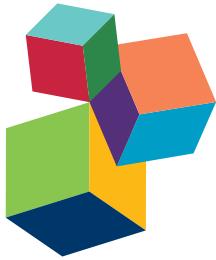


NATURAL DIVERSITY IN THE NEW MILLENNIUM

EDITED BY: Joanna M. Cross, Chiarina Darrah, Nnadozie Oraguzie,
Nourollah Ahmadi and Aleksandra Skirycz

PUBLISHED IN: *Frontiers in Plant Science*



Frontiers Copyright Statement

© Copyright 2007-2016 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88919-952-5

DOI 10.3389/978-2-88919-952-5

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: researchtopics@frontiersin.org

NATURAL DIVERSITY IN THE NEW MILLENNIUM

Topic Editors:

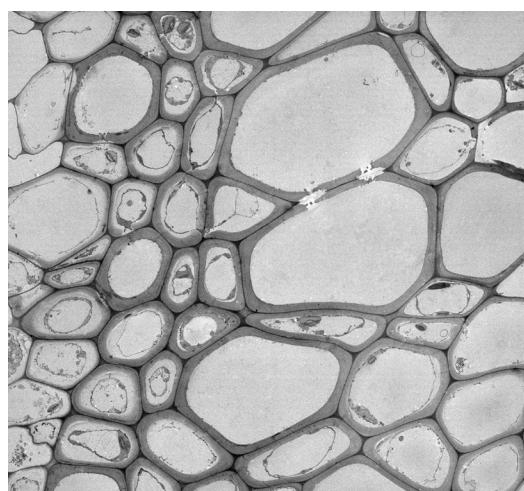
Joanna M. Cross, Inonu University, Turkey

Chiarina Darrah, Eunomia Research and Consulting, UK

Nnadozie Oraguzie, Washington State University, Prosser, USA

Nourollah Ahmadi, Centre de coopération internationale en recherche agronomique pour le développement, France

Aleksandra Skirycz, Instituto Tecnológico Vale Desenvolvimento Sustentável/Vale Institute of Technology Sustainable Development, Belem, Brazil



Electron microscopy of a transversal cut of the stem of an *Arabidopsis Columbia* plant.

Image by Joanna M. Cross

Natural diversity has been extensively used to understand plant biology and improve crops. However, studies were commonly based on visual phenotypes or on a few measurable parameters. Nowadays, a large number of parameters can be measured thanks to next generation sequencing, metabolomics, proteomics, and transcriptomics thus providing an unprecedented resolution in the detection of natural diversity. This enhanced resolution offers new possibilities in terms of understanding plant biology. Technology advances also contribute to a better assessment of the biodiversity loss currently taking place. Hence, the topic presents an overview on efforts for maintaining biological diversity in crops, on possibilities offered by recent technologies in the assessment of natural variation, and ends with examples of the diversity found even at the cellular level.

Citation: Cross, J. M., Darrah, C., Oraguzie, N., Ahmadi, N., Skirycz, A., eds. (2016). *Natural Diversity in the New Millennium*. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-952-5

Table of Contents

05 Editorial: Natural diversity in the new millennium
Joanna M. Cross, Chiarina Darrah, Nnadozie Oraguzie, Nourollah Ahmadi and Aleksandra Skirycz

1) Natural diversity in a populated world

07 Canga biodiversity, a matter of mining
Aleksandra Skirycz, Alexandre Castilho, Cristian Chaparro, Nelson Carvalho, George Tzotzos and Jose O. Siqueira

16 Oil palm natural diversity and the potential for yield improvement
Edson Barcelos, Sara de Almeida Rios, Raimundo N. V. Cunha, Ricardo Lopes, Sérgio Y. Motoike, Elena Babiychuk, Aleksandra Skirycz and Sergei Kushnir

32 Genetic diversity in tef [*Eragrostis tef* (Zucc.) Trotter]
Kebebew Assefa, Gina Cannarozzi, Dejene Girma, Rizqah Kamies, Solomon Chanyalew, Sonia Plaza-Wüthrich, Regula Blösch, Abiel Rindisbacher, Suhail Rafudeen and Zerihun Tadele

45 Genetic diversity and genomic resources available for the small millet crops to accelerate a New Green Revolution
Travis L. Goron and Manish N. Raizada

2) Natural diversity assessment with modern technologies

63 Comparative analysis of fruit aroma patterns in the domesticated wild strawberries "Profumata di Tortona" (*F. moschata*) and "Regina delle Valli" (*F. vesca*)
Alfredo S. Negri, Domenico Allegra, Laura Simoni, Fabio Rusconi, Chiara Tonelli, Luca Espen and Massimo Galbiati

76 Extensive sequence variation in rice blast resistance gene Pi54 makes it broad spectrum in nature
Shallu Thakur, Pankaj K. Singh, Alok Das, R. Rathour, M. Variar, S. K. Prashanthi, A. K. Singh, U. D. Singh, Duni Chand, N. K. Singh and Tilak R. Sharma

90 Analysis of genetic variation and diversity of Rice stripe virus populations through high-throughput sequencing
Lingzhe Huang, Zefeng Li, Jianxiang Wu, Yi Xu, Xiuling Yang, Longjiang Fan, Rongxiang Fang and Xueping Zhou

97 A Genome-wide Combinatorial Strategy Dissects Complex Genetic Architecture of Seed Coat Color in Chickpea
Deepak Bajaj, Shouvik Das, Hari D. Upadhyaya, Rajeev Ranjan, Saurabh Badoni, Vinod Kumar, Shailesh Tripathi, C. L. Laxmipathi Gowda, Shivali Sharma, Sube Singh, Akhilesh K. Tyagi and Swarup K. Parida

3) Natural diversity at the molecular level

115 *Mitogen-activated protein kinase cascades in Vitis vinifera*

Birsen Çakır and Ozan Kılıçkaya

133 *Biodiversity of genes encoding anti-microbial traits within plant associated microbes*

Walaa K. Mousa and Manish N. Raizada

158 *Target or barrier? The cell wall of early- and later-diverging plants vs cadmium toxicity: differences in the response mechanisms*

Luigi Parrotta, Gea Guerriero, Kjell Sergeant, Giampiero Cai and Jean-Francois Hausman



Editorial: Natural diversity in the new millennium

Joanna M. Cross^{1*}, Chiarina Darrah², Nnadozie Oraguzie³, Nourollah Ahmadi⁴ and Aleksandra Skirycz⁵

¹ Horticulture Department, Inonu University, Malatya, Turkey, ² Eunomia Research and Consulting, Bristol, UK, ³ Irrigated Agriculture Research and Extension Center (IAREC), Washington State University, Prosser, WA, USA, ⁴ Centre de coopération internationale en recherche agronomique pour le développement, Montpellier, France, ⁵ Instituto Tecnológico Vale Desenvolvimento Sustentável/Vale Institute of Technology Sustainable Development, Belém, Brazil

Keywords: natural diversity, crop improvement, next generation sequencing, biodiversity, plant biotechnology

Natural diversity is a recurrent theme in everyday life and in research. Thus, the objective of this topic was to highlight progress or novelties that occurred in this new millennium. The topic includes many different fields as can be seen by the variety of the articles. Indeed, subjects covered include variations in strawberry aroma, diversity of protein families, and peculiar ecosystems. However, several themes connect this apparently unrelated set of articles.

Unfortunately, one cannot ignore the issue of biodiversity and loss thereof. The United Nations Climate Change Conference was held in Copenhagen in 2009 to discuss ways of mediating environmental changes. No simple agreement was signed by all participating countries. That, combined with press articles has left a negative feeling. Thus, the general opinion is that the world is divided into two, the Western group environmentally friendly, and the rest of the world environmentally unfriendly. The articles reveal a more complex picture as many countries are trying to combine sustainable industries with economic realities. For instance, the Brazilian government has issued laws to safeguard the peculiar Canga ecosystems while allowing mining (Skirycz et al., 2014). Efforts are underway to produce improved oil palms to spare land (Barcelos et al., 2015). Moreover, many ecological studies can be found in the literature which highlights the international concern for loss of biodiversity. Unfortunately, biodiversity loss also applies to crops. Many landraces are abandoned as farmers adopt a few high yielding varieties. This results in a significant loss of genetic variability. Thus, more diversity was found in a disease resistance gene in rice landraces than in common varieties (Thakur et al., 2015). Once again, the articles illustrate an emphasis on collecting cultivars even for orphan crops. For instance, the Ethiopian Institute of Biodiversity houses 5000 accessions of tef, a major staple crop for the country (Assefa et al., 2015). The number represents a four-fold increase over the last 20 years. Likewise, significant resources exist for millet (Goron and Raizada, 2015).

