutrition and yield productivity primarily enabled by vast genetic resources of the gene banks. The transfer or replacement of dwarfing genes into cultivated crops such as rice and wheat resulted in shorter straws allowing diversion of more nutrients into grain, besides making heavier ears that allowed higher yields and better agronomic performance (Gale and Youssefian, 1985). The GR traits can be exploited for teff (*Eragrostis tef*) and finger millets that are known to flourish and grow well in East African climatic and soil conditions (Tadele and Assefa, 2012), where lodging leads to a considerable loss in their harvest. In rice, several quantitative trait loci (QTLs) that contribute to lodging stress tolerance have been identified and successfully integrated into the development of improved varieties (Liu et al., 2018). Similarly, the revolutionary gene, *sd-1*, that encodes for *gibberellin-20 oxidase*, provided rice varieties with lodging resistance without affecting the grain quality (Monna et al., 2002; Sasaki et al., 2002; Spielmeyer et al., 2002). Similarly, in wheat *Rht-B1* (*reduced height-B1*) and *Rht-D1* genes imparted lodging tolerance via dwarf plant development (Würschum et al., 2017). In maize, a close homolog of *dw3*, *Br2*, was identified which is an ATP-binding cassette-type B1 (ABCB1) auxin efflux transporter (Hilley et al., 2017).

The dwarfing trait in sorghum has been bred using *dw* (*1–4*) genes (Multani et al., 2003). The possibility of CRISPR/Cas9-mediated targeted gene modification has been demonstrated to be efficient in sorghum (Jiang et al., 2013b; Char et al., 2020). Editing of an alpha-Kafirin gene family that form protein bodies with poor digestibility was shown to increase digestibility and protein quality in sorghum grains following the CRISPR/Cas9 approach (Li A. et al., 2018) The diploid genome of foxtail millet (*Setaria italica*) has recently been sequenced and annotated (Bennetzen et al., 2012; Jia et al., 2013) that could serve as a model system for C4 plants. More recently, Cheng et al. (2021) have reported the induction of haploid embryos through seed by CRISPR-Cas9 mediated mutagenesis of the *SiMTL* gene that is orthologous to the maize MATRILINEAL/NOT-LIKE-DAD/PHOSPHOLIPASE A (*MTL/NLDZmPLA*) gene that generated haploids in maize (Liu et al., 2017). This study

paves ways for utilization of the double haploid breeding for enhancing genetic gains in dryland cereal crops.

The identification of height related and root architecture genes in the dryland cereal crops provide a foundation for evolutionary and functional analysis of specific proteins defining a comprehensive view of *Rht*, *dw3* or *Br2* family genes (Zhu et al., 2012; Zanke et al., 2014; Wei et al., 2018) in nutri-cereals such as millets and teff. For example, double knockout maize mutants of *ZmPHYC1* and *ZmPHYC2* created using the CRISPR/Cas9 technology displayed a moderate early-flowering phenotype under long-day conditions, while the overexpression of *ZmPHYC2* exhibit a moderately reduced plant height and ear height (Li et al., 2020b). A recent review summarizes genome editing efforts on plant architectural phenotypes in cereals and their manipulation to optimize their architecture towards the concept of ideotype for crop improvement (Huang et al., 2021).

Cassava-disease resistance and quality traits

Cassava (*Manihot esculenta*) is a very important crop which is not only vital to food security in tropics and subtropics, but also a predominant raw material of starch industry (Zhou et al., 2015). Cassava, an important staple food, is grown globally for the calories, of which it provides up to 50% intake of calories (Bredeson et al., 2016) for over 800 million people worldwide (Prochnik et al., 2012). Grown in marginal environments and provides for one of the most important sources of carbohydrate globally, this gluten-free carbohydrate source has seen up to 60% increase in global harvest between 2000 and 2012. There is a continuing need to improve the yields and adaptation of elite cassava varieties (Bull et al., 2018).

Cassava encounters some of the most devastating diseases caused by brown streak virus and cassava mosaic virus causing up to 50% crop yield losses (López and Bernal, 2012). Cassava mosaic virus disease (CMD) is caused by three innate types of Gemini virus; *CMD1*, known to be recessive and governed by multiple genes (Rabbi et al., 2014), *CMD2* possesses a single dominant locus on chromosome number 12 (Akano et al., 2002; Rabbi et al., 2014; Wolfe et al., 2016) and *CMD3* contains a QTL conferring resistance (Houngue et al., 2019). The development of resistant cultivars using somatic embryogenesis in CMD1 was ineffective due to loss of resistance in subsequent generations (Beyene et al., 2016). Hence, *CMD2* and *CMD3* could be the potential candidates for further exploitation via CRISPR/Cas-mediated site-specific targeting (Baltes et al., 2015; Bart and Taylor, 2017). Similarly, simultaneous CRISPR/Cas9-mediated editing of two isoforms of host translation factors, nCBP-1 and nCBP-2 conferred significant resistance to Cassava brown streak disease (CBSD) (Gomez et al., 2019). Suppression of interaction of viral genome-linked protein (VPg) with mutant alleles ncbp-1, ncbp-2, and ncbp-1/ncbp-2 resulted in delayed and attenuated CBSD aerial symptoms, as well as reduced severity and incidence of storage root necrosis.

In addition to disease resistance traits, herbicide tolerance was achieved in cassava by deploying HR and NHEJ DNA repair pathways (Hummel et al., 2018). For quality traits, efforts have been made to improve the quality of its starch, for developing suitable starch properties for cooking and processing. CRISPR-Cas9 mediated targeted mutagenesis of two genes protein targeting to starch (*PTST1*) and granule bound starch synthase (*GBSS*, involved in amylose biosynthesis), have been reduce amylose content in cassava root starch (Bull et al., 2018). In addition to improving the quality of starch, several research groups have been making efforts to develop a cynogenic-free cassava by using gene editing approaches for blocking the production of cyanide. Cassava contains potentially toxic levels of cyanogenic glycosides (Linamarin and Lotaustralin) which if not efficiently removed through processing, may cause various neurological disorders and in some cases may be fatal. The biosynthetic pathway of cyanide in cassava was already well understood and *CYP79D1/D2* gene that encode two cytochrome P450s catalyze the first-dedicated step in cyanogenic glycoside synthesis. Selective inhibition of this gene by antisense expression in leaves and roots have demonstrated a 99% reduction in root cyanogen levels providing road map for using genome editing methods for complete knockdown (Otun et al., 2022).

Grain size and plant architecture traits in teff and fonio

Another set of dryland crops including teff (*Eragrostis tef*) and fonio (*Digitaria* sp.) which despite their applications in food and feed, high nutrient content and high durability, are among the most under-utilized crop species in the African region (Lee, 2018). Teff is considered as “risk crop” due to its high adaptivity even under extreme conditions of drought and waterlogging and is now in high demand as a forage crop (Miller, 2007). Fonio, on the other hand is considered as the “*grain of life*” and is known for its high nutrient content and contains all 20 amino acids including methionine and cysteine (NRC, 1996; Taylor, 2017).

Mining the homologs of rice genes associated with grain size and weight (Li et al., 2011) in teff could be an effective way of achieving larger grain size in this nutritious cereal and will be a crucial step towards their genetic improvement (Valentine et al., 2017). Fonio shares a close synteny with sorghum, and mutations in genes such as DeSh1-9A, that have shown partial selective sweep but reduced seed shattering in sorghum, can also result in another beneficial architectural trait (Abrouk et al., 2020).

Improving the plant architecture of these underutilized crops is a major breeding goal towards the concept of “ideotype for crop improvement” (Huang et al., 2021). Green revolution saw

transfer or replacement of dwarfing genes into cultivated crops such as rice and wheat resulted in shorter straws allowing diversion of more nutrients into grain, besides making heavier ears that allowed higher yields and better agronomic performance (Gale and Youssefian, 1985). There is a vast potential to exploit GR traits for dryland cereals such as teff and finger millets, the major staples of east Africa (Tadele and Assefa, 2012), where lodging leads to a considerable loss in their harvest. Several quantitative trait loci (QTLs) that contribute to lodging stress tolerance have been identified and successfully integrated into breeding programs for improved rice and wheat varieties (Monna et al., 2002; Sasaki et al., 2002; Spielmeyer et al., 2002; Würschum et al., 2017; Liu et al., 2018). Genetic variations in *sd1* and *RHT* genes (Peng et al., 1999) have shown significant lodging tolerance in these major staples and translating these to crops like teff and fonio offer tremendous opportunities. More recently, Beyene et al. (2022) created CRISPR-induced knockout mutations in the tef orthologue of the rice *SEMIDWARF-1* (*SD-1*) gene that conferred semi-dwarfism and significantly higher resistance to lodging resistance in tef. Similarly, homologs of *OsSPL14* (squamosa promoter binding protein-like 14) gene and microRNA "*OsmiR397*" that have been reported to confer panicle branching trait in rice has potential to be explored and targeted in teff and fonio (Miura et al., 2010; Zhang et al., 2013; Numan et al., 2021).

In addition, the grain size of teff and fonio has been a major cause of reduced yield where not much progress has been made in terms of hybrid development through conventional breeding. However, since this trait is being extensively explored in other staple cereals such as maize and rice and with the availability of annotated genomic sequences of sorghum and foxtail millet there are emerging opportunities to identify candidate genes that might share genomic synteny with teff and fonio (Saha et al., 2016; Ayenan et al., 2018). Functional analysis and identification of homologs of these genes in teff and fonio will further help to form a basis for developing lines with enhanced grain size.

Quality traits-quinoa and pearl millet

Quinoa (*Chenopodium quinoa*), a pseudo cereal belonging to *Amaranthus* family originated in the Andean region grows in the marginal lands is one of the best food choices due to its balanced amino acid profile, vitamins, minerals, ions, and antioxidants, quinoa received a "superfood" status and contributes to the economic and global nutritional security (Vega-Gálvez et al., 2010). However, despite being nutri-climate-resilient, it is still an underutilized crop with major breeding objectives including, improved plant architecture, compact seed heads, increased heat tolerance, photoperiod and heat sensitivity. A well-annotated and high-quality reference genome sequence has recently been made available (Jarvis et al., 2017), thereby offering opportunities for allele mining for trait prospecting efforts. However, precision breeding in this crop requires establishing genome-scale engineering platforms and toolkits to understand gene functions and their interactions.

Quinoa seeds contain a mixture of triterpene glycosides called saponins that contribute to plant growth to a certain extent. However, this anti-nutritional property must be removed prior to human consumption as these saponins cause hemolysis in humans and a bitter flavor that are undesirable traits. Reducing or eliminating the saponin through physical and traditional approaches is costly and often water-intensive and negatively affects the quality of nutritional elements. Identification of the candidate genes and their genetic variations underlying the saponin biosynthetic pathway have been investigated in different germplasms of quinoa with the help of existing sequencing data (Jarvis et al., 2017) that would provide a platform for further studies in the generation of genotypes with sweetness and low saponin and their introgression into commercial varieties.

While CRISPR/Cas tool provides a robust platform for targeted quinoa breeding, the lack of an efficient transformation system in quinoa would be another objective in developing the next generation quinoa plants. Genetic transformation methods, including *Agrobacterium*-mediated transformation, hairy root and leaf agroinfiltration techniques have been used for quinoa (Wang et al., 2021). However, the transformation efficiency at this stage may not be sufficient for any meaningful genetic engineering and genome editing strategies. Nevertheless, the possibility of generation of transformed quinoa plants through *Agrobacterium*-mediated transformation is not a vague reality. This technique can be improved further by using the booster genes such as *WUSCHEL*, *BABY BOOM* and *LEAFY COTYLEDON1* which have been previously known to improve the transformation efficiency in other crops like maize and sorghum (Nelson-Vasilchik et al., 2022). The *de novo* induction of meristems could also be an alternative approach along with the expression of booster genes to avoid complications in tissue-culture strategies. Other challenge in quinoa genome editing could be due to its allotetraploid nature, where targeting all four copies of these genes could be challenging. Multiplex genome editing would be an ideal solution in this scenario as it has been carried out in other polyploidy crops such as wheat, canola, sugarcane, and banana (Vats et al., 2019).

In addition to quinoa, another food and nutritional security crop, pearl millet (*Pennisetum glaucum*) that grows in some of the most hostile-to-farm landscapes despite its many superior attributes, has an unsolved quality issue of flour rancidity, posing a hindrance to its wider acceptability. Rapid development of off-flavor in pearl millet flour within 5–7 days of milling hinders the commercial use of this crop besides creating additional drudgery for women of the household, as the amount that can be pounded is limited to a few days of household use, thereby necessitating that the grain be milled immediately prior to use. A collaborative

effort between the CGIAR and industry outlined a direct mechanism for hydrolytic and oxidative rancidity in millet flour, allelic variation two candidate lipase genes, Pg*TAGLip1* and Pg*TAGLip2* were identified, that correlated with the rancidity profile, confirming their function. Mutations in these key TAG lipases in pearl millet have potential in protection of lipids from TAG hydrolysis and fatty acid oxidation, leading to a reduction in off-flavour volatiles (Aher et al., 2022). In addition, since pearl millet has abundance of unsaturated fatty acids (>78%) representing the reactive center that produces odor-active volatiles, major markers for lipid oxidation (Sharma, 2015). Hence, shifting the fatty acid profile in pearl millet from poly unsaturated fatty acids (PUFA) to monounsaturated ones (MUFA) by generating inactive or partially active *Fad2* alleles, will serve to not only increase the shelf life but also deliver health dividends because of the positive health benefits of the monounsaturated fatty acids.

Abiotic stress component traits-Bambara groundnut (*Vigna subterranea*)

Bambara groundnut is an underutilized legume found mainly in the African sub-continent. Due to its high content of complex carbohydrates, unsaturated fatty-acids, minerals such as magnesium, iron, zinc and potassium, fiber, and plant-based proteins, it holds the potential for providing food security through a sustainable approach, especially in the arid and semi-arid region (Olanrewaju et al., 2022). However due to lack of knowledge, appropriate policies and resource limitation, Bambara groundnut is often overlooked and therefore, is categorized as an underutilized crop (Travella et al., 2019).

The first genome sequence of the bambara groundnut was released by Chang et al. (2018) which opened avenues for improvement of the crop through genetic approaches. Major traits of importance in Bambara groundnut are drought-resilience, photoperiod response, cooking quality and time, and nutritional value (Muhammad et al., 2020). Along with this, pipelines of other crops have also been utilized to develop translational frameworks are being used to provide gene orthologues in this legume crop (Popoola et al., 2019). For example, massively parallel signature sequencing (MPSS) strategy employed for expression profile analysis of Bambara groundnut under water-deficit conditions led to the revelation that major transcription factors like MYC, WRKY protein and DREB were absent in the dataset. A recent study assessed the genetic diversity and structure among Bambara groundnut landraces collected across South Africa and other regions in southern Africa using SSR markers for the cultivation and improvement of Bambara groundnut (Minnaar-Ontong et al., 2021). More recently, KUP genes have gained attention for their role in abiotic stress tolerance and hence offer opportunities for precision genetic interventions in Bambara groundnuts. This provides scope for further improvements and genome editing tool has potential to deploy these novel traits and aid precision breeding of Bambara groundnut (Paliwal et al., 2021).

Photoperiod sensitivity-pigeonpea (*Cajanus cajan* L.)

Pigeonpea is an important climate resilient annual legume grown in parts of Asia, Africa and Latin America grown with other legumes and cereals. Genetic studies on the essential traits of pigeonpea such as maturity, photosensitivity, breeding behavior and disease and pest resistance have implied that the major agronomic traits are mainly additive in nature. The first pigeonpea hybrid was developed in the 1990s based on cytoplasmic-male sterility-based breeding system. Advances in next-generation sequencing (NGS) has revolutionized GAB by facilitating development of markers for unique agronomic traits (Pazhamala et al., 2015; Singh et al., 2020) and have played a significant role in building breeding programs. However, modern technologies such as CRISPR/Cas9 based editing are integral for unravelling mechanisms of other important traits and enhancing pigeonpea program.

Being a short-day legume differential genotypic sensitivity to photoperiod has major implications in adaptation of pigeonpea with respect to latitude, altitude and season. Most of the traditionally grown pigeonpea cultivars and landraces are represented by varieties from the medium- and long-duration maturity groups that mature in 150–280 days. To expand pigeonpea cultivation into new crop improvement programs, the manipulation of flowering time is likely to contribute greatly to crop yields through tailoring of cultivars to specific climates or to changes in climate that are anticipated to occur. Certain SSRs and SNPs have been identified which shed light on the pleiotropic relationship between photosensitivity and flowering time (Bohra et al., 2020).

The manipulation of flowering time is likely to contribute greatly to crop yields through tailoring of cultivars to specific climates or to changes in climate that are anticipated to occur.

However, to accomplish this, an understanding of the genes associated with transition from photoperiodic sensitivity to photoperiodic insensitivity is required. Such knowledge can be used to develop pigeonpea germplasm that can be grown for yield gains under both long- and short-day conditions and provide sustainable production of grain legumes. A recent report provided detailed characterization of the genes involved in photoperiodic regulation of flowering in *C. cajan* offering clues to the role of PEBP (FT) family genes, based on genome-wide analyses and expression profiling. *CcFT6* and *CcFT8*, were identified as probable *Flowering locus T* genes that are responsible for the production of florigen in pigeonpea. While *CcFT6* upregulates under SD in photoperiod

sensitive, MAL3 genotype, *CcFT6* and *CcFT8* upregulate in photoperiod insensitive genotype (ICP20338) under SD and LD conditions, respectively. The presence of *CcFT8* as an additional florigen producing gene, having ability to flower in a photoperiod independent manner under LD conditions provide some clues on its photoperiod insensitive nature (Tribhuvan et al., 2020). More recently, two candidate genes coding for pentatricopeptide repeat (PPR) and cell division protein *FtsZ* homolog have been investigated in pigeonpea. These two candidate genes and previously reported genes such as *CcTFL1* and *EARLY FLOWERING3* (Saxena et al., 2017; Varshney et al., 2017) could be validated at a functional level for their specific roles. Tailoring of CcFT8 and other candidate FT genes, using genome editing has a potential to provide answers to the understanding molecular mechanisms associated with the trait. Moreover, precision targeting of the identified candidate genes involved in flowering, would play a crucial role in extending the cropping area of pigeonpea, a photoperiod sensitive major grain legume into new cropping systems.

Accelerating genetic gains through genome editing

Ex-situ collection of plant germplasm and its maintenance is crucial to protect the vast genetic diversity in crops that are fast deteriorating due to the development of domesticated cultivars over traditional landraces (Pérez-Jaramillo et al., 2016). Systematic phenotypic evaluation of the available resources would help researchers gain perspective about the underlying potential of these landraces. Genetic gains (Falconer and Mackay, 1996) in a species occurs when the frequency of desirable genes is increased usually achieved by selection of elite parental varieties based upon their phenotypic or genotypic characteristics. Since developing homozygous lines could take at least 10 years through conventional breeding, it alone will not be sufficient to bridge the gap between current level of crop production. Hence the rate of genetic gain has remained considerably low with time (Bhattacharya et al., 2021). Several strategies to create and unlock favourable genetic variations through molecular and genomic approaches including mutation, gene mapping and discovery, transgenics, and genome editing to enhance genetic gains in crops have been reviewed (Xu et al., 2017).

On the breeder's equation, the rate of genetic gain in any crop is related to the selection intensity that is applied, which is in turn related to the size of the breeding program, the accuracy of the data or the selections made. Essentially it is about the parental selections, genetic diversity, and the breeding cycle time. So, when it comes to breeding cycle time, many breeding programs have a cycle of more than 10 years and some could be even more up to 25 years. So, driving down, breeding cycle time on the denominator has really a massive impact on the genetic gains. That's when advanced breeding technologies that can improve the accuracy have the potential to create their greatest impact. Since the collective impact is greater than the sum of the parts, a synergistic integration of conventional breeding and gene editing approaches can deliver the highest possible rate of genetic gains bringing down the age of varieties in farmer's fields.

Additionally, changes in the zygote and germline cells would give rise to further heritable changes which are maintained across generations (Gonen et al., 2017). However, it was observed that to achieve persistent variance, one to two generations of editing were required due to the segregation of non-desirable alleles within the non-edited parents. Genome editing holds the potential for facilitating the identification of essential genetic variations and their deployment in breeding programs. Due to the availability of high-throughput screening technologies, the desirable phenotypes can now be identified and employed in pre-breeding strategies to obtain genetic variations. Such data allows the identification of core traits and sometimes in the discovery of specific genes that could aid in understanding relevant, novel and useful variations in elite varieties (Mascher et al., 2019). An interesting concept which has been proposed is the "re-domestication" of crops using CRISPR/Cas9 mediated knockouts. Such targeted gene modifications are being considered to induce genomic selection as well as transfer beneficial traits between domesticated crops and their wild varieties which otherwise is a time and labor-intensive process (Lemmon et al., 2018).

Additionally, quick domestication of annual crops is a real challenge because the crop would be sown each year in the same agricultural land, thereby deteriorating soil fertility that would eventually lead to lower nutrient and mineral uptake. Therefore, the domestication of perennial crops such as wheat could be a significant steppingstone towards achieving sustainable agricultural practices. Some unsuccessful attempts have been made to turn a wheat variety into a perennial crop by hybridizing with the wild varieties of some grasses. In such cases, the process can be accelerated by CRISPR/Cas tools by targeting the domestication of homologues genes for their successful knockout (Venske et al., 2019).

There is a need to explore the wider domestication opportunities for less researched and invested crops such as sweet potato, groundnut, cassava, teff, fonio, banana and quinoa, which are locally crucial for their extensive nutritional values. However, some undesirable characters such as lower grain yields, sprawling growth and fruit drop limit calls for a more comprehensive cultivation. Therefore, the demonstration to control plant architecture, flower production and grain size by

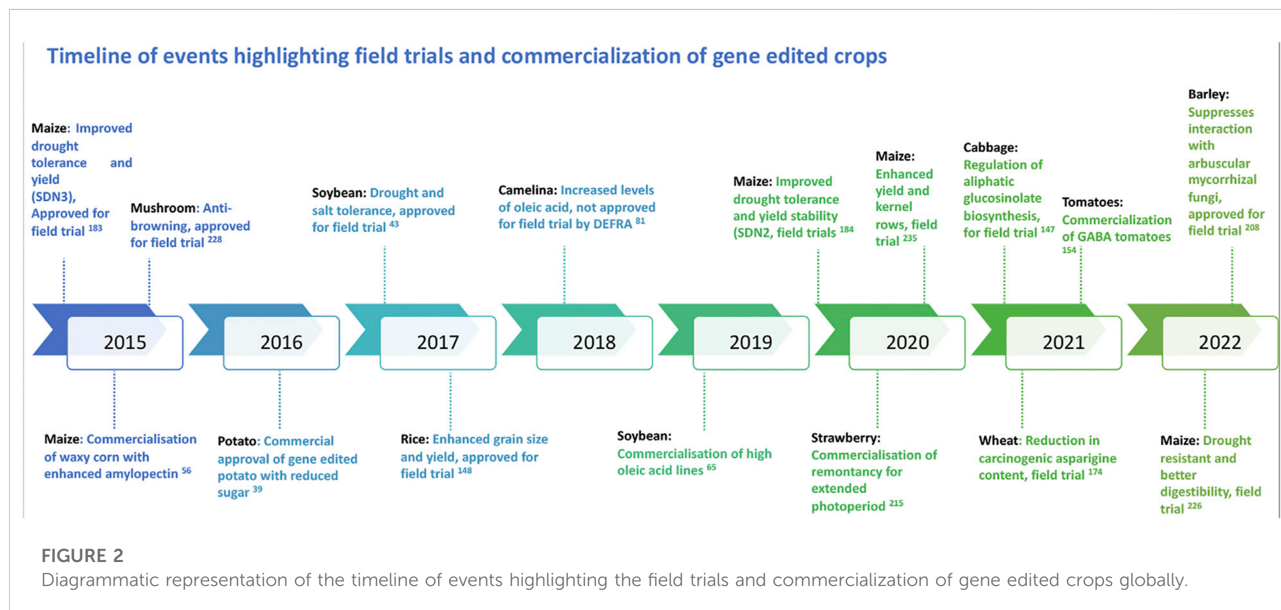


FIGURE 2
Diagrammatic representation of the timeline of events highlighting the field trials and commercialization of gene edited crops globally.

CRISPR/Cas technology in ground cherry, which is semi-domesticated orphan crop opens a wide array of applications of accelerating the genetic gains by editing multiple sites and modifying gene regulation (Lemmon et al., 2018; Zhang et al., 2020).

Current global regulations for genome edited crops

With the discovery of genome editing tools, a wide array of applications has been introduced and experimented in various organisms including viruses, bacteria, humans, animals and plants. In the light of recent developments in genome editing, product trials are ongoing in several crops across many countries and regions. While the application of genome editing for genetic gains and crop improvement has a highly potential, it is subject to immense societal resistance. As with any new technology, there are apprehensions around gene editing technologies. To fully comprehend the ethical debates and concerns on genome editing, it is important to understand the process and possible outcomes (Brokowski, 2018; Lassoued et al., 2021). Efficient science communication around the edited traits also may help in improved application and acceptability of these new breeding techniques. Regulatory policies of genome-edited plants in various countries adopt two major frameworks such as the process or the final product. Currently, while very few countries have developed the regulatory frameworks, a majority are yet to develop or declare their regulation process. Decision to either regulate or not to regulate the genome edited crops mainly depend on the type of regulatory system that already exists in a country. A recent review provides an update on the regulatory status of new breeding techniques and biosafety approaches in select countries (Obukosia et al., 2020). Nevertheless, with the evolving regulatory framework on genome editing, certain crops have surpassed the regulations to be now under field trial or on the road to commercialization (Figure 2). The “Am I Regulated” process of the USDA (now SECURE Rule’s Exemption and Confirmation Process beginning on 17 August 2020) allows for developers to determine whether their genetically modified or gene edited organism meets the regulations or not. With the introduction of this process, several inquiries have been submitted to the USDA for gene edited crops of specific traits, some of which have been duly approved (Figure 3).

The United States Department of Agriculture (USDA) declares that genome editing is almost equivalent to conventional breeding and therefore, does not require any regulatory process within the United States. Some speculations from USDA define that gene-edited plants can be considered as a separate category (Menz et al., 2020). In a recent development in the United States , ‘USDA APHIS’ announced the first comprehensive revision to 7 CFR part 340 which is referred to as “SECURE” rule to regulate biotechnology (USDA, 2020a). It provides three very important exemptions for single genetic modifications including products that would be categorized as SDN-1 or SDN-2 in terms of the outcomes of genome editing. The third exemption would also include the introduction of a gene that is known to occur in a plants’ gene pool or allele replacement. In Canada, the regulatory framework is based on the risk of the products comprising a policy of regulating the novelty of new traits in plants or the novel characteristics of new foods or livestock feeds. Hence, whether genome edited products

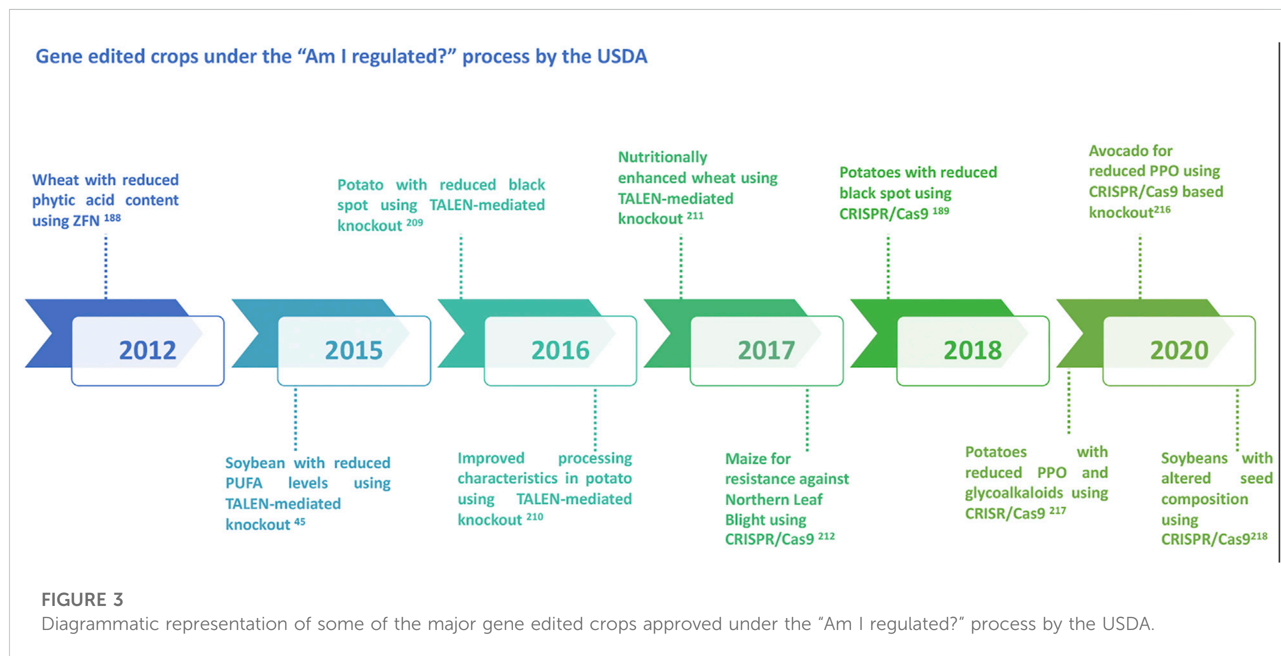


FIGURE 3
Diagrammatic representation of some of the major gene edited crops approved under the "Am I regulated?" process by the USDA.

will be regulated or not will depend upon the characteristics of the final product and not on the technology that was used (Smyth, 2017). Till date, two products including the non-browning apples and non-dark spot potatoes developed through gene editing have cleared the regulatory process in Canada.

Argentina's regulatory process is in accordance with the Cartagena Protocol on Biosafety and evaluated on case-by-case assessment, irrespective of the method used for product development (Lema, 2019). If the edited product is transgene free, the product is classified as non-transgenic and does not require any regulatory process. Countries like Chile, Brazil and Colombia follow the Argentinian model for their regulatory policies, evaluating such products on a case-by-case basis and exempting them from regulation when there is no insertion of a foreign gene (Tsuda et al., 2019).

While in Australia and New Zealand, products developed through CRISPR/Cas9 and other editing tools are excluded from the regulatory process (FSAN, 2018), the European Union (EU) countries follow unique regulatory process where the Court of Justice of the European Union (ECJ) has declared gene-edited crops as subject for stringent regulations as conventional genetically modified (GM) organisms (Laaninen, 2019). Amongst the Asian countries, Japan recently declared that foods derived from genome editing technologies which do not contain any foreign genes and/or fragments are not considered as GMOs and do not require any regulatory clearances (USDA, 2020b).

In India, a recent notification has exempted Site Directed Nuclease (SDN) 1 and 2 types (SDN1 and SDN2) of products which do not carry any vector DNA and are like the products of spontaneous or induced mutations from the transgenic regulation and risk assessment under Rules 1989. Guidelines for the safety assessment of genome edited plants 2022 have been released in May 2022, that define various categories of genome edited plants and determine regulatory requirement for appropriate category and provide the regulatory framework and scientific guidance on data requirement (DBT, 2022).

Similarly, Philippines has moved ahead with, a policy discussion paper under review and consideration on how products of new plant breeding technologies should be treated under existing regulatory regime, the benefits that may be derived and the capacity of the country to utilize such techniques. The policy framework will rely on a case-to-case and crop-to-crop based decision or regulatory pathways which will be the entry point of any genome edited plant products with or without involving the insertion of genes from non-sexually compatible species. However, in regions where the technology and infrastructure has not advanced enough, containment and monitoring measures are expected to be comparatively strict.

In South Africa, SDN-1 involving "small, targeted and untargeted inserts or deletions based on non-homologous end joining (NHEJ)" resulting from ZFNs, MNs, TALENs and CRISPR/Cas and considered to be exempt from GM Act (ASSAf, 2016). The regulatory guidelines for specific countries in Africa are at various stages of development (Travella et al., 2019; Obukosia et al., 2020).

While CRISPR/Cas9 is an inexpensive and flexible technology, international harmonization of the regulatory frameworks needs to be developed to ensure that these are based on sound science and the community of practices

developed around the world (El-Mounadi et al., 2020). More deliberative and worldwide conversation is expected to reexamine and rethink about the existing prohibitive rules and devise strategies to grow more logical and specialized overall models for genome editing applications in food and agriculture for betterment of farmers' livelihoods (Qaim, 2020; Beumer and Swart, 2021).

Conclusion and way forward

To achieve sustainable increase in the rate of genetic gains in food crops, transformative strategies for accelerated crop breeding pipelines need to be embraced. Several national agricultural research system (NARS) initiatives are ongoing under several major initiatives for Africa including Modernizing Ethiopian Research on Crop Improvement (MERCI). Accelerated Varietal Improvement and Seed Delivery of Legumes and Cereals in Africa (AVISA) (https://www.avisaproject.org/), Excellence in Breeding (EiB) Platform (https://excellenceinbreeding.org/), Crops to End Hunger (CtEH) (https://www.cgiar.org/excellence-breeding-platform/crops-to-end-hunger/) and various CGIAR Agri-food research programs (https://www.cgiar.org/research/research-portfolio/). These initiatives put major emphasis on modernizing breeding mainly through developing specific regional product profiles, mechanized operations and databases, besides focusing on infra-structure/human capacity development for efficient breeding and seed systems. Nonetheless, crops face intractable problems not easily solved by traditional breeding and hence there is a need for future breakthroughs in global agriculture. These ongoing breeding modernization agendas integrate innovations in advanced breeding tools (ABTs) such as CRISPR/Cas that are increasingly becoming relevant to fill gaps in the pipeline research required to deliver high yielding, nutritious and climate resilient crop varieties as per the regional demands. Integrating the ABTs such as CRISPR, reverse breeding, double haploids etc. in the "modernized crop breeding platforms" will not only provide game changing solutions to some of the most "intractable" traits but may also be used for enhancing the expression of superior alleles and removal of deleterious effect alleles. Furthermore, these tools and methodologies may be deployed to reverse domestication by editing genes related to domestication traits in wild species making superior lines with enhanced stress resistance for crop improvement. However, accomplishing these desired impacts would require having curated crop genotyping data sets integrated with the trait data from various crop germplasm panels to assist the discovery of trait-specific SNPs and haplotypes for further excavation of superior genes/alleles that may be subsequently deployed for gene editing applications. To support these endeavors, adaptive and user-friendly allele mining platforms need to be in place to manage and mine the massive datasets that have been generated by sequencing reference genomes and re-sequencing efforts on hundreds of new accessions and large transcriptome datasets.

CRISPR/Cas technology has made remarkable progress in recent years for its practical applicability for targeted genome editing in plant species including crop plants. However, certain obstacles such as transformation efficiency and off-target mutations still need to be overcome. For underutilized crops that are less researched, *in vitro* regeneration and transformation pose a major challenging obstacle. Moreover, the genotypic effects on plant regeneration and transformation can be very challenging. To overcome the problems posed by tissue culture and low transformation efficiencies in important crop species, several plant transformation systems such as RNP based systems or transformation free systems need to be established to increase the precision and editing efficiency of plant genome editing.

There have been continuous efforts in development of the tools and applications which has helped us discover newer technologies with each passing decade. The addition of these advanced tools and technologies in the breeder's tool kit holds tremendous potential to bring changes precisely and efficiently in the genetic makeup of the ruling elite varieties, significantly reducing the need for long breeding cycles for incremental traits speeding up the rate of genetic gains. In addition to the CRISPR/Cas system, several other recently developed systems such as base editing and prime editing have revolutionized the conventional breeding approaches and provided a new direction to the crop improvement programs. With the advancement in new prediction system for on-target activity such as sgRNA CNN (Niu et al., 2021), an array of wider application range has opened leading to an increase in efficiency of crop gene editing and crop improvement programs.

In conclusion, the evolution of genome editing tool kit over the decade has been escalating since the discovery of Cas9 from *Archaea* and undoubtedly, has emerged as the most powerful technology due to its precision, cost effectiveness, and uniqueness to overcome the shortcomings of crop breeding. While ensuing climate change, exploration and creation of additional genetic diversity of underutilized crops require using these precision genetic tools, to create impact on ground, such efforts need to be framed within a breeding pipeline mindset and should be included in the product design process. Although the regulatory pathway for gene edited products is expected to be less complex than for GMOs in several geographies, issues such as freedom to operate and securing the social license need consideration at the intervention design stage. Although CRISPR has potential to deliver disruptive innovations, the trait prioritization should consider the breeding product profiles and market segments for designing and accelerated delivery of locally adapted and preferred crop varieties for the drylands.

Author contributions

PB-M, KS, and SP were involved in the content and design of the manuscript and its review. PB-M, KS, SP, JB, and AS helped with literature and writing of the manuscript. All authors read and approved the manuscript.

Funding

This work was partly supported by CGIAR Research Program on Grain Legumes and Dryland Cereals (CRP-GLDC). ICRISAT and TERI provided funds for open access publication fee.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abrouk, M., Ahmed, H. I., Cubry, P., Šimoníková, D., Cauet, S., Pailles, Y., et al. (2020). Fonio millet genome unlocks African orphan crop diversity for agriculture in a changing climate. *Nat. Commun.* 11 (1), 1–13. doi:10.1038/s41467-020-18329-4

Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., et al. (2017). RNA targeting with CRISPR–Cas13. *Nature* 550, 280–284. doi:10.1038/nature24049

Aher, R. R., Reddy, P. S., Bhunia, R. K., Flyckt, K. S., Shankhapal, A. R., Ojha, R., et al. (2022). *Loss-of-function mutations in novel triacylglycerol lipase genes are associated with low rancidity in pearl millet flour*. bioRxiv. doi:10.1101/2022.04.02.486827

Ahmar, S., Mahmood, T., Fiaz, S., Mora-Poblete, F., Shafique, M. S., Chattha, M. S., et al. (2021). Advantage of nanotechnology-based genome editing system and its application in crop improvement. *Front. Plant Sci.* 12, 663849. doi:10.3389/fpls.2021.663849

Akano, A. O., Dixon, A. G. O., Mba, C., Barrera, E., and Fregene, M. (2002). Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theor. Appl. Genet.* 105, 521–525. doi:10.1007/s00122-002-0891-7

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157. doi:10.1038/s41586-019-1711-4

ASSAf (2016). *The regulatory implications of new breeding techniques*. Academy of Science of South Africa. doi:10.17159/assaf.2016/0011

Ayenan, M. A. T., Sodedji, K. A. F., Nwankwo, C. I., Olodo, K. F., and Alladassi, M. E. B. (2018). Harnessing genetic resources and progress in plant genomics for fonio (*Digitaria* spp.) improvement. *Genet. Resour. Crop Evol.* 65, 373–386. doi:10.1007/s10722-017-0565-6

Baltes, N. J., Hummel, A. W., Konecna, E., Cegan, R., Bruns, A. N., Bisaro, D. M., et al. (2015). Conferring resistance to geminiviruses with the CRISPR–Cas prokaryotic immune system. *Nat. Plants* 1 (10), 1–4. doi:10.1038/nplants.2015.145

Bao, A., Chen, H., Chen, L., Chen, S., Hao, Q., Guo, W., et al. (2019). CRISPR/Cas9-mediated targeted mutagenesis of *GmSPL9* genes alters plant architecture in soybean. *BMC Plant Biol.* 19 (12), 1–12. doi:10.1186/s12870-019-1746-6

Bart, R. S., and Taylor, N. J. (2017). New opportunities and challenges to engineer disease resistance in cassava, a staple food of African small-holder farmers. *PLoS Pathog.* 13, e1006287–7. doi:10.1371/journal.ppat.1006287

Bennetzen, J. L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A. C., et al. (2012). Reference genome sequence of the model plant *Setaria*. *Nat. Biotechnol.* 30, 555–561. doi:10.1038/nbt.2196

Beumer, K., and Swart Jac., A. A. (2019). Who is the african farmer? The importance of actor representations in the debate about biotechnology crops in Africa. *J. Agric. Environ. Ethics.* 34, 1. doi:10.1007/s10806-021-09841-8

Beyene, G., Chauhan, R. D., Villmer, J., Husic, N., Wang, N., Gebre, E., et al. (2022). CRISPR/Cas9-mediated tetra-allelic mutation of the 'Green Revolution' semidwarf-1 (SD-1) gene confers lodging resistance in tef (*Eragrostis tef*). *Plant Biotechnol. J.* 20, 1716–1729. doi:10.1111/pbi.13842

Beyene, G., Chauhan, R. D., Wagaba, H., Moll, T., Alicai, T., Miano, D., et al. (2016). Loss of CMD2-mediated resistance to cassava mosaic disease in plants regenerated through somatic embryogenesis. *Mol. Plant Pathol.* 17, 1095–1110. doi:10.1111/mpp.12353

Bhattacharya, A., Parkhi, V., and Char, B. (2021). Genome editing for crop improvement: A perspective from India. *Vitro Cell. Dev. Biol. Plant.* 57, 565–573. doi:10.1007/s11627-021-10184-2

Bilichak, A., and Kovalchuk, I. (2016). Transgenerational response to stress in plants and its application for breeding. *J. Exp. Bot.* 67, 2081–2092. doi:10.1093/jxb/erw066

Boch, J., and Bonas, U. (2010). *Xanthomonas* AvrBs3 family-type III effectors: Discovery and function. *Annu. Rev. Phytopathol.* 48, 419–436. doi:10.1146/annurev-phyto-080508-081936

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., et al. (2009). Breaking the code of DNA binding specificity of TAL-Type III effectors. *Science* 326, 1509–1512. doi:10.1126/science.1178811

Bohra, A., Saxena, K. B., Varshney, R. K., and Saxena, R. K. (2020). Genomics-assisted breeding for pigeonpea improvement. *Theor. Appl. Genet.* 133, 1721–1737. doi:10.1007/s00122-020-03563-7

Bredeson, J. V., Lyons, J. B., Prochnik, S. E., Wu, G. A., Ha, C. M., Edsinger-Gonzales, E., et al. (2016). Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* 34, 562–570. doi:10.1038/nbt.3535

Brokowski, C. (2018). Do CRISPR germline ethics statements cut it? *CRISPR J* 1, 115–125. doi:10.1089/crispr.2017.0024

Bull, S. E., Seung, D., Chanez, C., Mehta, D., Kuon, J. E., Truernit, E., et al. (2018). Accelerated *ex situ* breeding of GBSS-and PTST1-edited cassava for modified starch. *Sci. Adv.* 4, eaat6086. doi:10.1126/sciadv.aat6086

Burmistrz, M., Krakowski, K., and Krawczyk-Balska, A. (2020). RNA-targeting CRISPR–cas systems and their applications. *Int. J. Mol. Sci.* 21, 1122. doi:10.3390/ijms21031122

Carroll, D. (2011). Genome engineering with zinc-finger nucleases. *Genetics* 188, 773–782. doi:10.1534/genetics.111.131433

Ceasar, S. A., Baker, A., and Ignacimuthu, S. (2017). Functional characterization of thePHT1 family transporters of foxtail millet with development of a novel *Agrobacterium*-mediated transformation procedure. *Sci. Rep.* 7, 14064. doi:10.1038/s41598-017-14447-0

Ceasar, S. A., and Ignacimuthu, S. (2011). *Agrobacterium*-mediated transformation of finger millet (*Eleusine coracana* (L.) Gaertn.) using shoot apex explants. *Plant Cell Rep.* 30, 1759–1770. doi:10.1007/s00299-011-1084-0

Chang, Y., Liu, H., Liu, M., Liao, X., Sahu, S. K., Fu, Y., et al. (2018). Genomic data of the Bambara groundnut (*Vigna subterranea*). *GigaScience* 10, 1055.

Chapman, M. A., He, Y., and Zhou, M. (2022). Beyond a reference genome: Pangenomes and population genomics of underutilized and orphan crops for future food and nutrition security. *New Phytol.* 234, 1583–1597. doi:10.1111/nph.18021

Char, S. N., Lee, H., and Yang, B. (2020). Use of CRISPR/Cas9 for targeted mutagenesis in Sorghum. *Curr. Protoc. Plant Biol.* 5, e20112. doi:10.1002/cppb.20112

Chen, Y., Wang, Z., Ni, H., Xu, Y., Chen, Q., and Jiang, L. (2017). CRISPR/Cas9-mediated base-editing system efficiently generates gain-of-function mutations in *Arabidopsis*. *Sci. China. Life Sci.* 60, 520–523. doi:10.1007/s11427-017-9021-5

Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2013). CCRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* 70, 667–697.

Cheng, Z., Sun, Y., Yang, S., Zhi, H., Yin, T., Ma, X., et al. (2021). Establishing in planta haploid inducer line by edited *SiMTL* in foxtail millet (*Setaria italica*). *Plant Biotechnol. J.* 19, 1089–1091. doi:10.1111/pbi.13584

Chennakesavulu, K., Singh, H., Trivedi, P. K., Jain, M., and Yadav, S. R. (2021). State-of-the-art in CRISPR technology and engineering drought, salinity, and thermo-tolerant crop plants. *Plant Cell Rep.* 41, 815–831. doi:10.1007/s00299-021-02681-w

Chinnusamy, V., and Zhu, J. K. (2009). Epigenetic regulation of stress responses in plants. *Curr. Opin. Plant Biol.* 12, 133–139. doi:10.1016/j.pbi.2008.12.006

Cong, L., Ann Ran, F., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi:10.1126/science.1231143

Cox, D. B., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J., et al. (2017). RNA editing with CRISPR-Cas13. *Science* 358, 1019–1027.

Cui, X., Balcerzak, M., Schernthaner, J., Babic, V., Datla, R., Brauer, E. K., et al. (2019). An optimised CRISPR/Cas9 protocol to create targeted mutations in homoeologous genes and an efficient genotyping protocol to identify edited events in wheat. *Plant Methods* 15 (1), 1–12. doi:10.1186/s13007-019-0500-2

DBT (2022). Guidelines for the safety assessment of genome edited plants. Available at: https://dbtindia.gov.in/sites/default/files/Final_%2011052022_AnnexureI%2C%20Genome_Edited_Plants_2022_Hyperlink.pdf.

De Clercq, J., Zambre, M., Van Montagu, M., Dillen, W., and Angenon, G. (2002). An optimized *Agrobacterium*-mediated transformation procedure for *Phaseolus acutifolius* A. Gray. *Plant Cell Rep.* 21, 333–340. doi:10.1007/s00299-002-0518-0

Dillen, W., De Clercq, J., Goossens, A., Van Montagu, M., and Angenon, G. (1997). *Agrobacterium*-mediated transformation of *Phaseolus acutifolius* A. *Gray. Theor. Appl. Genet.* 94, 151–158. doi:10.1007/s001220050394

El-Mounadi, K., Morales-Floriano, M. L., and Garcia-Ruiz, H. (2020). Principles, applications, and biosafety of plant genome editing using CRISPR-cas9. *Front. Plant Sci.* 11, 56. doi:10.3389/fpls.2020.00056

Ellison, E. E., Nagalakshmi, U., Gamo, M. E., Huang, P. J., Dinesh-Kumar, S., and Voytas, D. F. (2020). Multiplexed heritable gene editing using RNA viruses and mobile single guide RNAs. *Nat. Plants* 6, 620–624. doi:10.1038/s41477-020-0670-y

Endo, A., Masafumi, M., Kaya, H., and Toki, S. (2016). Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from *Francisella novicida*. *Sci. Rep.* 6, 38169. doi:10.1038/srep38169

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to quantitative genetics*. 4th edn. London: Longman Group Ltd.

Fan, D., Liu, T., Li, C., Jiao, B., Li, S., Hou, Y., et al. (2015). Efficient CRISPR/Cas9-mediated targeted mutagenesis in *Populus* in the first generation. *Sci. Rep.* 5, 12217–7. doi:10.1038/srep12217

FSAN (2018). Preliminary report -Review of food derived using new breeding techniques –consultation outcomes. Available at: https://www.foodstandards.gov.au/consumer/gmfood/Documents/NBT%20Preliminary%20report.pdf (Accessed May 20, 2022).

Gaj, T., Gersbach, C. A., and Barbas, C. F., III (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405. doi:10.1016/j.tibtech.2013.04.004

Gale, M. D., and Youssefian, S. (1985). "Dwarfing genes in wheat," in *Progress in plant breeding-1*. Editor G. E. Russel (Elsevier), 1–35. doi:10.1016/B978-0-407-00780-2.50005-9

Gao, W., Long, L., Tian, X., Xu, F., Liu, J., Singh, P. K., et al. (2017). Genome editing in cotton with the CRISPR/Cas9 system. *Front. Plant Sci.* 8, 1364. doi:10.3389/fpls.2017.01364

Ghorbani, A., Hadifar, S., Salari, R., Izadpanah, K., Burmistrz, M., Afsharifar, A., et al. (2021). A short overview of CRISPR-Cas technology and its application in viral disease control. *Transgenic Res.* 30, 221–238. doi:10.1007/s11248-021-00247-w

Ghosh, G., Ganguly, S., Purohit, A., Chaudhuri, R. K., Das, S., and Chakraborti, D. (2017). Transgenic pigeonpea events expressing *Cry1Ac* and *Cry2Aa* exhibit resistance to *Helicoverpa armigera*. *Plant Cell Rep.* 36, 1037–1051. doi:10.1007/s00299-017-2133-0

Glass, Z., Lee, M., Li, Y., and Xu, Q. (2018). Engineering the delivery system for CRISPR-based genome editing. *Trends Biotechnol.* 36, 173–185. doi:10.1016/j.tibtech.2017.11.006

Gomez, M. A., Lin, Z. D., Moll, T., Chauhan, R. D., Hayden, L., Renninger, K., et al. (2019). Simultaneous CRISPR/Cas9-mediated editing of cassava *eIF 4E* isoforms *nCBP*-1 and *nCBP*-2 reduces cassava Brown streak disease symptom severity and incidence. *Plant Biotechnol. J.* 17, 421–434. doi:10.1111/pbi.12987

Gonen, S., Jenko, J., Gorjanc, G., Mileham, A. J., Whitelaw, C. B. A., and Hickey, J. M. (2017). Potential of gene drives with genome editing to increase genetic gain in livestock breeding programs. *Genet. Sel. Evol.* 49, 3–14. doi:10.1186/s12711-016-0280-3

Hamada, H., Liu, Y., Nagira, Y., Miki, R., Taoka, N., and Imai, R. (2018). Biolistic-delivery-based transient CRISPR/Cas9 expression enables in planta genome editing in wheat. *Sci. Rep.* 8, 14422–14427. doi:10.1038/s41598-018-32714-6

Han, Y., Broughton, S., Liu, L., Zhang, X. Q., Zeng, J., He, X., et al. (2021). Highly efficient and genotype-independent barley gene editing based on anther culture. *Plant Commun.* 2, 100082. doi:10.1016/j.xplc.2020.100082

Hema, R., Vemanna, R. S., Sreeramulu, S., Reddy, C. P., Senthil-Kumar, M., and Udayakumar, M. (2014). Stable expression of *mtlD* gene imparts multiple stress tolerance in finger millet. *PLoS ONE* 9, e99110. doi:10.1371/journal.pone.0099110

Hess, G. T., Tycko, J., Yao, D., and Bassik, M. C. (2017). Methods and applications of CRISPR-mediated base editing in eukaryotic genomes. *Mol. Cell* 68, 26–43. doi:10.1016/j.molcel.2017.09.029

Hickey, L. T., Hafeez, A. N., Robinson, H., Jackson, S. A., Leal-Bertioli, S., Tester, M., et al. (2019). Breeding crops to feed 10 billion. *Nat. Biotechnol.* 37, 744–754. doi:10.1038/s41587-019-0152-9

Hilley, J. L., Weers, B. D., Truong, S. K., McCormick, R. F., Mattison, A. J., McKinley, B. A., et al. (2017). Sorghum *Dw2* encodes a protein kinase regulator of stem internode length. *Sci. Rep.* 7 (1), 1–13. doi:10.1038/s41598-017-04609-5

Hilscher, J., Bürstmayr, H., and Stoger, E. (2017). Targeted modification of plant genomes for precision crop breeding. *Biotechnol. J.* 12 (1), 1600173. doi:10.1002/biot.201600173

Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., et al. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* 33, 510–517. doi:10.1038/nbt.3199

Houngue, J. A., Zandjanakou-Tachin, M., Ngalle, H. B., Pita, J. S., Cacaï, G. H. T., Ngatat, S. E., et al. (2019). Evaluation of resistance to cassava mosaic disease in selected African cassava cultivars using combined molecular and greenhouse grafting tools. *Physiol. Mol. Plant Pathol.* 105, 47–53. doi:10.1016/j.pmpp.2018.07.003

Hua, K., Jiang, Y., Tao, X., and Zhu, J. K. (2020). Precision genome engineering in rice using prime editing system. *Plant Biotechnol. J.* 18, 2167–2169. doi:10.1111/pbi.13395

Huang, L., Zhang, R., Huang, G., Li, Y., Melaku, G., Zhang, S., et al. (2018). Developing superior alleles of yield genes in rice by artificial mutagenesis using the CRISPR/Cas9 system. *Crop J.* 6, 475–481. doi:10.1016/j.cj.2018.05.005

Huang, X., Hilscher, J., Stoger, E., Christou, P., and Zhu, C. (2021). Modification of cereal plant architecture by genome editing to improve yields. *Plant Cell Rep.* 40, 953–978. doi:10.1007/s00299-021-02668-7

Hummel, A. W., Chauhan, R. D., Cermak, T., Mutka, A. M., Vijayaraghavan, A., Boyher, A., et al. (2018). Allele exchange at the EPSPS locus confers glyphosate tolerance in cassava. *Plant Biotechnol. J.* 16, 1275–1282. doi:10.1111/pbi.12868

Ignacimuthu, S., and Kannan, P. (2013). *Agrobacterium*-mediated transformation of pearl millet (*Pennisetum typhoides* (L.) R.Br.) for fungal resistance. *Asian J. Plant Sci.* 12, 97–108. doi:10.3923/ajps.2013.97.108

Jacob, P., Avni, A., and Bendahmane, A. (2018). Translational research: Exploring and creating genetic diversity. *Trends Plant Sci.* 23, 42–52. doi:10.1016/j.tplants.2017.10.002

Jain, M. (2015). Function genomics of abiotic stress tolerance in plants: A CRISPR approach. *Front. Plant Sci.* 6, 375. doi:10.3389/fpls.2015.00375

Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schmöckel, S. M., Li, B., Borm, T. J. A., et al. (2017). The genome of *Chenopodium quinoa*. *Nature* 542, 307–312. doi:10.1038/nature21370

Ji, X., Yang, B., and Wang, D. (2020). Achieving plant genome editing while bypassing tissue culture. *Trends Plant Sci.* 25, 427–429. doi:10.1016/j.tplants.2020.02.011

Jia, G., Huang, X., Zhi, H., Zhao, Y., Zhao, Q., Li, W., et al. (2013). A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nat. Genet.* 45, 957–961. doi:10.1038/ng.2673

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. A. (2013a). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 13 (3), 233–239.

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., and Weeks, D. (2013b). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res.* 41 (20), e188.

Jiri, O., Mafongoya, P. L., and Chivenge, P. (2017). Climate smart crops for food and nutritional security for semi-arid zones of Zimbabwe. *Afr. J. Food Agric. Nutr. Dev.* 17 (3), 12280–12294. doi:10.18697/ajfand.79.16285

Kakoulodou, I., Avramidou, E. V., Baranek, M., Brunel-Muguet, S., Farrona, S., Johannes, F., et al. (2021). Epigenetics for crop improvement in times of global change. *Biology* 10, 766. doi:10.3390/biology10080766

Kang, S., Jeon, S., Kim, S., Chang, Y. K., and Kim, Y. C. (2020). Development of a pVEC peptide-based ribonucleoprotein (RNP) delivery system for genome editing using CRISPR/Cas9 in *Chlamydomonas reinhardtii*. *Sci. Rep.* 10 (1), 1–11. doi:10.1038/s41598-020-78968-x

Kim, H., Choi, J., and Won, K. H. (2020). A stable DNA-free screening system for CRISPR/RNPs-mediated gene editing in hot and sweet cultivars of *Capsicum annuum*. *BMC Plant Biol.* 20 (1), 1–12. doi:10.1186/s12870-020-02665-0

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533 (7603), 420–424.

Koné, M., Paice, A. G., and Touré, Y. (2011). "Bambara groundnut [*Vigna subterranea* (L.) Verdc. Fabaceae)] usage in human health," in *Nuts and seeds in health and disease prevention*. Editors V. R. Preedy, R. R. Watson, and V. B. Patel (UK: Elsevier), 189–196. doi:10.1016/B978-0-12-375688-6.10022-2

Konermann, S., Lotfy, P., Brideau, N. J., Oki, J., Shokhirev, M. N., and Hsu, P. D. (2018). Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. *Cell* 173, 665–676. doi:10.1016/j.cell.2018.02.033

Koonin, E. V., Makarova, K. S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* 37, 67–78. doi:10.1016/j.mib.2017.05.008

Ku, H. K., and Ha, S. H. (2020). Improving nutritional and functional quality by genome editing of crops: Status and perspectives. *Front. Plant Sci.* 11, 1514. doi:10.3389/fpls.2020.577313

Kumar, S., AlAbed, D., Worden, A., Novak, S., Wu, H., Ausmus, C., et al. (2015). A modular gene targeting system for sequential transgene stacking in plants. *J. Biotechnol.* 207, 12–20. doi:10.1016/j.jbiotec.2015.04.006

Kumara Charyulu, D., Bantilan, C., Rajalaxmi, A., Rai, K. N., Yadav, O. P., Gupta, S. K., et al. (2014). *Development and diffusion of pearl millet improved cultivars in India: Impact on growth and yield stability*. Working Paper Series No. 52. Patancheru 502 324. Telangana, India: International Crops Research Institute for the Semi-Arid Tropics, 76.

Laaninen, T. (2019). *New plant-breeding techniques Applicability of EU GMO rules*. EPRS | European Parliamentary Research Service. Available at: https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/642235/EPRS_BRI(2019)642235_EN.pdf.

Lämke, J., Brzezinka, K., Altmann, S., and Bäurle, I. (2016). A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. *EMBO J.* 35, 162–175. doi:10.15252/embj.201592593

Lassoued, R., Phillips, P. W. B., Macall, D. M., Hesseln, H., and Smyth, S. J. (2021). Expert opinions on the regulation of plant genome editing. *Plant Biotechnol. J.* 19, 1104–1109. doi:10.1111/pbi.13597

Lee, H., Seon, H. J., and Choi, Y. D. (2018). Radiological-pathological correlation of malignant teratoma with liposarcomatous transformation: Proven by repeated transthoracic needle biopsy. *Thorac. Cancer* 12, 185–188. doi:10.1111/1759-7714.12553

Lee, K., Zhang, Y., Kleinstiver, B. P., Guo, J. A., Aryee, M. J., Miller, J., et al. (2019). Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. *Plant Biotechnol. J.* 17, 362–372. doi:10.1111/pbi.12982

Lema, M. (2019). Regulatory aspects of gene editing in Argentina. *Transgenic Res.* 28, 147–150. doi:10.1007/s11248-019-00145-2

Lemmon, Z. H., Reem, N. T., Dalrymple, J., Soyk, S., Swartwood, K. E., Rodriguez-Leal, D., et al. (2018). Rapid improvement of domestication traits in an orphan crop by genome editing. *Nat. Plants* 4, 766–770. doi:10.1038/s41477-018-0259-x

Li, A., Jia, S., Yobi, A., Ge, Z., Sato, S. J., Zhang, C., et al. (2018a). Editing of an alpha-kafirin gene family increases, digestibility and protein quality in sorghum. *Plant Physiol.* 177, 1425–1438. doi:10.1104/pp.18.00200

Li, C., Zong, Y., Wang, Y., Jin, S., Zhang, D., Song, Q., et al. (2018b). Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol.* 19, 59–9. doi:10.1186/s13059-018-1443-z

Li, G., Liu, Y. G., and Chen, Y. (2019a). Genome-editing technologies: The gap between application and policy. *Sci. China. Life Sci.* 62, 1534–1538. doi:10.1007/s11427-019-1566-1

Li, Q., Wu, G., Zhao, Y., Wang, B., Zhao, B., Kong, D., et al. (2020b). CRISPR/Cas9-mediated knockout and overexpression studies reveal a role of maize phytochrome C in regulating flowering time and plant height. *Plant Biotechnol. J.* 18, 2520–2532. doi:10.1111/pbi.13429

Li, S., and Xia, L. (2020a). Precise gene replacement in plants through CRISPR/cas genome editing technology: Current status and future perspectives. *aBIOTECH* 1, 58–73. doi:10.1007/s42994-019-00009-7

Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., et al. (2011). Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. *Nat. Genet.* 43, 1266–1269. doi:10.1038/ng.977

Li, Z., Liu, Z. B., Xing, A., Moon, B. P., Koellhoffer, J. P., Huang, L., et al. (2015). Cas9-guide RNA directed genome editing in soybean. *Plant Physiol.* 169, 960–970. doi:10.1104/pp.15.00783

Li, T., Hu, J., Sun, Y., Li, B., Zhang, D., Li, W., et al. (2021). Highly efficient heritable genome editing in wheat using an RNA virus and bypassing tissue culture. *Mol. Plant* 14 (11), 1787–1798.

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., et al. (2017). Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nat. Commun.* 8, 14261–14265. doi:10.1038/ncomms14261

Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., et al. (2020). Prime genome editing in rice and wheat. *Nat. Biotechnol.* 38, 582–585. doi:10.1038/s41587-020-0455-x

Liu, C., Li, X., Meng, D., Zhong, Y., Chen, C., Dong, X., et al. (2017). A 4-bp insertion at *ZmPLA1* encoding a putative phospholipase A generates haploid induction in maize. *Mol. Plant* 10, 520–522. doi:10.1016/j.molp.2017.01.011

Liu, J., Zheng, Q., Ma, Q., Gadidasu, K. K., and Zhang, P. (2011). Cassava genetic transformation and its application in breeding. *J. Integr. Plant Biol.* 53, 552–569. doi:10.1111/j.1744-7909.2011.01048.x

Liu, T., Li, R., Zhong, X., Jiang, M., Jin, X., Zhou, P., et al. (2018). Estimates of rice lodging using indices derived from UAV visible and thermal infrared images. *Agric. For. Meteorol.* 252, 144–154. doi:10.1016/j.agrformet.2018.01.021

Liu, Y., Luo, W., Linghu, Q., Abe, F., Hisano, H., Sato, K., et al. (2021). *In planta* genome editing in commercial wheat varieties. *Front. Plant Sci.* 12, 648841. doi:10.3389/fpls.2021.648841

Lloyd, A., Plaisier, C. L., Carroll, D., and Drews, G. N. (2005). Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2232–2237. doi:10.1073/pnas.0409339102

López, C. E., and Bernal, A. J. (2012). Cassava bacterial blight: Using genomics for the elucidation and management of an old problem. *Trop. Plant Biol.* 5, 117–126. doi:10.1007/s12042-011-9092-3

Lu, Y., Tian, Y., Shen, R., Yao, Q., Zhong, D., Zhang, X., et al. (2021). Precise genome modification in tomato using an improved prime editing system. *Plant Biotechnol. J.* 19, 415–417. doi:10.1111/pbi.13497

Lu, Y., and Zhu, J. K. (2017). Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. *Mol. Plant* 10, 523–525. doi:10.1016/j.molp.2016.11.013

Ma, X., Mau, M., and Sharbel, T. F. (2018). Genome editing for global food security. *Trends Biotechnol.* 36, 123–127. doi:10.1016/j.tibtech.2017.08.004

Mahas, A., Ali, Z., Tashkandi, M., and Mahfouz, M. M. (2019). Virus-mediated genome editing in plants using the CRISPR/Cas9 system. *Methods Mol. Biol.* 1917, 311–326. doi:10.1007/978-1-4939-8991-1_23

Maher, M. F., Nasti, R. A., Vollbrecht, M., Starker, C. G., Clark, M. D., and Voytas, D. F. (2020). Plant gene editing through *de novo* induction of meristems. *Nat. Biotechnol.* 38, 84–89. doi:10.1038/s41587-019-0337-2

Majola, N. G., Gerrano, A. S., and Shimelis, H. (2021). Bambara groundnut (Vigna subterranea [L.] Verdc.) production, utilisation and genetic improvement in Sub-Saharan Africa. *Agronomy* 11 (7), 1345. doi:10.3390/agronomy11071345

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., et al. (2020). Evolutionary classification of CRISPR–cas systems: A burst ofclass 2 and derived variants. *Nat. Rev. Microbiol.* 18, 67–83. doi:10.1038/s41579-019-0299-x

Martin-Ortigosa, S., Peterson, D. J., Valenstein, J. S., Lin, V. S., Trewyn, B. G., Lyznik, L. A., et al. (2014). Mesoporous silica nanoparticle-mediated intracellular cre protein delivery for maize genome editing via loxP site excision. *Plant Physiol.* 164, 537–547. doi:10.1104/pp.113.233650

Martins, P. K., Ribeiro, A. P., Cunha, B., Kobayashi, A. K., and Molinari, H. B. C. (2015). A simple and highly efficient *Agrobacterium*-mediated transformation protocol for *Setaria viridis*. *Biotechnol. Rep.* 6, 41–44. doi:10.1016/j.btre.2015.02.002

Marx, V. (2018). Base editing a CRISPR way. *Nat. Methods* 15, 767–770. doi:10.1038/s41592-018-0146-4

Marzec, M., and Hensel, G. (2020). Prime editing: Game changer for modifying plant genomes. *Trends Plant Sci.* 25, 722–724. doi:10.1016/j.tplants.2020.05.008

Mascher, M., Schreiber, M., Scholz, U., Graner, A., Reif, J. C., and Stein, N. (2019). Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nat. Genet.* 51, 1076–1081. doi:10.1038/s41588-019-0443-6

Mayes, S., Massawe, F. J., Alderson, P. G., Roberts, J. A., Azam-Ali, S. N., and Hermann, M. (2012). The potential for underutilized crops to improve security of food production. *J. Exp. Bot.* 63, 1075–1079. doi:10.1093/jxb/err396

Menz, J., Modrzejewski, D., Hartung, F., Wilhelm, R., and Sprink, T. (2020). Genome edited crops touch the market: A view on the global development and regulatory environment. *Front. Plant Sci.* 11, 586027. doi:10.3389/fpls.2020.586027

Miladinovic, D., Antunes, D., Yildirim, K., Bakhsh, A., Cvejić, S., Kondic-Spika, A., et al. (2021). Targeted plant improvement through genome editing: From laboratory to field. *Plant Cell Rep.* 40, 935–951. doi:10.1007/s00299-020-02655-4

Miller, D. R. (2007). Management guide for "tiffany teff" forage grass. Available at: http://www.targetseed.com.

Minnaar-Ontong, A., Gerrano, A. S., and Labuschagne, M. T. (2021). Assessment of genetic diversity and structure of Bambara groundnut [*Vigna subterranean* (L.) verdc.] landraces in South Africa. *Sci. Rep.* 11, 7408. doi:10.1038/s41598-021-86977-7

Mishra, R., Joshi, R. K., and Zhao, K. (2020). Base editing in crops: Current advances, limitations and future implications. *Plant Biotechnol. J.* 18, 20–31. doi:10.1111/pbi.13225

Miura, K., Ikeda, M., Matsubara, A., Song, X. J., Ito, M., Asano, K., et al. (2010). *OsSPL14* promotes panicle branching and higher grain productivity in rice. *Nat. Genet.* 42, 545–549. doi:10.1038/ng.592

Monna, L., Kitazawa, N., Yoshino, R., Suzuki, J., Masuda, H., Maehara, Y., et al. (2002). Positional cloning of rice semidwarfing gene, sd-1: Rice "green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. *DNA Res.* 9, 11–17. doi:10.1093/dnares/9.1.11

Monsur, M. B., Shao, G., Lv, Y., Ahmad, S., Wei, X., Hu, P., et al. (2020). Base editing: The ever expanding clustered regularly interspaced short palindromic repeats (CRISPR) tool kit for precise genome editing in plants. *Genes* 11, 466. doi:10.3390/genes11040466

Moradpour, M., and Abdulah, S. N. A. (2020). CRISPR/dCas9 platforms in plants: Strategies and applications beyond genome editing. *Plant Biotechnol. J.* 18, 32–44. doi:10.1111/pbi.13232

Muhammad, I., Rafii, M. Y., Ramlee, S. I., Nazli, M. H., Harun, A. R., Oladosu, Y., et al. (2020). Exploration of Bambara groundnut (*Vigna subterranea* (L.) Verdc.), an underutilized crop, to aid global food security: Varietal improvement, genetic diversity and processing. *Agronomy* 10, 766. doi:10.3390/agronomy10060766

Multani, D. S., Briggs, S. P., Chamberlin, M. A., Blakeslee, J. J., Murphy, A. S., and Johal, G. S. (2003). Loss of an MDR transporter in compact stalks of maize *br2* and sorghum *dw3* mutants. *Science* 302, 81–84. doi:10.1126/science.1086072

Nekrasov, V., Wang, C., Win, J., Lanz, C., Weigel, D., and Kamoun, S. (2017). Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Sci. Rep.* 7, 482–486. doi:10.1038/s41598-017-00578-x

Nelson-Vasilchik, K., Hague, J. P., Tilelli, M., and Kausch, A. P. (2022). "Rapid transformation and plant regeneration of sorghum (*Sorghum bicolor* L.) mediated by altruistic *Baby boom* and *Wuschel2*," *in Vitrocell* (Dev. Biol.-Plant.). doi:10.1007/s11627-021-10243-8

Ngalamu, T., Odra, J., and Tongun, N. (2015). *Cowpea production handbook*. IFS/AGRA.

Niu, M., Lin, Y., and Zou, Q. (2021). sgRNACNN: identifying sgRNA on-target activity in four crops using ensembles of convolutional neural networks. *Plant Mol. Biol.* 105, 483–495. doi:10.1007/s11103-020-01102-y

Nishizawa-Yokoi, A., and Toki, S. (2021). A piggyBac-mediated transgenesis system for the temporary expression of CRISPR/Cas9 in rice. *Plant Biotechnol. J.* 19 (7), 1386–1395.

NRC (1996). *Lost crops of Africa: Volume I: Grains*. Washington D.C. USA: National Research Council: The National Academy of Sciences, National Academies Press, 408. doi:10.17226/2305

Numan, M., Khan, A. L., Asaf, S., Salehin, M., Beyene, G., Tadele, Z., et al. (2021). From traditional breeding to genome editing for boosting productivity of the ancient grain tef [*Eragrostis tef* (Zucc.) Trotter]. *Plants* 10, 628. doi:10.3390/plants10040628

Oberkofler, V., and Bäurle, I. (2022). Inducible epigenome editing probes for the role of histone H3K4 methylation in *Arabidopsis* heat stress memory. *Plant Physiol.* 189, 703–714. doi:10.1093/plphys/kiac113

Obukosia, S., Akinbo, O., Sinebo, W., Savadogo, M., Timpo, S., Adeyemo, M., et al. (2020). Update of regulatory options of new breeding techniques and biosafety approaches among selected countries: A review. *Asian J. Biotechnol. Bioresour. Technol.* 6, 18–35. doi:10.9734/AJB2T/2020/v6i330083

Olanrewaju, O. S., Oyatomi, O., Babalola, O. O., and Abberton, M. (2022). Breeding potentials of Bambara groundnut for food and nutrition security in the face of climate change. *Front. Plant Sci.* 12, 798993. doi:10.3389/fpls.2021.798993

Oliva, R., Ji, C., Atienza-Grande, G., Huguet-Tapia, J. C., Perez-Quintero, A., Li, T., et al. (2019). Broad-spectrum resistance to bacterial blight in rice using genome editing. *Nat. Biotechnol.* 37, 1344–1350. doi:10.1038/s41587-019-0267-z

Otun, S., Escrich, A., Achilonu, I., Rauwane, M., Lerma-Escalera, J. A., Morones-Ramirez, J. R., et al. (2022). The future of cassava in the era of biotechnology in Southern Africa. *Crit. Rev. Biotechnol.* 2022, 1–19. doi:10.1080/07388551.2022.2048791

Paliwal, R., Adegboyega, T. T., Abberton, M., Faloye, B., and Oyatomi, O. (2021). Potential of genomics for the improvement of underutilized legumes in sub-Saharan Africa. *Legum. Sci.* 3, e69. doi:10.1002/leg3.69

Pazhamala, L., Saxena, R. K., Singh, V. K., Sameerkumar, C. V., Kumar, V., Sinha, P., et al. (2015). Genomics-assisted breeding for boosting crop improvement in pigeonpea (*Cajanus cajan*). *Front. Plant Sci.* 6, 50. doi:10.3389/fpls.2015.00050

Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., et al. (1999). 'Green revolution'genes encode mutant gibberellin response modulators. *Nature* 400, 256–261. doi:10.1038/22307

Pérez-Jaramillo, J. E., Mendes, R., and Raaijmakers, J. M. (2016). Impact of plant domestication on rhizosphere microbiome assembly and functions. *Plant Mol. Biol.* 90, 635–644. doi:10.1007/s11103-015-0337-7

Petolino, J. F. (2015). *Vitro Cell*, 51, 1–8.Genome editing in plants via designed zinc finger nucleases*Dev. Biol.-Plant*

Popoola, J., Ojuederie, O., Omonhinmin, C., and Adegbite, A. (2019). "Neglected and underutilized legume crops: Improvement and future prospects," in *Recent advances in grain crops research* (IntechOpen). doi:10.5772/intechopen.87069

Prochnik, S., Marri, P. R., Desany, B., Rabinowicz, P. D., Kodira, C., Mohiuddin, M., et al. (2012). The cassava genome: Current progress, future directions. *Trop. Plant Biol.* 5, 88–94. doi:10.1007/s12042-011-9088-z

Qaim, M. (2020). Role of new plant breeding technologies for food security and sustainable agricultural development. *Appl. Econ. Perspect. Policy* 42, 129–150. doi:10.1002/aepp.13044

Rabbi, I. Y., Hamblin, M. T., Kumar, P. L., Gedil, M. A., Ikpan, A. S., Jannink, J. L., et al. (2014). High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. *Virus Res.* 186, 87–96. doi:10.1016/j.virusres.2013.12.028

Ramineni, R., Sadumpati, V., Khareedu, V. R., and Vudem, D. R. (2014). Transgenic pearl millet male fertility restorer line (ICMP451) and hybrid (ICMH451) expressing *Brassica juncea* nonexpressor of pathogenesis related genes 1 (BjNPR1) exhibit resistance to downy mildew disease. *PLoS One* 9, e90839. doi:10.1371/journal.pone.0090839

Ran, Y., Liang, Z., and Gao, C. (2017). Current and future editing reagent delivery systems for plant genome editing. *Sci. China. Life Sci.* 60, 490–505. doi:10.1007/s11427-017-9022-1

Ren, Q., Sretenovic, S., Liu, G., Zhong, Z., Wang, J., Huang, L., et al. (2021). Improved plant cytosine base editors with high editing activity, purity, and specificity. *Plant Biotechnol. J.* 19, 2052–2068. doi:10.1111/pbi.13635

Rojas-Vasquez, R., and Gatica-Arias, A. (2020). Use of genome editing technologies for genetic improvement of crops of tropical origin. *Plant Cell, Tissue and Organ Culture (PCTOC)* 140 (1), 215–244. doi:10.1111/pbi.13635

Saha, D., Gowda, M. C., Arya, L., Verma, M., and Bansal, K. C. (2016). Genetic and genomic resources of small millets. *CRC. Crit. Rev. Plant Sci.* 35, 56–79. doi:10.1080/07352689.2016.1147907

Sanjana, N. E., Cong, L., Zhou, Y., Cunniff, M. M., Feng, G., and Zhang, F. (2012). A transcription activator-like effector toolbox for genome engineering. *Nat. Protoc.* 7, 171–192. doi:10.1038/nprot.2011.431

Santos, C. M., Romeiro, D., Silva, J. P., Basso, M. F., Molinari, H. B. C., and Centeno, D. C. (2020). An improved protocol for efficient transformation and regeneration of *Setaria italica*. *Plant Cell Rep.* 39, 501–510. doi:10.1007/s00299-019-02505-y

Sardesai, N., and Subramanyam, S. (2018). *Agrobacterium*: A genome-editing tool-delivery system. *Curr. Top. Microbiol. Immunol.* 418, 463–488. doi:10.1007/82_2018_101

Sasaki, A., Ashikari, M., Ueuchi-Tanaka, M., Itoh, H., Nishimura, A., Swapan, D., et al. (2002). Green revolution: A mutant gibberellin-synthesis gene in rice. *Nature* 416, 701–702. doi:10.1038/416701a

Saxena, R. K., Obala, J., Sinjushin, A., Sameer Kumar, C. V., Saxena, K. B., and Varshney, R. K. (2017). Characterization and mapping of *Dt1* locus which co-segregates with CcTFL1 for growth habit in pigeonpea. *Theor. Appl. Genet.* 130, 1773–1784. doi:10.1007/s00122-017-2924-2

Saxena, R. K., Saxena, K. B., Pazhamala, L. T., Patel, K., Parupalli, S., Sameerkumar, C. V., et al. (2015). Genomics for greater efficiency in pigeonpea hybrid breeding. *Front. Plant Sci.* 6, 793. doi:10.3389/fpls.2015.00793

Sharma, B. (2015). *Biochemical characterization of pearl millet (*Pennisetum glaucum *(L.) R. Br.) genotypes for identifying off-odour factors*. Hisar, India: CCS Haryana Agricultural University. Ph.D. Thesis.

Sharma, K. K., Sreelatha, G., and Dayal, S. (2006). "Pigeonpea [*Cajanus cajan* L. (Millsp.)]," in *Methods in Molecular Biology Vol. 343:* Agrobacterium Protocols, 2/e, *volume 1*. Editor K. Wang (Totowa, U.S.A. Humana Press), 359–367.

Shimatani, Z., Kashojiya, S., Takayama, M., Terada, R., Arazoe, T., Ishii, H., et al. (2017). Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* 35, 441–443. doi:10.1038/nbt.3833

Shin, H., Choi, W. L., Lim, J. Y., and Huh, J. H. (2022). Epigenome editing: Targeted manipulation of epigenetic modifications in plants. *Genes Genomics* 44, 307–315. doi:10.1007/s13258-021-01199-5

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., et al. (2017). Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* 15, 169–182. doi:10.1038/nrmicro.2016.184

Singh, D., Chaudhury, P., Taunk, J., Singh, C. K., Sharma, S., Singh, V. J., et al. (2021). Plant epigenomics for extenuation of abiotic stresses: Challenges and future perspectives. *J. Exp. Bot.* 72, 6836–6855. doi:10.1093/jxb/erab337

Singh, R. K., Prasad, A., Muthamilarasan, M., Parida, S. K., and Prasad, M. (2020). Breeding and biotechnological interventions for trait improvement: Status and prospects. *Planta* 252, 54. doi:10.1007/s00425-020-03465-4

Smyth, S. J. (2017). Canadian regulatory perspectives on genome engineered crops. *Gm. Crops Food* 8, 35–43. doi:10.1080/21645698.2016.1257468

Spielmeyer, W., Ellis, M. H., and Chandler, P. M. (2002). Semidwarf (sd-1), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. *Proc. Natl. Acad. Sci. U. S. A.* 99, 9043–9048. doi:10.1073/pnas.132266399

Sretenovic, S., Yin, D., Levav, A., Selengut, J. D., Mount, S. M., and Qi, Y. (2021). Expanding plant genome-editing scope by an engineered iSpyMacCas9 system that targets A-rich PAM sequences. *Plant Commun.* 2, 100101. doi:10.1016/j.xplc.2020.100101

Sun, C., Ali, K., Yan, K., Fiaz, S., Dormatey, R., Bi, Z., et al. (2021). Exploration of epigenetics for improvement of drought and other stress resistance in crops: A review. *Plants* 10, 1226. doi:10.3390/plants10061226

Tadele, Z., and Assefa, K. (2012). Increasing food production in Africa by boosting the productivity of understudied crops. *Agronomy* 2, 240–283. doi:10.3390/agronomy2040240

Tang, Y., and Fu, Y. (2018). Class 2 CRISPR/cas: An expanding biotechnology toolbox for and beyond genome editing. *Cell Biosci.* 8, 59–13. doi:10.1186/s13578-018-0255-x

Taylor, J. R. N. (2017). "Millets: Their unique nutritional and health-promoting attributes," in *Gluten-free ancient grains*. Editors J. R. N. Taylor and J. M. Awika (Woodhead Publishing), 55–103. doi:10.1016/B978-0-08-100866-9.00004-2

Tian, S., Jiang, L., Cui, X., Zhang, J., Guo, S., Li, M., et al. (2018). Engineering herbicide-resistant watermelon variety through CRISPR/Cas9-mediated base-editing. *Plant Cell Rep.* 37, 1353–1356. doi:10.1007/s00299-018-2299-0

Tra, M. V. T., Yin, X., Bajal, I., Balahadia, C. P., Quick, W. P., and Bandyopadhyay, A. (2021). Single base editing using cytidine deaminase to change grain size and seed coat color in rice. *Methods Mol. Biol.* 2238, 135–143. doi:10.1007/978-1-0716-1068-8_9

Travella, S., De Oliveira, D., De Buck, S., Lambein, F., Ngudi, D. D., De Bauw, V., et al. (2019). Scientific innovation for the sustainable development of african agriculture. *Afr. Focus* 32, 117–133. doi:10.1163/2031356x-03202010

Tribhuvan, K. U., Das, A., Srivastava, H., Kumar, K., Durgesh, K., Mithra, S. V., et al. (2020). Identification and characterization of PEBP family genes reveal *CcFT8* a probable candidate for photoperiod insensitivity in *C. cajan*. *3 Biotech.* 10, 194. doi:10.1007/s13205-020-02180-x

Tripathi, L., Dhugga, K. S., Ntui, V. O., Runo, S., Syombua, E. D., Muiruri, S., et al. (2022). Genome editing for sustainable agriculture in Africa. *Front. Genome Ed.* 4, 876697. doi:10.3389/fgeed.2022.876697

Tripathi, L., Ntui, V. O., and Tripathi, J. N. (2019). Application of genetic modification and genome editing for developing climate-smart banana. *Food Energy secur.* 8, e00168. doi:10.1002/fes3.168

Tsuda, M., Watanabe, K. N., and Ohsawa, R. (2019). Regulatory status of genome-edited organisms under the Japanese Cartagena Act. *Front. Bioeng. Biotechnol.* 387. doi:10.3389/fbioe.2019.00387

USDA (2020b). Policies and procedures for genome edited food and agricultural products. Available at: https://apps.fas.usda.gov/newgainapi/api/Report/DownloadReportByFileName?fileName=Policies%20and%20Procedures%20for%20Genome%20Edited%20Food%20and%20Agricultural%20Products_Tokyo_Japan_11-02-2020.

USDA (2020a). USDA SECURE rule paves way for agricultural innovation. Available at: https://www.aphis.usda.gov/aphis/newsroom/news/sa_by_date/sa-2020/secure.

Valentine, O. N., Edak, A. U., Ikuo, N., and Masahiro, M. (2017). Agrobacterium-mediated genetic transformation of fonio (*Digitaria exilis* (L.) stapf). *Afr. J. Biotechnol.* 16, 1302–1307. doi:10.5897/ajb2017.15903

Varshney, R. K., Saxena, R., Upadhyaya, H., Khan, A. W., Yu, Y., Kim, C., et al. (2017). Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nat. Genet.* 49, 1082–1088. doi:10.1038/ng.3872

Vats, S., Kumawat, S., Kumar, V., Patil, G. B., Joshi, T., Sonah, H., et al. (2019). Genome editing in plants: Exploration of technological advancements and challenges. *Cells* 8, 1386. doi:10.3390/cells8111386

Vega-Gálvez, A., Miranda, M., Vergara, J., Uribe, E., Puente, L., and Martínez, E. A. (2010). Nutrition facts and functional potential of quinoa (*Chenopodium quinoa* willd.), an ancient andean grain: A review. *J. Sci. Food Agric.* 90, 2541–2547. doi:10.1002/jsfa.4158

Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M. P., Guyon-Debast, A., Chauvin, J. E., et al. (2019). Transgene-free genome editing in tomato and potato plants using *Agrobacterium*-mediated delivery of a CRISPR/Cas9 cytidine base editor. *Int. J. Mol. Sci.* 20, 402. doi:10.3390/ijms20020402

Veley, K. M., Okwuonu, I., Jensen, G., Yoder, M., Taylor, N. J., Meyers, B. C., et al. (2021). Gene tagging via CRISPR-mediated homology-directed repair in cassava. *G3* 11, jkab028. doi:10.1093/g3journal/jkab028

Venske, E., Dos Santos, R. S., Busanello, C., Gustafson, P., and Costa de Oliveira, A. (2019). Bread wheat: A role model for plant domestication and breeding. *Hereditas* 156, 16–11. doi:10.1186/s41065-019-0093-9

Vetriventhan, M., Azevedo, V. C., Upadhyaya, H. D., Nirmalakumari, A., Kane-Potaka, J., Anitha, S., et al. (2020). Genetic and genomic resources, and breeding for accelerating improvement of small millets: Current status and future interventions. *Nucl. (Camb).* 63 (3), 217–239. doi:10.1007/s13237-020-00322-3

Wang, C., Liu, Q., Shen, Y., Hua, Y., Wang, J., Lin, J., et al. (2019). Clonal seeds from hybrid rice by simultaneous genome engineering of meiosis and fertilization genes. *Nat. Biotechnol.* 37, 283–286. doi:10.1038/s41587-018-0003-0

Wang, Y., Zhang, Y., Dai, C., Ma, J., Zhou, Y., Jaikishun, S., et al. (2021). *The establishment of two efficient transformation systems in quinoa*. Research Square. doi:10.21203/rs.3.rs-364280/v1

Wei, L., Zhang, X., Zhang, Z., Liu, H., and Lin, Z. (2018). A new allele of the *Brachytic2* gene in maize can efficiently modify plant architecture. *Heredity* 121, 75–86. doi:10.1038/s41437-018-0056-3

Wolfe, M. D., Rabbi, I. Y., Egesi, C. N., Hamblin, M., Kawuki, R., Kulakow, P., et al. (2016). Genome-wide association and prediction reveals genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. *Plant Genome* 9, 2. doi:10.3835/plantgenome2015.11.0118

Wolt, J. D., Wang, K., Sashital, D., and Lawrence-Dill, C. J. (2016). Achieving plant CRISPR targeting that limits off-target effects. *Plant Genome* 9, 2016–2105. doi:10.3835/plantgenome2016.05.0047

Würschum, T., Langer, S. M., Longin, C. F. H., Tucker, M. R., and Leiser, W. L. (2017). A modern green revolution gene for reduced height in wheat. *Plant J.* 92, 892–903. doi:10.1111/tpj.13726

Xu, R., Li, J., Liu, X., Shan, T., Qin, R., and Wei, P. (2020). Development of plant prime-editing systems for precise genome editing. *Plant Commun.* 1, 100043. doi:10.1016/j.xplc.2020.100043

Xu, Y., Li, P., Zou, C., Lu, Y., Xie, C., Zhang, X., et al. (2017). Enhancing genetic gain in the era of molecular breeding. *J. Exp. Bot.* 68, 2641–2666. doi:10.1093/jxb/erx135

Yan, W. X., Chong, S., Zhang, H., Makarova, K. S., Koonin, E. V., Cheng, D. R., et al. (2018). Cas13d is a compact RNA-targeting type VI CRISPR effector positively modulated by a WYL-domain-containing accessory protein. *Mol. Cell* 70, 327–339. doi:10.1016/j.molcel.2018.02.028

Yang, B., Li, X., Lei, L., and Chen, J. (2017). Apobec: From mutator to editor. *J. Genet. Genomics* 44, 423–437. doi:10.1016/j.jgg.2017.04.009

Zaman, Q. U., Li, C., Cheng, H., and Hu, Q. (2019). Genome editing opens a new era of genetic improvement in polyploid crops. *Crop J.* 7, 141–150. doi:10.1016/j.cj.2018.07.004

Zambre, M., Goossens, A., Cardona, C., Van Montagu, M., Terryn, N., and Angenon, A. (2005). A reproducible genetic transformation system for cultivated *Phaseolus acutifolius* (tepary bean) and its use to assess the role of arcelins in resistance to the Mexican bean weevil. *Theor. Appl. Genet.* 110, 914–924. doi:10.1007/s00122-004-1910-7

Zanke, C. D., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., et al. (2014). Whole genome association mapping of plant height in winter wheat (*Triticum aestivum* L.). *PloS One* 9, e113287. doi:10.1371/journal.pone.0113287

Zdruli, P. (2014). Land resources of the mediterranean: Status, pressures, trends and impacts on future regional development. *Land Degrad. Dev.* 25, 373–384. doi:10.1002/ldr.2150

Zhang, C. L., Xu, S. Y., Belopolski, I., Yuan, Z., Lin, Z., Tong, B., et al. (2016). Signatures of the adler–bell–jackiw chiral anomaly in a weyl fermion semimetal. *Nat. Commun.* 7, 10735–10739. doi:10.1038/ncomms10735

Zhang, Y. C., Yu, Y., Wang, C. Y., Li, Z. Y., Liu, Q., Xu, J., et al. (2013). Overexpression of microRNA *OsmiR397* improves rice yield by increasing grain size and promoting panicle branching. *Nat. Biotechnol.* 31, 848–852. doi:10.1038/nbt.2646

Zhang, Y., Iaffaldano, B., and Qi, Y. (2021). CRISPR ribonucleoprotein-mediated genetic engineering in plants. *Plant Commun.* 2, 100168. doi:10.1016/j.xplc.2021.100168

Zhang, Y., Pribil, M., Palmgren, M., and Gao, C. (2020). A CRISPR way for accelerating improvement of food crops. *Nat. Food* 1, 200–205. doi:10.1038/s43016-020-0051-8

Zhou, W., Yang, J., Hong, Y., Liu, G., Zheng, J., Gu, Z., et al. (2015). Impact of amylose content on starch physicochemical properties in transgenic sweet potato. *Carbohydr. Polym.* 122, 417–427. doi:10.1016/j.carbpol.2014.11.003

Zhu, Q., Smith, S. M., Ayele, M., Yang, L., Jogi, A., Chaluvadi, S. R., et al. (2012). High-throughput discovery of mutations in tef semi-dwarfing genes by next-generation sequencing analysis. *Genetics* 192, 819–829. doi:10.1534/genetics.112.144436

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., et al. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* 35, 438–440. doi:10.1038/nbt.3811

Zuris, J. A., Thompson, D. B., Shu, Y., Guilinger, J. P., Bessen, J. L., Hu, J. H., et al. (2015). Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing *in vitro* and *in vivo*. *Nat. Biotechnol.* 33, 73–80. doi:10.1038/nbt.3081

Frontiers in Genetics

TYPE Review
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.989199

Check for updates

OPEN ACCESS

EDITED BY
Deepmala Sehgal,
Syngenta, United Kingdom

REVIEWED BY
Muntazir Mushtaq,
National Bureau of Plant Genetic Resources (ICAR), India
Rupesh Deshmukh,
Plaksha University, India

*CORRESPONDENCE
Palakolanu Sudhakar Reddy,
p.sudhakarreddy@cgiar.org,
palakolanusreddy@gmail.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

RECEIVED 08 July 2022
ACCEPTED 15 November 2022
PUBLISHED 04 January 2023

CITATION
Kor SD, Chowdhury N, Keot AK, Yogendra K, Chikkaputtaiah C and Sudhakar Reddy P (2023), RNA Pol III promoters—key players in precisely targeted plant genome editing.
Front. Genet. 13:989199.
doi: 10.3389/fgene.2022.989199

COPYRIGHT
© 2023 Kor, Chowdhury, Keot, Yogendra, Chikkaputtaiah and Sudhakar Reddy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

RNA Pol III promoters—key players in precisely targeted plant genome editing

Sakshi Dharmendra Kor[1†], Naimisha Chowdhury[2†], Ajay Kumar Keot[2,3], Kalenahalli Yogendra[1], Channakeshavaiah Chikkaputtaiah[2,3] and Palakolanu Sudhakar Reddy[1]*

[1]International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, Telangana, India, [2]Biological Sciences and Technology Division, CSIR-North East Institute of Science and Technology (CSIR-NEIST), Jorhat, Assam, India, [3]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India

The clustered regularly interspaced short palindrome repeat (CRISPR)/CRISPR-associated protein Cas) system is a powerful and highly precise gene-editing tool in basic and applied research for crop improvement programs. CRISPR/Cas tool is being extensively used in plants to improve crop yield, quality, and nutritional value and make them tolerant to environmental stresses. CRISPR/Cas system consists of a Cas protein with DNA endonuclease activity and one CRISPR RNA transcript that is processed to form one or several short guide RNAs that direct Cas9 to the target DNA sequence. The expression levels of Cas proteins and gRNAs significantly influence the editing efficiency of CRISPR/Cas-mediated genome editing. This review focuses on insights into RNA Pol III promoters and their types that govern the expression levels of sgRNA in the CRISPR/Cas system. We discussed Pol III promoters structural and functional characteristics and their comparison with Pol II promoters. Further, the use of synthetic promoters to increase the targeting efficiency and overcome the structural, functional, and expressional limitations of RNA Pol III promoters has been discussed. Our review reports various studies that illustrate the use of endogenous U6/U3 promoters for improving editing efficiency in plants and the applicative approach of species-specific RNA pol III promoters for genome editing in model crops like Arabidopsis and tobacco, cereals, legumes, oilseed, and horticultural crops. We further highlight the significance of optimizing these species-specific promoters' systematic identification and validation for crop improvement and biotic and abiotic stress tolerance through CRISPR/Cas mediated genome editing.

KEYWORDS

CRISPR/Cas9, RNA pol III promoters, U6 and U3 snRNA promoters, TATA-box, USE, synthetic promoter

Introduction

Promoters are the key regulatory elements present upstream of the transcription start site that controls the transcription of a gene through the involvement of the TFs and RNA polymerases. Promoters play a critical role in regulating gene expression that can be greatly modified by identifying and applying specific promoter systems, such as constitutive or inducible, for genetic manipulation of an organism for a desired trait/s. Promoters are classified as constitutive, tissue-specific, stage/temporal-specific, or inducible based on their ability to control gene expression (Kummari et al., 2020). However, recent advancements in transgene expression studies have led to the development of synthetic promoters consisting of repeats of *cis*-elements to drive the desired gene of interest. A synthetic promoter should be optimized for precise specificity, immediate inducibility, versatile applications, and efficient editing (Ali and Kim, 2019). Promoters are also classified as pol II and pol III, based on their ability to recognize RNA polymerases. The pol II promoter is the region that involves the binding of RNA polymerase II to initiate DNA transcription (Venter and Botha, 2010).

On the other hand, polymerase III aids the exclusive transcription of small non-coding RNAs, including 5S rRNA, tRNAs, and another type 3 RNAs such as the U6 snRNA (Cramer et al., 2008). The promoter elements are present internally in type 1 and type 2 genes of the polymerase III promoters. In contrast, the type 3 Pol III promoters typically utilize the upstream regulatory elements with a distinct +1 transcription start site and distinguished stretches of "thymine" as a termination signal (Schramm and Hernandez, 2002). Several studies have been conducted to understand the polymerase activity of the commonly used type 3 Pol III promoters, such as U6, 7SK, and H1. Recent studies of Gao et al. (2018), provide functional evidences of Pol II and Pol III competing for usage of promoter like human H1 promoter (Myslinski et al., 2001; Gao et al., 2018).

Nevertheless, type 3 Pol III promoters have found their application in the expression of small RNAs, like short hairpin RNAs in RNAi, and guide RNAs in the breakthrough CRISPR/Cas system. Typically, the Pol III type 3 promoters, like U6, comprises a ~21 bp proximal sequence element, and a ~8 bp TATA box located upstream of the transcription start site (+1) are reported to be conserved among species (Dahlberg and Lund, 1988). However, RNA polymerase specificity may be attributed to the minor sequence differences in the 3' end of the proximal sequence element (Hernandez et al., 2007).

U6 promoters have reportedly been used to drive small hairpin RNA (shRNA) expression in vector-based RNA interference (Nie et al., 2010), and in identifying and characterizing U6 promoters from the genome of *Plutella xylostella* for enabling genome editing in non-model organisms (Huang et al., 2017). However, both the U3 and U6 promoters have been highly exploited in plants for efficient guide RNA activity (Belhaj et al., 2013). The U3 and U6 promoters in plants have a discrete transcription start site with adenine (A) and guanine (G), respectively. Therefore, the consensus sequence of A(N)19-22 for the U3 promoter and G(N) 19–22 for the U6 promoter is considered ideal for designing the guide sequences of the sgRNAs (Belhaj et al., 2013). The U6 snRNAs contribute to the intron splicing of pre-mRNA in the nucleus (Li et al., 2007), while the U3 snRNAs are involved in pre-rRNA processing (Marz and Stadler, 2009).

The revolutionary platform of genome editing with CRISPR/Cas has unlocked opportunities to explore the genetic makeup of all plant species. The sole influence of the Cas protein and the single guide RNA can profoundly affect the editing efficiency of the CRISPR/Cas9 system. The plant promoters like U3 and U6 have established their place for driving the expression of single guide RNA due to their proficiency in producing high levels of sgRNA, with a length of ~200 nucleotides (Shockey, 2020). The commonly used promoters in plant genome editing are the *Arabidopsis* (AtU3 and AtU6) and rice (OsU3 and OsU6) promoters used specifically for dicots and monocots, respectively (Lowder et al., 2015; Ma et al., 2015). Moreover, studies on the applicability of species-specific Pol III promoters provided some significant insight into the improved editing efficiencies due to increased sgRNA expression (Sun et al., 2015; Ng and Dean, 2017; Long et al., 2018) (Figure 1).

These type 3 Pol III promoters have a dual polymerase activity, making their usage more attractive for the concurrent expression of a small RNA and a protein. This expression system can complement the CRISPR/Cas9 genome-editing system, which involves the Cas9 protein and the single guide RNA. Ren et al. (2022) proposed that using Pol III promoter would abate the complexities involving transgene cassettes and aid the construction of viral vectors with limited packaging capacity. Nevertheless, exploring the pol III promoters from different plant species and their characterization can lead to milestones in the field of CRISPR/Cas9 genome editing (Ren et al., 2022).

Structural features of Pol III promoters

In all eukaryotes, the genes are transcribed by one of the three RNA polymerases: RNA Pol I, II, and III. Each type of RNA polymerase performs a different function of transcribing. The RNA Pol I is responsible for transcribing a single set of ribosomal RNA genes with a single recognizable promoter structure. The second is RNA Pol II, which transcribes the protein-coding (mRNA) genes along with some small nuclear RNA (snRNA) genes, while the third class of polymerase, RNA Pol III, transcribes small, non-coding RNA set of genes like 5S rRNA and tRNA. RNA pol III transcribed genes which are involved in cellular metabolic processes like t-RNA processing, mRNA splicing, and protein synthesis (Schramm and Hernandez, 2002). The RNA Pol II

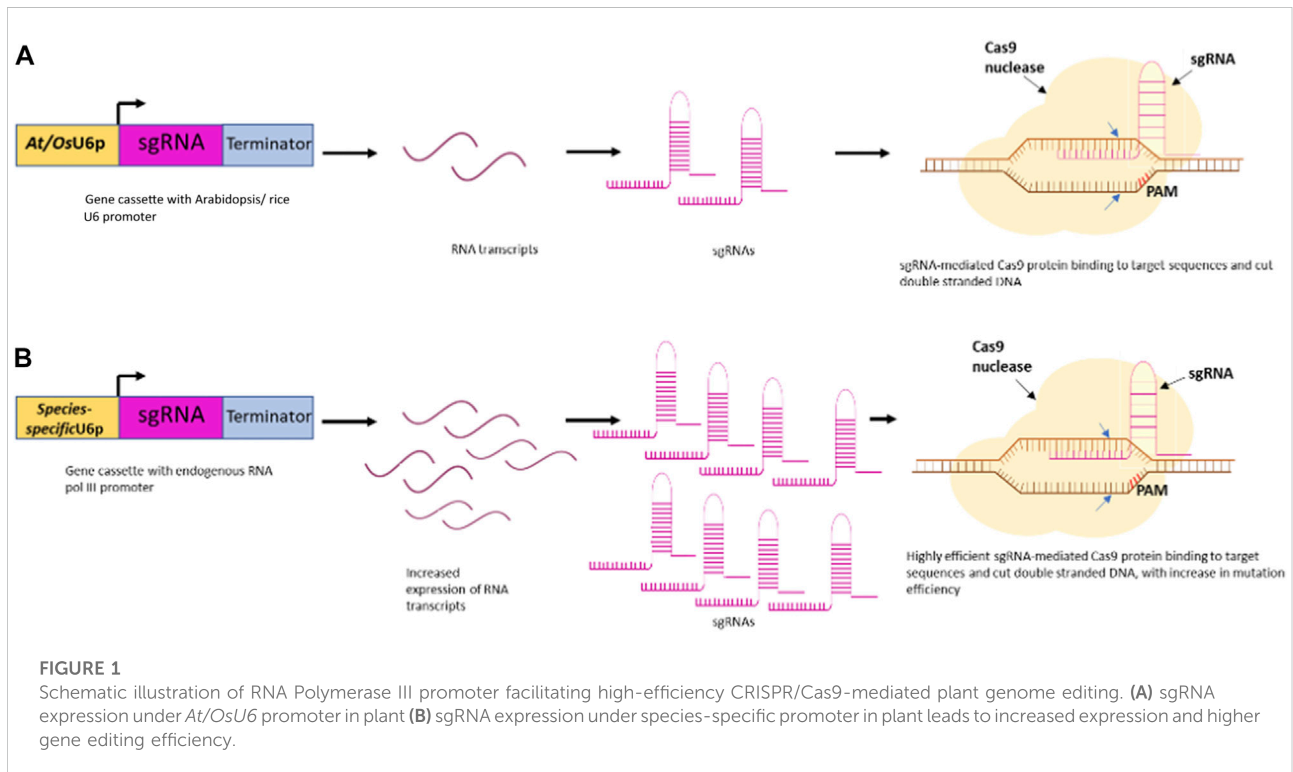


FIGURE 1
Schematic illustration of RNA Polymerase III promoter facilitating high-efficiency CRISPR/Cas9-mediated plant genome editing. **(A)** sgRNA expression under *At/OsU6* promoter in plant **(B)** sgRNA expression under species-specific promoter in plant leads to increased expression and higher gene editing efficiency.

and III promoters share common features, such as a similar TATA-box as a recognition site, while have some distinct features like the presence of poly-T (thymine) tail at 3′end as terminator whereas, Pol II promoters have *cis*-elements as their 3′-terminal signal and such Poly-T sites present endogenously in the sequence (Richard and Manley, 2009). There are three sub-classes of RNA Pol III promoter, namely type 1, 2, and 3, which are classified based on the position of the promoter with respect to the gene and the existence of the TATA-box. The type 1 and 2 promoters are gene-internal and TATA-less box, which assists in the transcription of 5S rRNA genes and Adenovirus (Ad2) VAI gene, tRNA set of genes, respectively (Fowlkes and Shenk, 1980; Bogenhagen et al., 1982; Schramm and Hernandez, 2002). The third class of Pol III promoter is distinct from the other two sub-classes by the presence of TATA-box and promoter sequence being present at the 5′ flanking ends or upstream to the transcription start site (+1TSS). This class of promoter includes U1 to U6, signal recognition particle (SRP), mitochondrial RNA processing (MRP) snRNAs, H1, etc. (Kunkel and Pederson, 1988; Baer et al., 1990; Topper and Clayton, 1990). The type 3 Pol III promoters have conserved regions as that of snRNA Pol II promoters, like Distal Sequence Element (DSE) and Proximal Sequence Element (PSE) in mammals or Upstream Sequence Element (USE) in plants, in addition to TATA box located at 30 bp upstream to the TSS (+1). The TATA-box has all required information to cluster together the elements for RNA Pol III transcription initiation (Mitchell et al., 1992; Roberts et al., 1995; Wang and Stumph, 1995; Schramm and Hernandez, 2002).