OPEN ACCESS

Edited and reviewed by:

Richard A. Jorgensen,
University of Arizona, USA

***Correspondence:**

Joanna M. Cross
joanna.cross@inonu.edu.tr

Specialty section:

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

Received: 20 September 2015

Accepted: 09 October 2015

Published: 29 October 2015

Citation:

Cross JM, Darrah C, Oraguzie N, Ahmadi N and Skirycz A (2015) Editorial: Natural diversity in the new millennium. *Front. Plant Sci.* 6:897.
doi: 10.3389/fpls.2015.00897

Assessment of biodiversity requires a combination of phenotypic and molecular techniques. Technological improvements have been significant in all areas. Hence, metabolomics can now detect many complex molecules such as aroma compounds. As a result, factors determining the taste of strawberry (Negri et al., 2015) and other fruits can be elucidated. However, the most spectacular change comes from the decrease in cost and enhanced speed of Next Generation Sequencing. Fifteen years ago, international consortiums sequenced a few model species and major crops. Now many universities are acquiring their own sequencing equipment. Most plants covered in the articles harbor some sequencing resources either as transcriptomics or as an annotated genome. This opens enormous possibilities for identifying new genetic variations, assessing the variability of isozymes, and associating a given phenotype with regions in the genome. Thus, sequencing reveals the enormous evolutionary potential of viruses (Huang et al., 2015). In addition, extensive diversity was found in the MAP, MAPP, MAPPP kinase families in grapevine (Çakir and Kılıçkaya, 2015).

Finally, these articles illustrate the spectacular biological diversity of plants. They seem to survive in any environment, by any means. Hence, a seemingly fixed structure such as the cell wall shows remarkable flexibility in response to environmental, physiological, and genetic cues (Parrotta et al., 2015). Moreover, the richness of plant metabolism is further enhanced by symbiotic relations with microorganisms (Mousa and Raizada, 2015). However, while well described, this diversity harbors many

mysteries as to biological functions, evolutionary adaptations, and physiological mechanisms. With the enhanced technologies, let's hope we can understand and safeguard our beautiful world.

AUTHOR CONTRIBUTIONS

JC, CD, and AS contacted participants; JC and AS edited manuscripts; NA and NO partly edited articles.

REFERENCES

Assefa, K., Cannarozzi, G., Girma, D., Kamies, R., Chanyalew, S., Plaza-Wüthrich, S., et al. (2015). Genetic diversity in tef [Eragrostis tef (Zucc.) Trotter]. *Front. Plant Sci.* 6:177. doi: 10.3389/fpls.2015.00177

Barcelos, E., De Almeida Rios, S., Cunha, R. N. V., Lopes, R., Motoike, S. Y., Babiychuk, E., et al. (2015). Oil palm natural diversity and the potential for yield improvement. *Front. Plant Sci.* 6:190. doi: 10.3389/fpls.2015.00190

Çakir, B., and Kılıçkaya, O. (2015). Mitogen-activated protein kinase cascades in *Vitis vinifera*. *Front. Plant Sci.* 6:556. doi: 10.3389/fpls.2015.00556

Goron, T. L., and Raizada, M. (2015). Genetic diversity and genomic resources available for the small millet crops to accelerate a New Green Revolution. *Front. Plant Sci.* 6:157. doi: 10.3389/fpls.2015.00157

Huang, L., Li, Z., Wu, J., Xu, Y., Yang, X., Fan, L., et al. (2015). Analysis of genetic variation and diversity of rice stripe virus populations through high-throughput sequencing. *Front. Plant Sci.* 6:176. doi: 10.3389/fpls.2015.00176

Mousa, W. K., and Raizada, M. N. (2015). Biodiversity of genes encoding antimicrobial traits within plant associated microbes. *Front. Plant Sci.* 6:231. doi: 10.3389/fpls.2015.00231

Negri, A. S., Allegra, D., Simoni, L., Rusconi, F., Tonelli, C., Espen, L., et al. (2015). Comparative analysis of fruit aroma patterns in the domesticated wild strawberries "Profumata di Tortona" (*F. moschata*) and "Regina delle Valli" (*F. vesca*). *Front. Plant Sci.* 6:56. doi: 10.3389/fpls.2015.00056

Parrotta, L., Guerrero, G., Sergeant, K., Cai, G., and Hausman, J.-F. (2015). Target or barrier? The cell wall of early- and later- diverging plants vs cadmium toxicity: differences in the response mechanisms. *Front. Plant Sci.* 6:133. doi: 10.3389/fpls.2015.00133

Skirycz, A., Castilho, A., Chaparro, C., Carvalho, N., Tzotzos, G., and Siqueira, J. O. (2014). Canga biodiversity, a matter of mining. *Front. Plant Sci.* 5:653. doi: 10.3389/fpls.2014.00653

Thakur, S., Singh, P. K., Das, A., Rathour, R., Variar, M., Prashanthi, S. K., et al. (2015). Extensive sequence variation in rice blast resistance gene Pi54 makes it broad spectrum in nature. *Front. Plant Sci.* 6:345. doi: 10.3389/fpls.2015.00345

Conflict of Interest Statement: 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.

Copyright © 2015 Cross, Darrah, Oraguzie, Ahmadi and Skirycz. 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) or licensor 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.



Canga biodiversity, a matter of mining

Aleksandra Skirycz¹, Alexandre Castilho², Cristian Chaparro¹, Nelson Carvalho¹, George Tzotzos¹ and Jose O. Siqueira^{1*}

¹ Department of Sustainable Development, Vale Institute of Technology, Belém, Brazil

² Vale S.A., AP Supervisao PCM, Carajas, Brazil

Edited by:

Elena R. Alvarez-Buylla, Universidad Nacional Autónoma de Mexico, Mexico

Reviewed by:

Luis Enrique Eguiarte, Universidad Nacional Autónoma de México, Mexico

Alma Pineyro-Nelson, University of California Berkeley, USA

***Correspondence:**

Jose O. Siqueira, Department of Sustainable Development, Vale Institute of Technology, Rua Boaventura da Silva, 955 (Nazaré), 66055-090 Belém, PA, Brazil
e-mail: jose.oswaldo.siqueira@itv.org

Brazilian name canga refers to the ecosystems associated with superficial iron crusts typical for the Brazilian state of Minas Gerais (MG) and some parts of Amazon (Flona de Carajas). Iron stone is associated with mountain plateaux and so, in addition to high metal concentrations (particularly iron and manganese), canga ecosystems, as other rock outcrops, are characterized by isolation and environmental harshness. Canga inselbergs, all together, occupy no more than 200 km² of area spread over thousands of km² of the Iron Quadrangle (MG) and the Flona de Carajas, resulting in considerable beta biodiversity. Moreover, the presence of different microhabitats within the iron crust is associated with high alpha biodiversity. Hundreds of angiosperm species have been reported so far across remote canga inselbergs and different micro-habitats. Among these are endemics such as the cactus *Arthrocereus glaziovii* and the medicinal plant *Pilocarpus microphyllus*. Canga is also home to iron and manganese metallophytes; species that evolved to tolerate high metal concentrations. These are particularly interesting to study metal homeostasis as both iron and manganese are essential plant micro-elements. Besides being models for metal metabolism, metallophytes can be used for bio-remediation of metal contaminated sites, and as such are considered among priority species for canga restoration. “Biodiversity mining” is not the only mining business attracted to canga. Open cast iron mining generates as much as 5–6% of Brazilian gross domestic product and dialog between mining companies, government, society, and ecologists, enforced by legal regulation, is ongoing to find compromise for canga protection, and where mining is unavoidable for ecosystem restoration. Environmental factors that shaped canga vegetation, canga biodiversity, physiological mechanisms to play a role, and ways to protect and restore canga will be reviewed.