In plants, the two basal promoter elements required for Pol III transcribed snRNA genes are the −70 bp highly conserved plant snRNA gene-specific element, USE (consensus RTCCCACATCG) and −28 to −30 bp TATA-box (Figure 2A) (Marshallsay et al., 1990; Waibel and Filipowicz, 1990). The U6 and U3 snRNA gene promoters have the USE element placed one helical turn closer to the TATA box than that in Pol II specific genes, which have the USE and TATA box positioned four helical turns apart (Marshallsay et al., 1990). In dicots, the sequences present upstream to USE have no significance in snRNA gene transcription but, an extra element located upstream to the USE in monocots, known as the monocot-specific promoter element (MSP), increases the efficiency of transcription (Figure 2B). These MSP(s) (consensus, RGCCCR) is present in one to three copies in the monocot snRNA gene promoter region. In monocots, the efficiency of snRNA gene transcription is determined by the strength of the MSP element/s present in the promoter region while that of dicot is measured using the quality of the USE element (Marshallsay et al., 1992). The AT-rich region of RNA Pol III resembles the TATA box found in Pol II, but the AT-rich box distinguishes Pol III promoters from that of mRNA promoters (Pol II) by initiating the transcription in downstream of the “forward” TATA box, whereas transcription initiated by Pol III is in downstream of the “reverse” TATA sequence (Mattaj et al., 1988; Lobo and Hernandez, 1989; Waibel and Filipowicz, 1990; Wang and

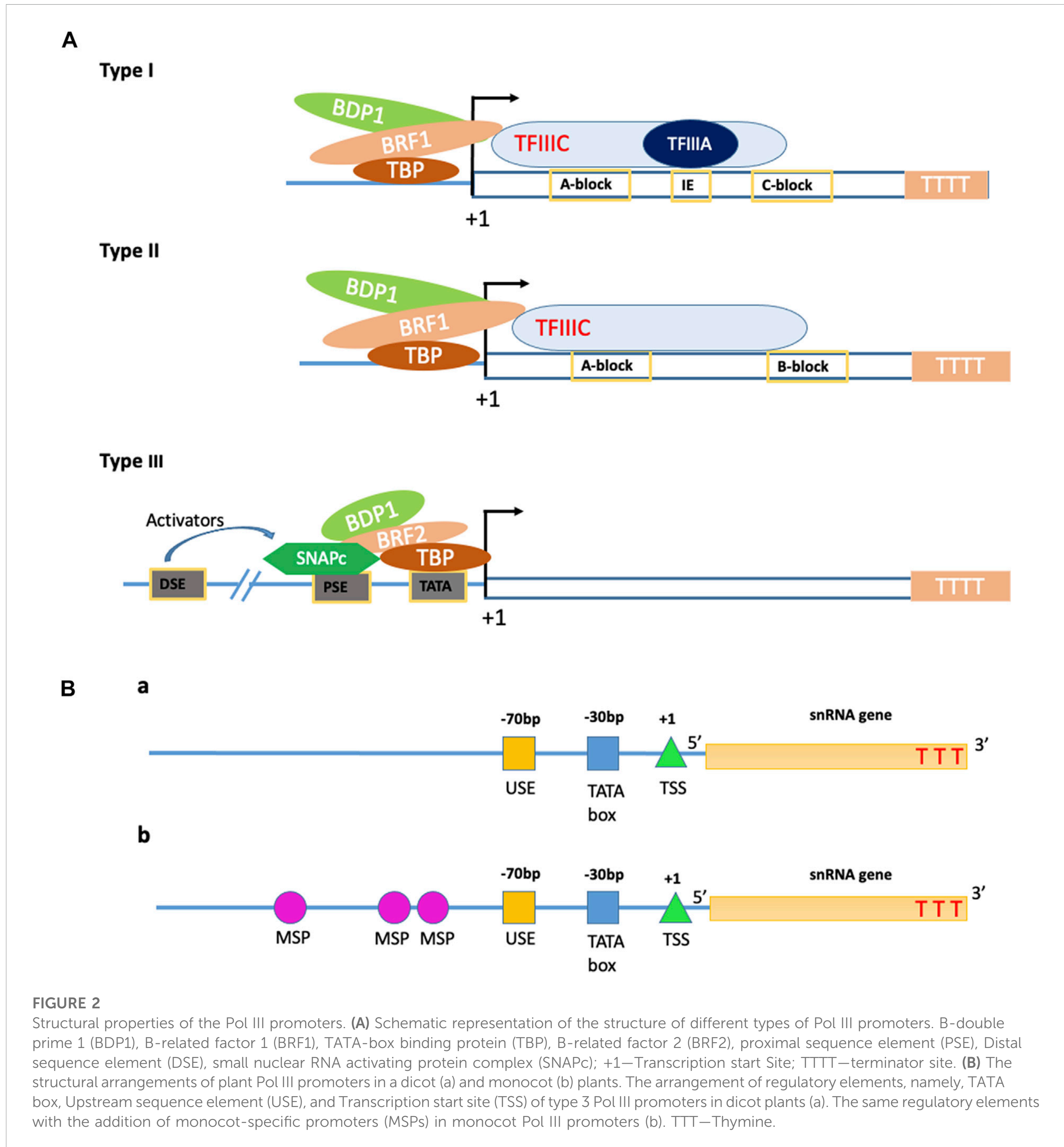


FIGURE 2
Structural properties of the Pol III promoters. **(A)** Schematic representation of the structure of different types of Pol III promoters. B-double prime 1 (BDP1), B-related factor 1 (BRF1), TATA-box binding protein (TBP), B-related factor 2 (BRF2), proximal sequence element (PSE), Distal sequence element (DSE), small nuclear RNA activating protein complex (SNAPc); +1—Transcription start Site; TTTT—terminator site. **(B)** The structural arrangements of plant Pol III promoters in a dicot (a) and monocot (b) plants. The arrangement of regulatory elements, namely, TATA box, Upstream sequence element (USE), and Transcription start site (TSS) of type 3 Pol III promoters in dicot plants (a). The same regulatory elements with the addition of monocot-specific promoters (MSPs) in monocot Pol III promoters (b). TTT—Thymine.

Stumph, 1995). The similarities and differences between RNA Pol II and Pol III promoters are given in Table 1.

Functional characteristics of the RNA Pol III promoters

The snRNA promoters have different characteristics which differentiate them from other classes of gene coding promoters. The U-snRNA class (U1, U2, U4, and U5) are transcribed by RNA Polymerase II, while genes like H1, U3, and U6 snRNAs are transcribed by RNA Polymerase III (Dahlberg and Lund, 1988). The genes encoding snRNAs in plants and vertebrates are unique in a way that some transcribed by Pol II and some by Pol III, but both classes of genes have similar promoter elements (Murphy et al., 1987; Filipowicz et al., 1990). These set of genes have their promoters for the recruitment of RNA polymerase III. In monocots and dicots, the Pol III promoters are used for

TABLE 1 Similarities and differences between RNA Pol II and RNA Pol III promoters.

RNA polymerase II	RNA polymerase III
It transcribes mRNA encoding genes, long and some small non-coding RNAs genes	It transcribes short non-coding RNA genes, like 5S rRNA, tRNAetc.
Pol II promoters have such Poly-T sites present internally	Pol III promoters have Poly-T sites at the 3′end
AT-rich region that resembles the TATA box in Pol II promoters	AT-rich box acts as TATA box to distinguish from Pol II promoter
RNA pol II is sensitive to 1 μg/ml of α-amanitin	RNA pol III is sensitive to 10 μg/ml of α-amanitin

transcription of *U3* and *U6* snRNA genes. These Pol III promoters are generally used in expression of small nuclear RNA, short hairpin RNA, and guide RNA in the CRISPR/Cas9 genome editing system (Ma et al., 2014). In most of the CRISPR/Cas9 constructs, the RNA polymerase III -type 3- U3 or U6 promoters are used for expression of sgRNA in monocots and dicots. Further, Pol III promoters are extensively used for expression of polycistronic tRNA-sgRNA construct involved in multiple gene-targeted genome editing (Jiang et al., 2013). These Pol III promoters need a very specific 5′ nucleotide, U6 promoter requires 5′-Guanine (G) and U3 needs 5′-Adenine (A) to start the transcription (Jiang et al., 2013). Thus, specificity can be increased by addition of specific nucleotide at 5′end of the target sequence or the gRNA sequences. These U6 and U3 promoters drive the expression of gRNAs in plants but may not always work for all targeted genes due to the absence of spatial and temporal specific control, as it is ubiquitously expressed in all tissues and at all stages of growth and development (Gao and Zhao, 2014; Xie et al., 2015).

CRISPR/Cas9 plant genome editing system uses two sets of RNA polymerases. Expression of Cas9 gene under RNA polymerase II promoter while the sgRNA cassette is driven by the RNA Polymerase III (U6 or U3 promoters) (Jiang et al., 2013). The two types of promoters control the co-expression of Cas9 and gRNA. While targeting the expression of multiple gRNAs in a single cassette, there will be a corresponding number of Pol III promoter sequences, further leading to increased cassette size, which is the limitation for cloning into the vectors (Li et al., 2013; Nekrasov et al., 2013; Shan et al., 2013). For efficient cloning of multiple guide RNAs, Collonnier et al. (2017) suggested using a different combination of promoters such as a U3 promoter for one sgRNA and a U6 promoter for the second sgRNA in the same vector backbone to avoid the hairpin structure formation and smooth DNA synthesis (Collonnier et al., 2017).

Activation of RNA polymerase III promoters

The transcription initiation leads to the polymerase complex formation in the promoter region. Protein factors or TFs that recognize the sequence motifs of RNA polymerase III transcribed genes are well studied in yeast and animals. The multi-subunit complex for activation of RNA polymerase III promoter includes TFIIIC, A and B boxes, TFIIIB, and TATA-binding protein (TBP) in yeast and vertebrates (Orioli et al., 2012). The recruitment of RNA Pol III to the promoter region in a plant is sketchily known. The Pre-initiation complex (PIC) assembly is formed about 50 bp upstream region of the transcription start site (TSS) of the pol III-transcribed gene. Thus, it is the prime surface of interaction with TFIIIB. The TFIIIB is responsible for the recruitment of the RNA polymerase III enzyme. The TBP is another protein involved in Pol III-dependent transcription and is a component of TFIIIB. The TATA box element of the U6 snRNA genes is a core promoter element for the transcription of RNA Pol III (Zhang et al., 2011). This TATA box governs the TSS selection by Pol III, with the aid of the A-box-bound τA sub-complex of TFIIIC. This τA is said to be extended within which TBP chooses the TATA-like sequence. This TATA-like sequences provide surface for assembly of TFIIIB and thus recruits TFIIIB and initiates the process of transcription downstream of 5′end of TATA element (Orioli et al., 2012).

The tCAAca sequence is another core promoter element in fungi, and plant Pol III transcribed genes are involved in TSS selection by RNA Polymerase III. In this tCAAca sequence, the uppercase letters indicate the least variable positions, and the first A is the TSS (Giuliodori et al., 2003; Yukawa et al., 2011; Zhang et al., 2011). The shared mechanistic characteristic between the tCAAca sequence and the initiator by RNA Polymerase II is the presence of two pyrimidines before A of TSS, this being the common feature between Pol II and Pol III transcription elements (Orioli et al., 2012).

Assembly of transcriptional initiation complex of Pol III promoters

RNA polymerase III is responsible for the bulk of transcriptional activity, including all the important non-coding RNAs (whole set of transfer RNAs, U6 spliceosomal RNA, and 5S ribosomal RNA) (Abascal-Palacios et al., 2018). The RNA polymerases III enzyme share numerous subunits with RNA polymerase II, but it identifies a different set of promoters with distinct transcription factor proteins (Shen, 2019). The most

notable and uncommon aspect of pol III promoters is that most of them require sequence components downstream of the transcription start site (+1), i.e., they have promoters entirely within the genes (Shen, 2019). The typical and classical RNA polymerase III promoters are type 1 (e.g., 5S rRNA gene), type 2 (e.g., tRNA gene), and type 3 (promoter of the *Homo sapiens* U6 snRNA) (Kummari et al., 2020).

Type 1 promoters of RNA polymerase III require two internal sequence elements for efficient transcription, an A block located from +50 to +70 and a C block from +80 to +90, and an intermediate element (IE) between blocks A and B (Figure 2A). Type 2 of RNA polymerase III promoters comprises two sequence blocks (A and B) present within the gene transcription region and are very conserved (Goodfellow and White 2005). Distinct from type 1 and 2, the type 3 promoter of RNA polymerase III (example: U6 snRNA gene) falls under the non-classical category. They have a transcription factor binding site upstream of the transcription start site (+1 site), and it consists of a TATA box (located amid −30 to −25 from +1 Site) and another upstream control sequence element named proximal sequence element (PSE), and finally upstream to PSE is a distal sequence element (DSE) (Goodfellow and White 2005; Arimbasseri and Maraia, 2016).

Transcription with RNA polymerase III takes 3 general steps: initiation, elongation, and termination. Positioning of eukaryotic RNA polymerase III enzyme to the transcription start site (TSS) requires many transcription factors that work synergistically. To initiate transcription on type 1 promoters, the RNA polymerases III complex relies on a different set of transcription factors (TFIIIA, TFIIIB, and TFIIIC) as it has less affinity for promoter sequence elements (Goodfellow and White 2005; Park et al., 2017). Briefly, TFIIIC interacts with internal promoter sequences (block A and B) and recruits the TFIIIB complex. TFIIIA binds specifically to the intermediate element (IE) of type 1 and recruits TFIIIC to its site and RNA polymerase to promote transcription initiation from the +1 site. In this case, the TFIIIB is a complex of three proteins, TBP (TATA-box binding protein), BDP1 (B double prime 1), and BRF1 (B-related factor 1) (Park et al., 2017). The assembly of transcription factors on type 2 promoter (e.g., tRNA) differ from that of type 1 promoters. The TFIIIC of the type 2 promoters (same set of protein as type 1) recognizes and binds to the A and B blocks of type 2 internal promoter and recruits the TFIIIB (B double prime 1, BDP1; B-related factor 1, BRF1; and TATA-box binding protein, TBP) and RNA polymerase to the transcription start site (Figure 2A) (Arimbasseri and Maraia, 2016; Park et al., 2017).

In the case of type 3 promoters (e.g., U6 snRNA gene), assembly of a transcription factor on the promoter sequence occurs upstream of the transcription start site (TSS). Here SNAPc (small nuclear RNA activating protein complex), an activating protein complex, binds to the upstream promoter element, proximal sequence element (PSE) to promote the TFIIIB recruitment and RNA polymerase III loading for transcription initiation. In this case, TFIIIB consist of TBP (TATA-box binding protein), BDP1 (B double prime 1), and BRF2 (B-related factor 2) (Arimbasseri and Maraia, 2016; Park et al., 2017).

RNA polymerase III has a high and steady small nuclear RNA transcriptional activity, accounting for approximately 40% of total RNA, which is validated by the fact that Pol III promoter has a primary role in RNA-guided genome editing strategies like CRISPR/Cas technology (Paule and White, 2000). The guide RNA used in the CRISPR/Cas9 technology is usually driven by RNA polymerase III (Ma et al., 2014). Another striking fact is that RNA polymerase III has defined sites for transcription initiation and termination, making them good candidates for genome editing techniques like CRISPR/Cas (Brummelkamp et al., 2002). Promoters like U6 and U3 are reported to work efficiently in plants where RNA Pol III transcribes them ubiquitously and constitutively to express guide RNAs (Li et al., 2013; Bortesi and Fischer, 2015). Types and features of RNA Pol III promoters are given in Table 2.

Synthetic RNA Pol III promoters

A synthetic promoter is a sequence of DNA which is artificially designed in order to regulate the expression of the target gene. The *cis*-regulatory element sequences of a promoter that exist in nature are used as fundamental blocks for synthesizing these artificial promoters. These can be created using rational design or ligation (Roberts, 2011). Synthetic promoters are important in studying the *cis*-motif elements' orientation, strength-weight, and position to understand gene regulation *in vivo*. These strategies can be used in designing of the expression cassette for target genes in genome editing technology (Venter and Botha, 2010). Hao et al. (2020) modified the active rice U3 and U6 promoters by shortening the 5′ sequences but retaining the USE and TATA box elements and the native MSPs, along with adding two to three artificial MSPs in the upstream region of USE to increase the transcriptional efficiency. Synthetic promoters were used to improve the efficiency of gene transcription for activating the *GUS* reporter gene in pco-dCas9-VP64 coupled with multiple sgRNAs (Lowder et al., 2015). These synthetic promoters were designed to check the functionality of the pco-dCas9-VP64 transcriptional activator and pco-dCas9-3X repressor (Lowder et al., 2015). The same strategy was used to develop the Orthogonal Control System (OCS) based on orthogonal synthetic promoters driven by the Artificial Transcription Factor (ATF). The constitutively expressing Pol III promoters can be synthetically controlled to express in a specific tissue, thus widening the use of OCS for targeted genome editing. The synthetic promoter needs its own transcription factor to be constructed and characterized (Kar et al., 2022). They preferentially drive the expression of Cas9, in Arabidopsis egg cell (Durr et al., 2018). Also, in yeast *Yarrowia*

TABLE 2 Types of RNA Pol III promoters.

	Type 1	Type 2	Type 3
Transcribing genes	5S rRNA	VAI gene	U3/U6 snRNA
Location of promoter with respect to gene	Gene-internal	Gene-internal	Gene-external
TATA-box	Absent	Absent	Present
Upstream Sequence Element (USE)	Absent	Absent	Present
Conserved domains	A, E and B boxes	A and B box	TATA-box and USE
Transcription factors	T.F IIIA, T.F IIIC	T.F IIIC	T.F IIIB, T.F IIIC

lipolytica, the single gene disruption efficiency of 92% and more was obtained due to synthetic hybrid promoters, RPR1′-tRNA^{Gly}, SCR1′- tRNA^{Gly} and SNR52′- tRNA^{Gly} under native RNA Pol III promoter (Schwartz et al., 2016). Löbs et al. (2017), used CRISPR/Cas9 system from *S. pyogenes* for *Kluyveromyces marxianus* genome editing using hybrid RNA Pol III promoters like RPR$_1$-tRNA^{Gly}, SCR1′- tRNA^{Gly} and SNR52′-tRNA^{Gly} hybrid promoters for knocking out alcohol dehydrogenase (ADH) and alcohol-O-acetyltransferase genes.

Applications of U6/U3 promoters in CRISPR/Cas-mediated genome editing

The plant species-specific Pol III promoters like U6 and U3 have been extensively used for increased sgRNA levels to achieve efficient editing using the CRISPR/Cas technology. In the last two decades, there have been several reports addressing the use of species-specific U3/U6 promoters in targeting certain traits in cereals, legumes, oilseeds, and horticultural crops, the details of which are discussed hereunder. Endogenously identified species-specific RNA Pol III promoters to enhance the genome editing efficiency are represented in Table 3.

In a study conducted by Li et al. (2013a), protoplast transient expression system was used for exploring the use of sgRNA: Cas9 technology. The plant codon-optimized SpCas9 and the sgRNAs was transcribed from the hybrid constitutive 35SPPDK promoter and the Arabidopsis U6 Polymerase III promoter, respectively. The sgRNAs were designed for targeting the *A. thaliana* genes viz., phytoene desaturase (*PDS3*), flagellin sensitive (*AtFLS2*), and the *Nicotiana benthamiana* ortholog of *AtPDS3* (*NbPDS*). Moreover, the authors targeted two members of the Arabidopsis RACK1 (Receptor for Activated C Kinase 1) family with multiple sgRNAs expressed under the U6 promoter, thereby ensuring targeted mutagenesis and gene knockout (Li et al., 2013a). In a contemporary study by Nekrasov et al. (2013), they used the sgRNA:Cas9 system for targeting the *PDS* gene in *N. benthamiana* with the sgRNA expressed under an Arabidopsis U6 promoter (Nekrasov et al., 2013). Similarly, the targeted mutation in *PDS* and *PDR*-type transporter genes was achieved through the CRISPR/Cas platform where the chimeric guide RNA was driven by the AtU6-26 promoter (Gao et al., 2015). Other research in Arabidopsis includes the targeted mutagenesis of endogenous DNA glycosylase genes *ROS1* and *DME* using sgRNA driven by the AtU6 promoter (Miki et al., 2018). Successful heritable homozygous mutations were also reported in the T_2 generation by using the Arabidopsis U6-26 promoter (Fauser et al., 2014; Feng et al., 2014).

Jiang et al. (2013) demonstrated the CRISPR/Cas9 mediated genome editing in immature embryos of sorghum where the sgRNA was expressed under the rice U6 promoter (Jiang et al., 2013). Shan et al. (2013) reported the design of two sgRNA, SP1 and SP2 for disrupting the rice phytoene desaturase gene *OsPDS* along with specific sgRNAs for targeting the *OsBADH2*, *Os02g23823*, and *OsMPK2* genes in rice using the rice endogenous U3 promoters (Shan et al., 2013). Moreover, the wheat U6 promoter was used to drive the sgRNA for targeting the wheat ortholog of barley MLO protein, *TaMLO*. In an aim to target the Maize *IPK* gene, involved in the phytic acid biosynthetic pathway, Liang et al. (2014) designed the sgRNA to express under the Maize U3 promoter. They confirmed the mutation of Inositol 1,3,4,5,6-pentakisphosphate 2-kinase gene in *Zea mays* using the CRISPR/Cas genome editing (Liang et al., 2014). Svitashev et al. (2016) reported DNA-free genome editing in maize by targeting four genes viz., male fertility genes (*MS26* and *MS45*), liguleless1 (*LIG*) and acetolactate synthase (*ALS2*). Under the expression of maize U6 promoter, the *in vitro* transcribed gRNAs and the purified Cas9 protein were pre-assembled to initiate the targeted mutagenesis (Svitashev et al., 2016). Very recently, Char et al. (2020), through CRISPR/Cas9 system demonstrated targeted mutagenesis in two endogenous genes of Sorghum, *SbFT* and *SbGA2ox5*, responsible for flowering time and plant height. The designed sgRNAs were driven by two different rice U6 promoters, and the induced mutations were passed on to the T_1 generation (Char et al., 2020). In another contemporary study by Liu et al. (2020), the efficiency to drive single-guide RNA in wheat was observed for three different promoters from rice (OsU6a) as well as wheat (TaU3 and TaU6), through *Agrobacterium*-mediated

TABLE 3 Endogenously identified species-specific RNA Pol III promoters to enhance genome editing efficiency.

Type	Plant	Common name	Promoter (U3/U6)	Target gene	References
Monocots	*Oryza sativa*	Rice	OsU3	*ADH2*	Mikami et al. (2015)
	Zea mays	Maize	ZmU3	Argonaute 18 and anthocyaninless genes	Char et al. (2017)
	Sorghum bicolor	Sorghum	SbU6	PDS, GDH7, kafirin, Apetela2	Massel et al. (2022)
	Triticum aestivum	Wheat	TaU6.1, TaU6.3	GFP	Zhang et al. (2019a)
	Musa acuminata	Banana	MaU6	PDS, Luciferase reporter	Zhang et al. (2022)
	Phyllostachys edulis	Mosa bamboo	PeU3	PDS	Huang et al. (2022)
Dicots	*Arabidopsis thaliana*	Arabidopsis	AtU6-1	*BON*	Li et al. (2014)
			AtU6-26	*St16DOX* and *SIIAA9*	Nakayasu et al., 2018; Ueta et al., 2017
			AtU6-29	EOD3	Khan et al. (2020)
			AtU3	*EOD3*	Khan et al. (2020)
	Nicotiana benthamiana	Tobacco	NbU6	*PDS*	Li et al. (2014)
	Camelina sativa	False flex or linseed dodder	CsU3	*FAD*	Morineau et al. (2017)
	Malus domestica	Apple	MdU3	*PDS*	Charrier et al. (2019)
	Fragaria vesca	Wild strawberry	FveU6-2	Auxin biosynthesis gene (TAA1), auxin response factor 8 (ARF8)	Zhou et al. (2018)
	Vitis vinifera	Grapevine	VvU3	PDS	Ren et al. (2022)
	M. domestica	Apple	MdU6	*PDS* and *TFL1*	Charrier et al. (2019)
	C. sativa	False flex or linseed dodder	CsU6	*FAD*	Morineau et al. (2017)
	Cichorium intybus	Chicory	CiU6-1	*PDS*	Bernard et al. (2019)
	Coffea canephora	Coffee tree	CcU6	*PDS*	Breitler et al. (2018)
	Vigna unguiculata	Cowpea	VuU6	*SPO11*, *Rec8* and *OSD1*	Juranic et al. (2020)
	Glycine max	Soyabean	GmU6	Glyma06g14180, Glyma08g02290, Glyma12g37050	Sun et al. (2015)
	Gossypium hirsutum	Cotton	GhU6 3.3	*PDS*	Long et al. (2018)
	Hevea brasiliensis	Rubber tree	HbU6	*PDS*	Dai et al. (2021)
	Lotus japonicus	Lotus	LjU6-1	*LjSYMRK*	Wang et al. (2016)
	Medicago truncatula	Alfalfa	MtU6	*PDS*	Meng et al. (2017)
	Picea glauca	White spruce	PaU6	*DXS1*	Cui et al. (2021)
Bryophyte	*Marchantia polymorpha*	Liverwort	MpU6-*1pro*	Auxin response factor (*AF1*)	Sugano et al. (2014)
Gymnosperm	*Cryptomeria japonica*	Japanese cedar	CjU6	*CjChll*	Nanasato et al. (2021)

transformation. TaU3 promoter was found to be a better choice than OsU6a or TaU6 for driving the expression of sgRNA in wheat. A high editing efficiency of 80.5% was achieved by the optimized SpCas9 system using TaU3 and two sgRNAs for targeted mutagenesis of two endogenous genes, *TaWaxy* (granule-bound starch synthase I) and *TaMTL* (MATRILINEAL, a pollen-specific phospholipase) (Liu et al., 2020). In the above usage of OsU6 or TaU6 promoter for

sorghum, genome editing can be replaced by recently identified endogenous sorghum SbU6 promoters by Massel et al. (2022). They identified eight putative SbU6 promoters in the BTx623 genome and selected SbU6_2.3 and SbU6_3.1 promoters to target *β-kafirin* (major grain storage protein). Using SbU6_2.3 resulted in 80.0% of the mutation rate in the *β-kafirin* gene. Thus, endogenous pol III promoter employment leads to a higher and more efficient CRISPR/Cas editing system. (Massel et al., 2022).

In an attempt to demonstrate the application potential of CRISPR/Cas9 in a forage crop like *Medicago truncatula*, Michno et al. (2015) successfully mutated a GUS transgene in somatic cells of *M. truncatula* through root hair transformation, where the target guide RNA was expressed under the Arabidopsis U6 promoter (Michno et al., 2015). In subsequent research, Meng et al. (2017) targeted the second exon of the *phytoene desaturase* (*MtPDS*) gene using a sgRNA under the effect of the native MtU6 promoter (Meng et al., 2017). The *symbiosis receptor-like kinase* (*SYMRK*) gene is crucial for nodule and arbuscular mycorrhizal symbiosis in legumes. Targeted disruption of three targets of exon 2 of the *VuSYMRK* in Cowpea (*Vigna unguiculata*) through the CRISPR/Cas9 system was carried out by Ji et al. (2019). The respective gRNAs were designed to be expressed under the U6 promoter, resulting in approximately 67% mutagenesis (Ji et al., 2019). The *SYMRK* gene was also targeted for mutagenesis in *Lotus japonicus* along with three homologous leghemoglobin loci (LjLb1, LjLb2, LjLb3), the designed guide RNAs of which were placed under the effect of LjU6-1 promoter (Wang et al., 2016). In yet another study, Chen et al. (2020) established an 'allele-aware chromosome-level genome assembly' genome editing protocol in *Medicago sativa* L. The expression of the sgRNAs targeting the *PDS* and *PALM1* (encoding a Cys(2)His(2) zinc finger transcription factor) genes was driven by the MtU6 Polymerase III promoter (Chen et al., 2020).

Genome modification in soybean was demonstrated by Jacobs et al. (2015) by targeting the transgene Green Fluorescent Protein, a putative glucosyl-transferase endogenous gene (Glyma07g14530), and the orthologs of the *A. thaliana DDM1* gene (Glyma01g38150 and Glyma11g07220) (Jacobs et al., 2015). The single guide RNAs were driven by the *M. truncatula* U6.6 Polymerase III promoter. Michno et al. (2015) performed the hairy root transformation in soybean, where they designed the guide RNA to target the *Glutamine synthase (GS1)* and *chalcone-flavanone isomerase* (*CHI20*) genes under the effect of the Arabidopsis U6 promoter (Michno et al., 2015). To address the problem of seed shattering from mature fruits in tetraploid oilseed rape (*Brassica napus*), Braatz et al. (2017) used CRISPR/Cas9 construct by targeting two homologs of the *ALCATRAZ (ALC)* gene. The sgRNA was placed under the control of the Arabidopsis U6-26 promoter, where a single target sequence generated four *alc* mutant alleles in an edited T_1 plant (Braatz et al., 2017). Contemporary studies with CRISPR/Cas9 gene editing also altered the fatty acid composition in *Camelina sativa* seeds by targeting *the FAD2* gene responsible for synthesizing polyunsaturated fatty acids. *C. sativa,* being an allohexaploid, the three homoeologous *FAD2* genes were targeted using the same sgRNA, which was driven by the Arabidopsis U6 promoter (Jiang et al., 2017). The same *FAD2* gene was modulated using the CRISPR/Cas9 system in *B. napus* cv. Westar and in peanut (*Arachis hypogaea* L.) using sgRNAs, driven by the Arabidopsis U6 promoter and *M. truncatula* U6 promoter, respectively (Okuzaki et al., 2018; Yuan et al., 2019). The enzyme lysophosphatidic acid acyltransferase (*LPAT*) aids the catalysis of fatty acid chains into 3-phosphoglycerate, thereby enhancing oil production. The *BnLPAT2* and *BnLPAT5* genes from *B. napus* were targeted using the Arabidopsis U6-26 promoter to drive the sgRNA expression further establishing their role in oil biosynthesis (Zhang K. et al., 2019). In another study, Di et al. (2019) analysed the effect of multiple *G. max* U6 promoters by targeting three genes, *Glyma03g36470*, *Glyma14g04180*, and *Glyma06g136900* through *Agrobacterium rhizogenes* infection, while Zhang Z. et al. (2019) tested the ECp-Cas9/gRNA system by targeting the *GmAGO7a* (Glyma.01G053100) and *GmAGO7b* (Glyma.02G111600) using the Arabidopsis U3 or U6 promoters to drive expression of each gRNA (Zhang Z. et al., 2019; Di et al., 2019).

Wang et al. (2015) used potato U6 RNA (StU6P) for initiating the expression of sgRNA *via Agrobacterium tumefaciens* mediated transient expression of *phytoene desaturase (PDS)* gene in *N. benthamiana.* They further transformed the CRISPR/Cas9 construct in potato to make stable mutations in the *StIAA2* gene encoding an Aux/IAA protein in potato (Wang et al., 2015). A contemporary study of genome editing was reported in cucumber (*Cucumis sativus* L.) by Chandrasekaran et al. (2016) (Chandrasekaran et al., 2016). Targeted disruption of the *eIF4E* gene (eukaryotic translation initiation factor 4E), was demonstrated through Cas9/sgRNA editing. Two sgRNAs, expressed under the effect of Arabidopsis U6 promoter, were designed to target two sites of the *eIF4E* gene. A successful CRISPR/Cas9 editing of the *flavanone-3-hydroxylase (F3H)* gene was performed in the carrot. Two single-guide RNA (gRNAs) was expressed in the CRISPR/Cas9 vectors under the effect of the Arabidopsis U3 promoter for obstructing the biosynthesis of anthocyanin (Klimek-Chodacka et al., 2018).

To study the effect of CRISPR/Cas9 genome editing in tomato, Brooks et al. (2014) constructed sgRNA for targeting the tomato homolog of Arabidopsis ARGONAUTE7 (*SlAGO7*) through *Agrobacterium*-mediated transformation. The sgRNAs, expressed under the effect of the Arabidopsis U6 promoter, were in duplicates in order to create large and well-defined deletions.

The mutant plants had needle-like or wiry leaves as compared to the compound leaves in wild-type tomatoes (Brooks et al., 2014). First study for genome editing in apple was reported by Nishitani et al. (2016). Precise modification in the apple *phytoene desaturase* (*PDS*) gene was confirmed by the use of four sgRNAs that functioned under the effect of Arabidopsis U6 promoter, which resulted in an approximately 13.6% editing efficiency (Nishitani et al., 2016). In a proof-of-concept study by Charrier et al. (2019), the *PDS* and *Terminal Flower 1 (TFL1)* genes were successfully knocked-out in apple. Two guide RNAs were expressed with U3 and U6 apple promoters for targeted editing. Successful editing in the *MdPDS* gene was confirmed by distinctive albino phenotype in about 85% of the edited lines, while early flowering was observed in 93% of the edited lines where the *MdTFL1* was targeted (Charrier et al., 2019). Successful editing of the *auxin biosynthesis* (*TAA1*) and *auxin response factor 8 (ARF8)* genes of wild strawberry *Fragaria vesca* was achieved by Zhou et al. (2018). Two promoters viz., wild strawberry U6 promoter (FveU6-2) and Arabidopsis U6 promoter (AtU6-26) drove the expression of the sgRNAs targeting the two genes and both were reported to create efficient genome editing (Zhou et al., 2018). On the other hand, Kaur et al. (2018) successfully demonstrated the application of the CRISPR/Cas9 system in banana cv. Rasthali. Single guide RNAs targeting two phytoene desaturase genes (*RAS-PDS1* and *RAS-PDS2*), expressed under the rice U3 promoter, created albino phenotype and abnormalities in growth of the edited plants (Kaur et al., 2018). But Zhang et al. (2022) used endogenous MaU6 promoter instead of OsU6 promoter and also used banana codon-optimized Cas9, which increased the mutation frequency four times. Thus, developing a foundation for DNA-free genome editing technology in banana plants (Zhang et al., 2022). In a recent study, the efficacy and efficiency of four *Vitis vinifera* U3 and U6 promoters and two UBQ promoters in precise targeting of grape phytoene desaturase (PDS) gene was established by Ren et al. (2022). Further, the AtU6 promoter was replaced by the VvU6 promoter, for targeting multiple sgRNA and developing a multiplex genome editing system in grapes. The concurrent editing of the two genes viz., *TMT1* and *TMT2* (*tonoplastic monosaccharide transporter*) demonstrated the successful editing in grapes (Ren et al., 2022).

CRISPR/Cas9 technology has also been applied in ornamental flowers like Petunia, which also serve as a model system for comparative research. Zhang et al. (2016) targeted the *PDS* gene, with the sgRNA driven by the Arabidopsis U6 promoter, to make precise deletion in homozygous chromosomal fragment of the target gene. The lignocellulose biosynthesis process involves five genes viz., C_3H, C_4H, *4CL*, *CCR*, and *IRX* encoding coumarate 3-hydroxylase, cinnamate 4-hydroxylase, 4-coumarate: coenzyme a ligase, cinnamoyl coenzyme a reductase, and irregular xylem5 respectively (Zhang et al., 2016). Kui et al. (2017) designed 3 pairs of sgRNA for each gene, which were driven by the OsU3 promoter, to successfully apply the CRISPR/Cas9 tool for genome editing in *Dendrobium officinale* (Kui et al., 2017). The hexaploid, *Chrysanthemum morifolium* is an important ornamental plant where Kishi-Kaboshi et al. (2017) attempted genome editing using the CRISPR/Cas9 system. They targeted four sites of the transgene *CpYGFP* (yellowish-green fluorescent protein gene from *Chiridius poppei*) with four sgRNA under the control of the Arabidopsis U6 promoter (Kishi-Kaboshi et al., 2017).

Nanasato et al., 2021, performed targeted mutagenesis in Japanese cedar and used endogenous *CjU6* promoter to knock out the reporter *GFP* gene. They also mutated the endogenous magnesium chelates subunit I (*CjChlI*) gene using the CjU6 promoter to obtain the albino phenotype (Nanasato et al., 2021). Also, Dai et al. (2021), in the same year, developed an efficient method to validate the functionality of sgRNAs in rubber tree using endogenous five HbU6 promoters and reported the first plasmid-mediated genome editing report in *Hevea brasiliensis via* CRISPR/Cas9 system. This study targeted the *PDS* and *flowering time* (*FT*) related genes. The first report of an immature embryo plant regeneration system and genetic transformation system in *Phyllostachys edulis*, a monopodial bamboo species using two PeU3 promoters and targeting the *PePDS1* and *PePDS2* genes. The usage of endogenous pol III promoters led to higher editing efficiency (35%–39%) than editing with the OsU3 promoter (Huang et al., 2022). White spruce is one of the major sources of timber and pulpwood, having high economic and ecological importance. Cui et al. (2021) successfully knocked out the *DXS1* gene using *the PaU6* promoter in CRISPR/Cas9 toolbox to produce a high frequency of chimerism (Cui et al., 2021).

Conclusion

This review on RNA Polymerase III promoters in plants has illustrated the importance of the type 3 RNA Pol III promoters in specifically creating mutations in targeted gene editing using CRISPR/Cas system. These U3/U6 promoters require mainly two elements for its activity viz., TATA-box and USE. Monocot-specific promoters require extra element upstream to USE known as MSPs to increase the transcriptional efficiency. Not much is known about the Pre-initiation complex (PIC) formation of snRNA Pol III promoters in plants thus, this area of research needs to be explored to understand more of the transcriptional factors and regulatory elements. The review highlights the U3 and U6 promoters and their application in various plant systems. Recent studies show that the use of endogenous RNA Pol III promoter that transcribes single or multiple guide RNAs in

CRISPR/Cas9 system have increased the editing efficiency. Therefore, the researchers, presently are aiming to identify the species-specific U3/U6 promoters and to broaden the understanding of transcriptional assembly for more specific and efficient genome editing.

Author contributions

PS and CC: conceptualized the idea, prepared the background information, literature survey and critically evaluated the manuscript. SK, NC, AK, KY, CC, and PS: prepared the manuscript. All authors read the final version of the manuscript, provided necessary suggestions and approved it for publication.

Acknowledgments

PS acknowledges financial support from the Core Research Grant (CRG) (File No. CRG/2019/004305) and YK for Start-up Research Grant (SRG) (File No. SRG/2021/000422) from the Science and Engineering Research Board (SERB), Govt. of India. CC acknowledges financial support from the Focused Basic Research (FBR) grant, Council of Scientific and Industrial Research (CSIR), Govt. of India, grant file No. MLP-07.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abascal-Palacios, G., Ramsay, E. P., Beuron, F., Morris, E., and Vannini, A. (2018). Structural basis of RNA polymerase III transcription initiation. *Nature* 553, 301–306. doi:10.1038/nature25441

Ali, S., and Kim, W. C. (2019). A fruitful decade using synthetic promoters in the improvement of transgenic plants. *Front. Plant Sci.* 10, 1433. doi:10.3389/fpls.2019.01433

Arimbasseri, A. G., and Maraia, R. J. (2016). RNA polymerase III advances: Structural and tRNA functional views. *Trends biochem. Sci.* 41, 546–559. doi:10.1016/j.tibs.2016.03.003

Baer, M., Nilsen, T. W., Costigan, C., and Altman, S. (1990). Structure and transcription of a human gene for H1 RNA, the RNA component of human RNase P. *Nucleic Acids Res.* 18, 97–103. doi:10.1093/nar/18.1.97

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., and Nekrasov, V. (2013). Plant genome editing made easy: Targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods* 9, 39–10. doi:10.1186/1746-4811-9-39

Bernard, G., Gagneul, D., dos Santos, H. A., Etienne, A., Hilbert, J. L., and Rambaud, C. (2019). Efficient genome editing using CRISPR/Cas9 technology in chicory. *Int. J. Mol. Sci.* 20, 1155. doi:10.3390/ijms20051155

Bogenhagen, D. F., Michael Wormington, W., and Brown, D. D. (1982). Stable transcription complexes of Xenopus 5S RNA genes: A means to maintain the differentiated state. *Cell* 28, 413–421. doi:10.1016/0092-8674(82)90359-2

Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* 33, 41–52. doi:10.1016/j.biotechadv.2014.12.006

Braatz, J., Harloff, H. J., Mascher, M., Stein, N., Himmelbach, A., and Jung, C. (2017). CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed rape (*Brassica napus*). *Plant Physiol.* 174, 935–942. doi:10.1104/pp.17.00426

Breitler, J. C., Dechamp, E., Campa, C., Zebral Rodrigues, L. A., Guyot, R., Marraccini, P., et al. (2018). CRISPR/Cas9-mediated efficient targeted mutagenesis has the potential to accelerate the domestication of *Coffea canephora*. *Plant Cell Tissue Organ Cult.* 134, 383–394. doi:10.1007/s11240-018-1429-2

Brooks, C., Nekrasov, V., Lipppman, Z. B., and Van Eck, J. (2014). Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiol.* 166, 1292–1297. doi:10.1104/pp.114.247577

Brummelkamp, T. R., Bernards, R., and Agami, R. (2002). Stable suppression of tumorigenicity by virus-mediated RNA interference. *Cancer Cell* 2, 243–247. doi:10.1016/S1535-6108(02)00122-8

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., et al. (2016). Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* 17, 1140–1153. doi:10.1111/mpp.12375

Char, S. N., Neelakandan, A. K., Nahampun, H., Frame, B., Main, M., Spalding, M. H., et al. (2017). An Agrobacterium-delivered CRISPR/Cas9 system for high-frequency targeted mutagenesis in maize. *Plant Biotechnol. J.* 15, 257–268. doi:10.1111/pbi.12611

Char, S. N., Wei, J., Mu, Q., Li, X., Zhang, Z. J., Yu, J., et al. (2020). An Agrobacterium-delivered CRISPR/Cas9 system for targeted mutagenesis in sorghum. *Plant Biotechnol. J.* 18, 319–321. doi:10.1111/pbi.13229

Charrier, A., Vergne, E., Dousset, N., Richer, A., Petiteau, A., and Chevreau, E. (2019). Efficient targeted mutagenesis in apple and first time edition of pear using the CRISPR-Cas9 system. *Front. Plant Sci.* 10, 40. doi:10.3389/fpls.2019.00040

Chen, H., Zeng, Y., Yang, Y., Huang, L., Tang, B., Zhang, H., et al. (2020). Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nat. Commun.* 11, 2494. doi:10.1038/s41467-020-16338-x

Collonnier, C., Guyon-Debast, A., Maclot, F., Mara, K., Charlot, F., and Nogué, F. (2017). Towards mastering CRISPR-induced gene knock-in in plants: Survey of key features and focus on the model *Physcomitrella patens*. *Methods* 121-122, 103–117. doi:10.1016/j.ymeth.2017.04.024

Cramer, P., Armache, K. J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., et al. (2008). Structure of eukaryotic RNA polymerases. *Annu. Rev. Biophys.* 37, 337–352. doi:10.1146/annurev.biophys.37.032807.130008

Cui, Y., Zhao, J., Gao, Y., Zhao, R., Zhang, J., and Kong, L. (2021). Efficient multi-sites genome editing and plant regeneration via somatic embryogenesis in *Picea glauca*. *Front. Plant Sci.* 12, 2198. doi:10.3389/fpls.2021.751891

Dahlberg, J. E., and Lund, E. (1988). The genes and transcription of the major small nuclear RNAs. *Struct. Funct. Major Minor Small Nucl. Ribonucleoprotein Part.*, 38–70. doi:10.1007/978-3-642-73020-7_2

Di, Y. H., Sun, X. J., Hu, Z., Jiang, Q. Y., Song, G. H., Zhang, B., et al. (2019). Enhancing the CRISPR/Cas9 system based on multiple GmU6 promoters in

soybean. *Biochem. Biophys. Res. Commun.* 519, 819–823. doi:10.1016/j.bbrc.2019.09.074

Dai, X., Yang, X., Wang, C., Fan, Y., Xin, S., Hua, Y., et al. (2021). CRISPR/Cas9-mediated genome editing in *Hevea brasiliensis*. *Ind. Crops Prod.* 164, 113418. doi:10.1016/j.indcrop.2021.113418

Durr, J., Papareddy, R., Nakajima, K., and Gutierrez-Marcos, J. (2018). Highly efficient heritable targeted deletions of gene clusters and non-coding regulatory regions in Arabidopsis using CRISPR/Cas9. *Sci. Rep.* 8, 1–11. doi:10.1038/s41598-018-22667-1

Fauser, F., Schiml, S., and Puchta, H. (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant J.* 79, 348–359. doi:10.1111/tpj.12554

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D. L., et al. (2014). Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 111, 4632–4637. doi:10.1073/pnas.1400822111

Filipowicz, W., Kiss, T., Marshallsay, C., and Waibel, F. (1990). U-snRNA genes, U-snRNAs and U-snRNPs of higher plants. *Mol. Biol. Rep.* 14, 125–129. doi:10.1007/BF00360443

Fowlkes, D. M., and Shenk, T. (1980). Transcriptional control regions of the adenovirus VAI RNA gene. *Cell* 22, 405–413. doi:10.1016/0092-8674(80)90351-7

Gao, J., Wang, G., Ma, S., Xie, X., Wu, X., Zhang, X., et al. (2015). CRISPR/Cas9-mediated targeted mutagenesis in *Nicotiana tabacum*. *Plant Mol. Biol.* 87, 99–110. doi:10.1007/s11103-014-0263-0

Gao, Y., and Zhao, Y. (2014). Self-processing of ribozyme-flanked RNAs into guide RNAs *in vitro* and *in vivo* for CRISPR-mediated genome editing. *J. Integr. Plant Biol.* 56, 343–349. doi:10.1111/jipb.12152

Gao, Z., Herrera-Carrillo, E., and Berkhout, B. (2018). RNA polymerase II activity of type 3 Pol III promoters. *Mol. Therapy-Nucleic Acids* 12, 135–145. doi:10.1016/j.omtn.2018.05.001

Giuliodori, S., Percudani, R., Braglia, P., Ferrari, R., Guffanti, E., Ottonello, S., et al. (2003). A composite upstream sequence motif potentiates tRNA gene transcription in yeast. *J. Mol. Biol.* 333, 1–20. doi:10.1016/j.jmb.2003.08.016

Goodfellow, S. J., and White, R. J. (2005). "RNA polymerase III," in *Encyclopedic reference of Genomics and proteomics in molecular medicine* (Berlin, Heidelberg: Springer). doi:10.1007/3-540-29623-9_2140

Hao, Y., Zong, W., Zeng, D., Han, J., Chen, S., Tang, J., et al. (2020). Shortened snRNA promoters for efficient CRISPR/Cas-based multiplex genome editing in monocot plants. *Sci. China. Life Sci.* 63, 933–935. doi:10.1007/s11427-019-1612-6

Hernandez, G., Valafar, F., and Stumph, W. E. (2007). Insect small nuclear RNA gene promoters evolve rapidly yet retain conserved features involved in determining promoter activity and RNA polymerase specificity. *Nucleic Acids Res.* 35, 21–34. doi:10.1093/nar/gkl982

Huang, Y., Wang, Y., Zeng, B., Liu, Z., Xu, X., Meng, Q., et al. (2017). Functional characterization of Pol III U6 promoters for gene knockdown and knockout in *Plutella xylostella*. *Insect biochem. Mol. Biol.* 89, 71–78. doi:10.1016/j.ibmb.2017.08.009

Huang, B., Zhuo, R., Fan, H., Wang, Y., Xu, J., Jin, K., et al. (2022). An efficient genetic transformation and CRISPR/Cas9-Based genome editing system for moso bamboo (*Phyllostachys edulis*). *Front. Plant Sci.* 13, 822022–822111. doi:10.3389/fpls.2022.822022

Jacobs, T. B., LaFayette, P. R., Schmitz, R. J., and Parrott, W. A. (2015). Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol.* 15, 16–10. doi:10.1186/s12896-015-0131-2

Ji, J., Zhang, C., Sun, Z., Wang, L., Duanmu, D., and Fan, Q. (2019). Genome editing in cowpea *Vigna unguiculata* using CRISPR-cas9. *Int. J. Mol. Sci.* 20, 2471–2484. doi:10.3390/ijms20102471

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., and Weeks, D. P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res.* 41 (20), e188. doi:10.1093/nar/gkt780

Jiang, W. Z., Henry, I. M., Lynagh, P. G., Comai, L., Cahoon, E. B., and Weeks, D. P. (2017). Significant enhancement of fatty acid composition in seeds of the allohexaploid, *Camelina sativa*, using CRISPR/Cas9 gene editing. *Plant Biotechnol. J.* 15, 648–657. doi:10.1111/pbi.12663

Juranić, M., Nagahatenna, D. S. K., Salinas-Gamboa, R., Hand, M. L., Sánchez-León, N., Leong, W. H., et al. (2020). A detached leaf assay for testing transient gene expression and gene editing in cowpea (*Vigna unguiculata* [L.] Walp.). *Plant Methods* 16, 88–17. doi:10.1186/s13007-020-00630-4

Kar, S., Bordiya, Y., Rodriguez, N., Kim, J., Gardner, E. C., Gollihar, J. D., et al. (2022). Orthogonal control of gene expression in plants using synthetic promoters and CRISPR-based transcription factors. *Plant Methods* 18, 42–15. doi:10.1186/s13007-022-00867-1

Kaur, N., Alok, A., Kaur, N., Pandey, P., Awasthi, P., Tiwari, S., et al. (2018). CRISPR/Cas9-mediated efficient editing in phytoene desaturase (*PDS*) demonstrates precise manipulation in banana *cv.* Rasthali genome. *Funct. Integr. Genomics* 18, 89–99. doi:10.1007/s10142-017-0577-5

Khan, M. H. U., Hu, L., Zhu, M., Zhai, Y., Khan, S. U., Ahmar, S., et al. (2020). Targeted mutagenesis of EOD3 gene in *Brassica napus* L. regulates seed production. *J. Cell. Physiol.* 236, 1996–2007. doi:10.1002/jcp.29986

Kishi-Kaboshi, M., Aida, R., and Sasaki, K. (2017). Generation of gene-edited *Chrysanthemum morifolium* using multicopy transgenes as targets and markers. *Plant Cell Physiol.* 58, 216–226. doi:10.1093/pcp/pcw222

Klimek-Chodacka, M., Oleszkiewicz, T., Lowder, L. G., Qi, Y., and Baranski, R. (2018). Efficient CRISPR/Cas9-based genome editing in carrot cells. *Plant Cell Rep.* 37, 575–586. doi:10.1007/s00299-018-2252-2

Kui, L., Chen, H., Zhang, W., He, S., Xiong, Z., Zhang, Y., et al. (2017). Building a genetic manipulation tool box for orchid biology: Identification of constitutive promoters and application of CRISPR/Cas9 in the orchid, *dendrobium officinale*. *Front. Plant Sci.* 7, 2036–2049. doi:10.3389/fpls.2016.02036

Kummari, D., Palakolanu, S. R., Kishor, P. B. K., Bhatnagar-Mathur, P., Singam, P., Vadez, V., et al. (2020). An update and perspectives on the use of promoters in plant genetic engineering. *J. Biosci.* 45, 119–124. doi:10.1007/s12038-020-00087-6

Kunkel, G. R., and Pederson, T. (1988). Upstream elements required for efficient transcription of a human U6 RNA gene resemble those of U1 and U2 genes even though a different polymerase is used. *Genes Dev.* 2, 196–204. doi:10.1101/gad.2.2.196

Li, X., Jiang, D. H., Yong, K., and Zhang, D. B. (2007). Varied transcriptional efficiencies of multiple Arabidopsis U6 small nuclear RNA genes. *J. Integr. Plant Biol.* 49, 222–229. doi:10.1111/j.1744-7909.2007.00393.x

Li, J. F., Norville, J. E., Aach, J., McCormack, M., Zhang, D., Bush, J., et al. (2013a). Multiplex and homologous recombination-mediated genome editing in Arabidopsis and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat. Biotechnol.* 31, 688–691. doi:10.1038/nbt.2654

Li, W., Teng, F., Li, T., Zhou, Q., Chen, K., Liang, Z., et al. (2013b). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 686–688. doi:10.1038/nbt.2650

Li, J. F., Zhang, D., and Sheen, J. (2014). *Cas9-based genome editing in arabidopsis and tobacco. 1st ed.* Elsevier. doi:10.1016/B978-0-12-801185-0.00022-2

Liang, Z., Zhang, K., Chen, K., and Gao, C. (2014). Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. *J. Genet. Genomics* 41, 63–68. doi:10.1016/j.jgg.2013.12.001

Liu, H., Wang, K., Jia, Z., Gong, Q., Lin, Z. S., Du, L. P., et al. (2020). Efficient induction of haploid plants in wheat by editing of TaMTL using an optimized Agrobacterium-mediated CRISPR system. *J. Exp. Bot.* 71, 1337–1349. doi:10.1093/jxb/erz529

Lobo, S. M., and Hernandez, N. (1989). A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter. *Cell* 58, 55–67. doi:10.1016/0092-8674(89)90402-9

Löbs, A. K., Engel, R., Schwartz, C., Flores, A., and Wheeldon, I. (2017). CRISPR-Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. *Biotechnol. Biofuels* 10, 1–16. doi:10.1186/s13068-017-0854-5

Long, L., Guo, D. D., Gao, W., Yang, W. W., Hou, L. P., Ma, X. N., et al. (2018). Optimization of CRISPR/Cas9 genome editing in cotton by improved sgRNA expression. *Plant Methods* 14, 85–89. doi:10.1186/s13007-018-0353-0

Lowder, L. G., Zhang, D., Baltes, N. J., Paul, J. W., Tang, X., Zheng, X., et al. (2015). A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. *Plant Physiol.* 169, 971–985. doi:10.1104/pp.15.00636

Ma, H., Wu, Y., Dang, Y., Choi, J. G., Zhang, J., and Wu, H. (2014). Pol III promoters to express small RNAs: Delineation of transcription initiation. *Mol. Ther. Nucleic Acids* 3, e161. doi:10.1038/mtna.2014.12

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., et al. (2015). A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Mol. Plant* 8, 1274–1284. doi:10.1016/j.molp.2015.04.007

Marshallsay, C., Kiss, T., and Filipowicz, W. (1990). Amplification of plant U3 and U6 snRNA gene sequences using primers specific for an upstream promoter element and conserved intragenic regions. *Nucleic Acids Res.* 18, 3459–3466. doi:10.1093/nar/18.12.3459

Marshallsay, C., Connelly, S., and Filipowicz, W. (1992). Characterization of the U3 and U6 snRNA genes from wheat: U3 snRNA genes in monocot plants are transcribed by RNA polymerase III. *Plant Mol. Biol.* 19, 973–983. doi:10.1007/BF00040529

Marz, M., and Stadler, P. F. (2009). Comparative analysis of eukaryotic U3 snoRNA. *RNA Biol.* 6, 503–507. doi:10.4161/rna.6.5.9607

Massel, K., Lam, Y., Hintzsche, J., Lester, N., Botella, J. R., and Godwin, I. D. (2022). Endogenous U6 promoters improve CRISPR/Cas9 editing efficiencies in Sorghum bicolor and show potential for applications in other cereals. *Plant Cell Rep.* 41, 489–492. doi:10.1007/s00299-021-02816-z

Mattaj, I. W., Dathan, N. A., Parry, H. D., Carbon, P., and Krol, A. (1988). Changing the RNA polymerase specificity of U snRNA gene promoters. *Cell* 55, 435–442. doi:10.1016/0092-8674(88)90029-3

Meng, Y., Hou, Y., Wang, H., Ji, R., Liu, B., Wen, J., et al. (2017). Targeted mutagenesis by CRISPR/Cas9 system in the model legume *Medicago truncatula*. *Plant Cell Rep.* 36, 371–374. doi:10.1007/s00299-016-2069-9

Michno, J. M., Wang, X., Liu, J., Curtin, S. J., Kono, T. J., and Stupar, R. M. (2015). CRISPR/Cas mutagenesis of soybean and *Medicago truncatula* using a new web-tool and a modified Cas9 enzyme. *Gm. Crops Food* 6, 243–252. doi:10.1080/21645698.2015.1106063

Mikami, M., Toki, S., and Endo, M. (2015). Comparison of CRISPR/Cas9 expression constructs for efficient targeted mutagenesis in rice. *Plant Mol. Biol.* 88, 561–572. doi:10.1007/s11103-015-0342-x

Miki, D., Zhang, W., Zeng, W., Feng, Z., and Zhu, J. K. (2018). CRISPR/Cas9-mediated gene targeting in Arabidopsis using sequential transformation. *Nat. Commun.* 9, 1967–1969. doi:10.1038/s41467-018-04416-0

Mitchell, M. T., Hobson, G. M., and Benfield, P. A. (1992). TATA box-mediated polymerase III transcription *in vitro*. *J. Biol. Chem.* 267, 1995–2005. doi:10.1016/s0021-9258(18)46045-0

Morineau, C., Bellec, Y., Tellier, F., Gissot, L., Kelemen, Z., Nogué, F., et al. (2017). Selective gene dosage by CRISPR-Cas9 genome editing in hexaploid *Camelina sativa*. *Plant Biotechnol. J.* 15, 729–739. doi:10.1111/pbi.12671

Murphy, S., Di Liegro, C., and Melli, M. (1987). The *in vitro* transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. *Cell* 51, 81–87. doi:10.1016/0092-8674(87)90012-2

Myslinski, E., Amé, J. C., Krol, A., and Carbon, P. (2001). An unusually compact external promoter for RNA polymerase III transcription of the human H1RNA gene. *Nucleic Acids Res.* 29, 2502–2509. doi:10.1093/nar/29.12.2502

Nakayasu, M., Akiyama, R., Lee, H. J., Osakabe, K., Osakabe, Y., Watanabe, B., et al. (2018). Generation of α-solanine-free hairy roots of potato by CRISPR/Cas9 mediated genome editing of the St16*DOX* gene. *Plant Physiol. biochem.* 131, 70–77. doi:10.1016/j.plaphy.2018.04.026

Nanasato, Y., Mikami, M., Futamura, N., Endo, M., Nishiguchi, M., Ohmiya, Y., et al. (2021). CRISPR/Cas9-mediated targeted mutagenesis in Japanese cedar (*Cryptomeria japonica* D. Don). *Sci. Rep.* 11, 1–12. doi:10.1038/s41598-021-95547-w

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G., and Kamoun, S. (2013). Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 31, 691–693. doi:10.1038/nbt.2655

Ng, H., and Dean, N. (2017). Dramatic improvement of CRISPR/Cas9 editing in Candida albicans by increased single guide RNA expression. *mSphere* 2, e00385–16. doi:10.1128/mSphere.00385-16

Nie, L., Thakur, M. D., Wang, Y., Su, Q., Zhao, Y., and Feng, Y. (2010). Regulation of U6 promoter activity by transcriptional interference in viral vector-based RNAi. *Genomics Proteomics Bioinforma.* 8, 170–179. doi:10.1016/S1672-0229(10)60019-8

Nishitani, C., Hirai, N., Komori, S., Wada, M., Okada, K., Osakabe, K., et al. (2016). Efficient genome editing in apple using a CRISPR/Cas9 system. *Sci. Rep.* 6, 31481–31488. doi:10.1038/srep31481

Okuzaki, A., Ogawa, T., Koizuka, C., Kaneko, K., Inaba, M., Imamura, J., et al. (2018). CRISPR/Cas9-mediated genome editing of the fatty acid desaturase 2 gene in *Brassica napus*. *Plant Physiol. biochem.* 131, 63–69. doi:10.1016/j.plaphy.2018.04.025

Orioli, A., Pascali, C., Pagano, A., Teichmann, M., and Dieci, G. (2012). RNA polymerase III transcription control elements: Themes and variations. *Gene* 493, 185–194. doi:10.1016/j.gene.2011.06.015

Park, J-L., Lee, Y-S., Kunkeaw, N., Kim, S-Y., Kim, I-H., and Lee, Y. S. (2017). Epigenetic regulation of noncoding RNA transcription by mammalian RNA polymerase III. *Epigenomics* 9 (2), 171–187. doi:10.2217/epi-2016-0108

Paule, M. R., and White, R. J. (2000). Survey and summary: Transcription by RNA polymerases I and III. *Nucleic Acids Res.* 28, 1283–1298. doi:10.1093/nar/28.6.1283

Ren, C., Liu, Y., Guo, Y., Duan, W., Fan, P., Li, S., et al. (2022). Optimizing the CRISPR/Cas9 system for genome editing in grape by using grape promoters. *Hortic. Res.* 8, 52. doi:10.1038/s41438-021-00489-z

Richard, P., and Manley, J. L. (2009). Transcription termination by nuclear RNA polymerases. *Genes Dev.* 23, 1247–1269. doi:10.1101/gad.1792809

Roberts, M. L. (2011). "The use of functional genomics in synthetic promoter design," in *Compt. Biol. Appl. Bioinform. Intech*. Editor H. S. Lopes, 375–396.

Roberts, S., Colbert, T., and Hahn, S. (1995). TFIIIC determines RNA polymerase III specificity at the TATA-containing yeast U6 promoter. *Genes Dev.* 9, 832–842. doi:10.1101/gad.9.7.832

Schramm, L., and Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. *Genes Dev.* 16, 2593–2620. doi:10.1101/gad.1018902

Schwartz, C. M., Hussain, M. S., Blenner, M., and Wheeldon, I. (2016). Synthetic RNA polymerase III promoters facilitate high-efficiency CRISPR-cas9-mediated genome editing in Yarrowia lipolytica. *ACS Synth. Biol.* 5, 356–359. doi:10.1021/acssynbio.5b00162

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 686–688. doi:10.1038/nbt.2650

Shen, C. H. (2019). Gene expression: Transcription of the genetic code. *Diagn. Mol. Biol.*, 59–86. doi:10.1016/b978-0-12-802823-0.00003-1

Shockey, J. (2020). Gene editing in plants : Assessing the variables through a simplified case study. *Plant Mol. Biol.* 103, 75–89. doi:10.1007/s11103-020-00976-2

Sugano, S. S., Shirakawa, M., Takagi, J., Matsuda, Y., Shimada, T., Hara-nishimura, I., et al. (2014). CRISPR/Cas9-mediated targeted mutagenesis in the liverwort Marchantia polymorpha L. *Plant Cell Physiol.* 55, 475–481. doi:10.1093/pcp/pcu014

Sun, X., Hu, Z., Chen, R., Jiang, Q., Song, G., Zhang, H., et al. (2015). Targeted mutagenesis in soybean using the CRISPR-Cas9 system. *Sci. Rep.* 5, 1–10. doi:10.1038/srep10342

Svitashev, S., Schwartz, C., Lenderts, B., Young, J. K., Cigan, A. M., and Mark CigAn, A. (2016). Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. *Nat. Commun.* 7, 13274–13277. doi:10.1038/ncomms13274

Topper, J. N., and Clayton, D. A. (1990). Characterization of human MRP/Th RNA and its nuclear gene: Full length MRP/Th RNA is an active endoribonuclease when assembled as an RNP. *Nucleic Acids Res.* 18, 793–799. doi:10.1093/nar/18.4.793

Ueta, R., Abe, C., Watanabe, T., Sugano, S. S., Ishihara, R., Osakabe, Y., et al. (2017). Rapid breeding of parthenocarpic tomato plants using CRISPR/Cas9. *Sci. Rep.* 7, 507–508. doi:10.1038/s41598-017-00501-4

Venter, M., and Botha, F. C. (2010). Synthetic promoter engineering. *Plant Dev. Biol.* 2, 393–414. doi:10.1007/978-3-642-04670-4

Waibel, F., and Filipowicz, W. (1990). U6 snRNA genes of Arabidopsis are transcribed by RNA polymerase III but contain the same two upstream promoter elements as RNA polymerase II-transcribed U-snRNA genes. *Nucleic Acids Res.* 18, 3451–3458. doi:10.1093/nar/18.12.3451

Wang, S., Zhang, S., Wang, W., Xiong, X., Meng, F., and Cui, X. (2015). Efficient targeted mutagenesis in potato by the CRISPR/Cas9 system. *Plant Cell Rep.* 34, 1473–1476. doi:10.1007/s00299-015-1816-7

Wang, L., Wang, L., Tan, Q., Fan, Q., Zhu, H., Hong, Z., et al. (2016). Efficient inactivation of symbiotic nitrogen fixation related genes in *Lotus japonicus* using CRISPR-cas9. *Front. Plant Sci.* 7, 1333. doi:10.3389/fpls.2016.01333

Wang, Y. A. N., and Stumph, W. E. (1995). RNA polymerase II/III transcription specificity determined by TATA box orientation. *Proc. Natl. Acad. Sci. U. S. A.* 92, 8606–8610. doi:10.1073/pnas.92.19.8606

Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci.* 112, 3–8. doi:10.1073/pnas.1420294112

Yuan, M., Zhu, J., Gong, L., He, L., Lee, C., Han, S., et al. (2019). Mutagenesis of *FAD2* genes in peanut with CRISPR/Cas9 based gene editing. *BMC Biotechnol.* 19, 1–7. doi:10.1186/s12896-019-0516-8

Yukawa, Y., Dieci, G., Alzapiedi, M., Hiraga, A., Hirai, K., Yamamoto, Y. Y., et al. (2011). A common sequence motif involved in selection of transcription start sites of Arabidopsis and budding yeast tRNA genes. *Genomics* 97, 166–172. doi:10.1016/j.ygeno.2010.12.001

Zhang, G., Lukoszek, R., Mueller-Roeber, B., and Ignatova, Z. (2011). Different sequence signatures in the upstream regions of plant and animal tRNA genes shape distinct modes of regulation. *Nucleic Acids Res.* 39, 3331–3339. doi:10.1093/nar/gkq1257

Zhang, B., Yang, X., Yang, C., Li, M., and Guo, Y. (2016). Exploiting the CRISPR/Cas9 system for targeted genome mutagenesis in Petunia. *Sci. Rep.* 6, 20315–20318. doi:10.1038/srep20315

Zhang, K., Nie, L., Cheng, Q., Yin, Y., Chen, K., Qi, F., et al. (2019a). Biotechnology for Biofuels Effective editing for lysophosphatidic acid acyltransferase 2/5 in allotetraploid rapeseed (*Brassica napus* L .) using CRISPR - Cas9 system. *Biotechnol. Biofuels* 1, 1–18. doi:10.1186/s13068-019-1567-8

Zhang, Z., Hua, L., Gupta, A., Tricoli, D., Edwards, K. J., Yang, B., et al. (2019b). Development of an Agrobacterium -delivered CRISPR/Cas9 system for wheat genome editing. *Plant Biotechnol. J.* 17, 1623–1635. doi:10.1111/pbi.13088

Zhang, S., Wu, S., Hu, C., Yang, Q., Dong, T., Sheng, O., et al. (2022). Increased mutation efficiency of CRISPR/Cas9 genome editing in banana by optimized construct. *PeerJ* 9, 12664. doi:10.7717/peerj.12664

Zhou, J., Wang, G., and Liu, Z. (2018). Efficient genome editing of wild strawberry genes, vector development and validation. *Plant Biotechnol. J.* 16, 1868–1877. doi:10.1111/pbi.12922

| Frontiers in Genetics

TYPE Review
PUBLISHED 06 January 2023
DOI 10.3389/fgene.2022.866976

Check for updates

OPEN ACCESS

EDITED BY
Deepmala Sehgal,
Syngenta, United Kingdom

REVIEWED BY
Keunsub Lee,
Iowa State University, United States
Kalyani Makarand Barbadikar,
Indian Institute of Rice Research (ICAR), India

*CORRESPONDENCE
Shafat Ahmad Ahanger,
shafatahanger99@gmail.com
Sajad Ali,
sajadmicro@yu.ac.kr
Muntazir Mushtaq,
muntazirbt@gmail.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

RECEIVED 31 January 2022
ACCEPTED 21 November 2022
PUBLISHED 06 January 2023

CITATION
Basu U, Riaz Ahmed S, Bhat BA, Anwar Z, Ali A, Ijaz A, Gulzar A, Bibi A, Tyagi A, Nebapure SM, Goud CA, Ahanger SA, Ali S and Mushtaq M (2023), A CRISPR way for accelerating cereal crop improvement: Progress and challenges. *Front. Genet.* 13:866976.
doi: 10.3389/fgene.2022.866976

COPYRIGHT
© 2023 Basu, Riaz Ahmed, Bhat, Anwar, Ali, Ijaz, Gulzar, Bibi, Tyagi, Nebapure, Goud, Ahanger, Ali and Mushtaq. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A CRISPR way for accelerating cereal crop improvement: Progress and challenges

Umer Basu[1], Syed Riaz Ahmed[2†], Basharat Ahmad Bhat[3†], Zunaira Anwar[2†], Ahmad Ali[4], Aqsa Ijaz[2], Addafar Gulzar[5], Amir Bibi[6], Anshika Tyagi[7†], Suresh M. Nebapure[1], Chengeshpur Anjali Goud[8], Shafat Ahmad Ahanger[5]*, Sajad Ali[7]* and Muntazir Mushtaq[9]*

[1]Division of Entomology, ICAR-Indian Agricultural Research Institute, New Delhi, India, [2]Nuclear Institute for Agriculture and Biology College, Pakistan Institute of Engineering and Applied Sciences (NIAB-C, PIEAS), Faisalabad, Pakistan, [3]Department of Bioresources, University of Kashmir, Srinagar, India, [4]National Key Laboratory of Crop Genetic Improvement, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China, [5]Division of Plant Pathology, Faculty of Agriculture, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Wadura Sopore, India, [6]Department of Plant Breeding and Genetics, Faculty of Agriculture Sciences, University of Agriculture Faisalabad, Faisalabad, Pakistan, [7]Department of Biotechnology, Yeungnam University, Gyeongsan, South Korea, [8]Institute of Biotechnology, Professor Jayashanker Telangana State Agriculture University, Hyderabad, India, [9]ICAR-National Bureau of Plant Genetic Resources, Division of Germplasm Evaluation, Pusa Campus, New Delhi, India

Humans rely heavily on cereal grains as a key source of nutrients, hence regular improvement of cereal crops is essential for ensuring food security. The current food crisis at the global level is due to the rising population and harsh climatic conditions which prompts scientists to develop smart resilient cereal crops to attain food security. Cereal crop improvement in the past generally depended on imprecise methods like random mutagenesis and conventional genetic recombination which results in high off targeting risks. In this context, we have witnessed the application of targeted mutagenesis using versatile CRISPR-Cas systems for cereal crop improvement in sustainable agriculture. Accelerated crop improvement using molecular breeding methods based on CRISPR-Cas genome editing (GE) is an unprecedented tool for plant biotechnology and agriculture. The last decade has shown the fidelity, accuracy, low levels of off-target effects, and the high efficacy of CRISPR technology to induce targeted mutagenesis for the improvement of cereal crops such as wheat, rice, maize, barley, and millets. Since the genomic databases of these cereal crops are available, several modifications using GE technologies have been performed to attain desirable results. This review provides a brief overview of GE technologies and includes an elaborate account of the mechanisms and applications of CRISPR-Cas editing systems to induce targeted mutagenesis in cereal crops for improving the desired traits. Further, we describe recent developments in CRISPR-Cas–based targeted mutagenesis through base editing and prime editing to develop resilient cereal crop plants, possibly providing new dimensions in the field of cereal crop genome editing.

KEYWORDS

CRISPR/Cas, cereals, food security, genome editing, crop improvement, TALENs, base editing, prime editing

Introduction

Natural disasters and climate change have significantly harmed our agricultural systems in recent decades. In the near future, agriculture may face immense challenges in feeding a population, that is, likely to rise to 9 billion by 2050. Therefore, in order to fulfill the requirements of food supply to the global human community, scientific communities have largely focused on modern technological interventions to modify major crops for improved yield and resilient qualities. Cereal crops are considered to be the main energy and protein source for humans because they provide approximately 50% of dietary energy globally, especially in developing countries, where the contribution is higher (Borrill, 2020; Poutanen et al., 2022). Cereals are high in dietary fiber and contain adequate carbohydrates, protein, lipids, fats, vitamins, and minerals. With these nutritional values, health benefits, and production, cereals have been a staple in our diet since the establishment of agriculture farming. Therefore, cereals are vital to global food and nutritional security. Abiotic and biotic stresses are the most devastating factors for cereal crop production, affecting all growth stages and posing serious threat to global food security. In this regard, modern developments in GE technology have accelerated a transition to precision breeding for crop improvement, by making selective and precise genetic alterations in crops. A series of technologies known as GE enable researchers to change any DNA. GE makes it possible to change, add, or remove a particular sequence from the genome of any living organism. Homologous recombination is the basis of genome engineering; however, its occurrence at low frequencies limits the editing efficiency (Chen et al., 2019). To improve the editing frequency, researchers have improved the utility of programmable endonucleases that generate DNA double-stranded breaks (DSB) at target sites. The evolution of various GE technologies such as transcription activator-like effector nucleases (TALEN) and zinc finger nuclease (ZFN) have been previously used in the intended modification of human, animal, and plant cell genomes (Shukla et al., 2009; Zhang et al., 2010; Davies et al., 2017; Adli, 2018; Kannan et al., 2018; Manghwar et al., 2019; Arimura et al., 2020; Yasumoto et al., 2020; Dong and Ronald, 2021). ZFN is a site-specific GE approach which combines the DNA binding domains of zinc-fingers (ZFs) with the restriction endonucleases *Fok*I (Kim et al., 1997). ZF domains that have been custom engineered are important for site-specific mutagenesis (Carroll, 2008; Urnov et al., 2010). ZFs have been developed to target unique DNA sequences at specific loci in order to decrease off-target effects since ZFs have been implicated in target site recognition and binding efficacy for a wide range of DNA sequences. *Fok*Tandem array of Cys_2-His_2 zinc fingers (ZFNs) have been generated, with each unit containing −30 amino acids bound to a single atom of zinc; each domain aids in the recognition of, and binding to particular nucleotide triplets in the target sequence. Combining several ZFs to generate an array of DNA-binding ZFs might improve the affinity and selectivity for recognition of target DNA sequences (Choo and Isalan, 2000; Pabo et al., 2001; Segal et al., 2003). Although the use of ZFN for GE is accompanied by numerous limitations, it has been successfully used in cereals such as maize (Shukla et al., 2009) and rice (Cantos et al., 2014). Moreover, Sangamo BioSciences and Sigma Aldrich (licensed ZFN provider) have designed ZFNs with minimal off-targeting, as illustrated by the efficacy and specificity of the ZFNs to several crop plants including corn, canola, and wheat (Davies et al., 2017). Owing to its complexity and off-targeting, ZFN-based gene editing entails optimization of the stability, and targeted mutagenesis devoid of any off-target risks. Interestingly, specific sequence nucleases such as TALENs and Cas9 have been proposed, with a simplistic construct design and superior efficiencies than ZFN.

TALENs are synthetic hybrid proteins comprising a TALE DNA-binding domain linked to a *Fok*I nuclease domain (Zhang et al., 2015). TALEs are proteins containing a DNA-binding domain made up of a string of tandem repeats that are secreted by plant bacterial pathogens of the genus *Xanthomonas* after infection of the host (Mak et al., 2013). Each domain is made up of a sequence of 33–35 repeating amino acids that differ significantly at positions 12 and 13, exhibiting hypervariability. The 13th amino acid is responsible for interactions with a specific DNA base, and the 12th amino acid stabilizes this bonding (Deng et al., 2012). These sites are referred to as RVDs (repeat variable diresidues) (Mak et al., 2013; Wei et al., 2013; Lau et al., 2014). The RVD type and order (in the TALE repeat) determine the target specificity of TALE. If the TALE repeats are interchanged with different RVDs, this results in novel specificities. In 2009, two separate research groups demonstrated that RVDs were accountable for the attachment of certain nucleotides at the TALE target site, in accordance with a simple code (Boch et al., 2009; Moscou and Bogdanove, 2009). Each RVD identifies a 1-bp sequence rather than the 3-bp motif identified by zinc fingers, hence the sequence specificity of TALENs may be designed more accurately than in ZFNs (Son and Park., 2022). The RVDs Asn-Ile (NI), Asn-Asn (NN), His-Asp (HD), and Asn-Gly (NG) recognize the nucleotides A, G/A, C, and T, respectively (Christian et al., 2012). In accordance with the DNA-binding selectivity code, TALEs can be customized to attach to any arbitrary DNA sequence and joined to the endonuclease site of *Fok*I to produce a TALEN. To create site-specific DNA double strand breaks (DSBs), twin TALENs

addressing sense and antisense strands are required, as *Fok*I requires the formation of a dimer for DNA cleavage. Site-specific indel mutations can be induced by localized DSBs through the error-prone non-homologous end-joining (NHEJ) repair pathway. By using a sister chromatid or an external homologous DNA template, homologous recombination becomes another method for repairing DSBs that enables highly precise editing such as the insertion or replacement of genes at the target areas. However, due to their reliance on a restricted number of loci, the production of specialized enzymes, the high cost of a particular protein domain assembly, the usage of specific monomers in vector creation, and the accompanying off-target implications, ZFN and TALEN have become obsolete. Moreover, as molecular biology and plant breeding have changed dramatically, new CRISPR tools (CRISPR/Cpf1, prime editing, and base editing) have been created to modify the genomes of plants accurately, effectively, and swiftly (Tan et al., 2019; Haroon et al., 2022).

The CRISPR/Cas9 system leverages RNA-guided DNA cleavage to execute genome editing and is extremely efficient compared with prior genome editing systems such as ZFNs and TALENs, which rely on protein-guided sequence-specific DNA recognition and cleavage (He and Zhao, 2020; Li et al., 2020; Mushtaq et al., 2021b). Prior genome editing CRISPR-Cas reagents such as sgRNA, Cas proteins, and DNA must be delivered to the plants. Transfection of protoplasts, biolistic transformations, or *Agrobacterium*-mediated processes are used as delivery systems. CRISPR systems may be categorized into two classes, each of which possesses six types and 19 sub-types (Shmakov et al., 2017). Class 2 systems have become mainstream in genome editing technology because they require a single Cas protein, whereas Class 1 systems employ a multi-subunit Cas complex. Hence, the most explored and utilized method is a Class 2, type II CRISPR/Cas9 system, which uses a single Cas protein from *Streptococcus pyogenes* (SpCas9). Cas9 is an endonuclease that was identified in *S. pyogenes* which possesses RuvC and HNH nuclease domains. Its cleavage specificity is determined by CRISPR RNA (crRNA), formed from a CRISPR array that encapsulates short segments of foreign DNA molecules encountered by the bacteria. This CRISPR system is then transformed into tiny crRNAs that drive Cas9 to the target sequence (such as foreign DNA), resulting in Cas9-directed cleavage of both non-target and target DNA strands inside the crRNA-target DNA complex. In this mechanism, trans-activating crRNA (tracrRNA), which acts as a connection between crRNA and Cas9, is required for maturation of the crRNA. The CRISPR/Cas9 system in *S. pyogenes* has been curtailed to just two components: Cas9 and a small RNA. A single-stranded, single-guide RNA (sgRNA) emulates the crRNA:tracrRNA duplex, and exhibits a unique 20-bp sequence before the adjacent protospacer motif (PAM) with the sequence NGG, which is required for Cas9 compatibility (Zhang et al., 2017). The sgRNA and Cas9 complex attaches to a specific target site present on genomic DNA, permitting the complex to cleave the complementary site, resulting in a double-stranded DNA break (DSB) (Son and Park., 2022). Following the creation of a DSB, two main paths exist: non-homologous end-joining (NHEJ) and homologous recombination (HR) (Moore and Habber, 1996; Song et al., 2021). Since NHEJ-mediated knockouts provide a very precise and efficient method of suppressing genes of interest, the CRISPR/Cas9 system is ideally attuned for plant breeding. Homology directed repair (HDR) can be employed in scientific research and agriculture for gene substitution, protein tagging, and gene stacking (Nouspikel, 2009; Malzahn et al., 2017).

The current review provides an in-depth understanding of GE technologies and their role in cereal crop improvement. A very deep insight into the applicability, precision, and efficiency of CRISPR/Cas GE techniques is offered. Moreover, we have also profoundly discussed recent advances in genome engineering through an understanding of base editing and prime editing as forefront technologies for crop improvement.

Technical prelude to evolution of GE technologies

Meganucleases, sometimes called homing endonucleases, are restriction enzymes almost always found in all microorganisms. Meganucleases were the first class of endonucleases used from 1970 to 1980 to produce site-specific double-strand DNA breaks (Jacquier and Dujon, 1985). These hybrid restriction enzymes, which bind the cleavage domain *Fok*I to a customized zinc-finger protein (ZFP), have been utilized to introduce a range of unique changes to eukaryotic cell genomes. They are known to recognize and cleave specific DNA sequences (18–30 bp) to produce double-strand breaks. The resultant double-strand DNA breaks lead to a wide range of DNA modifications such as point mutation, deletion, or insertions (Daboussi et al., 2015). This class of endonucleases is not highly efficient in recognizing site-specific sequences. A challenge for engineering meganucleases is the overlap of cleavage and DNA binding domains. If the sequence of an amino acid is altered in order to gain novel DNA sequence specificity, the catalytic activity of the enzyme is often compromised. However, in recent years, scientists have made tremendous efforts in engineering a variety of meganucleases to cleave specific DNA targets. Nowadays, a number of the engineered meganucleases are used to create genomic modifications in crops for agronomically important traits (Daboussi et al., 2015). ZFNs are hybrid endonucleases and powerful GE tools to introduce double-strand breaks (DSBs) in target genomes, which is usually followed by error-prone non-homologous end joining (NHEJ) repair to create insertions or deletions at the cleavage site. The first report of GE by ZFN in plants was described by knocking-in a herbicide tolerance gene *via* disruption at the *Inositol Phosphokinase1* (*IPK1*) locus to

TABLE 1 ZFNs and TALENs for nutritional and phenotype improvement in cereals.

Crop	Gene editor	Gene targeted	Improvement	Method	References
Rice (*Oryza sativa*)	ZFNs	*OsQQR*	Trait stacking	HDR	Cantos et al. (2014)
	TALEN	*OsDEP1*, *OsBADH2*, *OsCKX2*, and *OsSD1*	Gene knockout	NHEJ	Shan et al. (2013)
	TALEN	*OsBADH2*	Fragrant rice	NHEJ	Shan et al. (2015)
	1.1.1 TALEN	*OsMST7* and *OsMST8*	Gene knockout	NHEJ	Zhang et al. (2016)
	TALEN	*Os11N3* (*OsSWEET14*)	Disease resistance	NHEJ	Li et al. (2012)
Maize (*Zea mays*)	ZFNs	*ZmIPK1*	Herbicide tolerant and phytate-reduced maize	HDR	Shukla et al. (2009)
	ZFNs	*ZmTLP*	Trait stacking	HDR	Ainley et al. (2013)
	TALEN	*ZmMTL*	Induction of haploid plants	NHEJ	Kelliher et al. (2017)
	TALEN	*ZmIPK1A*, *ZmIPK*, and *ZmMRP4*	Phytic acid synthesis	NHEJ	Liang et al. (2014)
	TALEN	*ZmGL2*	Reduced epicuticular wax in leaves	NHEJ	Char et al. (2015)
Barley (*Hordeum vulgare*)	TALEN	*HvPAPhy-a*	Phytase activity	NHEJ	Wendt et al. (2013)
	TALEN	*BAR*	Bialaphos resistance	—	Gurushidze et al. (2014)

purposefully reduce inorganic phosphate levels in growing seeds as part of an effort to minimize phytate levels in plants (Shukla et al., 2009). The *SSIVa* locus was altered in transgenic rice, impacting grain fullness, starch content, and plant height (Jung, et al., 2018). Several reports demonstrate the successful application of ZFN to modify, add, and disrupt plant genes (Durai et al., 2005; Papworth et al., 2006). Imidazolinone herbicide resistance was accomplished by GE, based on the use of ZFNs in allohexaploid wheat to target the *acetohydroxyacid synthase* (*AHAS*) encoding gene (Ran et al., 2018).

For precise genome editing, TALENs have been used instead of ZFNs due to their simple assembly, high success rate, availability of powerful resources, and decreased off-targeting. The discoveries of the Transcription Activator-Like Effector (TALE-DNA binding domains) and TALENs (TALE nucleases) were important breakthroughs in the field of genetic engineering. TALENs have allowed scientists to create double stand DNA breaks, introducing DNA modifications, gene knockout, and gene knock-in. The speed and ease of creating TALEN reagents has made it possible for a large number of labs to make target-specific alterations in genes of interest, cells, or organisms using the available transformation methods (Cermak et al., 2015). TALEs (TAL effectors) can be virulence factors, plant-recognized avirulence factors, or both (Bogdanove et al., 2010). These proteins imitate transcription factors once they attach to the DNA sequence and can control the activation of target gene(s) (Becker and Boch. 2021; Saurabh, 2021). Researchers decided to use it as a tool for gene editing by creating two hybrid TALE nucleases, each containing a DBD and the catalytic domain *Fok*I. This hybrid chimeric nuclease attaches to DNA and produces double-strand DNA breaks (DSBs). The majority of these DSBs are fixed by the NHEJ mechanism with insertions or deletions (indels), leading to an altered genome. The use of TALENs for genome editing was shown to be effective in cereal plants, including maize (Yu et al., 2014; Kelliher et al., 2017), rice (Shan et al., 2015), wheat (Zong et al., 2017; Luo et al., 2019), barley (Gurushidze et al., 2014), and other cereal crops. Applications of ZFNs and TALENs in phenotypic and nutritional enhancement are summarized in tabulated form (Table 1). However, there are some limitations associated with TALENs such as problems in editing a methylated target site, successful transmission with a vector, off-target effects, non-specific binding ability, and large size, necessitating further development of this technology (Pennisi, 2013; Mahfouz et al., 2014; Mushtaq et al., 2018; Razzaq et al., 2019; Ansari et al., 2020).

The flexibility of ZF and TALE DNA binding domains allows them to assemble or reprogram in a specific fashion and to recognize a particular site in the targeted genome, provided a significant advantage to ZFN and TALEN tools for genetic engineering, compared with the CRISPR-Cas9 GE system (Musunuru, 2017). In plant GE for agricultural enhancement, these two techniques have been widely employed (Forsyth et al., 2016; Ran et al., 2018; Shan et al., 2018). However, due to off-target occurrences, tedious build designs, poor efficiency, and expensive cost, their applicability to plant GE has been confined (Cermak et al., 2011; Puchta, 2017; Khan., 2019). Because of these constraints, a new, low cost, precise, and specific technology called CRISPR/Cas9 was developed as a flexible tool for biological studies to understand gene functions and crop enhancement (He

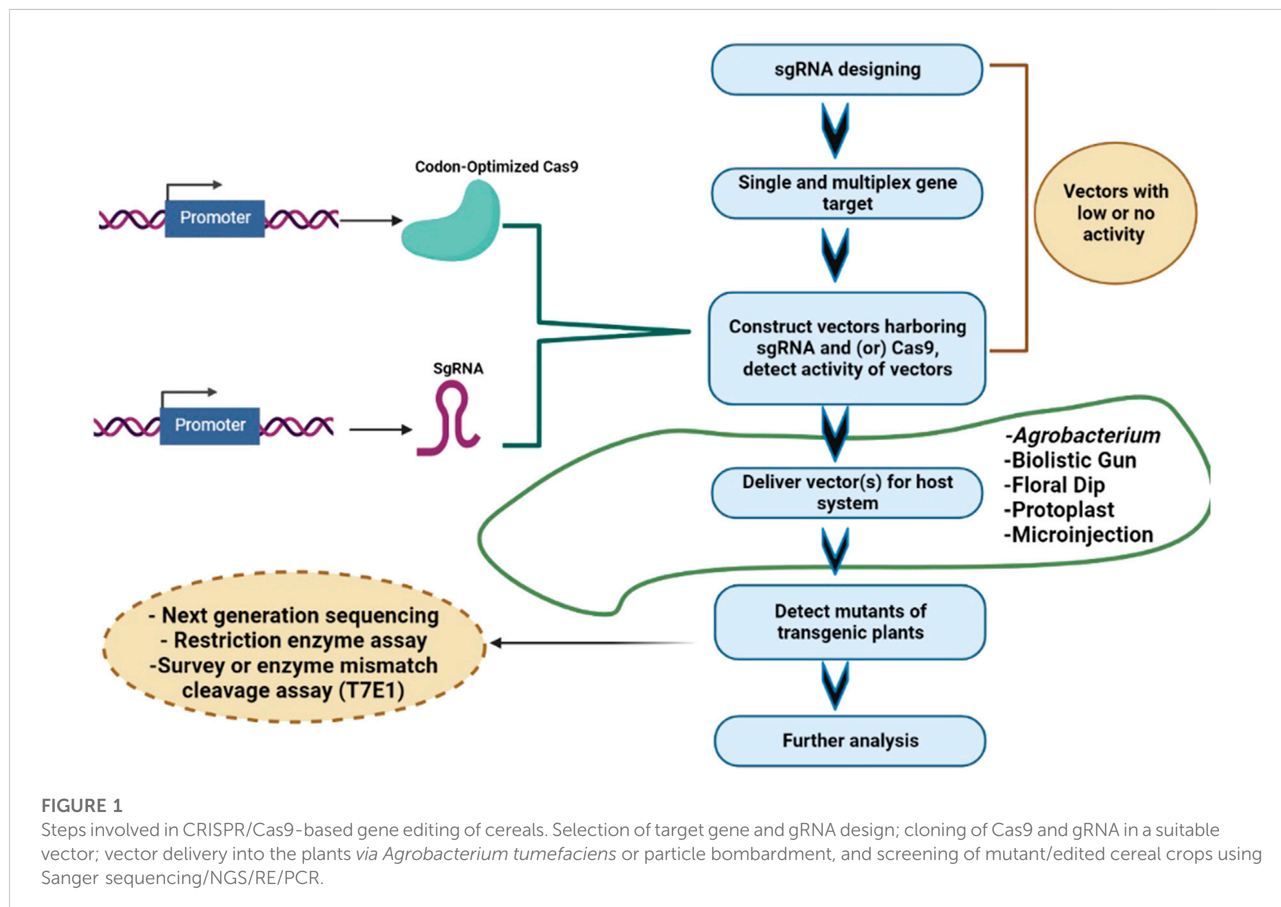


FIGURE 1
Steps involved in CRISPR/Cas9-based gene editing of cereals. Selection of target gene and gRNA design; cloning of Cas9 and gRNA in a suitable vector; vector delivery into the plants *via Agrobacterium tumefaciens* or particle bombardment, and screening of mutant/edited cereal crops using Sanger sequencing/NGS/RE/PCR.

et al., 2018; Lee et al., 2018; Mushtaq et al., 2018; Soyars et al., 2018; Manghwar et al., 2019; Mushtaq et al., 2019; Selma et al., 2019; Mushtaq et al., 2020; Mushtaq et al., 2021a; Mushtaq et al., 2021b; Mushtaq and Molla, 2021).

CRISPR for accelerated cereal crop improvement

Cereal crops are treated as a predominant food and a source of energy due to their supply of essential nutrients in the human diet. It has been estimated that more than 90% of global food production is derived from cereal crops. Rice and wheat are the staple foods of India, Bangladesh, Pakistan, and Afghanistan. CRISPR/Cas9 technology is the prime choice to address the growing demand for cereal crops, owing to its high accuracy and efficiency. CRISPR/Cas9 technology is capable of enhancing tolerance against biotic and abiotic stresses in cereal crops. A schematic workflow of CRISPR/Cas9-based GE in cereals is shown in Figure 1. Details of CRISPR/Cas9 technology applications in cereal crops are categorized in detail in Table 2. Cereal crops in which CRISPR/Cas-based GE has been used to modify different traits are also proposed (Figure 2). For clarity, and in-depth use of this versatile technology, we have considered the viability of CRISPR/Cas9 system-based GE in each individual cereal crop in the following sections.

Wheat

Wheat (*Triticum aestivum* L.) is one of the three main crops grown for human consumption, hence, wheat improvement initiatives for higher yields and improved resilience are crucial for ensuring global food security. Mildew-tolerant wheat varieties have been successfully developed *via* CRISPR/Cas9 to knockout the *TaMLO* gene coding for mildew resistance (Tripathi et al., 2020). This knockout approach resulted in up to 28.50% mutation frequency of the mildew-resistance locus and the crop successfully developed a tolerance (Shan et al., 2014). This experiment initiated interest in the CRISPR/Cas9 system among scientists worldwide to improve cereal crops. Another gene, *TaEDR1*, was known for developing tolerance against powdery mildew well before the advent of CRISPR/Cas9 technology; however, the expression level of this gene required enhancement to achieve improved results, and was later performed using CRISPR/Cas9 (Zhang et al., 2017). Char and Yang (2020) also successfully knocked-

TABLE 2 Applications of CRISPR/Cas9 system in cereal crops.

Crop	Type of study	Targeted gene	Method	sgRNA promoter	Cas9 promoter	Editing efficiency (%)	References
Wheat (*Triticum aestivum*)	Genome editing	*Tainox* and *Tapds*	*Agrobacterium*-mediated	CaMV 35S	CaMV 35S	11–12	Upadhyay et al. (2013)
	Genome editing	*TaGLW7*, *TaGW8*, *TaGW2*, and *TaCKX2-1*	*Agrobacterium*-mediated	TaU6.1, TaU6.2, TaU6.3, and TaU6.5	ZmUbi	64.3	Zhang et al. (2019a)
	Genome editing	*TaGASR7A1*, *TaGASR7 B1*, and *TaGASR7D1*	Biolistic	TaU6	CaMV 35S	5.2	Hamada et al. (2018)
	Knockout	*TdGASR7*	Biolistic	TaU6	2×CaMV 35S	1.8	Liang et al. (2018)
	Knockout	*TaGW2B1*, *TaGW2D1*, and *TaGW2A1*	Protoplast transfection using RNP	—	—	0–4.4	Liang et al. (2017)
	Gene editing	*TaABCC6*, *TaNFXL1*, and *TansLTP*	Protoplast transformation	TaU6	CaMV 35S	NA	Cui et al. (2019)
	knockout	*GW2-B*, *PinB-D*, and *ASN2-A*	Protoplast transformation using RNP	—	—	0–36	Brandt et al. (2020)
	Genome editing through transient expression	*TaGASR7A1*, *TaGASR7B1*, and *TaGASR7D1*	Particle bombardment	TaU6	Ubi	1.1–5	Zhang et al. (2016)
	Site-directed mutagenesis	*TaMLOD1*, *TaMLOA1*, and *TaMLOB1*	Particle bombardment	U6	Ubi1	3.4–6	Wang et al. (2014)
	Site-directed mutagenesis	*LOX2*	Protoplast transformation	TaU6	2×CaMV 35S	~1	Shan et al. (2014)
	Functional genomics	*TaPDS*	*Agrobacterium*- mediated	TaU6	ZmUbi	11–17	Howells et al. (2018)
Sorghum (*Sorghum bicolor*)	Gene editing	*TaLox2* and *TaUbiL1*	Electroporation	TaU6	Ubi1	2.2	Bhowmik et al. (2018)
	Gene editing	*CAD* and *PDS*	Particle bombardment	U3	ZmUbi	NA	Liu et al. (2020)
	Gene editing	*K1c*	*Agrobacterium*-mediated	TaU3	ZmUbi	14.1–78.3	Li et al. (2018)
	Knockout	*SbFT* and *SbGA2ox5*	*Agrobacterium*-mediated	U6P.1 and U6P.2	ZmUbi	33.3–83.3	Char and Yang. (2019)
	Targeted mutagenesis	*StALS1*	*Agrobacterium*-mediated	U6	2×CaMV 35S	5–60	Butler et al. (2015)
	Knockout	*Sb-CENH3*	*Agrobacterium*-mediated	U6	ZmUbi	37–40	Che et al. (2018)
	Functional genomics	*DsRED2*	*Agrobacterium*-mediated	U6	OsActin1	NA	Jiang et al. (2013)
Rice (*Oryza sativa*)	Genome editing	*TaLOX2*	*Agrobacterium*-mediated	OsU3	2×CaMV 35S	~1	Shan et al. (2014)
	Gene editing	*OsPMS3*, *OsYSAOsDERF1*, *OsMYB1*, *OsMSH1*, *OsPDS*, *OsSPP*, *OsEPSPS*, *OsMYB5*, and *OsROC5*	*Agrobacterium*-mediated	U3 and U6	ZmUbi. CaMV 35S	>35	Ma et al. (2015)
	Gene editing	*SBEI* and *SBEIIb*	*Agrobacterium*-mediated	OsU3	Ubi1	26.7–40	Sun et al. (2017)
	Knockout	*elF4G*	*Agrobacterium*-mediated	TaU6	ZmUbi1	30–64	Macovei et al. (2018)
	Knockout	*OsNramp5*	*Agrobacterium*-mediated	OsU6 and OsU3	Ubi1	13.6–35	Tang et al. (2017)
	Loss of function	*OsMORE1* and *OsMORE1a*	*Agrobacterium*-mediated	OsU3	ZmUbi	NA	Kim et al. (2022)

(Continued on following page)

TABLE 2 (*Continued*) Applications of CRISPR/Cas9 system in cereal crops.

Crop	Type of study	Targeted gene	Method	sgRNA promoter	Cas9 promoter	Editing efficiency (%)	References
	Knockout	*SAPK1* and *SAPK2*	*Agrobacterium*-mediated	OsU3 and OsU6a	CaMV 35S	NA	Lou et al. (2018)
	Gene editing	*OsPIN5b*, *gs3*, and *OsMYB30*	*Agrobacterium*-mediated	U6	CaMV 35S	42–66	Zeng et al. (2020)
	Deletion	*Waxya* and *Waxyb*	*Agrobacterium*-mediated	OsU3 or OsU6	CaMV 35S	8.6–11.85	Liu et al. (2022)
	Gene editing	*RLK*	*Agrobaterium*-mediated	U3	Ubi1	NA	Chen et al. (2022)
	Knockout	*ISA1*	*Agrobacterium*-mediated	U6	CaMV 35S	NA	Shufen et al. (2019)
	Knockout	*Waxy*	*Agrobacterium*-mediated	U6	CaMV 35S	82.7–86.9	Zhang et al. (2018)
	Knockout	*Waxy*	*Agrobacterium*-mediated	U3 and U6a	CaMV 35S	NA	Yunyan et al. (2019)
	Knockout	*OsRR22*	*Agrobacterium*-mediated	OsU6a	UbiH	64.3	Zhang et al. (2019a)
	Knockout	*OsCCD7*	*Agrobacterium*-mediated	OsU3	OsUbi	22.2–64.3	Butt et al. (2018)
	Knockout	*EPSPS*	Protoplast transformation	OsU3 and TaU3	—	2.0–2.2	Li et al. (2016)
	Site-directed mutagenesis	*OsROC5*, OsSPP, and Os*YSA*	*Agrobacterium*-mediated	OsU6-2	CaMV 35S	61.1–67.7	Zhang et al. (2014)
	Site-directed mutagenesis	*OsMPK5*	*Agrobacterium*-mediated	U3 and U6	CaMV 35S	3–8	Xie and Yang. (2013)
	Site-directed mutagenesis	*OsPDS*, *TaLOX2*, *OsBADH*, and *OsMPK2*	*Agrobacterium*-mediated	OsU3	CaMV 35S	~1	Shan et al. (2014)
	Site-directed mutagenesis	*OsMYB1*	Protoplast transformation	OsU3	CaMV 35S	NA	Miao et al. (2013)
	Multiplex editing capability with endogenous tRNA	*OsMPKs*	*Agrobacterium*-mediated	OsU3	OsUbi	6–100	Xie et al. (2015)
	Multiplex GE in dicot andmonocot plants	46 genomic targets	*Agrobacterium*-mediated	OsU3, OsU6 and OsU6c	OsUbi and CaMV 35S	24.7–90	Ma et al. (2015)
Barley (*Hordeum vulgare*)	Functional studies	*WDV1*, *WDV2*, *WDV3*, and *WDV4*	*Agrobacterium*-mediated	WDV	CaMV 35S and ZmUbi	NA	Kis et al. (2019)
	Fragment Deletions and Small Indels	*ENGase*	*Agrobacterium*-mediated	OsU6	ZmUbi	78	Kapusi et al. (2017)
	Gene editing	*NbPDS1*	*Agrobacterium*-mediated	U6	CaMV 35S	NA	Raitskin et al. (2019)
	Gene editing	*Hpt*	*Agrobacterium*- mediated	U6	ZmUbi	20–70	Watanabe et al. (2016)
	Knockout	*HptII*	*Agrobacterium*-mediated	U6	ZmUbi	NA	Lawrenson and Harwood. (2019)
	Knockout	*HvMORC1* and *HvMORC6a*	*Agrobacterium*-mediated	HvU3	CaMV 35S	8–81	Galli et al. (2022)
	Knockout	*HvCKX1*	*Agrobacterium*-mediated	HvU3	ZmUbi1	NA	Holubova et al. (2018)
	Knockout	*HvMORC1*	*Agrobacterium*-mediated	HvU3	ZmUbi	38–77.7	Kumar et al. (2018)

(Continued on following page)

TABLE 2 (*Continued*) Applications of CRISPR/Cas9 system in cereal crops.

Crop	Type of study	Targeted gene	Method	sgRNA promoter	Cas9 promoter	Editing efficiency (%)	References
	Knockout	*HvCKX1, HvCKX3,* and *Nud*	*Agrobacterium*-mediated	TaU6	ZmUbi	18–68	Gasparis et al. (2018)
	Knockout	*GPhsp70, GPhsp26, GPhsp16.9, GPgst, GPcrt, GPipi,* and *GPpdi*	*Agrobacterium*-mediated	TaU6	ZmUbi	NA	Panting et al. (2021)
Maize (*Zea mays*)	Knockout	*waxy*	Biolistic	TaU6	ZmUbi	1–55.4	Gao et al. (2020)
	Gene editing	*ZmIPK*	*Agrobacterium*-mediated	ZmU3	CaMV 35S	13.1	Liang et al. (2014)
	Gene editing	*ZmHKT1*	*Agrobacterium*-mediated	AtU6-26, OsU3 or TaU3	CaMV 35S or Ubi1	NA	Xing et al. (2014)
	Gene editing	*LIG1, MS26,* and *MS45*	*Agrobacterium*-mediated	ZmU6	ZmUbi	6–86	Svitashev et al. (2015)
	Gene editing	*Zmzb7*	*Agrobacterium*-mediated	ZmU3	CaMV 35S	19–31	Feng et al. (2016)
	Gene editing	*PSY1*	*Agrobacterium*-mediated	ZmU6	ZmUbi	10.67	Zhu et al. (2016)
	Gene editing	*CLE*	*Agrobacterium*-mediated	—	—	NA	Liu et al. (2021)

out the *TaEDR1* gene *via* CRISPR/Cas9 in 2020 to develop powdery mildew-tolerant cultivars. In addition to this, the CRISPR/Cas9 system was implemented in wheat to induce mutations in the *Tapx1* and *TaLox2* genes, with mutation rates of approximately 9 and 45% achieved, respectively (Shan et al., 2014). *TaDEP1*, *TaNAC2*, *TaGW2*, and *TaGASR7* genes in wheat were knocked-out *via* CRISPR/Cas9 to increase the grain length, grain width, grain area, and grain weight, compared to wild plants (Wang et al., 2018). We have summarized the application of CRISPR/Cas9 technology for targeting numerous genes in wheat to improve various traits (Table 2).

Rice

Owing to its small genome size, transformability, accessibility to genetic resources, and sequence data, rice (*Oryzae* spp.) was among the first crops to be extensively modified and studied (Biswas et al., 2020). Additionally, genome-wide association studies (GWAS), comparative genomics, and OMICS-based methods have been used to investigate a variety of genes and SNPs linked to agronomically desirable traits. This allows modification of target genes with greater efficiency. Numerous genome engineering experiments have been carried out, and more recently, the rice genome has been edited using CRISPR/Cas9 technology. The CRISPR/Cas9 technique was used to successfully modify the *OsPDS* (*phytoene desaturase*) gene in rice (Banakar et al., 2020). Two sgRNAs (namely *SP2* and *SP1*) were designed to disrupt the *OsPDS* gene, and mutation frequencies of approximately 9% in transgenic and 15% in protoplasts were observed (Shan et al., 2013). Similarly, the mitogen-activated *protein kinase5* (*OsMPK5*) gene in rice was knocked-out using CRISPR/Cas9 to enhance disease resistance, and a mutation frequency of 3–8% was observed (Ma et al., 2015). The CRISPR/Cas9 technology for multiplex GE in rice has been the subject of several investigations (Liu et al., 2020; Xu et al., 2020). An experiment was conducted where multiple sgRNAs were engineered to be expressed under U3 and U6 promoters. The results revealed that multiple GE using CRISPR/Cas9 technology is highly applicable in rice crops (Ma et al., 2015). A group of scientists conducted two experiments using CRISPR/Cas9 on different genes, with different approaches. In the first experiment, two sgRNAs were designed to target the *NAL1* (*narrow leaf*) gene; the results revealed a low mutation rate for this particular gene (Hu et al., 2016). The second experiment was conducted on *ACT1* (*ACTIN1*) and *UQ1* (*UBIQUITIN1*) genes using CRISPR/Cas9. The mutation frequency of both genes was high enough to develop disease-tolerant genotypes (Hu et al., 2018). Due to the outstanding potential of CRISPR/Cas9 to permanently maintain hybrid vigor, plant biologists and the seed industry have shown a strong interest in apomixis (Wang, 2020). Although, the introduction of apomixis traits

from wild relatives into major crops has remained ineffective, artificial apomixis has been used as an alternative to fix the hybrid vigor in rice (Biswas et al., 2020). The development of the *MiMe* (*mitosis* instead of *meiosis*) line in rice, which transforms meiosis into mitosis and results in the development of clonal gametes, has produced rice plants that generate functioning diploid gametes with the same genetic makeup as their parent. Apomixis-like clonal seeds are generated when the *MiMe* line in rice is combined with special genome elimination lines, which contain an altered, centromere-specific histone 3 (CENH3). Furthermore, the generation of haploid plants from egg cells can be achieved by either the egg cell–specific expression of *BABY BOOM1* (*BBM1*), or the disruption of *MATRILINEAL* (*MTL*) using CRISPR/Cas9 gene editing technology. Synthetic apomixis is established, and clonal seeds are produced by simultaneously engineering *MiMe* rice lines with altering *BBM1* expression or *MTL* disruption (Kumar et al., 2020). Additionally, multiple research teams have tried to mutate the genes related to cadmium (*OsNramp5*), drought (*OsSAPK2*), and salt (*OsRR22*) stresses, and the resulting altered lines exhibited improved resistances to the respective conditions (Tang et al., 2017; Liu et al., 2018; Liu et al., 2019). Many other studies have also focused on rice and CRISPR/Cas9-mediated genome editing (Table 2). Studies such as these prove that CRISPR/Cas9 can be successfully exploited for improving the tolerance of rice to stresses like salinity.

Maize

Maize (*Zea mays* L.) is the third most significant crop after rice and wheat, and is one of the most important cereals that can be cultivated in a wide variety of environmental circumstances (Liang et al., 2014). The first report of GE involved targeted disruption of the *IPK1* (*Inositol Phosphokinase1*) locus *via* knock-in of a herbicide tolerance gene using ZFNs (Shukla et al., 2009). On the other hand, the first use of TALENs in maize was a proof-of concept study to generate stable and heritable mutations at the *GL2* (*GLOSSY2*) locus (Char et al., 2020). Furthermore, GE in maize increased significantly with the advent of the CRISPR/Cas9 technology, and the initial investigations were ground-breaking since they were the first to demonstrate multiplex editing as well as DNA-free editing using Cas9/gRNA ribonucleoproteins (RNPs). In the first studies, five loci (*LG1*, *ALS1*, *ALS2*, *MS26*, and *MS45*) were targeted in maize embryos using DNA constructs, delivered by particle bombardment. Mutations were observed at all five target sites (upstream of *LG1*, in the acetolactate synthase genes *ALS1* and *ALS2*, and in the male fertility genes *MS45* and *MS26*), including multiplex mutations in *LG1*, *MS45*, and *MS26*. CRISPR/Cas9 was successfully used to knock out the *ZmIPK* gene in maize, which controls the formation of phytic acid, while two sgRNAs were utilized underneath the expression promoter U6 to knock out the *phytoene synthase* (*PSY1*) gene with a mutation frequency of 10.67% (Zhu et al., 2016). The authors also sequenced the mutated gene to confirm the effectiveness of the mutation. CRISPR/Cas9 was employed in T_0 maize lines to target the *albino marker* (*Zmzb7*) gene with a mutation frequency of 31% observed (Feng et al., 2016). By targeting the thermosensitive *male-sterile 5* (*ZmTMS5*) gene with three sgRNAs rather than one or two, researchers were able to perform protoplast alterations (Chen et al., 2018). The modified plants presented bi-allelic modification, demonstrating the potential of the CRISPR/Cas9 technology for intended mutagenesis in maize to improve particular traits (Char et al., 2020). Another major success of gene editing is the development of a maize variety with a higher grain yield under drought-prone environments, by employing precise insertion of a *GOS2* promoter inside the 5′-UTR of *ARGOS8* (Shi et al., 2017). These studies demonstrate the comprehensive applications of CRISPR/Cas9 systems for breeding approaches in maize.

Barley

In terms of global production, barley ranks as the fourth-most significant cereal crop. Due to its diploid genome structure, barley is used as a model plant for *Triticeae* crop species. Barley gene editing has proven to be a reliable, accurate, and affordable approach for quick plant breeding. Early attempts to establish GE in barley used TALENs and did not target a coding area, instead choosing to focus on the promoter region of the phytase *HvPAPhy-a* (Wendt et al., 2013). It was suggested that barley was receptive to GE, without producing a large number of primary transformants because, on average, one out of every four plants bearing the selection marker displayed editing activity. In some cases, editing efficiencies were even up to 88%; editing events were screened by methods other than sequencing, therefore, the reported efficiencies may be conservative estimates (Gasparis et al., 2018). Succeeding GE investigations, where targeted DSBs were mostly induced through *Agrobacterium*-mediated use of conventional Cas9, verified that the editing efficiency is not a constraint. Lawrenson et al. (2015) multicopy genes in barley (*Hordeum vulgare*) and *B. oleracea* to investigate the gRNA Cas9 editing method and target specificity requirements. The researchers targeted two copies of the *HvPM19* gene in *H. vulgare* and *B. oleracea* and found Cas9-induced mutations in 23% and 10% of lines, respectively; mutated plants were stunted in the first-generation. Stable Cas9-induced mutations were transferred to T_2 plants irrespective of the T-DNA composition in both *H. vulgare* and *B. oleracea.* Although the presence of at least one

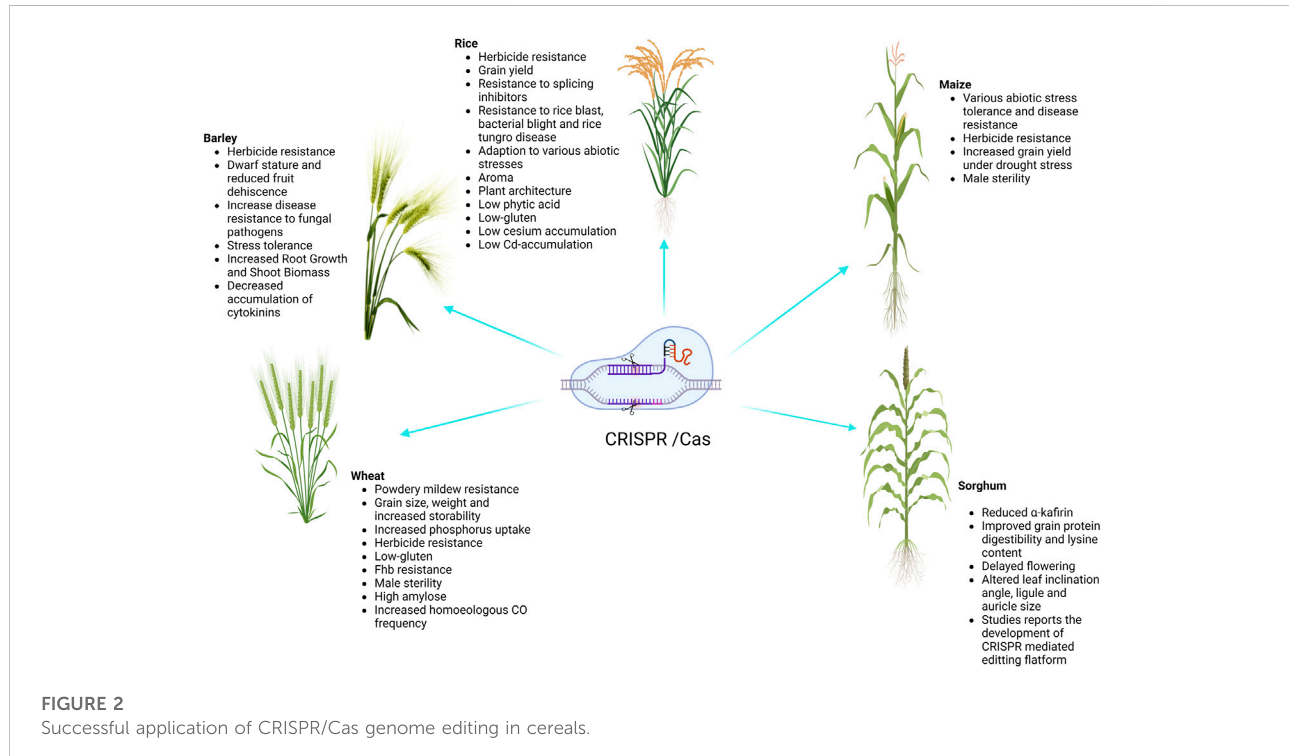


FIGURE 2
Successful application of CRISPR/Cas genome editing in cereals.

mismatch between both the sgRNA and the non-target gene sequences was observed, off-target activity across both species was discovered. A transgene-free *H. vulgare* plant exhibited mutations in both the target and non-target alleles of *HvPM19*. Multiple successful efforts have been made to alter the drought and other stress-related genes, specifically *TaDREB2* and *TaERF3*, in transient processes in protoplasts, indicating that this can be a quick method to identify specific and off-targets in the designed gRNAs of barley and wheat (Kim et al., 2018).

Technical advances in base editing and prime editing

Base editing

Genome-wide association studies (GWAS) revealed that single-base substitutions are often the best way to introduce excellent traits in crop plants. Based on this, several effective techniques have been employed to generate precise point mutations in crop plants to achieve desired results (Zhang et al., 2018). Numerous agronomic traits have been found to be influenced by single alterations in the bases of genes. Gene base conversion is unfortunately not possible using CRISPR/Cas9 technology. Due to this, finding a precise and reliable method for editing crop genomes is essential. Base editing is thought to be a substitute and a more effective strategy. In agricultural plants, base editing is utilized, replacing HDR-mediated gene editing in an efficient and systematic manner (Bharat et al., 2019). It can achieve automated nucleotide substitutions without disrupting genes. Base editing typically involves a combination of an inactive catalytic CRISPR-Cas9 domain (Cas9 variant, Cas9 nickase, or dCas9) and cytosine or an adenosine deaminase domain that transforms one nucleotide base into another (Mishra et al., 2022). Variations in the single base may produce excellent variant traits in crops, thereby helping to accelerate development in crop plants. Without destruction of genes, base editing can recover single nucleotides or base substitutions, thus reducing deletions and insertions. It is an efficient technology to design new characteristics in important crops for achieving global food and nutrition security (Eid et al., 2018).

A base editor is a chimeric protein composed of a catalytic region and a DNA-targeting module that can deaminate the nucleotide adenine or cytosine in the genome (Komor et al., 2016 and, 2017). In the base editing approach, a combination of the catalytic cytidine deaminase and dCas9 is directed by sgRNA molecules to conduct single-base changes without the formation of double-strand breaks (DSBs) in DNA molecules. The base editor may make single base substitutions, thereby minimizing the frequency of indels. The most commonly used DNA base editors are classified into two types: ABE (Adenine Base Editor) and CBE (Cytosine Base Editor). In recent years these have become effective tools for GE (from C to T and A to G) in eukaryotes (Liu et al., 2016; Qin et al., 2020; Bansal et al., 2021).

The base editing method has been effectively improved and verified in various cereal crops including wheat and maize (Table 3). A schematic representation of various Cas9-based base editors are highlighted in Figure 3.

Cytosine base editing system

A nuclease-deficient CRISPR system directs cytidine deaminase, which modifies the cytosine base. Deamination of cytosine produces uracil at the target location, which eventually transforms C-G into a T-A base pair without causing a double DNA strand break. The first-generation basic editor (BE1) was established in 2016 by Liu and others at Harvard University, United States. It is comprised of the cytosine deaminase APOBEC1 (from rat), which connects a linker with dCas9 through 16 unstructured XTEN amino acids. Base Editor (BE) and Target-AID (first cytosine base editing systems) employ rAPOBEC1 and PmCDA1, respectively, as deaminases and effectively insert alterations within the editing windows of 12–16 bases and 16–20 bases upstream of the PAM. The main limitation of BE1 is that uracil DNA glycosylase (UDG) often removes uracil, leading to a low editing efficacy. A series of upgraded basic editors have been designed, taking into account the limitations of BE1 and its low editing efficacy. When the C-terminal of the DNA component is combined with UGI (Uracil Glycosylase Inhibitor), a second-generation base editor (BE2) is formed: APOBEC-XTEN-dCas9-UGI (Komor et al., 2016). The activity of UDG is inhibited by the additive UGI, which catalyzes the deletion of UDNA from DNA inside the cells and commences the pathway of BER (base excision repair). The inhibition of BER generates a threefold increase in the efficacy of editing in human cells. Subsequently, a third-generation BE3 base editor was designed, which consists of an amalgamation of C-terminus and UGI *via* four amino acid linkers, and the fusion of the N-terminus of nickase Cas9 D10A with rAPOBEC1 *via* an XTEN linker (16 a.a) (Komor et al., 2016). Substitution of dCas9 with nCas9 (Cas9 nickase), cleaving the chain opposite decytidine is the main characteristic of the BE3 system. Therefore, the editing efficiency of BE3 is further improved sixfold compared with that of BE2. Compared with 0.1% in BE2, the application of nCas9 showed a 1.1% increase in indel frequency. The cytosine base editor conducts the automated conversion from C to T; however, the presence of multiple Cs in the catalytic window can cause off-target activity where C is converted into U. To minimize this error, multiple BE3 variants (using non-canonical PAM) were generated using different Cas9 variants. SpCas9 variants (such as VQR-BE3, VRER-BE3, EQR-BE3, and SaKKH-BE3) of *Staphylococcus aureus* target NGCG, NGAG, NGAN, and NNNRRT PAM sequences, respectively, and have improved the editing capability by 2.5 (Kim et al., 2017). In addition to SpCas9 variants, SaCas9 with an NNGRRT PAM sequence has also been applied in multiple research proposals, exhibiting higher efficacies. Various mutants of cytosine deaminase were produced, for example, YEE-BE3 and YEE-BE2, which increase the specificity of DNA and decrease off-target activity because of different editing window widths. YEE-BE3 showed the greatest editing efficiency within a narrow editing window width of (approximately 2 nt) (Kim et al., 2017). Activation-induced cytidine deaminase is another base editing method and Target-AID was developed and is composed of a cytidine deaminase pmCDA1 (from the southern eel) and a nickase (Cas9D10A) (Nishida et al., 2016). The Target-AID system, with increased efficiency, is used for targeted mutagenesis in human and mouse cells. Target-AID is a useful technique for generating numerous gene alterations in tomato and other crops in which a mutant population has been detected (Hunziker et al., 2020). The Target-AID technique may be used as an alternative whereby breeders can introduce allelic changes in many targets in a single line and generation. The editing effectiveness of BE3 and Target-AID are increased two- to threefold when UGI and nickase are used.

Further optimization of CBE was performed to reduce indel formation during base-editing, to improve editing efficiency, and to narrow the editing window. An improved fourth-generation base editing system (SaBE4 and BE4) was generated through the amalgamation of two UGI molecules with the N and C terminals of nCas9 *via* a 9-aa linker and with Cas9D10A and rAPOBEC1 *via* a 32-aa linker. The use of UGI prevents UNG from entering the uracil intermediate, inhibits the formation of BER, and limits unwanted products. Compared with SaBE4 and BE4, the average non-T product formation by SaBE4-Gam and BE4-Gam is reduced, and the C to T editing efficiency is improved. As a result, fourth-generation base editors may be used to successfully program from C to T, decreasing the creation of indels and improving product purity. Additionally, with the automated insertion of point mutations, deaminase is also used to build libraries of various point mutations located in target regions of genomes. To create local sequence diversity, two basic editing methods; TAM (Targeted AID-mediated mutagenesis) and CRISPR-X are utilized (Ma et al., 2016). Human AID is combined with dCas9 in the TAM system to obtain effective genetic diversity in animal cells. Excited AID variants are targeted by dCas9 in the CRISPR-X system to induce point mutations (local and diverse) (Hess et al., 2016). dCas9 was utilized as a DNA-targeting module and has been proved to be effective for gene editing. However, a major limitation is the requirement for G/C-rich PAM sequences. The first cytidine deaminase base editor (Cpf1-based) was developed by Li and others to improve the efficacy of base editing (Li et al., 2018). Cpf1 supports T-rich TTTN-PAM, produces 5 bp cohesive ends (Zetsche et al., 2015), and can analyze sgRNA, allowing it to be used in a variety of genome-targeting applications (Zetsche et al., 2017). A rat APOBEC1 domain is combined with UGI and catalytically inert dLbCpf1 (*Spirulina* bacterium Cpf1) to form

TABLE 3 List of genes targeted by cytidine and adenine base editors in cereal crops.

Cereal plant species	Trait improvement	Type of base editor used	Target gene	References(s)
Rice (*Oryza sativa*)	Nitrogen use efficiency	CBE	*NRT1.1B* and *SLR1*	Lu and Zhu (2017)
	Senescence and death	CBE	*OsCDC48*	Zong et al. (2017)
	Nutritional improvement	CBE	*OsPDS* and *OsSBEIIb*	Li et al. (2017)
	Herbicide resistant	CBE	*C287*	Shimatani et al. (2017)
	Pathogen-responsive gene	ABE	*OsMPK6*	Yan et al. (2018)
	Defense response	CBE	*OsRLCK185* and *OsCERK1*	Ren et al. (2018)
	Plant architecture and grain yield	ABE	*OsSPL14*	Hua et al. (2018)
	Herbicide resistance	ABE	*OsACC-T1*	Li et al. (2018)
	Della protein for plant height	ABE	*SLR1*	Hua et al. (2018)
	Herbicide resistance	CBE	*OsSPL14*	Tian et al. (2018)
	Blast resistance	CBE	*Pi-d2*	Ren et al. (2018)
	Herbicide resistance	CBE	*ALS*	Veillet et al. (2019)
	Grain size and yield	ABE	*GL2/OsGRF4* and *OsGRF3*	Hao et al. (2019)
	Rice amylose synthesis	ABE	*Wx*	Hao et al. (2019)
Wheat (*Triticum aestivum*)	Panicle length and grain weight	ABE	*TaDEP1* and *TaGW2*	Li et al. (2020)
	Lipid metabolism	CBE	*TaLOX2*	Zong et al. (2017)
Maize (*Zea mays*)	Chromosomal segregation	CBE	*ZmCENH3*	Zong et al. (2017)

dLbCpf1-BE0 (base editor). Before the PAM sequence, the base editor presents an editing window of 8–13 bps, with an efficiency of 20%–22%. As a result, Cpf1-based base editors can improve the base editing efficiency and provide various PAM sequence alternatives in the target gene (Mishra et al., 2020). Currently, the cytosine base editing technique is used for a variety of cereal plant traits (Table 3). The capability of the CBE3 method in rice was examined on three bases; firstly, (P2) in *OsPDS* that encodes phytoene desaturase; secondly, (S3); and thirdly, (S5) in OsSBEIIb that encrypts enzyme IIB restriction endonuclease (starch branching enzyme). Changes in a single nucleotide of the DNA were induced at P2, S3, and S5 target sites and the efficacies of inserted mutations were 1.0%, 10.5% and 19.2%, respectively. A high-amylose rice is generated during the destruction of the intron-exon boundary (Li et al., 2017; Lu and Zhu., 2017).

In rice, using the CBE3 system can produce stable SLR1 and NRT1.1B basic editing plants, with editing efficacies of 13.3% and 2.7%, respectively (Li et al., 2017). The use of NRT1.1B for effective editing in rice can increase the efficacy of nitrogen use (Hu et al., 2015). Through nCas9-cytidine deaminase fusion, the efficacy of targeted transformation from C to T in *ZmCENH3*, *TaLOX2*, *OsCDC48*, and *OsSPL14* genes as high as 43.5% in maize, wheat, and rice. rBE5 was created in rice through the linkage of Cas9n-NLS and hAID*D *via* a peptide linker, to edit *OSFLS2* and *Pi-d2* with efficacies of 57.0% and 30.8%, respectively. The *Pi-d2*-edited rice gene contains a point mutation that modulates the defense response to blast fungus, which is of significant agricultural importance (Ren et al., 2018). A3A-PBE (base editor) consists of UGI, human APOBEC3A, and nCas9. Unlike previous base editors, it can convert C to T in high GC content regions in maize, rice, and wheat within a window of 17 nt. (Zong et al., 2018). The CBE system converts C to T in rice with up to 80% frequency, using engineered SaCas9 and SpCas9 variants (Hua et al., 2019). These universal base modifying tools are expected to broaden the target range to include rice and other cereals.

Adenine base editing system

Nicole Gaudelli, a researcher in David Liu's laboratory, developed an adenine base editor that converts adenine to inosine (I), leading to an A to G conversion (Gaudelli et al., 2017). The first generation of ABE1.2 was produced using an XTEN linker (16a.a), with the fusion of a TadATadA* heterodimer and the N-terminus of nCas9. An NLS (nuclear localization signal) was combined with the C-terminus of nCas9. ABE, comprised of nCas9 and deoxy-adenosine deaminase. This connects with the target DNA sequence through guided RNA

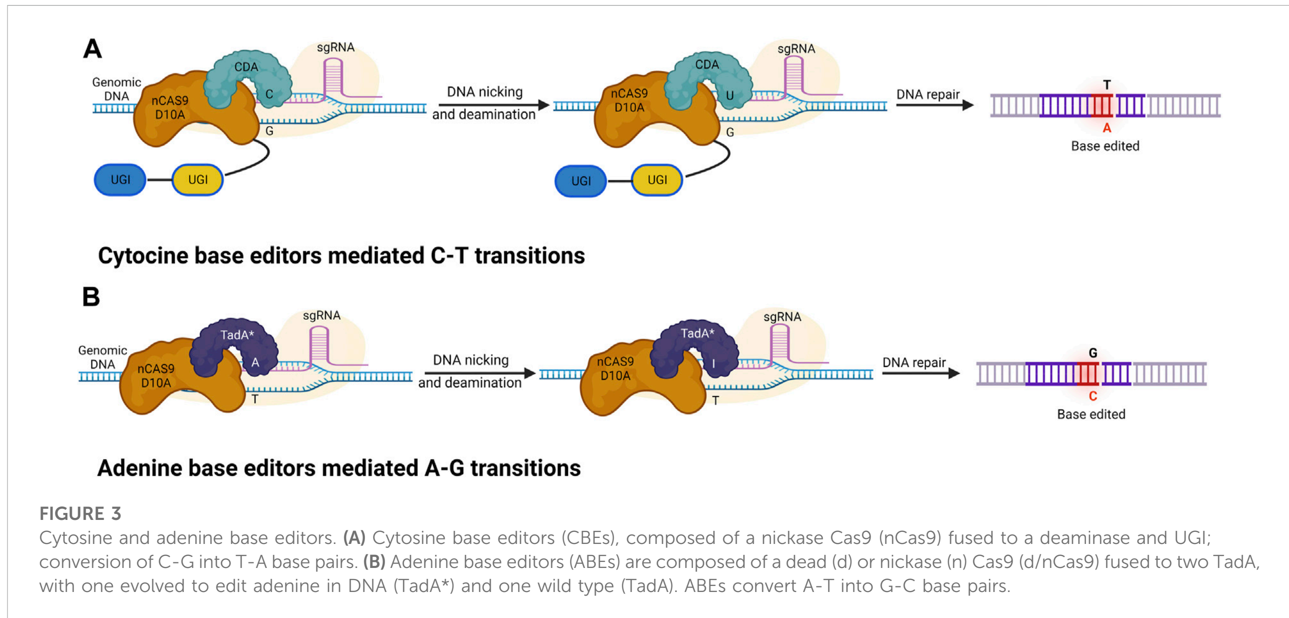


FIGURE 3
Cytosine and adenine base editors. **(A)** Cytosine base editors (CBEs), composed of a nickase Cas9 (nCas9) fused to a deaminase and UGI; conversion of C-G into T-A base pairs. **(B)** Adenine base editors (ABEs) are composed of a dead (d) or nickase (n) Cas9 (d/nCas9) fused to two TadA, with one evolved to edit adenine in DNA (TadA*) and one wild type (TadA). ABEs convert A-T into G-C base pairs.

programming, unveiling small bubbles of ssDNA, within which a putative deoxyadenosine deaminase domain catalyzes the conversion of A to I, which is ultimately converted to a G-C base pair at the target site, after DNA replication (Wolf et al., 2002; Li et al., 2018). ABE has been further optimized to improve the editing efficiency by the fusion of a TadA (2.1)* domain to the C-terminus of nCas9 (D10A), the use of different TadA mutations, the use of an N-terminally inactivated TadA* subunit, or by changing the gap between nCas9 (D10A) and TadA (2.1)* subunit (linker length). The seventh generation of ABE (e.g., ABE7.10) was developed through protein engineering and extensive directed evolution, and it effectively converts the target A to G (approximately 50%) in human cells, with extremely high product purity (≥99.9%) and a very low incidence of indels (≤0.1%) (Koblan et al., 2018). In human cells, SpCas9-NG can also efficiently generate selective mutations at distinct NG PAM positions (Hua et al., 2019), providing an opportunity to expand the application of ABE editing. Currently, ABE-P1S (base editor) containing ecTadA* 7.10-nSpCas9 (D10A) shows an increased editing efficacy in rice, compared with the widely implemented fusion of 7.10-nSpCas9 (D10A) *ecTadA-ecTadA. The editing efficacy of other ABE systems (including SaCas9 or SaKKH-Cas9 variants) can also be enhanced using a fusion protein (ecTadA* 7.10-nCas9) (Hua et al., 2020). More effective ABE will promote its use in crop productivity to improve the grain size and yield of rice (Tiwari et al., 2020). The ABE base editor effectively regulates the alteration of A to G in cereal plants (Table 3) (Li et al., 2018). Together with the abovementioned CBE, ABE can induce four types of conversions (from A-T to G-C or C-G to T-A) at specific target sites in the genome, improving the base editing potential. rBE14 (base editor) of a TadA:TadA7.10 heterodimer guided by nCas9 (D10A) has been developed. In rice, it easily and efficiently converted A to T in *OsWRKY45*, *OsSERK2*, and *OsMPK6*, with corresponding rates of 62.3%, 32.1%, and 16.7% (Yan et al., 2018).

A novel ABE plant, based on the fusion of nCas9 and an improved tRNA adenosine deaminase, permitted the transformation of A to G with up to 59.1% frequency in wheat and rice and 7.5% in protoplasts. The amalgamation of nCas9 (D10A) and the recombinant ecTadA* 7.10 protein resulted in the development of the ABE-P1 plant. The impact of editing on ABE-P1 was estimated at the *OsSLR1* and *OsSPL14* gene loci of the rice, with editing efficacies of 12.5% and 26.0%, respectively. Four plant-compatible ABE binary vectors (pcABE) were developed through the fusion of nCas9 and various modified ecTadAs (Kang et al., 2018). A novel ABE adenosine base editor was designed to increase the number of targeted sites in the rice genome with the help of a SpCas9 variant. The target genes (*OsSPL17* and *OsSPL14*) presented editing efficacies of 45% and 25%, respectively. These findings indicate that ABE with SpCas9-NG plays an effective role in rice, expands the compatibility of PAM, and expands the application of ABE in crop plants (Hua et al., 2019).

Glycosylase base editing system

The base editors discussed thus far (CBE and ABE) can only catalyze base transitions (C-T and A-G). These BEs cannot generate base transversions; instead, they can only produce base transitions such as C-T (or G-A) and A-G (or T-C) swaps. To overcome these technology limitations, the Zhang and Changhao groups developed new base editors, namely,

glycosylase base editors (GBE) (Zhao et al., 2021) (Figure 4). GBEs are made up of a uracil-DNA glycosylase (UNG), a Cas9 nickase, and a cytidine deaminases. UNG excises the U base produced by the deaminase, generating an apyrimidinic/apurinic (AP) site that begins the DNA repair procedure. As a new generation of base editing technology, GBE directly modifies the target base instead of relying on DNA replication. This technology further improves the base editing system, fills in any gaps in the different base editing systems, and realizes the arbitrary base editing of microbes for the first time. In wild-type *E. coli* strains, GBE editing technology has allowed the conversion of C-A with an accuracy of 93.8%. Any base editing (NBE) was also created, allowing any A, T, G, or C to be changed to any other base in a one, two, or three-step procedure. In addition to this, GBE allows the first C-G conversion in mammalian cells, with high position specificity and a narrow editing window. GBE achieved C-G conversions with a high specificity at the 6th C in an N20 sequence, which is different from other BE techniques (Zhao et al., 2021). A number of studies have been conducted to improve the performance of the C-G conversion (Liu et al., 2019; Koblan et al., 2021); however, in spite of this, the efficiency is still subpar and fluctuates greatly depending on the locus. Additionally, only a small number of GBEs with wider coverage were built. Thus, the continued development of GBE editors would facilitate various applications in genetic therapies and scientific research.

C-G base editing system

Only one or two types of base substitutions may be accomplished using single base editors or dual deaminase-mediated base editors, respectively. Recently, a novel glycosylase base editor system (CGBE) system was developed, where Uracil-DNA glycosylase (UNG) is used instead of the uracil glycosylase inhibitor (UGI), to effectively initiate multiple base conversions, including C-A, C-T, and C-G. CGBE consists of a Cas9 nickase fused to a uracil DNA glycosylase (UNG) and cytidine deaminase. Architecturally, CBE and CGBE are comparable, the difference being that UNG is used in place of UGI. In addition, UNG excises the U base produced by deaminase, generating an AP site that begins the DNA repair procedure, which introduces indel mutations *via* an error-prone repair and replication mechanism, resulting in preferred insertion of G at the AP site and hence leading to C-G editing. The indels and the C-A and C-T conversions produced by CGBE are regarded as undesirable by-products for accurate base editing (Kurt et al., 2021; Zhao et al., 2021). On the contrary, it is also believed that these two by-products are advantageous when CGBE is employed to produce a saturated mutagenic population in a gene, because they broaden the range of BE outcomes. However, the reasons why G is selected over the other two bases are still a mystery.

Zhao et al. linked the amino N-terminus of nCas9 to APOBEC1 cytidine deaminases as well as UNG to the carboxy C-terminus (APOBEC1-nCas9-UNG) (Zhao et al., 2021), while Kurt et al. coupled both APOBEC1 and UNG at the N-terminus (UNG-APOBEC1-nCas9) (Kurt et al., 2021) to achieve C-G editing in mammalian cells. Kurt et al. used a mutant variant of rAPOBEC1 (R33A) for linkage. MiniCGBE was also developed by removing UNG from the original CGBE, with a comparable but slightly lower efficiency. In 2022, the Liang group combined an ABE and CGBE to create an AGBE system; a new type of dual deaminase-mediated base editing system that could concurrently achieve four different base conversions (A-G, C-T, C-A, and C-G) in addition to indels with a single sgRNA. High-throughput screening may be utilized with AGBEs to create saturated mutants to evaluate the effects of various gene mutation patterns, including single-nucleotide variations (SNVs) and indels (Liang et al., 2022).

RNA base editors (ADAR)

Zhang and his group were the first to develop RNA base editing to perform conversion of bases at the RNA level by using a catalytically inactive Cas13 (dCas13) and a naturally occurring ADAR (adenosine deaminase acting on RNA) to direct adenosine to inosine conversion (Cox et al., 2017; Yarra and Sahoo, 2021). The RESCUE and REPAIR systems for RNA editing have been introduced for mammalian cells; however, in plants no REPAIR and RESCUE mechanisms for RNA editing have been employed (Bharat et al., 2019). These new technologies will greatly boost the application of the CRISPR system in plant RNA editing. The application of these two systems to crop enhancement requires future exploitation in rice and other crops (Bharat et al., 2019).

Targeting limitations of base editing

The target base must be present within a small base editing window for efficient base editing, and a specific PAM sequence is necessary (Gaudelli et al., 2017). This particular requirement for PAM is a strict restriction that reduces the editing efficiency in plant genomes. Modern ABE and CBE base editors are created with Cas9 variations, that can recognize PAM and NGG themes and because of this, the compatibility of PAM and the scope of basic editing has been increased (Endo et al., 2018; Nishimasu et al., 2018; Wang et al., 2019). The effectiveness of base editing is increased with these base editors, enhancing its applicability to a wide range of crops.

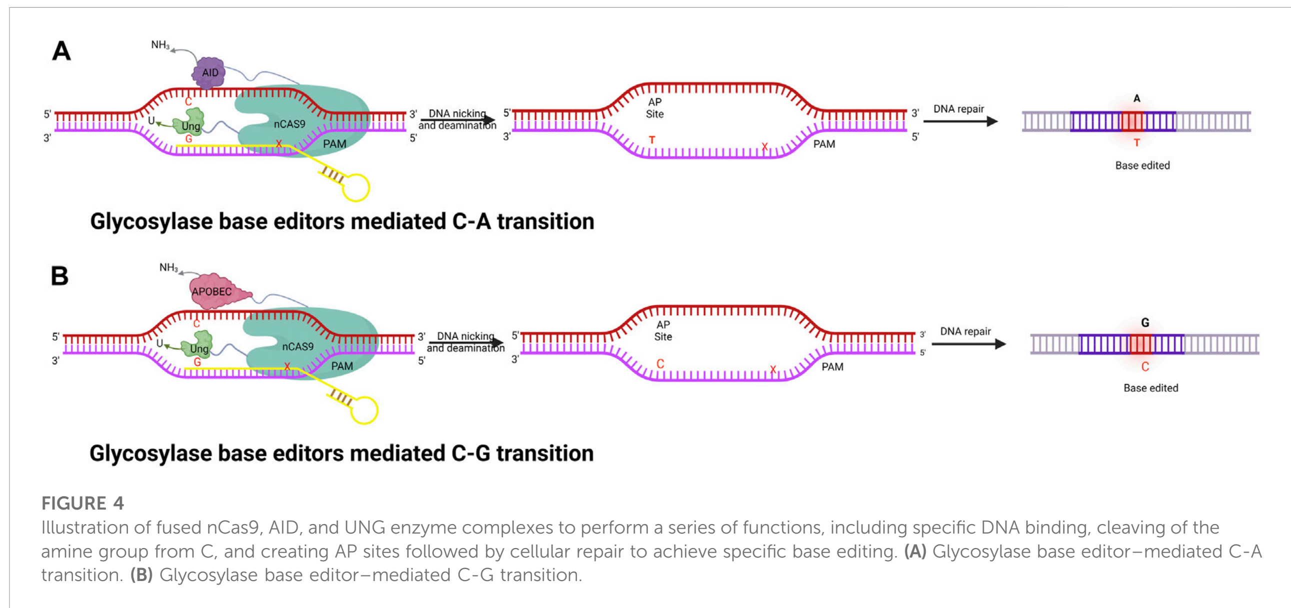


FIGURE 4
Illustration of fused nCas9, AID, and UNG enzyme complexes to perform a series of functions, including specific DNA binding, cleaving of the amine group from C, and creating AP sites followed by cellular repair to achieve specific base editing. **(A)** Glycosylase base editor–mediated C-A transition. **(B)** Glycosylase base editor–mediated C-G transition.

Size of catalytic window

Cytosine deaminase (base editor) may edit any C base pair over a wide range of nucleotides (5–9 nt) and this becomes a major concern, resulting in a low specificity and editing efficacy. Therefore, it is necessary to develop a highly precise base editor with a small window size that can efficiently edit a single C in a certain catalytic window. These probes are created by removing non-essential nucleotides from deaminase and evaluating different lengths of proline-rich linkers in order to narrow the catalytic range and improve the efficiency. Furthermore, GBE editing technology has allowed conversion with very high accuracy; however, only a small number of GBEs with wider target ranges were built. Therefore, these high-efficiency and high-precision basic editors are effective tools in crop breeding.

Off-target editing

Base editing (using CRISPR) is a recognized tool for base conversion. Previous research showed that the cleavage of on-target and off-target sites can be affected by different gRNA structures. Crystallography and single-molecule DNA curtain experiments showed that while the PAM site is essential to begin Cas9 binding, the sequence that corresponds to the 3′ end of the crRNA complementary recognition sequence, which is next to the PAM site, is also critical for subsequent Cas9 binding, activation of nuclease activities in Cas9, and R-loop formation. Off-target editing appears in this system when additional cytosines near the target base are edited. The activity of off-targets in BE3 is considerably decreased through the installment of mutations and the production of a high-fidelity base editor (HF-BE3) (Rees et al., 2017). CBEs, BE3, and HF1-BE3 have recently been found to cause unique and uncertain off-target alterations in rice (Jin et al., 2019). Such sudden alterations are commonly single nucleotide variants (SNV), from type C to type T. In order to mitigate these mutations, the literature suggests that it is obligatory to optimize UGI components and the cytidine deaminase domain. Furthermore, the modified CBE variant YEE-BE3 can be utilized in plants to decrease off-target editing (Jin et al., 2019).

Prime editing

A precise gene editing technique, prime editing (PE), is capable of carrying out targeted, small insertions, deletions, and base swapping. This seems quite similar to current CRISPR techniques. PE results have previously been attained in a variety of ways. The capability to remove base pairs is a hallmark of knock-outs, while the ability to add specific base pairs in a precise manner is the premise underlying knock-ins. The ability to perform focus editing without causing double-stranded DNA breaks is what distinguishes PE from standard CRISPR (Chen et al., 2019). Precise and dependable editing technologies are required to create non-DSB and template-free, genome-edited organisms. PE and BE can respond to the demand for precise and effective non-DSB and template-free editing systems. However, base editors cannot generate transversions, insertions, or removals. Precise insertions may be accomplished without donor DNA templates. The restricted range of the present base editing conversions (C>T, G>A, T>C, and A>G) is expanded with prime editing to include all 12 combination swaps. PE is a complete solution, with little

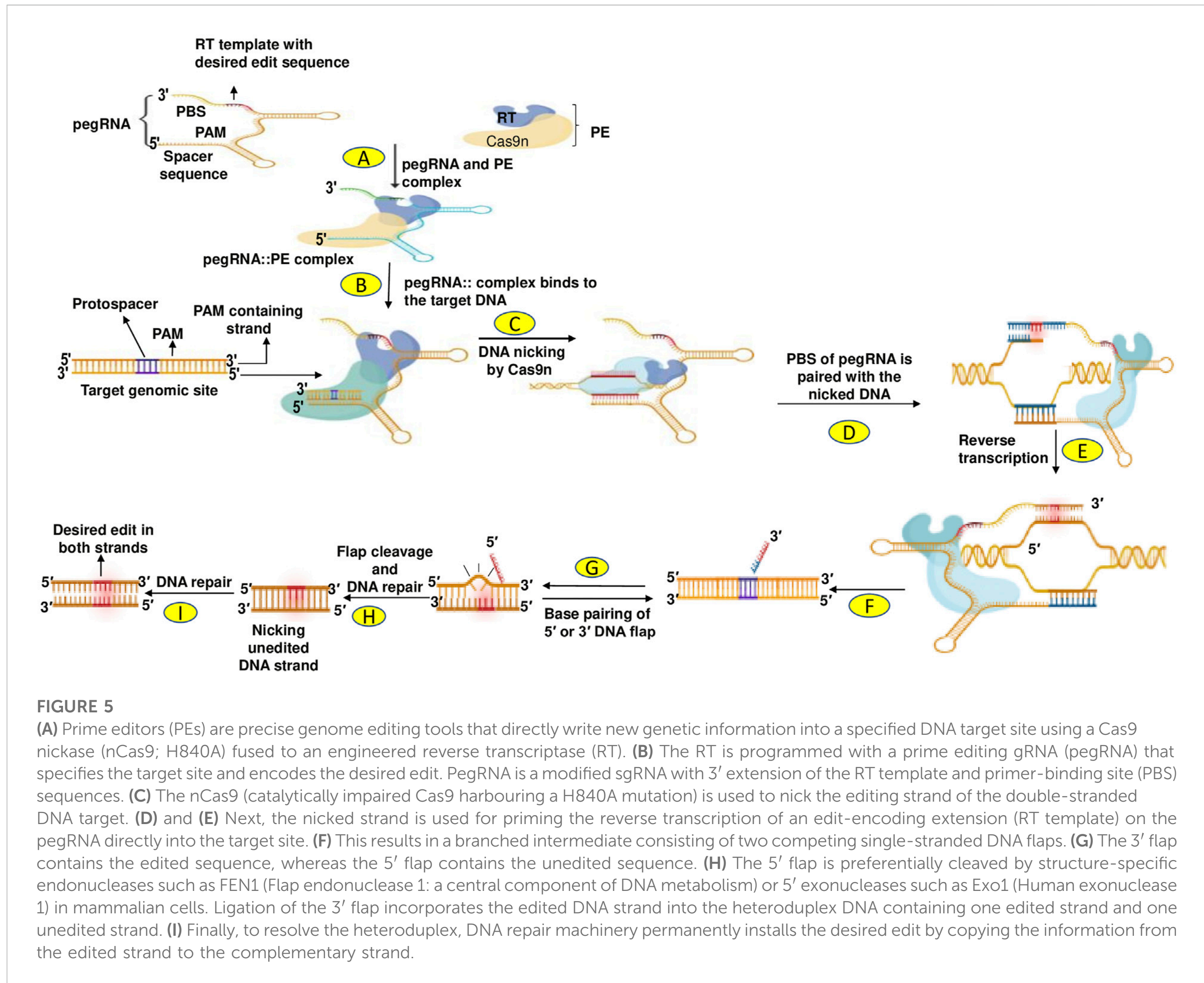


FIGURE 5
(A) Prime editors (PEs) are precise genome editing tools that directly write new genetic information into a specified DNA target site using a Cas9 nickase (nCas9; H840A) fused to an engineered reverse transcriptase (RT). **(B)** The RT is programmed with a prime editing gRNA (pegRNA) that specifies the target site and encodes the desired edit. PegRNA is a modified sgRNA with 3′ extension of the RT template and primer-binding site (PBS) sequences. **(C)** The nCas9 (catalytically impaired Cas9 harbouring a H840A mutation) is used to nick the editing strand of the double-stranded DNA target. **(D)** and **(E)** Next, the nicked strand is used for priming the reverse transcription of an edit-encoding extension (RT template) on the pegRNA directly into the target site. **(F)** This results in a branched intermediate consisting of two competing single-stranded DNA flaps. **(G)** The 3′ flap contains the edited sequence, whereas the 5′ flap contains the unedited sequence. **(H)** The 5′ flap is preferentially cleaved by structure-specific endonucleases such as FEN1 (Flap endonuclease 1: a central component of DNA metabolism) or 5′ exonucleases such as Exo1 (Human exonuclease 1) in mammalian cells. Ligation of the 3′ flap incorporates the edited DNA strand into the heteroduplex DNA containing one edited strand and one unedited strand. **(I)** Finally, to resolve the heteroduplex, DNA repair machinery permanently installs the desired edit by copying the information from the edited strand to the complementary strand.

CRISPR procedural enhancements that significantly influence the outcome—a typical case of the whole being more than the sum of its parts.

PE developed by Anzalone et al. (2019) allows all types of mutations, including insertions and deletions, to be implemented in base-to-base conversions (Marzec et al., 2020). The prime editing technique has been improved and successfully used in mammalian cells and plants, allowing targeted indels (insertions and deletions) and point mutations without breaking double strands or DNA donor repair templates (Lin P. et al., 2020; Tang et al., 2020). PE is a dynamic and precise GE technique that uses a Cas9 endonuclease with catalytic impairments complexed to a designed transcriptase, configured with prime edit RNA (pegRNA). This governs the target site and induces the desired edits to create new genetic modifications directly at a specific DNA site. PE proofreading showed maximum or comparable effectiveness with fewer by-products and fixes targets by homology which is complementary to the basic editing strengths and weaknesses and induces much lower off-target mutations compared to Cas9. PE significantly increases the range and capability of GE, and can fix up to 89% of the recognized gene mutations in humans (Anzalone et al., 2019). Because PE provides a wide range of different genome modification types, it has strong potential for a variety of purposes including yield improvement, quality enhancement of products, and resistance to various abiotic and biotic stresses (Hassan et al., 2021). The principal events in prime editing are highlighted in Figure 5.

Lin Q. et al. (2020) adapted prime editors for use in plants through codon, promoter, and editing-condition optimization. The resultant suite of plant prime editors enable point mutations, insertions, and deletions in rice and wheat protoplasts. Regenerated prime-edited rice plants were obtained at frequencies of up to 21.8%. Two parts in this system are pivotal: The prime editor and a PE guide RNA (pegRNA). The pegRNA contains a site that includes a complementary

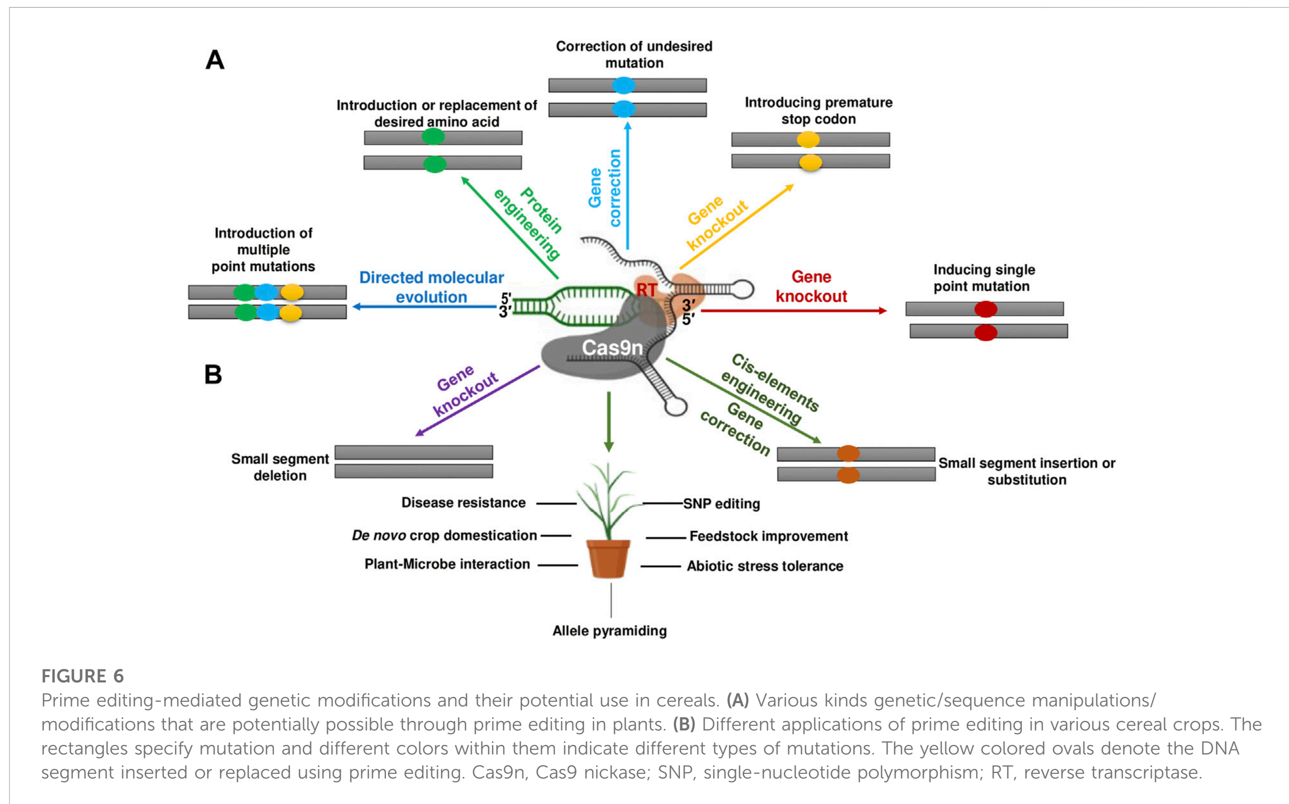


FIGURE 6
Prime editing-mediated genetic modifications and their potential use in cereals. **(A)** Various kinds genetic/sequence manipulations/modifications that are potentially possible through prime editing in plants. **(B)** Different applications of prime editing in various cereal crops. The rectangles specify mutation and different colors within them indicate different types of mutations. The yellow colored ovals denote the DNA segment inserted or replaced using prime editing. Cas9n, Cas9 nickase; SNP, single-nucleotide polymorphism; RT, reverse transcriptase.

DNA strand, a PBS (Primer Binding Site) (8–16 nt) sequence, and an RT-Template that carries the desired editing sequence, which may be replicated at the target location in the genome *via* reverse transcriptase. The prime editor, nickase Cas9 (Cas9n), possesses a mutant Cas9 protein that can cleave only one DNA strand. The editors also possess the necessary editing RT enzyme. The editor and pegRNA recombine during expression (transient or stable) and then travel toward the target site, led by pegRNA. The Cas9 nickase cleaves the PAM-containing strand to generate a single-strand DNA (ssDNA) flap; this process is directed by the target-specific pegRNA. The PBS, which itself is homologous to the ssDNA flap, intermixes with the RT blueprint and commences reverse transcription (RT), thus inserting sequences with the desired edit. After RT-mediated integration of the intended edit in the cleaved DNA molecule, the editing region comprises two duplicated ss-DNA flaps: an edited 3′ and un-edited 5′ DNA flaps. These ss-DNA flaps are eventually processed and integrated into the genome *via* endogenous DNA repair of the cell. The edited strand modifies the cleaved DNA strand by transferring sequence data from pegRNA, resulting in the development of a heteroduplex with one unedited and one edited strand. The second cleave is produced by using a standard guide, RNA, in the unmodified DNA strand, which is then fixed by transferring base pair information from the edited stand, resulting in the desired edit being integrated into both strands of the DNA. PE systems have the capacity to edit the genome efficiently and accurately, hence playing an important role in GE.

A wide range of changes at genomic sites can be efficiently produced in plants (rice and wheat). However, for effective and accurate edits it is important to optimize pegRNA designs and editing conditions. Using plant prime editors (PPEs) is an alternative way to induce mutations that cannot be generated by other plant GE tools. In plants PE is less effective at inducing transitional point mutations than base editors. PPEs can possibly produce insertions, deletions, replacements, and transversions. PE in plants is a versatile tool as it holds the potential to advance novel plant breeding and functional genomics research. Prime editing-mediated genetic modifications, and their potential use in cereals, are shown in Figure 6. The efficiency of PE in rice is demonstrated by developing herbicide resistance through targeting of the *OsALS* gene; furthermore, a PE2 editor was used to edit *OsIPA* and *OsTB1* (Butt et al., 2019). Prime editors are promising tools as they precisely edit endogenous genes and transgenic lines in rice; however, a low prime editor efficacy has been reported in some rice transgenic lines. Jiang et al. (2020) pioneered the editing of two non-allelic targets using PE in maize, and confirmed the hypothesis that enhanced pegRNA expression could improve the editing efficiency. In

summation, PE will broaden the scope and improve the capabilities of precise genome editing in important crops in future. Plant prime editing system optimization will empower the modification of crop genomes in a well-defined, cost-effective, and efficient manner, while fixing other superior agronomic traits. It should be noted that indels still occur in prime editing but with a frequency less than 1% in most cases. However, when using PE3, the indel frequency is generally less than 10%.

Conclusion

CRISPR has become one of the most flexible genetic engineering tools in recent decades and is used for a variety of genome editing applications. In comparison to traditional procedures and transgenic techniques, GE approaches are more cost-effective, faster, and accurate in attaining the desired crop improvements. This technology presents many other diverse advantages over traditional breeding techniques such as overcoming incompatibility barriers and efficiently modifying the genome. In recent years, the CRISPR/Cas9 genome editing technique has become widely used in crop research, especially to develop resilient cereal crops such as rice, barley, wheat, and maize. Genomic sequencing has been utilized to apply CRISPR/Cas systems to modify genomes for producing abiotic and biotic stress tolerant crops and to enhance crop yields as desired. Although off-target impacts must be considered, altering agriculturally important cereal crops may lead to a promising "ever green revolution" in the near future, addressing concerns such as nutrition uptake, nitrogen fixation, photosynthesis, climate change, and biofuel production. Precise gene insertion and sequence substitutions still remain a major obstacle for molecular breeding using CRISPR/Cas systems. These hindrances have been overcome using BE and PE to precisely and effectively introduce non-DSB and template-free publishing systems. With the progress of new BE technologies and the further enhancement of precision, unparalleled prospects are available for both plant agricultural advancement and biological research.

Future directions

Public acceptance and regulatory issues regarding CRISPR/Cas9 and its variants are still important issues to be resolved. The acceptance and wide application of this technology are still at the early stages and positive approaches to the related regulatory affairs may pave a way for global food and nutritional security. Looking ahead, at the global political level, people's interest in food and nutrition security is increasing, significantly impacting cereal research. In 2015, a sustainable development target to eliminate hunger by 2030 was set by the UN. As a result of this increased worldwide interest, funding for cereal research is expanding, which helps to promote the development of numerous novel methodologies. Recently, a directed evolution platform, based on CRISPR/Cas has been designed for plants. For example, the SF3B1 spliceosome protein resists splicing inhibitors in rice; different degrees of resistance to inhibitors are conferred by such mutant versions. To increase production yields and to improve resistance to abiotic and biotic stresses, the directed evolution platform is useful for engineering crops. It provides the possibility of cultivating weather-resistant crops and can enhance global food security. Resistance genes can now be cloned more rapidly owing to genomic approaches, as evidenced by the exponential growth in the number of resistance genes cloned in different crops and the simultaneous publication of multiple resistance genes. With the currently available resources, technologies, and those under development, there may be a similar expansion in understanding the molecular mechanisms of various traits in grains. The vast evolutionary genetic engineering-based modifications offered by the CRISPR/Cas technology has enhanced the pace of crop improvement and has reduced the threat of food insecurity at the global level.

To increase the purview of base editing, previous studies designed SpCas9-NG, xCas9, and SpCas9s variants in plants to expand the number of sites recognized by Cas9 (Endo et al., 2019). Further expansion of three optimized editors (AncBE4max, BE4max, and ABEmax) was completed with the help of bpNLS (codon-optimized dimerization nuclear localization signal) and was implemented in rice (Wang et al., 2019). Compared with known CBE and ABE editors, these base editors showed higher editing efficiencies. These upgraded base editors are beneficial for the molecular breeding approach. In many crops, DNA base editing technology is implemented to correct point mutations that are related to several traits. Therefore, in future, it is necessary to adopt new engineering variants in order to strengthen the current base editors, improve the efficacy of editing, and broaden the purview of basic editing, so as to be used in a variety of crops.

Genome editing-based PE technology aims to reduce the negative effects linked to other genome editing methods such as CRISPR-Cas9 or BE. PE does not require HDR or DSB when using exogenous donor DNA templates. Presently, advancements have been made in increasing the effectiveness of genome editing using the PE ribonucleoprotein complex. PE systems have evolved across four generations, each achieving a greater level of effectiveness. Recent data imply that *in vitro* screening of pegRNAs is crucial before conducting *in vivo* research, because this supports the potential application of PEs in repairing a wide range of mutations. However, PE also presents several difficulties, such as undesired mutations brought on by the double cleaving technique required by PE3, limitations regarding large DNA insertions, and the choice of ideal PBS and RT template combinations. Therefore, substantial advancements are required for generating more efficient

PPEs and extending their editing range. Because PE is still in its early developmental stages, much research has been focused on determining its efficiency and application in plant genome editing. The plant prime editing system can be used as an effective and universal technique in different crop species, providing a helpful tool for improving crops in a user-friendly manner. The modification of several precision genome editing tools for directed, accurate, and exact gene/allele replacement, in conjunction with classical breeding methods, will accelerate the breeding of diverse, superior crop varieties for maintainable agricultural development. Thus, we feel no hesitation in saying "To create a fully functional and high-precision genome editing tool, the prime editors must be optimized".

Author contributions

UB, SAA, MM, and SA conceived the idea and collected the material for writing this review. Each author contributed significantly, directly, and intellectually to the work, and gave their consent for it to be published.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* 9, 1911–1913. doi:10.1038/s41467-018-04252-2

Ainley, W. M., Sastry-Dent, L., Welter, M. E., Murray, M. G., Zeitler, B., Amora, R. V., et al. (2013). Trait stacking via targeted genome editing. *Plant Biotechnol. J.* 11 (9), 1126–1134. doi:10.1111/pbi.12107

Ansari, W. A., Chandanshive, S. U., Bhatt, V., Nadaf, A. B., Vats, S., Katara, J. L., et al. (2020). Genome editing in cereals: Approaches, applications andChallenges. *Int. J. Mol. Sci.* 21 (11), 4040. doi:10.3390/ijms21114040

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). Search-and-Replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157. doi:10.1038/s41586-019-1711-4

Arimura, S. I., Ayabe, H., Sugaya, H., Okuno, M., Tamura, Y., Tsuruta, Y., et al. (2020). Targeted gene disruption of ATP synthases 6-1 and 6-2 in the mitochondrial genome of *Arabidopsis thaliana* by mitoTALENs. *Plant J.* 104 (6), 1459–1471. doi:10.1111/tpj.15041

Banakar, R., Schubert, M., Collingwood, M., Vakulskas, C., Eggenberger, A. L., and Wang, K. (2020). Comparison of CRISPR-Cas9/Cas12a ribonucleoprotein complexes for genome editing efficiency in the rice phytoene desaturase (*OsPDS*) gene. *Rice* 13 (1), 4–7. doi:10.1186/s12284-019-0365-z

Bansal, M., Jindal, S., Wani, S. H., Ganie, S. A., and Singh, R. (2021). Genome editing and trait improvement in wheat. *Physiological, Mol. Genet. Perspect. Wheat Improv.* 2021, 263. doi:10.1007/978-3-030-59577-7_12

Becker, S., and Boch, J. (2021). TALE and TALEN genome editing technologies. *Gene Genome Ed.* 2, 100007–100283. doi:10.1016/j.ggedit.2021.100007

Bharat, S. S., Li, S., Li, J., Yan, L., and Xia, L. (2019). Base editing in plants: Current status and challenges. *Crop J.* 8 (3), 384–395. doi:10.1016/j.cj.2019.10.002

Bhowmik, P., Ellison, E., Polley, B., Bollina, V., Kulkarni, M., Ghanbarnia, K., et al. (2018). Targeted mutagenesis in wheat microspores using CRISPR/Cas9. *Sci. Rep.* 8 (1), 6502–6510. doi:10.1038/s41598-018-24690-8

Biswas, S., Tian, J., Li, R., Chen, X., Luo, Z., Chen, M., et al. (2020). Investigation of CRISPR/Cas9-induced SD1 rice mutants highlights the importance of molecularcharacterization in plant molecular breeding. *J. Genet. Genomics* 47 (5), 273–280. doi:10.1016/j.jgg.2020.04.004

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326 (5959), 1509–1512. doi:10.1126/science.1178811

Bogdanove, A. J., Schornack, S., and Lahaye, T. (2010). TAL effectors: Finding plant genes for disease and defense. *Curr. Opin. Plant Biol.* 13 (4), 394–401. doi:10.1016/j.pbi.2010.04.010

Borrill, P. (2020). Blurring the boundaries between cereal crops and model plants. *New Phytol.* 228 (6), 1721–1727. doi:10.1111/nph.16229

Brandt, K. M., Gunn, H., Moretti, N., and Zemetra, R. S. (2020). A streamlined protocol for wheat (*Triticum aestivum*) protoplast isolation and transformation with CRISPR-Cas ribonucleoprotein complexes. *Front. Plant Sci.* 11, 769. doi:10.3389/fpls.2020.00769

Butler, N. M., Atkins, P. A., Voytas, D. F., and Douches, D. S. (2015). Generation and inheritance of targeted mutations in potato (*Solanum tuberosum* L.) using the CRISPR/Cas system. *PloS one* 10 (12), e0144591. doi:10.1371/journal.pone.0144591

Butt, H., Eid, A., Momin, A. A., Bazin, J., Crespi, M., Arold, S. T., et al. (2019). CRISPR directed evolution of the spliceosome for resistance to splicing inhibitors. *Genome Biol.* 20 (1), 1–9. doi:10.1186/s13059-019-1680-9

Butt, H., Jamil, M., Wang, J. Y., Al-Babili, S., and Mahfouz, M. (2018). Engineering plant architecture via CRISPR/Cas9-mediated alteration of strigolactone biosynthesis. *BMC Plant Biol.* 18 (1), 174. doi:10.1186/s12870-018-1387-1

Cantos, C., Francisco, P., Trijatmiko, K. R., Slamet-Loedin, I., and Chadha-Mohanty, P. K. (2014). Identification of "safe harbor" loci in indica rice genome by harnessing the property of zinc-finger nucleases to induce DNA damage and repair. *Front. Plant Sci.* 5, 302. doi:10.3389/fpls.2014.00302

Carroll, D. (2008). Progress and prospects: Zinc-finger nucleases as gene therapy agents. *Gene Ther.* 15 (22), 1463–1468. doi:10.1038/gt.2008.145

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., et al. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 39 (12), e82. doi:10.1093/nar/gkr218

Cermak, T., Starker, C. G., and Voytas, D. F. (2015). "Efficient design and assembly of custom TALENs using the Golden Gate platform," in *Chromosomal mutagenesis* (New York, NY: Humana Press), 133–159.

Char, S. N., Unger-Wallace, E., Frame, B., Briggs, S. A., Main, M., Spalding, M. H., et al. (2015). Heritable site-specific mutagenesis using TALEN s in maize. *Plant Biotechnol. J.* 13 (7), 1002–1010. doi:10.1111/pbi.12344

Char, S. N., Wei, J., Mu, Q., Li, X., Zhang, Z. J., Yu, J., et al. (2020). An Agrobacterium-delivered CRISPR/Cas9 system for targeted mutagenesis in sorghum. *Plant Biotechnol. J.* 18 (2), 319–321. doi:10.1111/pbi.13229

Char, S. N., and Yang, B. (2020). Genome editing in grass plants. *aBIOTECH* 1 (1), 41–57. doi:10.1007/s42994-019-00005-x

Che, P., Anand, A., Wu, E., Sander, J. D., Simon, M. K., Zhu, W., et al. (2018). Developing a flexible, high-efficiency Agrobacterium-mediated sorghum transformation system with broad application. *Plant Biotechnol. J.* 16 (7), 1388–1395. doi:10.1111/pbi.12879

Chen, K., Ke, R., Du, M., Yi, Y., Chen, Y., Wang, X., et al. (2022). A FLASH pipeline for arrayed CRISPR library construction and the gene function discovery of rice receptor-like kinases. *Mol. Plant* 15 (2), 243–257. doi:10.1016/j.molp.2021.09.015

Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019). CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* 70, 667–697. doi:10.1146/annurev-arplant-050718-100049

Chen, R., Xu, Q., Liu, Y., Zhang, J., Ren, D., Wang, G., et al. (2018). Generation of transgene-free maize male sterile lines using the CRISPR/Cas9 system. *Front. Plant Sci.* 9, 1180. doi:10.3389/fpls.2018.01180

Choo, Y., and Isalan, M. (2000). Advances in zinc finger engineering. *Curr. Opin. Struct. Biol.* 10 (4), 411–416. doi:10.1016/s0959-440x(00)00107-x

Christian, M. L., Demorest, Z. L., Starker, C. G., Osborn, M. J., Nyquist, M. D., Zhang, Y., et al. (2012). Targeting G with TAL effectors: A comparison of activities of TALENs constructed with NN and NK repeat variable di-residues. *PLos One* 7, e45383. doi:10.1371/journal.pone.0045383

Cox, D. B., Gootenberg, J. S., Abudayyeh, O., Franklin, B., Kellner, M. J., Joung, J., et al. (2017). RNA editing with CRISPR-Cas13. *Science* 358 (6366), 1019–1027. doi:10.1126/science.aaq0180

Cui, X., Balcerzak, M., Schernthaner, J., Babic, V., Datla, R., Brauer, E. K., et al. (2019). An optimised CRISPR/Cas9 protocol to create targeted mutations in homoeologous genes and an efficient genotyping protocol to identify edited events in wheat. *Plant methods* 15 (1), 119–212. doi:10.1186/s13007-019-0500-2

Daboussi, F., Stoddard, T. J., and Zhang, F. (2015). “Engineering meganuclease for precise plant genome modification,” in *Advances in new technology for targeted modification of plant genomes* (New York, NY: Springer), 21–38.

Davies, J. P., Kumar, S., and Sastry-Dent, L. (2017). Use of zinc-finger nucleases for crop improvement. *Prog. Mol. Biol. Transl. Sci.* 149, 47–63. doi:10.1016/bs.pmbts.2017.03.006

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J. K., et al. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335 (6069), 720–723. doi:10.1126/science.1215670

Dong, O. X., and Ronald, P. C. (2021). Targeted DNA insertion in plants. *Proc. Natl. Acad. Sci. U. S. A.* 118 (22). doi:10.1073/pnas.2004834117

Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M. H., and Chandrasegaran, S. (2005). Zinc finger nucleases: Custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res.* 33 (18), 5978–5990.

Eid, A., Alshareef, S., and Mahfouz, M. M. (2018). CRISPR base editors: Genome editing without double-stranded breaks. *Biochem. J.* 475 (11), 1955–1964. doi:10.1042/BCJ20170793

Endo, M., Mikami, M., and Endo, A. (2018). Genome editing in plants by engineered CRISPR- Cas9 recognizing NG PAM. *Nat. Plants*, 21–27.

Endo, M., Mikami, M., Endo, A., Kaya, H., Itoh, T., Nishimasu, H., et al. (2019). Genome editing in plants by engineered CRISPR–Cas9 recognizing NG PAM. *Nat. Plants* 5 (1), 14–17. doi:10.1038/s41477-018-0321-8

Feng, C., Yuan, J., Wang, R., Liu, Y., Birchler, J. A., and Han, F. (2016). Efficient targeted genome modification in maize using CRISPR/Cas9 system. *J. Genet. Genomics* 43 (1), 37–43. doi:10.1016/j.jgg.2015.10.002

Forsyth, A., Weeks, T., Richael, C., and Duan, H. (2016). Transcription activator-like effector nucleases (TALEN)-mediated targeted DNA insertion in potato plants. *Front. Plant Sci.* 7, 1572. doi:10.3389/fpls.2016.01572

Galli, M., Martiny, E., Imani, J., Kumar, N., Koch, A., Steinbrenner, J., et al. (2022). CRISPR/Sp cas9-mediated double knockout of barley microrchidia MORC1 and MORC6a reveals their strong involvement in plant immunity, transcriptional gene silencing and plant growth. *Plant Biotechnol. J.* 20, 89–102. doi:10.1111/pbi.13697

Gao, H., Gadlage, M. J., Lafitte, H. R., Lenderts, B., Yang, M., Schroder, M., et al. (2020). Superior field performance of waxy corn engineered using CRISPR–Cas9. *Nat. Biotechnol.* 38 (5), 579–581. doi:10.1038/s41587-020-0444-0

Gasparis, S., Kała, M., Przyborowski, M., Łyżnik, L. A., Orczyk, W., and Nadolska-Orczyk, A. (2018). A simple and efficient CRISPR/Cas9 platform for induction of single and multiple, heritable mutations in barley (Hordeum vulgare L.). *Plant methods* 14 (1), 111. doi:10.1186/s13007-018-0382-8

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., et al. (2017). Programmable base editing of A● T to G● C in genomic DNA without DNA cleavage. *Nature* 551 (7681), 464–471. doi:10.1038/nature24644

Gurushidze, M., Hensel, G., Hiekel, S., Schedel, S., Valkov, V., and Kumlehn, J. (2014). True-breeding targeted gene knock-out in barley using designer TALE-nuclease in haploid cells. *PloS one* 9 (3), e92046. doi:10.1371/journal.pone.0092046

Hamada, H., Liu, Y., Nagira, Y., Miki, R., Taoka, N., and Imai, R. (2018). Biolistic-delivery- based transient CRISPR/Cas9 expression enables in planta genome editing in wheat. *Sci. Rep.* 8 (1), 14422–14427. doi:10.1038/s41598-018-32714-6

Hao, L. I., Ruiying, Q. I. N., Xiaoshuang, L. I. U., Shengxiang, L. I. A. O., Rongfang, X. U., Jianbo, Y. A. N. G., et al. (2019). CRISPR/Cas9-mediated adenine base editing in rice genome. *Rice Sci.* 26 (2), 125–128. doi:10.1016/j.rsci.2018.07.002

Haroon, M., Wang, X., Afzal, R., Zafar, M. M., Idrees, F., Batool, M., et al. (2022). Novel plant breeding techniques shake hands with cereals to increase production. *Plants* 11 (8), 1052. doi:10.3390/plants11081052

Hassan, M. M., Zhang, Y., Yuan, G., De, K., Chen, J. G., Muchero, W., et al. (2021). Construct design for CRISPR/Cas-based genome editing in plants. *Trends Plant Sci.* 26 (11), 1133–1152. doi:10.1016/j.tplants.2021.06.015

He, Y., and Zhao, Y. (2020). Technological breakthroughs in generating transgene-free and genetically stable CRISPR-edited plants. *aBIOTECH* 1 (1), 88–96. doi:10.1007/s42994-019-00013-x

He, Y., Zhu, M., Wang, L., Wu, J., Wang, Q., Wang, R., et al. (2018). Programmed self-elimination of the CRISPR/Cas9 construct greatly accelerates the isolation of edited and transgene-free rice plants. *Mol. Plant* 11 (9), 1210–1213. doi:10.1016/j.molp.2018.05.005

Hess, G. T., Frésard, L., Han, K., Lee, C. H., Li, A., Cimprich, K. A., et al. (2016). Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* 13 (12), 1036–1042. doi:10.1038/nmeth.4038

Holubová, K., Hensel, G., Vojta, P., Tarkowski, P., Bergougnoux, V., and Galuszka, P. (2018). Modification of barley plant productivity through regulation of cytokinin content by reverse-genetics approaches. *Front. Plant Sci.* 9, 1676. doi:10.3389/fpls.2018.01676

Howells, R. M., Craze, M., Bowden, S., and Wallington, E. J. (2018). Efficient generation of stable, heritable gene edits in wheat using CRISPR/Cas9. *BMC Plant Biol.* 18 (1), 215. doi:10.1186/s12870-018-1433-z

Hu, B., Wang, W., Ou, S., Tang, J., Li, H., Che, R., et al. (2015). Variation in NRT1.1B contributes to nitrate-use divergence between rice subspecies.1 B contributes to nitrate-use divergence between rice subspecies. *Nat. Genet.* 47 (7), 834–838. doi:10.1038/ng.3337

Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., et al. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 556 (7699), 57–63. doi:10.1038/nature26155

Hu, X., Wang, C., Fu, Y., Liu, Q., Jiao, X., and Wang, K. (2016). Expanding the range of CRISPR/Cas9 genome editing in rice. *Mol. Plant* 9 (6), 943–945. doi:10.1016/j.molp.2016.03.003

Hua, K., Tao, X., Han, P., Wang, R., and Zhu, J. K. (2019). Genome engineering in rice using Cas9 variants that recognize NG PAM sequences. Cas9 variants that recognize NG PAM sequences. *Mol. Plant* 12 (7), 1003–1014. doi:10.1016/j.molp.2019.03.009

Hua, K., Tao, X., Liang, W., Zhang, Z., Gou, R., and Zhu, J. K. (2020). Simplified adenine base editors improve adenine base editing efficiency in rice. *Plant Biotechnol. J.* 18 (3), 770–778. doi:10.1111/pbi.13244

Hua, K., Tao, X., Yuan, F., Wang, D., and Zhu, J. K. (2018). Precise A·T to G·C base editing in the rice genome. *Mol. Plant* 11 (4), 627–630. doi:10.1016/j.molp.2018.02.007

Hua, K., Tao, X., and Zhu, J. K. (2019). Expanding the base editing scope in rice by using Cas9 variants. *Plant Biotechnol. J.* 17 (2), 499–504. doi:10.1111/pbi.12993

Hunziker, J., Nishida, K., Kondo, A., Kishimoto, S., Ariizumi, T., and Ezura, H. (2020). Multiple gene substitution by Target-AID base-editing technology in tomato. *Sci. Rep.* 10 (1), 20471–20512. doi:10.1038/s41598-020-77379-2

Jacquier, A., and Dujon, B. (1985). An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* 41 (2), 383–394. doi:10.1016/s0092-8674(85)80011-8

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., and Weeks, D. P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res.* 41 (20), e188. doi:10.1093/nar/gkt780

Jiang, Z., Hong, X., Zhang, S., Yao, R., and Xiao, Y. (2020). CRISPR base editing and prime editing: DSB and template-free editing systems for bacteria and plants. *Synth. Syst. Biotechnol.* 5 (4), 277–292. doi:10.1016/j.synbio.2020.08.003

Jin, S., Zong, Y., Gao, Q., Zhu, Z., Wang, Y., Qin, P., et al. (2019). Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* 364 (6437), 292–295. doi:10.1126/science.aaw7166

Jung, Y. J., Nogoy, F. M., Lee, S. K., and Kang, K. K. (2018). Application of ZFN for site directed mutagenesis of rice SSIVa gene mutagenesis of rice SSIVa gene. *Biotechnol. Bioprocess Eng.* 23, 108–115. doi:10.1007/s12257-017-0420-9

Kang, B. C., Yun, J. Y., Kim, S. T., Shin, Y., Ryu, J., Choi, M., et al. (2018). Precision genome engineering through adenine base editing in plants. *Nat. Plants* 4 (7), 427–431. doi:10.1038/s41477-018-0178-x

Kannan, B., Jung, J. H., Moxley, G. W., Lee, S. M., and Altpeter, F. (2018). TALEN-mediated targeted mutagenesis of more than 100 COMT copies/alleles in highly polyploid sugarcane improves saccharification efficiency without compromising biomass yield. *Plant Biotechnol. J.* 16 (4), 856–866. doi:10.1111/pbi.12833

Kapusi, E., Corcuera-Gómez, M., Melnik, S., and Stoger, E. (2017). Heritable genomic fragment deletions and small indels in the putative ENGase gene induced by CRISPR/Cas9 in barley. *Front. Plant Sci.* 8, 540. doi:10.3389/fpls.2017.00540

Kelliher, T., Starr, D., Richbourg, L., Chintamanani, S., Delzer, B., Nuccio, M. L., et al. (2017). MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature* 542 (7639), 105–109. doi:10.1038/nature20827

Khan, S. H. (2019). Genome-editing technologies: Concept, pros, and cons of various genome- editing techniques and bioethical concerns for clinical application. *Mol. Ther. Nucleic Acids* 16, 326–334. doi:10.1016/j.omtn.2019.02.027

Kim, C. Y., Park, J. Y., Choi, G., Kim, S., Vo, K. T. X., Jeon, J. S., et al. (2022). A rice gene encoding glycosyl hydrolase plays contrasting roles in immunity depending on the type of pathogens. *Mol. Plant Pathol.* 23, 400–416. doi:10.1111/mpp.13167

Kim, Y. B., Komor, A. C., Levy, J. M., Packer, M. S., Zhao, K. T., and Liu, D. R. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* 35 (4), 371–376. doi:10.1038/nbt.3803

Kim, Y. G., Shi, Y., Berg, J. M., and Chandrasegaran, S. (1997). Site-specific cleavage of DNA-RNA hybrids by zinc finger/FokI cleavage domain fusions. RNA hybrids by zinc finger/*Fok*I cleavage domain fusions. *Gene* 203 (1), 43–49. doi:10.1016/s0378-1119(97)00489-7

Kis, A., Hamar, É., Tholt, G., Bán, R., and Havelda, Z. (2019). Creating highly efficient resistance against wheat dwarf virus in barley by employing CRISPR/Cas9 system. *Plant Biotechnol. J.* 17 (6), 1004–1006. doi:10.1111/pbi.13077

Koblan, L. W., Arbab, M., Shen, M. W., Hussmann, J. A., Anzalone, A. V., Doman, J. L., et al. (2021). Efficient C● G-to-G● C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nat. Biotechnol.* 39 (11), 1414–1425. doi:10.1038/s41587-021-00938-z

Koblan, L. W., Doman, J. L., Wilson, C., Levy, J. M., Tay, T., Newby, G. A., et al. (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* 36 (9), 843–846. doi:10.1038/nbt.4172

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533 (7603), 420–424. doi:10.1038/nature17946

Komor, A. C., Zhao, K. T., Packer, M. S., Gaudelli, N. M., Waterbury, A. L., Koblan, L. W., et al. (2017). Improved base excision repair inhibition and bacteriophage Mu gam protein yields C: G-to-T: A base editors with higher efficiency and product purity. *Sci. Adv.* 3 (8), eaao4774. doi:10.1126/sciadv.aao4774

Kumar, N., Galli, M., Ordon, J., Stuttmann, J., Kogel, K. H., and Imani, J. (2018). Further analysis of barley MORC 1 using a highly efficient RNA-guided Cas9 gene-editing system. *Plant Biotechnol. J.* 16 (11), 1892–1903. doi:10.1111/pbi.12924

Kumar, P., Alok, A., and Kumar, J. (2020). "Expanding the potential of CRISPR-cas9 technology for crops improvement," in *Advances in synthetic biology* (Singapore: Springer), 251–274.

Kurt, I. C., Zhou, R., Iyer, S., Garcia, S. P., Miller, B. R., Langner, L., et al. (2021). CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* 39 (1), 41–46. doi:10.1038/s41587-020-0609-x

Lau, C. H., Zhu, H., Tay, J. C. K., Li, Z., Tay, F. C., Chen, C., et al. (2014). Genetic rearrangements of variable di-residue (RVD)-containing repeat arrays in a baculoviral TALEN system. *Mol. Ther. Methods Clin. Dev.* 1, 14050. doi:10.1038/mtm.2014.50

Lawrenson, T., and Harwood, W. A. (2019). "Creating targeted gene knockouts in barley using CRISPR/Cas9," in *Barley* (New York, NY: Humana Press), 217–232.

Lawrenson, T., Shorinola, O., and Stacey, N. (2015). Induction of targeted, heritable mutations in barley and *Brassica oleracea* using RNA-guided Cas9 nuclease. *Nat. Commun.* 16, 3048. doi:10.1038/s41467-018-05477-x

Lee, J. K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y. H., et al. (2018). Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.* 9 (1), 1–10. doi:10.1038/s41467-018-05477-x

Li, A., Jia, S., Yobi, A., Ge, Z., Sato, S. J., Zhang, C., et al. (2018). Editing of an alpha-kafirin gene family increases, digestibility and protein quality in sorghum. *Plant Physiol.* 177 (4), 1425–1438. doi:10.1104/pp.18.00200

Li, C., Zong, Y., Wang, Y., Jin, S., Zhang, D., Song, Q., et al. (2018). Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol.* 19 (1), 59. doi:10.1186/s13059-018-1443-z

Li, H., Yang, Y., Hong, W., Huang, M., Wu, M., and Zhao, X. (2020). Applications of genome editing technology in the targeted therapy of human diseases: Mechanisms, advances and prospects. *Signal Transduct. Target. Ther.* 5 (1), 1–23. doi:10.1038/s41392-019-0089-y

Li, J., Meng, X., Zong, Y., Chen, K., Zhang, H., Liu, J., et al. (2016). Gene replacements and insertions in rice by intron targeting using CRISPR–Cas9. *Nat. Plants* 2 (10), 1–6. doi:10.1038/nplants.2016.139

Li, J., Sun, Y., Du, J., Zhao, Y., and Xia, L. (2017). Generation of targeted point mutations in rice by a modified CRISPR/Cas9 system. *Mol. Plant* 10 (3), 526–529. doi:10.1016/j.molp.2016.12.001

Li, S., Zhang, Y., Xia, L., and Qi, Y. (2020). CRISPR-Cas12a enables efficient biallelic gene targeting in rice. *Plant Biotechnol. J.* 18 (6), 1351–1353. doi:10.1111/pbi.13295

Li, T., Liu, B., Spalding, M. H., Weeks, D. P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.* 30 (5), 390–392. doi:10.1038/nbt.2199

Li, X., Wang, Y., Liu, Y., Yang, B., Wang, X., Wei, J., et al. (2018). Base editing with a Cpf1–cytidine deaminase fusion. *Nat. Biotechnol.* 36 (4), 324–327. doi:10.1038/nbt.4102

Liang, Y., Xie, J., Zhang, Q., Wang, X., Gou, S., Lin, L., et al. (2022). Agbe: A dual deaminase-mediated base editor by fusing CGBE with ABE for creating a saturated mutant population with multiple editing patterns. *Nucleic Acids Res.* 50 (9), 5384–5399. doi:10.1093/nar/gkac353

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., et al. (2017). Efficient DNA- free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nat. Commun.* 8 (1), 14261–14265. doi:10.1038/ncomms14261

Liang, Z., Chen, K., Zhang, Y., Liu, J., Yin, K., Qiu, J. L., et al. (2018). Genome editing of bread wheat using biolistic delivery of CRISPR/Cas9 *in vitro* transcripts or ribonucleoproteins. *Nat. Protoc.* 13 (3), 413–430. doi:10.1038/nprot.2017.145

Liang, Z., Zhang, K., Chen, K., and Gao, C. (2014). Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. *J. Genet. Genomics* 41 (2), 63–68. doi:10.1016/j.jgg.2013.12.001

Lin, P., Qin, S., Pu, Q., Wang, Z., Wu, Q., Gao, P., et al. (2020a). CRISPR-Cas13 inhibitors block RNA editing in bacteria and mammalian cells. *Mol. Cell* 78 (5), 850–861.

Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., et al. (2020b). Prime genome editing in rice and wheat. *Nat. Biotechnol.* 38 (5), 582–585. doi:10.1038/s41587-020-0455-x

Liu, G., Li, J., and Godwin, I. D. (2019). "Genome editing by CRISPR/Cas9 in sorghum through biolistic bombardment," in *Sorghum* (New York, NY: Humana Press), 169–183.

Liu, L., Gallagher, J., Arevalo, E. D., Chen, R., Skopelitis, T., Wu, Q., et al. (2021). Enhancing grain-yield-related traits by CRISPR–Cas9 promoter editing of maize CLE genes. *Nat. Plants* 7 (3), 287–294. doi:10.1038/s41477-021-00858-5

Liu, X., Ding, Q., Wang, W., Pan, Y., Tan, C., Qiu, Y., et al. (2022). Targeted deletion of the first intron of the Wxb allele via CRISPR/Cas9 significantly increases grain amylose content in rice. *Rice* 15 (1), 1–12. doi:10.1186/s12284-021-00548-y

Liu, X., Zhou, X., Li, K., Wang, D., Ding, Y., Liu, X., et al. (2020). A simple and efficient cloning system for CRISPR/Cas9-mediated genome editing in rice. *PeerJ* 8, e8491. doi:10.7717/peerj.8491

Liu, Y., Chen, L., Liu, Y., Dai, H., He, J., Kang, H., et al. (2016). Marker assisted pyramiding of two Brown planthopper resistance genes, Bph3 and Bph27 (t), into elite rice cultivars. *Rice* 9 (1), 27–7. doi:10.1186/s12284-016-0096-3

Liu, Z., Chen, S., Shan, H., Zhang, Q., Chen, M., Lai, L., et al. (2019). Efficient and precise base editing in rabbits using human APOBEC3A-nCas9 fusions. *Cell Discov.* 5 (1), 31–34. doi:10.1038/s41421-019-0099-5

Liu, Z., Lu, Z., Yang, G., Huang, S., Li, G., Feng, S., et al. (2018). Efficient generation of mouse models of human diseases via ABE-and BE-mediated base editing. *Nat. Commun.* 9 (1), 2338–8. doi:10.1038/s41467-018-04768-7

Liu, D., Wang, H., and Yu, D. (2018). The sucrose non-fermenting-1-related protein kinases SAPK1 and SAPK2 function collaboratively as positive regulators of

salt stress tolerance in rice. SAPK1 and SAPK2 function collaboratively as positive regulators of salt stress tolerance in rice. *BMC Plant Biol.* 18 (1), 203. doi:10.1186/s12870-018-1408-0

Lu, Y., and Zhu, J. K. (2017). Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. *Mol. Plant* 10 (3), 523–525. doi:10.1016/j.molp.2016.11.013

Luo, M., Li, H., Chakraborty, S., Morbitzer, R., Rinaldo, A., Upadhyaya, N., et al. (2019). Efficient TALEN-mediated gene editing in wheat. *Plant Biotechnol. J.* 17 (11), 2026–2028. doi:10.1111/pbi.13169

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., et al. (2015). A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Mol. Plant* 8 (8), 1274–1284. doi:10.1016/j.molp.2015.04.007

Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* 13 (12), 1029–1035. doi:10.1038/nmeth.4027

Macovei, A., Sevilla, N. R., Cantos, C., Jonson, G. B., Slamet-Loedin, I., Čermák, T., et al. (2018). Novel alleles of rice eIF4G generated by CRISPR/Cas9-targeted mutagenesis confer resistance to Rice tungro spherical virus. *Plant Biotechnol. J.* 16 (11), 1918–1927. doi:10.1111/pbi.12927

Mahfouz, M. M., Piatek, A., and Stewart, C. N., Jr (2014). Genome engineering via TALENs and CRISPR/Cas9 systems: Challenges and perspectives. *Plant Biotechnol. J.* 12 (8), 1006–1014. doi:10.1111/pbi.12256

Mak, A. N. S., Bradley, P., Bogdanove, A. J., and Stoddard, B. L. (2013). TAL effectors: Function, structure, engineering and applications. *Curr. Opin. Struct. Biol.* 23 (1), 93–99. doi:10.1016/j.sbi.2012.11.001

Malzahn, A., Lowder, L., and Qi, Y. (2017). Plant genome editing with TALEN and CRISPR. *Cell Biosci.* 7, 21. doi:10.1186/s13578-017-0148-4

Manghwar, H., Lindsey, K., Zhang, X., and Jin, S. (2019). CRISPR/Cas system: Recent advances and future prospects for genome editing. *Trends Plant Sci.* 24 (12), 1102–1125. doi:10.1016/j.tplants.2019.09.006

Marzec, M., Brąszewska-Zalewska, A., and Hensel, G. (2020). Prime editing: A new way for genome editing. *Trends Cell Biol.* 30 (4), 257–259. doi:10.1016/j.tcb.2020.01.004

Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., et al. (2013). Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res.* 23 (10), 1233–1236. doi:10.1038/cr.2013.123

Mishra, R., Joshi, R. K., and Zhao, K. (2020). Base editing in crops: Current advances, limitations and future implications. *Plant Biotechnol. J.* 18 (1), 20–31. doi:10.1111/pbi.13225

Mishra, R., Mushtaq, M., and andJoshi, R. K. (2022). "Expanding the scope of base editing in CropsUsing Cas9 variants," in *Genome editing technologies for crop improvement*. Editors K. Zhao, R. Mishra, and R. K. Joshi (Singapore: Springer). doi:10.1007/978-981-19-0600-8_8

Moore, J. K., and Haber, J. E. (1996). Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 16 (5), 2164–2173. doi:10.1128/mcb.16.5.2164

Moscou, M. J., and Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* 326 (5959), 1501. doi:10.1126/science.1178817

Mushtaq, M., Dar, A. A., Basu, U., Bhat, B. A., Mir, R. A., Vats, S., et al. (2021b). Integrating CRISPR-cas and next generation sequencing in plant virology. *Front. Genet.* 12, 735489. doi:10.3389/fgene.2021.735489

Mushtaq, M., Ahmad Dar, A., Skalicky, M., Tyagi, A., Bhagat, N., Basu, U., et al. (2021a). CRISPR-based genome editing tools: Insights into technological breakthroughs and future challenges. *Genes* 12 (6), 797. doi:10.3390/genes12060797

Mushtaq, M., Bhat, J. A., Mir, Z. A., Sakina, A., Ali, S., Singh, A. K., et al. (2018). CRISPR/Cas approach: A new way of looking at plant-abiotic interactions. *J. Plant Physiol.* 224, 156–162. doi:10.1016/j.jplph.2018.04.001

Mushtaq, M., and Molla, K. A. (2021). "CRISPR technologies for plant biotechnology innovation," in *Plant biotechnology* (Cham: Springer), 51–67.