Keywords: canga, iron, ecosystem, endemism, restoration

CANGA AND SIMILAR ECOSYSTEMS WORLD-WIDE

In Brazil term canga refers to the ecosystem associated with superficial iron crust and present in the states of Minas Gerais (MG; Quadrilatero Ferrifero) and Para (Serra de Carajas; **Figure 1A; Table 1**). Literally translated MG stands for “general mines” reflecting strong ties to the mining industry due to its rich mineral deposits. The iron quadrangle (IQ) covers approximately 7200 km² with the superficial iron crusts distributed on the tops of mountains at altitudes ranging from 1000 to 2000 m above the sea level. The total area covered by canga in the IQ is relatively small, approximately 100 km² (Dorr, 1964), and dispersed on mountain inselbergs (600–700 m above the sea level). Similarly, in the Amazon state of Para (**Figure 1A**) canga occupies only 2 or 3% of natural clearings on iron rocks of the Carajas mountains, approximately 90 km², with the surrounding area covered by dense tropical forest (Nunes, 2009).

Canga-like areas, characterized by the presence of ironstone, ferricrete soils, are not restricted to Brazil; however, in other parts of the world they would be recognized under a different name. Canga-like habitats, referred to as banded iron formations (BIFs) have been reported in South-Western Australia (Gibson et al., 2010, 2012; **Table 1**). Similarly to canga,

BIFs form isolated inselbergs (some 700 km apart) on tops of mountains and are characterized by high local and regional diversity. As canga they are endangered by the mining industry and their restoration provides a major challenge (Gibson et al., 2010).

ENVIRONMENTAL FORCES THAT SHAPED CANGA BIODIVERSITY

The canga ecosystem has been shaped by 1000s of years long evolution process that gave rise to communities thriving in unique and severe environmental conditions. As is the case with other mountain top outcrops, the canga biome had to adapt to high ultraviolet (UV) exposure, high daily temperatures, rapid water loss, strong winds and poorly developed soil cover (Jacobi et al., 2007). High transpiration and low soil retention capacity challenges plants during periods of drought, even during the wet season. In addition to canga soils being shallow, acid (pH~4.0, in comparison to pH~7.0 of for the optimal agricultural soils) and nutritionally poor, they also contain toxic levels of aluminum and heavy metals (**Table 1**). More specifically canga substrates contain little phosphorous (P), magnesium (Mg), and calcium (Ca) but plenty of iron (Fe) and manganese

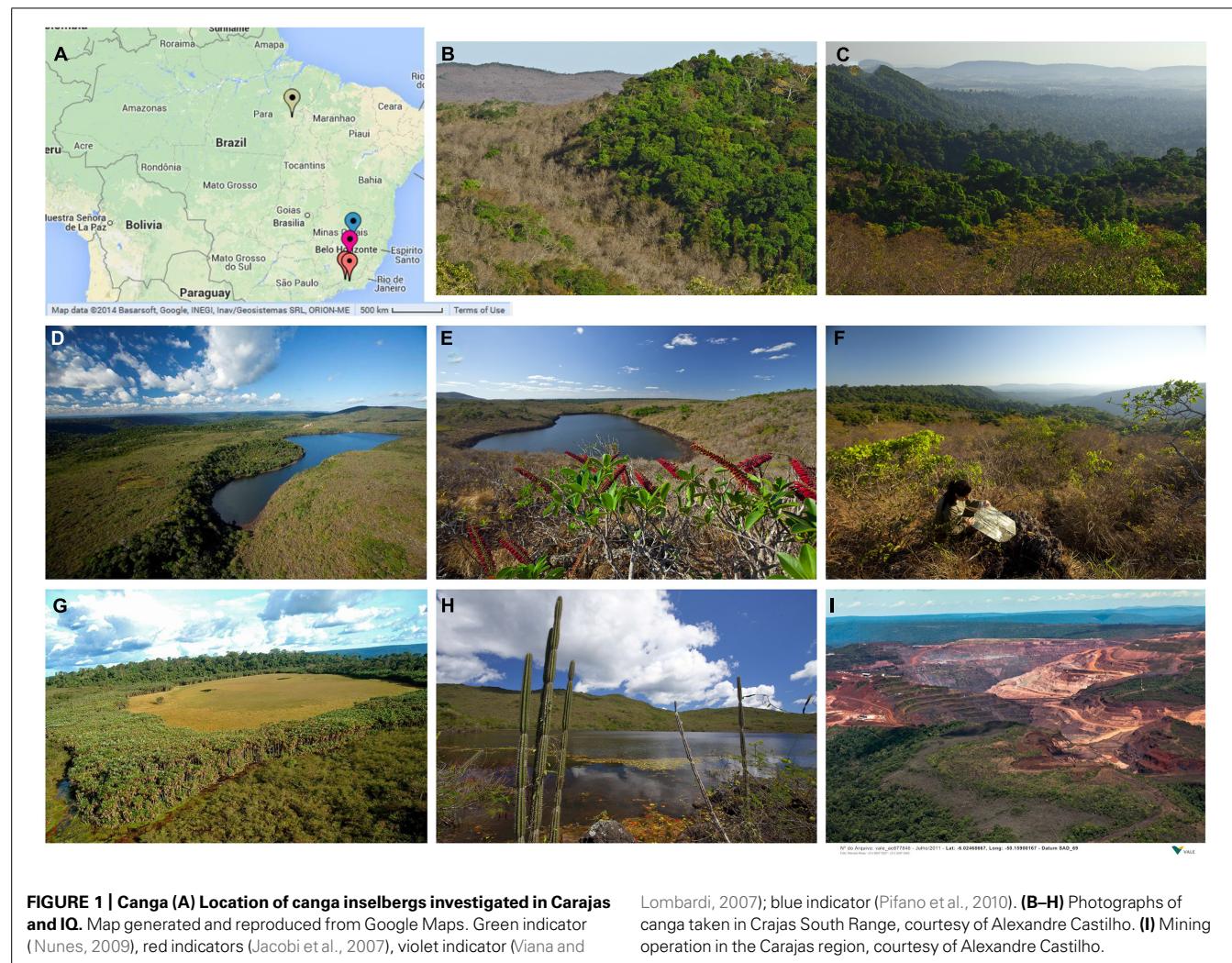


FIGURE 1 | Canga (A) Location of canga inselbergs investigated in Carajas and IQ. Map generated and reproduced from Google Maps. Green indicator (Nunes, 2009), red indicators (Jacobi et al., 2007), violet indicator (Viana and

Lombardi, 2007); blue indicator (Pifano et al., 2010). (B–H) Photographs of canga taken in Crajas South Range, courtesy of Alexandre Castilho. (I) Mining operation in the Carajas region, courtesy of Alexandre Castilho.

(Mn; Vincent and Meguro, 2008; Oliveira et al., 2009; Silva, 2013). Copper (Cu), nickel (Ni), zinc (Zn), chromium (Cr), and lead (Pb) accumulation has been also reported in plant samples harvested from the canga outcrop (Porto and da Silva, 1989; Silva, 1992). Importantly metal toxicity is aggravated by low pH, which increases metal mobility and so accessibility for plant uptake, e.g., under low pH iron is released from otherwise insoluble ferric oxides (Morrissey and Guerinot, 2009).

Adding to the severe abiotic conditions, the floral composition of canga is also a result of mineral and topographical heterogeneity giving rise to distinct microhabitats within canga (e.g., Jacobi et al., 2007; Viana and Lombardi, 2007). Iron-rich substrates may be totally fragmented or form a thick, solid crust. Cliffs, caves, grassland, rock fields, cracks in rock crevices, depressions, temporary or permanent ponds, temporally flooded areas all are part of canga. Finally canga biodiversity is affected by the surrounding ecosystems (Jacobi and Carmo, 2008a); in Carajas canga constitutes an enclave within an Amazonian forest whilst in IQ canga is located in the transition zone between the Cerrado and Atlantic Forest.

To summarize, harsh environmental conditions have supported evolution of hardy and often unique plants that can thrive in dry, metal rich but nutrient poor soil. High biodiversity is a result of canga being not one but rather a conglomerate of distinct microhabitats. It is dispersed along mountain tops rather than being a continuous habitat, and high biodiversity of the surrounding ecosystems.