Mushtaq, M., Mukhtar, S., Sakina, A., Dar, A. A., Bhat, R., Deshmukh, R., et al. (2020). Tweaking genome-editing approaches for virus interference in crop plants. *Plant Physiol. biochem.* 147, 242–250. doi:10.1016/j.plaphy.2019.12.022

Mushtaq, M., Sakina, A., Wani, S. H., Shikari, A. B., Tripathi, P., Zaid, A., et al. (2019). Harnessing genome editing techniques to engineer disease resistance in plants. *JAMA Cardiol.* 10, 550. doi:10.3389/fpls.2019.00550

Musunuru, K. (2017). The hope and hype of CRISPR-cas9 genome editing: A review. *JAMA Cardiol.* 2 (8), 914–919. doi:10.1001/jamacardio.2017.1713

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353 (6305), aaf8729. doi:10.1126/science.aaf8729

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., et al. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* 361 (6408), 1259–1262. doi:10.1126/science.aas9129

Nouspikel, T. (2009). DNA repair in mammalian cells. *Cell. Mol. Life Sci. Sci.* 66 (6), 994–1009. doi:10.1007/s00018-009-8737-y

Pabo, C. O., Peisach, E., and Grant, R. A. (2001). Design and selection of novel Cys2His2 zinc finger proteins. *Annu. Rev. Biochem.* 70 (1), 313–340. doi:10.1146/annurev.biochem.70.1.313

Panting, M., Holme, I. B., Björnsson, J. M., Zhong, Y., and Brinch-Pedersen, H. (2021). CRISPR/Cas9 and transgene verification of gene involvement in unfolded protein response and recombinant protein production in barley grain. *Front. Plant Sci.* 12, 755788. doi:10.3389/fpls.2021.755788

Papworth, M., Kolasinska, P., and andMinczuk, M. (2006). Designer zinc-finger proteins and their applications. *Gene* 366 (1), 27–38. doi:10.1016/j.gene.2005.09.011

Pennisi, E. (2013). The CRISPR craze. *Science* 341 (6148), 833–836. doi:10.1126/science.341.6148.833

Poutanen, K. S., Kårlund, A. O., Gómez-Gallego, C., Johansson, D. P., Scheers, N. M., Marklinder, I. M., et al. (2022). Grains–a major source of sustainable protein for health. *Nutr. Rev.* 80 (6), 1648–1663. doi:10.1093/nutrit/nuab084

Puchta, H. (2017). Applying CRISPR/cas for genome engineering in plants: The best is yet to come. *Curr. Opin. Plant Biol.* 36, 1–8. doi:10.1016/j.pbi.2016.11.011

Qin, L., Li, J., Wang, Q., Xu, Z., Sun, L., Alariqi, M., et al. (2020). High-efficient and precise base editing of C● G to T● A in the allotetraploid cotton (*Gossypiumhirsutum*) genome using a modified CRISPR/Cas9 system. *Plant Biotechnol. J.* 18 (1), 45–56. doi:10.1111/pbi.13168

Raitskin, O., Schudoma, C., West, A., and Patron, N. J. (2019). Comparison of efficiency and specificity of CRISPR-associated (Cas) nucleases in plants: An expanded toolkit for precision genome engineering. *PloS one* 14 (2), e0211598. doi:10.1371/journal.pone.0211598

Ran, Y., Patron, N., Kay, P., Wong, D., Buchanan, M., Cao, Y. Y., et al. (2018). Zinc finger nuclease-mediated precision genome editing of an endogenous gene in hexaploid bread wheat (*Triticum aestivum*) using a DNA repair template. *Plant Biotechnol. J.* 16 (12), 2088–2101. doi:10.1111/pbi.12941

Razzaq, A., Saleem, F., Kanwal, M., Mustafa, G., Yousaf, S., Imran Arshad, H. M., et al. (2019). Modern trends in plant genome editing: An inclusive review of the CRISPR/Cas9 toolbox. *Int. J. Mol. Sci.* 20 (16), 4045. doi:10.3390/ijms20164045

Rees, H. A., Komor, A. C., Yeh, W. H., Caetano-Lopes, J., Warman, M., Edge, A. S., et al. (2017). Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* 8 (1), 15790–15810. doi:10.1038/ncomms15790

Ren, B., Yan, F., Kuang, Y., Li, N., Zhang, D., Zhou, X., et al. (2018). Improved base editor for efficiently inducing genetic variations in rice with CRISPR/Cas9-guided hyperactive hAID mutant. *Mol. Plant* 11 (4), 623–626. doi:10.1016/j.molp.2018.01.005

Saurabh, S. (2021). Genome editing: Revolutionizing the crop improvement. *Plant Mol. Biol. Report.* 39, 1–21. doi:10.1007/s11105-021-01286-7

Segal, D. J., Beerli, R. R., Blancafort, P., Dreier, B., Effertz, K., Huber, A., et al. (2003). Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry* 42 (7), 2137–2148. doi:10.1021/bi026806o

Selma, S., Bernabé-Orts, J. M., Vazquez-Vilar, M., Diego-Martin, B., Ajenjo, M., Garcia-Carpintero, V., et al. (2019). Strong gene activation in plants with genome-wide specificity using a new orthogonal CRISPR/Cas9-based Programmable Transcriptional Activator. *Plant Biotechnol. J.* 17 (9), 1703–1705. doi:10.1111/pbi.13138

Shan, Q., Baltes, N. J., Atkins, P., Kirkland, E. R., Zhang, Y., Baller, J. A., et al. (2018). ZFN, TALEN and CRISPR-cas9 mediated homology directed gene insertion in arabidopsis: A disconnect between somatic and germinal cells. *J. Genet. genomics= Yi chuanxue bao* 45 (12), 681–684. doi:10.1016/j.jgg.2018.07.011

Shan, Q., Wang, Y., Chen, K., Liang, Z., Li, J., Zhang, Y., et al. (2013). Rapid and efficient gene modification in rice and Brachypodium using TALENs. *Mol. Plant* 6 (4), 1365–1368. doi:10.1093/mp/sss162

Shan, Q., Wang, Y., Li, J., and Gao, C. (2014). Genome editing in rice and wheat using the CRISPR/Cas system. *Nat. Protoc.* 9 (10), 2395–2410. doi:10.1038/nprot.2014.157

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* 31 (8), 686–688. doi:10.1038/nbt.2650

Shan, Q., Zhang, Y., Chen, K., Zhang, K., and Gao, C. (2015). Creation of fragrant rice by targeted knockout of the Os BADH 2 gene using TALEN technology. *Plant Biotechnol. J.* 13 (6), 791–800. doi:10.1111/pbi.12312

Shi, J., Gao, H., Wang, H., Lafitte, H. R., Archibald, R. L., Yang, M., et al. (2017). ARGOS 8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnol. J.* 15 (2), 207–216. doi:10.1111/pbi.12603

Shimatani, Z., Kashojiya, S., Takayama, M., Terada, R., Arazoe, T., Ishii, H., et al. (2017). Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* 35 (5), 441–443. doi:10.1038/nbt.3833

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., et al. (2017). Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* 15 (3), 169–182. doi:10.1038/nrmicro.2016.184

Shufen, C. H. A. O., Yicong, C. A. I., Baobing, F., Guiai, J. I. A. O., Zhonghua, S., Ju, L. U. O., et al. (2019). Editing of rice isoamylase gene ISA1 provides insights into its function in starch formation. *Rice Sci.* 26 (2), 77–87. doi:10.1016/j.rsci.2018.07.001

Shukla, V. K., Doyon, Y., Miller, J. C., DeKelver, R. C., Moehle, E. A., Worden, S. E., et al. (2009). Precise genome modification in the crop species Zea mays using zinc-finger nucleases. *Nature* 459 (7245), 437–441. doi:10.1038/nature07992

Son, S., and Park, S. R. (2022). Challenges facing CRISPR/Cas9-Based genome editing in *plants*. *Front. Plant Sci.* 13, 902413. doi:10.3389/fpls.2022.902413

Song, B., Yang, S., Hwang, G. H., Yu, J., and Bae, S. (2021). Analysis of NHEJ-based DNA repair after CRISPR-mediated DNA cleavage. *Int. J. Mol. Sci.* 22 (12), 6397. doi:10.3390/ijms22126397

Soyars, C. L., Peterson, B. A., Burr, C. A., and Nimchuk, Z. L. (2018). Cutting edge genetics: CRISPR/Cas9 editing of plant genomes. *Plant Cell Physiol.* 59 (8), 1608–1620. doi:10.1093/pcp/pcy079

Sun, Y., Jiao, G., Liu, Z., Zhang, X., Li, J., Guo, X., et al. (2017). Generation of high-amylose rice through CRISPR/Cas9-mediated targeted mutagenesis of starch branching enzymes. *Front. Plant Sci.* 8, 298. doi:10.3389/fpls.2017.00298

Svitashev, S., Young, J. K., Schwartz, C., Gao, H., Falco, S. C., and Cigan, A. M. (2015). Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and guide RNA. *Plant Physiol.* 169 (2), 931–945. doi:10.1104/pp.15.00793

Tan, J., Zhang, F., Karcher, D., and Bock, R. (2019). Author Correction: Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nat. Commun.* 10 (1), 2019–2110. doi:10.1038/s41467-019-10069-4

Tang, L., Mao, B., Li, Y., Lv, Q., Zhang, L., Chen, C., et al. (2017). Knockout of Osnramp5 using the crispr/cas9 system produces low Cd-accumulating indica rice without compromising yield. *Sci. Rep.* 7, 14438. doi:10.1038/s41598-017-14832-9

Tang, N., Cheng, C., Zhang, X., Qiao, M., Li, N., Mu, W., et al. (2020). TGF-β inhibition via CRISPR promotes the long-term efficacy of CAR T cells against solid tumors. *JCI insight* 5 (4), 1339777–e134249. doi:10.1172/jci.insight.133977

Tian, S., Jiang, L., Cui, X., Zhang, J., Guo, S., Li, M., et al. (2018). Engineering herbicide- resistant watermelon variety through CRISPR/Cas9-mediated base-editing. *Plant Cell Rep.* 37 (9), 1353–1356. doi:10.1007/s00299-018-2299-0

Tiwari, M., Kumar Trivedi, P., and Pandey, A. (2020). Emerging tools and paradigm shift of gene editing in cereals, fruits, and horticultural crops for enhancing nutritional value and food security. *Food Energy secur.* 10, e258. doi:10.1002/fes3.258

Tripathi, L., Ntui, V. O., and Tripathi, J. N. (2020). CRISPR/Cas9-based genome editing of banana for disease resistance. *Curr. Opin. Plant Biol.* 56, 118–126. doi:10.1016/j.pbi.2020.05.003

Upadhyay, S. K., Kumar, J., Alok, A., and Tuli, R. (2013). RNA-guided genome editing for target gene mutations in wheat. *G3* 3 (12), 2233–2238. doi:10.1534/g3.113.008847

Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., and Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* 11 (9), 636–646. doi:10.1038/nrg2842

Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M. P., Guyon-Debast, A., Chauvin, J. E., et al. (2019). Transgene-free genome editing in tomato and potato plants using Agrobacterium-mediated delivery of a CRISPR/Cas9 cytidine base editor. *Int. J. Mol. Sci.* 20 (2), 402. doi:10.3390/ijms20020402

Wang, J., Meng, X., Hu, X., Sun, T., Li, J., Wang, K., et al. (2019). xCas9 expands the scope of genome editing with reduced efficiency in rice. *Plant Biotechnol. J.* 17 (4), 709–711. doi:10.1111/pbi.13053

Wang, K. (2020). Fixation of hybrid vigor in rice: Synthetic apomixis generated by genome editing. *aBIOTECH* 1 (1), 15–20. doi:10.1007/s42994-019-00001-1

Wang, W., Pan, Q., He, F., Akhunova, A., Chao, S., Trick, H., et al. (2018). Transgenerational CRISPR-Cas9 activity facilitates multiplex gene editing in allopolyploid wheat. *CRISPR J.* 1 (1), 65–74. doi:10.1089/crispr.2017.0010

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* 32 (9), 947–951. doi:10.1038/nbt.2969

Watanabe, K., Breier, U., Hensel, G., Kumlehn, J., Schubert, I., and Reiss, B. (2016). Stable gene replacement in barley by targeted double-strand break induction. *J. Exp. Bot.* 67 (5), 1433–1445. doi:10.1093/jxb/erv537

Wei, C., Liu, J., Yu, Z., Zhang, B., Gao, G., and Jiao, R. (2013). TALEN or Cas9–rapid, efficientand specific choices for genome modifications. *J. Genet. genomics* 40 (6), 281–289. doi:10.1016/j.jgg.2013.03.013

Wendt, T., Holm, P. B., Starker, C. G., Christian, M., Voytas, D. F., Brinch-Pedersen, H., et al. (2013). TAL effector nucleases induce mutations at a pre-selected location in the genome of primary barley transformants. *Plant Mol. Biol.* 83 (3), 279–285. doi:10.1007/s11103-013-0078-4

Wolf, J., Gerber, A. P., and Keller, W. (2002). tadA, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. *EMBO J.* 21 (14), 3841–3851. doi:10.1093/emboj/cdf362

Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci. U. S. A.* 112 (11), 3570–3575. doi:10.1073/pnas.1420294112

Xie, K., and Yang, Y. (2013). RNA-guided genome editing in plants using a CRISPR–Cas system. *Mol. Plant* 6 (6), 1975–1983. doi:10.1093/mp/sst119

Xing, H. L., Dong, L., Wang, Z. P., Zhang, H. Y., Han, C. Y., Liu, B., et al. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* 14 (1), 1–12. doi:10.1186/s12870-014-0327-y

Xu, R., Li, J., Liu, X., Shan, T., Qin, R., and Wei, P. (2020). Development of plant prime- editing systems for precise genome editing. *Plant Commun.* 1 (3), 100043. doi:10.1016/j.xplc.2020.100043

Yan, F., Kuang, Y., Ren, B., Wang, J., Zhang, D., Lin, H., et al. (2018). Highly efficient A· T to G· C base editing by Cas9n-guided tRNA adenosine deaminase in rice. *Mol. Plant* 11 (4), 631–634. doi:10.1016/j.molp.2018.02.008

Yarra, R., and Sahoo, L. (2021). Base editing in rice: Current progress, advances, limitations, and future perspectives. *Plant Cell Rep.* 40, 595–604. doi:10.1007/s00299-020-02656-3

Yasumoto, S., Sawai, S., Lee, H. J., Mizutani, M., Saito, K., Umemoto, N., et al. (2020). Targeted genome editing in tetraploid potato through transient TALEN expression by Agrobacterium infection. *Plant Biotechnol.* 37, 205–211. doi:10.5511/plantbiotechnology.20.0525a

Yu, Z., Chen, H., Liu, J., Zhang, H., Yan, Y., Zhu, N., et al. (2014). Various applications of TALEN-and CRISPR/Cas9-mediated homologous recombination to modify the Drosophila genome. *Biol. Open* 3 (4), 271–280. doi:10.1242/bio.20147682

Yunyan, F. E. I., Jie, Y. A. N. G., Fangquan, W. A. N. G., Fangjun, F. A. N., Wenqi, L., Jun, W. A. N. G., et al. (2019). Production of two elite glutinous rice varieties by editing *wx* gene. *Rice Sci.* 26 (2), 118–124. doi:10.1016/j.rsci.2018.04.007

Zeng, Y., Wen, J., Zhao, W., Wang, Q., and Huang, W. (2020). Rational improvement of rice yield and cold tolerance by editing the three genes OsPIN5b, GS3, and OsMYB30 with the CRISPR–Cas9 system. *Front. Plant Sci.* 10, 1663. doi:10.3389/fpls.2019.01663

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of aclass 2 CRISPR-Cas system. *Cell* 163 (3), 759–771. doi:10.1016/j.cell.2015.09.038

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M., et al. (2017). Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* 35 (1), 31–34. doi:10.1038/nbt.3737

Zhang, A., Liu, Y., Wang, F., Li, T., Chen, Z., Kong, D., et al. (2019a). Enhanced rice salinity tolerance via CRISPR/Cas9-targeted mutagenesis of the OsRR22 gene. *Mol. Breed.* 39 (3), 47. doi:10.1007/s11032-019-0954-y

Zhang, H., Gou, F., Zhang, J., Liu, W., Li, Q., Mao, Y., et al. (2016). TALEN-mediated targeted mutagenesis produces a large variety of heritable mutations in rice. *Plant Biotechnol. J.* 14 (1), 186–194. doi:10.1111/pbi.12372

Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., et al. (2014). The CRISPR/C as9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol. J.* 12 (6), 797–807. doi:10.1111/pbi.12200

Zhang, J., Abadia, E., Refregier, G., Tafaj, S., Boschiroli, M. L., Guillard, B., et al. (2010). *Mycobacterium tuberculosis* complex CRISPR genotyping: Improving efficiency, throughput and discriminative power of 'spoligotyping'with new

spacers and a microbead-based hybridization assay. *J. Med. Microbiol.* 59 (3), 285–294. doi:10.1099/jmm.0.016949-0

Zhang, J., Zhang, H., Botella, J. R., and Zhu, J. K. (2018). Generation of new glutinous rice by CRISPR/Cas9-targeted mutagenesis of the Waxy gene in elite rice varieties. *J. Integr. Plant Biol.* 60 (5), 369–375. doi:10.1111/jipb.12620

Zhang, Y., Bai, Y., Wu, G., Zou, S., Chen, Y., Gao, C., et al. (2017). Simultaneous modification of three homoeologs of Ta EDR 1 by genome editing enhances powdery mildew resistance in wheat. *Plant J.* 91 (4), 714–724. doi:10.1111/tpj.13599

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., et al. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat. Commun.* 7 (1), 12617–12618. doi:10.1038/ncomms12617

Zhang, Y., Massel, K., Godwin, I. D., and Gao, C. (2018). Applications and potential of genome editing in crop improvement. *Genome Biol.* 19 (1), 210. doi:10.1186/s13059-018-1586-y

Zhang, Z., Lau, S. W., Zhang, L., and Ge, W. (2015). Disruption of zebrafish follicle-stimulating hormone receptor (fshr) but not luteinizing hormone receptor (lhcgr) gene by TALEN leads to failed follicle activation in females followed by sexual reversal to males. *Endocrinology* 156 (10), 3747–3762. doi:10.1210/en.2015-1039

Zhao, D., Li, J., Li, S., Xin, X., Hu, M., Price, M. A., et al. (2021). Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat. Biotechnol.* 39 (1), 35–40. doi:10.1038/s41587-020-0592-2

Zhu, J., Song, N., Sun, S., Yang, W., Zhao, H., Song, W., et al. (2016). Efficiency and inheritance of targeted mutagenesis in maize using CRISPR-Cas9. *Nat. Biotechnol.* 43 (1), 25–36. doi:10.1038/nbt.4261

Zong, Y., Song, Q., Li, C., Jin, S., Zhang, D., Wang, Y., et al. (2018). Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat. Biotechnol.* 36 (10), 950–953. doi:10.1038/nbt.4261

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., et al. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* 35 (5), 438–440. doi:10.1038/nbt.3811

| Frontiers in Genetics

TYPE Original Research
PUBLISHED 09 January 2023
DOI 10.3389/fgene.2022.1085332

Check for updates

OPEN ACCESS

EDITED BY
Ravi Valluru,
University of Lincoln, United Kingdom

REVIEWED BY
Hongliang Zhu,
China Agricultural University, China
Mohsen Hesami,
University of Guelph, Canada

*CORRESPONDENCE
Sanjeev Kumar,
sanjeevbunty@gmail.com

SPECIALTY SECTION
This article was submitted to Plant Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 31 October 2022
ACCEPTED 12 December 2022
PUBLISHED 09 January 2023

CITATION
Das J, Kumar S, Mishra DC, Chaturvedi KK, Paul RK and Kairi A (2023), Machine learning in the estimation of CRISPR-Cas9 cleavage sites for plant system.
Front. Genet. 13:1085332.
doi: 10.3389/fgene.2022.1085332

COPYRIGHT
© 2023 Das, Kumar, Mishra, Chaturvedi, Paul and Kairi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Machine learning in the estimation of CRISPR-Cas9 cleavage sites for plant system

Jutan Das[1], Sanjeev Kumar[2]*, Dwijesh Chandra Mishra[2], Krishna Kumar Chaturvedi[2], Ranjit Kumar Paul[2] and Amit Kairi[1]

[1]ICAR-Indian Agricultural Research Institute, New Delhi, India, [2]ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India

CRISPR-Cas9 system is one of the recent most used genome editing techniques. Despite having a high capacity to alter the precise target genes and genomic regions that the planned guide RNA (or sgRNA) complements, the off-target effect still exists. But there are already machine learning algorithms for people, animals, and a few plant species. In this paper, an effort has been made to create models based on three machine learning-based techniques [namely, artificial neural networks (ANN), support vector machines (SVM), and random forests (RF)] for the prediction of the CRISPR-Cas9 cleavage sites that will be cleaved by a particular sgRNA. The plant dataset was the sole source of inspiration for all of these machine learning-based algorithms. 70% of the on-target and off-target dataset of various plant species that was gathered was used to train the models. The remaining 30% of the data set was used to evaluate the model's performance using a variety of evaluation metrics, including specificity, sensitivity, accuracy, precision, F1 score, F2 score, and AUC. Based on the aforementioned machine learning techniques, eleven models in all were developed. Comparative analysis of these produced models suggests that the model based on the random forest technique performs better. The accuracy of the Random Forest model is 96.27%, while the AUC value was found to be 99.21%. The SVM-Linear, SVM-Polynomial, SVM-Gaussian, and SVM-Sigmoid models were trained, making a total of six ANN-based models (ANN1-Logistic, ANN1-Tanh, ANN1-ReLU, ANN2-Logistic, ANN2-Tanh, and ANN-ReLU) and Support Vector Machine models (SVM-Linear, SVM-Polynomial, SVM-Gaussian However, the overall performance of Random Forest is better among all other ML techniques. ANN1-ReLU and SVM-Linear model performance were shown to be better among Artificial Neural Network and Support Vector Machine-based models, respectively.

KEYWORDS

CRISPR, Cas9, SgRNA, genome editing, off-target, artificial neural network, support vector machine, random forest

1 Introduction

Genome editing (or gene editing) is nothing but the deletion, insertion and replacement of DNA at an explicit site in the genome of any organism. Molecular scissors, also known as designed nucleases, is used in the molecular laboratory to alter gene functions by editing or by modification of part of DNA (Urnov et al., 2010; Perez-Pinera et al., 2012).

Although there are many different gene editing methods (such as CRISPR-Cas9, ZFNs or TALENs etc.) available. Though techniques have been extensively used in a wide variety of cells, tissues and organisms (Sander and Joung 2014; Ma et al., 2017; Musunuru 2017) but CRISPR-Cas9 is the most widely used method by researchers worldwide.

CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats, is condensed segments of bacterial DNA that contain repetitive base sequences. It plays a critical role in providing natural immunity to bacteria against foreign DNA. With the event of identification of any viral DNA, the bacterium produces guide RNA, two strands of short RNA. Then, it forms a complex with an endonuclease enzyme, which is named Cas9 (CRISPR Associated Protein 9) (Barrangou et al., 2007; Terns and Terns 2011). The CRISPR-Cas9 complex targets and cuts out the viral DNA rendering the virus disabled. The Cas9 nuclease will not bind to the DNA if the target sequence is not followed by the Protospacer Adjacent Motif, or PAM, which helps the enzyme distinguish between the bacterial DNA and the viral DNA target. The CRISPR-Cas9 system then has the ability to store this viral data so that it will be able to recognize and eliminate future viral threats. CRISPRs are generally found in roughly 50% and 90% of sequenced genomes of bacteria and archaea, respectively (Sander and Joung 2014; Westra et al., 2014; Bortesi and Fischer 2015; Ma et al., 2017; Musunuru 2017).

The flexibility of the CRISPR-Cas9 system and its ability to find and modify particular genes can be used in research in the field of medicine, drug discovery and agriculture. The recent discovery of sequence-based genome editing technology for crop improvement (Georges and Ray, 2017). Particularly, CRISPR-Cas9 has shown the potential to address the emerging challenges of crop science and agriculture. This technology is capable of modifying any genomic sequence and can result in desired traits in organisms including crop species provided that the protospacer adjacent motif (PAM) sequence is available. CRISPR-Cas9 is an efficient, cost-effective, easier and highly precise genome editing tool as compared to other genome editing tools *viz.* zinc finger nucleases (ZFNs) and transcriptional activator-like effector nucleases (TALENs) (Wood et al., 2011). With the introduction and demonstration of the CRISPR-Cas9 system in 2012, it has been widely accepted among researchers across the globe. This genome editing system has been widely used and targeted many important genes of various cell lines and organisms, including bacteria, *C. elegans*, *Xenopus tropicalis*, yeast, zebrafish, *Drosophila*, rabbits, plants, monkeys, humans, rats and mice. Several workers have used this method and introduced single-point mutations, either deletions or insertions, into a target gene by using sgRNA (Sander and Joung 2014). Thus, CRISPR-Cas9 is one of the most emerging technology in the editing of plant genomes to cope up with emerging challenges of agriculture due to climate change and food security (Sovová et al., 2016; Haque et al., 2018).

Though sgRNA aims to target a specific segment of DNA, sometimes it is attached to other sites of DNA and unfortunately causes off-target mutations. Again it can tolerate mismatches in sgRNA-DNA at different positions but at the same time sensitive to the position, number and distribution of mismatches. Alter gene functions led by these off-target mutations can cause major genomic instability and pose a major threat while using CRISPR-Cas9 gene editing (Cho et al., 2014). It is imperative that altered but untargeted gene functions caused by off-targeted gene mutation lead to Genomic instability; it is one of the major problems associated with the application of CRISPR-Cas9 gene editing (Hsu et al., 2013; Cho et al., 2014). One way to safeguard from the ill effects of gene editing is to accurate prediction of off-target sites of the genome. Though, there are many off-target prediction methods available that works on the principles of calculation of scores based on the positions of the mismatches to the guide sequence (Haeussler et al., 2016; Xu et al., 2017). The score of each base pair in sgRNA-DNA is imitatively using the statistical analysis (Pearson correlation coefficient) of the mismatch effects based on prior gene editing experiments. Most of the current off-target prediction methods calculate scores, based on the positions of the mismatches to the guide sequence (Haeussler et al., 2016). The score of each base pair in sgRNA-DNA is imitatively using the statistical analysis of the mismatch effects based on prior gene editing experiments (Haeussler et al., 2016; Xu et al., 2017). For example, CCTop considers the distance of the mismatch from the PAM site when evaluating the specificity of candidate sgRNAs, "Optimized CRISPR Design" incorporates a position-specific mismatch penalty and additionally considers the spatial distribution of mismatches, and the CFD score penalizes each mismatch according to its specific substitution type and position (Zhang et al., 2014), MIT score only considered the positions and counts of the mismatched sites of sgRNA-DNA as the features to score the potential off-targets (Hsu et al., 2013) and the CFD score penalizes each mismatch according to its specific substitution type and position (Doench et al., 2016). Importantly, while these and other widely-used methods have been developed based on empirical data they mostly neglect the genomic context surrounding the target sequence and instead focus on predicting off-target effects for a given sgRNA using basic sequence features. It is significant that, even though these and other widely-used approaches were developed using empirical data, they primarily ignore the genomic context around the target sequence and instead concentrate on forecasting off-target effects for a given sgRNA using simple sequence properties (Sanjana et al., 2014). For accurately predicting cleavage sites, a variety of machine learning and deep learning method-based tools are available for humans (Abadi et al. 2017; Lin and Wong 2018) and plants (Hesami et al. 2021; Niu et al. 2021). These tools incorporate a wide range of features, including those that are specific to the genomic target, features that explain the sgRNA's thermodynamics, and features about the pair-wise

TABLE 1 Crop wise number of sgRNA, on-target and off-target.

Sl. No.	Crops name	No. of sgRNA	No. of on-target	No. of off-target	References
1	Rice	8	40	36	Li et al. (2016), Li et al. (2017b), Wang et al. (2017), Xu et al. (2014), Xu et al. (2015), Zhou et al. (2014)
2	Wheat	5	19	30	Shan et al. (2014), Zhang et al. (2016), Kim et al. (2018)
3	Soybean	2	10	5	Cai et al. (2015), Sun et al. (2015)
4	Cotton	8	20	33	Chen et al. (2017), Gao et al. (2017), Li et al. (2017a), Wang et al. (2018)
5	Cucumber	2	2	5	Chandrasekaran et al. (2016)
6	Tobacco	2	6	5	Gao et al. (2014)
7	Strawberry	2	5	7	Martín-Pizarro et al. (2019)
8	Watermelon	2	2	2	Tian et al. (2016)
9	Tomato	4	10	13	Brooks et al. (2014), Čermák et al. (2015), Pan et al. (2016)
10	Grape	2	8	10	Nakajima et al. (2017)
11	Potato	2	5	3	Butler et al. (2015), Wang et al. (2015), Andersson et al. (2017)
12	Apple	1	6	4	Malnoy et al. (2016)
13	Orange	4	15	20	Jia and Nian (2014)
14	Maize	6	31	23	Feng et al. (2016), Svitashev et al. (2016), Feng et al. (2018)
15	Barley	1	5	9	Kapusi et al. (2017)
Total		51	174	205	

TABLE 2 List of all the features used in this study.

Features derived from pair-wise sequence alignment	Features derived from nucleotide contents	Features derived from PAM sites
■ Pair-wise alignment score	■ 20th position nucleotide	■ PAM type
■ Wobble total	■ MGW (minor groove width) at the PAM NNGGN	■ In exon (non-NGG strand)
■ RNA bulges	■ DNA enthalpy—extended 223 nt	■ Downstream nt—position 1
■ Mismatches in positions 17–20	■ Nucleotide—position 2	■ Downstream nt—position 5
■ Mismatches	■ DHS (DNAse hypersensitive site) signal value	■ Downstream nt—position 2
■ DNA bulges	■ Guanine occupancy	■ In exon (NGG strand)
■ Tv (transversion mismatches) total	■ Distance from nucleosome	■ NGG strand expression
■ RR (purine-purine) total	■ Nucleotides—positions 4–5	■ Non-NGG strand expression
■ YY (pyrimidine-pyrimidine) total	■ Transcription region	■ PAM N nucleotide
	■ Coding region	
	■ GC content—extended	
	■ Nucleotide—position 4	

similarity between the sgRNA and the genomic target. To precisely determine the cleavage location of a gene, machine learning techniques are therefore quite advanced and effective.

Machine Learning is considered a subset of computational or Artificial Intelligence and provides the capacity for computers to learn from data without being explicitly programmed (Pedregosa, 2011). Compared to the other programming languages, it doesn't have explicit and defined steps or conditions to solve the problem. Rather, it enables to fit of the programs, algorithms, or methods to learn a specific task from

TABLE 3 Performance evaluation of developed six ANN models on plant dataset based on accuracy, precession, recall, FPR, specificity, F1 score, F2 score and AUC (figures in percentage).

Models	Accuracy	Precession	Recall	FPR	Specificity	F1 score	F2 score	AUC
ANN1-Logistic	91.65	89.00	96.49	14.05	85.95	92.59	94.89	97.26
ANN2-Logistic	86.87	81.78	97.44	25.57	74.43	88.93	93.85	90.54
ANN1-Tanh	90.33	88.29	94.68	14.80	85.20	91.37	93.33	93.18
ANN2-Tanh	87.66	84.23	94.98	20.95	79.05	89.28	92.61	95.34
ANN1-ReLU	90.65	90.44	92.49	11.52	88.48	91.45	92.07	96.94
ANN2-ReLU	77.39	73.50	91.00	38.65	61.35	81.32	86.86	91.20

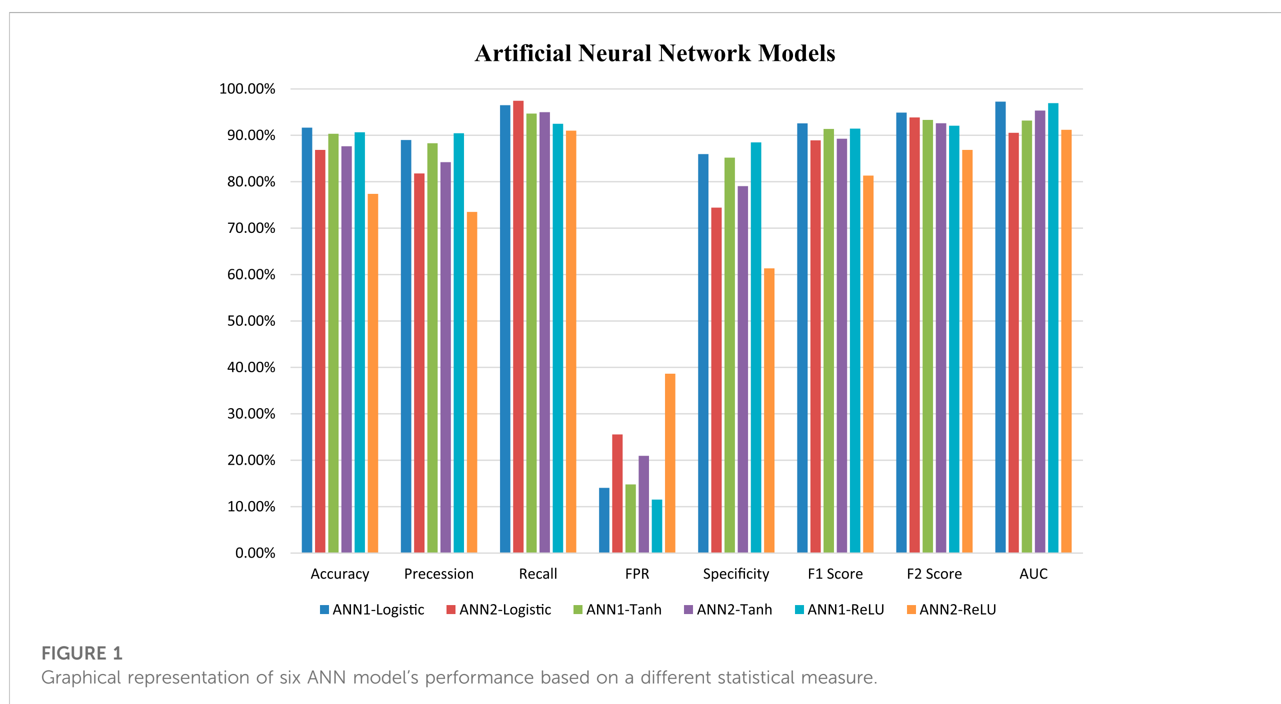


FIGURE 1
Graphical representation of six ANN model's performance based on a different statistical measure.

the experimental data set (Mitchell et al., 2003). These trained models help the machine to take decisions on different and variable situations based on the learning upon a dataset. Machine learning has been widely used in different fields of plant science such as plant breeding (van Dijk et al., 2021), *in vitro* culture (Hesami and Jones 2020), stress phenotyping (Singh et al., 2016), stress physiology (Jafari and Shahsavar 2020), plant system biology (Hesami et al., 2022), plant identification (Grinblat et al., 2016), and pathogen identification (Mishra et al., 2019).

The currently available machine learning- or deep learning-based algorithms for CRISPR off-target prediction are mostly based on data either from animal or human genomes. Their effectiveness on plant genomes has not been widely demonstrated. As a result, we used plant data to create machine learning-based models for plant genomes. The development of machine learning-based models for the prediction of CRISPR-cas9 off-target sites for plant genomes and for assessing the effectiveness of these models were the key contributions made in this study.

2 Materials and methods

2.1 Data collection

A thorough literature review has been conducted to gather information on off-target and on-target sequences, as well as associated sgRNA sequences, specific to crop species. We used Google Scholar as a search engine to look up published and accessible literature using terms like "off-target sites in crops," "off-target estimation," "CRISPR-Cas9 on-target and off-target sites," "off-target effect minimization in a plant cell," etc. Search

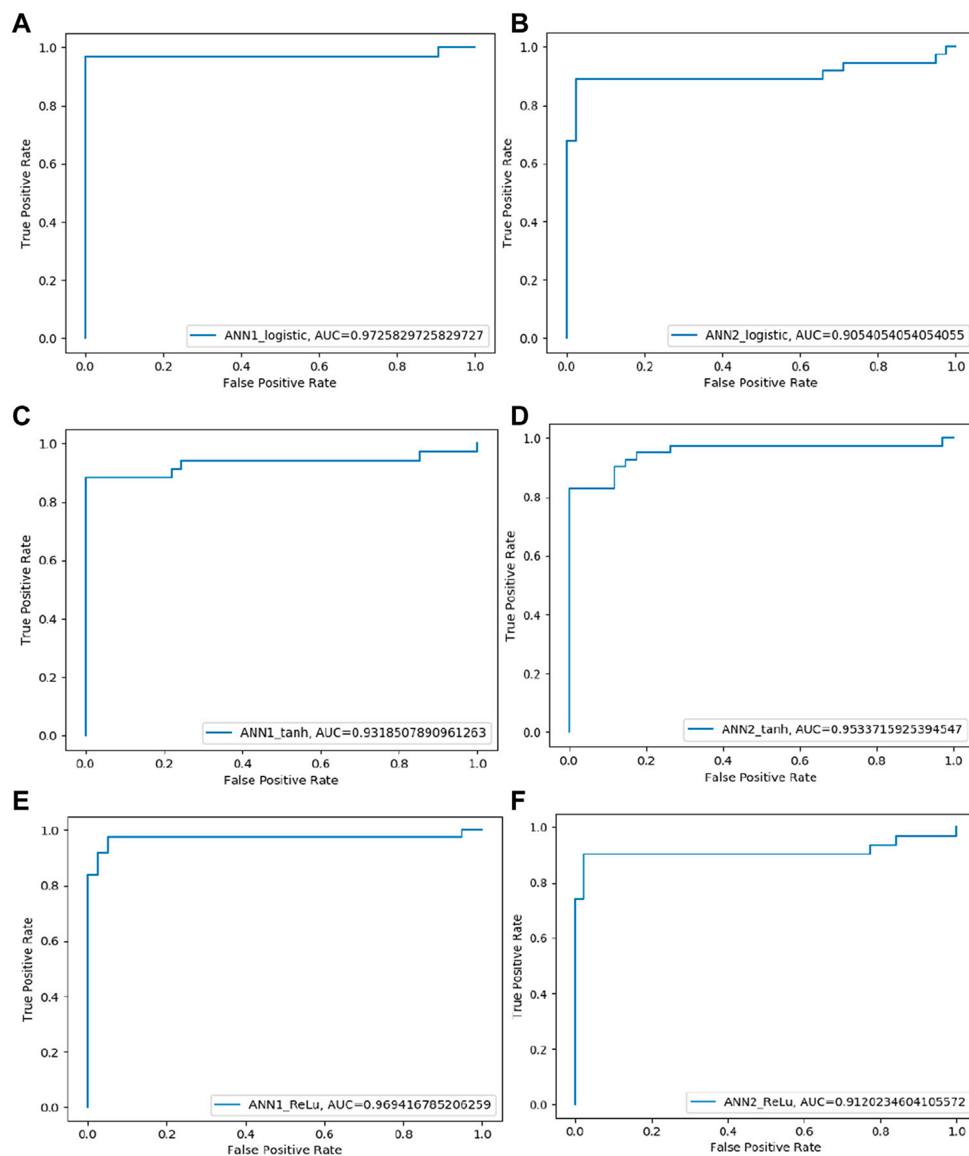


FIGURE 2
ROC curve of four SVM models performance based on AUC score: **(A)** SVM-Linear model ROC curve **(B)** SVM-Polynomial model ROC curve **(C)** SVM-rbf model ROC curve **(D)** SVM-Sigmoid model ROC curve.

results from the last 5 years were used to choose a few research papers that describe CRISPR-Cas9 experiments conducted on various crop species. Then, wherever it was available, the genomic sequence of the sgRNA, the on-targets, and the off-target sites were collected from the shortlisted articles.

2.2 Data preparation

A computer program was created in the Python programming language to extract the parameters from a large number of sequences based on the pairwise alignment of sgRNA and genome target sites, features regarding the nucleotide contents of 20 nucleotide sites and their contiguous genomic regions, and features regarding the PAM sites and nearby the nucleotides (Abadi et al., 2017). Then, using the constructed program, features based on the aforementioned criteria were extracted from the sequences of sgRNA, on-targets, and off-target sites of the genome. The creation of classification models based on machine learning uses these extracted characteristics as explanatory variables. The related site-specific on-target and off-target information were used to create a response variable, where respective on-targets were labeled as 1 and off-targets as 0.

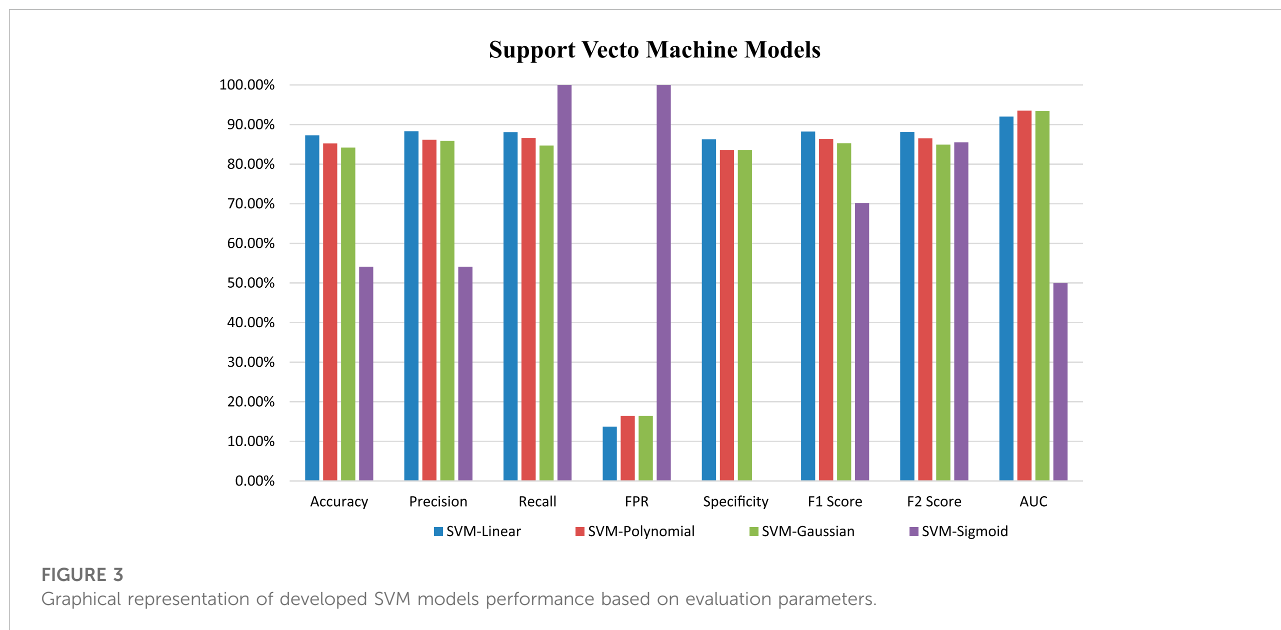


FIGURE 3
Graphical representation of developed SVM models performance based on evaluation parameters.

TABLE 4 Comparison of developed SVM models performance based on evaluation parameters (figures in percentage).

Models	Accuracy	Precision	Recall	FPR	Specificity	F1 score	F2 score	AUC
SVM-Linear	87.26	88.31	88.10	13.74	86.26	88.21	88.14	92.00
SVM-Polynomial	85.22	86.15	86.61	16.41	83.59	86.38	86.52	93.51
SVM-Gaussian	84.18	85.88	84.68	16.41	83.59	85.28	84.92	93.45
SVM-Sigmoid	54.09	54.09	100	100	0.00	70.21	85.49	50.00

2.3 Machine learning models/experiments

Recent machine learning-based classification modeling techniques have been employed to create robust classification models. The Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest machine learning techniques have all been investigated. In the process of creating a viable model, a variety of different variations and structures connected to the above modeling paradigm have also been tested.

2.3.1 Artificial neural network

Two Multi-layered perceptrons (MLP) structures (Chatterjee et al., 2022) of ANN were chosen with three and four hidden layers and named ANN1 and ANN2 respectively. The layer-wise number of perceptron was arbitrarily taken as 25:25:25 for each layer in ANN1 whereas 30:20:10:5 for each of the four layers in ANN2, starting from the input layer to the output layer. To train the above ANN models, three different activation functions have been considered here. They are Logistic, Tanh and ReLu, thus altogether six ANNs, namely ANN1-Logistics, ANN1-Tanh, ANN1-ReLu, ANN2-Logistics, ANN2-Tanh, and ANN2-ReLu, were used to model the training data. The MLPClassifier implemented in the python Scikit-learn module (Pedregosa, 2011) was used for the training of the ANN models using training data set. The following hyper-parameters of MLPClassifier were used during the training. To validate the model 5-fold cross-validation techniques (Refaeilzadeh et al., 2009) were used. The following parameters were used for developing the model.

hidden_layer_sizes: 25:25:25 and 30:20:10:5.

Activation: Used three activation functions i.e., logistic, tanh and relu.

Solver: Adam solver was used for optimizing the weights.

learning_rate_init: Used initial learning rate as 0.001.

2.3.2 Support vector machine

Depending on the type of kernel function used, four SVM models are developed, which are named as SVM-Linear, SVM-Polynomial, SVM-Gaussian and SVM-Sigmoid. The Support Vector Classifier (SVC) implemented in the python Scikit-learn module (Pedregosa, 2011) was used for training the SVM models using training data set. Polynomial kernel-based SVM model used

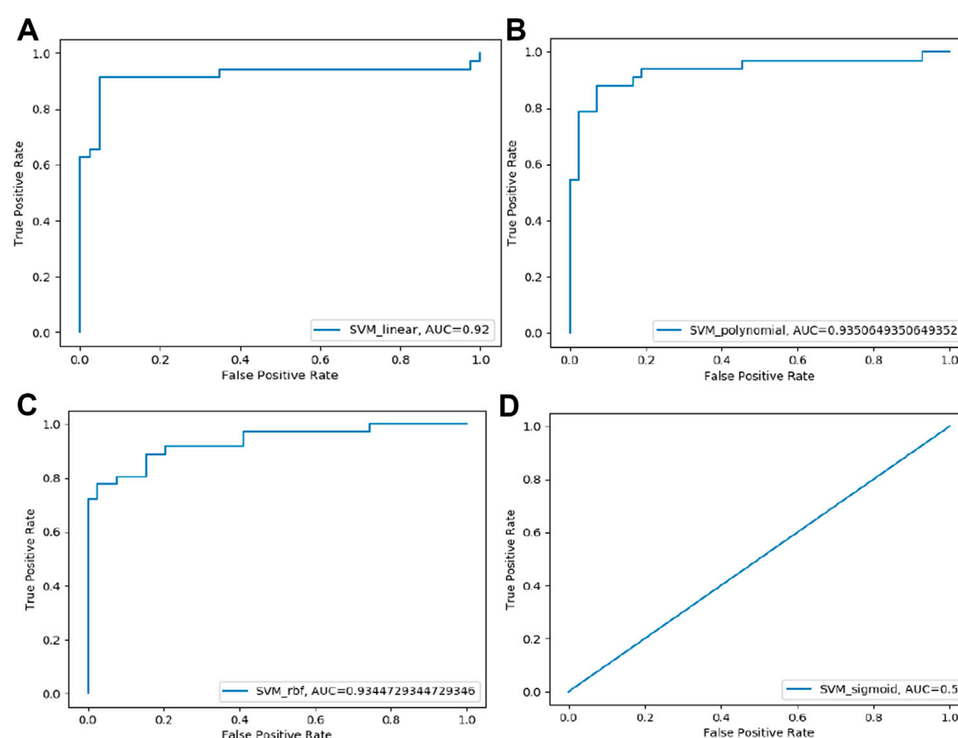


FIGURE 4
ROC curve of four SVM models performance based on AUC score: **(A)** SVM-Linear model ROC curve **(B)** SVM-Polynomial model ROC curve **(C)** SVM-rbf model ROC curve **(D)** SVM-Sigmoid model ROC curve.

TABLE 5 Different performance parameters of the random forest model (figures in percentage).

Parameters	RF model
Accuracy	96.27
Precision	94.75
Recall	98.56
FPR	6.44
Specificity	93.56
F1 score	96.62
F2 score	97.77
AUC	99.21

with the degree of 3. Here also, 5-fold cross-validation techniques were used for validation of the model.

2.3.3 Random forest

For the training dataset and a given set of features, we implemented the RF model using RandomForestClassifier implemented in the python scikit-learn module (Pedregosa, 2011). We used 5-fold cross-validation techniques for the validation of our model. In this study, we used it for 100 estimators and used the Gini index to measure the quality of a split and for building the trees, a bootstrap method was used.

4 Results

4.1 Crop species wise sgRNA related on-target and off-target sequence data

The research article containing sequence details about sgRNA related to on-target and off-target data was discovered in the following journals: Frontiers in plant science, Nature, Scientific Reports, Plant Biotechnology Journal, PloS one, Molecular plant pathology, Journal of Genetics and Genomics, Nucleic acids research, Nature communications, Rice, Molecular plant, Plant cell reports, Nature Biotechnology, Nature protocols, Cell, Journal of molecular biology, Journal of Molecular Biology, and Journal of Molecular Medicine. From a Google Scholar search using the stated keyword, a total of 64 research publications were found. Thirty two research publications detail the sequence of a crop-specific sgRNA and its

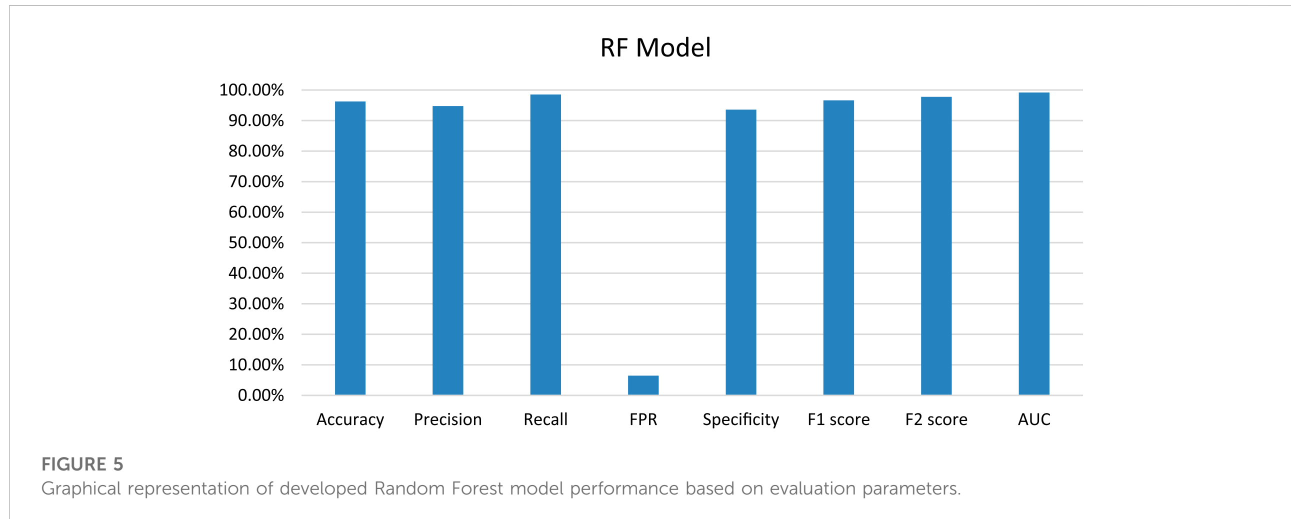


FIGURE 5
Graphical representation of developed Random Forest model performance based on evaluation parameters.

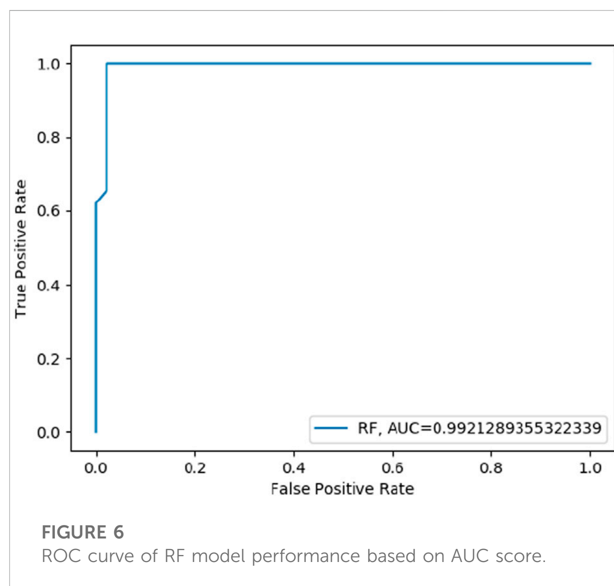


FIGURE 6
ROC curve of RF model performance based on AUC score.

associated on-target and off-target regions (Table 1). Table 1 perusal reveals that 15 important crop species were used to gather the sequencing data for 51 sgRNA and the related 174 and 205 on-targets and off-targets sites, respectively. SgRNA was between 17 and 20 nucleotides in length. The average length of the off-target and on-target was determined to be 23 or longer nucleotides. As a result, 51 sgRNA were the subject of a total of 379 data points collection, together with the matching on-target and off-target locations. The values of 48 explanatory variables and one response variable were obtained using a Python program. The complete 379-point data set, which contained 48 variables, was divided into two halves by chance, with 265 and 114 entries for each of the 31 variables being used for model building and model evaluation, respectively.

4.2 Importance of features

In addition to improving prediction abilities, the learning strategy enabled researchers to thoroughly comprehend the most crucial Cas9 traits. When the entire set of features was evaluated, three clusters emerged among the top 30 features: 1) a pair-wise similarity characteristic between the sgRNA and the DNA site Along with the pair-wise alignment score, the number of mismatches, the number of RNA/DNA bulges, and the kind of mismatch (transversion, transversion, or wobble) were all included in this cluster. 2) GC content, DNA enthalpy (Breslauer et al., 1986), and several measures of DNA spatial structure, such as minor groove width and bending stiffness (Zhou et al., 2013), were among these. These characteristics are related to the nucleotide composition of the 20-nt location and the genomic region surrounding it. 3) The DNA geometry scores calculated in and around this region, as well as PAM site characteristics like the PAM type (NGG or NAG), were among them. In this investigation, 30 feature numbers (Table 2) were used (Abadi et al., 2017).

4.3 Model results

Results from different machine learning-based developed models and their comparisons in various aspects are given below-

4.3.1 Artificial neural network model

Six ANN models were trained using training datasets, and the effectiveness of each trained model was assessed using a variety of evaluation parameters on the test dataset. In this research, we used k-Fold, or $k = 5$ cross-validation approaches, for validation purposes. All six evaluation parameter values have been shown against the implemented ANN models (Table 3), which makes it

TABLE 6 Comparison of all eleven developed models performance based on evaluation parameters (figures in percentage).

	ANN1_logistic	ANN2_logistic	ANN1_tanh	ANN2_tanh	ANN1_ReLu	ANN2_ReLu	SVM_linear	SVM_polynomial	SVM_rbf	SVM_sigmoid	Random forest
Accuracy	91.65	86.87	90.33	87.66	90.65	77.39	87.26	85.22	84.18	54.09	96.27
Precision	89.00	81.78	88.29	84.23	90.44	73.50	88.31	86.15	85.88	54.09	94.75
Recall	96.49	97.44	94.68	94.98	92.49	91.00	88.10	86.61	84.68	100.00	98.56
FPR	14.05	25.57	14.80	20.95	11.52	38.65	13.74	16.41	16.41	100.00	6.44
Specificity	85.95	74.43	85.20	79.05	88.48	61.35	86.26	83.59	83.59	0.00	93.56
F1 Score	92.59	88.93	91.37	89.28	91.45	81.32	88.21	86.38	85.28	70.21	96.62
F2 Score	94.89	93.85	93.33	92.61	92.07	86.86	88.14	86.52	84.92	85.49	97.77
AUC	97.26	90.54	93.18	95.34	96.94	91.20	92.00	93.51	93.45	50.00	99.21

simple to grasp similar comparisons in a graphical style (Figures 1, 2). This study is based on the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) (Hwang and Yoon 1981). From six ANN-based models, ANN1-Logistic obtained a TOPSIS analysis score of 0.85134 and is ranked one (Supplementary Material).

4.3.2 Support vector machine model

Four SVM models were trained on training datasets and the performance of each of the trained models was evaluated on test data set by using several evaluation parameters. For validation purposes in this paper, we used k-Fold i.e., $k = 5$ cross-validation techniques. In this paper, we compare all four developed SVM models with each other based on their evaluation parameters (Table 4). The SVM_ Linear model gives better accuracy (87.26%) and precision (88.31%) in comparison to other models. So, among all the developed SVM models the SVM-Linear model performs very well compared to the other three models (SVM-Polynomial, SVM-Gaussian, SVM-Sigmoid). The values of all six evaluation parameters have been plotted against the undertaken SVM models (Figures 3, 4) which show similar comparisons in a graphical format for easy understanding. This study fits under the TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) framework (Hwang and Yoon 1981). According to TOPSIS analysis result SVM.Linear model got 1st rank (Supplementary Material).

4.3.3 Random forest model

Random Forest (RF) model was trained on training datasets and the performance of the trained model was evaluated on test data set by using several evaluation parameters. For validation purposes in this paper, we used k-Fold i.e., $k = 5$ cross-validation techniques. The value of RF model accuracy is 96.27% and its AUC value is 99.21% (Table 5). The graphical representation of the model performance is also shown in Figure 5. Here, the RF model gives a very low score of false positive rate (FPR), which is good for any model. From the ROC curve of the RF model (Figure 6) it can be concluded that the developed model is performing very well in the plant data set.

4.4 Comparison among developed machine learning-based models

In this study, we developed three machine learning-based models for the estimation of off-target sites. The performance of these techniques is being evaluated by different statistical measures viz. sensitivity/recall, specificity, accuracy, precision, FPR, F1 score, F2 score and AUC (Table 6). Random Forest (RF) model achieves the best accuracy which is 96.27% compare to other models. RF model achieves the highest specificity value as compared to the other ten models which are 93.56%. According to the AUC score RF model cover, the maximum area under the ROC curve is 99.21% compare to the other seven models

FIGURE 7
Graphical representation of the performance of all models for off-target prediction developed using three groups of machine learning techniques.

TABLE 7 TOPSIS analysis result for all eleven machine learning-developed models on the plant dataset.

Models	Score	Rank
RF	0.961423	1
ANN1_logistic	0.789202	2
ANN1_ReLu	0.69774	3
ANN1_tanh	0.602147	4
ANN2_tanh	0.516367	5
SVM_linear	0.482942	6
ANN2_logistic	0.447746	7
SVM_polynomial	0.386691	8
SVM_rbf	0.327762	9
SVM_sigmoid	0.276047	10
ANN2_ReLu	0.166041	11

(Figure 7). From the above comparisons, Random Forest (RF) model perform comparatively better than the other ten models on plant datasets and got 1st rank according to TOPSIS analysis (Table 7).

5 Discussion

In CRISPR-Cas9 gene editing techniques, sgRNA seeks to target a specific DNA segment but occasionally it may attach to untargeted DNA locations, which regrettably results in off-target mutations. Altering gene functionalities brought on by these off-target mutations might generate significant genomic instability and constitute a serious concern. This off-target induced resultant genomic instability causes a major limitation in the use of the CRISPR-Cas9 gene editing technique. Therefore, it is essential to accurately predict the off-target site related to sgRNA.

In this study, we used a machine learning approach for predicting off-targets in the CRISPR-Cas9 gene editing technique. Here, a total of eleven machine learning-based models for CRISPR-Cas9 cleavage site prediction (6-ANN, 4-SVM, and 1-RF models) were constructed. Models were trained by using three main types of features pair-wise alignment features from the sgRNA-DNA sequence, features related to nucleotide composition and PAM site characteristics. Training of the various machine learning algorithms viz. ANN, SVM and RF were carried out with different combinations of layer counts, kernel types and tree counts respectively. The performance of the training models was evaluated based on selected statistics within and between groups of developed models.

In the case of Artificial Neural Networks (ANN), six models were developed using different activation functions, different hidden layers and different neuron numbers. Following activation functions Logistic, Tanh and ReLU has been applied with varying number of layers and neurons resulting in six ANN-based models named as ANN1-Logistic, ANN2-Logistic, ANN1-Tanh, ANN2-Tanh, ANN1-ReLU and ANN2-ReLU. These models were trained on plant data sets and performance was evaluated under 5-fold cross-validation. Although, they have achieved more or less similar

performance but according to TOPSIS analysis ANN1-Logistic model predicts off-target sites accurately as compared to the other five models *i.e.* ANN2-Logistic, ANN1-Tanh, ANN2-Tanh, ANN1-ReLU and ANN2-ReLU.

In the instance of a Support Vector Machine (SVM), a total of four models were developed using different kernel functions. The following kernel functions: Linear, Polynomial, Gaussian and Sigmoid have been used during the model training. Which results in four SVM-based models named as SVM-Linear, SVM-Polynomial, SVM-Gaussian and SVM-Sigmoid. These models were trained using data sets related to plants. The relative performance among the SVM-based models was evaluated using 5-fold cross-validation and TOPSIS analysis. Based on these evaluations, the SVM-Linear model predicts off-target sites more accurately than the other three SVM models i.e., SVM-Polynomial, SVM-Gaussian, and SVM-Sigmoid.

Random Forest (RF) experiments were carried out with different numbers of tree sizes and a model with optimal tree size was selected for further comparison with other developed models.

We evaluated the relative accuracy of the three groups of developed machine learning-based models for off-target prediction, using the 5-fold cross-validation method and TOPSIS analysis. The accuracy of the RF model was 96.27%, and its area under the ROC curve (AUC) was 99.21%, which is higher than that of the ANN and SVM models. Further, based on the TOPSIS analysis, the Random Forest model was scored highest among the group. This indicates the better performance of the Random Forest model over SVM and ANN-based models for the prediction of cleavage sites in the CRISPR-Cas9 gene editing technique for plant systems.

In the future, the latest machine-learning techniques such as deep learning etc. may be investigated further to enhance of the modeling process. It is generally accepted that such cutting-edge, computationally intelligent strategies will make future predictions of CRISPR-Cas9 off-target sites even more accurate.

6 Conclusion

Gene editing, commonly known as molecular scissors, is the process of insertion, deletion or replacement of DNA on a particular position in the genome of any organism. We demonstrated that the off-targets of CRISPR-Cas9 gene editing can be reliably predicted by machine learning approaches. In comparison to the other two conventional machine learning methods, ANN and SVM; our final Random Forest (RF) model better them all in terms of performance on the plant dataset. We think that these intelligent methods can make a significant contribution to CRISPR-Cas9 off-target predictions or other problems of a similar nature.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: Dataset will be made available on request. Requests to access these datasets should be directed to jutandas2468@gmail.com.

Author contributions

All authors contributed to the design of the approach. JD, SK, and AK performed the experiments. JD, DM, RP, and KC conceived the study and drafted the manuscript. All authors have read and approved the manuscript for publication.

Acknowledgments

ICAR-Indian Agricultural Statistics Research Institute, Centre for Agricultural Bioinformatics (CABin), New Delhi.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1085332/full#supplementary-material

References

Abadi, S., Yan, W. X., Amar, D., and Mayrose, I. (2017). A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action. *PLOS Comput. Biol.* 13 (10), e1005807. doi:10.1371/journal.pcbi.1005807

Andersson, M., Turesson, H., Nicolia, A., Fält, A. S., Samuelsson, M., and Hofvander, P. (2017). Efficient targeted multiallelic mutagenesis in tetraploid potato (Solanum tuberosum) by transient CRISPR-Cas9 expression in protoplasts. *Plant Cell. Rep.* 36 (1), 117–128. doi:10.1007/s00299-016-2062-3

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315 (5819), 1709–1712. doi:10.1126/science.1138140

Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* 33 (1), 41–52. doi:10.1016/j.biotechadv.2014.12.006

Breslauer, K. J., Frank, R., Blocker, H., and Marky, L. A. (1986). Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci.* 83 (11), 3746–3750. doi:10.1073/pnas.83.11.3746

Brooks, C., Nekrasov, V., Lipppman, Z. B., and van Eck, J. (2014). Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiol.* 166 (3), 1292–1297. doi:10.1104/pp.114.247577

Butler, N. M., Atkins, P. A., Voytas, D. F., and Douches, D. S. (2015). Generation and inheritance of targeted mutations in potato (Solanum tuberosum L.) using the CRISPR/cas system. *PLOS ONE* 10 (12), e0144591. doi:10.1371/journal.pone.0144591

Cai, Y., Chen, L., Liu, X., Sun, S., Wu, C., Jiang, B., et al. (2015). CRISPR/Cas9-Mediated genome editing in soybean hairy roots. *PLOS ONE* 10 (8), e0136064. doi:10.1371/journal.pone.0136064

Čermák, T., Baltes, N. J., Čegan, R., Zhang, Y., and Voytas, D. F. (2015). High-frequency, precise modification of the tomato genome. *Genome Biol.* 16 (1), 232–315. doi:10.1186/s13059-015-0796-9

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., et al. (2016). Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* 17 (7), 1140–1153. doi:10.1111/mpp.12375

Chatterjee, A., Saha, J., and Mukherjee, J. (2022). Clustering with multi-layered perceptron. *Pattern Recognit. Lett.* 155, 92–99. doi:10.1016/J.PATREC.2022.02.009

Chen, X., Lu, X., Shu, N., Wang, S., Wang, J., Wang, D., et al. (2017). Targeted mutagenesis in cotton (Gossypium hirsutum L.) using the CRISPR/Cas9 system. *Sci. Rep.* 7 (1), 44304–44307. doi:10.1038/srep44304

Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., et al. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 24 (1), 132–141. doi:10.1101/gr.162339.113

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34 (2), 184–191. doi:10.1038/nbt.3437

Feng, C., Su, H., Bai, H., Wang, R., Liu, Y., Guo, X., et al. (2018). High-efficiency genome editing using a dmc1 promoter-controlled CRISPR/Cas9 system in maize. *Plant Biotechnol. J.* 16 (11), 1848–1857.

Feng, C., Yuan, J., Wang, R., Liu, Y., Birchler, J. A., and Han, F. (2016). Efficient targeted genome modification in maize using CRISPR/Cas9 system. *J. Genet. Genomics* 43 (1), 37–43. doi:10.1016/J.JGG.2015.10.002

Gao, J., Wang, G., Ma, S., Xie, X., Wu, X., Zhang, X., et al. (2014). CRISPR/Cas9-mediated targeted mutagenesis in Nicotiana tabacum. *Plant Mol. Biol.* 87 (1), 99–110. doi:10.1007/s11103-014-0263-0

Gao, W., Long, L., Tian, X., Xu, F., Liu, J., Singh, P. K., et al. (2017). Genome editing in cotton with the CRISPR/Cas9 system. *Front. Plant Sci.* 8, 1364. doi:10.3389/fpls.2017.01364

Georges, F., and Ray, H. (2017). Genome editing of crops: A renewed opportunity for food security. *Taylor Francis* 8 (1), 1–12. doi:10.1080/21645698.2016.1270489

Grinblat, G. L., Uzal, L. C., Larese, M. G., and Granitto, P. M. (2016). Deep learning for plant identification using vein morphological patterns. *Comput. Electron. Agric.* 127, 418–424. doi:10.1016/J.COMPAG.2016.07.003

Haeussler, M., Schonig, K., Eckert, H., Eschstruth, A., Mianne, J., Renaud, J. B., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* 17 (1), 148–212. doi:10.1186/s13059-016-1012-2

Haque, E., Taniguchi, H., Hassan, M. M., Bhowmik, P., Karim, M. R., Śmiech, M., et al. (2018). Application of CRISPR/Cas9 genome editing technology for the improvement of crops cultivated in tropical climates: Recent progress, prospects, and challenges. *Front. Plant Sci.* 9, 617. doi:10.3389/fpls.2018.00617

Hesami, M., Alizadeh, M., Jones, A. M. P., and Torkamaneh, D. (2022). Machine learning: Its challenges and opportunities in plant system biology. *Appl. Microbiol. Biotechnol.* 106 (9), 3507–3530. doi:10.1007/s00253-022-11963-6

Hesami, M., and Jones, A. M. P. (2020). Application of artificial intelligence models and optimization algorithms in plant cell and tissue culture. *Appl. Microbiol. Biotechnol.* 104 (22), 9449–9485. doi:10.1007/s00253-020-10888-2

Hesami, M., Yoosefzadeh Najafabadi, M., Adamek, K., Torkamaneh, D., and Jones, A. M. P. (2021). Synergizing off-target predictions for in silico insights of CENH3 knockout in cannabis through CRISPR/cas. *Molecules* 26 (7), 2053. doi:10.3390/molecules26072053

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31 (9), 827–832. doi:10.1038/nbt.2647

Hwang, C.-L., and Yoon, K. (1981). "Methods for multiple attribute decision making," in *Multiple attribute decision making. Lecture notes in economics and mathematical systems* (Berlin, Heidelberg: Springer), Vol. 186, 58–191. doi:10.1007/978-3-642-48318-9_3

Jafari, M., and Shahsavar, A. (2020). The application of artificial neural networks in modeling and predicting the effects of melatonin on morphological responses of citrus to drought stress. *PLOS ONE* 15 (10), e0240427. doi:10.1371/journal.pone.0240427

Jia, H., and Nian, W. (2014). Targeted genome editing of sweet orange using cas9/sgRNA. *PLOS ONE* 9 (4), e93806. doi:10.1371/journal.pone.0093806

Kapusi, E., Corcuera-Gómez, M., Melnik, S., and Stoger, E. (2017). Heritable genomic fragment deletions and small indels in the putative ENGase gene induced by CRISPR/Cas9 in barley. *Front. Plant Sci.* 8, 540. doi:10.3389/fpls.2017.00540

Kim, D., Alptekin, B., and Budak, H. (2018). CRISPR/Cas9 genome editing in wheat. *Funct. Integr. Genomics* 18 (1), 31–41. doi:10.1007/s10142-017-0572-x

Li, C., Unver, T., and Zhang, B. (2017a2017). A high-efficiency CRISPR/Cas9 system for targeted mutagenesis in Cotton (Gossypium hirsutum L.). *Sci. Rep.* 7 (1), 43902–43910. doi:10.1038/srep43902

Li, J., Meng, X., Zong, Y., Chen, K., Zhang, H., Liu, J., et al. (2016). Gene replacements and insertions in rice by intron targeting using CRISPR–Cas9. *Nat. Plants* 2 (10), 16139–16145. doi:10.1038/nplants.2016.139

Li, J., Sun, Y., Du, J., Zhao, Y., and Xia, L. (2017b). Generation of targeted point mutations in rice by a modified CRISPR/Cas9 system. *Mol. Plant* 10 (3), 526–529. doi:10.1016/j.molp.2016.12.001

Lin, J., and Wong, K. C. (2018). Off-target predictions in CRISPR-Cas9 gene editing using deep learning. *Bioinformatics* 34 (17), i656–i663. doi:10.1093/bioinformatics/bty554

Ma, H., Marti-Gutierrez, N., Park, S. W., Wu, J., Lee, Y., Suzuki, K., et al. (2017). Correction of a pathogenic gene mutation in human embryos. *Nature* 548 (7668), 413–419. doi:10.1038/nature23305

Malnoy, M., Viola, R., Jung, M. H., Koo, O. J., Kim, S., Kim, J. S., et al. (2016). DNA-free genetically edited grapevine and apple protoplast using CRISPR/Cas9 ribonucleoproteins. *Front. Plant Sci.* 7 (12), 1904. doi:10.3389/fpls.2016.01904

Martín-Pizarro, C., Triviño, J. C., and Posé, D. (2019). Functional analysis of the TM6 MADS-box gene in the octoploid strawberry by CRISPR/Cas9-directed mutagenesis. *J. Exp. Bot.* 70 (3), 885–895. doi:10.1093/jxb/ery400

Mishra, B., Kumar, N., and Mukhtar, M. S. (2019). Systems biology and machine learning in plant–pathogen interactions. *Mol. Plant-Microbe Interact.* 32 (1), 45–55. doi:10.1094/MPMI-08-18-0221-FI

Mitchell, T., Buchanan, B., Dejong, G., Dietterich, T., Rosenbloom, P., and Waibel, A. (2003). Machine learning. *Annu. Rev. Comput. Sci.* 4 (1), 417–433. doi:10.1146/annurev.cs.04.060190.002221

Musunuru, K. (2017). Genome editing: The recent history and perspective in cardiovascular diseases. *J. Am. Coll. Cardiol.* 70 (22), 2808–2821. doi:10.1016/J.JACC.2017.10.002

Nakajima, I., Ban, Y., Azuma, A., Onoue, N., Moriguchi, T., Yamamoto, T., et al. (2017). CRISPR/Cas9-mediated targeted mutagenesis in grape. *PLOS ONE* 12 (5), e0177966. doi:10.1371/journal.pone.0177966

Niu, M., Lin, Y., and Zou, Q. (2021). sgRNACNN: identifying sgRNA on-target activity in four crops using ensembles of convolutional neural networks. *Plant Mol. Biol.* 105 (4), 483–495. doi:10.1007/s11103-020-01102-y

Pan, C., Ye, L., Qin, L., Liu, X., He, Y., Wang, J., et al. (2016). CRISPR/Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations. *Sci. Rep.* 6 (1), 24765–24769. doi:10.1038/srep24765

Pedregosa, F. (2011). Scikit-learn: Machine learning in Python gaël varoquaux bertrand thirion vincent dubourg alexandre passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu perrot. *J. Mach. Learn. Res.* 12, 2825–2830. Available at: http://scikit-learn.sourceforge.net (Accessed November 6, 2022).

Perez-Pinera, P., Ousterout, D. G., and Gersbach, C. A. (2012). Advances in targeted genome editing. *Curr. Opin. Chem. Biol.* 16 (3–4), 268–277. doi:10.1016/j.cbpa.2012.06.007

Refaeilzadeh, P., Tang, L., and Liu, H. (2009). Cross-validation. *Encycl. Database Syst.* 5, 532–538. doi:10.1007/978-0-387-39940-9_565

Sander, J. D., and Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32 (4), 347–355. doi:10.1038/nbt.2842

Sanjana, N. E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* 11 (8), 783–784. doi:10.1038/nmeth.3047

Shan, Q., Wang, Y., Li, J., and Gao, C. (2014). Genome editing in rice and wheat using the CRISPR/Cas system. *Nat. Protoc.* 9 (10), 2395–2410. doi:10.1038/nprot.2014.157

Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* 21 (2), 110–124. doi:10.1016/J.TPLANTS.2015.10.015

Sovová, T., Kerins, G., Demnerová, K., and Ovesná, J. (2016). Genome editing with engineered nucleases in economically important animals and plants: State of the art in the research pipeline. *Curr. Issues Mol. Biol.* 21, 41–62.

Stemmer, M., Thumberger, T., del Sol Keyer, M., Wittbrodt, J., and Mateo, J. L. (2015). CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLOS ONE* 10 (4), e0124633. doi:10.1371/journal.pone.0124633

Sun, X., Hu, Z., Chen, R., Jiang, Q., Song, G., Zhang, H., et al. (2015). Targeted mutagenesis in soybean using the CRISPR-Cas9 system. *Sci. Rep.* 5 (1), 10342–10410. doi:10.1038/srep10342

Svitashev, S., Schwartz, C., Lenderts, B., Young, J. K., and Mark Cigan, A. (2016). Genome editing in maize directed by CRISPR–Cas9 ribonucleoprotein complexes. *Nat. Commun.* 7 (1), 13274–13277. doi:10.1038/ncomms13274

Terns, M. P., and Terns, R. M. (2011). CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* 14 (3), 321–327. doi:10.1016/J.MIB.2011.03.005

Tian, S., Jiang, L., Gao, Q., Zhang, J., Zong, M., Zhang, H., et al. (2016). Efficient CRISPR/Cas9-based gene knockout in watermelon. *Plant Cell. Rep.* 36 (3), 399–406. doi:10.1007/s00299-016-2089-5

Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., and Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* 11 (9), 636–646. doi:10.1038/nrg2842

van Dijk, A. D. J., Kootstra, G., Kruijer, W., and de Ridder, D. (2021). Machine learning in plant science and plant breeding. *iScience* 24 (1), 101890. doi:10.1016/J.ISCI.2020.101890

Wang, M., Mao, Y., Lu, Y., Tao, X., and Zhu, J. (2017). Multiplex gene editing in rice using the CRISPR-cpf1 system. *Mol. Plant* 10, 1011–1013. doi:10.1016/j.molp.2017.03.001

Wang, P., Zhang, J., Sun, L., Ma, Y., Xu, J., Liang, S., et al. (2018). High efficient multisites genome editing in allotetraploid cotton (Gossypium hirsutum) using CRISPR/Cas9 system. *Plant Biotechnol. J.* 16 (1), 137–150. doi:10.1111/pbi.12755

Wang, S., Zhang, S., Wang, W., Xiong, X., Meng, F., and Cui, X. (2015). Efficient targeted mutagenesis in potato by the CRISPR/Cas9 system. *Plant Cell. Rep.* 34 (9), 1473–1476. doi:10.1007/s00299-015-1816-7

Westra, E. R., Buckling, A., and Fineran, P. C. (2014). CRISPR-cas systems: Beyond adaptive immunity. *Nat. Rev. Microbiol.* 12 (5), 317–326. doi:10.1038/nrmicro3241

Wood, A. J., Lo, T. W., Zeitler, B., Pickle, C. S., Ralston, E. J., Lee, A. H., et al. (2011). Targeted genome editing across species using ZFNs and TALENs. *Science* 333 (6040), 307.

Xu, R. F., Li, H., Qin, R. Y., Li, J., Qiu, C. H., Yang, Y. C., et al. (2015). Generation of inheritable and "transgene clean" targeted genome-modified rice in later generations using the CRISPR/Cas9 system. *Sci. Rep.* 5 (1), 11491–11501. doi:10.1038/srep11491

Xu, R., Li, H., Qin, R., Wang, L., Li, L., Wei, P., et al. (2014). Gene targeting using the Agrobacterium tumefaciens-mediated CRISPR-Cas system in rice. *Rice* 7 (1), 5–4. doi:10.1186/s12284-014-0005-6

Xu, X., Duan, D., and Chen, S. J. (2017). CRISPR-Cas9 cleavage efficiency correlates strongly with target-sgRNA folding stability: From physical mechanism to off-target assessment. *Sci. Rep.* 7 (1), 143–152. doi:10.1038/s41598-017-00180-1

Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., et al. (2014). The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol. J.* 12 (6), 797–807. doi:10.1111/pbi.12200

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., et al. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat. Commun.* 7 (1), 12617–12618. doi:10.1038/ncomms12617

Zhou, H., Liu, B., Weeks, D. P., Spalding, M. H., and Yang, B. (2014). Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. *Nucleic Acids Res.* 42 (17), 10903–10914. doi:10.1093/nar/gku806

Zhou, T., Yang, L., Lu, Y., Dror, I., Dantas Machado, A. C., Ghane, T., et al. (2013). DNAshape: A method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res.* 41 (W1), W56–W62. doi:10.1093/nar/gkt437

frontiers | Frontiers in Genetics

TYPE Mini Review
PUBLISHED 13 February 2023
DOI 10.3389/fgene.2023.1017388

OPEN ACCESS

EDITED BY
Deepmala Sehgal,
Syngenta, United Kingdom

REVIEWED BY
Babar Hussain,
University of Central Punjab, Pakistan

*CORRESPONDENCE
Yan Zhang,
zhangyan6032@163.com
Qiang Zhang,
zhangqiang1858@163.com

SPECIALTY SECTION
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

RECEIVED 12 August 2022
ACCEPTED 23 January 2023
PUBLISHED 13 February 2023

CITATION
Hou X, Guo X, Zhang Y and Zhang Q (2023), CRISPR/Cas genome editing system and its application in potato. *Front. Genet.* 14:1017388.
doi: 10.3389/fgene.2023.1017388

COPYRIGHT
© 2023 Hou, Guo, Zhang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

CRISPR/Cas genome editing system and its application in potato

Xin Hou, Xiaomeng Guo, Yan Zhang* and Qiang Zhang*

College of Plant Protection, Shandong Agricultural University, Tai'an, China

Potato is the largest non-cereal food crop worldwide and a vital substitute for cereal crops, considering its high yield and great nutritive value. It plays an important role in food security. The CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated) system has the advantages of easy operation, high efficiency, and low cost, which shows a potential in potato breeding. In this paper, the action mechanism and derivative types of the CRISPR/Cas system and the application of the CRISPR/Cas system in improving the quality and resistance of potatoes, as well as overcoming the self-incompatibility of potatoes, are reviewed in detail. At the same time, the application of the CRISPR/Cas system in the future development of the potato industry was analyzed and prospected.

KEYWORDS

potato, gene editing, CRISPR/Cas, genetic improvement, progress

1 Introduction

CRISPR/Cas is a part of the CRISPR adaptive immune system, including CRISPR and CRISPR-related protein genes (Sorek et al., 2008), that has been developed as an important tool in genome editing (Maximiano et al., 2021). Compared with transcription activator-like effector nuclease (TALEN) (Christian et al., 2010) and zinc-finger nuclease (ZFN) (Urnov et al., 2010), CRISPR/Cas has obvious advantages of simplicity, flexibility, efficiency, and economy (Cho et al., 2013; Wang et al., 2013; Castro et al., 2021). TALEN and ZFN technologies are technically difficult, making the construction of vectors time consuming; they cannot be efficiently utilized in routine laboratories, while CRISPR/Cas technologies are relatively simple to operate, making their use in the production of vectors inexpensive. They can be more widely used in general molecular biology laboratories. Therefore, CRISPR/Cas has become the most powerful genetic tool for crop character improvement and quality optimization. At present, it has been successfully applied to *Arabidopsis thaliana*, *Sorghum bicolor*, *Nicotiana tabacum*, *Oryza sativa*, and other plants (Jiang et al., 2013; Hussain et al., 2018; Jaganathan et al., 2018; Chen et al., 2019).

The potato is the third most important food crop in the world after rice and wheat in terms of human consumption (Hameed et al., 2018), and it directly affects the yield and quality of food production. Cultivated potato is tetraploid, highly heterozygous, and vegetatively propagated. Therefore, the development of new cultivars using traditional breeding methods is a long-term effort (Zhu et al., 2020). The CRISPR/Cas system has been applied to improve the genetic breeding and agricultural characteristics of potatoes and shown great potential in accelerating breeding, increasing yield, optimizing quality, and improving stress resistance. This paper introduces the types, principles, and applications of

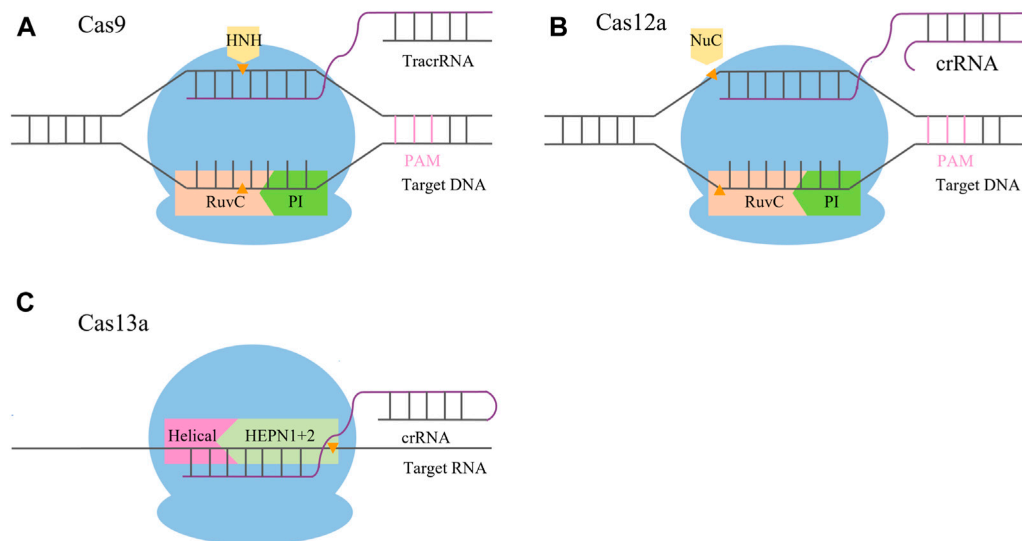


FIGURE 1
Schematic comparison of CRISPR/Cas systems. **(A)** CRISPR/Cas9 system mediates its function *via* Cas9 and two RNAs, the crRNA and tracrRNA. After hybridization, the crRNA/tracrRNA complex associates with the Cas9 nuclease and binds to its recognition site upstream of the PAM sequence. Recognition of the crRNA/tracrRNA/target complex is mediated by the REC (recognition) lobe; the PI (PAM interacting) domain is in charge of PAM recognition. The DSB is mediated by the HNH (orange) and RuvC nuclease domains (pink), with the HNH domain cleaving the target and the RuvC domain cleaving the non-target strand. **(B)** CRISPR/Cas12a system mediates its function *via* Cas12a and a single crRNA. After hybridization, the complex binds to its recognition site downstream of the PAM sequence. Recognition of the crRNA/target complex is mediated by the REC (recognition) lobe; the PI (PAM interacting, green) domain is in charge of PAM recognition. The DSB is mediated by the Nuc (orange) and RuvC domains (pink), with the Nuc domain cleaving the target strand and the RuvC domain cleaving the non-target strand. **(C)** CRISPR/Cas13 system mediates its function *via* Cas13 and a single crRNA. After hybridization, the complex binds to its recognition site within the target RNA mediated by the guide sequence of the crRNA. Recognition of the crRNA/target complex is mediated by the REC (recognition) lobe; the target RNA is cleaved by the HEPN domain (light green).

the CRISPR/Cas system in potatoes and suggests future applications of this technology in improving the characteristics of potatoes.

2 CRISPR-Cas structures and mechanisms

2.1 CRISPR/Cas9

The CRISPR/Cas9 system is a complex formed by guide RNA (gRNA) and endonuclease (Cas9). gRNA is a single-stranded RNA with a specific structure (Hille et al., 2018) that can combine with the target gene and guide the Cas protein to cut the target gene. Cas9 enzymes contain an HNH domain that cleaves the DNA strand that is complementary to the guide RNA sequence (target strand) and a RuvC nuclease domain required for cleaving the non-complementary strand (non-target strand), yielding double-strand DNA breaks (DSBs) (Jinek et al., 2014) (Figure 1A). In order to protect the genome from DSB damage, cells will undergo non-homologous end joining (NHEJ) and homology-directed repair (HR). Among them, NHEJ is the main repair pathway, which is efficient for direct repair by inserting or deleting a few bases at the end. However, the insertion or deletion of bases is random and cannot be edited accurately (Mladenov and Iliakis, 2011; Wu et al., 2013; Gantz and Bier, 2015). The HR pathway is an accurate repair mechanism that can insert specific targeted modification genes and generate accurate mutations at the cutting site for accurate editing, but the repair efficiency is low (Siebert and Puchta, 2002). Jinek et al. found that CRISPR/Cas9 could recognize a 5′-NGG-3′ protospacer adjacent motif (PAM) sequence. When the Cas9 binds with PAM and the target site pairs with the gRNA, a double-strand break (DSB) is caused between positions 17 and 18 of the 20-nt gRNA sequence (Jinek et al., 2012). This laid a theoretical foundation for studying the application of the CRISPR/Cas9 system in plants. In 2013, scientists successfully edited human cells using CRISPR/Cas9 (Cong et al., 2013; Mali et al., 2013). Subsequently, the CRISPR/Cas system was rapidly applied in the field of plants, and site-directed mutations of specific gene loci were achieved in rice, wheat, maize, *Arabidopsis*, and other plants (Li et al., 2013; Nekrasov et al., 2013; Shan et al., 2013; Matres et al., 2021).

2.2 CRISPR/Cas12a

The CRISPR/Cas9 system has the disadvantages of a limited selection of target sites and a high off-target rate that hinder studies using the system. The CRISPR/Cas12 system greatly expands the choice of target sites for genome editing and has significant advantages in safety and accuracy. Therefore, it may become a new gene editing technology to replace the CRISPR/Cas9 system (Dong et al., 2016; Fonfara et al., 2016; Shmakov et al., 2017). The novel CRISPR DNA endonuclease (Cpf1), also known as Cas12a protein, was first reported by the Zhang Feng lab (Zetsche et al., 2015; Wu et al., 2018) in 2015, and it was used in plant gene editing in 2016 (Endo et al., 2016). Cas12a processes its own pre-crRNA into mature crRNAs, without the requirement of a tracrRNA, making it a

more minimalistic system than Cas9. Cas12a possesses an RuvC domain and a nuclease (Nuc) domain. The RuvC domain of Cas12a cleaves both the non-complementary and the complementary strands of the double-stranded DNA (dsDNA), while the Nuc domain assists in the cleavage process. Therefore, DNA is cut in the same nuclease site by Cas12a, producing a staggered DSB that promotes HDR instead of NHEJ (Shmakov et al., 2015) (Figure 1B). In addition, it was proven that the off-target rate of CRISPR/Cas12a is lower than that of CRISPR/Cas9, which can ensure precise targeting of DNA (Bandyopadhyay et al., 2020).

2.3 CRISPR/Cas13

The CRISPR/Cas13 system is a novel gene editing technology with fascinating prospects due to its characteristic of specific RNA targeting. The CRISPR/Cas13 system is composed of two enzymatically active higher eukaryotes and prokaryotes nucleotide-binding (HEPN) *RNase* domains. One *RNase* is responsible for crRNA preprocessing, helping to form a mature interference complex, whereas the other one has two HEPN *endoRNase* domains that mediate the precise cleavage of RNA. Generally, the Cas13 protein families have been divided into four subtypes, namely, Cas13a, Cas13b, Cas13c, and Cas13d (Cox et al., 2017). Cas13a contains a Helical-1 domain and two HEPN domains. Upon the formation of guide-target RNA duplexes, Cas13a is activated by triggering the HEPN1 domain to move toward the HEPN2 domain and subsequently bind and cleave target RNA bearing a complementary sequence (Cox et al., 2017) (Figure 1C). The CRISPR/Cas13 system has been applied to confer modest interference against RNA viruses in mammalian cells and plants (Aman et al., 2018; Freije et al., 2019).

2.4 Base editors

The first generation of a cytosine base editor (CBE) was constructed by the fusion of rat cytosine deaminase (APOBEC) and dCas9 to form the APOBEC1-XTEN-dCas9 base editing system (Huang et al., 2021). APOBECs catalyze the deamination of cytosine bases in nucleic acids, which leads to a conversion of target cytosine (C) to uracil (U) and, consequently, a change in the single-stranded DNA/RNA sequence (Komor et al., 2016; Gaudelli et al., 2017). First-generation CBE was not effective in human cells due to cellular-mediated repair of the U-G intermediate in DNA by the base excision repair (BER) pathway. BER of U-G in DNA is initiated by uracil N-glycosylate (UNG), which recognizes the U-G mismatch and cleaves the glycosidic bond between the uracil and the deoxyribose backbone of DNA, resulting in reversion of the U-G intermediate created by the base editor back to the C-G base pair. The second generation of CBE (rAPOBEC1-XTEN-dCas9-UG) was improved by adding a uracil glycosylase inhibitor (UGI), inhibiting the activity of UDG. In the third generation of CBE, an APOBEC1-XTEN-Cas9 (D10A) base editing system was constructed by fusing the n-terminal of Rat APOBEC1 with Cas9 (D10A), which is a more efficient and convenient editing system. It has been successfully applied in crop breeding and is the most popular base editing tool at present. The fourth generation of CBE was generated by fusing an additional copy of UGI to the N-terminus of nCas9 with an optimized 27-bp linker, which greatly improved the accuracy of transformation and reduced the generation of non-target products (Gaudelli et al., 2017) (Figure 2A). The adenine base editor (ABE) is composed of the fusion of nCas9 (D10A) and adenosine deaminase. The mechanism of the ABE-mediated DNA editing operation is similar to that of CBE. The ABE-dCas9 fusion binds to a target DNA sequence in a guide RNA-programmed manner, and the deoxyadenosine deaminase domain catalyzes an adenine (A) to an inosine (I) transition. In the context of DNA replication, inosine is interpreted as guanine, and the original A-T base pair may be replaced with a G-C base pair at the target site (Gaudelli et al., 2017; Tabassum et al., 2021) (Figure 2B).

Prime editors (PEs) comprise a fusion protein consisting of a catalytically impaired SpCas9 nickase (H840A) fused to an engineered reverse transcriptase (RT) that can produce template-directed local sequence changes in the genome without the requirement for a DSB or exogenous donor DNA templates. Lin et al. adapted PEs for use in plants through codon, promoter, and editing-condition optimization (Lin et al., 2020). Anzalone et al. optimized the PEs to a "search-and-replace" genome editing technology that mediates targeted insertions, deletions, all 12 possible base-to-base conversions, and combinations thereof in human cells without requiring DSBs or donor DNA templates (Anzalone et al., 2019). Mok et al. recently developed DdCBE to enable targeted C•G-to-T•A conversions within mitochondria that have been applied for mitochondrial base editing in human embryos, mice, zebrafish, and plants. It opened a new door for base editors (Mok et al., 2022).

3 Applications of CRISPR/Cas in potato research

CRISPR/Cas system-mediated genome editing technology can selectively modify any genes controlling the stress resistance and nutritional quality of tuber crops to obtain desired traits. This technology plays an important role in accelerating the process of breeding, improving crop yield and quality, and enhancing stress resistance. It is a next-generation breeding technology. Potato is both an important food crop and also provides much raw material for the food processing industry (Sevestre et al., 2020). Its yield and quality are critical to global food security. The CRISPR/Cas system has been widely used in potato genetic improvement (Butler et al., 2015; Wang et al., 2015). It is an effective tool to promote the breeding of potatoes with excellent traits (Table 1).

3.1 Overcoming the self-incompatibility of potatoes

Despite the social and economic importance of potatoes, their breeding success has remained low because self-incompatibility hindered the development of inbred lines. *S-RNase* is a key gene controlling self-incompatibility in potatoes. Ye et al. knocked out *S-RNase* using the CRISPR Cas9 system to create self-compatible (SC) diploid potatoes, which provides a new tool for diploid potato breeding (Ye et al., 2018). Meanwhile, Enciso-Rodriguez et al. also

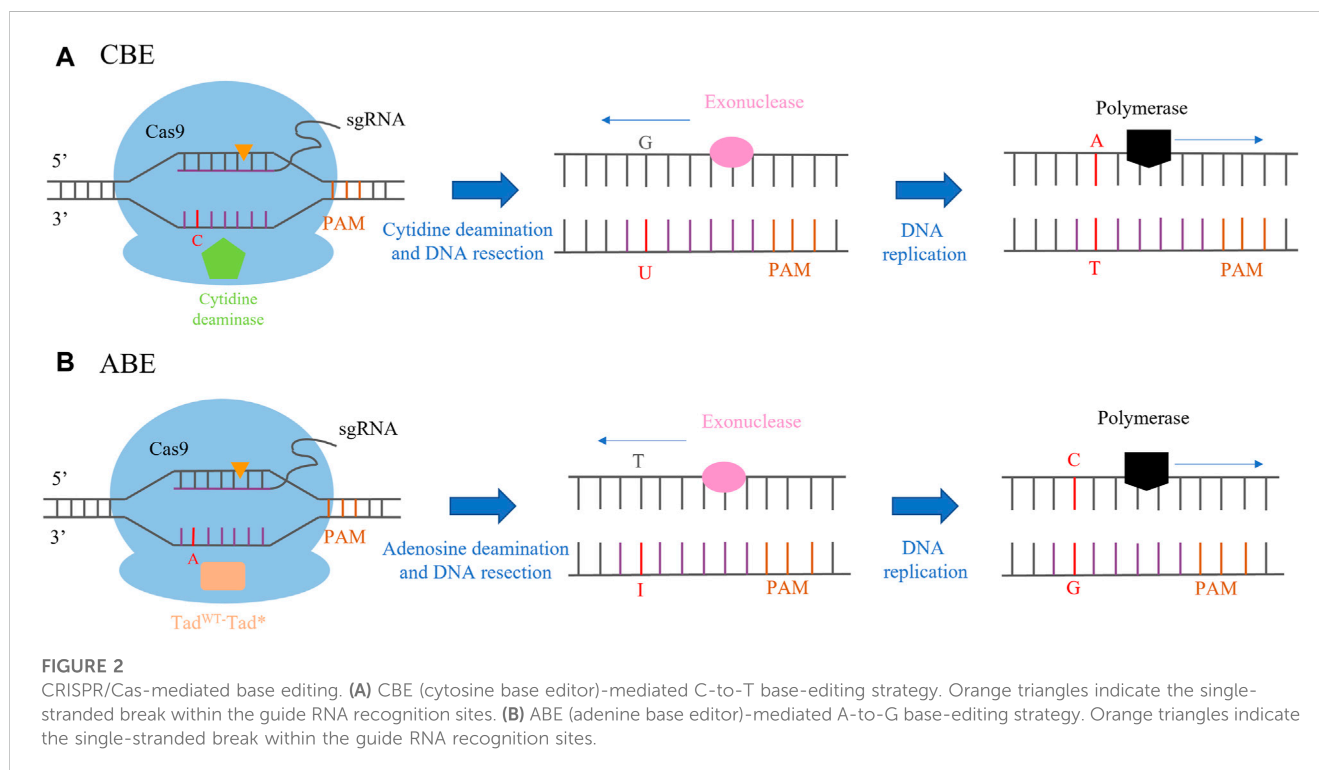


FIGURE 2
CRISPR/Cas-mediated base editing. **(A)** CBE (cytosine base editor)-mediated C-to-T base-editing strategy. Orange triangles indicate the single-stranded break within the guide RNA recognition sites. **(B)** ABE (adenine base editor)-mediated A-to-G base-editing strategy. Orange triangles indicate the single-stranded break within the guide RNA recognition sites.

generated SC diploid lines with stable self-compatibility by targeted mutagenesis of *S-RNase* using CRISPR-Cas9 (Taylor, 2018; Enciso-Rodriguez et al., 2019). This strategy accelerates the process of diploid potato breeding and will also be useful for studying other self-incompatible crops. In addition, the haplotype-resolved genome of heterozygous diploid potatoes and tetraploid potatoes was decoded, which lays an important foundation for genome editing-assisted breeding (Zhou et al., 2020; Sun et al., 2022). The *Sli* gene was knocked out by CRISPR/Cas9 to transform SC varieties into SI varieties, and its function was studied. In addition, a 533-bp insertion fragment was found in the promoter region of the *Sli* gene, enabling it to be expressed in pollen. In future breeding work, this region can be introduced into the promoter of the *Sli* gene in SI potatoes by directional insertion to make it become an *SC* gene (Eggers et al., 2021).

3.2 Improving the biotic and abiotic stress resistance of potatoes

Abiotic and biotic stresses are the main factors that affect plant growth and limit agricultural productivity (Deja-Sikora et al., 2020; Sharma and Gayen, 2021). Abiotic stresses such as salinity, drought, extreme temperatures, and heavy metals (Munns and Tester, 2008; Siddiqui et al., 2017; Yang and Guo, 2018; Liu et al., 2022) are important factors affecting plant growth and development that can lead to the destruction of the original physiological characteristics of plants (Mueller and Levin, 2020). Biotic stresses of plants are caused by viral, fungal, and bacterial infections. Potatoes are usually subjected to stress from various diseases and insect pests during growth (Savary et al., 2012), such as potato early blight, late blight, ring rot, bacterial wilt, and aphids, which can lead to 30%–60% economic losses (Brown, 2011).

The CRISPR/Cas gene editing system plays an important role in accelerating the breeding process of potatoes that are highly resistant to abiotic and biotic stresses. Zhou et al., (2017) obtained the *tMYB44* mutant through the gene editing system and demonstrated that this gene negatively regulates phosphate transport in potatoes by inhibiting the expression of *StPHO1*. Tiwari et al. (2021) found that potato late blight resistance genes (*R3a*, *RGA2*, *RGA3*, *R1B-16*, *Rpi-blb2*, *Rpi*, and *Rpi-vnt1*) and susceptibility genes (S genes) are of great significance in enhancing resistance to pathogens and provided a theoretical basis for studying resistance to late blight genes. Kieu et al. (2021) found that functional knockouts of S genes, that is, creating *StDND1*, *StCHL1*, and *StDMR6-1* by the CRISPR/Cas9 system, increased resistance against late blight in potatoes. It is the first report of increasing resistance to late blight in potatoes by editing the S gene. Hegde et al. (2021) used a gene editing system to achieve directed mutation of the *StCCoAOMT* gene in the potato, which greatly improved potato resistance to late blight. Gonzales et al. (2021) mutated the *StFLORE* promoter through CRISPR/Cas9 and detected a correlation between the transcription factor *StCDF1* and the antisense transcript *StFLORE* enhancing drought resistance. The mutant could better regulate the size and number of stomata to enhance drought resistance. CRISPR/Cas13a is an RNA-targeting CRISPR effector that provides protection against RNA phages. Zhan et al. reported the repurposing of CRISPR/Cas13a to protect potato plants from a eukaryotic virus, potato virus Y (PVY), by designing sgRNAs against conserved coding regions of three different PVY strains. The levels of viral resistance correlated with the expression levels of the Cas13a/sgRNA construct in the plants.

TABLE 1 Applications of CRISPR/Cas system-mediated genome editing technology in potato improvement.

Trait	Editing tool	The name of target genes	Type of edit	References
Genetic breeding	CRISPR/Cas9	*Sp3* and *Sp4*	Gene knockout	Ye et al. (2018)
	CRISPR/Cas9	*S-Rnase*	Gene knockout	Taylor (2018) and Enciso-Rodriguez et al. (2019)
	CRISPR/Cas9	*StD6PK* and *StSIEL*	Gene knockout	Zhou et al. (2020) and Sun et al. (2022)
	CRISPR/Cas9	*Sli*	Gene knockout	Eggers et al. (2021)
Stress resistance	CRISPR/Cas9	*StDND1*, *StCHL1*, and *StDMR6-1*	Gene knockout	Kieu et al. (2021)
	CRISPR/Cas9	*StFLORE*	Promoter mutation	Gonzales et al. (2021)
	CRISPR/Cas9	*StCCoAOMT*	Gene knockout	Hegde et al. (2021)
	CRISPR/Cas9	*tMYB44*	Gene knockout	(Zhou et al., 2017)
	CRISPR/Cas13a	*LshCas13a*	Gene knockout	Zhan et al. (2019)
Improved quality	CRISPR/Cas9	*StGBSS*	Gene knockout	Andersson et al. (2018)
	CRISPR/Cas9	*StGBSSI*	Gene knockout	Veillet et al. (2019)
	A3A-CBE	*StGBSSI*	Base editing	Zong et al. (2018)
	CRISPR/Cas9	*StGBSSI*	Gene knockout	Wang et al. (2019)
	CRISPR/Cas9	*StGBSSI*	Gene knockout	Johansen et al. (2019)
	PmCDA1-CBE	*StGBSSI*	Base editing	Veillet et al. (2020)
	CRISPR/Cas9	*StGBSSI*	Gene knockout	Toinga-Villafuerte et al. (2022)
	CRISPR/Cas9	*SBE1* and *SBE2*	Gene knockout	Tuncel et al. (2019)
	CRISPR/Cas9	*SBE1* and *SBE2*	Gene knockout	Zhao et al. (2021)
	CRISPR/Cas9	*AtCGS* and *StMGL*	Gene knockout	Helle et al. (2018)
	CRISPR/Cas9	*St16DOX*	Gene knockout	Nakayasu et al. (2018)
	CRISPR/Cas9	*StSSR2*	Gene knockout	Zheng et al. (2021)
	CRISPR/Cas9	*StPPO2*	Gene knockout	González et al. (2019)
Improved yield	CRISPR/Cas9	*StIT1*	Gene knockout	Tang et al. (2022)

This work showed the potential of the CRISPR/Cas13a system to confer stable resistance to an important viral disease in a major crop (Ji et al., 2015; Zhan et al., 2019).

3.3 Improving the tuber quality of potatoes

For potatoes, a pure amylopectin starch has advantages in facilitating sustainable downstream processing with decreased use of chemicals and energy compared to native starch. CRISPR-Cas9 was developed as a potato breeding method by implementing RNP delivery in protoplasts to decrease or eliminate the presence of unintended inserts in progeny. Andersson et al. created amylopectin starch potatoes by knocking out the gene of granule-bound starch synthase (GBSS) through this method, which shows the potential of CRISPR-Cas9 RNP technology as a future potato breeding method (Andersson et al., 2018). Another strategy to develop an amylopectin potato was described soon afterward, where the use of base editing (BE) to knock out the amylose-producing *StGBSSI* gene was found successful. Amylopectin starch is most likely the most progressed trait developed with genome editing in potatoes, and non-transgenic amylopectin lines are currently grown in the field for selection and seed multiplication (Veillet et al., 2019). Many other studies report using *GBSS* as a target gene to reduce amylose content (Zong et al., 2018; Johansen et al., 2019; Tuncel et al., 2019; Wang et al., 2019; Veillet et al., 2020; Zhao et al., 2021; Toinga-Villafuerte et al., 2022).

SS6, a recently discovered starch synthase isoform, was identified as a key enzyme of the starch biosynthetic pathway (Helle et al., 2018). Sevestre et al. used a BE genome editing system to knock out the SS6 gene in potatoes. The inactivation of this enzyme may lead to modifications of starch properties, potentially resulting in industrial applications (Helle et al., 2018). Apart from its nutrients, potato also contains some anti-nutrients, such as nitrates V, toxic glycoalkaloids, and nitrates III, which can damage its quality. Steroidal glycoalkaloids (SGAs) existing in most potato tissues confer a bitter taste and show toxicity when the fresh weight is over 200 mg kg^{-1}. Therefore, controlling the SGA levels in the tubers is an important focus of potato breeding. *St16DOX* encodes a steroid 16α-hydroxylase in SGA biosynthesis, which

exists as the single gene in the potato genome. Therefore, this is a preferable target for genome editing to generate an SGA-free potato. Nakayasu et al. knocked out *St16DOX* through CRISPR/Cas 9, resulting in the complete abolition of the SGA accumulation in potato hairy roots (Nakayasu et al., 2018). These results provided a reference for CRISPR/Cas9 to create crops without the potential gene loci of SGAs. Potato tubers and roots are also rich in SGAs. Zheng et al. conducted targeted mutagenesis of the sterol side chain reductase 2 gene (*StSSR2*) using the CRISPR/Cas9 system. The results revealed that the SGA level in leaves and tubers was decreased by 66% and 44%, respectively. In addition, the relative transcript levels of genes involved in SGA biosynthesis pathways were also reduced (Zheng et al., 2021). Enzymatic browning catalyzed by polyphenol oxidases (PPOs) leads to the formation of dark-colored precipitates in potatoes, causing undesirable changes in organoleptic properties and the loss of nutritional quality. González et al. induced mutations in the *StPPO2* gene using a CRISPR/Cas9 system in the tetraploid cultivar Desiree. The result showed that specific editing of the *StPPO2* gene resulted in a reduction of up to 69% in tuber PPO activity and a reduction of 73% in enzymatic browning, compared to the control, which demonstrates that the CRISPR/Cas9 system can be applied to develop potato varieties with reduced enzymatic browning in tubers (González et al., 2019).

The nutritional quality of potatoes can be improved from two aspects. On one hand, the expression of the potato nutrient synthesis gene is enhanced by gene editing technology so as to improve potato nutrient content. On the other hand, adverse substances in potatoes can be reduced by knocking out the genes of potato anti-nutritional compounds and toxins (Clasen et al., 2016). Improving the nutritional value of potatoes through gene editing technology can help potatoes become a dietary staple in many countries.

3.4 Improving yield of potatoes

The United Nations has long listed potatoes as one of the four main foods along with corn, wheat, and rice. Potatoes have been a staple food in Western developed countries for 50–60 years. Potatoes have a great potential to increase production, nutritional value, good taste, and storage; they have a long industrial chain, and strong processing and conversion ability. The use of biotechnological tools to improve potato yields has been increasingly reported. The exogenous sucrose phosphate synthetase gene was successfully introduced into the potato, which improved the supply of photosynthates from leaves (source) to tubers (pool), thus improving the yield and quality of potato tubers (Ishimaru et al., 2008). Tuber yield was increased by inserting purple acid phosphatase 2 of *Arabidopsis* (*AtPAP2*) (Zhang et al., 2014). The successful insertion of the *Agrobacterium auxin* biosynthesis gene increased the content of indoleacetic acid in tubers and tuber formation, thus increasing yield (Kolachevskaya et al., 2015). Tuber yield can be improved by down-regulating the sucrose transporter 4 (*stsu4*) gene, the negative regulator of tubers (Chincinska et al., 2008). Tang et al. proved that the *StIT1* gene was mutated by CRISPR/Cas9, resulting in stolon branching, which laid a foundation for increased yields (Tang et al., 2022).