PLANT BIODIVERSITY PORTFOLIO

Canga is home to 100s of plant species (Jacobi and Carmo, 2011). Close to 500 have been reported so far, sampled from only a small percentage of the total area occupied by canga. Three taxonomic studies published for the IQ (Jacobi et al., 2007; Viana and Lombardi, 2007; Pifano et al., 2010) and one for the Carajas Mountains (Nunes, 2009) will be referred to here.

Looking at plant species several characteristics emerge. Firstly, canga floristic composition significantly differs from the neighboring ecosystems. When compared with the Atlantic and Amazonian forest (Nunes, 2009; Pifano et al., 2010) the difference is obvious at first sight; canga landscape resembling savanna covered by shrubs and herbaceous plants with few trees and associated crawlers

Table 1 | Comparison of iron based ecosystems in Brasil (canga) and Australia (bounded iron formations).

	Canga in IQ	Canga in Carajas	BIFs
Location	Minais Gerais (MG), Brazil	Para, Brazil (Nunes, 2009)	South-Western Australia
Longitude	19.30, 20.31 S (Jacobi et al., 2007)	5.35, 5.57 S (Nunes, 2009)	25.30, 30.00 S (Gibson et al., 2012)
Latitude	43.00-44.30 W (Jacobi et al., 2007)	50.01, 51.04 W (Nunes, 2009)	116.00, 122.15 E (Gibson et al., 2012)
Approximate area	100km ² (Jacobi et al., 2007)	90 km ² (Nunes, 2009)	—
Spread across area/range	Area of 7200km ² (Jacobi et al., 2007)	Area of 4000 km ² (Nunes, 2009)	Across 700 km (Gibson et al., 2012)
Altitude	900–2000 m (Jacobi et al., 2007)	Max 620–660 m	Max 700 m (Gibson et al., 2012)
Climate	Tropical sub-humid (Jacobi et al., 2007)	Tropical humid (Nunes, 2009)	Arid (Gibson et al., 2012)
Minimum temperature	13–19°C	19°C (Nunes, 2009)	3.8–7.5°C (Gibson et al., 2012)
Maximum temperature	24–28°C	31°C (Nunes, 2009)	34–39°C (Gibson et al., 2012)
Mean precipitation	1500–1900 mm (Jacobi et al., 2007)	1900 mm (Nunes, 2009)	186–329 mm (Gibson et al., 2012)
Rain seasonality	Yes (Jacobi et al., 2007)	Yes (Nunes, 2009)	Yes (Gibson et al., 2012)
Soil characteristics*	Messias et al. (2013)	Nunes (2009)	Markey and Dillon (2010)
pH	4.7–5.15 (Messias et al., 2013)	3.76–4.37 (Nunes, 2009)	4.22–5.97 (Markey and Dillon, 2010)
Mn (mg/kg)	10 times higher (Messias et al., 2013)	Two times higher (Nunes, 2009)	4–20 times higher (Markey and Dillon, 2010)
Fe (mg/kg)	20–40 times higher (Messias et al., 2013)	20–50 times higher (Nunes, 2009)	2–6 times higher (Markey and Dillon, 2010)
P(mg/kg)	10–20 times lower (Messias et al., 2013)	5–20 times lower (Nunes, 2009)	10 times lower (Markey and Dillon, 2010)
Surrounding ecosystem	Atlantic rain-forest, Cerrado, rock outcrops (Jacobi et al., 2007)	Amazonian rain-forest (Nunes, 2009)	Arid and Transition Rainfall Zones (Gibson et al., 2012)
Plant Biodiversity***	Jacobi et al. (2007), Viana and Lombardi (2007), Pifano et al. (2010)	Nunes (2009)	Gibson et al. (2012)
Number of plant species	353 (Jacobi and Carmo, 2011)	125 (Nunes, 2009)	906 (Gibson et al., 2012)
Dominant genus 1	Asteraceae (Jacobi and Carmo, 2011)	Fabaceae (Nunes, 2009)	Fabaceae (Gibson et al., 2012)
Dominant genus 2	Poaceae (Jacobi and Carmo, 2011)	Euphorbiaceae (Nunes, 2009)	Myrtaceae (Gibson et al., 2012)
Dominant genus 3	Orchidaceae (Jacobi and Carmo, 2011)	Myrtraceae (Nunes, 2009)	Poaceae (Gibson et al., 2012)
Dominant life form 1	Herbs (Jacobi and Carmo, 2011)	Shrubs (Nunes, 2009)	Shrubs (Gibson et al., 2012)
Dominant life form 2	Shrubs (Jacobi and Carmo, 2011)	Herbs (Nunes, 2009)	Herbs (Gibson et al., 2012)
Presence of endemics	Yes (Jacobi et al., 2007)	Yes (Nunes, 2009)	Yes (Gibson et al., 2012)

*Approximate differences in element composition when compared to standard soil (Reganold et al., 2010). **Dominance calculated based on number of species.

(**Figures 1B–H**). Thus not surprisingly, 80% of species identified in canga was also found to be canga exclusive in comparison with the Atlantic forest, swamp and riparian forest (Pifano et al., 2010). Less obviously, the canga floristic composition also differs from other rocky outcrops present in IQ. A comparison of the 16 most represented angiosperm families found in sandstone and canga rocky outcrops revealed significant differences. For instance, whereas Solanaceae and Verbenaceae species are absent in the sandstone inselbergs they are abundant in canga. The opposite being true for Eriocaulaceae and Xyridaceae angiosperms (Jacobi and Carmo, 2008b).

Secondly, canga is characterized by high beta diversity, which refers to the differences measured between distinct canga inselbergs (Jacobi et al., 2007). In their work Jacobi et al. (2007) investigated two sites approximately 32 km apart. Out of 235

fern and angiosperm species only 27% were common for both sites. Even less overlap is found between the three studies from IQ, cangas 100s of kilometers apart. Only two species were identified in all studies and 18% were shared between any of the two. And so not surprisingly, the plant composition in the Carajas mountains, 1000 km away, shares only seven species common with those found in the IQ.

Thirdly, canga is characterized by high alpha biodiversity, which refers to differences measured within a single canga inselberg (Jacobi et al., 2007). High alpha biodiversity is strongly related to the presence of distinct microhabitats within canga; differences in the amount and composition of soil, and moisture being considered the main factors shaping diverse plant communities (Vincent and Meguro, 2008; Nunes, 2009; Silva, 2013). A comprehensive study of Viana and Lombardi (2007) reported 358

plant species representing 70 angiosperm plant families, which is approximately 15% of all known angiosperm families worldwide. Researchers examined and described four distinctive canga microhabitats. The majority of the plant species were found on so-called “grassy fields,” which as the name implies, form homogenous, in appearance, sea of grasses and sages mixed with subshrubs and shrubs, and growing on a fragmented iron crust. In contrast “rocky fields” represent more an alpine like ecosystem characterized by the diversity of perennial and annual herbs (both monocots and dicots) growing in rock crevices formed on a solid iron crust. Noteworthy, endemic species: cactus *Arthrocereus glaziovii*, bromeliads *Consimilis dyckia* and *Vriesea minarum*, orchid *Oncidium gracile*, *Sinningia rupicola*, and legume trees *Mimosa calodendron* are all found in the open “grassy” or “rocky” fields (Jacobi and Carmo, 2011). In places with deeper substrate and more organic matter forest islands are established with small tree species and shrubs covered with crawlers and epiphytes. Last microhabitat represents an area disturbed by human activity and is mostly similar to the grassy fields albeit with only one fourth of the species richness. Remarkably more than 60% of the plant species (235 out of 358) were unique to the given microhabitat and only five were present in all of them. This kind of microhabitat specificity was also reported by Jacobi et al. (2007) and Nunes (2009). Examples of other microhabitats found in canga and described by Jacobi et al. (2007) include small permanent rock pools inhabited by unicellular and filamentous algae, ponds formed during the wet season in shallow depressions and covered with rare monocotyledonous plants from the *Eriocaulon* genus, entrances of small caves underneath iron crust providing shade and humidity for mosses and delicate herbs.

Finally, it was demonstrated that the exact floristic composition of canga could vary between dry and wet seasons (Nunes, 2009; Silva, 2013). However, canga is home to relatively few annual plants (Jacobi and Carmo, 2011), partial or complete loss of above-ground organs is a common adaptation to dry season (Jacobi et al., 2007).

To our knowledge and to date, canga diversity was estimated exclusively based on taxonomic identification, which may lead to both under- or over-estimation (Frankham, 2010), the first being more common. To illustrate the point, suffice to mention a recent work of Nistelberger et al. (2014). Analysis of genomic and mitochondrial regions of the millipede species from the Australian BIFs pointed to the long-term isolation of distinctive populations, inhabiting remote inselbergs, and so the need to place each population into a separate conservation unit.

When looking at individual species canga biodiversity may seem overwhelming. However, there is uniformity when looking at vegetation type and plant families. In general canga vegetation can be described as rocky, ephilitic (growing on stone), shrub dominated with large number of sedges, grasses, and orchids. Plant species are evenly distributed between shrubs, subshrubs, herbs, and epiphytes. Trees and crawlers are by far less diverse (Jacobi and Carmo, 2011). In terms of families the richest are Poaceae (grasses and sedges), Asteraceae (daisy like herbs), Fabaceae (legumes), Myrtaceae (subshrubs and shrubs), Melastomataceae (shrubs and small trees), and Orchidaceae (orchids;

Jacobi et al., 2007; Viana and Lombardi, 2007; Nunes, 2009; Pifano et al., 2010). It is, however, important to mention that these families are not only rich in canga but also in other rocky outcrops (Jacobi and Carmo, 2008b), and they encompass 35% of the total ~31000 angiosperm species reported in Brazil (Forzza, 2010). Area-wise ironstone outcrops are dominated by shrubs, mainly dicots. Monocots register a large proportion of herbaceous plants (grasses and sedges) and subshrubs. Dicot herbs are represented by numerous species of relatively small population sizes (Jacobi et al., 2007). This general characteristic of canga outcrop is also shared with plant communities found in Australian BIFs (Table 1); BIFs landscape being dominated by shrubs and herbaceous species from the Fabaceae, Myrtaceae, and Poaceae families (Table 1), generously represented in canga (Gibson et al., 2012).

Summarizing, canga ecosystems are characterized by high local and regional diversity. To date, 100s of plant species, some endemic, have been identified across diverse canga microhabitats and geographical locations. The total area of canga is estimated to be approximately 200 km², of which approximately 1–2% was studied, pointing the need of additional investigation. Genomics tools can play a crucial role for the accurate estimation of genetic diversity.

CANGA AS SOURCE OF NEW METAL HYPER-ACCUMULATING SPECIES

To cope with the harsh environmental conditions canga plants developed a number of physiological adaptations, of which metal tolerance is by far the most interesting. Together with Australian BIFs, cangas may constitute one of few ecosystems worldwide to “mine” for plants adapted to high iron. We are reviewing the importance of this in this section.

Metallophytes, that is species that evolved on metalliferous soils, are commonly divided into those that can exclude metals and those that can accumulate or even hyper-accumulate metals in their shoots (Baker, 1981; reviewed by Krämer, 2010 and Leitnermaier and Küpper, 2013). The first class represents majority of the species (>95%) that evolved avoidance mechanisms to block metal uptake into cells or mediate metal efflux and storage into root vacuoles, far from the photosynthetic cells. In contrast few percentages of plants would undergo extensive adaptation that would allow them to accumulate one or few metals in their shoots, where they act as repellent for herbivores and pathogens (Boyd, 2012). Metal transport from soil into plant shoot, complexation with organic compounds to reduce toxicity and sequestration inside vacuoles are the key processes considered for metal accumulation and tolerance (Krämer, 2010).

Growing on metal rich soils canga plant community is considered a unique source “to mine” for novel metal hyper-accumulating taxa. Of these manganese and iron hyper-accumulators stand out for a number of reasons. (1) In contrast to several 100 known nickel, cadmium, and zinc hyper-accumulators, not more than 20 manganese hyper-accumulating species were reported to date (Fernando et al., 2013). To our knowledge no iron accumulating taxa have been reported and studied. (2) Manganese and iron are essential plant micro-nutrients and their deficiency impacts plant yield (Marschner, 2012). But in high concentrations,

manganese and iron are toxic to plants (Lynch and St Clair, 2004). As the unraveling molecular mechanisms underlying plant iron and manganese homeostasis is considered key to address both deficiency and toxicity issues reported for the agricultural crops, hyper-accumulators can be considered a good tool. For instance, analogous to zinc hyper-accumulators (Krämer, 2010), canga metallophytes are expected to encompass metal sensitive populations, which grow in ironstone outcrops along canga but in metal free substrate. Metal tolerant and metal sensitive populations within a species provide a great resource for genetic and molecular mining. On the one hand, recombinant inbred lines derived from parents differing in the trait of interest are the most common way to discover underlying QTLs. On the other hand, in system biology “omics” based analysis, differences in transcripts, proteins, and/or metabolites measured between tolerant and sensitive plants are good suspects behind observed phenotypes. (3) Soils with toxic levels of manganese are associated with mineral deposits but also with iron ore mining (strip mining generates manganese rich dust) and processing (steel industry uses big amounts of manganese). High concentration of manganese is toxic to plants but also to animals; in humans manganese poisons the nervous system resulting in Parkinson like symptoms (Taba, 2013). Because hyper-accumulators have the capacity to remove manganese from the soil they are considered a useful resource for phyto-remediation of post-industrial and post-mining areas (Whiting et al., 2004). (4) Finally, as metallophytes, similarly to endemics, are unique to canga they are a priority for post-mining canga restoration (see below).

Despite the obvious interest, we found only two rather old studies reporting metal content of canga plants. Porto and da Silva (1989) and Silva (1992) examined 24 canga plant species, both from IQ and Carajas, for foliar concentration of Mn, Ni, Cr, Cu, Pb, and Fe. Twenty of the investigated specimens accumulated single or several metals in levels several times higher than “standard reference plant.” Although this is not enough to

classify them as a hyper-accumulators (van der Ent et al., 2013) it is a good start for further canga research. Most interestingly, seven of the examined species accumulated at least double the iron or manganese leaf concentration of standard reference plants (Figure 2). Two more species accumulated iron and one more manganese in roots but not in their leaves (Figure 2). Concluding, studies of Porto and da Silva (1989) and Silva (1992) demonstrate that canga plants have developed various strategies to cope with high iron and manganese concentrations, through exclusion, or through accumulation, providing justification for additional studies.

To summarize quoting Krämer (2010): “metal hyper-accumulators will be instrumental in the development of systems biology approaches toward an integrated understanding of plant metal homeostasis” and Erskine et al. (2013): “to prevent continued biodiversity loss and to benefit from the unique adaptive mechanisms that exclude, tolerate, and even accumulate toxic metals in mine site rehabilitation, metallophytes must be recognized as vital asset at developmental stage of a mining operation.” From this perspective identification of canga metallophytes appears an obvious direction to pursue.