4 Future prospects

With the growth of world population, people pay more attention to food crops. Potato is one of the important food crops and has obvious advantages over other food crops in terms of per-mu yield, cost, and cultivation conditions (Zaheer and Akhtar, 2016). However, there are some major challenges facing potato production. The excellent traits of potato clones are difficult to maintain through sexual reproduction, due to the high heterozygosity in the tetraploid potato genome (Loebenstein, 2006; Nadakuduti et al., 2018). Various biotic and abiotic stresses may cause crop failure and yield loss (Xu and Gray, 2020); the toxic or anti-nutritional compounds in potatoes affect their consumption and processing (Scholthof et al., 2011; Wang et al., 2011). Classical improvement schemes comprise relatively long breeding cycles and are dictated by the genetic complexity and the sensitivity of potato to inbreeding depression. In order to meet the increasing demand for potato production, more efficient approaches for potato breeding are required. The CRISPR/Cas system, as the mainstream genome editing technique, can accelerate plant breeding by providing the means to modify genomes rapidly in a precise and predictable manner (Li M R et al., 2016; Li S C et al., 2016; Luo et al., 2016; Zhang et al., 2016; Liu et al., 2017; Lu et al., 2018; Zhang et al., 2018) and has achieved remarkable results in crop breeding. The CRISPR/Cas system has been applied to improving potato yield, quality, stress resistance, genetic breeding, and agricultural characteristics of potatoes and has shown great potential in accelerating breeding, increasing yield, optimizing quality, improving stress resistance, acquiring herbicide resistance, and reducing postharvest nutrient loss (Braatz et al., 2017; Li et al., 2017; Dong et al., 2018; Yao et al., 2018). The availability of the potato genome sequence allows scientists to precisely design its genome for crop improvement, which will facilitate the application of gene editing technology. In addition, the major genes that determine many important traits have been discovered or identified in wheat (Zong et al., 2017), maize (Svitashev et al., 2016; Chen et al., 2018; Teng et al., 2020; Jiang et al., 2021; Naik et al., 2022), rice (Zhou et al., 2019; Kuang et al., 2020; Tang et al., 2021), and soybean (Wang et al., 2020; Cai et al., 2021). The homologous genes of these genes in potato could be found and edited by the CRISPR/Cas system to construct transformants with expected traits.

The CRISPR/Cas9 gene editing system, while efficient, is not precise, leaving it vulnerable to knocking out a gene and producing many unwanted results. If the goal is to optimize the function of genes, rather than simply suppress or knock them out, the lack of predictability of the CRISPR/Cas9 system makes gene editing less feasible. Many important agronomic traits in plants are caused by single or a few base mutations. Single-site mutations in plants can be obtained based on traditional chemical mutagenesis, but this method is time-consuming, labor-intensive, random, and has a low mutation efficiency. In potato trait improvement, base editing technology can not only generate functionally acquired variants by changing individual bases but also achieve targeted evolution of specific genes by constructing sgRNA libraries to generate a library of mutations

for a single gene. Ultimately, the process of potato genetic breeding is accelerated.

Although gene editing technology has great potential in agricultural production, attitudes toward the technology vary globally. To avoid falling behind the rest of the world in the gene-editing race, more than 15 countries, including China, India, Argentina, and Australia, have open rules for crop gene-editing, but these countries distinguish gene-edited crops from conventional breeding products. Many countries, such as the United States, Canada, Brazil, and Japan, do not distinguish gene editing from traditional breeding. With the development of gene editing technology, it is reasonable that more countries will remove restrictions on gene editing in the future. Different parts of the plant may have different safety risks. As a root crop, the safety risks of transgenic potato need to be further determined.

The CRISPR/Cas system has already made significant gains in potato breeding, and we expect that this is just the beginning, with many more exciting developments to follow. With the development of second-generation sequencing, gene editing technology and target analysis technology based on the high-throughput sequencing method have a solid technical foundation, and the acquisition of high-throughput big data has become more common, convenient, and affordable. These advances will greatly promote the application of CRISPR/Cas in potato genetic improvement.

Author contributions

QZ and XH designed the article. YZ and XG drafted the manuscript and arranged the references. QZ and XH incorporated all necessary modifications. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript.

Funding

The study was supported by the Shandong Weifang Tobacco Company Science and Technology Project (2022-26), the China Tobacco Shandong Industrial Corporation Science and Technology Project (202102004), and the Shandong Province Modern Agricultural Technology System (SDAIT-25-02).

Acknowledgments

The authors thank all of the others in their laboratory for providing useful discussions and technical assistance. They are very grateful to the editor and reviewers for critically evaluating the manuscript and providing constructive comments for its improvement.

Conflict of interest

The authors declare that this study received funding from Shandong Weifang Tobacco Company and Tobacco Shandong Industrial Corporation. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aman, R., Ali, Z., Butt, H., Mahas, A., Aljedaani, F., Khan, M. Z., et al. (2018). RNA virus interference via CRISPR/Cas13a system in plants. *Genome Biol.* 19, 1. doi:10.1186/s13059-017-1381-1

Andersson, M., Turesson, H., Olsson, N., Falt, A. S., Ohlsson, P., Gonzalez, M. N., et al. (2018). Genome editing in potato via CRISPR-Cas9 ribonucleoprotein delivery. *Physiol. Plant.* 164, 378–384. doi:10.1111/ppl.12731

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149, doi:10.1038/s41586-019-1711-4

Bandyopadhyay, A., Kancharla, N., Javalkote, V. S., Dasgupta, S., and Brutnell, T. P. (2020). CRISPR-Cas12a (Cpf1): A versatile tool in the plant genome editing tool box for agricultural advancement. *Front. Plant Sci.* 11, 584151. doi:10.3389/fpls.2020.584151

Braatz, J., Harloff, H. J., Mascher, M., Stein, N., Himmelbach, A., and Jung, C. (2017). CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed Rape (brassica napus). *Plant Physiol.* 174, 935–942. doi:10.1104/pp.17.00426

Brown, C. R. (2011). The contribution of traditional potato breeding to scientific potato improvement. *Potato Res.* 54, 287–300. doi:10.1007/s11540-011-9199-9

Butler, N. M., Atkins, P. A., Voytas, D. F., and Douches, D. S. (2015). Generation and inheritance of targeted mutations in potato (solanum tuberosum L.) using the CRISPR/Cas system. *Plos One* 10, e0144591. doi:10.1371/journal.pone.0144591

Cai, Z. D., Xian, P. Q., Cheng, Y. B., Ma, Q. B., Lian, T. X., Nian, H., et al. (2021). CRISPR/Cas9-mediated gene editing of *GmJAGGED1* increased yield in the low-latitude soybean variety Huachun 6. *Plant Biotechnol. J.* 19, 1898–1900. doi:10.1111/pbi.13673

Castro, N. G., Bjelic, J., Malhotra, G., Huang, C., and Alsaffar, S. H. (2021). Comparison of the feasibility, efficiency, and safety of genome editing technologies. *Int. J. Mol. Sci.* 22, 10355. doi:10.3390/ijms221910355

Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019). CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* 70, 667–697. doi:10.1146/annurev-arplant-050718-100049

Chen, R. R., Xu, Q. L., Liu, Y., Zhang, J. J., Ren, D. T., Wang, G. Y., et al. (2018). Generation of transgene-free maize male sterile lines using the CRISPR/Cas9 system. *Front. Plant Sci.* 9, 1180. doi:10.3389/fpls.2018.01180

Chincinska, I. A., Liesche, J., Krügel, U., Michalska, J., Geigenberger, P., Grimm, B., et al. (2008). Sucrose transporter StSUT4 from potato affects flowering, tuberization, and shade avoidance response. *Plant Physiol.* 146, 515–528. doi:10.1104/pp.107.112334

Cho, S. W., Kim, S., Kim, J. M., and Kim, J. S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 31, 230–232. doi:10.1038/nbt.2507

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., et al. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757–761. doi:10.1534/genetics.110.120717

Clasen, B. M., Stoddard, T. J., Luo, S., Demorest, Z. L., Li, J., Cedrone, F., et al. (2016). Improving cold storage and processing traits in potato through targeted gene knockout. *Plant Biotechnol. J.* 14, 169–176. doi:10.1111/pbi.12370

Cong, L., Ran, F. A., Cox, D., Lin, S. L., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi:10.1126/science.1231143

Cox, D. B. T., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J., et al. (2017). RNA editing with CRISPR-Cas13. *Science* 358, 1019–1027. doi:10.1126/science.aaq0180

Deja-Sikora, E., Kowalczyk, A., Trejgell, A., Szmidt-Jaworska, A., Baum, C., Mercy, L., et al. (2020). Arbuscular mycorrhiza changes the impact of potato virus Y on growth and stress tolerance of solanum tuberosum L. *in vitro*. *Front. Microbiol.* 10, 2971. doi:10.3389/fmicb.2019.02971

Dong, D., Ren, K., Qiu, X. L., Zheng, J. L., Guo, M. H., Guan, X. Y., et al. (2016). The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* 532, 522, doi:10.1038/nature17944

Dong, L., Li, L. N., Liu, C. L., Liu, C. X., Geng, S. F., Li, X. H., et al. (2018). Genome editing and double-fluorescence proteins enable robust maternal haploid induction and identification in maize. *Mol. Plant* 11, 1214–1217. doi:10.1016/j.molp.2018.06.011

Eggers, E. J., Van Der Burgt, A., Van Heusden, S. a. W., De Vries, M. E., Visser, R. G. F., Bachem, C. W. B., et al. (2021). Neofunctionalisation of the Sli gene leads to self-compatibility and facilitates precision breeding in potato. *Nat. Commun.* 12, 4141. doi:10.1038/s41467-021-24267-6

Enciso-Rodriguez, F., Manrique-Carpintero, N. C., Nadakuduti, S. S., Buell, C. R., Zarka, D., and Douches, D. (2019). Overcoming self-incompatibility in diploid potato using CRISPR-Cas9. *Front. Plant Sci.* 10, 376. doi:10.3389/fpls.2019.00376

Endo, A., Masafumi, M., Kaya, H., and Toki, S. (2016). Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from Francisella novicida. *Sci. Rep.* 6, 38169. doi:10.1038/srep38169

Fonfara, I., Richter, H., Bratovic, M., Le Rhun, A., and Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* 532, 517, doi:10.1038/nature17945

Freije, C. A., Myhrvold, C., Boehm, C. K., Lin, A. E., Welch, N. L., Carter, A., et al. (2019). Programmable inhibition and detection of RNA viruses using Cas13. *Mol. Cell* 76, 826, doi:10.1016/j.molcel.2019.09.013

Gantz, V. M., and Bier, E. (2015). Genome editing. The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science* 348, 442–444. doi:10.1126/science.aaa5945

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., et al. (2017). Programmable base editing of A.T to G.C in genomic DNA without DNA cleavage. *Nature* 551, 464–471. doi:10.1038/nature24644

Gonzales, L. R., Shi, L., Bergonzi, S. B., Oortwijn, M., Franco-Zorrilla, J. M., Solano-Tavira, R., et al. (2021). Potato cycling dof factor 1 and its lncRNA counterpart *StFLORE* link tuber development and drought response. *Plant J.* 105, 855–869. doi:10.1111/tpj.15093

González, M. N., Massa, G. A., Andersson, M., Turesson, H., Olsson, N., Fält, A. S. L., et al. (2019). Reduced enzymatic browning in potato tubers by specific editing of a polyphenol oxidase gene via ribonucleoprotein complexes delivery of the CRISPR/Cas9 system. *Front. Plant Sci.* 10, 1649. doi:10.3389/fpls.2019.01649

Hameed, A., Ssea Zaidi, S. S., and Mansoor, S. (2018). Applications of new breeding technologies for potato improvement. *Front. Plant Sci.* 9, 925. doi:10.3389/fpls.2018.00925

Hegde, N., Joshi, S., Soni, N., and Kushalappa, A. C. (2021). The caffeoyl-CoA O-methyltransferase gene SNP replacement in russet burbank potato variety enhances late blight resistance through cell wall reinforcement. *Plant Cel. Rep.* 40 (1), 237–254. doi:10.1007/s00299-020-02629-6

Helle, S., Bray, F., Verbeke, J., Devassine, S., Courseaux, A., Facon, M., et al. (2018). Proteome analysis of potato starch reveals the presence of new starch metabolic proteins as well as multiple protease inhibitors. *Front. Plant Sci.* 9, 746. doi:10.3389/fpls.2018.00746

Hille, F., Richter, H., Wong, S. P., Bratovic, M., Ressel, S., and Charpentier, E. (2018). The biology of CRISPR-cas: Backward and forward. *Cell* 172, 1239–1259. doi:10.1016/j.cell.2017.11.032

Huang, T. P., Newby, G. A., and Liu, D. R. (2021). Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nat. Protoc.* 16, 1089–1128. doi:10.1038/s41596-020-00450-9

Hussain, B., Lucas, S. J., and Budak, H. (2018). CRISPR/Cas9 in plants: At play in the genome and at work for crop improvement. *Briefings Funct. Genomics* 17, 319–328. doi:10.1093/bfgp/ely016

Jaganathan, D., Ramasamy, K., Sellamuthu, G., Jayabalan, S., and Venkataraman, G. (2018). CRISPR for crop improvement: An update review. *Front. Plant Sci.* 9, 985. doi:10.3389/fpls.2018.00985

Ji, X., Zhang, H. W., Zhang, Y., Wang, Y. P., and Gao, C. X. (2015). Establishing a CRISPR-Cas-like immune system conferring DNA virus resistance in plants. *Nat. Plants* 1 (10), 15144. doi:10.1038/nplants.2015.144

Jiang, W. Z., Zhou, H. B., Bi, H. H., Fromm, M., Yang, B., and Weeks, D. P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res.* 41, e188. doi:10.1093/nar/gkt780

Jiang, Y. L., Li, Z. W., Liu, X. Z., Zhu, T. T., Xie, K., Hou, Q. C., et al. (2021). *ZmFAR1* and *ZmABCG26* regulated by microRNA are essential for lipid metabolism in maize anther. *Int. J. Mol. Sci.* 22, 7916. doi:10.3390/ijms22157916

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi:10.1126/science.1225829

Jinek, M., Jiang, F. G., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E. B., et al. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343:1247997, doi:10.1126/science.1247997

Johansen, I. E., Liu, Y., Jørgensen, B., Bennett, E. P., Andreasson, E., Nielsen, K. L., et al. (2019). High efficacy full allelic CRISPR/Cas9 gene editing in tetraploid potato. *Sci. Rep.* 9, 17715. doi:10.1038/s41598-019-54126-w

Kieu, N. P., Lenman, M., Wang, E. S., Petersen, B. L., and Andreasson, E. (2021). Mutations introduced in susceptibility genes through CRISPR/Cas9 genome editing confer increased late blight resistance in potatoes. *Sci. Rep.* 11, 4487. doi:10.1038/s41598-021-83972-w

Kolachevskaya, O. O., Alekseeva, V. V., Sergeeva, L. I., Rukavtsova, E. B., Getman, I. A., Vreugdenhil, D., et al. (2015). Expression of auxin synthesis gene tms1 under control of tuber-specific promoter enhances potato tuberization *in vitro*. *J. Integr. Plant Biol.* 57, 734–744. doi:10.1111/jipb.12314

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420, doi:10.1038/nature17946

Kuang, Y. J., Li, S. F., Ren, B., Yan, F., Spetz, C., Li, X. J., et al. (2020). Base-editing-mediated artificial evolution of *OsALS1* in planta to develop novel herbicide-tolerant rice germplasms. *Mol. Plant* 13, 565–572. doi:10.1016/j.molp.2020.01.010

Li, J. F., Norville, J. E., Aach, J., McCormack, M., Zhang, D. D., Bush, J., et al. (2013). Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nat. Biotechnol.* 31, 688–691. doi:10.1038/nbt.2654

Li, M. R., Li, X. X., Zhou, Z. J., Wu, P. Z., Fang, M. C., Pan, X. P., et al. (2016). Reassessment of the four yield-related Genes *Gn1a, DEP1, GS3,* and *IPA1* in rice using a CRISPR/Cas9 system. *Front. Plant Sci.* 7, 377. doi:10.3389/fpls.2016.00377

Li, S. C., Gao, F. Y., Xie, K. L., Zeng, X. H., Cao, Y., Zeng, J., et al. (2016). The OsmiR396c-OsGRF4-OsGIF1 regulatory module determines grain size and yield in rice. *Plant Biotechnol. J.* 14, 2134–2146. doi:10.1111/pbi.12569

Li, X. F., Zhou, W. J., Ren, Y. K., Tian, X. J., Lv, T. X., Wang, Z. Y. J., et al. (2017). High-efficiency breeding of early-maturing rice cultivars via CRISPR/Cas9-mediated genome editing. *J. Genet. Genomics* 44, 175–178. doi:10.1016/j.jgg.2017.02.001

Lin, Q. P., Zong, Y., Xue, C. X., Wang, S. X., Jin, S., Zhu, Z. X., et al. (2020). Prime genome editing in rice and wheat. *Nat. Biotechnol.* 38, 582, doi:10.1038/s41587-020-0455-x

Liu, J. F., Chen, J., Zheng, X. M., Wu, F. Q., Lin, Q. B., Heng, Y. Q., et al. (2017). *GW5* acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. *Nat. Plants* 3, 17043. doi:10.1038/nplants.2017.43

Liu, Z. G., Ma, C. Y., Hou, L., Wu, X. Z., Wang, D., Zhang, L., et al. (2022). Exogenous SA affects rice seed germination under salt stress by regulating Na+/K+ balance and endogenous GAs and ABA homeostasis. *Int. J. Mol. Sci.* 23, 3293. doi:10.3390/ijms23063293

Loebenstein, G. (2006). *Potato virus diseases, their diagnosis and preparation of virus-tested seed potatoes by rapid propagation*. Bursa, TURKEY. 3rd Balkan Symposium on Vegetables and Potatoes.

Lu, K., Wu, B. W., Wang, J., Zhu, W., Nie, H. P., Qian, J. J., et al. (2018). Blocking amino acid transporter *OsAAP3* improves grain yield by promoting outgrowth buds and increasing tiller number in rice. *Plant Biotechnol. J.* 16, 1710–1722. doi:10.1111/pbi.12907

Luo, M., Gilbert, B., and Ayliffe, M. (2016). Applications of CRISPR/Cas9 technology for targeted mutagenesis, gene replacement and stacking of genes in higher plants. *Plant Cell Rep.* 35, 1439–1450. doi:10.1007/s00299-016-1989-8

Mali, P., Yang, L. H., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., et al. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826. doi:10.1126/science.1232033

Matres, J. M., Hilscher, J., Datta, A., Armario-Najera, V., Baysal, C., He, W. S., et al. (2021). Genome editing in cereal crops: An overview. *Transgenic Res.* 30, 461–498. doi:10.1007/s11248-021-00259-6

Maximiano, M. R., Tavora, Ftpk, Prado, G. S., Mehta, S. C. Dias, A., and Franco, O. L. (2021). CRISPR genome editing technology: A powerful tool applied to developing agribusiness. *J. Agric. Food Chem.* 69, 6379–6395. doi:10.1021/acs.jafc.1c01062

Mladenov, E., and Iliakis, G. (2011). Induction and repair of DNA double strand breaks: The increasing spectrum of non-homologous end joining pathways. *Mutat. Research-Fundamental Mol. Mech. Mutagen.* 711, 61–72. doi:10.1016/j.mrfmmm.2011.02.005

Mok, B. Y., Kotrys, A. V., Raguram, A., Huang, T. P., Mootha, V. K., and Liu, D. R. (2022). CRISPR-free base editors with enhanced activity and expanded targeting scope in mitochondrial and nuclear DNA. *Nat. Biotechnol.* 40, 1378–1387. doi:10.1038/s41587-022-01256-8

Mueller, E. A., and Levin, P. A. (2020). Bacterial cell wall quality control during environmental stress. *Mbio* 11, e02456-20. doi:10.1128/mBio.02456-20

Munns, R., and Tester, M. (2008). Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol.* 59, 651–681. doi:10.1146/annurev.arplant.59.032607.092911

Nadakuduti, S. S., Buell, C. R., Voytas, D. F., Starker, C. G., and Douches, D. S. (2018). Genome editing for crop improvement - applications in clonally propagated polyploids with a focus on potato (solanum tuberosum L.). *Front. Plant Sci.* 9, 1607. doi:10.3389/fpls.2018.01607

Naik, B. J., Shimoga, G., Kim, S. C., Manjulatha, M., Reddy, C. S., Palem, R. R., et al. (2022). CRISPR/Cas9 and nanotechnology pertinence in agricultural crop refinement. *Front. Plant Sci.* 13, 843575. doi:10.3389/fpls.2022.843575

Nakayasu, M., Akiyama, R., Lee, H. J., Osakabe, K., Osakabe, Y., Watanabe, B., et al. (2018). Generation of alpha-solanine-free hairy roots of potato by CRISPR/Cas9 mediated genome editing of the *St16DOX* gene. *Plant Physiology Biochem.* 131, 70–77. doi:10.1016/j.plaphy.2018.04.026

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G., and Kamoun, S. (2013). Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 31, 691–693. doi:10.1038/nbt.2655

Savary, S., Ficke, A., Aubertot, J. N., and Hollier, C. (2012). Crop losses due to diseases and their implications for global food production losses and food security. *Food Secur.* 4, 519–537. doi:10.1007/s12571-012-0200-5

Scholthof, K. B. G., Adkins, S., Czosnek, H., Palukaitis, P., Jacquot, E., Hohn, T., et al. (2011). Top 10 plant viruses in molecular plant pathology. *Mol. Plant Pathol.* 12, 938–954. doi:10.1111/j.1364-3703.2011.00752.X

Sevestre, F., Facon, M., Wattebled, F., and Szydlowski, N. (2020). Facilitating gene editing in potato: A single-nucleotide polymorphism (SNP) map of the Solanum tuberosum L. Cv. Desiree genome. *Sci. Rep.* 10, 2045. doi:10.1038/s41598-020-58985-6

Shan, Q. W., Wang, Y. P., Li, J., Zhang, Y., Chen, K. L., Liang, Z., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 686–688. doi:10.1038/nbt.2650

Sharma, P., and Gayen, D. (2021). Plant protease as regulator and signaling molecule for enhancing environmental stress-tolerance. *Plant Cell Rep.* 40, 2081–2095. doi:10.1007/s00299-021-02739-9

Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., et al. (2015). Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol. Cell* 60, 385–397. doi:10.1016/j.molcel.2015.10.008

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., et al. (2017). Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* 15, 169–182. doi:10.1038/nrmicro.2016.184

Siddiqui, M. N., Mostofa, M. G., Akter, M. M., Srivastava, A. K., Abu Sayed, M., Hasan, M. S., et al. (2017). Impact of salt-induced toxicity on growth and yield-potential of local wheat cultivars: Oxidative stress and ion toxicity are among the major determinants of salt-tolerant capacity. *Chemosphere* 187, 385–394. doi:10.1016/j.chemosphere.2017.08.078

Siebert, R., and Puchta, H. (2002). Efficient repair of genomic double-strand breaks by homologous recombination between directly repeated sequences in the plant genome. *Plant Cell* 14, 1121–1131. doi:10.1105/tpc.001727

Sorek, R., Kunin, V., and Hugenholtz, P. (2008). Crispr - a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* 6, 181–186. doi:10.1038/nrmicro1793

Sun, H. Q., Jiao, W. B., Campoy, J. A., Krause, K., Goel, M., Folz-Donahue, C, et al. (2022). Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nat. Genet.* 54, 342, doi:10.1038/s41588-022-01015-0

Svitashev, S., Schwartz, C., Lenderts, B., Young, J. K., Cigan, A. M., and Mark CigAn, A. (2016). Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. *Nat. Commun.* 7, 13274. doi:10.1038/ncomms13274

Tabassum, J., Ahmad, S., Hussain, B., Mawia, A. M., Zeb, A., and Ju, L. (2021). Applications and potential of genome-editing systems in rice improvement: Current and future perspectives. *Agronomy* 11, 1359. doi:10.3390/agronomy11071359

Tang, D., Jia, Y., Zhang, J., Li, H., Cheng, L., Wang, P., et al. (2022). Genome evolution and diversity of wild and cultivated potatoes. *Nature* 606, 535–541. doi:10.1038/s41586-022-04822-x

Tang, Y. C., Abdelrahman, M., Li, J. B., Wang, F. J., Ji, Z. Y., Qi, H. X., et al. (2021). CRISPR/Cas9 induces exon skipping that facilitates development of fragrant rice. *Plant Biotechnol. J.* 19, 642–644. doi:10.1111/pbi.13514

Taylor, M. (2018). Routes to genetic gain in potato. *Nat. Plants* 4, 631–632. doi:10.1038/s41477-018-0223-9

Teng, C., Zhang, H., Hammond, R., Huang, K., Meyers, B. C., and Walbot, V. (2020). Dicer-like 5 deficiency confers temperature-sensitive male sterility in maize. *Nat. Commun.* 11, 2912. doi:10.1038/s41467-020-16634-6

Tiwari, J. K., Rawat, S., Luthra, S. K., Zinta, R., Sahu, S., Varshney, S. V., et al. (2021). Genome sequence analysis provides insights on genomic variation and late blight resistance genes in potato somatic hybrid (parents and progeny). *Mol. Biol. Rep.* 48, 623–635. doi:10.1007/s11033-020-06106-x

Toinga-Villafuerte, S., Vales, M. I., Awika, J. M., and Rathore, K. S. (2022). CRISPR/Cas9-Mediated mutagenesis of the granule-bound starch synthase gene in the potato variety yukon gold to obtain amylose-free starch in tubers. *Int. J. Mol. Sci.* 23, 4640. doi:10.3390/ijms23094640

Tuncel, A., Corbin, K. R., Ahn-Jarvis, J., Harris, S., Hawkins, E., Smedley, M. A., et al. (2019). Cas9-mediated mutagenesis of potato starch-branching enzymes generates a range of tuber starch phenotypes. *Plant Biotechnol. J.* 17, 2259–2271. doi:10.1111/pbi.13137

Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., and Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* 11, 636–646. doi:10.1038/nrg2842

Veillet, F., Chauvin, L., Kermarrec, M. P., Sevestre, F., Merrer, M., Terret, Z., et al. (2019). The Solanum tuberosum *GBSSI* gene: A target for assessing gene and base editing in tetraploid potato. *Plant Cell Rep.* 38, 1065–1080. doi:10.1007/s00299-019-02426-w

Veillet, F., Kermarrec, M. P., Chauvin, L., Chauvin, J. E., and Nogué, F. (2020). CRISPR-induced indels and base editing using the *Staphylococcus aureusStaphylococcus aureus* Cas9 in potato. *PLoS One* 15, e0235942. doi:10.1371/journal.pone.0235942

Wang, B. A., Ma, Y. L., Zhang, Z. B., Wu, Z. M., Wu, Y. F., Wang, Q. C., et al. (2011). Potato viruses in China. *Crop Prot.* 30, 1117–1123. doi:10.1016/j.cropro.2011.04.001

Wang, C. G., Wang, G. K., Gao, Y., Lu, G. H., Habben, J. E., Mao, G. F., et al. (2020). A cytokinin-activation enzyme-like gene improves grain yield under various field conditions in rice. *Plant Mol. Biol.* 102, 373–388. doi:10.1007/s11103-019-00952-5

Wang, H., Wu, Y., Zhang, Y., Yang, J., Fan, W., Zhang, H., et al. (2019). CRISPR/Cas9-Based mutagenesis of starch biosynthetic genes in sweet potato (Ipomoea batatas) for the improvement of starch quality. *Int. J. Mol. Sci.* 20, 4702. doi:10.3390/ijms20194702

Wang, H. Y., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., et al. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153, 910–918. doi:10.1016/j.cell.2013.04.025

Wang, S. H., Zhang, S. B., Wang, W. X., Xiong, X. Y., Meng, F. R., and Cui, X. (2015). Efficient targeted mutagenesis in potato by the CRISPR/Cas9 system. *Plant Cell Rep.* 34, 1473–1476. doi:10.1007/s00299-015-1816-7

Wu, H., Liu, Q. S., Shi, H., Xie, J. K., Zhang, Q. J., Ouyang, Z., et al. (2018). Engineering CRISPR/Cpf1 with tRNA promotes genome editing capability in mammalian systems. *Cell. Mol. Life Sci.* 75, 3593–3607. doi:10.1007/s00018-018-2810-3

Wu, Y. X., Liang, D., Wang, Y. H., Bai, M. Z., Tang, W., Bao, S. M., et al. (2013). Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell* 13, 659–662. doi:10.1016/j.stem.2013.10.016

Xu, Y., and Gray, S. M. (2020). Aphids and their transmitted potato viruses: A continuous challenges in potato crops. *J. Integr. Agric.* 19, 367–375. doi:10.1016/s2095-3119(19)62842-x

Yang, Y. Q., and Guo, Y. (2018). Elucidating the molecular mechanisms mediating plant salt-stress responses. *New Phytol.* 217, 523–539. doi:10.1111/nph.14920

Yao, L., Zhang, Y., Liu, C. X., Liu, Y. B., Wang, Y. L., Liang, D. W., et al. (2018). *OsMATL* mutation induces haploid seed formation in indica rice. *Nat. Plants* 4, 530–533. doi:10.1038/s41477-018-0193-y

Ye, M. W., Peng, Z., Tang, D., Yang, Z. M., Li, D. W., Xu, Y. M., et al. (2018). Generation of self-compatible diploid potato by knockout of S-RNase. *Nat. Plants* 4, 651–654. doi:10.1038/s41477-018-0218-6

Zaheer, K., and Akhtar, M. H. (2016). Potato production, usage, and nutrition-a review. *Crit. Rev. Food Sci. Nutr.* 56, 711–721. doi:10.1080/10408398.2012.724479

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas System. *Cell* 163, 759–771. doi:10.1016/j.cell.2015.09.038

Zhan, X. H., Zhang, F. J., Zhong, Z. Y., Chen, R. H., Wang, Y., Chang, L. R., et al. (2019). Generation of virus-resistant potato plants by RNA genome targeting. *Plant Biotechnol. J.* 17, 1814–1822. doi:10.1111/pbi.13102

Zhang, Y., Li, D., Zhang, D. B., Zhao, X. G., Cao, X. M., Dong, L. L., et al. (2018). Analysis of the functions of *TaGW2* homoeologs in wheat grain weight and protein content traits. *Plant J.* 94, 857–866. doi:10.1111/tpj.13903

Zhang, Y., Liang, Z., Zong, Y., Wang, Y. P., Liu, J. X., Chen, K. L., et al. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat. Commun.* 7, 12617. doi:10.1038/ncomms12617

Zhang, Y., Sun, F., Fettke, J., Schöttler, M. A., Ramsden, L., Fernie, A. R., et al. (2014). Heterologous expression of AtPAP2 in transgenic potato influences carbon metabolism and tuber development. *FEBS Lett.* 588, 3726–3731. doi:10.1016/j.febslet.2014.08.019

Zhao, X., Jayarathna, S., Turesson, H., Fält, A. S., Nestor, G., Gonzalez, M. N., et al. (2021). Amylose starch with no detectable branching developed through DNA-free CRISPR-Cas9 mediated mutagenesis of two starch branching enzymes in potato. *Sci. Rep.* 11, 4311. doi:10.1038/s41598-021-83462-z

Zheng, Z. Z., Ye, G. J., Zhou, Y., Pu, X. Q., Su, W., and Wang, J. (2021). Editing sterol side chain reductase 2 gene (*StSSR2*) via CRISPR/Cas9 reduces the total steroidal glycoalkaloids in potato. *All Life* 14, 401–413. doi:10.1080/26895293.2021.1925358

Zhou, J. P., Xin, X. H., He, Y., Chen, H. Q., Li, Q., Tang, X., et al. (2019). Multiplex QTL editing of grain-related genes improves yield in elite rice varieties. *Plant Cell Rep.* 38, 475–485. doi:10.1007/s00299-018-2340-3

Zhou, Q., Tang, D., Huang, W., Yang, Z. M., Zhang, Y., Hamilton, J. P., et al. (2020). Haplotype-resolved genome analyses of a heterozygous diploid potato. *Nat. Genet.* 52, 1018, doi:10.1038/s41588-020-0699-x

Zhou, X., Zha, M., Huang, J., Li, L., Imran, M., and Zhang, C. (2017). StMYB44 negatively regulates phosphate transport by suppressing expression of PHOSPHATE1 in potato. *J. Exp. Bot.* 68, 1265–1281. doi:10.1093/jxb/erx026

Zhu, H. C., Li, C., and Gao, C. X. (2020). Applications of CRISPR-Cas in agriculture and plant biotechnology. *Nat. Rev. Mol. Cell Biol.* 21, 661–677. doi:10.1038/s41580-020-00288-9

Zong, Y., Song, Q., Li, C., Jin, S., Zhang, D., Wang, Y., et al. (2018). Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat. Biotechnol.* 36, 950–953. doi:10.1038/nbt.4261

Zong, Y., Wang, Y. P., Li, C., Zhang, R., Chen, K. L., Ran, Y. D., et al. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* 35, 438, doi:10.1038/nbt.3811

| Frontiers in Genetics

TYPE Original Research
PUBLISHED 24 February 2023
DOI 10.3389/fgene.2023.1046624

Check for updates

OPEN ACCESS

EDITED BY
Sukhwinder Singh,
The International Maize and Wheat Improvement Center (CIMMYT), India

REVIEWED BY
Ramalingam Jegadeesan,
Tamil Nadu Agricultural University, India
M. Miftahudin,
IPB University, Indonesia

*CORRESPONDENCE
Hari Krishna,
harikrishna.agri@gmail.com
Kumble Vinod Prabhu,
kvinodprabhu@rediffmail.com

SPECIALTY SECTION
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

RECEIVED 16 September 2022
ACCEPTED 06 February 2023
PUBLISHED 24 February 2023

CITATION
Kumar PKC, Bellundagi A, Krishna H, Mallikarjuna MG, Thimmappa RK, Rai N, Shashikumara P, Sinha N, Jain N, Singh PK, Singh GP and Prabhu KV (2023), Development of bread wheat (*Triticum aestivum* L) variety HD3411 following marker-assisted backcross breeding for drought tolerance .
Front. Genet. 14:1046624.
doi: 10.3389/fgene.2023.1046624

COPYRIGHT
© 2023 Kumar, Bellundagi, Krishna, Mallikarjuna, Thimmappa, Rai, Shashikumara, Sinha, Jain, Singh, Singh and Prabhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Development of bread wheat (*Triticum aestivum* L) variety HD3411 following marker-assisted backcross breeding for drought tolerance

Prashanth K. C. Kumar[1], Amasiddha Bellundagi[2], Hari Krishna[1]*, Mallana Gowdra Mallikarjuna[1], Ramya K. Thimmappa[3], Neha Rai[1], P. Shashikumara[4], Nivedita Sinha[1], Neelu Jain[1], Pradeep K. Singh[1], Gyanendra Pratap Singh[5] and Kumble Vinod Prabhu[6]*

[1]ICAR-Indian Agricultural Research Institute, New Delhi, India, [2]ICAR-Indian Institute Millets Research, Hyderabad, India, [3]ICAR-Indian Institute of Oilseed Research, Hyderabad, India, [4]ICAR-Indian Grassland and Fodder Research Institute, Jhansi, India, [5]ICAR-Indian wheat and Barley Research Institute, Karnal, India, [6]PPV & FR Authority NAAS complex, New Delhi, India

Marker-assisted backcross breeding enables selective insertion of targeted traits into the genome to improve yield, quality, and stress resistance in wheat. In the current investigation, we transferred four drought tolerance quantitative trait loci (QTLs) controlling traits, *viz* canopy temperature, normalized difference vegetative index, chlorophyll content, and grain yield from the drought-tolerant donor line, C306, into a popular high-yielding, drought-sensitive variety, HD2733. Marker-assisted selection coupled with stringent phenotypic screening was used to advance each generation. This study resulted in 23 improved lines carrying combinations of four drought tolerance QTLs with a range of 85.35%–95.79% background recovery. The backcross-derived lines gave a higher yield under moisture-deficit stress conditions compared with the recipient parent. They also showed higher phenotypic mean values for physiological traits and stability characteristics of HD2733. A promising genotype, HD3411, derived from this cross was identified for release after national multi-location coordinating trials under the All India Coordinated Wheat Improvement Project. Our study is a prime example of the advantages of precision breeding using integrating markers and phenotypic selection to develop new cultivars with desirable traits like drought tolerance.

KEYWORDS

drought tolerance, wheat, MAS, foreground selection, background selection

Introduction

Wheat is a golden winter cereal grain and a major contributor to food and nutritional security, but the increase in drought due to climate can severely limit wheat production (Reynolds et al., 2001). Out of 12 distinct mega environments for wheat cultivation classified by the CIMMYT, only three are irrigated environments (Rajaram et al., 1995). The Indian sub-continent, which comes under the fourth mega environment, has a large area under wheat cultivation but only with residual soil moisture from the monsoon rains (Richards

et al., 2001). In the last few years, global warming has turned researchers' attention toward the drought tolerance of crop plants (Kang et al., 2009). It has been predicted that every 1°C increase in global warming leads to a decrease of 4.0%–6.5% in wheat production (Trnka et al., 2019). In the present context, threats of climate change and erratic behavior of climatic factors have raised awareness of the critical need to devise new ways to overcome drought stress.

Insufficient moisture affects wheat growth at phenological, reproductive, and grain-filling phases (Pradhan et al., 2012). Drought during the seedling stage will affect all the following growth stages and ultimately reduce grain yield (Sallam et al., 2022), and a prolonged mild drought during the flowering and grain-filling stages can cause 58%–92% reduction in grain yield (Matiu et al., 2017). Accurate field phenotyping for moisture stress adaption has been a critical issue in breeding for moisture stress tolerance (Mir et al., 2012; Tuberosa, 2012). At present, high-throughput phenotyping platforms are readily available for measuring phenotypic data. The normalized difference vegetative index (NDVI) and SPAD chlorophyll meter readings are related to plant health; higher values under moisture stress conditions are associated with greater vegetation and higher chlorophyll content. Such tools have been efficiently utilized in breeding for moisture stress tolerance in rice, maize, and wheat (Subash et al., 2011; Lu et al., 2011). The productivity of wheat under drought conditions is strongly associated with various physiological features such as leaf characteristics (low canopy temperature, staying green, chlorophyll content, early ground cover, NDVI, etc.), water use efficiency, and yield component traits (thousand-grain weight, spike length, grain number per spike, etc.). Additionally, a positive correlation between NDVI and chlorophyll content and grain yield under moisture stress has been reported in wheat (Harikrishna et al., 2016). Integration of physiological traits with grain yield component traits is necessary to improve wheat yield in moisture-stressed environments (Araus et al., 2008; Reynolds and Langridge, 2016). However, combining multicomponent drought tolerance-associated traits into a single cultivar using traditional breeding strategies is difficult in practice. The availability of markers for the genes underlying drought tolerance traits and marker-assisted backcross breeding (MABB) schemes can overcome the limitations associated with conventional introgression breeding.

Wheat breeding is rapidly changing, owing to advances in wheat genomics and molecular biology. The application of genomics technologies aims to realize faster and more efficient genetic gains of desirable traits. The discovery of RFLP marker technology led to the application of molecular markers in plant breeding (Tanksley et al., 1989), which are key tools for breeders in selecting desirable lines from germplasm and segregating generations (Raghavendra et al., 2020). The locking of genomic regions with the help of markers transformed the conventional backcross breeding into MABB, which is considered highly efficient in terms of time and cost and precision in selecting target traits. A plethora of QTLs and meta-QTLs for physiological and grain yield component traits have been reported by various studies (Quarrie et al., 2006; Kirigwi et al., 2007; Olivares-Villegas et al., 2007; Pinto et al., 2010; Kumar et al., 2012; Shukla et al., 2015; Gupta et al., 2017; Sunil et al., 2020; Puttamadanayaka et al., 2020; Khaled et al., 2022), which have enriched our knowledge of the genetic architecture of drought tolerance in wheat. The markers that have been validated, the stable QTLs, and meta-QTLs with more than 10% of explained phenotypic variance for drought tolerance traits have high practical utility in MABB to reconstruct drought-tolerant versions of important wheat cultivars. MABB is an accelerated approach for QTL introgression through marker-assisted foreground and background selection (Hospital and Charcosset, 1997). Marker-assisted foreground selection helps identify the gene of interest without extensive phenotypic assays (Tanksley, 1983; Melchinger, 1990), and marker-assisted background selection significantly expedites the rate of recovery of recurrent parent genomes with one or two backcrosses (Young and Tanksley, 1989; Visscher et al., 1996). MABB also produces a set of recombinant backcross inbred lines, which can be further evaluated to select the best recombinants from both parents (Manu et al., 2020; Shashikumara et al., 2020). So far, MABB has been exploited in all major crops: in rice, for bacterial blight resistance (Sundaram et al., 2008); in wheat, for powdery mildew (Zhou et al., 2003); and quality traits including high molecular weight (HMW) glutenins (De Bustos et al., 2001) and preharvest sprouting (Kumar et al., 2010). There are only a few reports on MABB application for drought tolerance in wheat, however (Rai et al., 2018; Gautam et al., 2020; Merchuk-Ovnat et al., 2016).

Keeping those mentioned previously in mind, we executed this study of the transfer of QTLs contributing to drought tolerance from the donor line C306 into the recipient variety HD2733 using the MABB approach. We successfully transferred QTLs for NDVI, chlorophyll content, low canopy temperature, and grain yield into the elite Indian wheat variety, HD2733. The improved lines with targeted QTLs were drought-resilient with higher grain yields under moisture stress conditions.

Materials and methods

The experiment material consisted of the HD2733 cultivar, one of India's most popular wheat varieties for the northeastern plain zone (NEPZ). However, HD2733 is susceptible to drought stress, resulting in substantial yield losses under field conditions of limited irrigation. The donor parent for the drought tolerance QTL was C306. This cultivar is suitable for drought-prone and rain-fed conditions and has been widely adopted in the NEPZ and NWPZ (northwestern plain zone).

Parental lines and targeted QTL regions for transfer

We targeted four QTLs associated with NDVI, chlorophyll content, canopy temperature, and grain yield for marker-assisted introgression. The details of the targeted QTL and the flanking markers are given in Table 1.

DNA isolation, PCR conditions, and parental polymorphism survey

Total genomic DNA was isolated by a micro-extraction protocol (Prabhu et al., 1998). A polymerase chain reaction (PCR) was performed in a 10 μL total reaction volume containing 2–3 μL (60–70 ng/μL) DNA,

TABLE 1 Molecular markers used for introgression into HD2733 through foreground selection.

QTL/trait	Primer	Chromosome	Sequence of the marker	Position (cM)	Annealing temperature (°C)	Reported R^2%	Observed R^2%	References
NDVI	*Xgdm* 93	2A	F: AAAAGCTGC TGGAGCATACA	170	61	20	18.2	Oliver et al., 2007
			R: GGAGCATGG CTACATCCTTC					
Canopy temperature (CT)	*Xbarc*68-*Xbarc*101	3B	F: CGATGCCAA CACACTGAGGT	33	55	35–40	28.3	Kumar et al., 2012
			R: AGCCGCATG AAGAGATAGGTA GAGA					
			F: GCTCCTCTC ACGATCACG CAAAG					
			R: GCGAGTCGA TCACACTATGAG CCAATC					
Yield and chlorophyll content	*Xgwm304*	5A	F: AGGAAACAG AAATATCGCGG	59	60	15	22.6	Pinto *et al.* (2010)
			R: AGGACTGTG GGGAATGAATG					
Chlorophyll content (CHL)	*Xgwm301*	2D	F: GAGGAGTAA GACACATGCCC	107	60	11.2	11.9	Pinto *et al.* (2010)
			R: GTGGCTGGA GATTCAGGTTC					

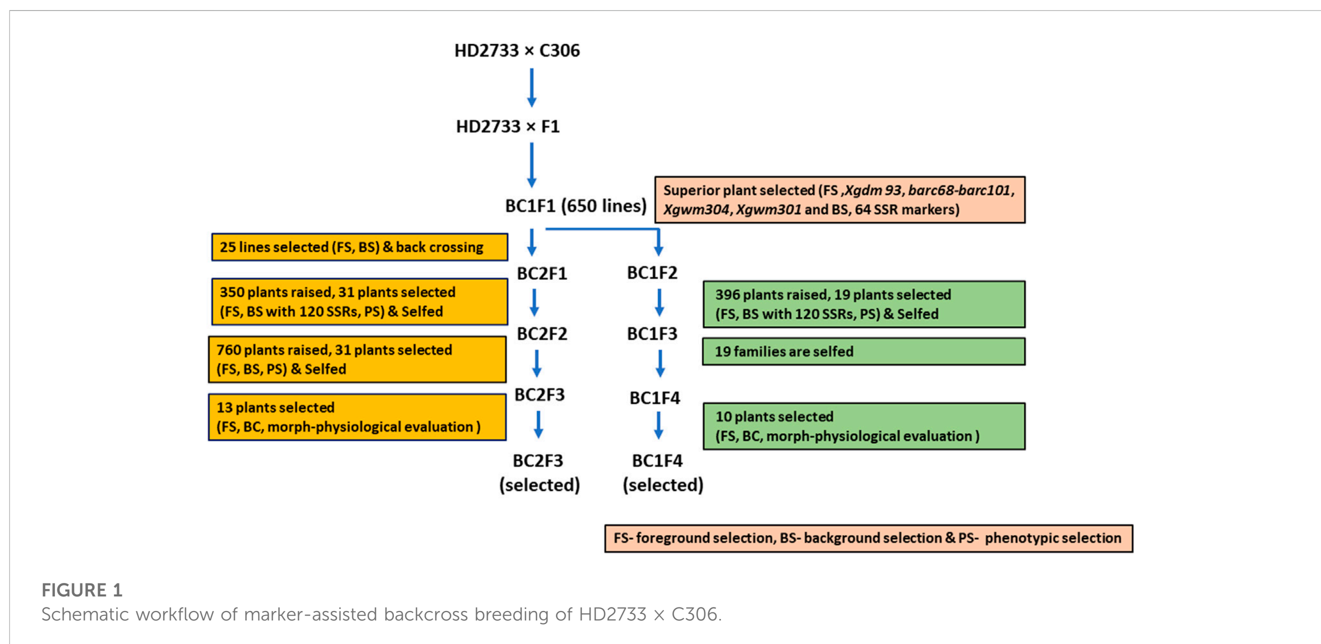


FIGURE 1
Schematic workflow of marker-assisted backcross breeding of HD2733 × C306.

2.0 µL 109 buffer with 25 mM $MgCl_2$, 0.5 µL dNTPs (10 mM) (Bangalore Genei, Bangalore, Karnataka, India), 1.0 µL each forward and reverse SSR primers (20 mM) (Sigma Inc., St. Louis, MO, United States), 0.3 µL Taq polymerase (3 U/µL) (Bangalore Genei, Bangalore, Karnataka, India), and 5.2 µL distilled water (sterile). Amplification of the template DNA was performed according to the annealing conditions for the wmc, gwm, barc, cfa, and cfd series of SSR markers used (Roder et al., 1998; Pestsova et al., 2000; Gupta et al., 2002; Somers et al., 2004; Quarrie et al., 2005; Kumar et al., 2012). Amplified products were resolved on a 3.2% agarose gel (MetaPhor, Lonza,

Rockland, ME, United States), along with a DNA ladder size standard (MBI, Fermentas), stained with 0.5 μg/mL ethidium bromide (Amresco, Solon, OH, United States), and documented with a gel documentation system (Bio-Rad, Hercules, CA, United States). The donor parent, C306, and recurrent parent, HD2733, were screened with 700 SSR markers, including markers associated with targeted traits.

Marker-assisted backcross breeding (MABB)

Initially, a cross was made between HD2733 and C306 to transfer drought stress tolerance QTLs into HD2733. The MABB procedure followed here is represented in Figure 1. The true F_1s were identified using foreground SSR markers and backcrossed to the recurrent parent. The BC_1F_1s were subjected to foreground and initial background selection with a set of 64 polymorphic markers. Twenty-five lines positive for target QTLs with maximum recurrent parent genome (RPG) recovery coupled with phenotypic similarity to the recipient parent were selected. The 21 selected lines were backcrossed to the recurrent parent and selfed to generate BC_2F_1 and BC_1F_2 seeds. BC_2F_1 and BC_1F_2 lines were repeated for the MABB process involving foreground and background selection with 120 polymorphic SSR markers.

The polymorphic SSR markers were used to construct a schematic map illustrating the genomic contributions of donor and recurrent parents with Graphical GenoType (GGT) v2.045 software to identify backcross-derived lines possessing the maximum recurrent parent genome. The positive foreground-selected plants genotyped for polymorphic markers at each backcross/selfing generation and recurrent parent genome recovery (G) were estimated using the following formula: $G = [(X + ½Y) \times 100]/N$; here, N is the total number of parental polymorphic markers screened, X is the number of markers showing homozygosity for recurrent parental alleles, and Y is the number of markers showing heterozygosity for parental alleles. Based on the recovery of the recurrent parental genome and the presence of targeted donor genomic regions, 50 lines were selected from BC_2F_1 and BC_1F_2 plants and advanced through selfing.

Thirteen plants were selected from advanced BC_2F_2 lines based on maximum recovery for RPG through background and foreground selection and visible phenotypic similarity with the recurrent parent strain, HD2733, while BC_1F_3 lines were selfed and advanced to BC_1F_4 generations. A total of 10 BC_1F_4 plants were again selected based on foreground selection and maximum background recovery of the recurrent parental genome. The selected 13 BC_2F_3 and 10 BC_1F_4 plants were evaluated for morphological and physiological traits and yield performance and further advanced through selfing for evaluation under a national testing trial.

Evaluation of morpho-physiological and yield component characteristics of the backcross-derived progenies for drought tolerance

The BC_1F_4 and BC_2F_3 families were raised under restricted irrigation conditions (two irrigations were carried out at 21 and 40 days after sowing), following an augmented design protocol, where parents were replicated as checks. The observations on introgressed progeny lines in the field for various traits contributing to drought tolerance and yield parameters were recorded as per CIMMYT guidelines published in "Physiological Breeding II: A field guide to wheat phenotyping" (Pask et al., 2012). The data were recorded for various morphological traits *viz*., 50% days to heading (DH), days to anthesis (DA), days to maturity (DM), plant height (PH), number of tillers (NT), spike length (SL), peduncle length (PL), 1,000 kernel weight (TKW), biomass, harvest index (HI), yield per plot (Y/P), and physiological traits like chlorophyll content, canopy temperature, and the normalized difference vegetation index were scored at three different stages of wheat development: the vegetative stage (late boot stage, Z49), the grain filling stage (early milk stage, Z73), and the grain maturity stage (late milk stage, Z85) according to Zadok's scale (Zadoks et al., 1974).

Results

Marker–trait association of targeted drought tolerance QTLs

Foreground selection using the QTL-linked markers, *Xbarc68-Xbarc101, Xgdm93, Xgwm165,* and *Xgwm301*, associated with moisture stress tolerance were carried out on the MABB population. Initially, targeted QTLs were validated in the segregating BC_1F_2 population under restricted irrigation. Individual lines were phenotyped for associated traits and genotyped for targeted donor parent alleles. Single marker analysis was performed to arrive at the presence of QTL. It was observed that *Xgdm93*, linked to *qNDVI* located on the 2A chromosome, showed a phenotypic variance of 18.2% (R^2 = 0.182). *QChl.ksu-3B* and *QLt.ksu-3B*, flanked by markers *Xbarc68-Xbarc101*, showed phenotypic variance of 28.3% (R^2 = 0.283) for canopy temperature. The microsatellite marker, *Xgmw304* associated with QTL for grain yield and chlorophyll content on chromosome 5A, showed 22.6% phenotypic variation (R^2 = 0.226). Another QTL associated with chlorophyll content co-segregated with *Xgwm301* on chromosome 2D and depicted 11.9% phenotypic variation (R^2 = 0.119) in the segregating BC_1F_2 population (Table 1).

Marker-assisted transfer of drought tolerance QTLs into HD2733

The crosses were made between HD2733 and C306 to improve drought tolerance in the recurrent parent HD2733, during *rabi* 2011. We simultaneously followed two approaches for the introgression of targeted QTLs: (1) where BC_1F_1 lines were allowed for the second backcross and advanced to generate BC_2F_3 and (2) where BC_1F_1 lines were selfed and advanced to the BC_1F_4 generation. The F_1s obtained were confirmed by screening *XBarc68-Xbarc101, Xgdm93, Xgwm301, and Xgwm304* microsatellite markers polymorphic between HD2733 and C306. The true F_1s were backcrossed with the recipient parent to produce 650 BC_1F_1s, which were confirmed for the presence of the QTL-linked markers, *Xgdm93, Xbarc68-Xbarc101, Xgwm304,* and *Xgwm301* (i.e., foreground selection). The

TABLE 2 Range of genome recovery (%) in different backcross and selfed generations.

S. No	Generation	No. of selected plants	Recurrent parent genome (%)
1	BC_1F_1	25	74.5–75.4
2	BC_2F_1	31	86.60–92.34
3	BC_2F_2	31	85.73–94.87
4	BC_2F_3	13	92.06–95.79
5	BC_1F_2	19	78.84–81.35
6	BC_1F_4	10	85.35–88.34

foreground selection resulted in the identification of 79 lines (20 lines with *Xbarc68-Xbarc101+Xgdm93* and *Xgwm301+Xgwm304*; 18 lines with *Xgwm304+Xgdm93*; 17 lines with *Xbarc68-Xbarc101* + *Xgwm304*; 12 plants with *Xgwm304+Xgwm301*, and 12 lines with *Xgdm93+Xgwm301*) for background selection. The selected lines were examined for phenotypic differences for targeted traits (QTL expression) and phenotypic similarity with the recipient parent. These lines were also subjected to background selection with a set of 64 polymorphic SSR markers, and 25 lines possessing the maximum recovery percentage of the recurrent parent genome (74.5%–75.4%, Table 2) were identified. The selected lines were selfed and crossed to the recipient parent to generate a population of 396 BC_1F_2s and 350 BC_2F_1s plants.

For the 350 BC_2F_1s plants, the foreground protocol as explained previously was repeated (11 lines positive for *Xbarc68-Xbarc101+Xgdm93+Xgwm301+Xgwm304*, 18 lines positive for *Xgwm304+Xgdm93*, 17 lines positive for *Xbarc68-Xbarc101+Xgwm304*, 18 lines positive for *Xgwm304+Xgwm301*, and 18 lines positive for *Xgdm93+Xgwm301*), and a total of 82 lines with different combinations of introgressed QTLs were selected for background screening. Applying background selection using 120 polymorphic markers and considering the phenotypic similarity of lines with the recurrent parent, 31 BC_2F_1 lines were selected with an RPG recovery of 86.60%–92.34% (Table 2). In BC_2F_1, among a total of 31 selected plants, six plants possessed four QTLs (*qCT* + *qNDVI* + *qCHL* + *qYield*); eight plants carried two QTLs (*qYield* + *qNDVI*); six plants carried two QTLs (*qYield* + *qCT*); four plants carried two QTLs (*qNDVI* + *qCHL*); and seven plants represented two QTLs (*qYield* + *qCHL*). The selected plants were advanced to generate 760 segregating progenies in the BC_2F_2 generation. Among the 760 progenies, 31 lines homozygous for two or more donor parent marker alleles with a maximum RPG of 85.73%–94.87% were advanced to the BC_2F_3 generation. After reconfirmation for targeted QTLs and lines ranging from 90.90% to 95.79%, the RPG recovery was evaluated under drought stress conditions, where the crop was irrigated only once at a critical stage, i.e., 21 days after sowing (Table 2). Based on the performance of lines under drought stress for different morpho-physiological traits, 13 BC_2F_3 lines with RPG ranging from 92.06% to 95.79% were selected (Table 3). The graphical genotype of selected 13 BC_2F_3 lines shows that the distribution of donor parents' genome segments was mostly restricted to targeted regions with an average of 88.27% of the recurrent parent genome (Figure 2).

Simultaneously, 396 BC_1F_2 lines were screened under foreground selection and targeted traits, including resemblance with the recurrent parent. A total of 37 BC_1F_2 plants, six with three QTLs (*qCT* + *qNDVI* + *qCHL*), eight with two QTLs (*qCT* + *qNDVI*), eight with two QTLs (*qCT* + *qCHL*), eight with one QTL (*qCHL*), and seven with one QTL (*qYield*) were selected for background screening using 120 polymorphic SSR markers. Of these, 19 BC_1F_2 plants with RPGs ranging from 78.84% to 81.35% were forwarded to the BC_1F_3 generation, and these 19 BC_1F_3 families were advanced to produce BC_1F_4 families. After confirming the presence of targeted QTLs with 85.35%–88.34% background recovery, the selected lines were examined under drought stress conditions (one irrigation at 21 days after sowing). Based on different morpho-physiological traits associated with the lines' performance, 10 BC_1F_4 lines were selected for further evaluation. Graphical genotyping of the 10 selected BC_1F_4 lines revealed an average genome recovery of 73.50% from the recipient parent (Figure 3). Among all the transferred lines, the best included line in BC_2F_3 (HD2733-33-59-141, 4QTLs) with an RPG recovery of 94.58% and the one in BC_1F_4 (HD2733-20-1, 3QTLs) with an RPG recovery of 86.67% (Table 3).

Evaluation of the improved lines of BC_2F_3 and BC_1F_4 for morpho-physiological traits

The improved lines containing QTLs in combination (four QTLs, two QTLs, and one QTL), derived from crossing HD2733 × C306, were evaluated in the augmented design along with parents as checks. The 13 BC_2F_3 and 10 BC_1F_4 lines selected were evaluated under drought stress for 26 different morpho-physiological traits and distinctiveness, uniformity, and stability (DUS) traits. DUS characterization was based on the plant's ear shape and density, glume pubescence, growth attitude, foliage color, peduncle attitude, sheath-blade waxiness, and ear peduncle waxiness. We observed significantly higher phenotypic performance than HD2733 and a clear difference between lines for the trait-targeted MABB lines introgressed with respective QTL combinations. The details of the differences in traits are presented in Table 4 and Table 5. In particular, the three improved lines were pyramided with different combination of QTLs (*qCT* + *qCHL*, *qCHL* + *qYield*, and *qCT* + *qNDVI* + *qCHL* + *qYield*) in BC_2F_3, and BC_1F_4 showed a lower canopy temperature ranging from 18.3°C to 19.27°C compared to the recurrent parent, HD2733, with a canopy temperature of 21.0°C at the late boot stage. With respect to NDVI values, the improved lines were on par with the donor parent, C306, at the LB, EM, and LM stages. The average chlorophyll content of HD2733 at the LM stage under drought conditions was 37.40. We observed an improvement of up to

TABLE 3 Percentage contribution of recurrent and donor parent alleles in selected MABB lines.

S. No.	Selected progeny	HD2733 allele %	Hetero. allele %	C-306 allele %	Total RPG %
BC_2F_3 lines					
1	HD2733-68-70-359	84.17	15.83	7.9	92.08
2	HD2733-33-59-141	89.17	10.83	5.4	94.58
3	HD2733-33-68-298	87.50	12.50	6.3	93.46
4	HD2733-571-296-642	89.17	10.83	5.4	94.39
5	HD2733-571-296-645	91.67	08.33	4.2	95.79
6	HD2733-33-64-203	89.17	10.83	5.4	94.86
7	HD2733-33-64-215	89.17	10.83	5.4	94.39
8	HD2733-365-262-622	88.33	11.67	5.8	94.86
9	HD2733-87-251-481	84.17	15.83	7.9	92.06
10	HD2733-571-296-657	89.17	10.83	5.4	94.39
11	HD2733-87-251-475	88.33	11.67	5.8	94.39
12	HD2733-361-37-97	88.33	11.67	5.8	93.46
13	HD2733-18-218-446	89.17	10.83	5.4	94.39
BC_1F_4 lines					
1	HD2733-20-1	73.33	26.7	13.3	86.67
2	HD2733-34-4	74.17	25.8	12.9	87.08
3	HD2733-34-5	75.00	25.0	12.5	87.50
4	HD2733-52-8	76.67	23.3	11.7	88.33
5	HD2733-56-9	72.50	27.5	13.8	86.25
6	HD2733-59-10	71.67	28.3	14.2	85.83
7	HD2733-62-11	70.83	29.2	14.6	85.42
8	HD2733-67-15	73.33	26.7	13.3	86.67
9	HD2733-70-17	75.83	24.2	12.2	87.92
10	HD2733-17-18	71.67	28.3	14.2	85.83

Note: C-306 allele % does not include the introgressed QTL regions of 2A, 2D, 3B, and 5A chromosomes.

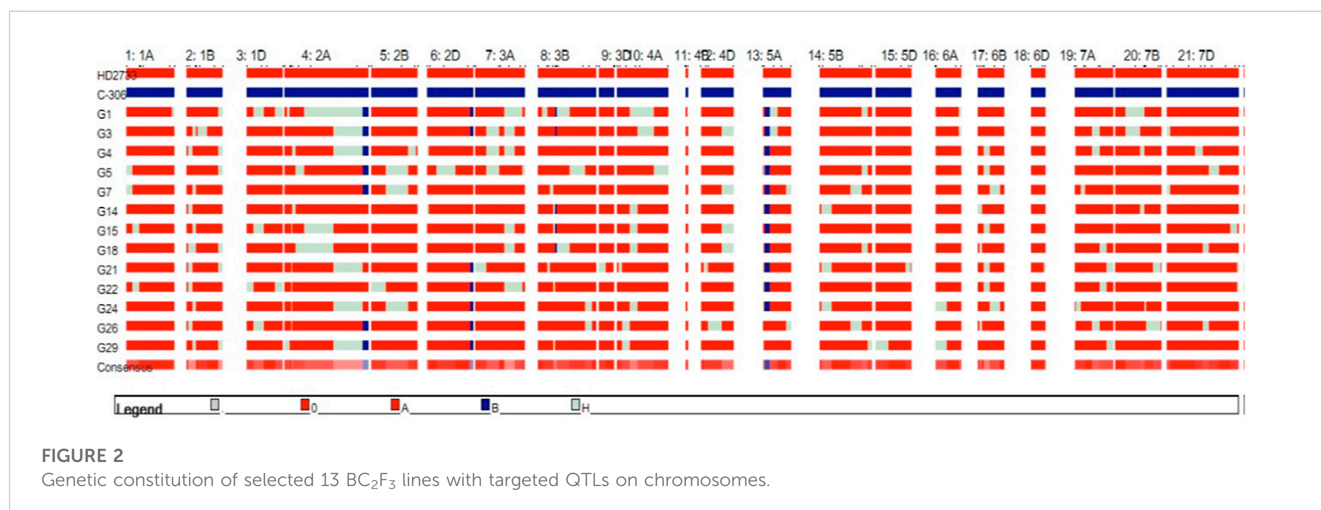


FIGURE 2
Genetic constitution of selected 13 BC_2F_3 lines with targeted QTLs on chromosomes.

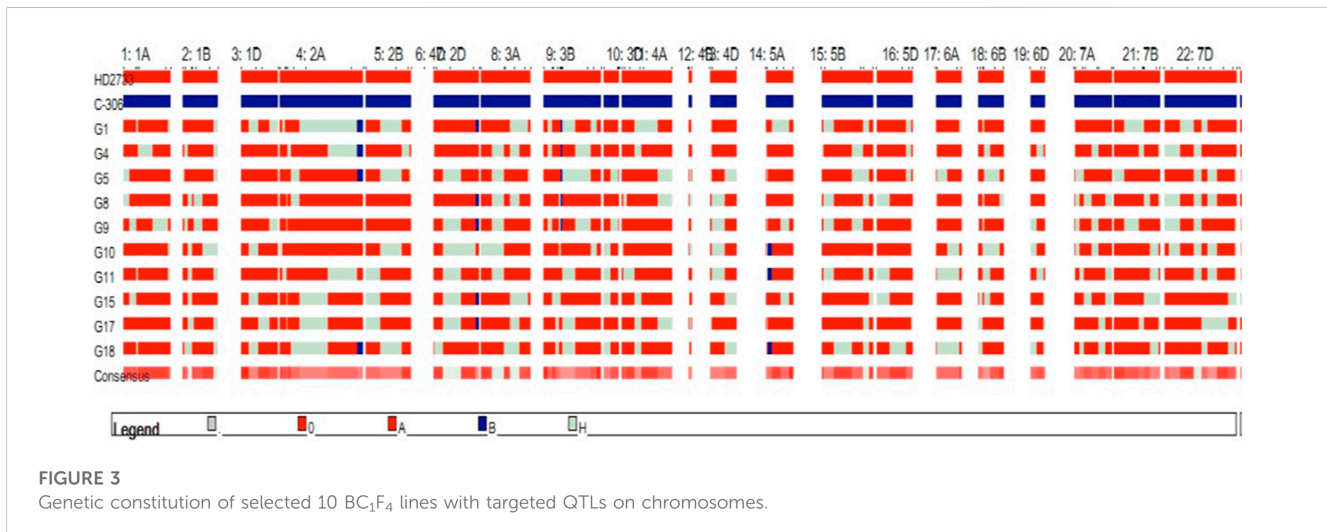


FIGURE 3
Genetic constitution of selected 10 BC_1F_4 lines with targeted QTLs on chromosomes.

TABLE 4 Mean performance of parents and selected MABB-derived lines with combinations of introgressed QTLs for phenological and agronomic traits.

Trait	HD2733 (rainfed)	C306 (rainfed)	BC_2F_3 lines					BC_1F_4 lines					
			qCT + qNDVI + qCHL + qYield	*qNDVI + qYield*	*qCT + qYield*	*qCHL + qYield*	*qCHL + qNDVI*	*qCT + qCHL + qNDVI*	*qCT + qNDVI*	*qCT + qCHL*	*qYield*	*qCHL*	*qYield + qNDVI*
DFL	80	71	80	78	78	79	81	73.67	74	76.33	75.75	75	76.5
DH	87	85	89	88	87	88	88	86.67	86	86.67	87.5	86.25	87
DA	90	88	92	91	91	91	91	88.33	88.67	89.67	91	88.75	91.5
DM	122	126	126	126	126	128	125	120.67	121	125.33	123.75	123.75	122.5
PH	78	97	78.3	80.1	77.1	83.2	77.1	77.53	71.93	78.83	77.88	84.63	90.75
SL	9.5	9.2	11.6	11.2	10.7	11.2	9.5	11.17	10.77	11.17	10.25	10.55	13
PL	54.5	44	39.7	40.4	37.6	44.8	39.6	37.3	39.27	36.83	38.5	38.13	39.15
SNPS	18.2	15.8	19.7	18.5	18.8	19.6	19	18.93	18.87	18.27	18.75	18.55	18
TNP	12.4	9.8	14.7	13.6	14.8	13.2	11.3	10	9.27	12.07	10.7	10.25	8.9
KNPS	200	177.00	221.70	213.70	225.90	244.8	221.1	203.33	236.67	196.67	222.25	238.5	210
TKW	46.9	45.45	45.9	45.5	46.4	42.3	44.7	38.57	38.39	43.79	42.31	46.77	42.98
HI	36.84	38.11	37.9	35.4	40.5	46.3	33.8	23.81	27.42	34.32	39.63	35.3	26.66
GY	53.68	60.97	79.3	78.1	80.1	81.4	65.3	58.46	55.37	69.58	64.48	74.03	54.95

DFL, days to flag leaf emergence; DH, days to heading; DA, days to anthesis; DM, days to maturity; PH, plant height; SL, spike length; PL, peduncle length; SNPS, spikelet number per spike; TNP, tillers per plant; KNPS, kernel number per five spikes; TKW, thousand kernel weight; HI, harvest index; GY, grain yield per five plants (grams); qCT, QTL related to canopy temperature; qNDVI, QTL related to the normalized difference vegetation index; qCHL, QTL related to chlorophyll content; qYield, QTL related to yield

50.84 in selected BC_1F_4 lines with introgression of *qYield* QTL. Overall, the introgressed lines have shown significant improvement over the recurrent parents for lower canopy temperature, NDVI, stomatal conductance, and chlorophyll content (Table 5). The mean grain yield values were significantly higher (54.9-81.4 g) in BC_2F_3 and (55.4-74.0 g) in BC_1F_4 lines compared to the recipient parent HD2733with an average grain yield of 53.68 g (Table 4). Overall, based on superior phenotyping performance, DUS scoring, and a high percentage of background recovery, a total of 13 progeny lines in BC_2F_3 and 10 progeny lines in BC_1F_4 were selected.

Evaluation of HD3411 under the All India Coordinated Research Project (AICRP) on wheat and barley

Among the superior lines selected from BC_1F_4, one line, HD3411 with three major QTLs, has been submitted for a coordinated multi-location varietal trial under the All India Coordinated Research Project (AICRP) on wheat and barley. HD3411 performed better than its recurrent parent HD2733 and reported yield superiority for 2 years over check varieties DBW 39,

TABLE 5 Mean performance of parents and selected MABB-derived lines with combinations of introgressed QTLs for physiological traits.

Trait	Stage	HD2733 (rainfed)	C306 (rainfed)	BC_2F_3 lines					BC_1F_4 lines						
				qCT + qNDVI + qCHL + qYield	*qNDVI + qYield*	*qCT + qYield*	*qCHL + qYield*	*qCHL + qNDVI*	*qCT + qCHL + qNDVI*	*qCT + qNDVI*	*qCT + qCHL*	*qYield*	*qCHL*	*qYield + qNDVI*	*CD @ 5%*
CT	LB	21.0	19.1	19.27	19.67	19.47	19.26	21.47	20.23	19.7	18.3	20.78	19.88	21.25	0.452
	EM	22.4	24.3	21.03	21.23	21.51	21.12	21.74	21	21.28	19.4	20.8	20.64	21.23	0.409
	LM	30.6	34.1	31.60	30.12	29.91	30.10	30.35	31.02	29.15	29.1	29.85	29.28	30.98	0.434
NDVI	LB	0.78	0.79	0.76	0.71	0.73	0.74	0.70	0.79	0.8	0.79	0.77	0.78	0.81	0.038
	EM	0.70	0.71	0.68	0.70	0.70	0.70	0.67	0.71	0.71	0.74	0.73	0.72	0.69	0.019
	LM	0.40	0.54	0.37	0.48	0.46	0.50	0.43	0.44	0.51	0.55	0.53	0.51	0.44	0.016
%GC	Seedling	46.16	47.68	25.68	18.07	19.16	14.56	16.47	20.01	20.2	20.52	21.37	19.5	17.95	1.012
Stomatal conductance	LM	337.4	359.2	477.53	377.63	462.29	407.56	387.00	438.4	584.8	416.6	371.23	445.25	474.2	8.521
	EM	500.7	364.3	478.89	504.76	481.48	516.07	470.10	508.89	583.56	406.33	335.25	545.58	481.67	20.24
Chlorophyll content	LB	41.42	47.12	49.72	49.65	48.84	49.74	49.25	50.94	45.38	47.76	48.06	49.32	49.3	0.314
	EM	50.40	50.54	52.76	51.89	53.09	52.49	51.99	53.43	50.98	50.24	52.9	54.36	53.38	0.521
	LM	37.40	42.08	28.97	39.61	46.38	43.91	37.39	40.23	47.96	38.86	50.84	47.89	41.14	1.112

CT, canopy temperature; NDVI, normalized difference vegetation index, %GC, percent ground cover; LB, late boot; EM, early milk; LM, late milk.

TABLE 6 Summarized yield data (q/ha) of All India wheat-coordinated varietal trials of HD3411.

	Year of testing	No. of trials	Proposed variety (HD 3411)	Check variety					Recurrent parent	C.D @ 5%
				HD3249	DBW187	DBW39	HD 2967	HD 3086	HD 2733	
Mean yield q/ha	2020–21	14	46.3	46.4	45.5	44.5	46.8	42.6	46.1	1.0
	2021–22	5	48.00	48.0	50.2	0.0	48.5	47.8	45.9	1.4
	Weighted Mean		46.75	46.8	46.74	44.5	47.2	44.0	46.0	
% Increase over the check cultivars	2020–21	14	-	−0.22	1.73	3.89	−1.08	7.99	0.43	-
	2021–22	5	-	0.00	−4.58	-	−1.04	0.42	4.38	-
	Weighted Mean			−0.16	0.02	4.81	−1.07	5.94	1.50	

CD, critical difference; locations, 2020–21 (14): Kanpur, Prayagraj, Ghaghra Ghat, Ayodhya, Gorakhpur, Sabour, Pusa Bihar, Cooch Behar, Kalyani, Burdwan, Manikchak, Ranchi, Chianki, and Dumka; locations, 2021–22 (5): Araol, Ayodhya, Gorakhpur, Burdwan, and Chianki

DBW 187, and HD 3086 by 4.81%, 0.02%, and 5.94%, respectively, in national multi-location trials. HD3411 had a potential yield of 65.8 q/ha with an average yield of 46.75 q/ha under timely sown, irrigated conditions (Table 6). This variety also revealed a high level of resistance against leaf rust and moderate resistance to leaf blight, powdery mildew, kernel bunt, and flag smut. Based on yield superiority to its recurrent parent, this variety has been identified as suitable for timely sown, irrigated conditions in the northeastern plain zone (NEPZ) of the wheat growing region in India (AICRP on wheat and barley).

Discussion

Drought stress alone causes greater yield loss in wheat than biotic stresses. Here, we have shown how the introgression of QTLs associated with water use efficiency performed through MABB can deliver improved stress-resilient products. The timely expression of introgressed QTLs in improved lines enhances water use efficiency and grain yield. Roots are the main organs that sense early moisture stress and try to compensate by extracting moisture from lower soil layers. Cooler canopy temperatures result in better evapotranspiration, which is an indirect measure of efficient water uptake by the roots from deep soil layers. Higher grain yield, chlorophyll content, and NDVI are key indicators of better photosynthetic efficiency of plants under moisture stress. So, we successfully infiltrated these key characteristics into the popular, but drought susceptible variety, HD2733.

The introgression of such complex traits with low heritability and unpredictable genotype × environment interactions is possible due to mapped QTLs associated with drought tolerance traits. MAS-based breeding is simple, efficient, robust, and accurate compared to conventional breeding methods that are time-consuming, laborious, and influenced by the environment. Numerous studies have reported QTLs for grain yield and component traits under drought stress conditions (Pinto et al., 2010; Gupta et al., 2012,, 2017; Puttamadanayaka et al., 2020). However, seldomly noted QTLs have been applied to improve drought tolerance in wheat (Merchuk-Ovnat et al., 2016; Rai et al., 2018; Gautam et al., 2020). In the present study, four drought tolerance QTLs (NDVI, *Xgdm 93*; canopy temperature, *Xbarc68-Xbarc101*; yield and chlorophyll content, *Xgwm304*; and chlorophyll content, *Xgwm301*, have been successfully introgressed into HD2733 using MABB along with combined phenotypic selection. Acuna-Galindo et al. (2015) reported major meta-QTL regions for drought and heat tolerance based on genomic regions identified by independent studies. Our three targeted QTLs were located in putative meta-QTL regions. MQTL26 co-localized with QTL *Xbarc68-Xbarc101* on chromosome 3B and MQTL38 co-localized with QTL *Xgwm304* on chromosome 5A. Another putative MQTL on chromosome 2D co-localized with *Xgwm301*. We transferred these three meta-QTL regions into the HD2733 background. Many drought tolerance component traits are associated with MQTL regions and are believed to be carrying genes underlying drought tolerance mechanisms. Therefore, such regions need to be fine-mapped and validated. The improved lines exhibited relatively higher grain yield under restricted irrigation/rain-fed conditions. The use of improved HD2733 in wheat breeding programs could disperse these QTLs into the backgrounds of genotypes derived from it.

Marker-assisted transfer of drought tolerance QTLs

The most effective way to carry out introgression is by following MABB with stringent phenotypic selection for QTL expression and recurrent parent phenology (Collard and Mackill, 2008). In executing MABB, we considered the important factors mentioned by Frisch and Melchinger (2005), like the number of targeted gene/QTLs to be transferred, the marker map, the applied selection strategy, and the crossing scheme for efficient conversion of the recurrent parent. The reported markers need to be validated before executing MAS (Xu and Crouch, 2008; Nicholas et al., 2008). Therefore, initially in the BC_1F_2 population, we validated the presence of introgressed QTLs in wheat chromosomes 2A, 2D, 3B, and 5A that explained phenotypic variance from 11.9% to 28.3%. In the present investigation, foreground selection with two backcrosses (BC_1F_1 and BC_2F_1) efficiently introgressed four targeted QTLs in a combination of 2–4 QTLs per improved line. The effectiveness of foreground selection was confirmed by the improvement in grain yield of 47.7% in the four QTL-combined backcross-derived lines, compared with HD2733. Similarly, a 48.7% yield increase was observed in two QTL combinations with a QTL for grain yield. Furthermore, those lines carrying chlorophyll and NDVI QTLs produced 21.6% more grain than the recurrent parents, and the significant phenotypic QTL expression in the introgressed lines indicated very low background effects of the recipient genome. Drought tolerance is a complex phenomenon, governed by the combined effect of several QTLs. The plant has to undergo modifications from roots to leaves to meet the altered evapotranspiration demands of moisture stress. Therefore, the breeder has to insert several traits to improve water use efficiency. In wheat, a significant positive association between grain yield, NDVI, chlorophyll content, canopy temperature depression, and thousand kernel weight has been successfully established by Lopes and Reynolds (2012), Harikrishna et al. (2016), and Ramya et al. (2016). Hence, the strategic coupling of NDVI, chlorophyll content, and canopy temperature with grain yield had a complementary effect on productivity under drought-stress conditions.

Prigge et al. (2009) showed that increasing marker density from early to advanced backcross generations resulted in maximum genome recovery with a minimum number of marker data points. Overall, 120 polymorphic SSR markers were sufficient to replace the recurrent parent genome in MABB. Additional backcrosses have produced benefits in increased background recovery to BC_2F_3 compared to BC_1F_4 (Table 3). Our targeted QTLs were dispersed over various chromosomes (2A, 3B, 5A, and 5D), which increased the chance of background recovery in respective chromosomes. Supporting markers assisted background selection by phenotypic selection for critical traits, which maximized the recipient parent genome reconstitution. Bhawar et al. (2011) reported 94.55% genome recovery in selected individuals of the BC_2F_2 generation in their study on pyramiding leaf rust-resistant genes into an elite cultivar, HD2687. Similarly, one MABB line, HD2733-571-296-645 from the BC_2F_3 generation had 95.79%

FIGURE 4
Field view of the variety HD 3411 with its parent ().

background recovery. Fusion of gametes with donor and recurrent heterotic allelic combinations in BC1F1 gives rises to increased homozygosity in the progenies. Since, the majority of gametes in BC1F1 were segregating, random fusion might lead to combinations containing recurrent parent genomes, which could lead to enhanced recovery of RPG at the cost of residual heterozygosity. An MABB study by Chukwu et al., 2020 reported a recurrent parent genome recovery of 80%–86.4% from BC_2F_1 to 93.2%–98.7% from BC_2F_2, after one generation of selfing. Similar reports were presented by Bellundagi et al. (2022) for wheat and Miah et al. (2015) in rice. The potential application of background selection in accelerating the recurrent parent genome was thoroughly studied and discussed widely by many researchers (Servin and Hospital, 2002; Bai et al., 2006; Bhawar et al., 2011; Singh et al., 2012). The HD3411, the improved version of HD2733, derived from this cross has shown yield superiority over the recurrent parent under timely sown, irrigated conditions. A field view of HD 3411 with donor and recipient parents is given in Figure 4. Since this variety has been introgressed with QTLs for drought tolerance traits, it is also recommended for restricted irrigation conditions (AICRP on wheat and barley). We conclude that the success of MABB in delivering a drought-tolerant version of HD2733 is attributed to efficient foreground selection for different targeted QTLs, the combined effect of QTLs on yield in recurrent parent backgrounds, the screening of a large segregating population, and the presence of complementing markers that assisted background selection by phenotypic selection.

Conclusion

The improved lines with different QTL combinations showed higher phenotypic mean values for respective traits (NDVI, chlorophyll content, low canopy temperature, and grain yield) compared to the recurrent parent. The backcross-derived lines carrying QTLs for both yield and physiological traits were superior in yield to the lines carrying QTL for either of the

physiological traits alone. The four QTLs introgressed through MABB led to the development of drought-tolerant HD2733. Phenological traits such as days to flag leaf emergence, days to heading, days to anthesis, and days to maturity of back cross-derived lines were equal to the recurrent HD 2733 parent and slightly higher than the recurrent parent for yield-contributing traits. A total of 13 progeny lines in BC_2F_3 and 10 progeny lines in BC_1F_4 generations were found promising in performance under drought stress. One superior line HD3411 has shown higher yield over selected cultivars ranging from 0.02% to 5.94% after 2 years of a multi-location trial at the national level. The variety, HD3411, has been identified for varietal release and testing in the northeastern plain zone of the wheat-growing region in India.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

KP, GS, PS, and NJ designed the study and provided critical inputs. PK performed experiments. PK, AB, HK, RT, and NR collected the phenotypic data. PK, AB, MM, HK, and NS contributed to genotyping data collection and analyzed the data. PK, MM, PS, HK, and KP wrote the manuscript. All authors contributed to the submitted manuscript.

Funding

Part of the research was supported by a grant from the Bill & Melinda Gates Foundation (Grant number # OPP53402) under a Generation Challenge Program (GCP) and ICAR-funded NICRA project.

Acknowledgments

PK acknowledges the ICAR-Indian Agricultural Research Institute (IARI), New Delhi, for scholarships to complete this work as a part of Ph.D. research work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Acuna-Galindo, M. A., Mason, R. E., Subramanian, N. K., and Hays, D. B. (2015). Meta-analysis of wheat QTL regions associated with adaptation to drought and heat stress. *Crop Sci.* 55, 477–492. doi:10.2135/cropsci2013.11.0793

AICRP on Wheat and Barley *All India coordinated research project (AICRP) on wheat & barley*. Available at: www.aicrpwheatbarleyicar.in/wp-content/uploads/2022/09/Final-VICproceedings-2022 (Accessed September 5, 2022).

Araus, J. L., Slafer, G. A., Royo, C., and Serret, M. D. (2008). Breeding for yield potential and stress adaptation in cereals. *Crit. Rev. Plant Sci.* 27, 377–412. doi:10.1080/07352680802467736

Bai, J. Y., Zhang, Q., and Jia, X. P. (2006). Comparison of different foreground and background selection methods in marker-assisted introgression. *Acta Genet. Sin.* 33, 1073–1080. doi:10.1016/S0379-4172(06)60144-3

Bellundagi, A., Ramya, K. T., Krishna, H., Jain, N., Shashikumara, P., Singh, P. K., et al. (2022). Marker-assisted backcross breeding for heat tolerance in bread wheat (*Triticum aestivum* L.). *Front. Genet.* 13, 1056783. doi:10.3389/fgene.2022.1056783

Bhawar, K. B., Vinod, Sharma, J. B., Singh, A. K., Sivasamy, M., Singh, M., et al. (2011). Molecular marker assisted pyramiding of leaf rust resistance genes Lr19 and Lr28 in bread wheat (*Triticum aestivum* L.) variety HD2687. *Indian J. Genet.* 71 (4), 304–311.

Chukwu, S. C., Rafii, M. Y., Ramlee, S. I., Ismail, S. I., Oladosu, Y., Muhammad, I. I., et al. (2020). Recovery of recurrent parent genome in a marker-assisted backcrossing against rice blast and blight infections using functional markers and SSRs. *Plants* 9 (11), 1411. doi:10.3390/plants9111411

Collard, B. C., and Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 363, 557–572. doi:10.1098/rstb.2007.2170

De Bustos, A., Rubio, P., Soler, C., Garcia, P., and Jouve, N. (2001). Marker assisted selection to improve HMW-glutenins in wheat. *Euphytica* 119, 69–73. doi:10.1023/A:1017534203520

Frisch, M., and Melchinger, A. E. (2005). Selection theory for marker assisted backcrossing. *Genetics* 170, 909–917. doi:10.1534/genetics.104.035451

Gautam, T., Saripalli, G., Kumar, A., Gahlaut, V., Gadekar, D. A., Oak, M., et al. (2020). Introgression of a drought insensitive grain yield QTL for improvement of four Indian bread wheat cultivars using marker assisted breeding without background selection. *J. Plant Biochem. Biotechnol.* 1, 172–183. doi:10.1007/s13562-020-00553-0

Gupta, P. K., Balyan, H. S., Edwards, K. J., Isaac, P., Korzun, V., R€oder, M., et al. (2002). Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat. *Theor. Appl. Genet.* 105, 413–422. doi:10.1007/s00122-002-0865-9

Gupta, P. K., Balyan, H. S., Gahlaut, V., and Kulwal, P. L. (2012). Phenotyping, genetic dissection, and breeding for drought and heat tolerance in common wheat: Status and prospects. *Plant Breed. Rev.* 36, 85–168. doi:10.1002/9781118358566

Gupta, P. K., Balyan, H. S., and Gahlaut, V. (2017). QTL analysis for drought tolerance in wheat: Present status and future possibilities. *Agronomy* 7, 5–21. doi:10.3390/agronomy7010005

Harikrishna,Singh, G. P., Jain., N., Singh, P. K., Prasad., S., Divya, A., et al. (2016). Physiological characterization and grain yield stability analysis of RILs under different moisture stress conditions in wheat (*Triticum aestivum* L.). *Ind. J. Plant Physiol.* 21, 576–582. doi:10.1007/s40502-016-0257-9

Hospital, F., and Charcosset, A. (1997). Marker assisted introgression of quantitative trait loci. *Genetics* 147, 1469–1485. doi:10.1093/genetics/147.3.1469

Kang, Y., Khan, S., and Ma, X. (2009). Climate change impacts on crop yield, crop water productivity and food security: A review. *Prog. Nat. Sci.* 19, 1665–1674. doi:10.1016/j.pnsc.2009.08.001

Khaled, K. A. M., Habiba, R. M. M., Bashasha, J. A., and El-Aziz, M. H. A. (2022). Identification and mapping of QTL associated with some traits related for drought tolerance in wheat using SSR markers. *Beni-Suef Univ. J. Basic Appl. Sci.* 11 (1), 38–13. doi:10.1186/s43088-022-00212-4

Kirigwi, F. M., Van Ginkel, M., Brown-Guedira, G., Gill, B. S., Paulsen, G. M., and Fritz, A. K. (2007). Markers associated with a QTL for grain yield in wheat under drought. *Mol. Breed.* 20, 401–413. doi:10.1007/s11032-007-9100-3

Kumar, J., Mir, R. R., Kumar, N., Kumar, A., Mohan, A., Prabhu, K. V., et al. (2010). Marker-assisted selection for pre-harvest sprouting tolerance and leaf rust resistance in bread wheat. *Plant Breed.* 129, 617–621. doi:10.1111/j.1439-0523.2009.01758.x

Kumar, S., Sehgal, S. K., Kumar, U., Vara Prasad, P. V., Joshi, A. K., and Gill, B. S. (2012). Genomic characterization of drought tolerance related traits in spring wheat. *Euphytica* 186, 265–276. doi:10.1007/s10681-012-0675-3

Lopes, M. S., and Reynolds, M. P. (2012). Stay-green in spring wheat can be determined by spectral reflectance measurements (normalized difference vegetation index) independently from phenology. *J. Exp. Bot.* 63, 3789–3798. doi:10.1093/jxb/ers071

Lu, Y., Hao, Z., Xie, C., Crossa, J., Araus, J. L., Gao, S., et al. (2011). Large-scale screening for maize drought resistance using multiple selection criteria evaluated under water-stressed and well-watered environments. *Field Crops Res.* 124, 37–45. doi:10.1016/j.fcr.2011.06.003

Manu, B., Shashikumara, P., Sunil, B., Divya, C., Rahul, P., Divya, A., et al. (2020). Genetic gain and morpho-physiological characterization of BILs (Backcross inbred lines) under different moisture regimes in wheat (*Triticum aestivum*. L). *Indian J. Genet.* 80, 84–93. doi:10.31742/IJGPB.80.1.11

Matiu, M., Ankerst, D. P., and Menzel, A. (2017). Interactions between temperature and drought in global and regional crop yield variability during 1961–2014. *PLoS ONE* 12, e0178339. doi:10.1371/journal.pone.0178339

Melchinger, A. E. (1990). Use of molecular markers in breeding for oligogenic disease resistance. *Plant Breed.* 104, 1–19. doi:10.1111/j.1439-0523.1990.tb00396.x

Merchuk-Ovnat, L., Barak, V., Fahima, T., Ordon, F., Lidzbarsky, G. A., Krugman, T., et al. (2016). Ancestral QTL alleles from wild emmer wheat improve drought resistance and productivity in modern wheat cultivars. *Front. Plant Sci.* 7, 452. doi:10.3389/fpls.2016.00452

Miah, G., Rafii, M. Y., Ismail, M. R., Puteh, A. B., Rahim, H. A., and Latif, M. A. (2015). Recurrent parent genome recovery analysis in a marker-assisted backcrossing program of rice (Oryza sativa L.). *Comptes rendus Biol.* 338 (2), 83–94. doi:10.1016/j.crvi.2014.11.003

Mir, R. R., Zaman-Allah, M., Sreenivasulu, N., Trethowan, R., and Varshney, R. K. (2012). Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops. *Theor. Appl. Genet.* 125, 625–645. doi:10.1007/s00122-012-1904-9

Nicholas, C. C., Francxois, T., and Roberto, T. (2008). Quantitative trait loci and crop performance under abiotic stress: Where do we stand? *Plant Physiol.* 147, 469–486. doi:10.1104/pp.108.118117

Olivares-Villegas, J. J., Reynolds, M. P., and McDonald, G. K. (2007). Drought-adaptive attributes in the Seri/Babax hexaploid wheat population. *Funct. Plant Biol.* 34, 189–203. doi:10.1071/FP06148

Pask, A. J. D., Pietragalla, J., Mullan, D. M., and Reynolds, M. P. (2012). *Physiological breeding II: A field guide to wheat phenotyping*. Mexico, DF: CIMMYT.

Pestsova, E., Ganal, M. W., and R€oder, M. S. (2000). Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. *Genome* 43, 689–697. doi:10.1139/g00-042

Pinto, R. S., Reynolds, M. P., Mathews, K. L., McIntyre, C. L., Olivares-Villegas, J. J., and Chapman, S. C. (2010). Heat and drought adaptive QTL in a wheat population designed to minimize confounding agronomic effects. *Theor. Appl. Genet.* 121, 1001–1021. doi:10.1007/s00122-010-1351-4

Prabhu, K. V., Somers, D. J., Rakow, G., and Gugel, R. K. (1998). Molecular markers linked to white rust resistance in mustard Brassica juncea. *Theor. Appl. Genet.* 97, 865–870. doi:10.1007/s001220050966

Pradhan, G. P., Prasad, P. V. V., Fritz, A. K., Mary, B. K., and Bikram, S. G. (2012). Effects of drought and high temperature stress on synthetic hexaploid wheat. *Funct. Plant. Biol.* 39, 190–198. doi:10.1071/FP11245

Prigge, V., Melchinger, A. E., Dhillon, B. S., and Frisch, M. (2009). Efficiency gain of marker-assisted backcrossing by sequentially increasing marker densities over generations. *Theor. Appl. Genet.* 119, 23–32. doi:10.1007/s00122-009-1013-6

Puttamadanayaka, S., HarikrishnaBalaramaiah, M., Biradar, S., Sunilkumar, V. P., Sinha, N., et al. (2020). Mapping genomic regions of moisture deficit stress tolerance using backcross inbred lines in wheat (*Triticum aestivum* L.). *Sci. Rep.* 10 (1), 21646–21717. doi:10.1038/s41598-020-78671-x

Quarrie, S. A., Quarrie, S. P., Radosevic, R., Rancic, D., Kaminska, A. S., Barnes, J. D., et al. (2006). Dissecting a wheat QTL for yield present in a range of environments: From the QTL to candidate genes. *J. Exp. Bot.* 57, 2627–2637. doi:10.1093/jxb/erl026

Quarrie, S. A., Steed, A., Calestani, C., Semikhodskii, A., Lebreton, C., Chinoy, C., et al. (2005). A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring x SQ1 and its use to compare QTLs for grain yield across a range of environments. *Theor. Appl. Genet.* 110, 865–880. doi:10.1007/s00122-004-1902-7

Raghavendra, P., Kumar, B. M. D., Shashikumara, P., Sachin, K. H. M., Krishnamurthy, S. L., and Shailaja, H. (2020). Marker-assisted identification of novel genetic lines for salinity tolerance and their categorization for utilization in development of hybrid rice (*Oryza sativa* L.). *Cereal Res. Commun.* 49, 171–180. doi:10.1007/s42976-020-00104-2

Rai, N., Bellundagi, A., Kumar, P. K., Kalasapura Thimmappa, R., Rani, S., Sinha, N., et al. (2018). Marker-assisted backcross breeding for improvement of drought tolerance in bread wheat (*Triticum aestivum* L. em Thell). *Plant Breed.* 137, 514–526. doi:10.1111/pbr.12605

Rajaram, S., Van Ginkel, M., and Fischer, R. A. (1995). "CIMMYT's wheat breeding mega environments (ME)," in Proceedings of the 8th International Wheat Genetics Symposium, Beijing, China.

Ramya, P., Singh, G. P., Neelu, J., Singh, P. K., Manoj, K. P., Kavita, S., et al. (2016). Effect of recurrent selection on drought tolerance and related morpho-physiological traits in bread wheat. *PloS one* 11, e0156869. doi:10.1371/journal.pone.0156869

Reynolds, M. P., and Langridge, P. (2016). Physiological breeding. *Curr. Opin. Plant Biol.* 31, 162–171. doi:10.1016/j.pbi.2016.04.005

Reynolds, M. P., Ortiz-Monasterio, J. I., and McNab, A. (2001). in *Application of physiology in wheat breeding* (Mexico, D.F.: CIMMYT), 88–100.

Richards, R. A., Condon, A. G., and Rebetzke, G. J. (2001). "Traits to improve yield in dry environments," in *Application of physiology in wheat breeding*. Editors M. P. Reynolds, J. I. Ortiz-Monasterio, and A. D. F. McNab (Mexico: CIMMYT), 88–100.

Roder, M. S., Korzun, V., Wendehake, K., Plaschke, J., Marie-Helene, T., Leroy, P., et al. (1998). A microsatellite map of wheat. *Genetics* 149, 2007–2023. doi:10.1093/genetics/149.4.2007

Sallam, A., Eltaher, S., Alqudah, A. M., Belamkar, V., and Baenziger, P. S. (2022). Combined GWAS and QTL mapping revealed candidate genes and SNP network controlling recovery and tolerance traits associated with drought tolerance in seedling winter wheat. *Genomics* 114 (3), 110358. doi:10.1016/j.ygeno.2022.110358

Servin, B., and Hospital, F. (2002). Optimal positioning of markers to control genetic background in marker assisted backcrossing. *J. Hered.* 93, 214–217. doi:10.1093/jhered/93.3.214

Shashikumara, P., Harikrishna, Neelu, J., Nivedita, S., Divya, C., Rahul, M. P., et al. (2020). Genetic variability, Genotype × Environment interaction for grain yield of wheat (*Triticum aestivum* L.) backcross inbred lines population under different moisture regimes. *Indian J. Agric. Sci.* 90 (9), 1678–1684. doi:10.56093/ijas.v90i9.106595

Shukla, S., Singh, K., Patil, R. V., Kadam, S., Bharti, S., Prasad, P., et al. (2015). Genomic regions associated with grain yield under drought stress in wheat (*Triticum aestivum* L.). *Euphytica* 203, 449–467. doi:10.1007/s10681-014-1314-y

Singh, A., Vikas, K. S., Singh, S. P., Pandian, R. T. P., Ranjith, K. E., Davinder, S., et al. (2012). Molecular breeding for the development of multiple disease resistance in Basmati rice. *AoB Plants* 29, pls029–13. doi:10.1093/aobpla/pls029

Somers, D. J., Isaac, P., and Edwards, K. (2004). A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 109, 1105–1114. doi:10.1007/s00122-004-1740-7

Subash, N., Ram Mohan, H. S., and Banu Kumar, K. (2011). Comparing water-vegetative indices for rice (*Oryza sativa* L.)–wheat (*Triticum aestivum* L.) drought assessment. *Comput. Electron. Agric.* 77, 175–187. doi:10.1016/j.compag.2011.05.001

Sundaram, R. M., Vishnupriya, M. R., Biradar, S. K., Laha, G. S., Reddy, G. A., Rani, N. S., et al. (2008). Marker-assisted introgression of bacterial blight resistance in Samba Mahsuri, an elite indica rice variety. *Euphytica* 160, 411–422. doi:10.1007/s10681-007-9564-6

Sunil, krishna, H., Deepali, U., Rahul, G., Shashikumara, P., Divya, C., et al. (2020). QTL mapping for heat tolerance related traits using backcross inbred lines in wheat (*Triticum aestivum* L.). *Indian J. Genet.* 80, 242–249. doi:10.31742/IJGPB.80.3.2

Tanksley, S. D. (1983). Mapping polygenes. *Annu. Rev. Genet.* 27, 205–233. doi:10.1146/annurev.ge.27.120193.001225

Tanksley, S. D., Young, N. D., Paterson, A. H., and Bonierbale, M. W. (1989). Rflp mapping in plant-breeding – new tools for an old science. *Nat. Biotechnol.* 7, 257–264. doi:10.1038/nbt0389-257

Trnka, M., Feng, S., Semenov, M. A., Olesen, J. E., Kersebaum, K. C., Rötter, R. P., et al. (2019). Mitigation efforts will not fully alleviate the increase in water scarcity occurrence probability in wheat-producing areas. *Sci. Adv.* 5 (9), eaau2406. doi:10.1126/sciadv.aau2406

Tuberosa, R. (2012). Phenotyping for drought tolerance of crops in the genomics era. *Front. Physiol.* 3, 347. doi:10.3389/fphys.2012.00347

Visscher, P. M., Haley, C. S., and Thompson, R. (1996). Marker-assisted introgression in backcross breeding programs. *Genetics* 144, 1923–1932. doi:10.1093/genetics/144.4.1923

Xu, Y., and Crouch, J. H. (2008). Marker-assisted selection in plant breeding: From publications to practice. *Crop Sci.* 48 (2), 391–407. doi:10.2135/cropsci2007.04.0191

Young, N. D., and Tanksley, S. D. (1989). RFLP analysis of the size of chromosomal segments retained around the Tm-2 locus of tomato during backcross breeding. *Theor. Appl. Genet.* 77, 353–359. doi:10.1007/BF00305828

Zadoks, J. C., Chang, T. T., and Cal Konzak, C. F. (1974). A decimal code for the growth stages of cereals. *Weed Res.* 14 (6), 415–421. doi:10.1111/j.1365-3180.1974.tb01084.x

Zhou, W. C., Kolb, F. L., Bai, G. H., Domier, L. L., Boze, L. K., and Smith, N. J. (2003). Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. *Plant Breed.* 122, 40–46. doi:10.1046/j.1439-0523.2003.00802.x

Frontiers in Genetics

TYPE Review
PUBLISHED 08 March 2023
DOI 10.3389/fgene.2023.1121462

Check for updates

OPEN ACCESS

EDITED BY
Vijay Rani Rajpal,
University of Delhi, India

REVIEWED BY
Shailendra Goel,
University of Delhi, India
Rakesh Kumar,
Central University of Karnataka, India
Gopala Krishnan S,
Indian Council of Agricultural Research (ICAR), India

*CORRESPONDENCE
Naveen Puppala,
npuppala@nmsu.edu

SPECIALTY SECTION
This article was submitted to Plant Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 11 December 2022
ACCEPTED 06 February 2023
PUBLISHED 08 March 2023

CITATION
Puppala N, Nayak SN, Sanz-Saez A, Chen C, Devi MJ, Nivedita N, Bao Y, He G, Traore SM, Wright DA, Pandey MK and Sharma V (2023), Sustaining yield and nutritional quality of peanuts in harsh environments: Physiological and molecular basis of drought and heat stress tolerance.
Front. Genet. 14:1121462.
doi: 10.3389/fgene.2023.1121462

COPYRIGHT
© 2023 Puppala, Nayak, Sanz-Saez, Chen, Devi, Nivedita, Bao, He, Traore, Wright, Pandey and Sharma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Sustaining yield and nutritional quality of peanuts in harsh environments: Physiological and molecular basis of drought and heat stress tolerance

Naveen Puppala[1]*, Spurthi N. Nayak[2], Alvaro Sanz-Saez[3], Charles Chen[3], Mura Jyostna Devi[4,5], Nivedita Nivedita[5], Yin Bao[6], Guohao He[7], Sy M. Traore[7], David A. Wright[8], Manish K. Pandey[9] and Vinay Sharma[9]

[1]Agricultural Science Center at Clovis, New Mexico State University, Las Cruces, NM, United States, [2]Department of Biotechnology, University of Agricultural Sciences, Dharwad, India, [3]Department of Crop, Soil and Environmental Sciences, Auburn University, Auburn, AL, United States, [4]USDA-ARS Vegetable Crops Research, Madison, WI, United States, [5]Department of Horticulture, University of Wisconsin-Madison, Madison, WI, United States, [6]Biosystems Engineering Department, Auburn University, Auburn, AL, United States, [7]Department of Plant and Soil Sciences, Tuskegee University, Tuskegee, AL, United States, [8]Department of Biotechnology, Iowa State University, Ames, IA, United States, [9]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Telangana, India

Climate change is significantly impacting agricultural production worldwide. Peanuts provide food and nutritional security to millions of people across the globe because of its high nutritive values. Drought and heat stress alone or in combination cause substantial yield losses to peanut production. The stress, in addition, adversely impact nutritional quality. Peanuts exposed to drought stress at reproductive stage are prone to aflatoxin contamination, which imposes a restriction on use of peanuts as health food and also adversely impact peanut trade. A comprehensive understanding of the impact of drought and heat stress at physiological and molecular levels may accelerate the development of stress tolerant productive peanut cultivars adapted to a given production system. Significant progress has been achieved towards the characterization of germplasm for drought and heat stress tolerance, unlocking the physiological and molecular basis of stress tolerance, identifying significant marker-trait associations as well major QTLs and candidate genes associated with drought tolerance, which after validation may be deployed to initiate marker-assisted breeding for abiotic stress adaptation in peanut. The proof of concept about the use of transgenic technology to add value to peanuts has been demonstrated. Advances in phenomics and artificial intelligence to accelerate the timely and cost-effective collection of phenotyping data in large germplasm/breeding populations have also been discussed. Greater focus is needed to accelerate research on heat stress tolerance in peanut. A suits of technological innovations are now available in the breeders toolbox to enhance productivity and nutritional quality of peanuts in harsh environments. A holistic breeding approach that considers drought and heat-tolerant traits to simultaneously address both stresses could be a successful strategy to produce climate-resilient peanut genotypes with improved nutritional quality.