OTHER PHYSIOLOGICAL ADAPTATION OF THE CANGA PLANTS

Besides metal tolerance and similar to the Australian BIFs (Table 1), canga plant community is known for its overall resilience to water scarcity (Jacobi et al., 2007). Rocky plant communities found in IQ are home to largest group of resurrection plants in Brazil, being able to tolerate almost complete desiccation by partial or complete loss of their above-ground organs during periods of severe drought. Less drastic adaptations include thick, waxy, imbricate leaves; guard cells sunken into pits; extensive and deep root systems; presence of water storing organs, e.g., pseudo-bulbs in orchids. Some plants, such as Clusa trees, are characterized by CAM photosynthesis, which allow them to keep guard cells shut

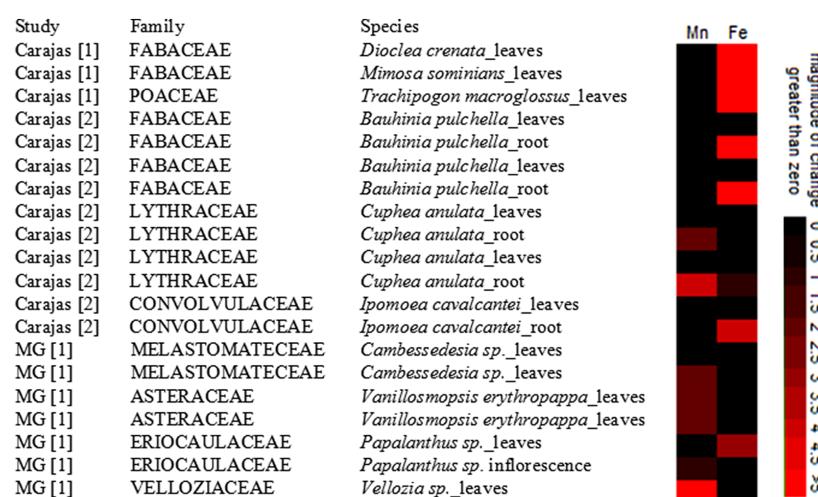


FIGURE 2 | Canga plant species accumulating at least double the standard concentration of manganese (200 ppm) and iron (300 ppm). From Porto and da Silva (1989) and Silva (1992). Two entries given to one species indicate populations collected in different area of occurrence.

during the day, by using night acquired CO₂ stored in the organic acid malate. Slow growth rates of the canga plant species may be seen as yet another adaptation to water shortage. In stressful environments plants accumulate rather than spend their resources (carbon and water) in expense of growth, to avoid starvation if the conditions get worse (Skirycz et al., 2011).

CANGA, METABOLITES, AND MEDICINAL PLANTS

To our knowledge, there is not an adequate molecular understanding of the physio-morphological traits that allow canga plants to thrive in their stressful environment. Of course parallels can be made from model plants but particulars need investigation. The potential of interesting findings can be demonstrated by available metabolite data. Many of the IQ plant species, present also in canga, were investigated for their metabolite profiles. These plants were those with reported medicinal properties and metabolite analysis was performed to investigate plausible agents responsible for the observed therapeutic affects (Table 2). Although these studies did not aim at understanding plant physiology, they revealed a richness of secondary metabolites associated with IQ ecosystem, with some of the compounds being well known for their involvement in plant stress responses. For example flavonols play a role in oxidative stress associated with drought and high UV (Skirycz et al., 2007) and past findings point to role of anthocyanin in metal complexion (Hale et al., 2001). Based on existing data, we believe that more comprehensive metabolite profiling of the canga plant species will shed light on the environmental adaptation of this unique rocky outcrop.

Touching upon medicinal plants, it is interesting to mention the small tree *Pilocarpus microphyllus* (popular name jaborandi) endemic to Brazil and found in Flona de Carajas, where its associated with forest clearings and poor soils such as those present in canga. Leaves of *Pilocarpus* are collected by local communities and sold to pharmaceutical companies as source of the FDA approved anti-glaucomé drug Timpilo (Costa, 2012). The main constituent of Timpilo is pilocarpine, an imidazole alkaloid inaccessible for cheap organic synthesis. To avoid irreversible depletion of the wild population, pilocarpine collection is regulated by the

environmental agencies. Because of its value for the local community jaborandi is considered one of the key plants for post-mining re-vegetation. Unfortunately, when cultivated *Pilocarpus* gives low levels of pilocarpine. It is thought that one of the reasons for this is the lack of environmental stressors (e.g., metals or low soil pH) that boost pilocarpine production in the wild. We believe that the molecular understanding of pilocarpine synthesis would shed light on jaborandi adaptation to disturbed environments. In turn, this may result in income generation for local communities and the pharmaceutical industry, as well as environmental benefits arising from more effective conservation and restoration practices of the mining industry.

HOW TO PROTECT AND RESTORE?

In 1992 during the Earth Summit in Rio de Janeiro, the Convention on Biological Diversity (CDS) was signed to help creating legislation for biodiversity protection. Mineral rich regions present a serious dilemma for the CDS (Gibson et al., 2010; Jacobi et al., 2011). On the one hand, they are home to unique ecosystems, often considered biodiversity hot spots. On the other hand the demand for metallic ores is growing rapidly and so is the number of mining permits, a trend that will continue in the future. Thus to sustain metalliferous ecosystems, while at the same time ensuring satisfactory returns from the mining activities compromise must be made. As long-term outcomes of any restoration efforts cannot be predicted (see below), it is indisputable that parts of canga must be protected; exactly how much is a subject of debate. In Brazil state protected reserves, which are exempted from mining rights, cover relatively small part of mineral regions. In MG 39 km² of canga lies within borders of Serra do Rola Moca State Park (Jacobi et al., 2008). In Carajas, iron ore is locked in two mountain ranges called Serra Sul and Serra Nord located at the territory of Natural Reserve Park, Flona de Carajas. Mining already takes place in Serra Nord (Figure 1I), and a new mining complex in Serra Sul, S11D, is planned to start producing in 2016. Importantly, parts of S11D will be exempted from mining, as a result of heavily negotiated agreements between environmental agencies and the mining company. For instance, of the 187

Table 2 | Examples of secondary metabolites found in plant species from Minas Gerais (MG) and identified in canga outcrop by Jacobi et al. (2007).

Metabolite class	Example of compounds	Plant species	Reference
Terpenoids	Thymol, myrcene, cymene, terpinene, parthenin, beta-bisabolene	<i>Lippia gracilis</i> , <i>Lantana camara</i> , <i>cabralea canjerana</i> , <i>Copaifera langsdorffii</i> , <i>Lychnophora pinaster</i> , <i>Eremanthus erythropappus</i>	Guzzo et al. (2008), Sarria et al. (2011), Souza et al. (2011), Abreu et al. (2013), de Melo et al. (2013), Ferraz et al. (2013), Pradeep et al. (2013), Silvério et al. (2013), Zimmermann-Franco et al. (2013)
Phenylpropanoids	Catechin,	<i>Myrsine umbellata</i> , <i>Marcetia taxifolia</i> ,	Ribeiro et al. (1997), Guzzo et al. (2008), Soberón et al. (2010), Leite et al. (2012), Abreu et al. (2013),
Flavonoids	methoxyflavanones, safrol	<i>Tripodanthus acutifolius</i> , <i>Trixis vauthieri</i> , <i>Lychnophora pinaster</i> , <i>Ocotea cf. pulchella</i>	Novaes et al. (2013), Leporatti et al. (2014)
Alkaloids		<i>Psychotria vellosiana</i>	Moraes et al. (2011)
Carotenoids	Lutein	<i>Myrsine umbellata</i>	Novaes et al. (2013)
Others	8-Methoxylapachanol, steroids, grifolin, grifolic acid, piperogalbin	<i>Sinningia allagophylla</i> , <i>Lychnophora pinaster</i> , <i>Peperomia gallioides</i>	Langfield et al. (2004), Guzzo et al. (2008), Abreu et al. (2013), Barbosa et al. (2013)

iron-stone caves 152 will be fully protected (Vale Sustainability Report, 2012). Moreover protection belt between mountain lakes, do Violão and do Amendoin, and the mining operation is planned to protect lakes and surrounding canga outcrop (Figures 1C,D).

In mining areas where protection is not possible ecosystem restoration is a legal requirement. In Brazil to obtain or renew mining permits, companies need to submit area restoration plans to be implemented upon mine closure. These include a plethora of steps that preceded the closure and sometimes even mining itself (see below). Moreover actions mitigating ecosystem destruction during mining operation need to be included. For example, as part of the compromise with the environmental agencies, a number of technological solutions have been implemented in the S11D Carajás complex, such as, truckless, conveyor belt operated delivery of the iron ore from the mining site to the processing plant outside Flona de Carajás reserve (Vale Sustainability Report, 2012, 2013). It is noteworthy that no such or similar solutions were required from mining companies, when exploitation started 30 years ago in Serra Nord. Similarly, in the past, it was legally acceptable to rehabilitate exhausted mining areas with plant cover from any available plant species (see Griffith and Toy, 2001 for the early re-vegetation strategies of iron mines in IQ). Nowadays the aim is to restore ecosystems relying exclusively on native taxa.

The dependence of canga on the unique biotic components mentioned earlier, makes restoration a difficult task. Iron ore open air mining is aggressive. Ironstone outcrops and associated biota are striped, so the iron ore deposits can be accessed with the subsequent excavation reaching even 300 m in depth. The landscape left at the end of the mining operation has an erased geographical topography, no previous terrain structure, biotic, and abiotic components left. Hydrology can be seriously disturbed due to changes in water table and erosion (Castro et al., 2011a,b). Despite these considerable difficulties, principal restoration steps do not differ when compared with recovery of other post-mining areas, and will be listed keeping canga in mind. (1) As mentioned previously, restoration goals are set before the mining begins and long before it ends. (2) Careful ecosystem reviews (e.g., Jacobi and Carmo, 2011) are used as restoration guidelines. Species information is collected together with habitat information (e.g., soil composition; for canga see, e.g., Nunes, 2009; Messias et al., 2013) and used for ecological modeling (reviewed by Memmott, 2009; Stouffer et al., 2012) to delineate and visualize relations between organisms (for canga seem, e.g., Araújo et al., 2006; Jacobi et al., 2008; Jacobi and Carmo, 2011). Based on the above, priority species and interactions are selected. For instance, detailed analysis of the angiosperm composition identified the following species as priority for canga recovery in IQ (1) grass *Andropogon ingratius* small shrub *Lychnospora pinaster*, and sedge *Bulbostylis fimbriata* as most abundant in terms of individual number and total area occupied (2) *Mimosa calodendron* as nitrogen fixing nursery taxa (3) *Vellozia* species for their metal tolerance and accumulation mechanisms (see previous section on metallophytes; Jacobi et al., 2008). Plants used by indigenous communities such as *Pilocarpus microphyllus* and endemic species such as *Arthrocereus glaziovii* have been also considered as a priority. (3) Mining areas are

prepared so as to minimize deforestation, e.g., as in the S11D complex by shifting all the ore processing and tailings outside Flona de Carajás territory. (4) Topsoil removal and storage is now being practiced, as it contains seed banks and associated micro-biomes, and was proven to be top priority for ecosystem restoration (Cooke and Johnson, 2002). For instance, work of Rezende et al. (2013) demonstrated successful growth of 14 out of 15 tested canga herbaceous, shrubby, and woody species, over the period of 42 months, using topsoil salvaged from iron ore mining sites in MG. Another published example is a pilot study done for a new iron-ore mining complex in IQ, called Minas Rio (Mello et al., 2014). Flora rescue operations from canga areas to be mined salvaged 45 plant species belonging to different families and life forms that were subsequently introduced together with 5 cm of top-soil to a different location. After a period of approximately 3 years almost 2000 individuals representing 38 plant species were recorded. Noteworthy, the majority of the species, such as the already mentioned sedge *Bulbostylis fimbriata*, were successfully reproduced within the experiment duration. Since 2010, a further 108 plant species have been rescued by the Minas Rio operation. (5) Landscape architecture to stabilize slopes and restore some of the original topography is also required to recreate functional ecosystem, as for example in case of Germano mine in MG (see below; Castro et al., 2011a) (6) Re-vegetation using available planting material. As described above, many canga species can be successfully rescued before mining begins and subsequently cultivated *ex situ* awaiting restoration. Seeds and plantings obtained from *ex situ* collections constitute the obvious starting material for re-vegetation. (7) During the early phases of restoration applying fertilizers and using chemical means to deal with invading plant species and a plethora of pests is often required. Introducing native legumes to fix nitrogen is not uncommon. (8) Monitoring criteria, which are essential to assess restoration progress, are usually decided at the early operation stages (Cooke and Johnson, 2002). Soil organic matter content and species richness are traditionally used, physiological parameters (Cooke and Suski, 2008), nowadays more and more accessible from the aerial photographs (e.g., ground cover, rate of biomass increase, chlorophyll content), are, however, increasingly more popular.

An often cited example of a successful restoration effort is the Jarrah forest in Western Australia (reviewed by Grant and Koch, 2007). Within 50 years of the mining operation closure the original ecosystem, with more than 700 native plant and 200 animal species, has been reproduced and nowadays is considered fully self-sustainable. It demonstrates that restoration is possible, where all involved parties commit to their obligations, main fear of the environmentalists. In Brazil, two examples of recently closed iron ore mines are Aguas Claras and Germano in MG (Castro et al., 2011a). Due to its location, area of Aguas Claras will be used to build a small town neighboring a new lake created in the depleted mining pit. In contrast, the Germano pit closure focuses on environmental rehabilitation; extensive landscape architecture being carried on at the moment. Closure time for the Minas Rio (MG) and S11D (Carajás), new projects that include canga restoration planning, is envisaged only in 30–40 years, respectively. As it is very long time before it will be possible to evaluate restoration success,

and as mentioned before, the need of canga protection remains indisputable.

SUMMARY, CANGA, AND MEANING OF BIODIVERSITY IN THE NEW MILLENNIUM

Although very small in size canga demonstrates biodiversity challenges of the ongoing Millennium. Due to its unique setting, canga significantly contributes to regional and global biodiversity. And as such is seen as a holy grail by conservationists and a treasure box by biologists. Laying on mineral deposits canga is, however, threatened by human activities, presenting an important dilemma between biodiversity and economical gains, ever more acute in the current millennium, with the human population and its prosperity rapidly growing. The difficult comprise between human economic activities and biodiversity is simply unavoidable. Protect and restore discussions are not easy and new concepts, such as environmental offsets, are emerging. Hopefully the current Millennium will bring new solutions to biodiversity preservation as it brings challenges. "The future is unknown but it can be invented."

REFERENCES

Abreu, V. G., Correa, G. M., Silva, T. M., Fontoura, H. S., Cara, D. C., Piló-Veloso, D., et al. (2013). Anti-inflammatory effects in muscle injury by transdermal application of gel with *Lychnophora pinaster* aerial parts using phonophoresis in rats. *BMC Complement Altern. Med.* 13:270. doi: 10.1186/1472-6882-13-270

Araújo, V. A., Antonini, Y., and Araújo, A. P. (2006). Diversity of bees and their floral resources at altitudinal areas in the Southern Espinhaço Range, Minas Gerais, Brazil. *Neotrop. Entomol.* 35, 30–40. doi: 10.1590/S1519-566X2006000100005

Baker, A. J. M. (1981). Accumulators and excluders – strategies in response of plant to heavy metals. *J. Plant Nutr.* 3, 643–654. doi: 10.1080/01904168109362867

Barbosa, F. L., Mori, L. S., Riva, D., Stefanello, M. É., and Zampronio, A. R. (2013). Antinociceptive and anti-inflammatory activities of the ethanolic extract, fractions and 8-methoxylapachol from *Sinningia allagophylla*. *Basic Clin. Pharmacol. Toxicol.* 113, 1–7. doi: 10.1111/bcpt.12051

Boyd, R. S. (2012). Plant defense using toxic inorganic ions: conceptual models of the defensive enhancement and joint effects hypotheses. *Plant Sci.* 195, 88–95. doi: 10.1016/j.plantsci.2012.06.012

Castro, M., Lima, H., and Flores, J. (2011a). Overview of mine closure in Minas Gerais, Brazil. *Rev. Esc. Minas Ouro Preto* 64, 205–211. doi: 10.1590/S0370-44672011000200012

Castro, P., Nalini, H., and Lima, H. (2011b). *Understanding Mining around the Quadrilátero Ferrífero. First Edition*. Belo Horizonte: Ecológico Consultoria Ambiental.

Costa, F. (2012). *Os folgeiros do Jaborandi: Organizacao, Parerias e seu Lugar No Extrativismo Amazonico*. Universidade Federal de Para, Belém, 184.