KEYWORDS

drought, heat stress, physiological, molecular, high-throughput phenotyping, introgression, transgenes and CRISPR

Introduction

Climate change and hot weather extremes have perpetuated vulnerability in the ecosystem and agriculture sector, threatening food and nutritional security. Intergovernmental Panel on Climate Change (IPCC) assessment has estimated a 1.5°C rise in global warming in the near term (2021–2040). At the projected rising temperature of 2°C–3°C, the subsequent increase in frequency and severity of water scarcity (drought stress) will lead to severe loss in biodiversity and crop production in various geographic regions (IPCC, 2022). Drought is the single greatest abiotic stress, reducing yield under rainfed and irrigated cropping systems (Boyer, 1982; Araus et al., 2002). Drought produces the reduction of transpiration and thus photosynthesis which results in decreased biomass accumulation and yield (Tardieu and Tuberosa, 2010). For example, in 2012, the drought that happend in the United States (US) during summer and fall, cost approximately 30 billion dollars to the US economy (Riley, 2015). Additionally the US peanut industry losses every year 50 million dollars due to drought stress (U.S. Department of Agriculture and Agricultural Research Service, 2019).

Peanut is an important oilseed crop, widely grown across continents in semi-arid tropics, and often exposed to drought and heat stresses, with severe losses in production and deterioration in peanut quality worldwide (Nigam et al., 2005; Hamidou et al., 2012; Hamidou et al., 2013). About 90% of the world's peanuts are cultivated in tropical and semi-arid regions, and ~65% of United States peanuts are grown in dryland, and rainfed conditions (Hamidou et al., 2013). While peanuts tolerate early drought stress, it is more sensitive to drought and heat stress toward the reproductive phase. A temperature range between 25°C and 30°C is optimum for peanut growth and productivity. Temperature above 32°C negatively impacts yield and total biomass in peanuts (Cox, 1979; Golombek and Johansen, 1997; Prasad et al., 2003). Peanuts under drought stress are vulnerable to aflatoxin contamination due to infection caused by *Aspergillus flavus* (Hamidou et al., 2014), a toxic substance harmful to human and animal health, impacting the peanut trade internationally. Drought and heat stress also alters compositional changes in seed chemistry, including adverse effects on minerals (Dwivedi et al., 2013).

A meta-analysis involving over 120 published case studies of crop responses to combined drought and heat stress reveals that the combined effect significantly impacts yield by reducing harvest index, shortening the life cycle of crops, and altering seed number, size, and composition. Moreover, such impacts are more severe when the stress combination occurs during the crops reproductive phase (Cohen et al., 2021).

Hence, understanding the physiological and molecular basis of drought and heat stress tolerance is the key to improving peanuts' productivity in harsh environments (Figure 1). Here we provide synthesis to a wide range of plant responses to these stresses to harness variation toward developing stress-tolerant and productive peanut germplasm, which may be recycled in breeding programs or could be deployed in commercial production after assessing their performance in each production system.

High-throughput phenomics to accelerate data collection in germplasm/breeding populations

Drought stress

Traditional screening for drought tolerance refers to conditions in which the germplasm/breeding populations are exposed to varying moisture stress levels in field environments. While the control plots receive optimal irrigation throughout the crop cycle, in the stressed plots, water is withheld at a critical stage (i.e., reproductive phase) for a specific period and then released similar to control (irrigated) plots. The difference in pod yield between irrigated and drought-stressed plots is measured as a response to drought stress. The genotypes that show the least reduction in pod yield under stress are classified as tolerant to drought (Craufurd et al., 2003; Krishnamurthy et al., 2007; Kakani et al., 2015; Akbar et al., 2017; Abadya et al., 2021). This type of stress is categorized as a mid- or end-of-season drought. The occurrence, frequency, and intensity of stress in natural field environments are difficult to predict, i.e., the crop may face stress at any given time during the rainy season. This type of stress is defined as intermittent drought. In a situation like this, the genotypes/breeding populations are exposed to intermittent drought stress while the corresponding control plots receive optimal irrigation throughout the crop cycle. Genotypes with the least difference in pod yield between stressed and control (irrigated) plots are identified as tolerant to intermittent drought (Gangappa et al., 2006; Ratnakumar and Vadez, 2011; Hamidou et al., 2012; Vadez and Ratnakumar, 2016). Such screening methods are time- and resource-intensive, and subject to bias due to genotype-by-environment interaction effects. However, screening only for yield response under drought or high temperature does not give us information regarding the physiological and genetic mechanisms that may be involved in the observed yield under drought tolerance.

Water use efficiency (WUE) is a critical trait in breeding for drought- and heat-stress tolerance in peanuts. However, long-term transpiration is challenging to measure under field conditions. Technically, it requires using lysimeters, which is economically unfeasible for typical peanut breeding programs. Surrogates for WUE have been identified, including carbon isotope discrimination ($\Delta^{13}C$), specific leaf area (SLA), SPAD chlorophyll meter reading (SCMR), canopy temperature depression (CTD), normalized difference vegetation index (NDVI), and visual rating of leaf wilting (Wright et al., 1994; Ravi et al., 2011; Luis et al., 2016; Vadez and Ratnakumar, 2016). Still, the labor and time required to collect these measurements are prohibitive for large populations and in multiple environments. High-throughput phenotyping in plants is thus a significant bottleneck in breeding programs.

TABLE 1 Exisiting high-throughput plant phenotyping studies related to drought tolerance screening in peanuts.

Sensor	Platform	Data analytics	Traits	References
RGB	UAV	VIs	leaf wilting rating	Balota and Oakes (2017)
		VIs + linear/ML regression	leaf area index, lateral growth	Sarkar et al. (2021)
		digital surface model	plant height	Sarkar et al. (2020)
	pushcart	CNN-based pod detection and counting	pod yield	Bidese et al. (2021)
	minirhizotron	CNN-based root semantic segmentation	root architecture	Xu et al. (2020)
multispectral	UAV	NDVI	disease rating, pod yield	Patrick et al. (2017), Chen et al. (2020)
hyperspectral	UAV	VI + ensemble ML	pod yield, pod count, biomass	Bagherian et al. (2022)
		VNIR reflectance + CNN		
thermal	handheld	canopy temperature	transpiration, pod yield	Balota and Oakes (2017)
LiDAR	UGV	point cloud analysis	plant height	Yuan et al. (2019)
GPR	pushcart	image thresholding	pod yield	Dobreva et al. (2021)

Vegetation index (VI), Machine learning (ML), convolutional neural network (CNN), unmanned aerial vehicle (UAV), unmanned ground vehicle (UGV), ground penetrating radar (GPR), light detection and ranging (LiDAR), visible near-infrared (VNIR).

High-throughput plant phenotyping (HTPP) employs unmanned aerial vehicle (UAV), unmanned ground vehicle (UGV), robotics, various imaging technologies, and advanced data analytics to enable efficient and effective characterization of complex plant traits for screening germplasm or breeding populations. Most HTPP research has been focused on major row crops such as cotton, maize, soybean, and wheat, whereas HTPP research in peanuts only started in recent years. Table 1 summarized a list of existing HTPP studies where the predicted traits were or could be used to screen drought tolerance in peanuts.

Infrared thermal imaging of canopy temperature is currently the most accurate and direct method, as drought-induced stomatal closure causes a reduction in transpiration and, thus, a decrease in canopy temperature. Typically, a thermal camera would be mounted on a UAV along with other imaging sensors for high-throughput multi-modal imagery acquisition over a large area. Balota and Oakes (2017) first evaluated UAV-based red-green-blue (RGB) and near-infrared (NIR) imaging and handheld RGB and thermal imaging for HTPP of 26 peanut cultivars in a drought experiment. RGB color indices, NDVI, and CTD, were found to have strong to moderate correlations with visual leaf wilting rating, pod yield, sound mature kernel, and crop value at the end of water stress imposition. Aerial RGB color indices coupled with statistical learning models have been reported to achieve a 90% accuracy in predicting visual leaf wilting ratings (Sarkar et al., 2021). CTD can detect drought stress before visible leaf wilting occurs (Balota and Oakes, 2017). Other related HTPP studies in peanuts were focused on peanut canopy morphology. Although they may not provide early detection of drought and heat stress, peanut canopy architecture traits can potentially influence plant water use. Peanut canopy height has been quantified accurately by both Light Detection and Ranging (LiDAR) sensors on a high-clearance motorized cart (Yuan et al., 2019) and digital surface models derived from a UAV-based RGB imaging platform (Sarkar et al., 2020). In addition, leaf area index (LAI) and lateral growth can be predicted by training statistical and machine learning models on aerial vegetation indices (Sarkar et al., 2021). Deep roots can increase plant water uptake capability, contributing to drought and heat tolerance. Minirhizotron imaging has been used to infield HTPP of peanut root architecture, and the UNet-based semantic segmentation method has been effective and robust in detecting root pixels (Xu et al., 2020).

HTPP of other agronomic traits in peanuts also has been standardized to accelerate peanut breeding efforts. Pod yield has been the most important trait to measure in peanut breeding programs. UAV remote sensing-based vegetation indices of peanut canopy at critical phenological stages, such as the pod-filling stage, have shown their value for early yield prediction (Balota and Oakes, 2017; Jewan et al., 2022). For direct sensing of peanut pods, ground penetrating radar has shown the potential to explain yield variability up to 51% (Dobreva et al., 2021). In addition, HTPP of infield peanut pods after inversion presents a low-cost approach for pod yield prediction at the end of the growing season. Bidese et al. (2021) employed a push-cart system to collect top-viewing and side-viewing RGB videos of inverted peanut plants in the field. They explored Mask R-CNN-based peanut pod detection coupled with multivariate linear regression for pod yield prediction. The imaged scenes were highly complex, with heavy occlusions between peanut pods, leaves, and vines. The potential of this approach needs further investigation to account for pod size and variability in visibility. Disease incidence may become a confounding factor for screening of drought and heat tolerant peanut genotypes and affect subsequent data analysis and selection process. UAV multispectral imaging-derived vegetation indices accurately predict visual ratings of tomato spot wilt virus and bacterial wilt in peanuts (Patrick et al., 2017; Chen et al., 2020).

In addition to the studies reviewed above, other HTPP technologies can facilitate breeding drought and heat tolerance in peanuts. Hyperspectral imaging provides both high resolutions in spatial and spectral dimensions for phant phenotyping applications (Sarić et al., 2022). Compared to a typical multispectral camera with five wide spectral bands (blue, green, red, red edge, and near-infrared), a

TABLE 2 HTPP methods that have not been studied for drought and heat stress phenotyping in peanuts. Shortwave infrared (SWIR).

Sensor	Platform	Data analytics	Traits	References
hyperspectral	UAV	VIs	leaf water content	Raj et al. (2021)
	pushcart	VNIR reflectance + PLSR	$V_{c,max}$, J_{1800}, $P_{\max}$, Chl *a*/*b* ratio, Chlorophyll content	Meacham-Hensold et al. (2020)
	handheld	VNIR + SWIR reflectance + Ensemble ML	$V_{c,max}$, $J_{\max}$	Fu et al. (2019)
chlorophyll fluorescence imaging	field-based gantry	kinetic chlorophyll fluorescence curve	F_V/F_M	Herritt et al. (2020)
	controlled imaging chamber		F_q'/F_m', F_V/F_M, Φ_{PSII}	Abdelhakim et al. (2021), Wang H. et al. (2018), Yao et al. (2018)

visible-near-infrared (VNIR) hyperspectral camera produces hundreds of narrow spectral bands between 400 and 1000 nm wavelengths, which can reveal a far more detailed spectral signature of plant organs. New normalized difference vegetation indices (i.e., FOSBNDI-1, FOSBNDI-2, and COSBNDI) derived from UAV-based hyperspectral data were effective at predicting maize leaf water content at the V6 stage in conjunction with a machine learning model (Raj et al., 2021). In contrast to engineering spectral features, all spectral bands can also be exploited by statistical or deep learning methods (i.e., partial least squares regression or deep convolutional neural networks) for maize leaf water content prediction with automatic feature selection or learning (Ge et al., 2016; Rehman et al., 2020). Similar approaches can help rapidly screen peanut genotypes with high water uptake capability.

Regarding remote sensing of peanut yield and yield components, Bagherian et al. (2022) recently evaluated predicting biomass, pod count, and pod yield using UAV-based hyperspectral imaging and machine learning techniques for single peanut plants of an F_1 population in a mid-season drought experiment using rainout shelters. Eighteen days after the drought was found to result in the highest prediction accuracies for the three agronomic traits (R^2 = 0.52–0.61). On the other hand, high photosynthetic capacity can act as a mechanism for drought and heat tolerance. Currently, photosynthetic parameters such as the maximum carboxylation rate of Rubisco ($V_{c,max}$), maximum electron transport rate (J_{1800}), maximum electron transport rate supporting RuBP regeneration ($J_{\max}$), maximal light-saturated photosynthesis ($P_{\max}$), and chlorophyll content are often measured using a portable photosynthesis instrument (e.g., LI-COR LI-6800), which can be extremely time- and labor-intensive. Hyperspectral imaging and machine learning have been found effective in predicting photosynthetic parameters (Fu et al., 2019; Meacham-Hensold et al., 2020). This can be applied to track the temporal dynamics of photosynthetic activity in peanuts for drought experiments under rainout shelters and identify superior genotypes that quickly recover from drought stress. Chlorophyll fluorescence imaging provides a two-dimensional image instead of a point measurement using a chlorophyll fluorimeter. The resultant high spatial resolution can reveal spatial variability in photosynthetic performance on a single leaf or between leaves on a plant. Chlorophyll fluorescence imaging has been used to study drought and heat stress response in tomato, *Arabidopsis*, and wheat in controlled environments (Wang C. et al., 2018; Yao et al., 2018; Abdelhakim et al., 2021) and grain sorghum under field conditions (Herritt et al., 2020). Chlorophyll fluorescence measurements, such as maximum and operating quantum efficiencies of photosystem II photochemistry (F_V/F_M, Φ_{PSII} = $(F_m'-F)/F_m'$), can indicate plant drought and heat stress earlier than the occurrence of morphological changes such as leaf wilting. Table 2 lists some HTPP studies for drought and/or heat stress in crop species that could be applied to peanuts.

With the advancements in HTPP and increasing availability of high-dimensional sensor-based phenotypic datasets, phenomic-assisted selection has recently been proposed and evaluated. This analysis uses HTPP data instead of genomic data as input to the statistical models in genomic selection. The phenomic choice can achieve comparable predictive accuracy compared to genomic selection for crops such as wheat, soybean, and maize (Rincent et al., 2018; Parmley et al., 2019; Weiß et al., 2022). The advantages include low cost and robustness across different environments (Rincent et al., 2018). The phenomic selection performed more accurately for complex traits such as grain yield than traits controlled by a few genes (Zhu et al., 2022). Since drought and heat tolerance are considered complex traits, peanut breeders are expected to benefit from phenomic selection and reduce the labor and time required for screening diverse populations.

Heat stress

A similar approach, like discarding of ultrasusceptible types to drought stress, is also recommended for screening for heat stress tolerance in peanut (Akbar et al., 2017; Table 3). Developing a reliable index and identifying traits for acquired thermotolerance in peanuts is necessary for breeding heat-tolerant varieties. Several reports observed the genotypic variability in peanut's heat tolerance for partitioning dry matter to pods and kernels, fruit set, membrane stability, and chlorophyll fluorescence (Srinivasan et al., 1996; Vara Prasad et al., 2001; Craufurd et al., 2003). The detached leaf assay method was used to screen the sixteen genotypes from US minicore accessions along with standard checks were evaluated for acquired thermoterance. Here, the change in the temperature sensitivity of chlorophyll accumulation was used as an indicator of acquired thermotolerance. However, in this study, there was no significance effect of thermotolerance on seed weight was observed, hence it was difficult to relate chlorophyll content with heat tolerance (Selvaraj et al., 2011; Table 3). In another study, peanut seedlings of diverse genotypes for heat tolerance were screened using temperature induction response techniques. About 2 days old peanut seedlings were exposed to sublethal temperature from 28°C to 54°C for 5 h, followed by the lethal temperature at 54°C for 3 h. The expression patterns of stress-responsive genes were analyzed in selected heat-tolerant genotypes;

TABLE 3 Genotypes exhibiting heat stress tolerance along with its responsive trait.

Genotype	Subspecies	Heat stress tolerance responsive/tolerant trait	References
COC038	*A. hypogaea*	Chlorophyll accumulation/HSP production	Selvaraj et al. (2011)
COC041	*A. fastigiata*	Chlorophyll accumulation/HSP production	Selvaraj et al. (2011)
COC050	*A. fastigiata*	Chlorophyll accumulation/HSP production	Selvaraj et al. (2011)
COC068	*A. fastigiata*	Chlorophyll accumulation/HSP production	Selvaraj et al. (2011)
ICGS 76	*A. hypogaea*	Chlorophyll accumulation/Acquired thermal tolerance	Selvaraj et al. (2011)
ICGS 44	*A. hypogaea*	High yield/HSP expression	Chakraborty et al. (2018)
ICG 8242	*A. hypogaea*	High Yield	Chakraborty et al. (2018)
796	*A. hypogaea*	Low relative injury and High Yield	Craufurd et al. (2003)
ICG 1236	*A. hypogaea*	Cardinal Temperature for pollen germination	Craufurd et al. (2003)
ICGV 86021	*A. hypogaea*	Crop growth rate, plant growth rate and partitioning	Craufurd et al. (2003)
ICGV 87281	*A. hypogaea*	Microsporogenesis, Flowering, Cellular membrane stability, Crop growth rate and Pod growth rate	Craufurd et al. (2003)
ICGV 92121	*A. hypogaea*	Microsporogenesis and Flowering	Craufurd et al. (2003)
SPT 06-07	*A. hypogaea*	Chlorophyll index, less membrane damage and pollen viability	Craufurd et al. (2003)
ICGV 97182	*A. hypogaea*	High stress tolerance index (STI) value	Akbar et al. (2017)
ICGV 01232	*A. hypogaea*	High STI value	Akbar et al. (2017)
ICGV 07013	*A. hypogaea*	High STI value	Akbar et al. (2017)
ICGV 07213	*A. hypogaea*	High STI value	Akbar et al. (2017)
ICGV 89280	*A. hypogaea*	High STI value	Akbar et al. (2017)
ICGV 00350	*A. hypogaea*	High STI value	Akbar et al. (2017)
ICGV 03057	*A. hypogaea*	High STI value	Akbar et al. (2017)
ICGV 06420	*A. hypogaea*	High STI value	Akbar et al. (2017)
ICGV 02266	*A. hypogaea*	High STI value	Akbar et al. (2017)
ICGV 03109	*A. hypogaea*	High STI value	Akbar et al. (2017)
ICGV 06099	*A. hypogaea*	High STI value and high kernel Fe- and Zn- content	Akbar et al. (2017)
ICGV 07273	*A. hypogaea*	High STI value	Akbar et al. (2017)
ICGV 00351	*A. hypogaea*	High STI value and drought-tolerant	Akbar et al. (2017)
ICGV 07268	*A. hypogaea*	High STI value	Akbar et al. (2017)
ICGV 06039	*A. hypogaea*	High STI value and Superior pod yield	Akbar et al. (2017)
ICGV 07148	*A. hypogaea*	High STI value	Akbar et al. (2017)
ICGV 03042	*A. hypogaea*	High STI value and Superior pod yield	Akbar et al. (2017)
ICGV 05032	*A. hypogaea*	High STI value	Akbar et al. (2017)
ICGV 07038	*A. hypogaea*	High STI value and Superior pod yield	Akbar et al. (2017)
ICGV 05155	*A. hypogaea*	High STI value	Akbar et al. (2017)
ICGV 06040	*A. hypogaea*	High STI value, Superior pod yield, and high kernel Fe- and Zn- content	Akbar et al. (2017)
ICGV 07012	*A. hypogaea*	High STI value and Superior pod yield	Akbar et al. (2017)
ICGV 06424	*A. hypogaea*	High STI value and Superior pod yield	Akbar et al. (2017)
ICGV 07246	*A. hypogaea*	High STI value and Superior pod yield	Akbar et al. (2017)

(Continued on following page)

TABLE 3 (*Continued*) Genotypes exhibiting heat stress tolerance along with its responsive trait.

Genotype	Subspecies	Heat stress tolerance responsive/tolerant trait	References
TG 37	*A. hypogaea*	High STI value	Akbar et al. (2017)
TAG 24	*A. hypogaea*	High STI value	Akbar et al. (2017)
ICG 4729	*A. fastigiata*	High yielding-High Temperature	Hamidou et al. (2013)
ICG 5236	*A. hypogaea*	High yielding-High Temperature	Hamidou et al. (2013)
ICG 12879	*A. hypogaea*	High yielding-High/moderate Temperature	Hamidou et al. (2013)
ICG 15042	*A. hypogaea*	High yielding-High Temperature	Hamidou et al. (2013)
ICG 862	*A. hypogaea*	High yielding-High/moderate Temperature	Hamidou et al. (2013)
ICG 1668	*A. hypogaea*	High yielding-High Temperature	Hamidou et al. (2013)
ICG 2925	*A. hypogaea*	High yielding-High Temperature	Hamidou et al. (2013)
ICG 8285	*A. hypogaea*	High yielding-High/moderate Temperature	Hamidou et al. (2013)
ICG 11219	*A. hypogaea*	High yielding-High Temperature	Hamidou et al. (2013)
Derived RILs from JL 24 × 55–437	*A. hypogaea*	Heat use efficiency, phenothermal indices	Sukanth (2022)

TABLE 4 Summary of available abiotic stress tolerant germplasm in cultivated peanuts.

Abiotic type	Tolerant germplam	Evaluated trait
Drought stress	PI502120, PI 493329, AU-NPL 17, TifNV-High O/L, Line-4, Line-8, Georgia 06, C76-16, AU16-28, AU18-35, SPT06-6, Tifrunner, and PI196635	Yield under stress, $\Delta^{13}C$, photosynthesis, and g_s
	#11, #34, #49, A596, Datangyou, Fenghua 1, Huayu 17, Huayu 21, Huayu 22, Huayu 25, Huayu 27, Ji 0212-4, Jihua 2, Jihua 4, L19, L121, L146, Luhua 14, NC6, Rugaoxiyangsheng, Shanhua 11, Tai 0125, Tai 0005, Taihua 4, Tangke 8, Xianghua 2008, Xianghua 55, Xuhua 13, Yuanza 9102, Yuanza 9307, Yueyou 7, Zhonghua 8	root depth, length, and density
	ICG 5891, ICG 6057, ICG 9777	pod yield and physiological traits
Heat stress	ICGVs 07246, 07012, 06039, 06040, 03042, 07038, and 06424	pod yield, hundred-seed weight, and pod growth rate under heat stress

genes related to HSP90, DREB2A, and *LEA4-2* were highly induced (Kokkanti et al., 2019) that can be used as markers for screening. Lipid peroxidation can cause severe membrane injury (MI) during abiotic stress. As such, it can be measured to assess the degree of stress in peanuts (Blum and Ebercon, 1981; Srinivasan et al., 1996; Bajji et al., 2002). Ribose, hydroxyproline, and saturated fatty acids were negatively correlated with MI, which can be used as stress tolerance parameters. Hence, there is a need to emphasize on the practical and robust screening methods to select for heat stress tolerance in peanut. Of late, studies have been carried out to utilize the thermal indices (growing degree days, phenothermal indices, heat use efficiency) for studying heat tolerance in peanut (Sukanth, 2022) and efforts are being made to map the heat tolerance related traits in groundnut (Sharma et al., unpublished).

Physiological basis of stress tolerance

Drought stress

Peanut shows different water needs at different developmental stages. The water demand is the highest at mid-pod filling stage because the peanut canopy covers all the ground and maintains open the stomata to maintain high photosynthesis to fill the growing pods (Stansell and Pallas, 1985; Rowland et al., 2012a; Rowland et al., 2012b). Understanding the main effects of drought on plant growth and yield may unfold the physiological basis of drought tolerance.

The drying of the soil due to drought and the subsequent reduction in leaf water potential and cell turgor leads to the inhibition of cell division and elongation that results in slower leaf growth rates aimed at reducing transpiration at the canopy level (Figure 2) (Ribaut et al., 2009; Avramova et al., 2015). To preserve water in the soil and maintain an acceptable leaf water potential, peanuts tend to decrease stomatal conductance (g_s) and transpiration resulting in reduced photosynthesis (Reddy et al., 2003; Pilon et al., 2018). Reduced leaf area expansion and lower photosynthesis per leaf area lead to a decline in canopy carbon assimilation that will reduce biomass accumulation and yield (Reddy et al., 2003). Plant traits that preserve soil moisture, such as high-water use efficiency (WUE) due to rapid stomatal closure, could increase drought tolerance (Devi et al., 2010; Devi and Sinclair, 2011; Shekoofa et al., 2015; Sinclair et al., 2017; Zhang et al., 2022). Contrarily, there are peanut cultivars that can maintain adequate plant water status and escape drought by collecting more water due to a more complex or deep root system (Rowland et al., 2012a; b;

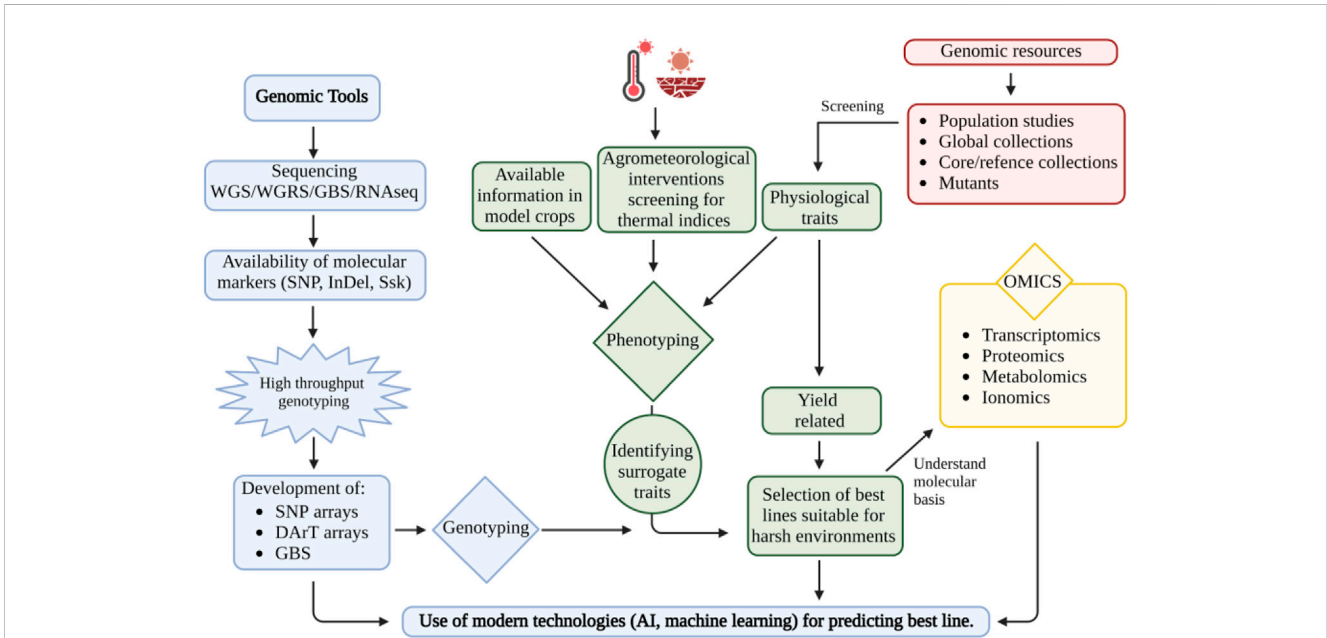


FIGURE 1
Utilization of genomic and genetic resources for developing peanuts for harsh environment conditions. This figure was created with BioRender.com.

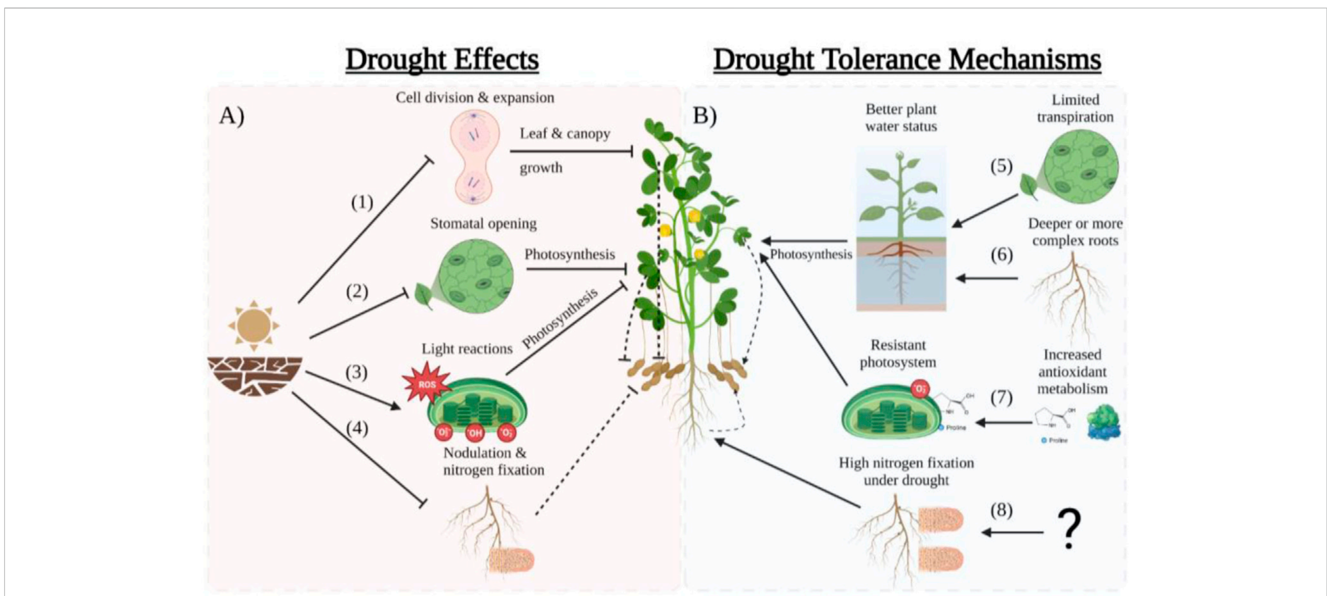


FIGURE 2
Scheme of drought effects **(A)** and tolerance mechanisms **(B)** in peanut. (1) Drought decreases the leaf water potential which inhibits cell division and expansion limiting leaf and canopy growth thus reducing pod yield. (2) The reduction of leaf water potential limits stomatal opening which reduces photosynthesis and therefore yield. (3) Drought produces reactive oxygen species (ROS) that damage the chloroplast membranes hindering the light reaction which result in decreased photosynthesis and yiled. (4) Drought inhibits nitrogen fixation in nodules by inhibiting the enzyme nitrogenase which reduces N availability resulting in lower yields. (5) Plants that reduces transpiration early in the drought period (water savers) are able to save water in the soil showing a better water status and therefore maintain moderate photosynthesis producing acceptable yields. (6) Plants with deeper or more dense root systems are able to extract more water to maintain good plant water status which allows the plant to photosynthetize more and maintain high yileds under drought. (7) Cultivars that produce more antioxidants such as proline are able to detoxify the ROS produced by drought maintaining a more healthy photosystems which results in higher photosynthesis and drought. (8) Cultivars that maintain high nitrogen fixation under drought are able to produce higher yields. However the underlying mechanisms of high nitrogen fixation under drought is unknown. This figure was created with BioRender.com.

Zhang, et al., 2022). Utilization of genomic and genetic resources for developing peanuts for harsh environment conditions are illustrated in Figure 1.

When drought is maintained for long periods, and CO_2 assimilation is reduced, the excess light not used for photosynthesis tends to produce reactive oxygen species (ROS) such as superoxide anion and hydrogen peroxide, H_2O_2 (Akcay et al., 2010; Laxa et al., 2019). ROS accumulation has been related to lipid peroxidation and thylakoid membrane damage (Lauriano et al., 2000; Quilambo, 2004). It decreases the effectiveness of Photosystems I and II (PSI and PSII) by producing non-stomatal limitations of drought that decrease photosynthetic efficiency resulting in reduced yield (Pilon et al., 2018). Peanut cultivars can tolerate these effects of drought by accumulating antioxidant substances that can reduce the accumulation of ROS to maintain higher photosynthetic rates during drought (Figure 2) (Akcay et al., 2010; Li et al., 2021).

Drought not only decreases plant growth and yield through a decrease in leaf and canopy photosynthesis (Reddy et al., 2003; Pilon et al., 2018) but also decreases nitrogen uptake by inhibiting the biological nitrogen fixation (BNF) in the roots (Pimratch et al., 2007). Nitrogen fixation is more sensitive to drought than photosynthesis because drought increases oxygen permeability to the nodule, inhibiting the activity of nitrogenase, the enzyme that catalyzes BNF (Parsons and Sunley, 2001). Drought also reduces the transport of the amino acid products of BNF from the nodule to the shoot, accumulates amino acids in the nodules, and creates a feedback mechanism that inhibits BNF (Peoples et al., 1986; Vessey et al., 2005). Thus, selecting genotypes with higher BNF under drought is another target to improve peanut yields (Sinclair et al., 1995; Sinclair, 2011; Devi et al., 2013). Peanut plants may therefore adapt diverse physiological attributes to balance productivity and stress tolerance as detailed herewith.

Limited Transpiration and High-Water Use Efficiency (WUE)

WUE (also referred to as transpiration efficiency, TE) is the amount of carbon assimilated as biomass per unit of water used by the crop (Medrano et al., 2015). When the atmosphere surrounding a plant gets drier, the plants tend to compensate by opening the stomata and increasing transpiration if they have enough water in the soil (Figure 2). This air drying can be simulated in a growth chamber by drying the air while measuring the plant transpiration by gravimetric methods (Devi et al., 2010). Under these conditions, some peanut cultivars can limit transpiration quicker than others when the environment gets drier by reducing their stomatal conductance, g_s (Devi et al., 2010; Devi and Sinclair, 2011; Shekoofa et al., 2015; Sinclair et al., 2017). This helps to save water in the soil that can "feed" the plant until maturity preserving biomass production and yield. This screening method has been used to select cultivars that later showed drought tolerance in field environments (Shekoofa et al., 2015). Reduced transpiration due to lower stomatal conductance maintains yield because of higher WUE. Vadez and Ratnakumar (2016) demonstrated that cultivars with high WUE can produce more yield under severe terminal drought conditions in a mini-lysimeter experiment under controlled field conditions. However, the high WUE trait may be a disadvantage under intermittent drought as the reduced stomatal conductance limits photosynthesis and biomass production compared with other crops that use more water (Blum, 2009; Polania et al., 2016). However, this yield penalty of high WUE cultivars have not been demonstrated until know and more research needs to be done in this area.

Effective Use of Water due to More Complex Root System

Effective use of water (EUW) refers to the amount of water that a plant can extract from the soil during the entire growing season and then use for transpiration, photosynthesis, biomass production and thus yield (Figure 2) (Blum, 2009). In common beans, cultivars with high EUW can maintain transpiration and photosynthesis for more time, resulting in higher yields under drought (Polania et al., 2016; Sanz-Saez et al., 2019). This mechanism has been detected and estimated by measuring the $\Delta^{13}C$ of the biomass and selecting for high $\Delta^{13}C$ in common beans (Farquhar et al., 1989; Polania et al., 2016; Sanz-Saez et al., 2019). Such genotypes use more water directly related to more profound or abundant root systems, as reported in common beans (White et al., 1990). Drought-tolerant peanut cultivars can exhibit high WUE or EUW. Peanut cultivars with high EUW do not show a yield advantage compared to high WUE under mid-season drought (Zhang et al., 2022). The high EUW capacity of these peanut cultivars has not yet been associated with a more profound or complex root system, as evidenced in the common bean (White et al., 1990). However, peanut cultivars with deeper or dense roots can extract more water to withstand drought (Songsri et al., 2008; Zhao et al., 2016). Zurweller et al. (2018) found that cultivars with more root development at deeper soil profiles (80 cm) do not result in drought tolerance in the mini-lysimeter environment. This conclusion may be affected by the fact that the roots are confined in a pot or only be relevant to the cultivars studied. For this reason, more research is needed to understand the role of different root morphological and anatomical characteristics on peanut drought tolerance.

Increased Antioxidant Metabolism to Reduce Adverse Effects of Reactive Oxygen Species (ROS)

Drought maintained for long periods can increase the production of ROS, which can damage proteins and lipids, ultimately reducing the efficiency of the photosynthetic system (Akcay et al., 2010; Laxa et al., 2019). To get protection from ROS, plants have evolved oxygen-scavenging systems consisting of non-enzyme antioxidant compounds such as proline, ascorbate, and glutathione and different antioxidant enzymes such as SOD, APX, CAT, POX, and GR (Bowler et al., 1992) (Figure 2). Drought-tolerant peanut cultivars showed high levels of CAT and APX that helped plants to decrease dangerous levels of H_2O_2. In contrast, high proline helped to maintain a higher osmotic potential to compensate for lower water potentials under drought (Akcay et al., 2010). In a more innovative approach, Banavath et al. (2018) produced a transgenic peanut line overexpressing a homeodomain-leucine zipper transcription factor (*AtHDG11*) which showed increased photosynthesis under drought conditions, probably due to more active antioxidant metabolism that reduces the ROS damage. In the U.S., the peanut industry does not encourage transgenic approaches as peanut is mostly used for

human consumption, and transgenic food crops do not have high consumer approval. Thus, screening of diverse lines with high antioxidant activity is needed to find and introgress genotypes that are tolerant to drought and produce high levels of ROS (Mittler, 2002).

Maintaining High Biological N_2-fixation (BNF) under Drought

Maintaining high BNF under drought has been documented as a tolerant trait for legumes (Figure 2). Using different physiological techniques, crop physiologists and breeders have been able to introgress this trait in soybean elite lines that resulted in commercial cultivars with high BNF and yield under drought (Sinclair, 2000; Chen et al., 2007; King et al., 2014). In peanuts, it has been also demonstrated that cultivars that maintain high BNF accumulate more biomass resulting in higher yields (Sinclair et al., 1995; Devi et al., 2013). There is a very little research in the literature that focuses on understanding the underlying mechanisms regulating nitrogen fixation under drought conditions for peanuts and therefore we do not know why these cultivars show high nitrogen fixation under drought. However there have been some new efforts to understand the regulation involved in the nodulation of peanuts (Peng et al., 2017, 2021). For example, there have been no reports in the literature in which a sizeable number of peanut cultivars have been screened for BNF and then introgressed in elite lines as in soybean (Sinclair, 2000; Chen et al., 2007; King et al., 2014). This is partly because determining BNF *in-situ* in the field is very difficult and costly. In soybean and common bean, tolerance of BNF to drought has been screened in diverse populations using the ^{15}N natural abundance method to find new breeding lines (Steketee et al., 2019; Oladzad et al., 2020). With the discovery of non-nodulating peanut lines (Peng et al., 2021), using the ^{15}N natural abundance method should facilitate the screening of diverse peanut populations under well-water and drought conditions to delineate genomic regions responsible for the maintenance of N_2-fixation under drought. Such an approach could also result in the discovery of lines with high N_2 fixation under well-water and drought conditions for use in introgression breeding programs in peanuts, as has been done in soybean.

Interaction between Drought Tolerant Traits

To our knowledge, no publications focus on understanding if there is any relationship between the drought, as mentioned earlier, tolerance mechanisms in peanuts. For BioRender obj example, cultivars with limited transpiration and high WUE, as they maintain a good water status in the plant, will probably show higher BNF as the plant is not suffering as much drought stress. This high BNF under drought is not a sign of direct tolerance caused by a more resistant nitrogenase activity to drought but the consequence of maintaining a better water status. Another example is the maintenance of a better antioxidant metabolism; the events that improved the antioxidant quality in transgenic plants also improved the water status of the plant by increasing WUE (Banavath et al., 2018). In this case, it is unclear if the overexpression of *AtHDG11* improves the antioxidant status of peanuts and then water status or *vice versa*. For these reasons, experiments that aim to separate between different drought-tolerance mechanisms would be important to identify parental lines that can introgress different drought-tolerant traits into breeding programs.

Heat stress

Higher temperatures can disrupt the physiological processes in plants, including a reduction in the rate of photosynthesis, degradation of chloroplast proteins, damage to PSII, lower relative water potential, ROS accumulation, and increase in lipid peroxidation (Crafts-Brandner and Salvucci, 2002; Dutta et al., 2009; Hasanuzzaman et al., 2020a). Heat stress affects the growth of male and female reproductive organs by impairing pollen tube growth, pollen viability, germination, egg viability, and fertilization (Figure 3). Microsporogenesis (3–6 days before flowering) and fruit set are two critical stages of peanut development which is affected by high temperatures (Craufurd et al., 2002, 2003). The late flowering to early seed setting stage was observed to be highly susceptible to high temperatures in peanuts (Prasad et al., 1999). However, the time of flower initiation at temperatures higher than 40°C/28°C day and night is the primary determinant of pod number in peanuts (Craufurd et al., 2000).

The impact of the elevated temperature is devastating as it affects membrane stability, inactivates chloroplast and mitochondrial enzyme function, causes protein degradation, reduces carbon metabolism, and alters cytoskeleton organization (Bita and Gerats, 2013) (Figure 3). Under heat stress, the thylakoid membrane and photosystem II (PSII) are severely damaged, disrupting the electron transport system and ATP synthesis during photosynthesis (Wang Q. L. et al., 2018). At temperatures higher than 38°C/32°C, the leaf chlorophyll content is reduced, which leads to less photosynthesis and low sugar content (Liu and Hang, 2000). Heat stress also affects the water status in the plant cells due to osmotic perturbation caused by low photosynthetic capacity, reduced sugar content, and higher transpiration rate (Hemantaranjan et al., 2018). High temperature under limited water conditions causes a reduction in relative water content and total water absorption rate, which contributes to total yield loss (Ashraf and Hafeez, 2004) (Figure 3). Stomatal conductance (g_s) is directly correlated to the rate of photosynthesis and transpiration rate. During heat stress, stomatal density and stomatal conductance were found to be decreased in susceptible peanut varieties (Dash et al., 2020).

The efficient partitioning and allocation of carbon assimilate and photosynthates from source to sink are essential for plant growth and yield. Heat stress reduces carbon assimilation and partitioning and affects the plant source-to-sink relationship. Seed weight and yield depend on net photosynthesis and re-translocation of water-soluble carbohydrates from vegetative organs during seed filling (Fischer, 2011). The partitioning of dry matter allocation to shell relative to seed was higher in peanuts during higher temperatures. Higher temperature affects the photosynthates partitioning to the pod leading to low pod yield. The tolerant peanut genotypes were found to have higher crop growth rate and pod growth rate under heat-stress conditions than in the non-stress environment. Further,

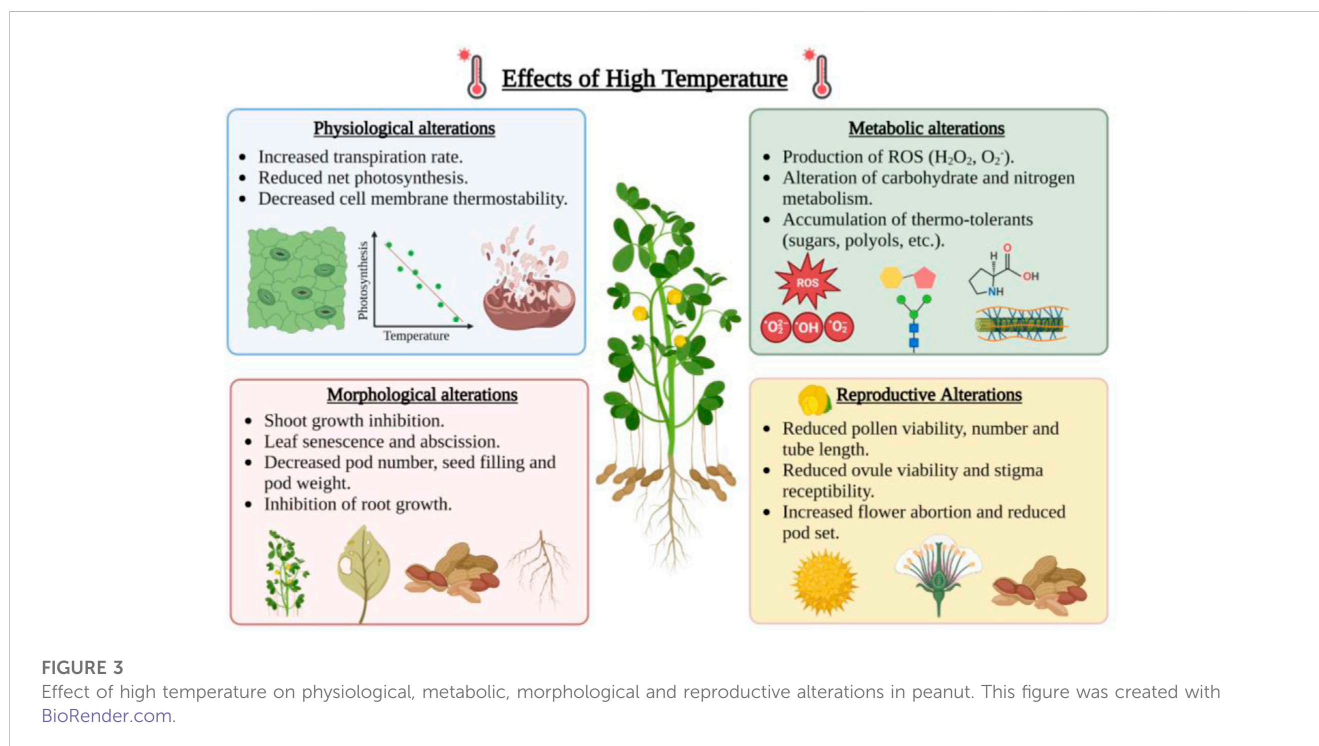


FIGURE 3
Effect of high temperature on physiological, metabolic, morphological and reproductive alterations in peanut. This figure was created with BioRender.com.

the photoperiod × temperature interaction influences the partitioning of dry matter into pods in peanuts (Nigam et al., 1994; Nigam et al., 1998; Akbar et al., 2017). Heat stress also damages the enzymes involved in nutrient metabolism and disrupts nutrient acquisition (Hungria and Kaschuk, 2014) (Figure 3).

Plants can tolerate the adverse effects of heat stress to some extent by modifying physiological and biochemical processes such as solute accumulation, osmotic adjustment, cellular homeostasis maintenance, and redox balance (Janská et al., 2010). One of the strategies for heat-stress tolerant cultivars is to have higher ceiling temperatures for pollen germination. Since heat stress significantly impacts peanut flowering, genotypes with a higher pollen germination ceiling temperature tend to be heat tolerant (Kakani et al., 2002; Chaudhary et al., 2022). Another effect of heat tolerance on peanuts or any other crop is membrane injury; increased unsaturated fatty acid levels caused by heat stress can disrupt membrane permeability, causing disrupted cellular homeostasis (Marcum, 1998). Photochemical changes during photosynthesis and reactive oxygen species (ROS) production due to heat stress can also affect membrane integrity (Bita and Gerats, 2013), causing membrane leakage. The heat-tolerant peanut varieties can be screened using an electrolyte leakage test or membrane injury test (MIT) by evaluating cell membrane thermostability (CMT) (Lauriano et al., 2000; Yeh and Lin, 2003; Akcay et al., 2010). Craufurd et al., 2002 reported the lower membrane injury in heat-tolerant peanut varieties. The higher heat tolerance is observed to be associated with higher gs value, photosynthesis rate, and stomatal conductance (Awasthi et al., 2014). The higher carotenoid content helps in preventing photo-oxidation of chlorophyll during stress. A study showed the heat tolerance peanut cultivars exhibited the higher carotenoid and higher chlorophyll level in the leaf (Dash et al., 2020). The protective mechanism of heat tolerance is associated with the activation of enzymatic and non-enzymatic ROS scavenging compounds. The higher activity of antioxidants such as SOD (Super oxide dismutase), CAT (Catalases), APX (Ascorbate peroxidase), and GR (Glutathione reductase) has been reported to provide plant thermostability (Kumar et al., 2013). Higher levels of total soluble sugar improve the thermotolerance of legume plants primarily by increasing water relations and gas exchange activities, thereby enhancing vegetative and reproductive growth (Ahmad et al., 2021a). Phytohormones such as abscisic acid (ABA) play a vital role in the stress tolerance of the plant by regulating its physiological processes. Plant growth regulators such as ABA, TU (Thiourea), GABA (Gamma-Amino butyric acid), and brassinosteroids help in enhancing the accumulation of soluble sugar, osmoprotectants, antioxidant enzymes, and gas exchange traits during heat stress tolerance (Ahmad et al., 2021b). Further, heat shock proteins (HSPs) play a crucial role in thermotolerance by maintaining protein structure and membrane integrity. Gene expression profile revealed that HSPs and heat shock factors (HSFs) are involved in tolerance during heat stress in legume crops (Zhang et al., 2015; Ma et al., 2016).

Genes/transcription factors that have targeted in model/crop plants for developing tolerant plant against heat/drought

Functional genomics and biotechnological techniques have been a valuable tools to identify and characterize genes associated with agronomic traits for the crop improvement. Differentially expressed genes related to storage proteins, fatty acid metabolism, oil production, biotic stress, *etc.*, have been identified and cloned using EST sequencing for the improvement of peanut variety (Jain et al., 2001; Chen et al., 2012). Candidate genes/QTLs for

thermotolerance such as HSPs, compatible osomoprotectants, antioxidants, *etc.*, have been identified which can be used to develop heat tolerant crops using transgenic approach (Chaudhary et al., 2022). An *E. coli* gene encoding trehalose-6-phosphate synthase/phosphatase (TPSP) overexpressed in tomato induced the expression of HsfA1, HsfA2, and HsfB1which further upregulated the heat tolerant HSPs, i.e., Hsp17.8, ER-sHsp, and Mt-sHsp (Lyu et al., 2018). Further, the overexpression of CaHsp25.9 in Capsicum improved the thermotolerance and increased the proline and SOD content in transformed lines (Feng et al., 2019). Pea plant overexpressing the heat shock factor gene HsfA1d from A. thaliana showed the higher activity of proteins related to antioxidative pathways such as SOD and APX activity and lower H_2O_2 accumulation during heat stress which further enhanced the thermotolerance of plant (Shah et al., 2020). In another study, the transgenic tomato overexpressing the cAPX helped in increasing the tolerance against heat stress (Wang et al., 2006). ROS generation disturbs the cellular processes during the heat stress. The activation of HSPs/HSFs involved in lowering the ROS accumulation *via* ROS scavenging pathway. The heat shock transcription factors HsfA2 and HsfA4 act as a H_2O_2 sensor and involved in the regulation of genes associated with ROS mitigation (Scarpeci et al., 2008). Sakuma et al., 2006 reported that the overexpression of a constitutive active form of transcription factor DREB2A CA induces drought-, salt-responsive as well as HSPs-related genes in *Arabidopsis* and provides significant tolerance to heat and water stress.

Sources of variation for drought and heat stress tolerance

Cultivated genepool

Plant genetic resources are the basic raw materials to empower crop improvement programs. The peanut gene pool consists of cultigen (*Arachis hypogaea*) with its many landraces, improved cultivars, and more than 83 wild species of the genus *Arachis* (Gregory et al., 1980). Cultivated peanut is an allotetraploid (2n = 4× = 40) that originated from natural hybridization of two diploid species, *Arachis duranensis* (A-genome) and *Arachis ipaensis* (B-genome) followed by spontaneous polyploidization of the hybrid and its subsequent domestication by Neolithic proto-farmers 6-10,000 years ago (Bertioli et al., 2016). Cultivated peanut germplasm is classified into two main subspecies: *A. hypogaea* and *A. fastigiata*. The subsp. *hypogaea* contains two botanical varieties: *hypogaea* (Virginia type) and *hirsuta*, while subsp. *fastigiata* contains four botanical varieties: *fastigiata* (Valencia type), *peruviana*, *aequatoriana*, and *vulgaris* (Spanish type). All six botanical varieties have unique morphological characteristics that separate them from one another (Krapovickas et al., 1994). Worldwide over 15,000 peanut accessions are preserved in the national and international genebanks, including 1823 accessions in N.I. Vavilov Research Institute of Plant Industry, Russia; 14,320 in ICRISAT, India; 7,432 in USDA, Griffin; and 9103 in China (Pandey et al., 2012). Assessment of such a large gene pool for agronomically beneficial traits is economically not feasible and also subject to high genotype by environment interaction. Smaller subsets representing a diversity of the entire collection of given species preserved in a genebank are ideal resources to evaluate for agronomic and stress tolerance traits.

Reduced subsets in the form of core (Frankel, 1984) and mini core (Upadhyaya and Ortiz, 2001) collections provide smaller subsets of germplasm that could be used to mine valuable traits from entire germplasm collections more efficiently instead of screening germplasm as a whole collection. Several such sets are reported for efficient utilization of genetic resources in peanut breeding and genetics (Ding et al., 2022; Dwivedi et al., 2008; Holbrook and Stalker, 2003; Holbrook and Dong, 2005; Upadhyaya et al., 2002a and; 2002b). The U.S. peanut mini-core collection has been effectively used for the identification of interesting alleles and traits for breeding programs for traits related to drought tolerance (Wang H, et al., 2016; Wang M. L. et al., 2016; Zhang et al., 2019; Li et al., 2022; Patel et al., 2022; Zhang et al., 2022). End-of-season drought tolerance was reported in 15 accessions after evaluating ICRISAT peanut mini core collection and selected based on pod yield, SPAD, and SLA measurements (Upadhyaya, 2005).

Assessment of 150 peanut genotypes under rainout shelters showed significant differences in pod yield, relative water content (RWC), SLA, leaf dry matter content (LDMC), chlorophyll fluorescence (CHF), $\Delta^{13}C$, photosynthesis, and stomatal conductance (g_s), and resulted in 13 accessions resistant to midseason drought stress. In addition, gas exchange parameters were measured regularly during the drought and recovery to monitor dynamic changes in photosynthesis and g_s under stress. Genotypes with high yield, $\Delta^{13}C$, photosynthesis, and g_s under stress were classified as water spenders while genotypes with equally high yields but with low $\Delta^{13}C$ and g_s and moderate photosynthesis under drought stress were classified as water savers (Zhang et al., 2022). The previous reports on screening US peanut mini core collection across three irrigation treatments over 2 years and two field locations unfolded five accessions (PI 502120, PI 493329, Line 8, Georgia-06G, AU-NPL-17) as resistant to drought. These accessions had high SPAD, flowering, and paraheliotropism (Selvaraj et al., 2009; Belamkar, 2010). Paraheliotropism refers to condition in plants wherein the plants orient their leaves parallel to incoming rays of light. Elsewhere several germplasms tolerant to drought were reported, which showed significant differences in root depth, length, and density. The tolerant germplasms (#11, #34, #49, A596, Datangyou, Fenghua 1, Huayu 17, Huayu 21, Huayu 22, Huayu 25, Huayu 27, Ji 0212-4, Jihua 2, Jihua 4, L19, L121, L146, Luhua 14, NC6, Rugaoxiyangsheng, Shanhua 11, Tai 0125, Tai 0005, Taihua 4, Tangke 8, Xianghua 2008, Xianghua 55, Xuhua 13, Yuanza 9102, Yuanza 9307, Yueyou 7, Zhonghua 8) display increased root to shoot ratio and the enhanced root length and density, particularly in the deep soil, in comparison to those grown under normal growth conditions. The dragon-type (runner) peanuts, such as 'A596' and 'Rugaoxiyangsheng,' were more tolerant, followed by Virginia, Spanish, intermediate, and Valencia peanuts (Yang et al., 2019).

A screen of 247 ICRISAT accessions tested under water-deficit environments in Africa and India identified a few most adapted genotypes [ICG 5891, ICG 6057, ICG 9777] across moderate- and high-temperature stressed environments. It showed significant genotype × environment interaction (Hamidou et al., 2012). Field assessment of advanced breeding lines under irrigated conditions

during the postrainy season in heat-stressed (air temperature 35°C and above during flowering) and non-stressed (air temperature below 35°C irrigated) environments unfolded large variability for pod yield and physiological traits among genotypes in a heat stress environment. A pod yield reduction of 1.5%–43.2% was recorded under a heat-stressed climate. The genotypes under heat stress either yield poorly stable or increased pod yield under high-temperature stress. The heat-tolerant genotypes are identified based on the stress tolerance index. GJG 31, ICGV 87846, ICGV 03057, ICGV 07038, and GG 20 showed an increase in pod yield of 9.0%–47.0% at high temperatures, with a 0.65%–3.6% increase in pod growth rate. ICGV 06420, ICGV 87128, ICGV 97182, TCGS 1043, and ICGV 03042 were stable for pod yield and recorded a 0.25%–3.1% increase in pod growth rate. Thus, pod yield, hundred-seed weight, and pod growth rate under heat stress can be used to select heat-stress tolerant genotypes. Based on stress tolerance indices and pod yield performance, ICGVs 07246, 07012, 06039, 06040, 03042, 07038, and 06424 were identified as heat-tolerant genotypes and are used as parents in breeding programs in India (Akbar et al., 2017).

Developing reliable indices and traits for acquired thermotolerance in peanuts is necessary for breeding heat-tolerant varieties. Several reports observed the genotypic variability in peanut's heat tolerance for partitioning dry matter to pods and kernels, fruit set, membrane stability, and chlorophyll fluorescence (Srinivasan et al., 1996; Vara Prasad et al., 2001; Craufurd et al., 2002).

Wild *Arachis* species

Wild *Arachis* species originated in South America, selected during evolution in a range of environments and biotic stresses, which provided opportunities for the evolution of a rich source of allele diversity for resistance to several pests, including fungal diseases and drought tolerance (Bertioli et al., 2021). Thoppurathu et al. (2022) conducted A transcriptome analysis for *A. duranensis* (drought tolerant) and *Arachis stenosperma* (drought susceptible) revealed *A. duranensis* genotype had a higher number of transcripts related to DNA methylation or demethylation, phytohormone signal transduction and flavonoid production, transcription factors, and responses to ethylene, indicating that it is tolerant to drought stress. Exposing A and B genome diploids under progressive drying to examine curves of vapor pressure deficit (VPD) against a fraction of transpirable soil water (FTSW) revealed that *A. magna* and *A. duranensis* accessions had superior ability to regulate transpiration under water deficit stress (Leal-Bertioli et al., 2012).

In summary, although we have large number of accessions of peanut germplasm collection, a very limited numbers have been identified as abiotic stress tolerant genotypes. Some of those tolerant genotypes have been used in breeding programs, and others are available for further investigation as shown in Table 4.

Molecular basis of stress tolerance

Abiotic stress tolerance is a complex phenomenon involving several small effect genes and their interaction with the environment. Understanding the molecular mechanisms of stress perception and adaptive/tolerance responses by the plants is essential for engineering crop plants for stress tolerance (Aravind et al., 2022).

Drought stress

Genomic approaches to study drought tolerance

With the availability of several sequencing platforms, it is possible to detect many sequence variations. The most abundant markers available are single nucleotide polymorphisms (SNPs) obtained by several genotyping platforms. SNP markers are extensively used to assess crops' genetic diversity and trait mapping, For high resolution trait mapping, a high-density SNP genotyping array having uniform genome coverage is required. Large genome size and low genetic diversity in the cultivated gene pool of peanuts driven the development of SNP arrays for high throughput genotyping. The availability of the Axiom_*Arachis* array of highly informative genome-wide SNPs, 58,233 SNPs after sequencing 41 diverse genotypes, allows for the generation of high throughput genotyping data to capture genetic diversity, high-resolution trait mapping and improve breeding efficiency (Pandey et al., 2017). Similarly, another 48K SNP array called "Axiom Archis2" was developed in which 1,674 haplotype-based SNP markers were included from 21 re-sequenced peanut accessions (Clevenger et al., 2018).

58K "*Axiom_Arachis*" array based bi-parental QTL mapping detected sixteen major main-effect QTLs (10.0%–33.9% PVE) for traits associated with drought tolerance, wherein the significant QTLs were detected for haulm weight (20.1% PVE) and SCMR (22.4% PVE) (Pandey et al., 2020). This study was also successful in identifying important candidate genes underlying these QTL regions, such as those encoding *glycosyl hydrolases*, *malate dehydrogenases*, *microtubule-associated proteins*, *transcription factors* such as *MADS-box*, *basic helix-loop-helix* (*bHLH*), *NAM*, *ATAF*, and *CUC* (*NAC*), and *myeloblastosis* (*MYB*).

Earlier literature reported 52 main-effect QTLs (M-QTLs) for nine different traits under two water regimes, accounting for low <12% PVE (Faye et al., 2015), while another study detected 153 main effect QTLs and 25 epistatic QTLs with low to moderate phenotypic variance for drought tolerance traits (Gautami et al., 2012). As the QTLs obtained for drought tolerance (Table 5) were showing low to moderate PVE, the attempts were not made to introgress the QTL regions for breeding for drought tolerance in groundnut.

Various genome-wide association studies (GWAS) are noteworthy for detecting significant associations for yield-related traits under abiotic stress. GWAS analysis performed in a reference set of peanuts reported a total of 152 significant marker-trait associations (MTAs) for different traits under well-watered (WW) and drought stress (DS) conditions. Eighty-four significant MTAs were detected under WW, explaining 8.83%–88.90% phenotypic variance, and sixty-eight significant MTAs were detected under DS, accounting for 8.24%–90.09% phenotypic variance (Pandey et al., 2014). A study involving ICRISAT peanut mini core collection (Upadhyaya et al., 2002) and four physiological traits (leaf area index, canopy temperature, SCMR, NDVI) resulted in 20 significant MTAs for the target traits, with 20% PVE for SCMR (Shaibu et al., 2020), while another study involving 58K Axiom_Arachis array data on 453 peanut accessions reported seven significant MTAs on six chromosomes and SNP AX-176820297 on Araip. B05 was associated with leaf chlorophyll content across the seasons. The gene Arahy. SDG4EV was found to be related to leaf chlorophyll content (Zou et al., 2022). The chlorophyll content is a critical component affecting photosynthesis

in plants and is associated with abiotic stress adaptation (Singh and Thakur, 2018).

Transcriptomic approaches to study drought tolerance

Transcription factors (TFs) play a major role in abiotic stress adaptation. The expression of certain TFs is regulated by hormonal signals that trigger the expression of several stress-responsive genes. During drought stress, the abscisic acid (ABA), ethylene, and salicylic acid signaling pathways are induced to produce osmoregulatory substances to maintain ROS homeostasis in the plant cells (Miller et al., 2010). The osmolytes and osmoprotectants play a key role in protecting the plant cells by scavenging free radicals. Mannitol, an osmoprotectant, plays an important role in scavenging hydroxyl radicals generated during abiotic stresses. Dehydration responsive element-binding (DREB) TFs enhance plant tolerance to abiotic stresses by specifically binding dehydration response element/C-repeat (DRE/CRT) cis-elements to control downstream gene expression (Liu et al., 1998). A study of the transcriptome of genotypes that show differential behavior during drought stress could provide insights into the molecular mechanisms of stress tolerance. A genome-wide transcriptome study in the peanut genotypes C76-16 (drought tolerant) and Valencia-C (drought sensitive) using RNA-sequencing have revealed the activation of key genes involved in ABA and sucrose metabolic pathways during moisture-stress conditions. The differentially expressed genes (DEGs) under water deficit conditions include *Gcn5-related N-acetyltransferase* (*GNAT*), *BON1- associated protein*, *the lateral organ boundary* (*LOB*), and the *late embryogenesis abundance* (*LEA*), *etc*., that are involved in the synthesis of osmoprotectants, photosynthates, abscisic acid, secondary metabolites, *etc.* (Bhogireddy et al., 2020). Another comparative transcriptome analysis in two peanut cultivars- NH5 (tolerant) and FH18 (sensitive), under drought stress, has indicated DEGs involved in pathways of GSH-related glutathione metabolism, glycolysis, glyoxylic acid, dicarboxylic acid ester metabolism, ABA and SA signal-transduction, ROS-scavenging, proline metabolism, cell wall sclerosis-related, and cutin and wax metabolism (Jiang et al., 2021). In the combined omics approach, the transcriptome and proteome of *Arachis duranensis*, the "A genome proginator" of cultivated peanut, was studied in water deficit conditions (Carmo et al., 2018). This study showed the downregulation of expression of certain genes [*Cht2*, *MLP-34*, *heat shock proteins* (*HS70*, *HS17.3*), *DOT-1,* and *MatK*] in the stressed root tissues. The information can further be utilized using appropriate genomics or transgenic/genome editing approaches to improve cultivated peanuts for drought tolerance.

Heat stress

Genomics approaches to study heat tolerance

A bulk segregant analysis using single marker analysis (SMA) on a mapping population (Tamrun OL01 × BSS 56) showed eight marker-trait associations with 9.19%–17.69% PVE (Selvaraj et al., 2009). Preliminary studies on single marker analysis using mapping population JL 24 × 55–437 have suggested that the traits like heat use efficiency, thermal indices, specific leaf area, membrane injury indices can be used as surrogate traits for selecting heat tolerant genotypes (Aravind, 2021; Sukanth, 2022). Of late, there are efforts to identify QTLs for high-temperature tolerance related traits (Sharma et al., unpublished).

Transcriptomic approaches to study heat tolerance

The membrane stability during stress determines the heat tolerance level. With their chaperon activity, the heat shock proteins (HSPs) help the cells to tolerate heat stress by protecting essential enzymes and nucleic acids from denaturation and misfolding due to high temperature (Jain, 2000). Rapid induction of small HSPs was found during high-temperature stress conditions in peanuts and imparts physiological adaptation to heat stress (Chakraborty et al., 2018). A genome-wide analysis of HSFs using the genomic sequences of wild peanut ancestors, *A. duranensis* and *A. ipaensis*, detected sixteen orthologous pairs of highly syntenic Hsfs, clustered into three groups, between AA and BB genomes. These HSFs were also shown to have fungal elicitor-responsive elements that showed differential expression in cultivated peanuts under abiotic stress and *Aspergillus flavus* infection (Wang et al., 2017). The reproductive parts of peanuts are most affected during heat stress. A lipidome study on peanut anthers revealed that phosphatidylcholine (PC), phosphatidylethanolamine (PE), and triacylglycerol (TAG) lipid species contributed towards more than 50% of total lipids both in ambient and heat stress conditions. A recent study involving another lipidome reports a decrease in unsaturated lipid species containing 18:3 fatty acid and downregulation of the fatty acid desaturase 3-2 gene (FAD3-2) in peanuts under high temperatures (Zoong Lwe et al., 2020). A similar report also indicated the possibility of utilizing the information on membrane lipid unsaturation as an indicator of heat tolerance in soybean and peanuts (Rustgi et al., 2021).

Genes related to HSP90, dehydration-responsive element binding -2A (DREB2A), and late embryogenesis abundant 4-2 (LEA4-2) were highly induced during heat stress in a few peanut genotypes (Kokkanti et al., 2019). Heat stress generates ROS, such as superoxide radicals and H_2O_2, which results in oxidative damage to biomolecules, lipid peroxidation, and reduced activities of ROS-scavenging enzymes (Dat et al., 2000). ROS signaling is linked to the activation of heat shock factors (HSFs) and heat shock proteins (HSPs) (Driedonks et al., 2015). The transcript and biochemical analysis demonstrated the higher expression and activities of gene encoding ascorbate peroxidase (APX), superoxide dismutase (SOD), and glutathione reductase (GR), whereas catalase (CAT) expression declined during heat stress combined with salinity and drought stress in peanut. An increase in lipid peroxidation was also observed during heat stress in peanuts (Patel et al., 2022).

Studies on heat tolerance in peanut is limited to physiological aspects and there is need to look into the molecular basis of heat tolerance. There is need to identify molecular markers and candidate genes with high PVE, that can potential be used in genomics-assisted breeding for abiotic stress tolerance.

Combining stress tolerance, productivity, and nutritional quality through plant breeding and biotechnological-assisted approaches

Plant breeding and selection

Breeding for drought and heat stress tolerance is extremely challenging due to the complexity associated with various stress-adaptive mechanisms, uncertainty in the onset and intensity of stress, and large genotype × environment interactions. Conventional crossing

TABLE 5 QTL and marker-trait association studies carried out for drought and heat tolerance in peanut.

S. No.	Abiotic stress	Mapping population	Molecular markers	Traits studied	QTLs/MTAs identified	References
1	Drought stress	TAG 24 × ICGV 86031	SSR	• Transpiration • Transpiration Efficiency • Specific Leaf Area • SCMR • Canopy conductance • Carbon discrimination ratio • Yield parameters	105 main effect QTLs with 3%–33% PVE (few major QTLs)	Ravi et al. (2011)
2	Drought stress	ICGS 76 × CSMG 84-1 ICGS 44 × ICGS 76	SSR	• transpiration efficiency • transpiration • SCMR • shoot dry weight • Yield parameters	153 Main effect QTS (No major QTLs)	Gautami et al. (2012)
3	Drought stress	300 genotypes (ICRISAT References collection)	DArT	• 50 agronomic traits including drought related traits	GWAS 84 MTAs under well-watered (8.83%–88.90% PVE) and 68 MTAs under drought stress (8.24%–90.09% PVE)	Pandey et al. (2014)
4	Drought stress	TAG 24 × ICGV 86031	SSR	• Yield parameters • SCMR	52 QTLs with <12 PVE%	Faye et al. (2015)
5	Drought stress	TAG 24 ICGV 86031	Axiom_Archis Array (SNP)	• Transpiration • Transpiration Efficiency • Specific Leaf Area • SCMR • Canopy conductance • Carbon discrimination ratio • Water use efficiency • Leaf area • Yield parameters	19 Major Main-effect QTLs with 10%–33.0% PVE	Pandey et al. (2020)
6	Drought stress	125 genotypes	GBS DArTseq (SNP)	• leaf area index • canopy temperature • SCMR • NDVI	GWAS 20 significant MTA (6.6–20.8%PVE)	Shaibu et al. (2020)
7	Drought stress	453 genotypes	Axiom_Arachis array (SNP)	• Leaf Chlorophyll Content	GWAS 5 MTA	Zou et al. (2022)
8	Heat stress	Tamrun OL01 × BSS 56	SSR	• Yield parameters • Pod and kernel traits • Oil content	Single marker analysis (SMA) on Bulks (8 MTAs with 9.19%–17.69% PVE)	Selvaraj et al. (2009)

(Continued on following page)

TABLE 5 (*Continued*) QTL and marker-trait association studies carried out for drought and heat tolerance in peanut.

S. No.	Abiotic stress	Mapping population	Molecular markers	Traits studied	QTLs/MTAs identified	References
9	Heat stress	JL 24 × 55–437	Transposable elements (AhTE) and SNP (GBS)	• Yield parameters • Thermal indices • Membrane injury indices • Heat Use Efficiency • Phenological parameters	SMA (39 marker-trait association with 2.19%–5.78% PVE)	Sukanth (2022)

and selection to improve drought and heat stress tolerance have been met with limited success. The focus on yield, lack of simple physiological traits as a measure of tolerance, and complex inheritance (polygenes with small effects) contributed to the low genetic gain in stress tolerance breeding. A better understanding of the physiological basis of stress tolerance contributes to identifying and manipulating traits associated with yield in water-deficit stressed field environments (reference needed). A systematic characterization of germplasm and breeding lines resulted in genetically diverse germplasm varying in response to drought and heat stress (Table 5). Such genetic stocks showed large variations for physiological traits such as SLA, chlorophyll content, amount of water transpired, WUE and harvest index in a drought-stressed environment. Both empirical (yield in the stressed environment) and trait-based selection (SLA, SCMR) have led to the development and release of drought-tolerant peanut cultivars in India and Australia (Rachaputi, 2003). The trait-based selection, however, did not show a consistent superiority over the empirical selection for drought tolerance (Nigam et al., 2005). Integrating physiological traits or their surrogates in the selection scheme would be advantageous in selecting segregants that utilize water more efficiently and partition photosynthates more effectively into economic yield. A few drought-tolerant cultivars of wheat bred by trait-based breeding without any yield penalty have been released in Australia where drought is intense and terminal (Rebetzke et al., 2002; Richards et al., 2002; Condon et al., 2004). A combination of trait-based selection in an early stage of breeding and yield assessment at a later stage of cultivar development in target environments is needed to select for abiotic stress adaptation and yield in peanuts. High yield potential and higher resistance are difficult to target together; therefore, to avoid the yield penalty, cultivars with high yield potential were bred to moderate levels of stress tolerance (Nigam et al., 1991). A few commercial heat-tolerant varieties have been released through conventional breeding, such as 55-437, 796, ICG 1236, ICGV 86021, ICGV 87281, and ICGV 92121 (Craufurd et al., 2003).

Introgression breeding using Wild *Arachis* species and their derivatives

Crop wild relatives are the source of variation for stress tolerance and productivity traits. Advanced backcross populations originating from synthetic amphidiploid as donors for wild alleles detected several QTLs with positive effects on pod/seed size and adaptation traits in water-limited environments (Essandoh et al., 2022). Progenies derived from a cross between synthetic allotetraploid (*A. duranensis* × *A. batizocoi*) and cultivated peanut improved photosynthetic traits and yield under water-deficit stress (Dutra et al., 2018). A root transcriptomic study involving drought tolerant (*A. duranensis*) and susceptible (*Arachis stenosperma*) wild species unfolded 1465 differentially expressed genes (DEGs) under drought stress and 366 polymorphic SNPs among DEGs. Three SNPs differentiated the two species and may be useful for selecting drought-tolerant lines (Thoppurathu et al., 2022). In addition, advanced backcross populations involving several synthetics (ISATGR 121250, ISATGR 278-18, ISATGR 265-5, ISATGR 40) and peanut cultivars (ICGV 91114, ICGV 87846, TMV 2, Tifrunner) exhibited considerable variability for morpho-agronomic traits (Sharma, 2017). They could be a good resource for screening for abiotic stress adaptation and use in peanut breeding. Synthetics are tetraploid derivatives originating from a cross between two diploid *Arachis* species from secondary gene pool with different genomes (Mallikarjuna et al., 2012).

Genomics-assisted breeding

Advances in peanut genomics unfolded several QTLs and markers (SSRs, SNPs) and candidate genes associated with drought tolerance surrogate traits in peanut. For example, NDVI effectively predicts biomass and yield, while CTD is associated with transpiration efficiency and carbon dioxide assimilation. These markers explained between 6.6% and 20.8% phenotypic variation, with most markers identified on the A subgenomes and respective homeologous chromosomes on the B subgenomes (Shaibu et al., 2020). Such markers, upon validation, may be deployed in marker-assisted breeding for drought tolerance in peanuts. A number of NAM- and MAGIC-based populations are being developed in peanuts which may provide useful genetic and genomic resources to study and implement genomic-assisted breeding for enhanced resistance to drought and heat stress in peanuts (Holbrook et al., 2013; Varshney, 2016; Gangurde et al., 2019). Efforts are underway to apply genomic selection (GS) for predicting the phenotypes by studying their genotypic architecture in multi-environment breeding trials (Pandey et al., 2020), but study of GS related to abiotic stress tolerance is not yet defined in peanut. Marker-assisted breeding has successfully introduced resistance to nematodes, rust

and leaf spot and improved oil quality in peanuts (Varshney et al., 2014; Bera et al., 2018; Ballén-Taborda et al., 2022).

Ascorbate peroxidase (APX), an antioxidant enzyme, contributes to ROS scavenging by decreasing hydrogen peroxide (H_2O_2) under environmental stresses. A comprehensive GWAS unfolded 166 *AhAPX* genes in the peanut genome, grouped into 11 main clusters, and have roles in oxidoreductase activity, catalytic activity, cell junction, cellular response to stimulus and detoxification, biosynthesis of metabolites, and phenylpropanoid metabolism. *AhAPX4/7/17/77/82/86/130/133* and *AhAPX160* showed significantly higher expression in diverse tissues/organs, while *AhAPX4/17/19/55/59/82/101/102/137* and *AhAPX140* were significantly upregulated under drought and cold stress, and phytohormones treatments. Functional characterization and validation of the *AhAPX* and SNPs may accelerate breeding programs to develop stress-tolerant peanut cultivars (Raza et al., 2022).

US peanut researchers at Clemson University reported lipid metabolism traits associated with heat tolerance which could be useful in determining lipid biomarkers to develop climate-resilient varieties. A targeted effort is underway in the US to develop new heat-resilient peanut cultivars using a range of heat-sensitive to heat-tolerant varieties such as "Bailey," "Georgia 12Y," "Phillips," "Sugg," "Tifguard," and "Wayne" and a breeding line SPT06-07 (Zoong Lwe et al., 2020). The development of molecular marker linkage maps and identification of markers and QTLs for target traits paved the way to develop efficient breeding methods to generate new, improved heat-tolerant peanut cultivars. The availability of peanut genome sequences and advanced genomics tools will aid in efficiently utilizing genetic resources toward a generation of sustainable crop yield.

The classical breeding methods employed to enhance drought and heat stress tolerance have had limited success. Advances in next-generation sequencing and phenomics, availability of genome sequences, and advances in bioinformatics and biotechnological tools may open new windows of opportunities to improve abiotic stress adaptation in food crops, including peanuts.

Transgenes and CRISPR/Cas9 Genome editing

Transgene and genome editing methods make up the core of the genetic engineering tool kit. These technologies alter a genome to create modified cell lines, new cultivars possessing valuable traits, or learn novel information about cellular processes or development. A transgene is a unit of genetic material inserted permanently or transiently into a cell where it is expressed to confer a phenotype. Efforts have been made to transfer genes of interest into the peanut genome *via Agrobacterium*-mediated transformation or particle bombardment. However, the successful production of transgenic plants has been limited since only a few genotypes were found to be transformable with relatively high efficiency. In peanuts, many factors, including a lack of efficient protocols to regenerate whole plants (Sharma and Anjaiah, 2000; Sharma and Ortiz, 2000; Geng et al., 2012; Chu et al., 2013) and bacterial interactions with peanut cells (Gelvin, 2003) may restrict transformation success and regeneration *via Agrobacterium*. The recalcitrance nature of many peanut cultivars to *Agrobacterium*-mediated transformation and regeneration is a challenging bottleneck for future peanut-based technology development. Therefore, there is a pressing need to explore suitable genotype-independent transformation methods, such as *in planta* transformation, which may avoid time-consuming tissue culture and regeneration processes.

Nevertheless, the technology has successfully deployed to create transgenic events in peanuts with enhanced drought, salt, and aflatoxin tolerance. Transgenic peanuts containing *AtDREB1A* confer tolerance to drought. Assessment of these events under varying moisture stress regimes and vapour pressure deficits (VPDs) yielded up to 24% improvement in seed yield largely due to increased harvest index and higher seed filling, and displayed 20%–30% lower pod yield reduction than WT under drought stress (Bhatnagar-Mathur et al., 2014). Another study led by Qin et al. (2011) reported that regulating the expression of the *IPT* gene by a water-deficit inducible promoter (P_{SARK}) performed much better, maintained higher photosynthetic rates and stomatal conductance (g_s), produced significantly more biomass, and yield under reduced irrigation conditions in greenhouse and field environments. Transgenic peanut plants overexpressing the *AtAVP1* were tolerant to both drought and salt stress, produced high biomass, and maintained higher photosynthetic and transpiration rates under reduced irrigation and saline conditions in the greenhouse. Additionally, transgenic peanuts expressing the transcription factors *AtNAC2* and *MuNAC4* from *Arabidopsis* and *Macrotyloma* showed high tolerance to drought, salt, and moisture stress and high yield in stressed environments (Pandurangaiah et al., 2014; Patil et al., 2014). Similarly, the expression of the *mtlD* (mannitol-1- phosphate dehydrogenase) in transgenic peanut plants conferred drought tolerance (Bhauso et al., 2014a; Bhauso et al., 2014b; Patel et al., 2017), and the overexpression of *GmMYB3a* into transgenic peanut plants displayed better physiological parameters with improved drought tolerance (He et al., 2020).

Since stress tolerance is a multigenic trait involving different signaling cascades, developing transgenic lines with more tolerance traits by transferring more than one gene is needed (Venkatesh et al., 2018). Co-expression of multiple genes in transgenic plants has shown improved stress tolerance compared to transgenic plants with single-gene. Using modified MultiSite Gateway approach (Vemanna et al., 2013) to simultaneously stack *Alfin1*, *PgHSF4,* and *PDH45* genes driven by individual promoters and terminators into a single vector resulted in transgenic peanut plants with improved stress tolerance, higher growth, and productivity than WT plants under drought-stress conditions (Table 6). Another successful example was that two-antifungal plant defensins *MsDef1* and *MtDef4.2* and two host-induced gene silencing of *aflM* and *aflP* genes were cloned into four binary vectors. These vectors were mobilized into *Agrobacterium*, resulting in transgenic peanuts with a near-immunity of aflatoxin contamination (Sharma et al., 2018). This gives hope that day is not far off to stacking genes cascade with suitable promoters for developing peanuts that combine aflatoxin resistance, tolerate drought as well productive. Conventional breeding has had limited success to achieving resistance to aflatoxin because of multiple mechanisms (*in-vitro* seed colonization, pre-harvest aflatoxin contamination and aflatoxin production) controlling aflatoxin contamination, phenotyping

TABLE 6 Genes used in transgenic peanut for tolerance of heat and drought stresses.

Gene	Function	Trait	References
IPT	Cytokinin biosynthesis	Drought	Li et al. (2013)
AtDREB1A	Transcription factor	Drought	Bhatnagar-Mathur et al. (2014); Sarkar et al. (2014)
AtDREB2A, AtHB7, AtABF3	Improve cellular tolerance	Drought and salt	Pruthvi et al. (2014)
Alfin1, PgHSF, PDH45	Stress-responsive transcription factor	Drought and oxidative stress	Ramu et al. (2016)
AtHDG11	Developmental regulator	Drought and salt	Banavath et al. (2018)
MuNAC4	Induce lateral root growth	Drought	Pandurangaiah et al. (2014), Venkatesh et al. (2022)

bottlenecks to measure different components of resistance, large genotype × environment interactions, and issues associated with pre- and post-harvest management of peanuts (Pandey et al., 2019).

Overcoming tradeoffs is a significant breeding challenge when combining stress tolerance and crops productivity as many genes of minor effects are involved. Identifying gene variants with diverse functions to overcome tradeoffs should receive a greater investment of time and resources to balance crop growth, stress tolerance and productivity (Dwivedi et al., 2021).

Genome editing involves transgenes or occasionally only proteins with or without an RNA, which can modify existing genetic material in a targeted manner, creating insertions, deletions, or base modifications. This technology has provided an alternative approach to plant breeding and has been efficient in producing new cultivars and genetic resources within a relatively short period. The clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated protein (Cas) type II systems provide methods for rapidly and efficiently editing plant genomes. This is accomplished through an RNA (CRISPR) guided nuclease (Cas) induced targeted double-strand DNA break, which can be repaired through several pathways that may lead to mutation. Furthermore, genomic DNA can be modified by tethering various enzymes to a nuclease-deficient Cas protein, which may also introduce targeted mutations (Xie et al., 2015; Samanta et al., 2016; Li et al., 2019). Specifically, CRISPR/Cas9-related technologies have tremendously impacted functional genomics by enabling selective and specific alteration of genomic DNA sequences *in vivo* (Li et al., 2019; Scheben et al., 2017; Zhang et al., 2019). Genome editing has various applications in plants, including basic and applied biological research for developing advanced biotechnology products (Zhang et al., 2019) through forward and reverse genetics, targeted gene insertion, promoter modification, and splice variant generation. This is a valuable technique for functional gene analysis or trait alteration, and its effectiveness has been demonstrated in many plant species.

Moreover, gain-of-function mutations through this same approach have been reported in plants (Wang et al., 2022). Additionally, identifying other type II Cas proteins, such as Cas12a (Cpf1), is another important avenue in genome editing research as it opens additional genomic regions to modification due to alternative PAM: TTTV (V=C, G, A) utilization. The recent development of Cas protein variants fused to a variety of enzymes has also widened the application of CRISPR/Cas technology from inducing indels, gene insertion, and gene modification through targeted double-strand DNA breaks to produce targeted single base changes through a system known as base editing (BE) or manipulating the gene expression through the promoter region.

Although CRISPR/Cas tools have been successful in a wide variety of plant species, their application in peanuts is currently limited. Realizing the full potential of CRISPR/Cas-based genome editing in peanuts will require the development of a toolbox of validated CRISPR/Cas constructs and protocols for their utilization. Several research programs have been focused on establishing these systems in peanuts by developing gene editing and base editing technology to include vectors using CRISPR/Cas9 and CRISPR/Cas12a variants to induce indels, make DNA alterations through base editing, or regulate gene expression. Work has progressed to develop delivery methods and validate construct functionality. For example, we have developed several constructs for genome editing using the peanut *FAD2* genes as proof-of-concept experiments. Two CRISPR/Cas9 constructs, pDW3872, and pDW3877, have induced indels in *FAD2* with an efficiency of up to 32% and 24%, respectively (unpublished). Two base editing constructs, pDW3873, and pDW3876, were developed using nCas9 fused to the cytosine deaminases APOEBEC1 or PmCDA1, respectively. These constructs successfully induced C to T changes with an overall efficiency of up to 21% and 42%, respectively (unpublished). Additional constructs (pDW3882 and pDW3886) expressing the enzymes AsCpf1 or LbCpf1 were investigated for their editing efficiency. The latter was more effective in peanuts than the former (unpublished). These preliminary results demonstrate that genome editing using CRISPR/Cas systems is feasible in peanuts.

Overall, using transgenes and genome editing technology in peanuts comes down to developing genotype-independent transformation protocols, identifying genes of interest and proof of concept using suitable genome editing constructs. Application of these advances will greatly accelerate genetic improvement in peanuts leading to the efficient generation of new lines with desirable traits, which will benefit peanut producers, industry, and consumers. On the other hand, the presence of foreign genes in transgenic plants triggers biosafety regulations. However, a comparison of transgenic Golden rice GR2E and conventional rice showed no statistically significant differences in the concentrations of phytic acid or the levels of trypsin inhibitor and no differences in pest and disease reactions between them (Mallikarjuna Swamy et al., 2019; Mallikarjuna Swamy et al., 2021). Comparative assessment of transgenic wheat containing the sunflower gene, *HaHB4* conferring drought tolerance and improved yield in driest environments, is nutritionally equivalent to non-transgenic wheat lines (González et al., 2019; Miranda et al., 2022). Because CRISPR/Cas9-based genome editing plants can be transgene-free by crossing

edited plant offspring, they are assigned a non-regulated status (Ahmad et al., 2021), such as the common button mushroom modified by the CRISPR system obtained a non-regulated status in 2016 (Waltz, 2016). 'Sanatech Seed' has launched the world's first genome-edited high-GABA tomato with enhanced nutritional benefits for consumption in Japan. This tomato contains high levels of gamma-aminobutyric acid (GABA), an amino acid believed to aid relaxation and help lower blood pressure (https://www.isaaa.org/kc/cropbiotechupdate/article/default.asp?ID=18668).

Concluding remarks

The evidence to date and predictions suggest the overall negative effect of climate change on agricultural production especially when more food, nutritious and safe, is required to feed the growing word population. In addition, the food and feed produced today is less nutritious, and faces an increased risk to contamination by mycotoxin producing fungi due to climate change.

Peanut at reproductive stage is ultra-susceptible to drought and heat stress, causing substantial loss to production and nutritional quality. The drought-stressed peanut is prone to aflatoxin contamination. Thus, aflatoxin contaminated peanut is hazardous to human and animal health, in addition to adversely impacting peanut trade.

Novel sources of resistance to drought in peanut gene pool has led researchers thoroughly investigate the physiological and molecular basis of stress tolerance, while the genes and markers associated with stress tolerance detected significant marker trait associations, which after validation may be deployed in genomics-assisted breeding in peanut. Greater resources are needed to unfold the genetic and molecular basis of heat stress tolerance as this trait in the past received less attention compared to drought research in peanut. A few reports indicate the feasibility of gene transfer by transgenic technology, with some events showing no growth-defense tradeoff, suggesting the transgene(s), a valid technology, to rapidly integrate stress tolerance gene(s) without yield penalty. Advances in developing peanut-based construct and their editing efficiency demonstrate that genome editing using CRISPR/Cas system is feasible in peanut. Public perception about the use of genetically modified and/or gene edited crops is gradually changing in favor of for food and feed uses and also for commercial cultivation, as the evidence to date suggests no significant changes in proximate composition between genetically engineered and conventionally-bred produce, except for the trait introduced.

Author contributions

NP and CC conceptualization of the article. SN, MP, and VS wrote the molecular basis of stress tolerance, AS-S, MD, and NN wrote the source of variation for drought and heat stress tolerance, NP and CC wrote the genomics-assisted breeding and introgression breeding, YB wrote the high-throughput phenomics, GH, ST, and DW wrote the Transgenes and CRISPR/Cs9 genome editiing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abadya, S., Shimelis, H., Pasupuleti, J., Mashilo, J., Chaudhari, S., and Manohar, S. S. (2021). Assessment of genetic diversity of groundnut (*Arachis hypogaea* L.) for kernel yield, oil and fodder quantity and quality under drought conditions. *Crop Sci.* 61, 1926–1943. doi:10.1002/csc2.20483

Abdelhakim, L. O. A., Rosenqvist, E., Wollenweber, B., Spyroglou, I., Ottosen, C. O., and Panzarová, K. (2021). Investigating combined drought-and heat stress effects in wheat under controlled conditions by dynamic image-based phenotyping. *Agronomy* 11, 364. doi:10.3390/agronomy11020364

Ahmad, A., Munawar, N., Khan, Z., Qusmani, A. T., Khan, S. H., Jamil, A., et al. (2021). An outlook on global regulatory landscape for genome-edited crops. *Int. J. Mol. Sci.* 22, 11753. doi:10.3390/ijms222111753

Ahmad, M., Waraich, E. A., Zulfiqar, U., Ullah, A., and Farooq, M. (2021a). Thiourea application improves heat tolerance in camelina (Camelina sativa L. Crantz) by modulating gas exchange, antioxidant defense and osmoprotection. *Ind. Crops Prod.* 170, 113826. doi:10.1016/j.indcrop.2021.113826

Ahmad, M., Waraich, E. A., Tanveer, A., and Anwar-ul-Haq, M. (2021b). Foliar applied thiourea improved physiological traits and yield of camelina and canola under normal and heat stress conditions. *J. Soil Sci. Plant Nutr.* 21, 1666–1678. doi:10.1007/s42729-021-00470-8

Akbar, A., Manohar, S. S., Variat, M. T., Kurapati, S., and Pasupuleti, J. (2017). Efficient partitioning of assimilates in stress-tolerant groundnut genotypes under high temperature stress. *Agronomy* 7, 30. doi:10.3390/agronomy7020030

Akcay, U. C., Ercan, O., Kavas, M., Yildiz, L., Yilmaz, C., Oktem, H. A., et al. (2010). Drought-induced oxidativ damage and antioxidant responses in peanut (*Arachis hypogaea* L.) seedlings. *Plant Growth Regul.* 1, 21–28. doi:10.1007/s10725-010-9445-1

Araus, J. L., Slafer, G. A., Reynolds, M. P., and Royo, C. (2002). Plant breeding and drought in C3 cereals: What should we breed for? *Ann. Bot.* 89 Spec No (7), 925–940. doi:10.1093/aob/mcf049(

Aravind, B., Nayak, S. N., Choudhary, R. S., Gandhadmath, S. S., Prasad, P. V. V., Pandey, M. K., et al. (2022). "Integration of genomics approaches in abiotic stress tolerance in groundnut (*Arachis hypogaea* L.): An overview," in *Genomic designing of abiotic stress resistant oilseed crops*. Editor C. Kole (Cham, Switzerland: Springer international), 149–197. ISBN: 9783030900434.

Aravind, B. (2021). *Molec.ular elucidation of heat stress in terms of physiological and biochemical parameters in groundnut (*Arachis hypogaea *L.)*. M.Sc Thesis. Dharwad: University of Agricultural Sciences.

Ashraf, M., and Hafeez, M. (2004). Thermo-tolerance of pearl millet and maize at early growth stages: Growth and nutrient relations. *Biol. Plant.* 48, 81–86. doi:10.1023/b:biop.0000024279.44013.61

Avramova, V., Abdelgawad, H., Zhang, Z., Fotschki, B., Casadevall, R., Vergauwen, L., et al. (2015). Drought induces distinct growth response, protection, and recovery mechanisms in the maize leaf growth zone. *Plant Physiol.* 169, 1382–1396. doi:10.1104/pp.15.00276

Awasthi, R., Kaushal, N., Vadez, V., Turner, N. C., Berger, J., Siddique, K. H. M., et al. (2014). Individual and combined effects of transient drought and heat stress on carbon

assimilation and seed filling in chickpea. *Funct. Plant Biol.* 41, 1148–1167. doi:10.1071/FP13340

Bagherian, K., Bidese, R., Bao, Y., Zhang, Q., Sanz-Saez, A., Chen, C., et al. (2022). "Phenotyping agronomic traits of peanuts using UAV-based hyperspectral imaging and deep learning," in 2022 ASABE Annual International Virtual Meeting (American Society of Agricultural and Biological Engineers), 1. doi:10.13031/aim.202200814

Bajji, M., Kinet, J. M., and Lutts, S. (2002). The use of the electrolyte leakage method for assessing cell membrane stability as a water stress tolerance test in durum wheat. *Plant Growth Regul.* 36, 61–70. doi:10.1023/A:1014732714549

Ballén-Taborda, C., Chu, Y., Ozias-Akins, P., Holbrook, C. C., Timper, P., Jackson, S. A., et al. (2022). Development and genetic characterization of peanut advanced backcross lines that incorporate root-knot nematode resistance from *Arachis stenosperma*. *Front. Plant Sci.* 12, 785358. doi:10.3389/fpls.2021.785358

Balota, M., and Oakes, J. (2017). "UAV remote sensing for phenotyping drought tolerance in peanuts," in *Autonomous air and ground sensing systems for agricultural optimization and phenotyping II*, 10218, 81–87. SPIE. doi:10.1117/12.2262496

Banavath, J. N., Chakradhar, T., Pandit, V., Konduru, S., Guduru, K. K., Akila, C. S., et al. (2018). Stress inducible overexpression of *AtHDG11* leads to improved drought and salt stress tolerance in peanut (*Arachis hypogaea* L.). *Front. Chem.* 6, 34. doi:10.3389/fchem.2018.00034

Belamkar, V. (2010). *A first insight into population structure, linkage disequilibrium, and association mapping of drought tolerance-related traits in the U.S. peanut mini core collection*. M.S. thesis. Lubbock, TX: Texas Tech University.

Bera, S. K., Kamdar, J. H., Kasundra, S. V., Dash, P., Maurya, A. K., Jasani, M. D., et al. (2018). Improving oil quality by altering levels of fatty acids through marker-assisted selection of *ahfad2* alleles in peanut (*Arachis hypogaea* L.). *Euphytica* 214, 162. doi:10.1007/s10681-018-2241-0

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K., et al. (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48 (4), 438–446. doi:10.1038/ng.3517

Bertioli, D. J., Clevenger, J., Godoy, I. J., Stalker, H. T., Wood, S., Santos, J. F., et al. (2021). Legacy genetics of Arachis cardenasii in the peanut crop shows the profound benefits of international seed exchange. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2104899118. doi:10.1073/pnas.2104899118

Bhatnagar-Mathur, P., Rao, J. S., Vadez, V., Dumbala, S. R., Rathore, A., Yamaguchi-Shinozaki, K., et al. (2014). Transgenic peanut overexpressing the DREB1A transcription factor has higher yields under drought stress. *Mol. Breed.* 33, 327–340. doi:10.1007/s11032-013-9952-7

Bhauso, T., Radhakrishnan, T., Kumar, A., Mishra, G., Dobaria, J., Patet, K., et al. (2014a). Over-expression of bacterial *mtlD* gene in peanut improves drought tolerance through accumulation of mannitol. *Sci. World J.* 125967, 125967. doi:10.1155/2014/125967

Bhauso, T. D., Thankappan, R., Kumar, A., Mishra, G. P., Dobaria, J. R., and Venkat Rajam, M. (2014b). Over-expression of bacterial *mtlD* gene confers enhanced tolerance to salt-stress and water-deficit stress in transgenic peanut (*Arachis hypogaea*) through accumulation of mannitol. *Aust. J. Crop Sci.* 8, 413. doi:10.3389/fpls.2016.00935

Bhogireddy, S., Xavier, A., Garg, V., Layland, N., Arias, R., Payton, P., et al. (2020). Genome-wide transcriptome and physiological analyses provide new insights into peanut drought response mechanisms. *Sci. Rep.* 10, 4071. doi:10.1038/s41598-020-60187-z

Bidese, R., Bao, Y., Sanz-Saez, A., and Chen, C. (2021). "Infield peanut pod counting using deep neural networks for yield estimation," in 2021 ASABE Annual International Virtual Meeting. (American Society of Agricultural and Biological Engineers). doi:10.13031/aim.202101080

Bita, C. E., and Gerats, T. (2013). Plant tolerance to high temperature in a changing environment: Scientific fundamentals and production of heat stress-tolerant crops. *Front. Plant Sci.* 4, 273. doi:10.3389/fpls.2013.00273

Blum, A., and Ebercon, A. (1981), Cell membrane stability as a measure of drought and heat tolerance in Wheat1. *Crop Sci.*, 21: 43–47. doi:10.2135/cropsci1981.0011183X002100010013x

Blum, A. (2009). Effective use of water (EUW) and not water-use efficiency (WUE) is the target of crop yield improvement under drought stress. *Field Crops Res.* 112, 119–123. doi:10.1016/j.fcr.2009.03.009

Bowler, C., Montagu, M. V., and Inze, D. (1992). Superoxide dismutase and stress tolerance. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 43, 83–116. doi:10.1146/annurev.pp.43.060192.000503

Boyer, J. S. (1982). Plant productivity and environment. *Science* 218 (4571), 443–448. doi:10.1126/science.218.4571.443

Carmo, L. S. T., Martins, A. C. Q., Martins, C. C. C., Passos, M. A. S., Silva, L. P., et al. (2018). Comparative proteomics and gene expression analysis in *Arachis duranensis* reveal stress response proteins associated to drought tolerance. *J. Proteomics* 192, 299–310. doi:10.1016/j.jprot.2018.09.011,

Chakraborty, K., Bishi, S. K., Singh, A. L., Zala, P. V., Mahatma, M. K., Kalariya, K. A., et al. (2018). Rapid induction of small heat shock proteins improves physiological adaptation to high temperature stress in peanut. *J. Agron. Crop Sci.* 204, 285–297. doi:10.1111/jac.12260

Chaudhary, S., Devi, P., HanumanthaRao, B., Jha, U. C., Sharma, K. D., Prasad, P. V. V., et al. (2022) Physiological and molecular approaches for developing thermotolerance in vegetable crops: A growth, yield and sustenance perspective. *Front. Plant Sci.* 13: 878498. doi:10.3389/fpls.2022.878498

Chen, N., Yang, Q., Su, M., Pan, L., Chi, X., Chen, M., et al. (2012). Cloning of six ERF family transcription factor genes from peanut and analysis of their expression during abiotic stress. *Plant molecular biology reporter* 30, 1415–1425.

Chen, P., Sneller, C., Purcell, L., Sinclair, T., King, C. A., and Ishibashi, T. (2007). Registration of soybean germplasm lines R01-416F and R01-581F for improved yield and nitrogen fixation under drought stress. *J. Plant Reg.* 1, 166–167. doi:10.3198/JPR2007.01.0046CRG

Chen, T., Yang, W., Zhang, H., Zhu, B., Zeng, R., Wang, X., et al. (2020). Early detection of bacterial wilt in peanut plants through leaf-level hyperspectral and unmanned aerial vehicle data. *Comp. Electron. Agric.* 177, 105708. doi:10.1016/j.compag.2020.105708

Chu, Y., Bhattacharya, A., Wu, C., Knoll, J. E., and Ozias-Akins, P. (2013). Improvement of peanut (*Arachis hypogaea* L.) transformation efficiency and determination of transgene copy number by relative quantitative real-time PCR. *Vitro Cell. Dev. Biol. Plant* 49, 266–275. doi:10.1007/s11627-013-9518-8

Clevenger, J. P., Korani, W., Ozias-Akins, P., and Jackson, S. (2018). Haplotype-based genotyping in polyploids. *Front. Plant Sci.* 9, 564. doi:10.3389/fpls.2018.00564

Cohen, I., Zandalinas, S. I., Huck, C., Fritschi, F. B., and Mittler, R. (2021). Meta-analysis of drought and heat stress combination impact on crop yield and yield components. *Physiol. Plant.* 171, 66–76. doi:10.1111/ppl.13203

Condon, A. G., Richards, R. A., Rebetzke, G. J., and Farquhar, G. D. (2004). Breeding for high water use efficiency. *J. Exp. Bot.* 55, 2447–2460. doi:10.1093/jxb/erh277

Cox, F. R. (1979). Effect of temperature treatment on peanut vegetative and fruit growth 1. *Peanut Sci.* 6, 14–17. doi:10.3146/i0095-3679-6-1-4

Crafts-Brandner, S. J., and Salvucci, M. E. (2002). Sensitivity of photosynthesis in a C4 plant, maize, to heat stress. *Plant Physiol.* 129, 1773–1780. doi:10.1104/pp.002170

Craufurd, P. Q., Wheeler, T. R., Ellis, R. H., Summerfield, R. J., and Prasad, V. P. V. (2000). Escape and tolerance to high temperature at flowering in groundnut (*Arachis hypogaea* L.). *J. Agric. Sci.* 135, 371–378. doi:10.1017/S0021859699008394

Craufurd, P. Q., Prasad, P. V. V., and Summerfield, R. J. (2002). Dry matter production and rate of change of harvest index at high temperature in peanut. *Crop Sci.* 42, 146–151. doi:10.2135/cropsci2002.1460

Craufurd, P. Q., Prasad, P. V., Kakani, G. V., Wheeler, T. R., and Nigam, S. N. (2003). Heat tolerance in groundnut. *Field crop. Res.* 80, 63–77. doi:10.1016/s0378-4290(02)00155-7

Dash, D., Vp, C., and Bo, K. (2020). Impact of heat stress on physiological and yield components under varied temperature regimes in groundnut cultivars. *J. Pharmacogn. Phytochemistry* 9, 1060–1066.

Dat, J., Vandenabeele, S., Vranová, E., Van Montagu, M., Inze, D., and Van Breusegem, F. (2000). Dual action of the active oxygen species during plant stress responses. *CMLS, Cell. Mol. Life Sci.* 57, 779–795. doi:10.1007/s000180050041

Devi, M. J., and Sinclair, T. R. (2011). Diversity in drought traits among commercial southeastern US peanut cultivars. *Int. J. Agron.* 2011, 1–7. doi:10.1155/2011/754658

Devi, J. M., Sinclair, T. R., and Vadez, V. (2010). Genotypic variation in peanut for transpiration response to vapor pressure deficit. *Crop Sci.* 50, 191–196. doi:10.2135/cropsci2009.04.0220

Devi, J. M., Rowland, D. L., Payton, P., Faircloth, W., and Sinclair, T. R. (2013). Nitrogen fixation tolerance to soil water deficit among commercial cultivars and breeding lines of peanut. *Field Crops Res.* 149, 127–132. doi:10.1016/j.fcr.2013.04.026

Ding, Y., Qiu, X., Luo, H., Huang, L., Guo, J., Yu, B., et al. (2022). Comprehensive evaluation of Chinese peanut mini-mini core collection and QTL mapping for aflatoxin resistance. *BMC Plant Biol.* 22, 207. doi:10.1186/s12870-022-03582-0

Dobreva, I. D., Ruiz-Guzman, H. A., Barrios-Perez, I., Adams, T., Teare, B. L., Payton, P., et al. (2021). Thresholding analysis and feature extraction from 3D ground penetrating radar data for noninvasive assessment of peanut yield. *Remote Sens.* 13, 1896. doi:10.3390/rs13101896

Driedonks, N., Xu, J., Peters, J. L., Park, S., and Rieu, I. (2015). Multi-level interactions between heat shock factors, heat shock proteins, and the redox system regulate acclimation to heat. *Front. Plant Sci.* 6, 999. doi:10.3389/fpls.2015.00999

Dutra, W. F., Guerra, Y. L., Ramos, J. P. C., Fernandes, P. D., Silva, C. R. C., Bertioli, D. J., et al. (2018). Introgression of wild alleles into the tetraploid peanut crop to improve water use efficiency, earliness and yield. *PLOS ONE* 13, e0198776. doi:10.1371/journal.pone.0198776

Dutta, S., Mohanty, S., and Tripathy, B. C. (2009). Role of temperature stress on chloroplast biogenesis and protein import in pea. *Plant Physiol.* 150, 1050–1061. doi:10.1104/pp.109.137265

Dwivedi, L., Puppala, N., Upadhyaya, H. D., Manivannan, N., and Singh, S. (2008). Developing a core collection of peanut pecific to Valencia market type. *Crop Sci.* 48, 625–632.

Dwivedi, S., Sahrawat, K., Upadhyaya, H., and Ortiz, R. (2013). Food, nutrition and agrobiodiversity under global climate change. *Adv. Agron.* 120, 1–128. doi:10.1016/B978-0-12-407686-0.00001

Dwivedi, S. L., Reynolds, M. P., and Ortiz, R. (2021). Mitigating tradeoffs in plant breeding. *iScience* 24, 102965. doi:10.1016/j.isci.2021.102965

Essandoh, D. A., Odong, T., Okello, D. K., Fonceka, D., Nguepjop, J., Sambou, A., et al. (2022). Quantitative trait analysis shows the potential for alleles from the wild species*Arachis batizocoi* and *A. duranensis* to improve groundnut disease resistance and yield in east Africa. *Agronomy* 12, 2202. doi:10.3390/agronomy12092202

Farquhar, G. D., Ehleringer, J. R., and Hubick, K. T. (1989). Carbon isotope discrimination and photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 40, 40503–40537. doi:10.1146/annurev.pp.40.060189.002443

Faye, I., Pandey, M. K., Hamidou, F., Rathore, A., Ndoye, O., Vadez, V., et al. (2015). Identification of quantitative trait loci for yield and yield related traits in groundnut (*Arachis hypogaea* L.) under different water regimes in Niger and Senegal. *Euphytica* 206, 631–647. doi:10.1007/s10681-015-1472-6

Feng, X. H., Zhang, H. X., Ali, M., Gai, W. X., Cheng, G. X., Yu, Q., et al. (2019). A small heat shock protein CaHsp25. 9 positively regulates heat, salt, and drought stress tolerance in pepper (Capsicum annuum L.). *Plant Physiol. biochem.* 142, 151–162. doi:10.1016/j.plaphy.2019.07.001

Fischer, R. A. (2011). Wheat physiology: A review of recent developments. *Crop Pasture Sci.* 62, 95–114. doi:10.1071/cp10344

Frankel, O. H. (1984). "Genetic perspectives of germplasm conservation," in *Genetic manipulation: Impact on man and society*. Editors W. K. Arber, K. Llimensee, W. J. Peacock, and P. Starlinger (Cambridge: Cambridge Univ. Press), 161–170.

Fu, P., Meacham-Hensold, K., Guan, K., and Bernacchi, C. J. (2019). Hyperspectral leaf reflectance as proxy for photosynthetic capacities: An ensemble approach based on multiple machine learning algorithms. *Front. Plant Sci.* 10, 730. doi:10.3389/fpls.2019.00730

Gangappa, E., Ravi, N., and Veerakumar, G. N. (2006). Evaluation of groundnut (*Arachis hypogaea* L.) genotypes for temperature tolerance based on temperature induction response (TIR) technique. *Indian J. Genet. Plant Breed.* 66, 127–130.

Gangurde, S. S., Wang, H., Yaduru, S., Pandey, M. K., Fountain, J. C., Chu, Y., et al. (2019). Nested association mapping (NAM)-based genetic dissection uncovers candidate genes for seed and pod weights in peanut (*Arachis hypogaea*). *Plant Biotechnol. J.* 18, 1457–1471. doi:10.1111/pbi.13311

Gautami, B., Pandey, M. K., Vadez, V., Nigam, S. N., Ratnakar, P., Krishnamurthy, L., et al. (2012). Quantitative trait locus analysis and construction of consensus genetic map for drought folerance traits based on three recombinant inbred line populations in cultivated groundnut (*Arachis hypogaea* L.). *Mol. Breed.* 30, 757–772. doi:10.1007/s11032-011-9660-0

Ge, Y., Bai, G., Stoerger, V., and Schnable, J. C. (2016). Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging. *Comp. Electron. Agric.* 127, 625–632. doi:10.1016/j.compag.2016.07.028

Gelvin, S. B. (2003). Agrobacterium-mediated plant transformation: The biology behind the "gene-jockeying" tool. *Microbiol. Mol. Biol. Rev.* 67, 16–37. doi:10.1128/mmbr.67.1.16-37.2003

Geng, L., Niu, L., Gresshoff, P. M., Shu, C., Song, F., Huang, D., et al. (2012). Efficient production of Agrobacterium rhizogenes-transformed roots and composite plants in peanut (*Arachis hypogaea* L.). *Plant Cell Tiss. Organ Cult.* 109, 491–500. doi:10.1007/s11240-012-0113-1

Golombek, S. D., and Johansen, C. (1997). Effect of soil temperature on vegetative and reproductive growth and development in three Spanish genotypes of peanut (*Arachis hypogaea l.*). *Peanut Sci.* 24, 67–72. doi:10.3146/i0095-3679-24-2-1

González, F. G., Capella, M., Ribichich, K. F., Curín, F., Giacomelli, J. I., Ayala, F., et al. (2019). Field-grown transgenic wheat expressing the sunflower gene *HaHB4* significantly outyields the wild type. *J. Exp. Bot.* 70, 1669–1681. doi:10.1093/jxb/erz037

Gregory, W. C., Krapovickas, A., and Gregory, M. P. (1980). "Structure, variation, evolution, and classification in Arachis," in *Structure, variation, evolution, and classification in Arachis*, 469–481.

Hamidou, F., Ratnakumar, P., Halilou, O., Mponda, O., Kapewa, T., Monyo, E., et al. (2012). Selection of intermittent drought stress tolerant lines across years and locations in the reference collection of groundnut (*Arachis hypogaea* L.). *Field Crops Res.* 126, 189–199. doi:10.1016/j.fcr.2011.10.009

Hamidou, F., Halilou, O., and Vadez, V. (2013). Assessment of groundnut under combined heat and drought stress. *J. Agron. Crop Sci.* 199, 1–11. doi:10.1111/j.1439-037X.2012.00518.x

Hamidou, F., Rathore, A., Waliyar, F., and Vadez, V. (2014). Although drought intensity increases aflatoxin contamination, drought tolerance does not lead to less aflatoxin contamination. *Field Crops Res.* 156, 103–110. doi:10.1016/j.fcr.2013.10.019

Hasanuzzaman, M., Bhuyan, M. H. M., Zulfiqar, F., Raza, A., Mohsin, S. M., Mahmud, J. A., et al. (2020a). Reactive oxygen species and antioxidant defense in plants under abiotic stress: Revisiting the crucial role of a universal defense regulator. *Antioxidants* 9, 681. doi:10.3390/antiox9080681

He, Y., Mu, S., He, Z., Wang, B., and Li, Y. (2020). Ectopic expression of MYB repressor *GmMYB3a* improves drought tolerance and productivity of transgenic peanuts (*Arachis hypogaea* L.) under conditions of water deficit. *Trans. Res.* 29, 563–574. doi:10.1007/s11248-020-00220-z

Hemantaranjan, A., Malik, C. P., and Bhanu, A. N. (2018). Physiology of heat stress and tolerance mechanisms-an overview. *J. Plant Sci. Res.* 34, 51–64. doi:10.32381/jpsr.2018.34.01.7

Herritt, M. T., Pauli, D., Mockler, T. C., and Thompson, A. L. (2020). Chlorophyll fluorescence imaging captures photochemical efficiency of grain sorghum (*Sorghum bicolor*) in a field setting. *Plant Methods* 16, 109. doi:10.1186/s13007-020-00650-0

Holbrook, C. C., and Dong, W. (2005). Development and evaluation of a mini core collection for the U. S. peanut germplasm collection. *Crop Sci.* 45, 1540–1544. doi:10.2135/cropsci2004.0368

Holbrook, C. C., and Stalker, H. T. (2003). Peanut breeding and genetic resources. *Plant Breed. Rev.* 22, 297–356. doi:10.1002/9780470650202.ch6

Holbrook, C. C., Isleib, T. G., Ozias-Akins, P., Chu, Y., Knapp, S. J., Tillman, B., et al. (2013). Development and phenotyping of recombinant inbred line (RIL) populations for peanut (*Arachis hypogaea* L.). *Peanut Sci.* 40, 89–94. doi:10.3146/ps13-5.1

Hungria, M., and Kaschuk, G. (2014). Regulation of N_2 fixation and NO3−/NH4+ assimilation in nodulated and N-fertilized *Phaseolus vulgaris* L. exposed to high temperature stress. *Environ. Exp. Bot.* 98, 32–39. doi:10.1016/j.envexpbot.2013.10.010

IPCC (2022). "Climate change (2022): Impacts, adaptation, and vulnerability," in *Contribution of working group II to the sixth assessment report of the intergovernmental panel on climate change*. Editors H.-O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría, et al. (Cambridge, UK and New York, NY, USA: Cambridge University Press), 3056. doi:10.1017/9781009325844

Jain, A. K., Basha, S. M., and andHolbrook, C. C. (2001). Identification of drought responsive transcripts in peanut(*Arachishypogaea* L.). *Mol. Biotechnol. Genet.* 4, 2. doi:10.2225/vol4-issue2-fulltext-2

Jain, V. K. (2000). *Fundamental of plant physiology*. New Delhi xS. Chand Publishing. ISBN: 8121904625.

Janská, A., Marsík, P., Zelenková, S., Ovesná, J., and Janska, A. (2010). Cold stress and acclimation - what is important for metabolic adjustment? *Plant Biol. (Stuttg)* 12 (3), 395–405. doi:10.1111/j.1438-8677.2009.00299.x

Jewan, S. Y. Y., Pagay, V., Billa, L., Tyerman, S. D., Gautam, D., Sparkes, D., et al. (2022). The feasibility of using a low-cost near-infrared, sensitive, consumer-grade digital camera mounted on a commercial UAV to assess Bambara groundnut yield. *Int. J. Remote Sens.* 43, 393–423. doi:10.1080/01431161.2021.1974116

Jiang, C., Li, X., Zou, J., Ren, J., Jin, C., Zhang, H., et al. (2021). Comparative transcriptome analysis of genes involved in the drought stress response of two peanut (*Arachis hypogaea* L.) varieties. *BMC Plant Biol.* 21, 64. doi:10.1186/s12870-020-02761-1

Kakani, V. G., Prasad, P. V. V., Craufurd, P. Q., and Wheeler, T. R. (2002). Response of*in vitro* pollen germination and pollen tube growth of groundnut (*Arachis hypogaea* L.) genotypes to temperature. *Plant, Cell Environ.* 25, 1651–1661. doi:10.1046/j.1365-3040.2002.00943.x

Kakani, V. G., Timothy, R., Wheeler, P., Craufurd, Q,, and Rao, C. N. R. (2015). "Effect of high temperature and water stress on groundnuts under field conditions. combined stresses in forests," in *Combined stresses in plants. Physiological, molecular and biochemical aspects*. Editor R. Mahalingham (Springer), 159–180. doi:10.1007/978-3-319-07899-1_8.2015

King, Z., Serrano, J., Roger Boerma, H., and Li, H. (2014). Non-toxic and efficient DNA extractions for soybean leaf and seed chips for high-throughput and large-scale genotyping. *Biotechnol. Lett.* 36, 1875–1879. doi:10.1007/s10529-014-1548-8

Kokkanti, R. R., Hindu, V., Latha, P., Vasanthi, R. P., Sudhakar, P., and Usha, R. (2019). Assessment of genetic variability and molecular characterization of heat stress tolerant genes in *Arachis hypogaea* L. through qRT-PCR. *Biocatal. Agric. Biotechnol.* 20, 101242. doi:10.1016/j.bcab.2019.101242

Krapovickas, A., Gregory, W. C., Williams, D. E., and Simpson, C. E. (1994). Taxonomia del genero *Arachis* (*Leguminoae*). *Bonplandia* VIII, 1–187.

Krishnamurthy, L., Vadez, V., Jyotsna, D. M., Serraj, R., Nigam, S. N., Sheshshayee, M. S., et al. (2007). Variation in transpiration efficiency and its related traits in a groundnut (*Arachis hypogaea* L.) mapping population. *Field Crops Res.* 103, 189–197. doi:10.1016/j.fcr.2007.06.009

Kumar, S., Thakur, P., Kaushal, N., Malik, J. A., Gaur, P., and Nayyar, H. (2013). Effect of varying high temperatures during reproductive growth on reproductive function, oxidative stress and seed yield in chickpea genotypes differing in heat sensitivity. *Archives Agron. Soil Sci.* 59 (6), 823–843. doi:10.1080/03650340.2012.683424

Lauriano, J. A., Lidon, F. C., Carvalho, C. A., Campos, P. S., and do Céu Matos, M. (2000). Drought effects on membrane lipids and photosynthetic activity in different peanut cultivars. *Photosynthetica* 38, 7–12. doi:10.1023/A:1026775319916

Laxa, M., Liebthal, M., Telman, W., Chibani, K., and Dietz, K. J. (2019). The role of the plant antioxidant system in drought tolerance. *Antioxidants* 8, 94. doi:10.3390/antiox8040094

Leal-Bertioli, S. C. M., Bertioli, D. J., Guimaraes, P. M., Pereira, T. D., Galhardo, I., Silva, J. P., et al. (2012). The effect of tetraploidization of wild*Arachis*on leaf morphology and other drought-related traits. *Environ. Exp. Bot.* 84, 17–24. doi:10.1016/j.envexpbot.2012.04.005

Li, S., Hu, R., Shen, G., and Zhang, H. (2013). Genetic engineering peanut for higher drought- and salt-tolerance. *Food Nutr. Sci.* 4, 1–7. doi:10.4236/fns.2013.46a001

Li, J., Li, Y., and Ma, L. (2019). CRISPR/Cas9-Based genome editing and its applications for functional genomic analyses in plants. *Small Methods* 3 (3), 1800473. doi:10.1002/smtd.201800473

Li, H., Yang, M., Zhao, C., Wang, Y., and Zhang, R. (2021). Physiological and proteomic analyses revealed the response mechanisms of two different drought-resistant maize varieties. *BMC Plant Biol.* 21, 513. doi:10.1186/s12870-021-03295-w

Li, L., Cui, S., Dang, P., Yang, X., Wei, X., Chen, K., et al. (2022). GWAS and bulked segregant analysis reveal the Loci controlling growth habit-related traits in cultivated peanut (*Arachis hypogaea* L.). *BMC Genom* 23, 403. doi:10.1186/s12864-022-08640-3

Liu, X. Z., and Hang, B. R. (2000). Heat stress injury in relation to membrane lipid peroxidation in creeping bentgrass. *Crop Sci.* 40, 503–510. doi:10.2135/cropsci2000.402503x

Liu, X., Kasuga, M., Sakuma, Y., AbeMiura, S., Yamaguchi-Shinozaki, K., et al. (1998). Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in Arabidopsis**.** *Plant Cell* 10, 1391–1406. doi:10.2307/3870648

Luis, J. M., Ozias-Akins, P., Holbrook, C. C., Kemerait, R. C., Jr, Snider, J. L., and Liakos, V. (2016). Phenotyping peanut genotypes for drought tolerance. *Peanut Sci.* 43, 36–48. doi:10.3146/0095-3679-43.1.36

Lyu, J. I., Park, J. H., Kim, J. K., Bae, C. H., Jeong, W. J., Min, S. R., et al. (2018). Enhanced tolerance to heat stress in transgenic tomato seeds and seedlings overexpressing a trehalose-6-phosphate synthase/phosphatase fusion gene. *Plant Biotechnol. Rep.* 12, 399–408. doi:10.1007/s11816-018-0505-8

Ma, W., Zhao, T., Li, J., Liu, B., Fang, L., Hu, Y., et al. (2016). Identification and characterization of the GhHsp20 gene family in *Gossypium hirsutum*. *Sci. Rep.* 6, 32517. doi:10.1038/srep32517

Mallikarjuna, N., Senthilvel, S., and Hoisington, D. (2012). Development of new sources of tetraploid Arachis to broaden the genetic base of cultivated groundnut (*Arachis hypogaea* L.). *Genet. Resour. Crop Evol.* 58, 889–907. doi:10.1007/s10722-010-9627-8

Mallikarjuna Swamy, B. P., Samia, M., Boncodin, R., Marundan, S., Rebong, D. B., Ordonio, R. L., et al. (2019). Compositional analysis of genetically engineered GR2E "Golden Rice" in comparison to that of conventional rice. *J. Agri. Food Chem.* 67, 7986–7994. doi:10.1021/acs.jafc.9b01524

Mallikarjuna Swamy, B. P., Marundan, S., Jr., Samia, M., Ordonio, R. L., Rebong, D. B., Miranda, R., et al. (2021). Development and characterization of GR2E Golden rice introgression lines. *Sci. Rep.* 11, 2496. doi:10.1038/s41598-021-82001-0

Marcum, K. B. (1998), Cell membrane thermostability and whole-plant heat tolerance of Kentucky bluegrass. *Crop Sci.* 38, 1214–1218. doi:10.2135/cropsci1998.0011183X003800050017x

Meacham-Hensold, K., Fu, P., Wu, J., Serbin, S., Montes, C. M., Ainsworth, E., et al. (2020). Plot-level rapid screening for photosynthetic parameters using proximal hyperspectral imaging. *J. Exp. Bot.* 71, 2312–2328. doi:10.1093/jxb/eraa068

Medrano, H., Tomás, M., Martorell, S., Flexas, J., Hernández, E., Rosello, J., et al. (2015). From leaf to whole-plant water use efficiency (WUE) in complex canopies: Limitations of leaf WUE as a selection target. *Crop J.* 3 (3), 220–228. doi:10.1016/j.cj.2015.04.002

Miller, G., Suzuki, N., Ciftci-Yilmaz, S., and Mittler, R. (2010). Reactive oxygen species homeostasis and signalling during drought and salinity stresses. *Plant Cell Environ.* 33, 453–467. doi:10.1111/j.1365-3040.2009.02041.x

Miranda, P. V., Iglesias, B. F., Charriere, M. V., and Burachik, M. (2022). Drought tolerant wheat IND-ØØ412-7 is nutritionally equivalent to its non-transgenic comparator. *Gm. Crops Food* 13, 119–125. doi:10.1080/21645698.2022.2079179

Mittler, R. (2002). Oxidative stress, antioxidants and stress tolerance. *Trends Plant Sci.* 7, 405–410. doi:10.1016/s1360-1385(02)02312-9

Nigam, S. N., Dwivedi, S. L., and Gibbons, R. W. (1991). Groundnut breeding: Constraints, achievements and future possibilities. *Plant Breed. Abstr.* 61 (10), 1127–1136.

Nigam, S. N., Rao, R. C. N., Wynne, J. C., Williams, J. H., Fitzner, M., and Nagabhushanam, G. V. S. (1994). Effect and interaction of temperature and photoperiod on growth and partitioning in three groundnut (*Arachis hypogaea* L.) genotypes[1]. *Ann. Appl. Biol.* 125, 541–552. doi:10.1111/j.1744-7348.1994.tb04991.x

Nigam, S. N., Nageswara Rao, R. C., and Wynne, J. C. (1998). Effects of temperature and photoperiod on vegetative and reproductive growth of groundnut (*Arachis hypogaea* L.). *J. Agron. Crop Sci.* 181, 117–124. doi:10.1111/j.1439-037X.1998.tb00406.x

Nigam, S., Chandra, S., Sridevi, K. R., Bhukta, M., Reddy, A. G. S., Rachaputi, N. R., et al. (2005). Efficiency of physiological trait-based and empirical selection approaches for drought tolerance in groundnut. *Ann. Appl. Biol.* 146, 433–439. doi:10.1111/j.1744-7348.2005.040076.x

Oladzad, A., González, A., Macchiavelli, R., de Jensen, C. E., Beaver, J., Porch, T., et al. (2020). Genetic factors associated with nodulation and nitrogen derived from atmosphere in a middle American common bean panel. *Front. Plant Sci.* 11. doi:10.3389/fpls.2020.576078

Pandey, M. K., Monyo, E., Ozias-Akins, P., Liang, X., Guimarães, P., Nigam, S. N., et al. (2012). Advances in Arachis genomics for peanut improvement. *Biotechnol. Adv.* 30 (3), 639–651. doi:10.1016/j.biotechadv.2011.11.001

Pandey, M. K., Upadhyaya, H. D., Rathore, A., Vadez, V., Sheshshayee, M. S., Sriswathi, M., et al. (2014). Genome-wide association studies for 50 agronomic traits in peanut using the 'reference set'comprising 300 genotypes from 48 countries of the semi-arid tropics of the world. *PLoS one* 9, e105228. doi:10.1371/journal.pone.0105228

Pandey, M., Agarwal, G., Kale, S. M., Clevenger, J., Nayak, S. N., Sriswathi, M., et al. (2017). Development and evaluation of a high density genotyping 'Axiom_Arachis' array with 58 K SNPs for accelerating genetics and breeding in groundnut. *Sci. Rep.* 7, 40577 doi:10.1038/srep40577

Pandey, M. K., Kumar, R., Pandey, A. K., Soni, P., Gangurde, S. S., Sudini, H. K., et al. (2019). Mitigating aflatoxin contamination in groundnut through A combination of genetic resistance and post-harvest management practices. *Toxins (Basel)* 11, 315. doi:10.3390/toxins11060315

Pandey, M. K., Gangurde, S. S., Sharma, V., Pattanashetti, S. K., Naidu, G. K., Faye, I., et al. (2020). Improved genetic map identified major QTLs for drought tolerance-and iron deficiency tolerance-related traits in groundnut. *Genes* 12, 37. doi:10.3390/genes12010037

Pandurangaiah, M., Rao, L. G., Sudhakarbabu, O., Nareshkumar, A., Kiranmai, K., Lokesh, U., et al. (2014). Overexpression of horsegram (*Macrotyloma uniflorum* Lam. Verdc.) NAC transcriptional factor (*MuNAC4*) in groundnut confers enhanced drought tolerance. *Mol. Biotechnol.* 56, 758–769. doi:10.1007/s12033-014-9754-0

Parmley, K., Nagasubramanian, K., Sarkar, S., Ganapathysubramanian, B., and Singh, A. K. (2019). Development of optimized phenomic predictors for efficient plant breeding decisions using phenomic-assisted selection in soybean. *Plant Phenomics* 2019, 5809404. doi:10.34133/2019/5809404

Parsons, R., and Sunley, R. J. (2001). Nitrogen nutrition and the role of root–shoot nitrogen signalling particularly in symbiotic systems. *J. Exp. Bot.* 52, 435–443. doi:10.1093/jexbot/52.suppl_1.435

Patel, K. G., Thankappan, R., Mishra, G. P., Mandaliya, V. B., Kumar, A., and Dobaria, J. R. (2017). Transgenic peanut (*Arachis hypogaea* L.) overexpressing mtlD gene showed improved photosynthetic, physio-biochemical, and yield-parameters under soil-moisture deficit Stress in lysimeter system. *Front.Plant Sci.* 8, 1881. doi:10.3389/fpls.2017.01881

Patel, J. D., Wang, M. L., Dang, P., Butts, C., Lamb, M., and Chen, C. Y. (2022). Insights into the genomic architecture of seed and pod quality traits in the U.S. Peanut mini-core diversity panel. *Plants* 11, 837. doi:10.3390/plants11070837

Patil, M., Ramu, S. V., Jathish, P., Sreevathsa, R., Reddy, P. C., Ganesh, P., et al. (2014). Overexpression of AtNAC2 (ANAC092) in groundnut (*Arachis hypogaea* L.) improves abiotic stress tolerance. *Plant Biotechnol. Rep.* 8, 161–169. doi:10.1007/s11816-013-0305-0

Patrick, A., Pelham, S., Culbreath, A., Holbrook, C. C., De Godoy, I. J., and Li, C. (2017). High throughput phenotyping of tomato spot wilt disease in peanuts using unmanned aerial systems and multispectral imaging. *IEEE Instru. Meas. Mag.* 20, 4–12. doi:10.1109/MIM.2017.7951684

Peng, Z., Liu, F. X., Wang, L. P., Zhou, H., Paudel, D., Tan, L. B., et al. (2017). Transcriptome profiles reveal gene regulation of peanut (*Arachis hypogaea* L.) nodulation. *Sci. Rep.* 7, 40066. doi:10.1038/srep40066

Peng, Z., Chen, H., Tan, L., Shu, H., Varshney, R. K., Zhou, Z., et al. (2021). Natural polymorphisms in a pair of NSP2 Homoeologs can cause loss of nodulation in peanut. *J. Exp. Bot.* 72, 1104–1118. doi:10.1093/jxb/eraa505

Peoples, M. B., Pate, J. S., Atkins, C. A., and Bergersen, F. J. (1986). Nitrogen nutrition and xylem sap composition of peanut (*Arachis hypogaea* L. cv Virginia Bunch) 1. *Plant Physiol.* 82, 946–951. doi:10.1104/pp.82.4.946

Pilon, C,, Snider, J. L., Sobolev, V., Chastain, D. R., Sorensen, R. B., Meeks, C. D., et al. (2018). Assessing stomatal and non-stomatal limitations to carbon assimilation under progressive drought in peanut (*Arachis hypogaea* L.). *J. Plant Physiol.* 231, 124–134. doi:10.1016/j.jplph.2018.09.007

Pimratch, S., Jogloy, S., Vorasoot, N., Toomsan, B., Patanothai, A., and Holbrook, C. C. (2007). Relationship between biomass production and nitrogen fixation under drought-stress conditions in peanut genotypes with different levels of drought resistance. *J. Agron. Crop Sci.* 194, 15–25. doi:10.1111/j.1439-037X.2007.00286.x

Polania, J. A., Poschenrieder, C., Beebe, S., and Rao, I. M. (2016). Effective use of water and increased dry matter partitioned to grain contribute to yield of common bean improved for drought resistance. *Front. Plant Sci.* 7, 660. doi:10.3389/fpls.2016.00660

Prasad, P. V. V., Craufurd, P. Q., and Summerfield, R. J. (1999). Sensitivity of peanut to timing of heat stress during reproductive development. *Crop Sci.* 39, 1352–1357. doi:10.2135/cropsci1999.3951352x

Prasad, P. V. V., Boote, K. J., Hartwell, A. L., and Thomas, J. M. G. (2003). Super-optimal temperatures are detrimental to peanut (*Arachis hypogaea* L.) reproductive processes and yield at both ambient and elevated carbon dioxide. *Glob. Change Biol.* 9, 1775–1787. doi:10.1046/j.1365-2486.2003.00708.x

Pruthvi, V., Narasimhan, R., and Nataraja, K. N. (2014). Simultaneous expression of abiotic stress responsive transcription factors, *AtDREB2A, AtHB7* and *AtABF3* improves salinity and drought tolerance in peanut (*Arachis hypogaea* L.). *PLoS ONE* 9, e111152. doi:10.1371/journal.pone.0111152

Qin, H., Gu, Q., Zhang, J., Sun, L., Kuppu, S., Zhang, Y., et al. (2011). Regulated expression of an isopentenyltransferase gene (*IPT*) in peanut significantly improves drought tolerance and increases yield under field conditions. *Plant Cell Physiol.* 52, 1904–1914. doi:10.1093/pcp/pcr125

Quilambo, O. A. (2004). Proline content, water retention capability and cell membrane integrity as parameters for drought tolerance in two peanut cultivars. *South Afr. J. Bot.* 70, 227–234. doi:10.1016/S0254-6299(15)30239-8

Rachaputi, N. C. (2003). "Environmental characterization of experimental sites in India and Australia,". Report of a workshop held at ICRISAT Center, Andhra Pradesh, India, 25-27 February 2002 in *Breeding for drought-resistant peanuts*. Editors A. W. Cruickshank, N. C. Rachaputi, G. C. Wright, and S. N. Nigam, 61–66. ACIAR Proceedings No. 112.

Raj, R., Walker, J. P., Vinod, V., Pingale, R., Naik, B., and Jagarlapudi, A. (2021). Leaf water content estimation using top-of-canopy airborne hyperspectral data. *Int. J. Appl. Earth Obs. Geoinf.* 102, 102393. doi:10.1016/j.jag.2021.102393

Ramu, V. S., Swetha, T. N., Sheela, S. H., Babitha, C. K., Rohini, S., Reddy, M. K., et al. (2016). Simultaneous expression of regulatory genes associated with specific drought-adaptive traits improves drought adaptation in peanut. *Plant Biotechnol. J.* 14, 1008–1020. doi:10.1111/pbi.12461

Ratnakumar, P., and Vadez, V. (2011). Groundnut (*Arachis hypogaea*) genotypes tolerant to intermittent drought maintain a high harvest index and have small leaf canopy under stress. *Funct. Plant Biol.* 38, 1016–1023. doi:10.1071/FP11145

Ravi, K., Vadez, V., Isobe, S., Mir, R. R., Guo, Y., Nigam, S. N., et al. (2011). Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (*Arachis hypogaea* L.). *Theor. Appl. Genet.* 122, 1119–1132. doi:10.1007/s00122-010-1517-0

Raza, A., Sharif, Y., Chen, K., Wang, L., Fu, H., Zhuang, Y., et al. (2022). Genome-wide characterization of ascorbate peroxidase gene family in peanut (*Arachis hypogea* L.) revealed their crucial role in growth and multiple stress tolerance. *Front. Plant Sci.* 13, 962182. doi:10.3389/fpls.2022.962182

Rebetzke, G. J., Condon, A. G., Richards, R. A., and Farquhar, G. D. (2002). Selection for reduced carbon isotope discrimination increases aerial biomass and grain yield of rainfed bred wheat. *Crop Sci.* 42, 739–745. doi:10.2135/cropsci2002.0739

Reddy, T. Y., Reddy, V. R., and Anbumozhi, V. (2003). Physiological responses of groundnut (*Arachis hypogaea* L.) to drought stress and its amelioration: A critical review. *Plant Growth Regul.* 41, 75–88. doi:10.1556/aagr.51.2003.2.9

Rehman, T. U., Ma, D., Wang, L., Zhang, L., and Jin, J. (2020). Predictive spectral analysis using an end-to-end deep model from hyperspectral images for high-throughput plant phenotyping. *Comp. Electron. Agric.* 177, 105713. doi:10.1016/j.compag.2020.105713

Riley, K. (2015). NOAA National Centers for Environmental Information, Monthly Drought Report for Annual 2015, published online January 2016. Available at: https://www.ncei.noaa.gov/access/monitoring/monthly-report/drought/201513.

Ribaut, J.-M., Betran, J., Monneveux, P., and Setter, T. (2009). "Drought tolerance in maize," in *Handbook of maize: Its biology*. Editors J. L. Bennetzen and S. C. Hake (New York, NY: Springer), 311–344.

Richards, R. A., Rebetzke, G. J., Condon, A. G., and van Herwaarden, A. F. (2002). Breeding opportunities for increasing the efficiency of water use and crop yield in temperate cereals. *Crop Sci.* 42, 111–121. doi:10.2135/cropsci2002.0111

Rincent, R., Charpentier, J. P., Faivre-Rampant, P., Paux, E., Le Gouis, J., Bastien, C., et al. (2018). Phenomic selection is a low-cost and high-throughput method based on indirect predictions: Proof of concept on wheat and poplar. *Genes|Genomes|Genetics* 8, 3961–3972. doi:10.1534/g3.118.200760

Rowland, D. L., Faircloth, W. H., Payton, P., Tissue, D. T., Ferrell, J. A., Sorensen, R. B., et al. (2012a). Primed acclimation of cultivated peanut (*Arachis hypogaea* L.) through the use of deficit irrigation timed to crop developmental periods. *Agric. Water Manag.* 113, 85–95. doi:10.1016/j.agwat.2012.06.023

Rowland, D., Puppala, N., Beasley, J., Burow, M., Gorbet, D., Jorden, D., et al. (2012b). Variation in carbon isotope ratio and its relation to other traits in peanut breeding lines and cultivars from U.S. Trials. *J. Plant Breed. Crop Sci.* 4, 144–155. doi:10.5897/JPBCS12.031

Rustgi, S., Kakati, J. P., Jones, Z. T., Zoong Lwe, Z. S., and Narayanan, S. (2021). Heat tolerance as a function of membrane lipid remodeling in the major US oilseed crops (soybean and peanut). *J. Plant Biochem. Biotechnol.* 30, 652–667. doi:10.1007/s13562-021-00729-2

Sakuma, Y., Maruyama, K., Qin, F., Osakabe, Y., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2006). Dual function of an Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 103, 18822–18827. doi:10.1073/pnas.0605639103

Samanta, M. K., Dey, A., and Gayen, S. (2016). CRISPR/Cas9: An advanced tool for editing plant genomes. *Trans. Res.* 25, 561–573. doi:10.1007/s11248-016-9953-5

Sanz-Saez, A., Maw, M. J. W., Polania, J. A., Rao, I. M., Beebe, S. E., and Fritschi, F. B. (2019). Using carbon isotope discrimination to assess genotypic differences in drought resistance of parental lines of common bean. *Crop Sci.* 59, 2153–2166. doi:10.2135/cropsci2019.02.0085

Sarić, R., Nguyen, V. D., Burge, T., Berkowitz, O., Trtílek, M., Whelan, J., et al. (2022). Applications of hyperspectral imaging in plant phenotyping. *Trends Plant Sci.* 27 (3), 301–315. doi:10.1016/j.tplants.2021.12.003

Sarkar, T., Thankappan, R., Kumar, A., Mishra, G. P., and Dobaria, J. R. (2014). Heterologous expression of the AtDREB1A gene in transgenic peanut-conferred tolerance to drought and salinity stresses. *PLoS ONE* 9 (12), e110507. doi:10.1371/journal.pone.0110507

Sarkar, S., Cazenave, A. B., Oakes, J., McCall, D., Thomason, W., Abbot, L., et al. (2020). High-throughput measurement of peanut canopy height using digital surface models. *Plant Phenome J.* 3, e20003. doi:10.1002/ppj2.20003

Sarkar, S., Cazenave, A. B., Oakes, J., McCall, D., Thomason, W., Abbott, L., et al. (2021). Aerial high-throughput phenotyping of peanut leaf area index and lateral growth. *Sci. Rep.* 11, 21661. doi:10.1038/s41598-021-00936-w

Scarpeci, T. E., Zanor, M. I., Carrillo, N., Mueller-Roeber, B., and Valle, E. M. (2008). Generation of superoxide anion in chloroplasts of *Arabidopsis thaliana* during active photosynthesis: A focus on rapidly induced genes. *PlantMol. Biol.* 66, 361–378. doi:10.1007/s11103-007-9274-4

Scheben, A., Wolter, F., Batley, J., Puchta, H., and Edwards, D. (2017). Towards CRISPR/Cas crops–bringing together genomics and genome editing. *New Phytol.* 216, 682–698. doi:10.1111/nph.14702

Selvaraj, G. M., Narayana, M., Schubert, A. M., Ayers, J. L., Baring, M. R., and Burow, M. D. (2009). Identification of QTLs for pod and kernel traits in cultivated peanut by bulked segregant analysis. *Electron. J. Biotechnol.* 12 (2), 3–4.

Selvaraj, M. G., Burow, G., Burke, J. J., Belamker, V., Puppala, N., and Burow, M. D. (2011). Heat stress screening of peanut (*Arachis hypogaea* L.) seedlings for acquired thermotolerance. *Plant Growth Regul.* 65, 83–91. doi:10.1007/s10725-011-9577-y

Shah, Z., Iqbal, A., Khan, F. U., Khan, H. U., Durrani, F., and Ahmad, M. Z. (2020). Genetic manipulation of pea (Pisum sativum L.) with Arabidopsis's heat shock factor HsfA1d improves ROS scavenging system to confront thermal stress. *Genet. Resour. Crop Evol.* 67, 2119–2127. doi:10.1007/s10722-020-00966-9

Shaibu, A. S., Sneller, C., Motagi, B. N., Chepkoech, J., Chepngetich, M., Miko, Z. L., et al. (2020). Genome-wide detection of SNP markers associated with four physiological traits in groundnut (*Arachis hypogaea* L.) mini core collection. *Agronomy* 10, 192. doi:10.3390/agronomy10020192

Sharma, K. K., and Anjaiah, V. (2000). An efficient method for the production of transgenic plants of peanut (*Arachis hypogaea* L.) through *Agrobacterium tumefaciens*-mediated genetic transformation. *Plant Sci.* 159, 7–19. doi:10.1016/s0168-9452(00)00294-6

Sharma, K. K., and Ortiz, R. (2000). Program for the application of genetic transformation for crop improvement in the semi-arid tropics. *Vitro Cell. Dev. Biol. Plant* 36, 83–92. doi:10.1007/s11627-000-0019-1

Sharma, K. K., Pothana, A., Prasad, K., Shan, D., Kaur, J., Bhatnagar, D., et al. (2018). Peanuts that keep aflatoxin at bay: A threshold that matters. *Plant Biotechnol. J.* 16, 1024–1033. doi:10.1111/pbi.12846

Sharma, S. (2017). Prebreeding using wild species for genetic enhancement of grain legumes at ICRISAT. *Crop Sci.* 57, 1132–1144. doi:10.2135/cropsci2017.01.0033

Shekoofa, A., Rosas-Anderson, P., Sinclair, T. R., Balota, M., and Isleib, T. G. (2015). Measurement of limited-transpiration under high vapor pressure deficit for peanut in chambers and in field. *Agron. J.* 107, 1019–1024. doi:10.2134/agronj14.0570

Sinclair, T. R., Leilah, A. A., and Schreffler, A. K. (1995). Peanut nitrogen fixation (C2H2 reduction) response to soil dehydration. *Peanut Sci.* 22, 162–166. doi:10.3146/i0095-3679-22-2-17

Sinclair, T., Devi, J., Shekoofa, A., Choudhary, S., Sadok, W., Vadez, V., et al. (2017). Limited transpiration response to high vapor pressuredDeficit in crop ppecies. *J. Plant Sci.* 260, 109–118. doi:10.1016/j.plantsci.2017.04.007

Sinclair, T. R. (2000). Model analysis of plant traits leading to prolonged crop survival during severe drought. *Field Crops Res.* 68, 211–217. doi:10.1016/S0378-4290(00)00125-8

Sinclair, T. R. (2011). Challenges in breeding for yield increase for drought. *Trends Plant Sci.* 16, 289–293. doi:10.1016/j.tplants.2011.02.008

Singh, J., and Thakur, J. K. (2018). "Photosynthesis and abiotic stress in plants," in *Biotic and abiotic stress tolerance in plants* (Singapore: Springer), 27–46.

Songsri, P., Jogloy, S., Kesmala, T., Vorasoot, N., Akkasaeng, C., Patanothai, A., et al. (2008). Heritability of drought resistance traits and correlation of drought resistance and agronomic traits in peanut. *Crop Sci.* 48, 2245–2253. doi:10.2135/cropsci2008.04.0228

Srinivasan, A., Takeda, H., and Senboku, T. (1996). Heat tolerance in food legumes as evaluated by cell membrane thermostability and chlorophyll fluorescence techniques. *Euphytica* 88, 35–45. doi:10.1007/bf00029263

Stansell, J. R., and Pallas, J. E., Jr. (1985). Yield and quality response of florunner peanut to applied drought at several growth Stages1. *Peanut Sci.* 12, 64–70. doi:10.3146/pnut.12.2.0005

Steketee, C. J., Sinclair, T. R., Riar, M. K., Schapaugh, W. T., and Li, Z. (2019). Unraveling the genetic architecture for carbon and nitrogen related traits and leaf hydraulic conductance in soybean using genome-wide association analyses. *BMC Genom* 20, 811. doi:10.1186/s12864-019-6170-7

Sukanth, B. S. (2022). *Molecular analysis of physiological and productivity traits under high-temperature stress in groundnut (*Arahis hypogaea *L.)*. MSc thesis. Dharwad, India: University of Agricultural Sciences.

Tardieu, F., and Tuberosa, R. (2010)). Dissection and modelling of abiotic stress tolerance in plants. *Curr. Opin. Plant Biol.* 13 (2), 206–212. doi:10.1016/j.pbi.2009.12.012

Thoppurathu, F. J., Ghorbanzadeh, Z., Vala, A. K., Hamid, R., and Joshi, M. (2022). Unravelling the treasure trove of drought-responsive genes in wild-type peanut through transcriptomics and physiological analyses of root. *Funct. Integr. Genom.* 22, 215–233. doi:10.1007/s10142-022-00833-z

U.S. Department of Agriculture; Agricultural Research Service. (2019), Available at: https://www.ars.usda.gov/oc/fnrb/2019/

Upadhyaya, H. D., and Ortiz, R. (2001). A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. *Theor. Appl. Genet.* 102, 1292–1298.

Upadhyaya, H. D., Bramel, P. J., Ortiz, R., and Singh, S. (2002). Geographical patterns of diversity for morphological and agronomic traits in the groundnut germplasm collection. *Euphytica* 128, 191–204. doi:10.1023/A:1020835419262

Upadhyaya, H. D., Bramel, P. J., Ortiz, R., and Singh, S. (2002a). Developing a mini core of peanut for utilization of genetic resources. *Crop Sci.* 42, 2150–2156. doi:10.2135/cropsci2002.2150

Upadhyaya, H. D., Ortiz, R., Bramel, P. J., and Singh, S. (2002b). Development of a groundnut core collection using taxonomical, and morphological descriptors. *Genet. Resour. Crop. Evol.* 50, 39–148. doi:10.1023/a:1022945715628

Upadhyaya, H. D. (2005). Variability for drought resistance related traits in the mini core collection 0of peanut. *Crop Sci.* 45, 1432–1440. doi:10.2135/cropsci2004.0389

Vadez, V., and Ratnakumar, P. (2016). High transpiration efficiency increases pod yield under intermittent drought in dry and hot atmospheric conditions but less so under wetter and cooler conditions in groundnut (*Arachis hypogaea* (L.)). *Field Crops Res.* 193, 16–23. doi:10.1016/j.fcr.2016.03.001

Vara Prasad, P. V., Craufurd, P. Q., Kakani, V. G., Wheeler, T. R., and Boote, K. J. (2001). Influence of temperature during pre- and post anthesis stages of floral development on fruit-set and pollen germination in peanut. *Aust. J. Plant Physiol.* 28, 233–240.

Varshney, R. K., Pandey, M. K., Janila, P., Nigam, S. N., Sudini, H., Gowda, M. V. C., et al. (2014). Marker-assisted introgression of a QTL region to improve rust resistance in three elite and popular varieties of peanut (*Arachis hypogaea* L.). *Theor. Appl. Genet.* 127, 1771–1781. doi:10.1007/s00122-014-2338-3

Varshney, R. K. (2016). Exciting journey of 10 years from genomes to fields and markets: Some success stories of genomics-assisted breeding in chickpea, pigeonpea and groundnut. *Plant Sci.* 242, 98–107. doi:10.1016/j.plantsci.2015.09.009

Vemanna, R. S., Chandrashekar, B. K., Rao, H. M. H., Sathyanarayanagupta, S. K., Sarangi, K. S., Nataraja, K. N., et al. (2013). A modified multisite gateway cloning strategy for consolidation of genes in plants. *Mol. Biotechnol.* 53, 129–138. doi:10.1007/s12033-012-9499-6

Venkatesh, B., Reddy, V. A., Lokesh, U., Kiranmai, K., Johnson, A. A. M., Pandurangaiah, M., et al. (2018). Multigenic groundnut transgenics: An advantage over traditional single gene traits in conferring abiotic stress tolerance: A review. *J. Agri. Allied Sci.* 7, 113–120.

Venkatesh, B., Vennapusa, A. R., Kumar, N. J., Jayamma, N., Reddy, B. M., Johnson, A. M. A., et al. (2022). Co-expression of stress-responsive regulatory genes, MuNAC4, MuWRKY3 and MuMYB96 associated with resistant-traits improves drought adaptation in transgenic groundnut (*Arachis hypogaea* L.) plants. *Front. Plant Sci.* 13, 1055851. doi:10.3389/fpls.2022.1055851

Vessey, J. K., Pawlowski, K., and Bergman, B. (2005). Root-based N_2-fixing symbioses: Legumes, actinorhizal plants, Parasponia sp. and cycads. *Plant Soil* 274, 51–78. doi:10.1007/s11104-005-5881-5

Waltz, E. (2016). Gene-edited CRISPR mushroom escapes US regulation. *Nature* 532, 293. doi:10.1038/nature.2016.19754

Wang, C., Wan, Y., Liu, F., and Zhang, K. (2018). Study on drought resistance of peanut varieties at seedling stage under PEG6000 osmotic stress. *Shandong Agric. Sci.* 50, 65–71.

Wang, Y., Wisniewski, M., Meilan, R., Cui, M., and Fuchigami, L. (2006). Transgenic tomato (Lycopersicon esculentum) overexpressing cAPX exhibits enhanced tolerance to UV-B and heat stress. *J. Appl. Hortic.* 8, 87–90. doi:10.37855/jah.2006.v08i02.21

Wang, P., Song, H., Li, C., Li, P., Li, A., Guan, H., et al. (2017). Genome-wide dissection of the heat shock transcription factor family genes in Arachis. *Front. Plant Sci.* 8, 106. doi:10.3389/fpls.2017.00106

Wang, W., Wang, W., Pan, Y., Tan, C., Li, H., Chen, Y., et al. (2022). A new gain-of-function OsGS2/GRF4 allele generated by CRISPR/Cas9 genome editing increases rice grain size and yield. *Crop J* 10, 1207–1212. doi:10.1016/j.cj.2022.01.004

Wang, H., Lei, Y., Wan, L., Yan, L., Lv, J., Dai, X., et al. (2016). Comparative transcript profiling of resistant and susceptible peanut post-harvest seeds in response to aflatoxin production by *Aspergillus flavus*. *BMC Plant Biol.* 16, 54. doi:10.1186/s12870-016-0738-z

Wang, H., Qian, X., Zhang, L., Xu, S., Li, H., Xia, X., et al. (2018). A method of high throughput monitoring crop physiology using chlorophyll fluorescence and multispectral imaging. *Front. Plant Sci.* 9, 407. doi:10.3389/fpls.2018.00407

Wang, M. L., Grusak, M. A., Chen, C. Y., Tonnis, B., Barkley, N. A., and Evans, E. (2016). Seed protein percentage and mineral concentration variability and their correlation with other seed quality traits in the U.S. peanut mini-core collection. *Peanut Sci.* 43, 119–125. doi:10.3146/PS15-15.1

Wang, Q. L., Chen, J. H., He, N. Y., and Guo, F. Q. (2018). Metabolic reprogramming in chloroplasts under heat stress in plants. *Int. J. Mol. Sci.* 19, 849. doi:10.3390/ijms19030849

Weiß, T. M., Zhu, X., Leiser, W. L., Li, D., Liu, W., Schipprack, W., et al. (2022). Unraveling the potential of phenomic selection within and among diverse breeding material of maize (*Zea mays* L.). *G3 Genes, Genomes, Genet.* 12, jkab445. doi:10.1093/g3journal/jkab445

White, J. W., Castillo, J. A., and Ehleringer, J. (1990). Associations between productivity, root growth and carbon isotope discrimination in *Phaseolus vulgaris* under water deficit. *Funct. Plant Biol.* 17, 189–198. doi:10.1071/pp9900189

Wright, G. C., Rao, R. C. N., and Farquhar, G. D. (1994). Water-use efficiency and carbon isotope discrimination in peanut under water deficit conditions. *Crop Sci.* 34, 92–97. doi:10.2135/cropsci1994.0011183X003400010016x

Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *PNAS* 112, 3570–3575. doi:10.1073/pnas.1420294112

Xu, W., Yu, G., Zare, A., Zurweller, B., Rowland, D. L., Reyes-Cabrera, J., et al. (2020). Overcoming small minirhizotron datasets using transfer learning. *Comp. Electron. Agric.* 175, 105466. doi:10.1016/j.compag.2020.105466

Yang, X., Luo, L., Yu, W., Mo, B., and Liu, L. (2019). Recent advances in the acclimation mechanisms and genetic improvement of peanut for drought tolerance. *Agric. Sci.* 10, 1178–1193. doi:10.4236/as.2019.109088

Yao, J., Sun, D., Cen, H., Xu, H., Weng, H., Yuan, F., et al. (2018). Phenotyping of *Arabidopsis* drought stress response using kinetic chlorophyll fluorescence and multicolor fluorescence imaging. *Front. Plant Sci.* 9, 603. doi:10.3389/fpls.2018.00603

Yeh, D. M., and Lin, H. F. (2003). Thermostability of cell membranes as a measure of heat tolerance and relationship to flowering delay in chrysanthemum. *J. Am. Soc. Hortic. Sci.* 128, 656–660. doi:10.21273/JASHS.128.5.0656

Yuan, H., Bennett, R. S., Wang, N., and Chamberlin, K. D. (2019). Development of a peanut canopy measurement system using a ground-based lidar sensor. *Front. Plant Sci.* 10, 203. doi:10.3389/fpls.2019.00203

Zhang, Y., Xia, H., Yuan, M., Zhao, C., Li, A., and Wang, X. (2011). Cloning and expression analysis of peanut (*Arachis hypogaea* L.) CHI gene. *Electron. J. Biotechnol.* 15, 9–12. doi:10.2225/vol15-issue1-fulltext-6

Zhang, J., Liu, B., Zhang, L., Wang, Y., Zheng, H., et al. (2015). *Hsf* and *hsp* gene families in *populus*: Genome-wide identification, organization and correlated expression during development and in stress responses. *BMC Genomics* 16, 181. doi:10.1186/s12864-015-1398-3

Zhang, Q., Dang, P., Chen, C., Feng, Y., Batchelor, W., Lamb, M., et al. (2022). Tolerance to mid-season drought in peanut can be achieved by high water use efficiency or high efficient use of water. *Crop Sci.* 62, 1948–1966. doi:10.1002/csc2.20806

Zhang, H., Wang, M. L., Schaefer, R., Dang, P., Jiang, T., and Chen, C. (2019). GWAS and coexpression network reveal ionomic variation in cultivated peanut. *J. Agric. Food Chem.* 67, 12026–12036. doi:10.1021/acs.jafc.9b04939,

Zhao, Y., Chan, Z., Gao, J., Xing, L., Cao, M., Chunmei, Y., et al. (2016). ABA receptor PYL9 promotes drought resistance and leaf senescence. *Proc. Natl. Acad. Sci.* 113, 1949–1954. doi:10.1073/pnas.1522840113

Zhu, X., Maurer, H. P., Jenz, M., Hahn, V., Ruckelshausen, A., Leiser, W. L., et al. (2022). The performance of phenomic selection depends on the genetic architecture of the target trait. *Theor. Appl. Genet.* 135, 653–665. doi:10.1007/s00122-021-03997-7

Zoong Lwe, Z. S., Welti, R., Anco, D., Rustgi, S., and Narayanan, S. (2020). Heat stress elicits remodeling in the anther lipidome of peanut. *Sci. Rep.* 10, 22163. doi:10.1038/s41598-020-78695-3

Zou, K., Kim, K.-S., Kang, D., Kim, M.-C., Ha, J., Moon, J.-K., et al. (2022). Genome-wide association study of leaf chlorophyll content using high-density SNP array in peanuts (*Arachis hypogaea* L.). *Agronomy* 12, 152. doi:10.3390/agronomy12010152

Zurweller, B. A., Rowland, D. L., Tillman, B. L., Payton, P. P., Migliaccio, K., Wright, D., et al. (2018). Assessing above- and below-ground traits of disparate peanut genotypes for determining adaptability to soil hydrologic conditions. *Field Crops Res.* 219, 98–105. doi:10.1016/j.fcr.2018.01.020

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 18 April 2023
DOI 10.3389/fgene.2023.1085024

OPEN ACCESS

EDITED BY
Deepmala Sehgal,
Syngenta, United Kingdom

REVIEWED BY
Muhammad Uzair,
National Institute for Genomics and Advanced Biotechnology (NIGAB), Pakistan
Palakolanu Sudhakar Reddy,
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

*CORRESPONDENCE
Rajendra Kumar,
rajendrak64@yahoo.co.in
Aladdin Hamwieh,
a.hamwieh@cgiar.org

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Plant Genomics,
a section of the journal Frontiers in Genetics

RECEIVED 31 October 2022
ACCEPTED 24 March 2023
PUBLISHED 18 April 2023

CITATION
Singh C, Kumar R, Sehgal H, Bhati S, Singhal T,
Gayacharan, Nimmy MS, Yadav R, Gupta SK, Abdallah NA, Hamwieh A and Kumar R (2023), Unclasping potentials of genomics and gene editing in chickpea to fight climate change and global hunger threat.
Front. Genet. 14:1085024.
doi: 10.3389/fgene.2023.1085024

COPYRIGHT
© 2023 Singh, Kumar, Sehgal, Bhati, Singhal, Gayacharan, Nimmy, Yadav, Gupta, Abdallah, Hamwieh and Kumar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Unclasping potentials of genomics and gene editing in chickpea to fight climate change and global hunger threat

Charul Singh[1†], Ramesh Kumar[2†], Hansa Sehgal[3†], Sharmista Bhati[4], Tripti Singhal[5], Gayacharan[6], M. S. Nimmy[7], Renu Yadav[8], Santosh Kumar Gupta[9], Naglaa A. Abdallah[10], Aladdin Hamwieh[11]* and Rajendra Kumar[5]*

[1]USBT, Guru Govind Singh Indraprastha University, Delhi, India, [2]Department of Biochemistry, University of Allahabad Prayagraj, Prayagraj, India, [3]Department of Biological Sciences, Birla Institute of Technology and Sciences, Pilani, India, [4]School of Biotechnology, Gautam Buddha University, Greater Noida, India, [5]Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi, India, [6]Division of Germplasm Evaluation, ICAR- National Bureau of Plant Genetic Resources, New Delhi, India, [7]ICAR-National Institute for Plant Biotechnology, New Delhi, India, [8]AIOA, Amity University, Noida, India, [9]DBT-National Institute of Plant Genome Research, New Delhi, India, [10]Faculty of Agriculture, Cairo University, Cairo, Egypt, [11]The International Center for Agricultural Research in the Dry Areas (ICARDA), Cairo, Egypt

Genomics and genome editing promise enormous opportunities for crop improvement and elementary research. Precise modification in the specific targeted location of a genome has profited over the unplanned insertional events which are generally accomplished employing unadventurous means of genetic modifications. The advent of new genome editing procedures viz; zinc finger nucleases (ZFNs), homing endonucleases, transcription activator like effector nucleases (TALENs), Base Editors (BEs), and Primer Editors (PEs) enable molecular scientists to modulate gene expressions or create novel genes with high precision and efficiency. However, all these techniques are exorbitant and tedious since their prerequisites are difficult processes that necessitate protein engineering. Contrary to first generation genome modifying methods, CRISPR/Cas9 is simple to construct, and clones can hypothetically target several locations in the genome with different guide RNAs. Following the model of the application in crop with the help of the CRISPR/Cas9 module, various customized Cas9 cassettes have been cast off to advance mark discrimination and diminish random cuts. The present study discusses the progression in genome editing apparatuses, and their applications in chickpea crop development, scientific limitations, and future perspectives for biofortifying cytokinin dehydrogenase, nitrate reductase, superoxide dismutase to induce drought resistance, heat tolerance and higher yield in chickpea to encounter global climate change, hunger and nutritional threats.

KEYWORDS

chickpea improvement, climate change, CRISPR/Cas-9, genome editing, hunger threat, TALENs, zinc finger nuclease

1 Introduction

Since their origin, land plants have evolved in an essentially hostile environment. These factors deleteriously disturb plant productivity, growth, development and are referred to as stress in plants. Plant stress is due to drastic changes in salinity, temperatures, heavy metals, soil moisture levels, and ultraviolet (UV) emissions. Stresses including both abiotic and biotic are posturing a great menace to agriculture, ecosystems, and noteworthy production losses (Wang et al., 2003; Wani et al., 2016). According to a published report (FAO, 2019), abiotic stress affects roughly 96.5 percent of worldwide rural land areas (Cramer et al., 2011). Crop yields in lower latitude regions are currently declining, whereas yields in higher latitude regions are increasing (Iizumi et al., 2018; IPCC, 2019). Extreme weather occurrences, according to the Intergovernmental Panel on Climate Change (IPCC, 2019), will interrupt and reduce the global food supply resulting in higher food costs. The current estimates of a report by UN reveals that after a continuous decline over a decade, numbers of people suffering from hunger crisis have gradually increased since 2015. Data reveals that at present there are around 690 million people who are hungry which equates to 8.9% of the world population. The report further states that a majority of undernourished population have been found living in Asia and more than 250 million live in Africa, where the numbers are increasing at a very fast rate than anywhere else in the world. On the other hand, there are an estimated 2 billion people who lack access to safe, nutritious and adequate food and are exposed to food insecurity. The report explains that if the present trend persists, the number of people affected by hunger and undernourishment will exceed 840 million, i.e., 9.8% of total population (Arora and Mishra, 2022). The Global Hunger Index (GHI) shows that the number of people who lack regular intake of sufficient calories is increasing. India has ranked poorly for GHI position amongst 107 countries as 100th in 2017, 102nd in 2019, and 94th in 2020. This ranking was counterintuitive considering the fifth rank of India in the world economy. However, Indian policymakers have argued that hunger is an emotional subject and there have been many criticisms and rebuttals of the GHI. Thus, GHI is a misleading hunger index as its methodology ignores genetic factors wherein international norms on stunting and wasting may not be applicable to India (Singh et al., 2021).

During the last two decades, stress has increased by more than two folds, majorly attributed to temperature rise, drought, and salinization of agricultural lands. According to a new meta-analysis study, the worldwide average temperature will rise by almost 5°C by 2,100 (Raftery et al., 2017). Increased heavy metal poisoning of agricultural areas is restricting food output while also posing major health dangers to humans (Rehman et al., 2018). Besides abiotic stresses, biotic stresses also induce stresses through infestations with insects, bacteria, fungi, viruses, and nematodes. Although plants have evolved with various kinds of defence systems to survive, such as halophytes have developed a specific organ to emit salt, as seen by Limonium bicolor's salt gland (Yuan et al., 2013; 2016). The available basic information on chickpea for the genomic structure (Singh et al., 2013), genetic resources for Dof genes (Yadav et al., 2016), salinity (Mittal et al., 2015), drought (Singh et al., 2012; Singh et al., 2013; Bhardwaj et al., 2014; Mittal et al., 2014; Kumar et al., 2017; Yadav et al., 2019; Bhardwaj et al., 2021), nitrate reductase (Katoch et al., 2016), superoxide dismutase (Singh A. P. et al., 2022) and appropriate strategies (Chandana et al., 2022; Singh R. K. et al., 2022) are necessary and will facilitate the deployment of biotechnological approaches to develop heritably engineered transgenic chickpea plants with upgraded stress resistance. To combat food scarcity, an amalgamation of outdated plant breeding and novel methodologies such as molecular plant breeding and gene editing must be applied. Targeted genome editing boosted grain size related metrics viz; the number of tillers, and protein quality in rice and corn including several monocots and dicots (Shan et al., 2014; Sedeek et al., 2019). The introgression of quantitative trait loci (QTLs) genomic regions implicated for stress tolerance resulted in the introduction and/or over expression of selected genes into genetically altered plants and appear to be a promising alternative for hastening the breeding of "better" crop plants including chickpea. Thus, genetic engineering, often known as genetically modified (GM) crop technology allows scientists to transfer valuable genes from a completely separate gene pool into the crop plants with the least amount of disturbance to the plant genome and is frequently advocated as an answer for raising yields in crops including chickpea around the world, predominantly in under-developed areas where food insecurity and low crop production are major concerns (Nelson et al., 2007).

Chickpea a member of the fabacean family, one of the extremely significant and second largest leguminous food crops across the globe, has an extraordinary mandate due to the high dietary value of the grain. Today, chickpea ranks third among leguminous food plants for global production, behind field pea (*Pisum sativum L*) and beans (*Phaseolus* spp.) (FAO, 2019). It is cultivated in more than 55 countries across the globe on an estimated 14.56 million hectares area generating 14.78 million tons of total production. Chickpea production, on the other hand, is insufficient to supply the protein requirements of an ever-increasing human population (Reddy et al., 2016; Henchion et al., 2017). A foremost task for crop breeders is enhancing crop production to feed probably ~10 billion worldwide civilization by 2050 (Hickey et al., 2019). Among legumes family members, after common bean, Chickpea is the economically as well as nutritionally important crop plant. However, cultivation of chickpea is limited due to the various abiotic and biotic stress factors. Being rabi crop, it also faces low temperature stress especially during reproductive stage leading to significant loss in its production. Recently, a detailed review focusing on impact of various stresses on chickpea showed how slightest change in condition can alter the development of the plant (Rani et al., 2020; Akinlade et al., 2022). Thus, various strategies have been applied to improve the tolerance of chickpea employing various conventional breeding techniques but time consuming and laborious processes are the challenges faced by breeders in developing a cultivar tolerant to stresses (Jha et al., 2014).

However, genome editing technologies have tremendous effects on plant breeding techniques to guard crop plants against numerous tasks and augment crop yield (Taranto et al., 2018). Editing the target DNA sequence by adding, selecting, or substituting nucleotide bases is a cutting-edge molecular biology technique. The techniques such as ZFNs, TALENs, Base Editors, CRISPR/Cas9, and Primer Editors are currently being used for genome editing. The CRISPR/Cas9 technologies corroborate the utmost operational GE machineries since these are precise, less expensive, speedy, and

consent for numerous site-specific genome editing (Zhu et al., 2017). Hence, in this review article, we are focusing on genetic engineering approaches as comprehensive efforts for biofortifying cytokinin dehydrogenase, nitrate reductase, superoxide dismutase to induce drought resistance, heat tolerance and higher yielding diversities that will upsurge chickpea productivity, usefulness for chickpea growing farmers to encounter global climate change, hunger and nutritional threats.

2 Bottlenecks in chickpea gene editing applications

Presently, India is the world largest producer of chickpea (Khine et al., 2022). Yet we dawdle behind other chickpea growing countries in productivity. Hence, it is important to improve the productivity of chickpea. To sustain chickpea production development of climate resilient cultivars are needed. Scientific community around the globe had put lots of effort to enhance yield of chickpea still not able to reach at significant level. The primary reason is that chickpeas have inherently narrow genetic base as they have been extorted to natural selection, domestication syndrome, founder effect, etc. (Abbo et al., 2003).

Chickpea transformation which was accomplished using cutting-edge biotechnological techniques, is a crucial part for genetic enhancement and a prerequisite for genome editing. The efficient production of transgenic chickpeas is hampered by tissue cultures that refuse to cooperate and the occasional chimerism that is found during transformation. Legumes including chickpea are well known to be both resistant to the uptake and integration of introduced DNA (Yadav et al., 2017) and recalcitrant in terms of regeneration (Ochattet al., 2018). Being recalcitrant in nature chickpea transformation is difficult and a robust transformation method is a prerequisite for researchers to carry out the genetic transformation studies in the crop. Although several labs have reported chickpea transformation, the limitations associated with the reproducibility of the technique (Huda et al., 2000; Das Bhowmik et al., 2019), poor *in vitro* rooting (Polowick et al., 2004), low transformation efficiency, regeneration capacity and non-transmission of genes to subsequent generations (Sarmah et al., 2004) remain problematic. The reports available till date indicate majority of the chickpea transformation works have been carried out using *Bacillus thuringiensis* genes for pod borer resistance (Das et al., 2017). However, recently an agrobacterium mediated transformation system in six cultivars of chickpea with 8.6% efficiency has been established (Sadhu et al., 2022). Further, major factors leading to narrowing of genetic base, utilization of available genetic resources for devising strategies to broadening the genetic base, facilitating the transformation strategies and also provide opportunities for genome editing applications in chickpea have been explained (Singh A. P. et al., 2022).

Chimerism is another challenge due to which recovery of stable transgenic lines decline. For instance, earlier researchers have revealed that the percentages of non-transmitting, chimeric lines in chickpea and lentil, were 22% and 29% respectively (Christou, 1990; Dillen et al., 1997; Sarmah et al., 2004; Celikkol Akcay et al., 2009). The effectiveness of recovering stable transgenic lines is decreased by the presence of chimeric tissues (Christou, 1990; Dillen et al., 1997; Sarmah et al., 2004; Celikkol Akcay et al., 2009). Measures for removing chimerism in legumes including chickpea yet have not been published except a single report on lentil (Celikkol Akcay et al., 2009), which showed reduced chimerism and stable expression of a GUS reporter in successive generations.

The carotenoid biosynthesis candidate genes have been identified as a knockout target to increase the carotenoid concentration in chickpea (Rezaei et al., 2016). Developing different abiotic stress tolerant lines would be the future genome editing target in case of chickpea. First report of CRISPR/Cas9-mediated editing of chickpea protoplasts was recently published. Scientists from Australia's Royal Melbourne Institute of Technology (RMIT) have demonstrated the feasibility of gene editing in chickpea laying a technical foundation for future trait discovery and improvement by creating knockouts of 4-coumarate ligase (4CL) and Reveille 7 (RVE7) genes associated with drought tolerance in chickpea (Badhan et al., 2021). Thus, it is evident that though genome editing is progressing well, the recalcitrant nature of the crop for *in vitro* gene transfer and regeneration is a major challenge and successful chickpea traits improvement will remain dependent on the efficient plant transformation and regeneration protocols.

3 Genome editing (GE) tools and strategies

GE technologies have been continuously in use for dissimilar plants including species such as *Arabidopsis* and major crops such as rice, maize, wheat, and economically less important crops such as strawberries and peanuts. In the majority of the cases, these techniques have been employed for fundamental research as proof-of-concept or to examine gene functions. Several market-oriented qualities such as improved agronomic properties, upgraded quality of food and feed, higher endurance to abiotic and biotic stresses and herbicide tolerance have been addressed. The traditional genetic engineering strategies have several flaws and limitations, one of which is the difficulty of manipulating big genomes in higher plants (Nemudryi et al., 2014). The development of revolutionary tools for procreation and biotechnology, a genetic engineering application area, has attracted a lot of attention, resulting in the rapid development of valuable tools. Genetic modification for targeted gene augmentation is widely used in the field of plant science for both fundamental research and the development of desirable characteristics in commercial crops.

The generations of GM crops rely on randomly inserting new stretches of DNA sequences into the genome. The inserted genome may affect or inactivate other neighbouring genes' activity which is one of the major concerns of this strategy. However, genome editing makes advantage of more contemporary knowledge and technology to allow for the alteration of a definite area of the genome, enhancing the preciseness of the insertion, avoiding cell death, and providing flawless duplication (Voytas, 2013; Voytas and Gao, 2014). Genome editing, also known as genome engineering, is one of the utmost talented machineries applied in biological investigation (Hu et. al., 2008), engineering revolutions and right now a sophisticated tool that allows for precise changes to the genome, using only some of the nucleotides in a living cell's genome sequence. Despite those facilitations, various obstacles exist which include public scepticism about GM crops, which is heightened when "foreign"

TABLE 1 Genetic transformation of Chickpea.

Genotype	Explant	Transgene	Promoter	Gene delivery system	Aim	References
C 235, BG 256, Pusa 362 and Pusa 372	Cotyledonary node	*cry1Ac*	*CaMV35S*	Agrobacterium-mediated	Insect resistance against *H. armigera*	Sanyal et al. (2005)
ICCC37	Epicotyl	*cryIAc*	*CaMV35S*	Agrobacterium-mediated	Insect resistance against *H. armigera*	Indurker et al. (2010)
Annigeri	Cotyledonary node	*P5CS*	*CaMV35S*	Agrobacterium-mediated	Salinity tolerance	Ghanti et al. (2011)
P-362	Cotyledonary node	*cry1Ab* and *cry1Ac*	*CaMV35S* and synthetic constitutive expression promoter (*Pcec*)	Agrobacterium-mediated	Insect resistance	Mehrotra et al. (2011)
DCP 92–3	Embryonic axis	*cry1Ab/cry1Ac*	Rice *actin1* and soybean *msg*	Agrobacterium-mediated	Insect resistance	Ganguly et al. (2014)
Gokce	Mature embryo	miR408	CaMV35S	Agrobacterium-mediated	Drought tolerance	Hajyzadeh et al. (2015)
ICCV 89,314	Single cotyledon with half embryo	*cry1Ac*	*RuBisCO* small subunit and ubiquitin	Agrobacterium-mediated	Insect resistance to target *H. armigera*	Chakraborty et al. (2016)
DCP 92–3	Axillary meristem	*cry1Aabc*	*CaMV35S*	Agrobacterium-mediated	Insect resistance	Das et al. (2017)
PBA HatTrick	Half-embryonic axis	nicotianamine synthase 2 and ferritin	*CaMV35S* and nopaline synthase	Agrobacterium-mediated	Iron biofortifcation	Tan et al. (2018)

genes from remotely related creatures are introduced, as this is viewed as "unnatural," despite mounting shreds of evidences to the differences as natural sweet potato variations are well recognized to include T-DNA from the bacteria *Agrobacterium tumefaciens* and can be seen as "natural GMO" (Rastogi Verma, 2013; Lucht, 2015). GM crop production is costly, and the biosafety education required to come across controlling criteria adds significantly to the cost, which is predicted over $120 million per trait (Lusser et al., 2012). As a result, GM technology could not be utilized to its potential, except in a few crops by a few countries. Similarly, in chickpeas very limited efforts on transformation for a few selected target traits have been accomplished (Table 1). Monitoring of necessities also cause significant delays in product introductions. Targeting gene expression with homologous recombination is a valuable way for obtaining facts on genetic expressions (Capecchi, 2005; Gaj et al., 2013). However, the technique's implementation has been limited owing to its low efficiency, extended study duration, mutagenesis consequences, and off-target impacts. Here, various approaches have been discussed that are/may be used in chickpea genome editing like site-specific recombinase or Site-Specific Nucleases that could be used to modify the genome.

3.1 GE through site-specific recombinase (SSR): A molecular machine for genetic reformation

SSR is a frequently used genetic engineering technique for permanently altering the target genome. Lots of site-specific recombinase systems have been developed to accomplish DNA reorganizations including Cre/loxP and Flp/FRT(Araki et al., 1995; Allen and Weeks, 2005; Allen et al., 2009; Wang et al., 2011). SSRs can be used to manipulate genomes and stimulate or deactivate gene expression in numerous organisms (Wang et al., 2011). Recombinase has been widely utilized to modify the DNA of mammals, yeast, plants, and bacteria by introducing knockout or knock-in mutations into their genomes (Abdallah et al., 2015). One of the benefits of recombinases is that they are not reliant on intracellular repair mechanisms (Abdallah et al., 2015).

SSRs are molecular machines that allow DNA molecules to be cut, paste and editing by adding, removing, or inverting precisely defined DNA segments (Grindley et al., 2006; Gaj et al., 2014). The mechanism incorporates and eliminates the bacteriophage DNA from a definite location in its host genome. *Escherichia coli* was the first example of site-specific recombination in bacteria (Landy, 2015). Each strand of recombining DNA has two core-type sites, which are inverted repeat recombinase binding sites, that flank an identical 7 bp "overlap region" called as O in both DNAs (Rutkai et al., 2006). The two active Ints on one side of the "overlap region" cleave and interchange the top strands of the DNA to form a four-way DNA junction called Holliday junction (HJ), which is subsequently resolved to recombinant products by the other pair of Ints cleaving and trading the bottom strands of the overlap region. Additional DNA sequences that encode binding sites for the second (NTD) DNA binding domain of Int and the accessory DNA bending proteins, IHF, Xis, and Fis are added to two of the four core-type sites. However, some sites are considered necessary either only for excisive recombination between the attL and attR sites, or for integrative recombination between attP (on the phage chromosome) and attB (on the bacterial chromosome), or needed for both reactions (Landy, 2015). Two short DNA sequences are brought together at different positions inside one DNA or in distinct molecules; the DNA fragments are damaged at specified phosphodiester links inside DNA, and the damaged ends are re-

TABLE 2 Application of CRISPR based genome editing approach in plants for biotic, abiotic, and nutritional traits.

Crop	Method	Target gene	Stress/trait	References
Biotic stress				
A. thaliana/	NHEJ	dsDNA of virus (A7, B7, and C3	Beet severe curly top virus resistance	Ji et al. (2014)
Rice	NHEJ	OsERF922 (ethylene responsive	Blast Resistance	Wang et al. (2016)
Bread wheat	NHEJ	TaMLO-A1, TaMLO-B1, and	Powdery mildew resistance	Wang et al. (2014)
Cucumber	NHEJ	eIF4E (eukaryotic translation		Chandrasekaran et al. (2016)
Abiotic stress				
Maize	HDR	ARGOS8	Increased grain yield under drought stress	Shi et al. (2017)
Tomato	NHEJ	SlMAPK3	Drought tolerance	Wang et al. (2017)
A. thaliana	NHEJ	UGT79B2, UGT79B3	Susceptibility to cold, salt, and drought stresses	Zhao et al. (2016)
Rice	HDR, NHEJ	OsPDS, OsMPK2, OsBADH2	Involved in various abiotic stress tolerance	Xie and Yang (2013)
Rice	NHEJ, HDR	OsMPK2, OsDEP1	Yield under stress	Shan et al. (2014)
Nutritional Traits				
Rice	NHEJ	25604 gRNA for 12802 genes	Creating genome wide mutant library	Meng et al. (2017)
Maize	NHEJ	ZmIPK1A ZmIPK andZmMRP4	Phytic acid synthesis	Liang et al. (2014)
Wheat	HDR	TaVIT2	Fe content	Connorton et al., 2017
Soybean	NHEJ	GmPDS11 and GmPDS18	Carotenoid biosynthesis	Du et al. (2016)
Tomato	NHEJ	Rin	Fruit ripening	Ito et al. (2015)
Potato	HDR	ALS1	Herbicide resistance	Butler et al. (2016)
Cassava	NHEJ	MePDS	Carotenoid biosynthesis	Odipio et al. (2017)

joined in a new configuration to generate recombinants (Figure 1A). The identification of sequence and biochemical catalytic phases of this procedure is carried out by the site-specific recombinase, a system-specific enzyme. It is frequently discussed as conventional particular recombination to discriminate it from procedures like homologous recombination (HR), transposition, and non-homologous end-joining, since it does not need DNA synthesis, fragmentation, or cofactors. In more complicated systems, the SSR dimer's "crossover site" neighbouring to "accessory" sequences is acknowledged and assured by the SSR and/or additional proteins (Figure 1B).

The first and foremost application, which was established more than two decades before, is the elimination of a targeted gene from a locus (Dale and Ow, 1991; Russell et al., 1992), which has been monitored quickly by site-specific integration (SSI) to construct accurate one-copy transgene loci and determining complex loci to one copy (De Buck et al., 2007; Srivastava and Ow, 2015). These applications were established for the first time using the Cre-lox system and then lengthened to include additional SSR systems such as Par A, FLP-FRT, phiC31, R-RS, Cin H, and Bxb1 (Sugita et al., 2000; Li Z. et al., 2009; Moon et al., 2011; Thomson et al., 2012). SSR systems can knock down the genome liable on the positioning of the definite sites adjoining the target site. These systems are applicable in numerous plant species and can be used in chickpeas for genetic modification tasks: 1) marker gene elimination and 2) particular external gene insertion *via* site-specific integration.

3.1.1 Basic steps involved in site-specific recombination systems

The three SSR systems identified in the initial 1990s, namely, Cre-lox from bacteriophage P1 of *E. coli*, FLP-FRT from *Saccharomyces cerevisiae*, and R-RS from *Zygosaccharomyces rouxii* are still in use for incorporating diversities in crop plants and can be employed also in chickpea in order to encounter the global climate change and hunger threat. Enzyme recombinase Cre, FLP, or R catalyses recombination in between its analogous recombination sites lox, FRT, or RS, in these recombination systems. Each recombination target site (RTS) up to 34 bp in length has an unequal core/spacer section flanks inverted repeats (RE and LE) that act as recombinase binding locations. Several regions confer the cross over sites, while their unevenness provides the recombination site directions. The reaction steps are 1) identification along with the binding of recombinase dimers to

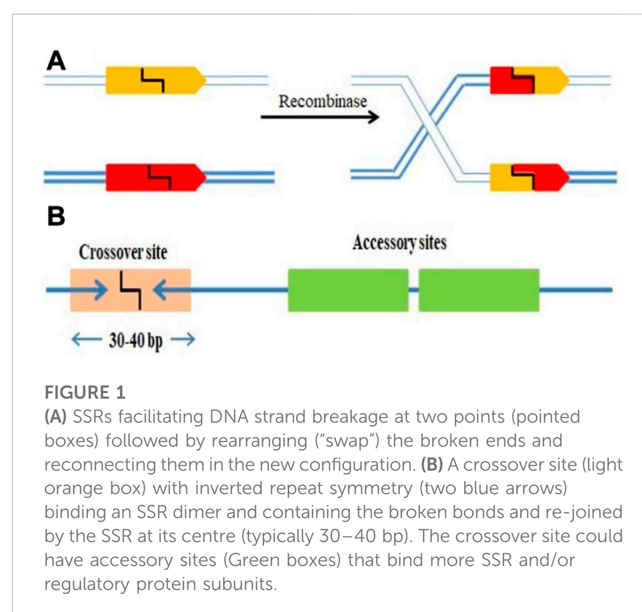


FIGURE 1
(A) SSRs facilitating DNA strand breakage at two points (pointed boxes) followed by rearranging ("swap") the broken ends and reconnecting them in the new configuration. **(B)** A crossover site (light orange box) with inverted repeat symmetry (two blue arrows) binding an SSR dimer and containing the broken bonds and re-joined by the SSR at its centre (typically 30–40 bp). The crossover site could have accessory sites (Green boxes) that bind more SSR and/or regulatory protein subunits.

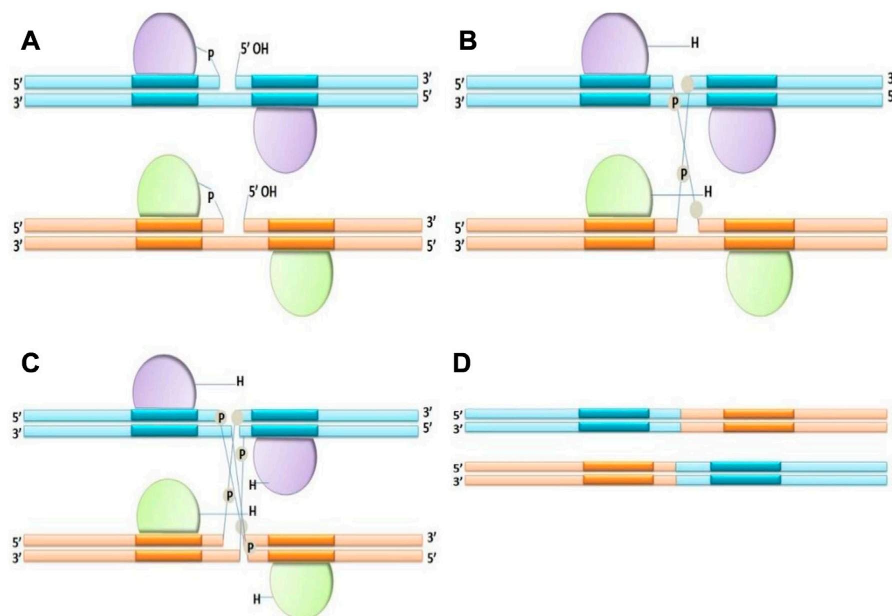


FIGURE 2
Mechanism of Tyrosine recombinase **(A–D)**, making an intermediate Holliday junction by expurgating and interchanging one pair of DNA strands, followed by cutting and interchanging the other couple of strands.

mandatory sites, 2) synaptic complex formation between bound positions, 3) strand exchange and fusion proceedings mediated by recombinase, 4) synaptic complex segregation (Whiteson and Rice, 2008). Some other SSR systems for plant transformation developed in recent years, such as the ΦC31- att and λ-att systems, which consist of a recombinase protein, phiC31 or λ integrase (Int), and catalyse recombination between unrelated recombination sites identified as attB and attP to produce fusion sites attL and attR. The translocation, co-integration, inversion, and deletion can occur subject to the location of attB and attP. However, to catalyze the reverse reaction and reproduce attL and attR hybrid sites from attB and attP, a supplementary excision/resolvase protein is required; therefore, the incorporation reaction is one way in absence of protein.

3.1.2 Site-specific recombinase families

System-specific SSR organization unfolds that the majority of the thousands known site-specific recombination systems are divided into 2 families. These are recombinases of serine and tyrosine, termed after the identification of amino acid residues at the nucleophilic active sites. The side chain of serine or tyrosine breaks a strand by attacking the phosphodiester bond of the DNA and covalently links at the damaged DNA strand end. The phosphodiester link between the DNA preserves bond energy, allowing recombinant strands to be re-joined without the use of cofactors like ATP or additional polymerase or ligase processing. The common features between these two families are crossover site recognition by SSR dimer and catalysis within the SSR tetramer, although their mode of action is different and the proteins have no sequence and organizational similarities (Castillo et al., 2017).

1. ***Tyrosine recombinases:*** Tyrosine side chain attacks a specific phosphodiester bond in the recombination site. When tyrosine recombinase attacks the DNA strands, the hydroxyl group of tyrosine residue covalently bonds to each 3′end of the damaged DNA. Tyrosine recombinases interchange, disrupt and rejoin two DNA strands at once; their reactions continue through a “Holiday” or 4-way connection intermediary, in which 2 strands are non-recombinant while the remaining 2 are recombinant (Figure 2). In experimental genetics and biotechnology, several tyrosine recombinases have been utilized; in fact, the most extensively used SSRs such as Cre (Sauer and Henderson, 1989) and FLP (Golic and Lindquist, 1989) and R (Onouchi et al., 1991) are members of this family.
2. ***Serine recombinases:*** Serine recombinase breaks the DNA strand by the aggression of the phosphodiester with the OH group of serine amino acid and covalently attaches the recombinant DNA to the 5′end at the breakdown. During recombination, serine recombinases create instantaneous double strand breaks in both recombining sites and there is no Holliday junction. A unique subunit rotation mechanism causes recombination by swapping the locations of the cut DNA ends (Figure 3). The upper and lower strand breaks are always 2 bp apart and proportionally positioned in the midpoint of the crossover sites (Smith and Thorpe, 2002; Marshall Stark, 2015). The serine recombinases family contains phiC31 Integrase and phi C31 excisionase (Thorpe and Smith, 1998).

3.1.3 Applications of site-specific recombination systems

The advantages of using SSR over other methods for DNA rearrangement are concise due to its specificity, efficiency, and

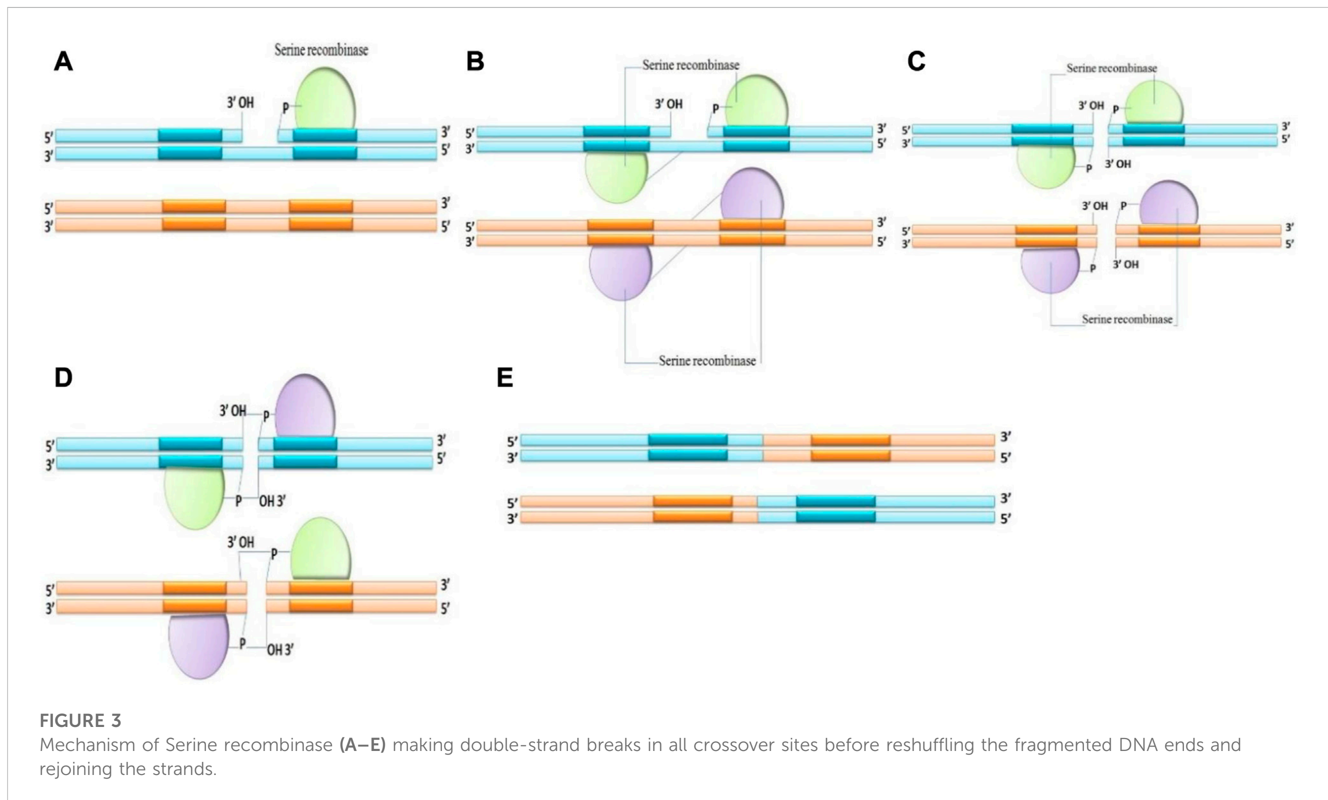


FIGURE 3
Mechanism of Serine recombinase **(A–E)** making double-strand breaks in all crossover sites before reshuffling the fragmented DNA ends and rejoining the strands.

simplicity as SSR is rigorously restricted to a particular DNA sequence consisting of a site of 30–40 amino acids (Gaj et al., 2013; Carroll, 2014). *In vitro* and *in vivo*, site-specific recombination could be exceedingly quick and effectual under optimum conditions (Nash et al., 1996). The SSR encourages full recombination by breaking and re-joining of all 4 DNA filaments at the recombination sites. There are no additional cofactors required (Olorunniji et al., 2016).

The first application of Cre to catalyze the exclusion of selectable marker genes from transgenic tobacco (Dale and Ow, 1991; Albert et al., 1995) happened in the early 90s followed by several other reports such as in rice (Hoa et al., 2002; Radhakrishnan and Srivastava, 2005; Hu et al., 2008), wheat (Srivastava et al., 1999), tomato (Stuurman et al., 1996), barley (Kapusi et al., 2012), soybean (Li et al., 2009), Arabidopsis (Vergunst and Hooykaas, 1998), maize (Zhang et al., 2003; Kerbach et al., 2005; Anand et al., 2019) and in Chickpea-Rhizobium Rcd301 utilizing site-specific homologous recombination, the hup gene fragment from cosmid pHU52 was incorporated into the genome followed by addition of two fragments of the strain Rcd301's own genomic DNA to flank the cloned hup genes for successful integration (Vijaya Bhanu et al., 1994).

3.2 GE through oligonucleotide directed mutagenesis (ODM)

ODM, which dates back to the early 1980s, is a gene editing tool that is a base pair specific, precise and non-transgenic that has been greatly advanced to create unique and commercially relevant features in agriculturally important crops and can also be employed in chickpea. ODM, after its successful application in mammalian systems, has set off as an alternative novel gene edition method for plants (Abdurakhmonov, 2016; Sauer et al., 2016). ODM is a technique for targeted mutagenesis that employs a 20–100 base oligo nucleotide whose sequence is alike to the target sequence in the genome excluding a unit base pair change to achieve site-specified editing of the sequence of interest (Rádi et al., 2021). When these short oligonucleotide sequences are temporarily exposed to cultured plant cells, the repair template matches and binds to the homologous plant DNA sequence. The cell's inherent repair mechanism recognizes the single base mismatch between its DNA and the repair template once it has been attached. The cell will restore its DNA sequence by replicating the discrepancy in its DNA sequence. As a result, the oligo nucleotide is destroyed by the cell, and the required particular alteration in the plant's DNA is created. Plants with the precise mutation are then regenerated using tissue culture techniques, and standard breeding techniques are used to efficiently breed the desired features into elite plant varieties while removing undesired characteristics.

ODM has been greatly advanced using Rapid Trait Development System (RTDS). The RTDS™ machinery deals with a quick, explicit, and non-transgenic breeding substitute for traits enhancement to create unique commercially relevant features in agriculturally essential crops (Gocal et al., 2015). The RTDS method uses the cell's regular DNA repair system to alter particularly targeted bases in the genome for utilization of chemically generated oligo nucleotides. These oligo nucleotides serve as restoration templates causing DNA mismatches at the target location.

3.2.1 Applications of ODM technology

The ODM approach has been used successfully in a variety of plant crops, including herbicide tolerance (Zhu et al., 1999; Okuzaki and Toriyama, 2004; Dong et al., 2006; Rádi et al., 2021). Single point mutations are one of the ways of ODM applications in plants to transform endogenous loci(s) by targeting Aceto-Hydroxy Acid Synthase (AHAS) gene. Herbicides that block this enzyme such as imidazolinones (Imis), chlorsulfuron (CS), pyrimidinyl thiobenzoates, sulfonylureas (SUs), and bispyribacsodium (BS) make mutant enzymes easily selective (Tan et al., 2005). HRAC Group B and Australian Group B herbicides are classified as Group 2 herbicides in the Canadian herbicide classification system. One of three amino acid sites P197, S653, and W574 were targeted based on numbering on the Arabidopsis AHAS protein sequence to accomplish struggles to the afore mentioned herbicide chemistries. The study defining the fruitful applications of ODM was first conducted in the tobacco Nt-1 cell suspensions (Beetham et al., 1999; Ruiter et al., 2003), henceforth on maize (Zhu et al., 1999; Zhu et al., 2000). Other crops such as rice (Okuzaki and Toriyama, 2004), rapeseed (Ruiter et al., 2003; Gocal et al., 2015), including Arabidopsis (Kochevenko and Willmitzer, 2003) were also studied and tested. The transformation rates are liable to the crop, its cellular biology, the type of oligonucleotide and its concentration, the strand being directed, and the specific mutation taking place, which makes it difficult to compare different oligonucleotide delivery systems. In many aspects, the application of a fluorescence conversion approach, in which a BFP that is a Blue Fluorescent Protein could be converted into green fluorescent protein (GFP) just by editing a unit nucleotide of the Blue Fluorescent Protein gene, has improved oligo nucleotide mediated conversions. For example, oligo nucleotide length optimization and end protective chemistries have shown the potentials in boosting conversion rates (Sauer et al., 2016).

The protoplasts, generated through a BFP transgenic strain, were evaluated for the BFP to GFP gene edit for demonstrating the efficiency of oligo nucleotide mediated conversions in Arabidopsis. The findings show that oligo nucleotide mediated conversions have an excellent way to induce precise alterations in Arabidopsis. Moreover, these oligo nucleotide optimizations can have a big impact on the frequency of targeted modifications (Sauer et al., 2016). Furthermore, ODM has the potentials to improve crops without introducing additional genetic material by utilizing the plant's genome to boost abiotic (heat, drought, salinity) and biotic disease resistance (insect, bacterial, and virus), nutritional value, as well as its yield. ODM is presented as one of the numerous innovative breeding approaches that have set about the commercialization of food plants due to its capacity to accurately change sequences in genomes. Some commercial crops have been exploited *via* ODM such as maize, wheat, rice and rapeseed for herbicide tolerance as mentioned above. In 2016, A US based company Cibus put forward a herbicide tolerant rapeseed in several EU countries as a test case by using ODM in Rapid Trait Development System (RTDS) (Fladung, 2016). So far, no work has been reported in chickpea using ODM. Nevertheless, it is equally applicable in chickpea as well and may be expected to be done in near future.

3.3 GE mediated through site specific nucleases (SSNs)

Sequence-specific nuclease-based mutagenesis was first employed in plant research 15 years ago in 2006 (Razzaq et al., 2019) where engineered nucleases (ENs) were primarily used (Bruce Wallace et al., 1981) and engineered nucleases are divided into four categories: Zinc-Finger Nucleases (ZFNs), Transcription Activator Like Effector Nucleases (TALENs), Mega-nucleases and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Systems. SSNs work by building endonucleases that can cleave DNA onto a specific sequence in the genome. SSNs can have DNA or RNA binding pockets that attach to specific target sequences (Gaj et al., 2013; Carroll, 2014). These evolving technologies are progressing at breakneck speed, particularly in the realm of CRISPR-based genome editing (Abdallah et al., 2015; Kamburova et al., 2017) and are equally applicable in food legumes including chickpea.

3.3.1 GE mediated through zinc finger nucleases (ZFNs)

Currently, scientists have access to several techniques that can assist them to tackle difficulties related to precise genome editing in plants. Kim et al. (1996) discovered for the first time that protein domains like "zinc fingers" combine with FokI endonuclease domains, which act as site-responsive ZFNs and cleave DNA *in vitro* in well-defined locations (Miller et al., 2007). The chimeric protein has a modular structure because each one "zinc finger" domain recognizes nucleotides in the form of a triplet. This approach was used to alter cultured cells including both model and non-model plants (Cai et al., 2014). These were the first class of proteins to target a specific region of DNA and make double-stranded breaks. For their action *Flavobacterium okeanokoites I* (Fok1) nuclease enzymes assist them (Khandagale and Nadaf, 2016). The Cys2His2 type Zinc fingers are considered as most common eukaryotic transcription factors, whereas zinc finger nucleases are engineered restriction enzymes. It comprises 30 amino acids present in ββα fold and the inking of zinc provides more stability to the structure (Chen et al., 2014). The crystalline form of Zinc finger protein showed that it binds to major grooves of target DNA (Aslam et al., 2019). Structurally, its monomer consists of two important domains, namely, the DNA binding domain and DNA cleavage domain or nuclease domain. Out of an array of 4-6, zinc finger domains each of them recognizes 3bp of DNA sequence as shown in Figure 4. Using the phage display method wide range of ZFNs domains recognizing specific DNA triplets are identified. Knowing distinct domain recognized by ZFNs allow us to fuse them in tandem *via* linker peptide to form polydactyly zinc finger proteins that can target a wide range of DNA sequences (Gaj et al., 2016). Recent studies have tried to include more fingers to recognize longer and cleave rare targets (Urnov et al., 2010). The specificity of adherence to DNA is influenced by interaction with adjacent domains too (Petolino, 2015). For high specificity two ZFN monomers are required as the FokI nuclease domain act in dimerized form. Furthermore, the amino acids positioned at first, second, third, and +6 at the starting of the zinc finger alpha helix, contribute to peculiar binding to sites (Osakabe and Osakabe, 2015).