Cooke, J. A., and Johnson, M. S. (2002). Ecological restoration of land with particular reference to the monitoring of metals and industrial minerals: a review of theory and practice. *Environ. Rev.* 10, 41–71. doi: 10.1139/a01-014

Cooke, J. A., and Suski, C. D. (2008). Ecological restoration and physiology: an overdue integration. *Bioscience* 58, 957–968. doi: 10.1641/B581009

Dorr, J. N. (1964). Supergene iron ores of Minas Gerais, Brazil. *Econ. Geol.* 59, 1203–1240. doi: 10.2113/gsecongeo.59.7.1203

de Melo, J. O., Bitencourt, T. A., Fachin, A. L., Cruz, E. M., de Jesus, H. C., Alves, P. B., et al. (2013). Antidermatophytic and antileishmanial activities of essential oils from *Lippia gracilis* Schauer genotypes. *Acta Trop.* 128, 110–115. doi: 10.1016/j.actatropica.2013.06.024

Erskine, P., van der Ent, A., and Fletcher, A. (2013). Sustaining metal-loving plants in mining regions. *Science* 337, 1172–1173. doi: 10.1126/science.337.6099.1172-b

Fernando, D. R., Marshall, A., Baker, A. J., and Mizuno, T. (2013). Microbeam methodologies as powerful tools in manganese hyperaccumulation research: present status and future directions. *Front Plant Sci.* 4:319. doi: 10.3389/fpls.2013.00319

Ferraz, R. P., Bomfim, D. S., Carvalho, N. C., Soares, M. B., da Silva, T. B., Machado, W. J., et al. (2013). Cytotoxic effect of leaf essential oil of *Lippia gracilis* Schauer (Verbenaceae). *Phytomedicine* 15, 615–621. doi: 10.1016/j.phymed.2013.01.015

Forzza, R. C. (2010). *Catálogo das Plantas e Fungos do Brasil*, Vol. 2. Rio de Janeiro: Andrea Jakobsson Estúdio and Rio de Janeiro Botanical Garden.

Frankham, B. (2010). Where are we in conservation genetics and where do we need to go? *Conserv. Genet.* 11, 661–663. doi: 10.1007/s10592-009-0010-2

Gibson, N., Meissner, R., Markey, A. S., and Thompson, W. A. (2012). Patterns of plant diversity in ironstone ranges in arid south Western Australia. *J. Arid Environ.* 77, 25–31. doi: 10.1016/j.jaridenv.2011.08.021

Gibson, N., Yates, C. J., and Dillon, R. (2010). Plant communities of the ironstone ranges of south-western Australia: hotspots for plant diversity and mineral deposits. *Biodivers. Conserv.* 19, 3951–3962. doi: 10.1007/s10531-010-9939-1

Grant, C., and Koch, J. (2007). Decommissioning Western Australia's first bauxite mine: co-evolving vegetation restoration techniques and targets. *Ecol. Manag. Restor.* 8, 92–105. doi: 10.1111/j.1442-8903.2007.00346.x

Griffith, J. J., and Toy, T. J. (2001). Evolution in revegetation of iron-ore mines in Minas Gerais State, Brazil. *Unasylva* 52, 9–15.

Guzzo, L. S., Saúde-Guimarães, D. A., Silva, A. C., Lombardi, J. A., Guimarães, H. N., and Grabe-Guimarães, A. (2008). Antinociceptive and anti-inflammatory activities of ethanolic extracts of *Lychnophora* species. *J. Ethnopharmacol.* 116, 120–124. doi: 10.1016/j.jep.2007.11.006

Hale, K. L., McGrath, S. P., Lombi, E., Stack, S. M., Terry, N., Pickering, I. J., et al. (2001). Molybdenum sequestration in *Brassica* species. A role for anthocyanins? *Plant Physiol.* 126, 1391–1402. doi: 10.1104/pp.126.4.1391

Jacobi, C. M., and Carmo, F. F. (2008a). The contribution of ironstone outcrops to plant diversity in the iron quadrangle, a threatened Brazilian landscape. *Ambio* 37, 324–326. doi: 10.1579/0044-7447(2008)37[324:TCIOIT]2.0.CO;2

Jacobi, C. M., and Carmo, F. F. (2008b). Diversidade dos campos rupestres ferruginosos no Quadrilátero Ferrífero, MG. *Megadiversidade* 4, 25–32.

Jacobi, C. M., and Carmo, F. F. (2011). Life-forms, pollination and seed dispersal syndromes in plant communities on ironstone outcrops, SE Brazil. *Acta Botanica Brasiliensis* 25, 395–412. doi: 10.1590/S0102-33062011000200016

Jacobi, C. M., Carmo, F. F., and Campos, I. C. (2011). Soaring extinction threats to endemic plants in Brazilian metal-rich regions. *Ambio* 40, 540–543. doi: 10.1007/s13280-011-0151-7

Jacobi, C. M., Carmo, F. F., and de Castro, V. R. (2008). Estudo fitossociológico de uma comunidade vegetal sobre canga como subsídio para a reabilitação de áreas mineradas no quadrilátero ferrífero, MG. *Revista Árvore* 32, 345–353. doi: 10.1590/S0100-67622008000200017

Jacobi, C. M., Carmo, F. F., Vincent, R. C., and Stehmann, J. R. (2007). Plant communities on the ironstone outcrops – a diverse and endangered Brazilian ecosystem. *Biodivers. Conserv.* 16, 2185–2200. doi: 10.1007/s10531-007-9156-8

Krämer, U. (2010). Metal hyperaccumulation in plants. *Annu. Rev. Plant Biol.* 61, 517–534. doi: 10.1146/annurev-applant-042809-112158

Langfield, R. D., Scarano, F. J., Heitzman, M. E., Kondo, M., Hammond, G. B., and Neto, C. C. (2004). Use of a modified microplate bioassay method to investigate antibacterial activity in the Peruvian medicinal plant *Peperomia galloides*. *J. Ethnopharmacol.* 94, 279–281. doi: 10.1016/j.jep.2004.06.013

Leite, T. C., de Sena, A. R., Dos Santos Silva, T. R., Dos Santos, A. K., Uetanabaro, A. P., and Branco, A. (2012). Antimicrobial activity of *Marctetia* DC species (Melastomataceae) and analysis of its flavonoids by reverse phase-high performance liquid chromatography coupled-diode array detector. *Pharmacogn. Mag.* 8, 209–214. doi: 10.4103/0973-1296.99286

Leitenmaier, B., and Küpper, H. (2013). Compartmentation and complexation of metals in hyperaccumulator plants. *Front Plant Sci.* 4:374. doi: 10.3389/fpls.2013.00374

Leporatti, M. L., Pintore, G., Foddai, M., Chessa, M., Piana, A., Petretto, G. L., et al. (2014). Chemical, biological, morphoanatomical and antimicrobial study of *Ocotea puchury-major* Mart. *Nat. Prod. Res.* 28, 294–300. doi: 10.1080/14786419.2013.858338

Lynch, J. P., and St Clair, S. B. (2004). Mineral stress: the missing link in understanding how global climate change will affect plant in the real world soils. *Field Crop Res.* 90, 101–115. doi: 10.1016/j.fcr.2004.07.008

Markey, A. S., and Dillon, S. J. (2010). Flora and vegetation of the banded iron formations of the Yilgarn Craton: the Booylgo Range. *Conserv. Sci. W. Aust.* 7, 503–529.

Marschner, P. (2012). *Marschner's mineral Nutrition of Higher Plants*. Boston, MA: Academic Press.

Mello, L., Camargo, R., Medeiros, D., Almeida, R., Vasconcelos, R., Toledo, F., et al. (2014). "Partnerships and early planning with good science: the key to long-term ecological and socio-economic success," in *Proceedings of Mine Closure Solutions* (Ouro Preto: InfoMine), 26–30.

Memmott, J. (2009). Food webs: a ladder for picking strawberries or a practical tool for practical problems? *Phil. Trans. R. Soc. Bot.* 364, 1693–1699. doi: 10.1098/rstb.2008.0255

Messias, M., Leite, M., Neto, J., Kozovits, A., and Tavares, R. (2013). Soil-vegetation relationship in quartzitic and ferruginous Brazilian rocky outcrops. *Folia Geobot.* 48, 509–521. doi: 10.1007/s12224-013-9154-4

Moraes, T. M., de Araujo, M. H., Bernardes, N. R., de Oliveira, D. B., Lasunskia, E. B., Muzitano, M. F., et al. (2011). Antimycobacterial activity and alkaloid prospectus of *Psychotria species* (Rubiaceae) from the Brazilian Atlantic Rainforest. *Planta Med.* 77, 964–970. doi: 10.1055/s-0030-1250656

Morrissey, J., and Guerinot, M. L. (2009). Iron uptake and transport in plants: the good, the bad, and the ionome. *Chem. Rev.* 109, 4553–4567. doi