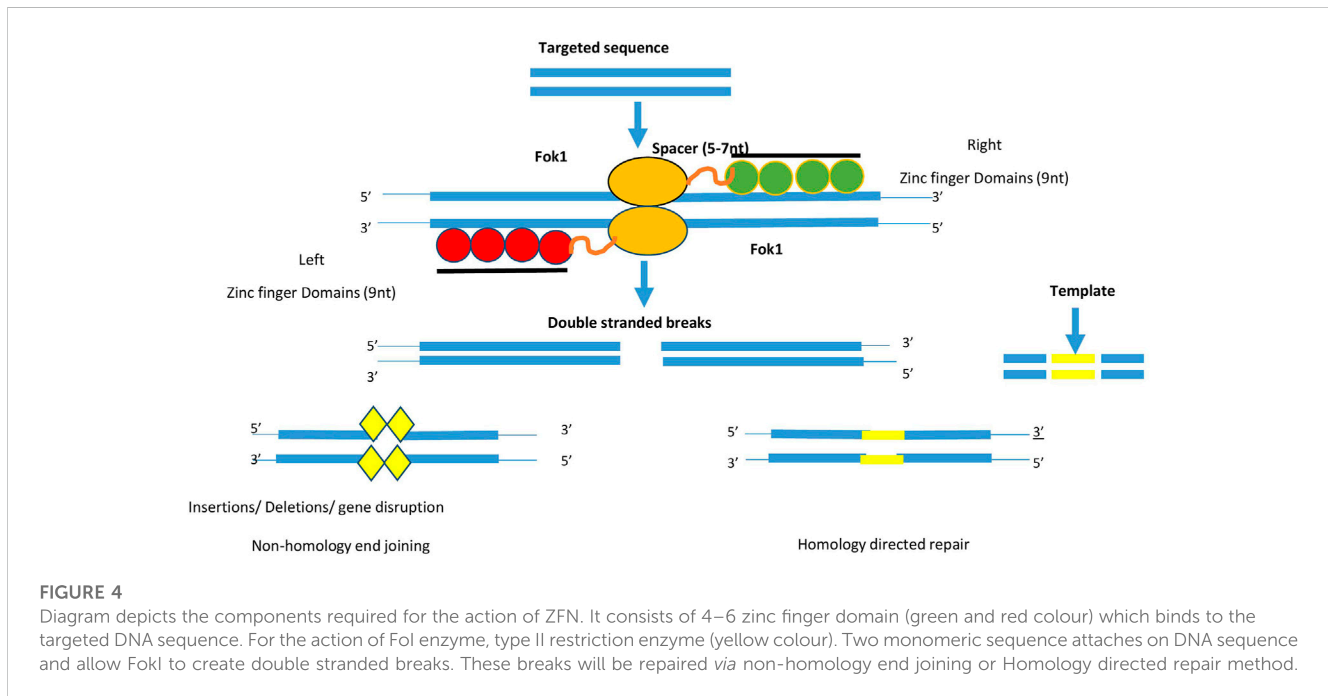


FIGURE 4
Diagram depicts the components required for the action of ZFN. It consists of 4–6 zinc finger domain (green and red colour) which binds to the targeted DNA sequence. For the action of FoI enzyme, type II restriction enzyme (yellow colour). Two monomeric sequence attaches on DNA sequence and allow FokI to create double stranded breaks. These breaks will be repaired *via* non-homology end joining or Homology directed repair method.

To reduce off-site cleavage, FoKI variants have been developed which require heterodimerization between two monomers of ZFN (Ran et al., 2018). Engineering methods are widely used for the construction of engineered ZFNs, identification of triplet sequences, modular assembly, and oligomerized pool. The drawback of this approach is that ZFNs can bind to neighbouring fingers as well as to bases present outside of the proximity of the targeted DNA triplet (Urnov et al., 2010). GE through ZFNs yields modification with efficiencies of more than 10% by creating double-stranded breaks (Miller et al., 2007). The efficiency of mutagenesis was reported in Arabidopsis and it was found to be 78% in case of simple deletions, 13% in simple insertions, and approximately 8% in deletions with long insertions (Lloyd et al., 2005). In another study, the constitutive expression of ZFN resulted in a 2% mutation and deletion of sequence ranging from 1 to 80 bp (de Pater et al., 2009).

3.3.1.1 Application of ZFN technology

Despite of challenges faced during the construction of ZFNs, they have been widely used to modify genes of cultivated crops *Arabidopsis*, tobacco, maize, soybean, and canola (Mushtaq et al., 2019). In maize disruption of endogenous inositol phosphatase kinase 1 gene by the introduction of PAT gene cassettes lead to the development of herbicide-tolerant cultivars and simultaneously alteration in inositol phosphate of developing seeds (Zhang et al., 2018). In another approach, ABA INSENSITIVE-4 (ABI4) gene was mutagenized in *Arabidopsis*, and the frequency of insertion and deletion was a maximum of 3% in nine transgenic lines. However, when estrogen-inducible ZFNs were used to create mutations in *Arabidopsis,* in the first generation the rate of mutations was 7% and 16% in the two genes, namely, alcohol dehydrogenase1 and transparent test4 (Zhang F. et al., 2010). In the oil seed family, ZFN was performed in soybean and brassica to improve agronomic traits. A similar approach was made to create mutations in dicer-like (DCL) genes in soybean to develop the Zinger finger consortium by context-dependent assembly (Curtin et al., 2013). In *Brassica napus*, the method has been used for activation of β-ketoac- ACP synthase II, resulting in a decrease in the production of palmitic acid and entire saturated fatty acid content (Gupta et al., 2012). Recently, purified ZFN monomer proteins were isolated from bacterial cultures and delivered into unmodified microspores to edit the inositol pentakiphosphatase kinase1 gene, which is found to be involved in catalysing the end step of phytic acid production (Bilichak et al., 2020). In the populous, the heat-inducible ZFN system mutagenizes floral genes at a rate of 0.3% (Lu et al., 2016). In tobacco, mutations were targeted in SuRA and SuRB conferring herbicidal resistance to imidazolinone and sulfonylurea compounds (Maeder et al., 2008; Townsend et al., 2009). ZFN approach can be used to facilitate multiple knockouts of the gene as seen in wheat, three homologous copies of the acetohydroxy acid synthase gene were targeted simultaneously (Ran et al., 2018). Against biotic stress, plants develop resistance against the pathogen, and ZFNs were artificially designed to bind against the circular single-stranded DNA of begomovirus (Chen et al., 2014). Earlier, an artificial zinc finger protein (AFP) without a nuclease domain was designed to block the transcription of viral replication protein of beet severe curly top virus, 80% of transgenic *Arabidopsis* showed no symptoms against BSCTV. Similarly, the Rep gene of tomato yellow leaf curl China virus and tobacco curly shoot Yunnan virus were targeted to increase the resistance against these viruses (Yin and Qiu, 2019). Peer et al. (2015) reported the use of ZFN for the induction of targeted mutagenesis in perennial fruits including apples and fig.

The creation of lines of chickpeas with only two transgenes has been described so far (Mehrotra et al., 2011). As a consequence of the limited cloning sites inside the cassettes expressing the gene, the binary vectors employed for this transformation process have

limited contribution to the transfer of more than 1-2 genes. As a result, binary vectors must be improved to transfer multiple genes in chickpeas. Dual-gene binary vectors have been created using zinc finger nucleases, which can bind and cleave lengthy DNA sequences (Zeevi et al., 2012). In chickpeas, similar procedures can be used to create a binary vector for many transgenes insertion.

Despite successful examples, various challenges are certain limitations viz; the need for DNA/protein interaction, redesigning of protein for a different DNA sequence every time is a difficult task, costly and time taking approach (Piatek et al., 2018).

3.3.2 GE mediated through homing (mega) endonucleases

Site-specific restriction endonucleases can be employed to make site-directed double-strand breaks (DSBs) in the genome. Mega nucleases also known as homing endonucleases are unique enzymes with high activity and long recognition sequences (>14 bp) that digest target DNA in a site-specific manner (Epinat et al., 2003; Smith et al., 2006). Epinat et al. (2003) described the manufacture of hybrid enzymes utilizing two mega nucleases that identify new target sequences, I-Cre I and I-Dmo I. Novel mega nuclease variants that detect unique sequences with enhanced nuclease activities have also been created using specialized mutagenesis and high-throughput screening approaches (Smith et al., 2006; Arnould et al., 2007; Grizot et al., 2009).

In comparison to other SSN systems, mega nuclease has the disadvantage of being more expensive and time-demanding to develop sequence-specific enzymes for all conceivable sequences. As a result, each new genome-engineering target necessitates a first round of protein engineering to create a bespoke mega nuclease. As a result, working with mega nucleases has been difficult, and patent battles have hampered the progress (Smith et al., 2012).

3.3.3 GE mediated through transcription activator like effector nucleases (TALENs)

The area of GE is rapidly expanding as new approaches and technologies emerge. GE will be required to enhance crop production since the global population is expected to reach 9.6 billion by 2050 (IPCC, 2019), while arable land shrinks. In 2009, TALEN effectors for DNA targeting were revealed. The discovery of distinctive transcription activator like effector (TALE) protein in 2011 that recognizes and activates certain plant developed through a sequence of tandem repeats led to the development of a new GE method based on chimeric nucleases dubbed TALENs (Jankele and Svoboda, 2014). TALENs are easier to construct and more widely used than ZFNs. Non-etheless, repeating sequences in the TALEN composition can enhance the probability of homologous recombination. ZFNs and TALENs are structurally and functionally identical because both of them contain the restriction endonuclease FokI.

TALE protein's DNA binding central repeat domain is composed of a few to 33.5 repeats, each of which is made up of 34 amino acids that triggers the transcription of the target gene. Structurally, it is composed of a monomer, which binds at one specific region in the target nucleotide sequence. Monomers are found positioned at 12 and 13 repeats of 34 amino acids and are extremely variable (that are repeat variable di-residue, RVD), and are responsible for the identification of a specific nucleotide. This code degenerates and some RVDs bind to multiple nucleotides with vastly differing efficiency degrees. The targeted DNA molecule always contains the same nucleotide, that is the thymidine, before the 5′- end of a sequence, which is bound by a TALE monomer and affects the binding efficacy. The rear most tandem repeat that clips to the nucleotide at the 3′- end of the recognition site contains approximately 20 amino acid residues and is known as a half repeat (Nemudryi et al., 2014).

TALEs show high specificity towards sequence in the presence of magnesium and calcium divalent cations. However, when potassium and sodium monovalent ions are present, the TALEs are strapped to a specific as well as the non-specific region of DNA with nearly equal affinity. In comparison to monovalent ions, divalent ions in turn bind to DNA which attenuates the non-specific reciprocity between TALEs and DNA which further leads to a balanced complex (Cuculis et al., 2020).

TALENs are developed by fusing the restriction endonuclease Fok-I, a nuclease entity to a TALE DNA binding domain. To carry out precise genome editing TALEN work in pairs, binding to the DNA sequence in an opposite orientation such that the FokI domain could dimerize and cut the DNA sequence present within the spacer in between the two different binding sites. Half of the targeted sites of TALEN are conscripted in a way that the pairs are presented in an opposing intention on contradictory sides of dsDNA with an optimal sequence that acts as a spacer between them (Figure 5). In yeast, the activities of TALENs were demonstrated by combining the N- or C-terminal of TALEs with the catalytic domain of the Fok-I protein, which leads to cleavage of DNA with efficiencies equivalent to ZFN. As for as, the activity of the TALEs C-terminal domain is concerned, it is not vital. Hence, shortening the C-terminal by amino acids at +17, +28, or +63 and then fusing to the Fok-1 catalytic domain is possible that increases the efficiency too. Fok-I-based TALEN also works similarly to ZFN. Based on the length of the C-terminal TALE domain optimal spacer length is selected (Miller et al., 2011).

When the DNA-binding domains of two identical FokI nucleases come into contact, they dimerize and cut the DNA target. When these halves are created using a homodimer Fok-I, they can interact in three different ways. The left halves or right halves can combine to form a functional nuclease just as easily as the calculated interlinkage between the left and right halves of a nuclease set, which increases the likelihood that a TALEN will bind to sites with properties resembling those of the targeted DNA. Correspondingly, TALEN molecules may be linked to various parts of the genome in various combinations. It becomes more likely that a cell will be overrun by DSBs, leading to cell death and collateral loss to the DNA of surviving cells. Several obligatory heterodimer variations of FokI have been created to lessen off target toxicity. The created versions are based on mutagenesis, DNA shuffling, and structure-guided design (Joung and Sander, 2013).

This approach was created to improve genome editing efficiency, safety, and accessibility (Boch and Bonas, 2010; Urnov et al., 2010; Miller et al., 2011). The proteins imparting the effects are members of the DNA binding protein family and, like transcription factors in eukaryotic genomes, can be utilized to induce the expression of the targeted heat tolerance genes. TAL effectors (TALEs) are produced naturally by the phytopathogen *Xanthomonas oryzae* (Xanthomonas), which penetrates and

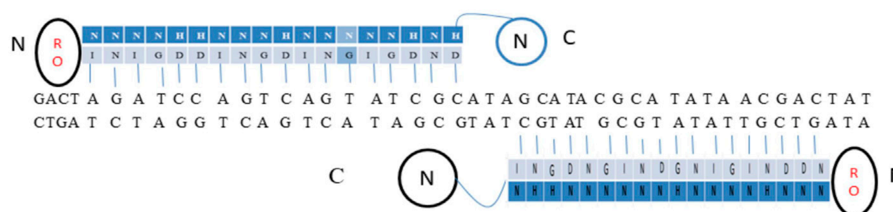


FIGURE 5
TALE activator along with a pair of TALENs.

reaches the nucleus of the cell and modify the transcription process to provide benefit to the pathogen (Cermak et al., 2011a). TALEs consist of a core where DNA-binding repeats are presented that regulate the binding specificity of DNA *via* an one-to-one repeated base pair binding relationship (Cermak et al., 2011a; Deng et al., 2012). TALEs can be generated to fuse any DNA sequence by modifying the number and kind of repeats (Li et al., 2013). *In vitro* and *in vivo*, fusing a TALE to nuclease results in an enzyme that is capable of creating site specific DSBs (Christian et al., 2010; Mahfouz et al., 2011; Deng et al., 2012). RVDs of the TALE repeat sequence enhance and stabilize the contact with the amino acid at the 13 positions, which give binding specificity, which are the structural foundations of TALE-DNA binding (Boch et al., 2009; Deng et al., 2012).

Because of their DNA-binding specificities, TALEs can be employed as DNA binding modules in the creation of synthetic transcriptional and epigenetic regulators. TALENs have catalysed much amusement and excitement among researchers as they can be designed easily and rapidly that ally modular DNA binding of TALE repeat domains to discrete bases in a target binding site. The primitive building blocks are used to design the domain of TALENs where DNA binds are highly conserved. Recently, co-crystal structures of TALE showed that DNA binding domains were bound to their coupled sites in the major groove of DNA (Joung and Sander, 2013).

For TALEs, several engineering platforms have been created. Furthermore, researchers examined the genetic makeup of bacteria besides, Xanthomonas and discovered that *Ralstonia solanacearum* has Ralstonia TALE-like proteins (that is RTLs) which have corresponding structure but distinct repeats with specificity as determined by numbers of RVD presence (Bogdanove et al., 2010; Remigi et al., 2011).

3.3.3.1 Application of TALEN technology

The TALEN mediated genome editing approach was applied for crop enhancement for the first time in rice by disrupting the bacterial blight susceptibility gene Os SWEET14 and producing a mutant rice to show resistance towards bacterial blight (Li et al., 2012). TALENs have also been utilized to knock out three TaMLO homeologs in wheat to develop powdery mildew resistant wheat (Wang et al., 2014). Char et al. (2015) generated mutants of maize with the glossy phenotype, reduced amount of epicuticular wax in the leaves, and the ability to be surface manured by eliminating the maize GL2 gene. TALEN mediated mutagenesis has increased the composition of the cell wall and saccharification effectiveness in sugarcane (Jung and Altpeter, 2016; Kannan et al., 2018). During cold storage, product quality declines majorly because of the accumulation of reducing sugars. As observed in potato tubers, knocking down the vacuolar invertase (VInv) gene resulted in tubers with undetectable amounts of harmful reducing sugars (Clasen et al., 2016). Integrating TALENs and donor DNA in Gemini virus replicons markedly escalate the copy number and homologous recombination efficiency *via* introducing a powerful promoter upstream of the gene regulating anthocyanin biosynthesis resulting in purple tomatoes with an increased amount of anthocyanin (Čermák et al., 2015). Recently, one of the mitochondrial *orf* genes, *orf 312* (CMS-associated gene), knocked out by this approach showed that it is responsible for pollen abortion and leads to cytoplasmic male sterility in rice (Takatsuka et al., 2022). These examples show how TALEN technology can be used to improve crops including chickpea heat tolerance and yield traits in a variety of ways. However, the production of TALE repeats remains a difficult path to follow and harness the efficacy of gene targeting.

3.3.4 GE mediated through Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)

CRISPR technology was introduced 2 years later, after the discovery of the TALEN proteins. CRISPR, which consists of non-coding RNAs and Cas proteins, was developed and has since become widely employed. Unlike first generation genome editing approaches, CRISPR/Cas9 is easy to design, clone, and the similar Cas9 protein theoretically can be used with a variety of guide RNAs to target several locations throughout the genome. The most commonly used genome editing tools are TALENs and CRISPR associated Cas9. Each represents a type of engineered nuclease that can be customized to recognize, bind, and cleave a specific sequence in the genome. TALENs are entirely protein-based, and CRISPR/Cas9 has both protein and RNA components (Musunuru, 2017). Unlike the chimeric TALEN proteins, the CRISPR/Cas9 system recognizes the DNA site which needs to be altered by a complementary interaction between a non-coding RNA and the targeted site. Hence, it leads to the formation of a complex consisting of non-coding RNA and Cas9 proteins having nuclease activity. The generalized mechanism of CRISPR technology is depicted below as Figure 6.

CRISPR associated Cas9 system, is the most prominent and innovative genome editing approach which has recently become popular. CRISPR/CAS-9 has been widely accepted due to its preciseness, high efficiency, and utility to ameliorate abiotic and

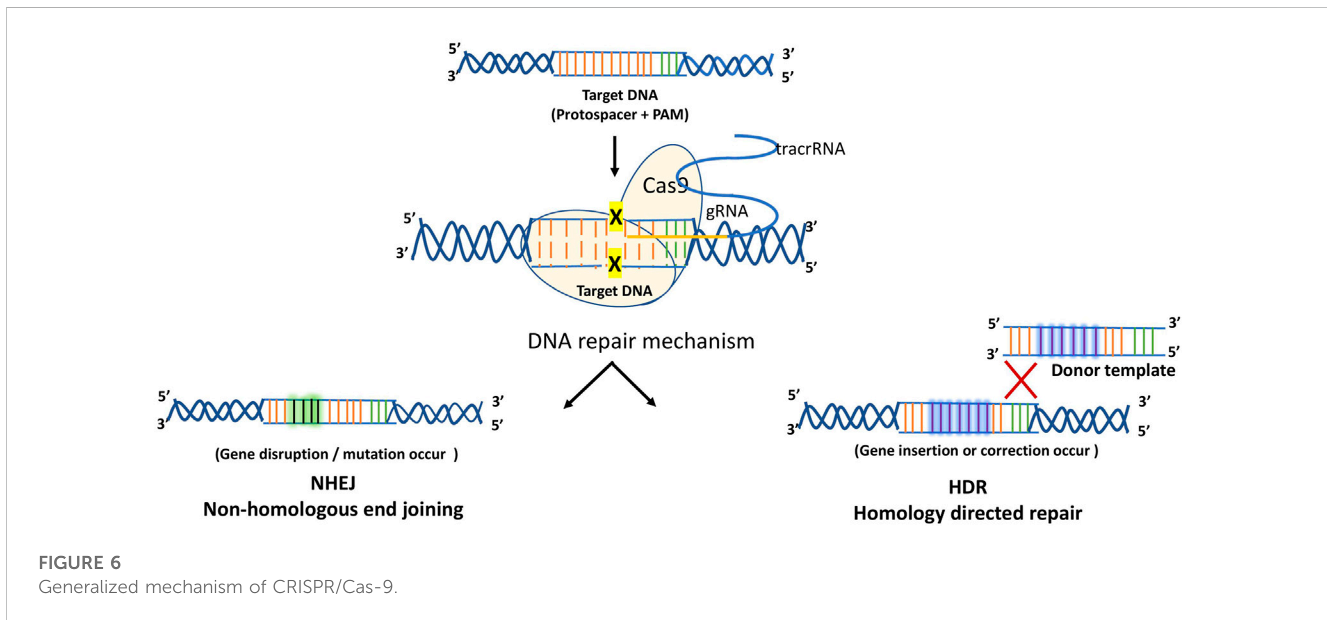


FIGURE 6
Generalized mechanism of CRISPR/Cas-9.

TABLE 3 Summary of CRISPR-Cas enzymes.

Class	Type	Subtype	Effector	Target	Nuclease domains	TracrRNA requirement	PAM/PFS
1 (Multi-Cas proteins)	I	A,B,C,D,E,F,U	Cascade	dsDNA	HD fused to Cas3	No	-
1 (Multi-Cas proteins)	III	A,B,C,D	Cascade	ssRNA	HD fused to Cas10	No	–
1 (Multi-Cas proteins)	III	A,B,C,D	Cascade	ssRNA	HD fused to Cas10	No	–
1 (Multi-Cas proteins)	IV	A, B	Cascade	dsDNA	unknown	No	–
2(Single-Cas protein)	II	A	SpCas9	dsDNA	RuvC, HNH	Yes	NGG
2 (Single-Cas protein)	II	A	SaCas9	dsDNA	RuvC, HNH	Yes	NNGRRT
2(Single-Cas protein)	II	B	FnCas9	dsDNA/ ssRNA	RuvC, HNH	Yes	NGG
2(Single-Cas protein)	II	C	NmCas9	dsDNA	RuvC, HNH	Yes	NNNNGATT
2(Single-Cas protein)	V	A	Cas12a (Cpf1)	dsDNA	RuvC, Nuc	No	5˚ AT-rich PAM
2(Single-Cas protein)	V	B	Cas12b (C2c1)	dsDNA	RuvC	Yes	5˚ AT-rich PAM
2(Single-Cas protein)	V	C	Cas12c (C2c3)	dsDNA	RuvC	Yes	5˚ AT-rich PAM
2(Single-Cas protein)	VI	A	Cas13a (C2c2)	ssRNA	2xHEPN	No	3˚ PFS: non-G
2(Single-Cas protein)	VI	B	Cas13b (C2c4)	ssRNA	2xHEPN	No	5˚ PFS: non-C; 3˚ PFS: NAN/NNA
2(Single-Cas protein)	VI	C	Cas13c (C2c7)	ssRNA	2xHEPN	No	–
2(Single-Cas protein)	VI	D	Cas13d	ssRNA	2xHEPN	No	–

biotic stress tolerance in plants as detailed mentioned in Table 2. CRISPR is palindromic repeat sequences found in the bacterial genome separated by a spacer of 32–36 base pairs. There are several CRISPR/Cas9 systems but primarily classified into three types; type I, II, and III. For plant genome editing, CRISPR/Ca9 type II is frequently used. It is an adaptation of the Gram-positive *Streptococcus pyogenes* system (Le Rhun et al., 2019). Presently, it has been believed to be an efficient and precise *in vitro* as well as *in vivo* genome editing tool and many tailored Cas9 complexes have been utilized to increase the frequency of selectivity of target and reduce the chances of off target cleavage after proof-of-concept demonstrations by core CRISPR/Cas9 module (viz- Nmcas9, Sacas9, and Stcas9) in plants. Additionally, utilization of Cas9 enzymes from different bacterial strains have increased the specificity and efficacy of gene editing procedures as presented in Table 3 (Jaganathan et al., 2018).

3.3.4.1 Applications of CRISPR/Cas9 systems

CRISPR can make deliberate changes in genome structures hence it has a tremendous impact on bioengineering and molecular biology. The technology was used to improve the colour, shelf life, and commercial attractiveness of fruits and vegetables by reducing the amount of toxic steroidal glycoalkaloids. A boost in amylose, starch, aroma, good fats like oleic acid, etc., and a decrease in gluten proteins and unsaturated fatty acid content and so on were among the other modifications (Jiang et al., 2017; Sun et al., 2017). Thus, in crop plants, the CRISPR/cas9 technique can be exploited to improve the yield and quality by increasing the shell life, amending colour, size, texture, etc. (Xing et al., 2020).

To develop biotic resistant crops an attempt was made, where initiation factor elF4E of cucumber was inactivated using CRISPR/Cas9 system, resulting in plants found to be resistant towards cucumber vein yellowing virus. Similarly, grape knockout of VvWRKY52 increased tolerance against fungal infection. In another experiment conducted on rice, CRISPR/Cas9 knocked out the LAZY1 gene resulting in a tiller-spreading phenotype that may boost yield in a certain environment (Miao et al., 2013; Li et al., 2016). In another study, three different genes including Grain Number 1a (Gn1a), dense and erect panicle (dep1), and grain size (GS3) of the rice cultivar Zhonghua 11 were mutated by the CRISPR/Cas9 system those showed a greater number of grains with an increase in size and dense erect panicles. Recently, the role of *Oryza sativa* senescence associated protein during drought has been explored by editing drought induced genes (Park et al., 2022).

Chickpea production is hampered by drought, low and high temperatures, and other abiotic conditions (Gaur et al., 2008; Mantri et al., 2010; Jha et al., 2014). Recently, two potential genes, 4 coumarate ligase (4CL) playing important role in phenylpropanoid metabolism, and Reveille 7 (RVE7) involved in circadian rhythm were chosen for CRISPR/Cas9 editing in chickpea protoplast, both of which are linked to drought tolerance. The 4CL enzyme is engaged in the phenylpropanoid metabolism pathway during the production of lignin. To knock off these targeted genes in chickpeas, researchers used DNA free CRISPR/Cas9 editing tool. In chickpeas, protoplast editing is a revolutionary technique for accomplishing targeted mutagenesis. In comparison to the 4CL gene, the RVE7 gene showed excellent *in vivo* editing effectiveness. According to Ninan et al. (2019), in the leaves of chickpeas, cytokines have increased sink activities. Isopentenyl transferase controls the earliest steps in the synthesis of cytokines (IPT). The cytokinin dehydrogenase or oxidase is now in charge of controlling cytokinin breakdown. Root-specific promoter CaWRKY31 of chickpeas could be used to explore the mechanism behind how cytokinin diminution impacts the development of root architecture and tolerance towards drought. In Arabidopsis and chickpeas to study definite and indeterminate growth patterns, a root specific promoter CaWRKY31 can be used. In the model plant Arabidopsis and chickpea, it is observed that root-specific CaCKX6 expression increased the proliferation of lateral roots plant biomass without impairing the vegetative and reproductive development. Root cytokinin oxidase/dehydrogenase (CKX) gene activity was seen to be increased in transgenic chickpea strains. CKX gene functional characterization studies in chickpeas have only recently begun. Gene editing tools such as TALENs and CRISPR/Cas9 approach can be quite useful in this situation (Mahto et al., 2022). Gene editing technologies can help with knock-ins in addition to knockouts.

Heat, drought, floods, temperature extremes, salt, heavy metals, radiation, and other factors can contribute to abiotic stress. Stress has a significant impact on the yield of crops. Several crops have been mutated to defend against abiotic (Shan et al., 2013; Klap et al., 2017). To boost drought tolerance in maize, researchers employed CRISPR/Cas9 to introduce a promoter at a specific region (Shi et al., 2017). Site specific genomic change has previously been accomplished using gene editing tools like zinc finger nuclease and transcription activator like effector nucleases, but these tools have limitations (Gupta and Musunuru, 2014).

Biotic stress, on the other hand, is caused by microbes like fungi, bacteria, and viruses. Several crops have been mutated to defend against biotic stresses (Lu et al., 2018). Hybrid breeding, which includes improvements in hybrid wheat seed production, is another approach to increasing crop output. Hybrid crops are effective high yielding cultivars today, yet hybrid seed production requires emasculation to avoid self-pollination.

These gene editing technologies like TALENs or CRISPR/Cas9 can be quite useful in the creation of non-genetically modified crops that have the desired trait, boosting yield potential under biotic and abiotic stress situations (Mahto RK et al., 2022).

However, a major drawback of CRISPR technology compared to other genome editing tools is the high frequency of off target mutations even to the extent of up to 50%. (Zhang et al., 2015). The most difficult problem so far has been getting the CRISPR system into the target cells. Each crop including chickpea that uses CRISPR/Cas9 has intrinsic restrictions. At first, it is impossible to determine potential editing targets of interest or evaluate gRNAs off target behaviour without access to or incomplete assembly of a genome sequence (Hahn and Nekrasov, 2019). There is a need for additional research in this field due to technical challenges in creating viable transgenic chickpeas and the lack of a stable transient system of expression for quick study of gene expression and function (Badhan et al., 2021). The generalized limitations and benefits of CRISPR technology is depicted below as Figure 7.

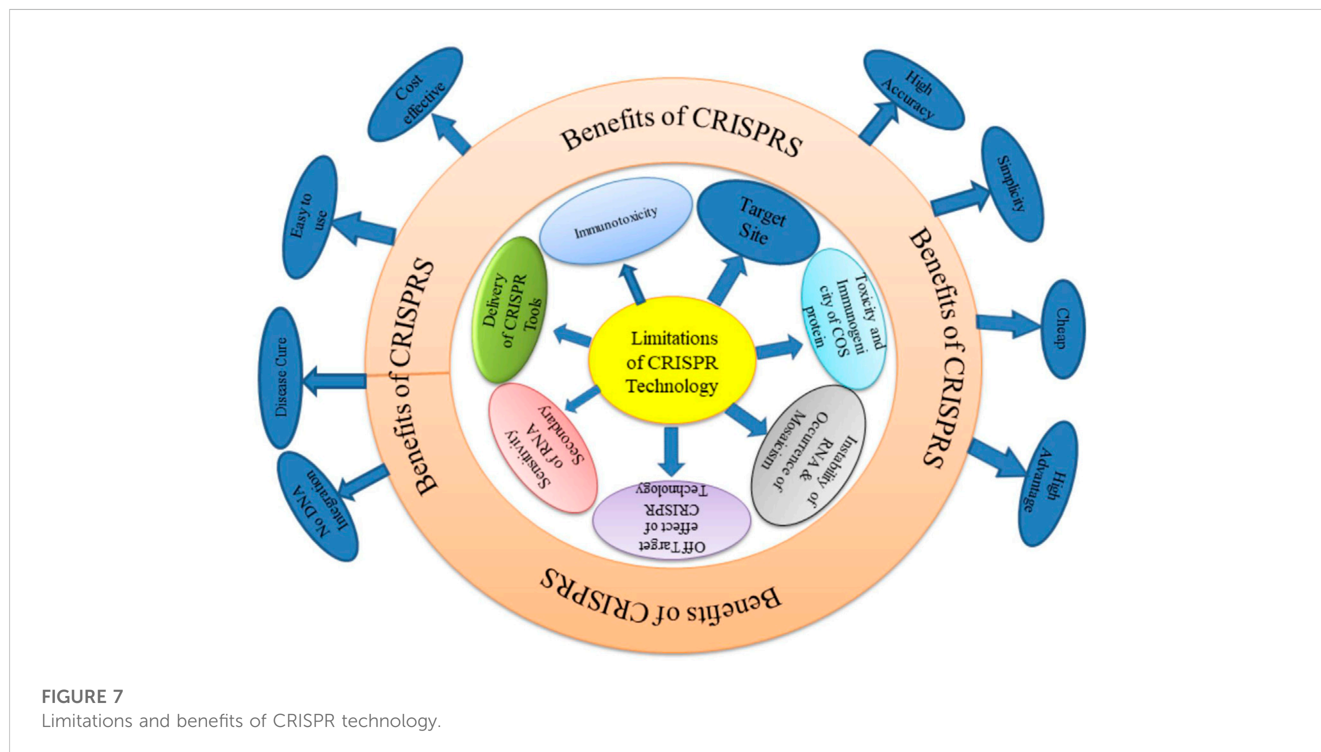


FIGURE 7
Limitations and benefits of CRISPR technology.

3.4 GE mediated through base and prime advanced approaches

Over the past few years, numerous prime editing (PE) and base editing (BE) variants have been created and experimentally validated in plants (Molla et al., 2021). These are two recently established genome engineering techniques that can rapidly insert specific modifications into target regions without the use of donor DNA templates or DSB creations. Applications like controlling cis-elements, altering RNA splice sites, including synthetic miRNAs, or customizing miRNA binding sites are made possible using PE and BE technologies. The binding locations of effectors produced by fungal infections to target plant susceptibility genes may also be altered by these methods and heritable resistance may be passed down in this manner (Van Vu et al., 2022). Both base editing and prime editing have been tested on a variety of plant types and proven to be effective.

3.4.1 Base editing

BE is a game changing method for precisely implanting point mutations at the appropriate places without the use of donor DNA templates or the production of double strand breaks (Rees and Liu, 2018). First cytosine base editor (CBE) was produced using a SpCas9 (D10A) nickase in combination with a cytidine deaminase and an uracil glycosylase inhibitor (UGI) to make the transition from C G to T A (Komor et al., 2016). Following that cytidine deaminase will deaminate the exposed non-target DNA strand changing cytosine (C) to uracil (U) resulting in a C to T base change during DNA repair and replication. Structurally, the adenine base editor (ABE) is analogous to the CBE, and using *E. coli* transfer RNA adenosine deaminase (ecTadA), it converts adenine A) to inosine (I) in the non-target strand (Gaudelli et al., 2017). Moreover, in a variety of plant species, CBEs and ABEs have been employed to research the function of genes undiscovered and improve crop qualities (Molla and Yang, 2019; Mishra et al., 2020). To handle CT and AG conversion in a genomic area of interest, an interesting approach was applied, and a dual base kind of editor was constructed by fusing cytidine and adenosine deaminases into Cas protein. Discretely, CBE, ABE as well as dual base editors, have a similar mode of action: deamination of C and A by cytidine and adenosine deaminase, respectively (Abdallah et al., 2021). The generalized mechanism of base editing technology is depicted below as Figure 8.

Although CRISPR based precision genome editing technologies have evolved and flourished fast, these tools have been unable to reach organelle genomes because of the non-availability of guide RNA as well as Cas proteins inside organelles. Hence, it is important and needs to explore the possible ways to approach organelle specific gene editing of monocots and dicots to decipher the function of the gene and limit the off targets' chance. However, very recently, organellar genome engineering has been described (Mok et al., 2020) and the group has discovered the deaminase domain of the bacterial toxin DddA which is structurally similar to that of APOBEC enzymes and deaminates the cytosines in double stranded DNA (dsDNA). DddAtox is being integrated with organelle focused transcription activator like effector (TALE) repeat arrays, which directly deaminates dsDNA in organellar genomes. Despite the efficiency of DdCBEs in a variety of species of the plants, various issues such as DddAtox deaminase sequence preference and likely editing of off target sites must be directed before precise organellar genome editing in plants can be carried out (Azameti and Dauda, 2021).

3.4.2 Prime editing

PE is a non DSB genome editing method that results in all feasible base conversions, tiny indels, and combinations of them at

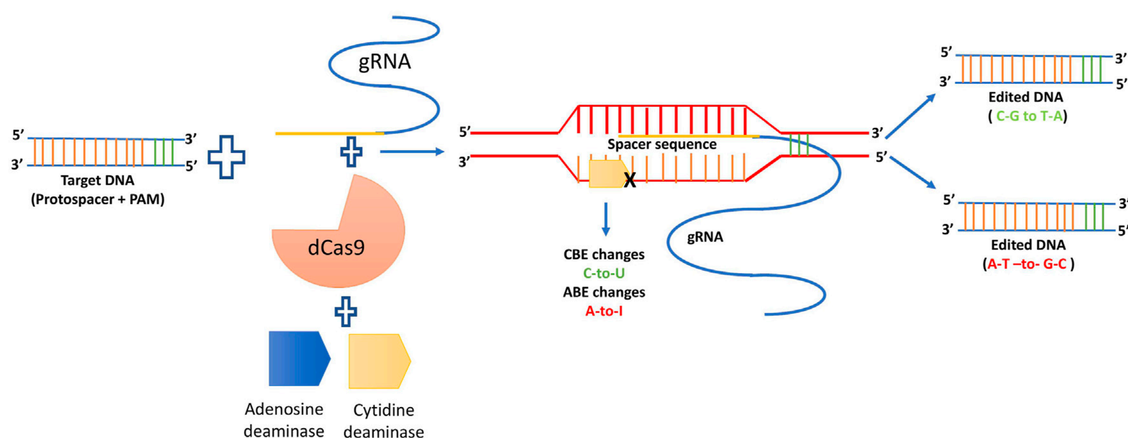


FIGURE 8
Generalized mechanism of base editing.

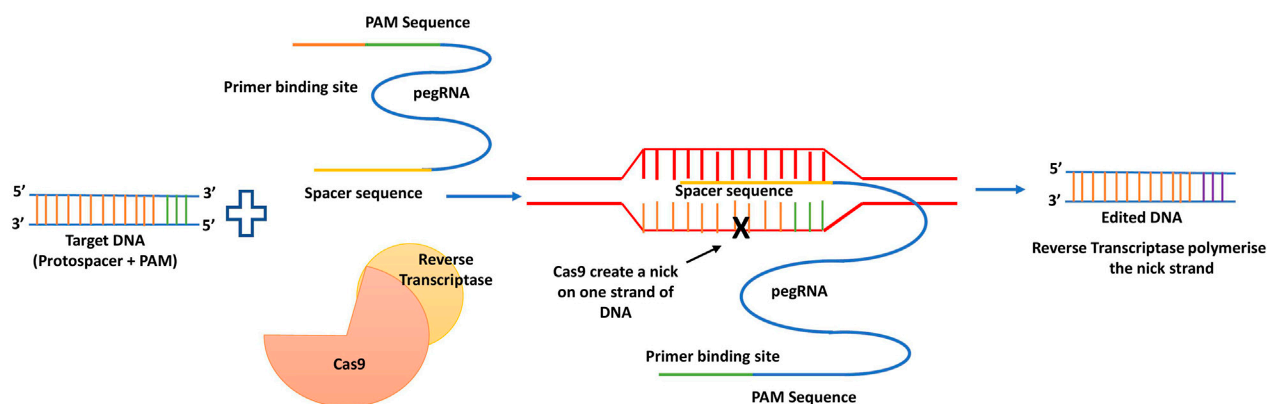


FIGURE 9
Mechanism of prime editing.

selected regions (Anzalone et al., 2019). The target site is specified using guide RNA with a 5′spacer sequence. The Cas9 nickase reverse transcriptase and fusion proteins are the prerequisites. The prime editing guide RNA called pegRNA, which guides the fusion of proteins to identify the target site before causing a nick on the non-target strand, after which it anneals with primer binding site (PBS) and finally primes the reverse transcriptase of the reverse transcriptase template, which then copies the right sequence into the target after a lengthy DNA repair mechanism (Anzalone et al., 2019). The generalized mechanism of prime editing technology is depicted below as Figure 9.

The PE method has been used with a variety of plants (Xu R. et al., 2020; Butt et al., 2020; Xu W. et al., 2020; Hua et al., 2020; Jiang et al., 2020; Li et al., 2020). In comparison to mammalian cells editing frequencies are lower in monocot plants and in dicot species not at all (Lu et al., 2021; Wang et al., 2021). PE events have been observed in stable transgenic lines of two important crops *Oryza sativa* and *Solanum Lycopersicon*, however, the ratio of homozygous in comparison to biallelic edits is significantly low (Xu R. et al., 2020; Xu W. et al., 2020; Hua et al., 2020; Li et al., 2020; Lin et al., 2020; Lu et al., 2021), indicating PE's inefficiency in plants (Hua et al., 2020). Further, Biswas et al. (2022) have shown a low range of prime editing efficiency in legumes, ranging from 0.2% to 0.5% of protoplast cells showing the targeted edits, a higher editing efficiency is expected once transgenic plants are developed. However, further optimization of the prime editing system should improve editing efficiency in legumes including chickpea.

3.4.2.1 Application of base editors and prime editors

Research articles related to DSB independent genome editing tools, base editing, and prime editing considered them to be more predictive than DSB dependent genome editing tools, which have various advantages including knowing about the function of gene and precision crop breeding (Komor et al., 2016; Gaudelli et al., 2017; Anzalone et al., 2019). Bes, PEs can interrupt genes by incorporating stop codons, alternately inactivating, splicing sites,

which are highly conserved in coding regions of genes for thwarting undesired mutations in the genome, synthesis of aberrant proteins, (Billon et al., 2017; Kluesner et al., 2021; Ren et al., 2021). In addition, BEs and PEs can precisely alter possible gene regulatory regions including sites where miRNA or transcription factors bind or modifies post transitional regions and can act on the open reading frame to infer their activities (Xing et al., 2020; Ren et al., 2021).

4 Conclusion and future perspective

The crop genome engineering inclusive of genomics and genome editing tools have already been successfully employed in several crops, although it is still in its early phase for production enhancement and abiotic stresses including heat tolerance, drought, salinity, etc in chickpea. Various genomic approaches viz; multi-omics, transcriptomics, proteomics, metabolomics, pan and genome editing technologies have tremendous potentials to influence the plant breeding techniques to guard crop plants against numerous abiotic/biotic stresses and augment crop yield. Editing the target DNA sequence by adding, deleting, or substituting nucleotide bases are cutting edge molecular biology techniques and Genome amending procedures viz; SSRs, ODMs, SSNs inclusive of ZFNs, TALENs, Mega nucleases, CRISPR/Cas9 and advanced approaches viz; Base Editors, Primer Editors are used. The CRISPR/Cas9 technologies corroborate the utmost operational GE machinery since these are precise, less expensive, speedy, and consent for numerous site-specific genome editing. SSNs have been utilized to elucidate the activities of many essential genes in plants that could be exploited to boost agricultural yield and often SSN induced NHEJ were used in polyploidy plants to investigate gene function and trait development which resulted in gene deletions. Recently, scientists are focusing on fabricating plant genomes to make them withstand climatic changes. In defiance of its success in the laboratory, gene editing technology for climate change has yet to demonstrate a significant impact in the real world as regulations, societal hurdles, and proscriptive policies, among other externalities outside the technical limits stated have hampered the adoption of these technological advancements. However, current technical advances are rapidly expanding and thanks go to the continued efforts of both public and commercial organizations. Genetic engineering approaches as mentioned above that alter minimal DNA/chromatin configurations, but exact modifications in the genome or precise insertion of small DNA fragments are attractive possibilities for worldwide regulatory overhaul, policy improvements, and increased consumer acceptance. Naturally, the advantages of gene editing applications will only be recognized once farmers and producers have access to these revolutionary technologies. Despite technological restrictions, socio-political barriers must overcome and gene-modified products should be widely adopted. Thus, CRISPR gene editing tool is an essential forward step for agricultural adaptability in the face of negative climate impact and holds the great possibilities for harnessing the betterment of future agriculture including chickpea enhanced capabilities for cytokinin dehydrogenase, nitrate reductase, superoxide dismutase to induce drought resistance, heat tolerance and higher yield higher yield to encounter global climate change, hunger and nutritional threats.

Author contributions

RK and AH conceptualized, interpreted and refined the manuscript. CS, HS, RKr and SB executed literature collections and drafting of the manuscript. TS, G, NMS, RY and SG contributed to the improvement of manuscript. NA helped in interpretation and refinement of the manuscript. All authors meticulously perused and approved the manuscript.

Acknowledgments

The all authors gratefully acknowledge professional editing contribution of Prof Ranjana Aggarwal, Director, CSIR-National Institute of Science Communication and Policy Research, New Delhi; ICAR-Indian Agricultural Research Institute, National Institute for Plant Biotechnology and National Bureau of Plant Genetic Resources, New Delhi, India; Guru Govind Singh Indraprastha University, Delhi, India; Birla institute of technology and Sciences, Pilani, Rajasthan, India; University of Allahabad Prayagraj, Uttar Pradesh, India; Gautam Buddha University, Greater Noida, Greater Noida Gautam Budh Nagar, Uttar Pradesh, India; Amity University, Noida, UP, India; DBT-National Institute of Plant Genome Research, New Delhi, India; Cairo University and The International Center for Agricultural Research in the Dry Areas (ICARDA) Egypt for providing necessary support for the study. The corresponding author RK also gratefully acknowledges the DST-SERB, New Delhi.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abbo, S., Berger, J., and Turner, N. C. (2003). Evolution of cultivated chickpea: Four bottlenecks limit diversity and constrain adaptation. *Funct. Plant Biol.* 30 (10), 1081–1087.

Abdallah, N. A., Hamwieh, A., Radwan, K., Fouad, N., and Prakash, C. (2021). Genome editing techniques in plants: A comprehensive review and future prospects toward zero hunger. *GM Crops Food* 12, 601–615. doi:10.1080/21645698.2021.2021724

Abdallah, N. A., Prakash, C. S., and McHughen, A. G. (2015). Genome editing for crop improvement: Challenges and opportunities. *GM Crops Food* 6, 183–205. doi:10.1080/21645698.2015.1129937

Abdurakhmonov, I. Y. (2016). Genomics Era for plants and crop species – advances made and needed tasks ahead. *IntechOpen*. doi:10.5772/62083

Akinlade, O. J., Voss-Fels, K., Costilla, R., Kholova, J., Choudhary, S., Varshney, R. K., et al. (2022). Designing chickpea for a hotter drier world. *Euphytica* 218, 100. doi:10.1007/s10681-022-03048-2

Albert, H., Dale, E. C., Lee, E., and Ow, D. W. (1995). Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *Plant J.* 7 (4), 649–659. doi:10.1046/j.1365-313x.1995.7040649.x

Allen, B. G., and Weeks, D. L. (2009). "Bacteriophage φC31 integrase mediated transgenesis in *Xenopus laevis* for protein expression at endogenous levels," in *Microinjection*. Editor D. J. Carroll (Totowa, NJ: Humana Press), 113–122. doi:10.1007/978-1-59745-202-1_9

Allen, B. G., and Weeks, D. L. (2005). Transgenic *Xenopus laevis* embryos can be generated using phiC31 integrase. *Nat. Methods* 2, 975–979. doi:10.1038/nmeth814

Anand, A., Wu, E., Li, Z., TeRonde, S., Arling, M., Lenderts, B., et al. (2019). High-efficiency Agrobacterium-mediated site-specific gene integration in maize utilizing the FLP - FRT recombination system. *Plant Biotechnol. J. pbi* 13089, 1636–1645. doi:10.1111/pbi.13089

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157. doi:10.1038/s41586-019-1711-4

Araki, K., Araki, M., Miyazaki, J., and Vassalli, P. (1995). Site-specific recombination of a transgene in fertilized eggs by transient expression of Cre recombinase. *Proc. Natl. Acad. Sci. U.S.A.* 92, 160–164. doi:10.1073/pnas.92.1.160

Arnould, S., Perez, C., Cabaniols, J.-P., Smith, J., Gouble, A., Grizot, S., et al. (2007). Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J. Mol. Biol.* 371, 49–65. doi:10.1016/j.jmb.2007.04.079

Arora, N. K., and Mishra, I. (2022). Current scenario and future directions for sustainable development goal 2: A roadmap to zero hunger. *Environ. Sustain.* 5, 129–133. doi:10.1007/s42398-022-00235-8

Aslam, S., Khan, S. H., Ahmed, A., and Dandekar, A. M. (2019). Genome editing tools: Need of the current era. *AJMB* 09, 85–109. doi:10.4236/ajmb.2019.93008

Azameti, M. K., and Dauda, W. P. (2021). Base editing in plants: Applications, challenges, and future prospects. *Front. Plant Sci.* 12, 664997. doi:10.3389/fpls.2021.664997

Badhan, S., Ball, A. S., and Mantri, N. (2021). First report of CRISPR/Cas9 mediated DNA-free editing of4CL and RVE7 genes in chickpea protoplasts. *Int. J. Mol. Sci.* 22 (1), 396. doi:10.3390/ijms22010396

Beetham, P. R., Kipp, P. B., Sawycky, X. L., Arntzen, C. J., and May, G. D. (1999). A tool for functional plant genomics: Chimeric RNA/DNA oligonucleotides cause *in vivo* gene-specific mutations. *Proc. Natl. Acad. Sci. U.S.A.* 96, 8774–8778. doi:10.1073/pnas.96.15.8774

Bhardwaj, J., Kumari, N., Ford, R., Yadav, R., Choi, I., and Rajendra, K. (2014). *In silico* development and validation of EST derived new SSR markers for drought tolerance in Cicer arietinum L. *Indian J. Genet. Plant Breed.* 74 (2), 254–256. doi:10.5958/0975-6906.2014.00164.3

Bhardwaj, J., Kumari, N., Mittal, N., Yadav, R., Singh, R. K., and Kumar, R. (2021). Quantitative expression analysis through transcript profiling for drought stress in Cicer arietinum L. *Res. J. Biotechnol.* 16 (3), 26–32.

Bilichak, A., Sastry-Dent, L., Sriram, S., Simpson, M., Samuel, P., Webb, S., et al. (2020). Genome editing in wheat microspores and haploid embryos mediated by delivery of ZFN proteins and cell-penetrating peptide complexes. *Plant Biotechnol. J.* 18, 1307–1316. doi:10.1111/pbi.13296

Billon, P., Bryant, E. E., Joseph, S. A., Nambiar, T. S., Hayward, S. B., Rothstein, R., et al. (2017). CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons. *Mol. Cell* 67, 1068–1079.e4. doi:10.1016/j.molcel.2017.08.008

Biswas, S., Bridgeland, A., Irum, S., Thomson, M. J., and Septiningsih, E. M. (2022). Optimization of prime editing in rice, peanut, chickpea, and cowpea protoplasts by restoration of GFP activity. *Int. J. Mol. Sci.* 23, 9809. doi:10.3390/ijms23179809

Boch, J., and Bonas, U. (2010). Xanthomonas AvrBs3 family-type III effectors: Discovery and function. *Annu. Rev. Phytopathol.* 48, 419–436. doi:10.1146/annurev-phyto-080508-081936

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509–1512. doi:10.1126/science.1178811

Bogdanove, A. J., Schornack, S., and Lahaye, T. (2010). TAL effectors: Finding plant genes for disease and defense. *Curr. Opin. Plant Biol.* 13, 394–401. doi:10.1016/j.pbi.2010.04.010

Bruce Wallace, R., Schold, M., Johnson, M. J., Dembek, P., and Itakura, K. (1981). Oligonucleotide-directed mutagenesis of the human β-globin gene: A general method for producing specific point mutations in cloned DNA. *Nucl. Acids Res.* 9, 3647–3656. doi:10.1093/nar/9.15.3647

Butler, N. M., Baltes, N. J., Voytas, D. F., and Douches, D. S. (2016). Gemini virus-mediated genome editing in potato (Solanum tuberosum L.) using sequence-specific nucleases. *Front. Plant Sci.* 7, 1045. doi:10.3389/fpls.2016.01045

Butt, H., Rao, G. S., Sedeek, K., Aman, R., Kamel, R., and Mahfouz, M. (2020). Engineering herbicide resistance via prime editing in rice. *Plant Biotechnol. J.* 18, 2370–2372. doi:10.1111/pbi.13399

Cai, Y., Bak, R. O., and Mikkelsen, J. G. (2014). Targeted genome editing by lentiviral protein transduction of zinc-finger and TAL-effector nucleases. *Elife* 3, e01911. doi:10.7554/eLife.01911

Capecchi, M. R. (2005). Gene targeting in mice: Functional analysis of the mammalian genome for the twenty-first century. *Nat. Rev. Genet.* 6, 507–512. doi:10.1038/nrg1619

Carroll, D. (2014). Genome engineering with targetable nucleases. *Annu. Rev. Biochem.* 83, 409–439. doi:10.1146/annurev-biochem-060713-035418

Castillo, F., Benmohamed, A., and Szatmari, G. (2017). Xer site specific recombination: Double and single recombinase systems. *Front. Microbiol.* 8, 453. doi:10.3389/fmicb.2017.00453

Celikkol Akcay, U., Mahmoudian, M., Kamci, H., Yucel, M., and Oktem, H. A. (2009). Agrobacterium tumefaciens-mediated genetic transformation of a recalcitrant grain legume, lentil (Lens culinaris Medik). *Plant Cell Rep.* 28, 407–417. doi:10.1007/s00299-008-0652-4

Čermák, T., Baltes, N. J., Čegan, R., Zhang, Y., and Voytas, D. F. (2015). High-frequency, precise modification of the tomato genome. *Genome Biol.* 16, 232. doi:10.1186/s13059-015-0796-9

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., et al. (2011a). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 39, e82. doi:10.1093/nar/gkr218

Chakraborty, J., Sen, S., Ghosh, P., Sengupta, A., Basu, D., and Das, S. (2016). Homologous promoter derived constitutive and chloroplast targeted expression of synthetic cry1Ac in transgenic chickpea confers resistance against helicoverpa armigera. *Plant Cell Tiss. Org. Cult.* 125, 521–535. doi:10.1007/s11240-016-0968-7

Chandana, B. S., Mahto, R. K., Singh, R. K., Ford, R., Vaghefi, N., Kumar, R., et al. (2022). Epigenomics as potential tools for enhancing magnitude of breeding approaches for developing climate resilient chickpea. *Front. Genet.* 13, 900253. doi:10.3389/fgene.2022.900253

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., et al. (2016). Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* 17, 1140–1153. doi:10.1111/mpp.12375

Char, S. N., Unger-Wallace, E., Frame, B., Briggs, S. A., Main, M., Spalding, M. H., et al. (2015). Heritable site-specific mutagenesis using TALENs in maize. *Plant Biotechnol. J.* 13, 1002–1010. doi:10.1111/pbi.12344

Chen, W., Qian, Y., Wu, X., Sun, Y., Wu, X., and Cheng, X. (2014). Inhibiting replication of begomoviruses using artificial zinc finger nucleases that target viral-conserved nucleotide motif. *Virus Genes* 48, 494–501. doi:10.1007/s11262-014-1041-4

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., et al. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757–761. doi:10.1534/genetics.110.120717

Christou, P. (1990). Morphological description of transgenic soybean chimeras created by the delivery, integration and expression of foreign DNA using electric discharge particle acceleration. *Ann. Bot.* 66, 379–386. doi:10.1093/oxfordjournals.aob.a088039

Clasen, B. M., Stoddard, T. J., Luo, S., Demorest, Z. L., Li, J., Cedrone, F., et al. (2016). Improving cold storage and processing traits in potato through targeted gene knockout. *Plant Biotechnol. J.* 14, 169–176. doi:10.1111/pbi.12370

Cramer, G. R., Urano, K., Delrot, S., Pezzotti, M., and Shinozaki, K. (2011). Effects of abiotic stress on plants: A systems biology perspective. *BMC Plant Biol.* 11, 163. doi:10.1186/1471-2229-11-163

Cuculis, L., Zhao, C., Abil, Z., Zhao, H., Shukla, D., and Schroeder, C. M. (2020). Divalent cations promote TALE DNA-binding specificity. *Nucleic Acids Res.* 48, 1406–1422. doi:10.1093/nar/gkz1174

Curtin, S. J., Anderson, J. E., Starker, C. G., Baltes, N. J., Mani, D., Voytas, D. F., et al. (2013). "Targeted mutagenesis for functional analysis of gene duplication in legumes," in *Legume genomics*. Editor R. J. Rose (Totowa, NJ: Humana Press), 25–42. doi:10.1007/978-1-62703-613-9_3

Dale, E. C., and Ow, D. W. (1991). Gene transfer with subsequent removal of the selection gene from the host genome. *Proc. Natl. Acad. Sci. U.S.A.* 88, 10558–10562. doi:10.1073/pnas.88.23.10558

Das, A., Datta, S., Thakur, S., Shukla, A., Ansari, J., Sujayanand, G. K., et al. (2017). Expression of a chimeric gene encoding insecticidal crystal protein Cry1Aabc of Bacillus thuringiensis in chickpea (Cicer arietinum L.) confers resistance to gram pod borer (Helicoverpa armigera Hubner.). *Front. Plant Sci.* 8, 1423. doi:10.3389/fpls.2017.01423

Das Bhowmik, S. S., Cheng, A. Y., Long, H., Tan, G. Z. H., Hoang, T. M. L., Karbaschi, M. R., et al. (2019). Robust genetic transformation system to obtain non-chimeric transgenic chickpea. *Front. plant Sci.* 10, 524. doi:10.3389/fpls.2019.00524

De Buck, S., Peck, I., De Wilde, C., Marjanac, G., Nolf, J., De Paepe, A., et al. (2007). Generation of single-copy T-DNA transformants in Arabidopsis by the CRE/loxP recombination-mediated resolution system. *Plant Physiol.* 145, 1171–1182. doi:10.1104/pp.107.104067

de Pater, S., Neuteboom, L. W., Pinas, J. E., Hooykaas, P. J. J., and van der Zaal, B. J. (2009). ZFN-induced mutagenesis and gene-targeting in Arabidopsis through Agrobacterium-mediated floral dip transformation. *Plant Biotechnol. J.* 7, 821–835. doi:10.1111/j.1467-7652.2009.00446.x

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.-K., et al. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335, 720–723. doi:10.1126/science.1215670

Dillen, W., Clercq, J. D., Goossens, A., Montagu, M. V., and Angenon, G. (1997). Agrobacterium-mediated transformation of Phaseolus acutifolius A. Gray. *Gray. Theor. Appl. Genet.* 94, 151–158. doi:10.1007/s001220050394

Dong, C., Beetham, P., Vincent, K., and Sharp, P. (2006). Oligonucleotide-directed gene repair in wheat using a transient plasmid gene repair assay system. *Plant Cell Rep.* 25, 457–465. doi:10.1007/s00299-005-0098-x

Du, H., Zeng, X., Zhao, M., Cui, X., Wang, Q., Yang, H., et al. (2016). Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. *J. Biotechnol.* 217, 90–97. doi:10.1016/j.jbiotec.2015.11.005

Epinat, J.-C., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., et al. (2003). A novel-engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res. Spec. Publ.* 31, 2952–2962. doi:10.1093/nar/gkg375

FAO (2019). FAOSTAT. Available at: https://www.fao.org/faostat/en/#home (Accessed October 14, 2022).

Fladung, M. (2016). Cibus' herbicide-resistant canola in European limbo. *Nat. Biotechnol.* 34 (5), 473–474. doi:10.1038/nbt.3558

Gaj, T., Gersbach, C. A., and Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405. doi:10.1016/j.tibtech.2013.04.004

Gaj, T., Sirk, S. J., and Barbas, C. F. (2014). Expanding the scope of site-specific recombinases for genetic and metabolic engineering: Expanding the Scope of Site-Specific Recombinases. *Biotechnol. Bioeng.* 111, 1–15. doi:10.1002/bit.25096

Gaj, T., Sirk, S. J., Shui, S., and Liu, J. (2016). Genome-editing technologies: Principles and applications. *Cold Spring Harb. Perspect. Biol.* 8, a023754. doi:10.1101/cshperspect.a023754

Ganguly, M., Molla, K. A., Karmakar, S., Datta, K., and Datta, S. K. (2014). Development of pod borer-resistant transgenic chickpea using a pod-specific and a constitutive promoter-driven fused cry1Ab/Ac gene. *Theor. Appl. Genet.* 127 (12), 2555–2565. doi:10.1007/s00122-014-2397-5

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., et al. (2017). Programmable base editing of AT to GC in genomic DNA without DNA cleavage. *Nature* 551, 464–471. doi:10.1038/nature24644

Gaur, P. M., Krishnamurthy, L., and Kashiwagi, J. (2008). Improving drought-avoidance root traits in chickpea (cicer arietinum L.) -current status of research at ICRISAT. *Plant Prod. Sci.* 11, 3–11. doi:10.1626/pps.11.3

Ghanti, S. K. K., Sujata, K. G., Kumar, B. V., Karba, N. N., JanardhanReddy, K., Rao, M. S., et al. (2011). Heterologous expression of P5CS gene in chickpea enhances salt tolerance without affecting yield. *Biol. Plant.* 55, 634. doi:10.1007/s10535-011-0161-0

Gocal, G. F. W., Schöpke, C., and Beetham, P. R. (2015). "Oligo-mediated targeted gene editing," in *Advances in new technology for targeted modification of plant genomes*. Editors F. Zhang, H. Puchta, and J. G. Thomson (New York, NY: Springer), 73–89. doi:10.1007/978-1-4939-2556-8_5

Golic, K. G., and Lindquist, S. (1989). The FLP recombinase of yeast catalyzes site-specific recombination in the drosophila genome. *Cell* 59, 499–509. doi:10.1016/0092-8674(89)90033-0

Grindley, N. D. F., Whiteson, K. L., and Rice, P. A. (2006). Mechanisms of site-specific recombination. *Annu. Rev. Biochem.* 75, 567–605. doi:10.1146/annurev.biochem.73.011303.073908

Grizot, S., Smith, J., Daboussi, F., Prieto, J., Redondo, P., Merino, N., et al. (2009). Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res. Spec. Publ.* 37, 5405–5419. doi:10.1093/nar/gkp548

Gupta, M., DeKelver, R. C., Palta, A., Clifford, C., Gopalan, S., Miller, J. C., et al. (2012). Transcriptional activation of Brassica napus β-ketoacyl-ACP synthase II with an engineered zinc finger protein transcription factor: Transcriptional activation in Brassica napus. *Plant Biotechnol. J.* 10, 783–791. doi:10.1111/j.1467-7652.2012.00695.x

Gupta, R. M., and Musunuru, K. (2014). Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-cas9. *J. Clin. Invest.* 124, 4154–4161. doi:10.1172/JCI72992

Hahn, F., and Nekrasov, V. (2019). CRISPR/Cas precision: Do we need to worry about off-targeting in plants? *Plant Cell Rep.* 38, 437–441. doi:10.1007/s00299-018-2355-9

Hajyzadeh, M., Turktas, M., Khawar, K. M., and Unver, T. (2015). miR408 overexpression causes increased drought tolerance in chickpea. *Gene* 555, 186–193. doi:10.1016/j.gene.2014.11.002

Henchion, M., Hayes, M., Mullen, A., Fenelon, M., and Tiwari, B. (2017). Future protein supply and demand: Strategies and factors influencing a sustainable equilibrium. *Foods* 6, 53. doi:10.3390/foods6070053

Hickey, L. T., Hafeez, N., Robinson, H., Jackson, S. A., Leal-Bertioli, S. C. M., Tester, M., et al. (2019). Breeding crops to feed 10 billion. *Nat. Biotechnol.* 37, 744–754. doi:10.1038/s41587-019-0152-9

Hoa, T. T. C., Bong, B. B., Huq, E., and Hodges, T. K. (2002). Cre/lox site-specific recombination controls the excision of a transgene from the rice genome. *Theor. Appl. Genet.* 104, 518–525. doi:10.1007/s001220100748

Hu, Q., Kononowicz-Hodges, H., Nelson-Vasilchik, K., Viola, D., Zeng, P., Liu, H., et al. (2008). FLP recombinase-mediated site-specific recombination in rice. *Plant Biotechnol. J.* 6, 176–188. doi:10.1111/j.1467-7652.2007.00310.x

Hua, K., Jiang, Y., Tao, X., and Zhu, J. (2020). Precision genome engineering in rice using prime editing system. *Plant Biotechnol. J.* 18, 2167–2169. doi:10.1111/pbi.13395

Huda, S., Islam, R., and Bari, M. (2000). Shoot regeneration from internode derived callus of chickpea (Cicer arietinum L.). *Int. Chickpea Pigeonpea Newslett.* 7, 28–29.

Iizumi, T., Shiogama, H., Imada, Y., Hanasaki, N., Takikawa, H., and Nishimori, M. (2018). Crop production losses associated with anthropogenic climate change for 1981-2010 compared with preindustrial levels. *Int. J. Climatol.* 38, 5405–5417. doi:10.1002/joc.5818

Indurker, S., Misra, H. S., and Eapen, S. (2010). Agrobacterium-mediated transformation in chickpea (cicer arietinum L.) with an insecticidal protein gene: Optimisation of different factors. *Physiol. Mol. Biol. Plants* 16 (3), 273–284. doi:10.1007/s12298-010-0030-x

Ito, Y., Nishizawa-Yokoi, A., Endo, M., Mikami, M., and Toki, S. (2015). CRISPR/Cas9-mediated mutagenesis of the RIN locus that regulates tomato fruit ripening. *Biochem. Biophys. Res. Commun.* 467, 76–82. doi:10.1016/j.bbrc.2015.09.117

Jaganathan, D., Ramasamy, K., Sellamuthu, G., Jayabalan, S., and Venkataraman, G. (2018). CRISPR for crop improvement: An update review. *Front. plant Sci.* 9, 985. doi:10.3389/fpls.2018.00985

Jankele, R., and Svoboda, P. (2014). TAL effectors: Tools for DNA targeting. *Brief. Funct. Genom* 13, 409–419. doi:10.1093/bfgp/elu013

Jha, U. C., Chaturvedi, S. K., Bohra, A., Basu, P. S., Khan, M. S., and Barh, D. (2014). Abiotic stresses, constraints and improvement strategies in chickpea. *Plant Breed.* 133 (2), 163–178. doi:10.1111/pbr.12150

Ji, X., Zhang, H., Zhang, Y., Wang, Y., and Gao, C. (2015). Establishing a CRISPR-Cas-like immune system conferring DNA virus resistance in plants. *Nat. Plants* 1, 15144. doi:10.1038/nplants.2015.144

Jiang, W. Z., Henry, I. M., Lynagh, P. G., Comai, L., Cahoon, E. B., and Weeks, D. P. (2017). Significant enhancement of fatty acid composition in seeds of the allohexaploid, Camelina sativa, using CRISPR/Cas9 gene editing. *Plant Biotechnol. J.* 15, 648–657. doi:10.1111/pbi.12663

Jiang, Y.-Y., Chai, Y.-P., Lu, M.-H., Han, X.-L., Lin, Q., Zhang, Y., et al. (2020). Prime editing efficiently generates W542L and S621I double mutations in two ALS genes in maize. *Genome Biol.* 21, 257. doi:10.1186/s13059-020-02170-5

Joung, J. K., and Sander, J. D. (2013). TALENs: A widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* 14, 49–55. doi:10.1038/nrm3486

Jung, J. H., and Altpeter, F. (2016). TALEN mediated targeted mutagenesis of the caffeic acid O-methyltransferase in highly polyploid sugarcane improves cell wall composition for production of bioethanol. *Plant Mol. Biol.* 92, 131–142. doi:10.1007/s11103-016-0499-y

Kamburova, V. S., Nikitina, E. V., Shermatov, S. E., Buriev, Z. T., Kumpatla, S. P., Emani, C., et al. (2017). Genome editing in plants: An overview of tools and applications. *J. Agron.* 2017, 1–15. doi:10.1155/2017/7315351

Kannan, B., Jung, J. H., Moxley, G. W., Lee, S.-M., and Altpeter, F. (2018). TALEN-mediated targeted mutagenesis of more than 100 COMT copies/alleles in highly polyploid sugarcane improves saccharification efficiency without compromising biomass yield. *Plant Biotechnol. J.* 16, 856–866. doi:10.1111/pbi.12833

Kapusi, E., Kempe, K., Rubtsova, M., Kumlehn, J., and Gils, M. (2012). phiC31 integrase-mediated site-specific recombination in barley. *PLoS One* 7, e45353. doi:10.1371/journal.pone.0045353

Katoch, O., Chauhan, U. S., Yadav, R., Yadav, S. S., Kumar, A., Kumar, R., et al. (2016). Nitrate Reductase based phylogenetic analysis in chickpea. *Res. J. Chem. Environ.* 20 (7), 1–8.

Kerbach, S., Lörz, H., and Becker, D. (2005). Site-specific recombination in Zea mays. *Theor. Appl. Genet.* 111, 1608–1616. doi:10.1007/s00122-005-0092-2

Khandagale, K., and Nadaf, A. (2016). Genome editing for targeted improvement of plants. *Plant Biotechnol. Rep.* 10, 327–343. doi:10.1007/s11816-016-0417-4

Khine, N. A., Kundu, K. K., Malik, D. P., and Papang, J. S. (2022). Production and trade of chickpea (cicer aneityum) in India and Myanmar. *Indian J. Appl. Bus. Econ. Res.* 3 (1), 81–97. doi:10.47509/IJABER.2022.v03i01.06

Kim, Y. G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U.S.A.* 93, 1156–1160. doi:10.1073/pnas.93.3.1156

Klap, C., Yeshayahou, E., Bolger, A. M., Arazi, T., Gupta, S. K., Shabtai, S., et al. (2017). Tomato facultative parthenocarpy results from SlAGAMOUS-LIKE 6 loss of function. *Plant Biotechnol. J.* 15, 634–647. doi:10.1111/pbi.12662

Kluesner, M. G., Lahr, W. S., Lonetree, C., Smeester, B. A., Qiu, X., Slipek, N. J., et al. (2021). CRISPR-Cas9 cytidine and adenosine base editing of splice-sites mediates highly-efficient disruption of proteins in primary and immortalized cells. *Nat. Commun.* 12, 2437. doi:10.1038/s41467-021-22009-2

Kochevenko, A., and Willmitzer, L. (2003). Chimeric RNA/DNA oligonucleotide-based site-specific modification of the tobacco acetolactate syntase gene. *Plant Physiol.* 132, 174–184. doi:10.1104/pp.102.016857

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424. doi:10.1038/nature17946

Kumar, R., Yadav, R., Soi, S., SrinivasanYadav, S. S., Mishra, J. P., et al. (2017). Morpho-molecular characterization of landraces/wild genotypes of cicer for biotic/abiotic stresses. *Legum. Res.* 40 (6), 974–984. doi:10.18805/lr.v0iOF.9100/LR-3701

Landy, A. (2015). The λ integrase site-specific recombination pathway. *Microbiol. Spectr.* 3 (2), MDNA3–2014. doi:10.1128/microbiolspec.MDNA3-0051-2014

Le Rhun, A., Escalera-Maurer, A., Bratovič, M., and Charpentier, E. (2019). CRISPR-Cas in Streptococcus pyogenes. *RNA Biol.* 16, 380–389. doi:10.1080/15476286.2019.1582974

Li, H., Li, J., Chen, J., Yan, L., and Xia, L. (2020). Precise modifications of both exogenous and endogenous genes in rice by prime editing. *Mol. Plant* 13, 671–674. doi:10.1016/j.molp.2020.03.011

Li, L., Atef, A., Piatek, A., Ali, Z., Piatek, M., Aouida, M., et al. (2013). Characterization and DNA-binding specificities of Ralstonia TAL-like effectors. *Mol. Plant* 6, 1318–1330. doi:10.1093/mp/sst006

Li, M., Li, X., Zhou, Z., Wu, P., Fang, M., Pan, X., et al. (2016). Reassessment of the four yield-related genes Gn1a, DEP1, GS3, and IPA1 in rice using a CRISPR/Cas9 system. *Front. Plant Sci.* 7, 377. doi:10.3389/fpls.2016.00377

Li, T., Liu, B., Spalding, M. H., Weeks, D. P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.* 30, 390–392. doi:10.1038/nbt.2199

Li, Z., Xing, A., Moon, B. P., McCardell, R. P., Mills, K., and Falco, S. C. (2009). Site-specific integration of transgenes in soybean via recombinase-mediated DNA cassette exchange. *Plant Physiol.* 151, 1087–1095. doi:10.1104/pp.109.137612

Liang, Z., Zhang, K., Chen, K., and Gao, C. (2014). Targeted mutagenesis in Zea mays using TALENs and the CRISPR/Cas system. *J. Genet. Genomics* 41, 63–68. doi:10.1016/j.jgg.2013.12.001

Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., et al. (2020). Prime genome editing in rice and wheat. *Nat. Biotechnol.* 38, 582–585. doi:10.1038/s41587-020-0455-x

Lloyd, A., Plaisier, C. L., Carroll, D., and Drews, G. N. (2005). Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2232–2237. doi:10.1073/pnas.0409339102

Lu, H., Klocko, A. L., Dow, M., Ma, C., Amarasinghe, V., and Strauss, S. H. (2016). Transient hypoxia reprograms differentiating adipocytes for enhanced insulin sensitivity and triglyceride accumulation. *Mol. Breed.* 36, 121–128. doi:10.1038/ijo.2015.137

Lu, H., Luo, T., Fu, H., Wang, L., Tan, Y., Huang, J., et al. (2018). Resistance of rice to insect pests mediated by suppression of serotonin biosynthesis. *Nat. Plants.* 4, 338–344. doi:10.1038/s41477-018-0152-7

Lu, Y., Tian, Y., Shen, R., Yao, Q., Zhong, D., Zhang, X., et al. (2021). Precise genome modification in tomato using an improved prime editing system. *Plant Biotechnol. J.* 19, 415–417. doi:10.1111/pbi.13497

Lucht, J. M. (2015). Public acceptance of plant biotechnology and GM crops. *Viruses* 7, 4254–4281. doi:10.3390/v7082819

Lusser, M., Parisi, C., Plan, D., and Rodríguez-Cerezo, E. (2012). Deployment of new biotechnologies in plant breeding. *Nat. Biotechnol.* 30, 231–239. doi:10.1038/nbt.2142

Maeder, M. L., Thibodeau-Beganny, S., Osiak, A., Wright, D. A., Anthony, R. M., Eichtinger, M., et al. (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell.* 31, 294–301. doi:10.1016/j.molcel.2008.06.016

Mahfouz, M. M., Li, L., Shamimuzzaman, Md., Wibowo, A., Fang, X., and Zhu, J.-K. (2011). De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2623–2628. doi:10.1073/pnas.1019533108

Mahto, R. K., AmbikaSingh, C., Chandana, B. S., Singh, R. K., Verma, S., et al. (2022). Chickpea biofortification for cytokinin dehydrogenase via genome editing to enhance abiotic-biotic stress tolerance and food security. *Front. Genet.* 13, 900324. doi:10.3389/fgene.2022.900324

Mantri, N. L., Ford, R., Coram, T. E., and Pang, E. C. K. (2010). Evidence of unique and shared responses to major biotic and abiotic stresses in chickpea. *Environ. Exp. Bot.* 69, 286–292. doi:10.1016/j.envexpbot.2010.05.003

Marshall Stark, W. (2015). "The serine recombinases," in *Mobile DNA III*. Editors N. L. Craig, M. Chandler, M. Gellert, A. M. Lambowitz, P. A. Rice, and S. B. Sandmeyer (Washington, DC, USA: ASM Press), 73–89. doi:10.1128/9781555819217.ch3

Mehrotra, M., Singh, A. K., Sanyal, I., Altosaar, I., and Amla, D. V. (2011). Pyramiding of modified cry1Ab and cry1Ac genes of Bacillus thuringiensis in transgenic chickpea (Cicer arietinum L.) for improved resistance to pod borer insect Helicoverpa armigera. *Euphytica* 182, 87. doi:10.1007/s10681-011-0501-3

Meng, X., Yu, H., Zhang, Y., Zhuang, F., Song, X., Gao, S., et al. (2017). Construction of a genome-wide mutant library in rice using CRISPR/Cas9. *Mol. Plant* 10 (9), 1238–1241. doi:10.1016/j.molp.2017.06.006

Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., et al. (2013). Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res.* 23, 1233–1236. doi:10.1038/cr.2013.123

Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y.-L., Rupniewski, I., et al. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* 25, 778–785. doi:10.1038/nbt1319

Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., et al. (2011). A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* 29, 143–148. doi:10.1038/nbt.1755

Mishra, R., Joshi, R. K., and Zhao, K. (2020). Base editing in crops: Current advances, limitations, and future implications. *Plant Biotechnol. J.* 18, 20–31. doi:10.1111/pbi.13225

Mittal, N., Saxena, R. S., Bhardwaj, J., and Kumar, R. (2014). Construction of genetic linkage map and QTL mapping for agro-morphological traits in Chickpea (Cicer aritenium L.) under two contrasting water treatment conditions. *Prog. Agric. Int. J.* 14 (2), 224–232.

Mittal, N., Saxena, R. S., Yadav, R., Bhardwaj, J., Yadav, N., Kumar, R., et al. (2015). Comparative quantitative gene expression analysis between salinity tolerant and susceptible varieties of chickpea. *Vegetos- Int. J. Plant Res.* 28 (2), 103–110. doi:10.5958/2229-4473.2015.00044.0

Mok, B. Y., de Moraes, M. H., Zeng, J., Bosch, D. E., Kotrys, A. V., Raguram, A., et al. (2020). A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* 583, 631–637. doi:10.1038/s41586-020-2477-4

Molla, K. A., Sretenovic, S., Bansal, K. C., and Qi, Y. (2021). Precise plant genome editing using base editors and prime editors. *Nat. Plants* 7, 1166–1187. doi:10.1038/s41477-021-00991-1

Molla, K. A., and Yang, Y. (2019). CRISPR/Cas-Mediated base editing: Technical considerations and practical applications. *Trends Biotechnol.* 37, 1121–1142. doi:10.1016/j.tibtech.2019.03.008

Moon, H. S., Abercrombie, L. L., Eda, S., Blanvillain, R., Thomson, J. G., Ow, D. W., et al. (2011). Transgene excision in pollen using a codon-optimized serine resolvase CinH-RS2 site-specific recombination system. *Plant Mol. Biol.* 75, 621–631. doi:10.1007/s11103-011-9756-2

Mushtaq, M., Sakina, A., Wani, S. H., Shikari, A. B., Tripathi, P., Zaid, A., et al. (2019). Harnessing genome editing techniques to engineer disease resistance in plants. *Front. Plant Sci.* 10, 550. doi:10.3389/fpls.2019.00550

Musunuru, K. (2017). The hope and hype of CRISPR-cas9 genome editing: A review. *JAMA Cardiol.* 2, 914–919. doi:10.1001/jamacardio.2017.1713

Nash, H. A., Rice, P. A., Yang, S., and Mizuuchi, K. (1996). Crystal structure of an IHF- DNA complex: A protein-induced DNA U-turn. *Cell* 87 (7), 1295–1306. doi:10.1016/s0092-8674(00)81824-3

Nelson, D. R., Adger, W. N., and Brown, K. (2007). Adaptation to environmental change: Contributions of a resilience framework. *Annu. Rev. Environ. Res.* 32 (1), 395–419. doi:10.1146/annurev.energy.32.051807.090348

Nemudryi, A. A., Valetdinova, K. R., Medvedev, S. P., and Zakian, S. M. (2014). TALEN and CRISPR/cas genome editing systems: Tools of discovery. *Acta Naturae* 6, 19–40. doi:10.32607/20758251-2014-6-3-19-40

Ninan, A., Grant, J., Song, J., and Jameson, P. (2019). Expression of genes related to sugar and amino acid transport and cytokinin metabolism during leaf development and senescence in *Pisum sativum* L. *L. Plants* 8, 76. doi:10.3390/plants8030076

Ochatt, S., Conreux, C., Moussa Mcolo, R., Despierre, G., Magnin-Robert, J.-B., and Raffiot, B. (2018). Phytosulfokine-alpha, an enhancerof *in vitro* regeneration competence in recalcitrant legumes. *PCTOC* 135, 189–201. doi:10.1007/s11240-018-1455-0

Odipio, J., Alicai, T., Ingelbrecht, I., Nusinow, D. A., Bart, R., and Taylor, N. J. (2017). Efficient CRISPR/Cas9 genome editing of phytoene desaturase in cassava. *Front. Plant Sci.* 8, 1780. doi:10.3389/fpls.2017.01780

Okuzaki, A., and Toriyama, K. (2004). Chimeric RNA/DNA oligonucleotide-directed gene targeting in rice. *Plant Cell Rep.* 22, 509–512. doi:10.1007/s00299-003-0698-2

Olorunniji, F. J., Rosser, S. J., and Stark, W. M. (2016). Site-specific recombinases: Molecular machines for the genetic revolution. *Biochem. J.* 473, 673–684. doi:10.1042/BJ20151112

Onouchi, H., Yokoi, K., Machida, C., Matsuzaki, H., Oshima, Y., Matsuoka, K., et al. (1991). Operation of an efficient site-specific recombination system of Zygosaccharomyces rouxii in tobacco cells. *Nucleic Acids Res.* 19, 6373–6378. doi:10.1093/nar/19.23.6373

Osakabe, Y., and Osakabe, K. (2015). Genome editing with engineered nucleases in plants. *Plant Cell Physiol.* 56, 389–400. doi:10.1093/pcp/pcu170

Park, J.-R., Kim, E.-G., Jang, Y.-H., Jan, R., Farooq, M., Ubaidillah, M., et al. (2022). Applications of CRISPR/Cas9 as new strategies for short breeding to drought gene in rice. *Front. Plant Sci.* 13, 850441. doi:10.3389/fpls.2022.850441

Peer, R., Rivlin, G., Golobovitch, S., Lapidot, M., Gal-On, A., Vainstein, A., et al. (2015). Targeted mutagenesis using zinc-finger nucleases in perennial fruit trees. *Planta* 241, 941–951. doi:10.1007/s00425-014-2224-x

Petolino, J. F. (2015). Genome editing in plants via designed zinc finger nucleases. *Dev. Biol.Plant* 51, 1–8. doi:10.1007/s11627-015-9663-3

Piatek, A. A., Lenaghan, S. C., and Neal Stewart, C. (2018). Advanced editing of the nuclear and plastid genomes in plants. *Plant Sci.* 273, 42–49. doi:10.1016/j.plantsci.2018.02.025

Polowick, P. L., Baliski, D. S., and Mahon, J. D. (2004). Agrobacterium tumefaciens-mediated transformation of chickpea (cicer arietinum L.): Gene integration, expression and inheritance. *Plant Cell Rep.* 23, 485–491. doi:10.1007/s00299-004-0857-0

Radhakrishnan, P., and Srivastava, V. (2005). Utility of the FLP-FRT recombination system for genetic manipulation of rice. *Plant Cell Rep.* 23, 721–726. doi:10.1007/s00299-004-0876-x

Rádi, F., Nagy, B., Ferenc, G., Török, K., Nagy, I., Zombori, Z., et al. (2021). In planta test system for targeted cellular mutagenesis by injection of oligonucleotides to apical meristem of maize seedlings. *Acta Physiol. Plant* 43, 79. doi:10.1007/s11738-021-03252-y

Raftery, A. E., Zimmer, A., Frierson, D. M. W., Startz, R., and Liu, P. (2017). Less than 2 °C warming by 2100 unlikely. *Nat. Clim. change* 7, 637–641. doi:10.1038/nclimate3352

Ran, Y., Patron, N., Kay, P., Wong, D., Buchanan, M., Cao, Y.-Y., et al. (2018). Zinc finger nuclease-mediated precision genome editing of an endogenous gene in hexaploid bread wheat (*Triticum aestivum*) using a DNA repair template. *Plant Biotechnol. J.* 16, 2088–2101. doi:10.1111/pbi.12941

Rani, A., Devi, P., Jha, U. C., Sharma, K. D., Siddique, K. H., and Nayyar, H. (2020). Developing climate-resilient chickpea involving physiological and molecular approaches with a focus on temperature and drought stresses. *Front. plant Sci.* 10, 1759. doi:10.3389/fpls.2019.01759

Rastogi Verma, S. (2013). Genetically modified plants: Public and scientific perceptions. *Int. Sch. Res. Not.* 2013, e820671. doi:10.5402/2013/820671

Razzaq, A., Saleem, F., Kanwal, M., Mustafa, G., Yousaf, S., Imran Arshad, H. M., et al. (2019). Modern trends in plant genome editing: An inclusive review of the CRISPR/Cas9 toolbox. *Int. J. Mol. Sci.* 20, 4045. doi:10.3390/ijms20164045

Reddy, A. A., Rani, C. R., Cadman, T., Kumar, S. N., and Reddy, A. N. (2016). Towards sustainable indicators of food and nutritional outcomes in India. *World Rev. Sci. Technol. sustain. Dev.* 13 (2), 128–142. doi:10.1108/wjstsd-10-2015-0049

Rees, H. A., and Liu, D. R. (2018). Base editing: Precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* 19, 770–788. doi:10.1038/s41576-018-0059-1

Rehman, K., Fatima, F., Waheed, I., and Akash, M. S. H. (2018). Prevalence of exposure of heavy metals and their impact on health consequences. *J. Cell. Biochem.* 119, 157–184. doi:10.1002/jcb.26234

Remigi, P., Anisimova, M., Guidot, A., Genin, S., and Peeters, N. (2011). Functional diversification of the GALA type III effector family contributes to Ralstonia solanacearum adaptation on different plant hosts. *New Phytol.* 192, 976–987. doi:10.1111/j.1469-8137.2011.03854.x

Ren, Q., Sretenovic, S., Liu, G., Zhong, Z., Wang, J., Huang, L., et al. (2021). Improved plant cytosine base editors with high editing activity, purity, and specificity. *Plant Biotechnol. J.* 19, 2052–2068. doi:10.1111/pbi.13635

Rezaei, M. K., Deokar, A., and Tar'an, B. (2016). Identification and expression analysis of candidate genes involved in carotenoid biosynthesis in chickpea seeds. *Front. Plant Sci.* 7, 1867. doi:10.3389/fpls.2016.01867

Ruiter, R., Van Den Brande, I., Stals, E., Delauré, S., Cornelissen, M., and D'Halluin, K. (2003). Spontaneous mutation frequency in plants obscures the effect of chimeraplasty. *Plant Mol. Biol.* 53, 675–689. doi:10.1023/B:PLAN.0000019111.96107.01

Russell, S. H., Hoopes, J. L., and Odell, J. T. (1992). Directed excision of a transgene from the plant genome. *Mol. Genet. Genom.* 234, 49–59. doi:10.1007/BF00272344

Rutkai, E., György, A., Dorgai, L., and Weisberg, R. A. (2006). Role of secondary attachment sites in changing the specificity of site-specific recombination. *J. Bacteriol.* 188 (9), 3409–3411. doi:10.1128/JB.188.9.3409-3411.2006

Sadhu, S. K., Jogam, P., Gande, K., Banoth, R., Penna, S., and Peddaboina, V. (2022). Optimization of different factors for an Agrobacterium-mediated genetic transformation system using embryo axis explants of chickpea (Cicer arietinum L.). *J. Plant Biotechnol.* 49 (1), 61–73. doi:10.5010/jpb.2022.49.1.061

Sanyal, I., Singh, A. K., Kaushik, M., and Amla, D. V. (2005). Agrobacterium-mediated transformation of chickpea (cicer arietinum L.) with Bacillus thuringiensis cry1Ac gene for resistance against pod borer insect helicoverpa armigera. *Plant Sci.* 168 (4), 1135–1146. doi:10.1016/j.plantsci.2004.12.015

Sarmah, B. K., Moore, A., Tate, W., Molvig, L., Morton, R. L., Rees, D. P., et al. (2004). Transgenic chickpea seeds expressing high levels of a bean-amylase inhibitor. *Mol. Breed.* 14, 73–82. doi:10.1023/b:molb.0000037996.01494.12

Sauer, B., and Henderson, N. (1989). Cre-stimulated recombination at loxP -containing DNA sequences placed into the mammalian genome. *Nucleic Acids Res.* 17, 147–161. doi:10.1093/nar/17.1.147

Sauer, N. J., Mozoruk, J., Miller, R. B., Warburg, Z. J., Walker, K. A., Beetham, P. R., et al. (2016). Oligonucleotide-directed mutagenesis for precision gene editing. *Plant Biotechnol. J.* 14, 496–502. doi:10.1111/pbi.12496

Sedeek, K. E. M., Mahas, A., and Mahfouz, M. (2019). Plant genome engineering for targeted improvement of crop traits. *Front. plant Sci.* 10, 114. doi:10.3389/fpls.2019.00114

Shan, Q., Wang, Y., Li, J., and Gao, C. (2014). Genome editing in rice and wheat using the CRISPR/Cas system. *Nat. Protoc.* 9, 2395–2410. doi:10.1038/nprot.2014.157

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 686–688. doi:10.1038/nbt.2650

Shi, J., Gao, H., Wang, H., Lafitte, H. R., Archibald, R. L., Yang, M., et al. (2017). ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnol. J.* 15, 207–216. doi:10.1111/pbi.12603

IPCC (2019). "Summary for policymakers," in *Climate change and land: An IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*. Editors P. R. Shukla, J. Skea, E. Calvo Buendia, V. Masson Delmotte, H. O. Portner, D. C. Roberts, et al. (Geneva, Cambridge: IPCC, Cambridge University Press).

Singh, A. P., Singh, R. K., Hegde, V. S., Shukla, N., Saini, P., Kumar, R., et al. (2022). Unclasping potential chickpea resources for the antioxidant enzyme Superoxide Dismutase. *J. Sci. Food Agric.* 2 (7), 320–331. doi:10.1002/jsf2.65

Singh, P., Kurpad, A., Verma, V., Davendra, N., Arun, K., Bhargava, B., et al. (2021). Global Hunger Index does not really measure hunger - an Indian perspective. *Indian J. Med. Res.* 154 (3), 455–460. doi:10.4103/ijmr.ijmr_2057_21

Singh, R. K., PrernaSenger, R. S., Bhatnagar, S. K., and Kumar, R. (2012). Molecular diversity analysis of selected drought resistant chickpea (cicer arietinum L) genotypes. *Vegetos* 25 (1), 111–116.

Singh, R. K., Singh, C., AmbikaChandana, B, S., Mahto, R, K., Kumar, R., Patial, R., et al. (2022). Exploring chickpea germplasm diversity for broadening the genetic base utilizing genomic resourses. *Front. Genet.* 13, 905771. doi:10.3389/fgene.2022.905771

Singh, R., Kumari, N., Upadhyaya, H. D., Yadav, R., Choi, I., Kumar, R., et al. (2013). Molecular analysis for genetic structure of biotic and abiotic stress resistant genotypes in chickpea (Cicer arietinum L.). *Indian J. Biotechnol.* 12 (4), 537–540.

Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J.-C., Chames, P., et al. (2006). A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.* 34, e149. doi:10.1093/nar/gkl720

Smith, J. J., Jantz, D., and Hellinga, H. W. (2012). *Methods of cleaving DNA with rationally-designed meganucleases*. U.S. Patent No 8,143,051. Washington, DC: U.S. Patent and Trademark Office.

Smith, M. C. M., and Thorpe, H. M. (2002). Diversity in the serine recombinases: Diversity in the serine recombinases. *Mol. Microbiol.* 44, 299–307. doi:10.1046/j.1365-2958.2002.02891.x

Srivastava, V., Anderson, O. D., and Ow, D. W. (1999). Single-copy transgenic wheat generated through the resolution of complex integration patterns. *Proc. Natl. Acad. Sci. U.S.A.* 96, 11117–11121. doi:10.1073/pnas.96.20.11117

Srivastava, V., and Ow, D. W. (2015). "Simplifying transgene locus structure through cre-lox recombination," in *Plant gene silencing*. Editors K. S. Mysore and M. Senthil-Kumar (New York, NY: Springer New York), 95–103. doi:10.1007/978-1-4939-2453-0_6

Stuurman, J., de Vroomen, M. J., Nijkamp, H. J. J., and van Haaren, M. J. J. (1996). Single-site manipulation of tomato chromosomes *in vitro* and *in vivo* using Cre-lox site-specific recombination. *Plant Mol. Biol.* 32, 901–913. doi:10.1007/BF00020487

Sugita, K., Kasahara, T., Matsunaga, E., and Ebinuma, H. (2000). A transformation vector for the production of marker-free transgenic plants containing a single copy transgene at high frequency: The GST-MAT vector system. *Plant J.* 22, 461–469. doi:10.1046/j.1365-313X.2000.00745.x

Sun, Y., Jiao, G., Liu, Z., Zhang, X., Li, J., Guo, X., et al. (2017). Generation of high-amylose rice through CRISPR/Cas9-Mediated targeted mutagenesis of starch branching enzymes. *Front. Plant Sci.* 8, 298. doi:10.3389/fpls.2017.00298

Takatsuka, A., Kazama, T., Arimura, S., and Toriyama, K. (2022). TALEN -mediated depletion of the mitochondrial gene orf312 proves that it is a Tadukan-type cytoplasmic male sterility-causative gene in rice. *Plant J.* 110, 994–1004. doi:10.1111/tpj.15715

Tan, S., Evans, R. R., Dahmer, M. L., Singh, B. K., and Shaner, D. L. (2005). Imidazolinone-tolerant crops: History, current status and future. *Pest. Manag. Sci.* 61, 246–257. doi:10.1002/ps.993

Tan, G. Z., Das, B. S. S., Hoang, T. M., Karbaschi, M. R., Long, H., Cheng, A., et al. (2018). Investigation of baseline iron levels in Australian chickpea and evaluation of a transgenic biofortification approach. *Front. Plant Sci.* 9, 788. doi:10.3389/fpls.2018.00788

Taranto, F., Nicolia, A., Pavan, S., De Vita, P., and D'Agostino, N. (2018). Biotechnological and digital revolution for climate-smart plant breeding. *Agronomy* 8, 277. doi:10.3390/agronomy8120277

Thomson, J. G., Chan, R., Smith, J., Thilmony, R., Yau, Y.-Y., Wang, Y., et al. (2012). The Bxb1 recombination system demonstrates heritable transmission of site-specific excision in Arabidopsis. *BMC Biotechnol.* 12, 9. doi:10.1186/1472-6750-12-9

Thorpe, H. M., and Smith, M. C. M. (1998). *In vitro* site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc. Natl. Acad. Sci. U.S.A.* 95, 5505–5510. doi:10.1073/pnas.95.10.5505

Townsend, J. A., Wright, D. A., Winfrey, R. J., Fu, F., Maeder, M. L., Joung, J. K., et al. (2009). High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* 459, 442–445. doi:10.1038/nature07845

Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., and Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* 11, 636–646. doi:10.1038/nrg2842

Van Vu, T., Das, S., Hensel, G., and Kim, J.-Y. (2022). Genome editing and beyond: What does it mean for the future of plant breeding? *Planta* 255, 130. doi:10.1007/s00425-022-03906-2

Vergunst, A. C., and Hooykaas, P. J. J. (1998). Cre/lox-mediated site-specific integration of Agrobacterium T-DNA in *Arabidopsis thaliana* by transient expression of cre. *Plant Mol. Biol.* 38, 393–406. doi:10.1023/A:1006024500008

Vijaya Bhanu, N., Khanuja, S. P. S., and Lodha, M. L. (1994). Integration ofhup genes into the genome of chickpea-rhizobium through site-specific recombination. *J. Plant Biochem. Biotechnol.* 3, 19–24. doi:10.1007/BF03321943

Voytas, D. F., and Gao, C. (2014). Precision genome engineering and agriculture: Opportunities and regulatory challenges. *PLoS Biol.* 12, e1001877. doi:10.1371/journal.pbio.1001877

Voytas, D. F. (2013). Plant genome engineering with sequence-specific nucleases. *Annu. Rev. Plant Biol.* 64, 327–350. doi:10.1146/annurev-arplant-042811-105552

Wang, F., Wang, C., Liu, P., Lei, C., Hao, W., Gao, Y., et al. (2016). Enhanced rice blast resistance by CRISPR/Cas9-targeted mutagenesis of the ERF transcription factor gene OsERF922. *PLoS One* 11, e0154027. doi:10.1371/journal.pone.0154027

Wang, L., Chen, L., Li, R., Zhao, R., Yang, M., Sheng, J., et al. (2017). Reduced drought tolerance by CRISPR/Cas9-mediated SlMAPK3 mutagenesis in tomato plants. *J. Agric. Food Chem.* 65, 8674–8682. doi:10.1021/acs.jafc.7b02745

Wang, L., Kaya, H. B., Zhang, N., Rai, R., Willmann, M. R., Carpenter, S. C. D., et al. (2021). Spelling changes and fluorescent tagging with prime editing vectors for plants. *Front. Genome Ed.* 3, 617553. doi:10.3389/fgeed.2021.617553

Wang, W., Vinocur, B., and Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: Towards genetic engineering for stress tolerance. *Planta* 218, 1–14. doi:10.1007/s00425-003-1105-5

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* 32, 947–951. doi:10.1038/nbt.2969

Wang, Y., Yau, Y.-Y., Perkins-Balding, D., and Thomson, J. G. (2011). Recombinase technology: Applications and possibilities. *Plant Cell. Rep.* 30, 267–285. doi:10.1007/s00299-010-0938-1

Wani, S. H., Kumar, V., Shriram, V., and Sah, S. K. (2016). Phytohormones and their metabolic engineering for abiotic stress tolerance in crop plants. *Crop J.* 4, 162–176. doi:10.1016/j.cj.2016.01.010

Whiteson, K. L., and Rice, P. A. (2008). Binding and catalytic contributions to site recognition by Flp recombinase. *J. Biol. Chem.* 283, 11414–11423. doi:10.1074/jbc.M800106200

Xie, K., and Yang, Y. (2013). RNA-guided genome editing in plants using a CRISPR-Cas system. *Mol. Plant* 6 (6), 1975–1983. doi:10.1093/mp/sst119

Xing, S., Chen, K., Zhu, H., Zhang, R., Zhang, H., Li, B., et al. (2020). Fine-tuning sugar content in strawberry. *Genome Biol.* 21, 230. doi:10.1186/s13059-020-02146-5

Xu, R., Li, J., Liu, X., Shan, T., Qin, R., and Wei, P. (2020). Development of plant prime-editing systems for precise genome editing. *Plant Commun.* 1, 100043. doi:10.1016/j.xplc.2020.100043

Xu, W., Zhang, C., Yang, Y., Zhao, S., Kang, G., He, X., et al. (2020). Versatile nucleotides substitution in plant using an improved prime editing system. *Mol. Plant* 13, 675–678. doi:10.1016/j.molp.2020.03.012

Yadav, D., Malviya, N., Nasim, J., and Kumar, R. (2016). Bioinformatics intervention in elucidating structural and functional attributes of plant specific transcription factors. *Res. J. Biotechnol.* 11 (7), 83–96.

Yadav, R., Jain, V., Hegde, V., Yadav, N., and Kumar, R. (2019). Bio-physico-chemical response of drought tolerant chickpeas to nickel. *Legum. Res.* 43 (3), 345–352. doi:10.18805/LR-4179

Yadav, R., Mehrotra, M., Singh, A. K., Niranjan, A., Singh, R., Sanyal, I., et al. (2017). Improvement inAgrobacterium-mediated transformation of chickpea (Cicer arietinum L.) by the inhibi-tion of polyphenolics released during wounding of cotyledonary node explants. *Protoplasma* 254, 253–269. doi:10.1007/s00709-015-0940-0

Yin, K., and Qiu, J.-L. (2019). Genome editing for plant disease resistance: Applications and perspectives. *Phil. Trans. R. Soc. B* 374, 20180322. doi:10.1098/rstb.2018.0322

Yuan, F., Chen, M., Leng, B. Y., and Wang, B. S. (2013). An efficient autofluorescence method for screening Limonium bicolor mutants for abnormal salt gland density and salt secretion. *S. Afr. J. Bot.* 88, 110–117. doi:10.1016/j.sajb.2013.06.007

Yuan, F., Leng, B., and Wang, B. (2016). Progress in studying salt secretion from the salt glands in recretohalophytes: How do plants secrete salt? *Front. Plant Sci.* 7, 977. doi:10.3389/fpls.2016.00977

Zeevi, V., Liang, Z., Arieli, U., and Tzfira, T. (2012). Zinc finger nuclease and homing endonuclease-mediated assembly of multigene plant transformation vectors. *Plant Physiol.* 158, 132–144. doi:10.1104/pp.111.184374

Zhang, F., Maeder, M. L., Unger-Wallace, E., Hoshaw, J. P., Reyon, D., Christian, M., et al. (2010). High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12028–12033. doi:10.1073/pnas.0914991107

Zhang, W., Subbarao, S., Addae, P., Shen, A., Armstrong, C., Peschke, V., et al. (2003). Cre/lox-mediated marker gene excision in transgenic maize (Zea mays L.) plants. *Theor. Appl. Genet.* 107, 1157–1168. doi:10.1007/s00122-003-1368-z

Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S., and Yang, S.-H. (2015). Off-target effects in CRISPR/Cas9-mediated genome engineering. *Mol. Ther. Nucleic Acids* 4, e264. doi:10.1038/mtna.2015.37

Zhang, Y., Massel, K., Godwin, I. D., and Gao, C. (2018). Applications and potential of genome editing in crop improvement. *Genome Biol.* 19, 210. doi:10.1186/s13059-018-1586-y

Zhao, Y., Zhang, C., Liu, W., Gao, W., Liu, C., Song, G., et al. (2016). An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. *Sci. Rep.* 6, 23890. doi:10.1038/srep23890

Zhu, C., Bortesi, L., Baysal, C., Twyman, R. M., Fischer, R., Capell, T., et al. (2017). Characteristics of genome editing mutations in cereal crops. *Trends Plant Sci.* 22, 38–52. doi:10.1016/j.tplants.2016.08.009

Zhu, T., Mettenburg, K., Peterson, D. J., Tagliani, L., and Baszczynski, C. L. (2000). Engineering herbicide-resistant maize using chimeric RNA/DNA oligonucleotides. *Nat. Biotechnol.* 1 18, 555–558. doi:10.1038/75435

Zhu, T., Peterson, D. J., Tagliani, L., St. Clair, G., Baszczynski, C. L., and Bowen, B. (1999). Targeted manipulation of maize genes *in vivo* using chimeric RNA/DNA oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 96, 8768–8773. doi:10.1073/pnas.96.15.8768

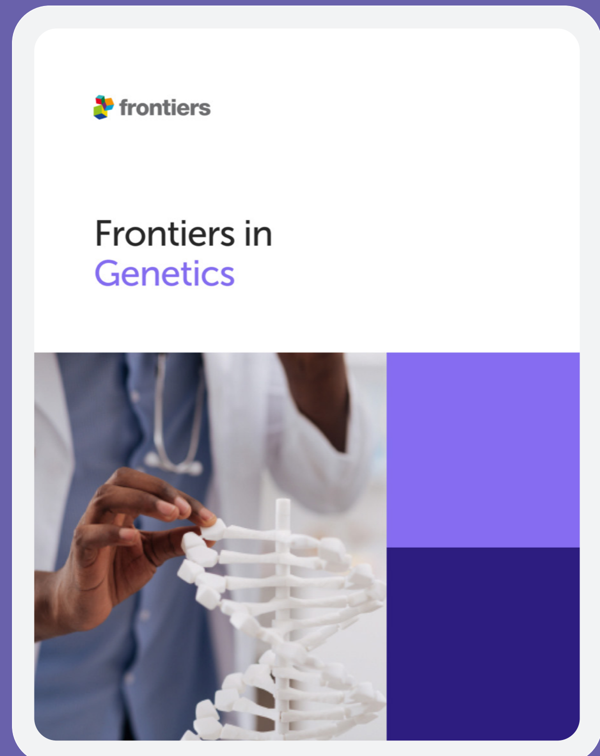
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact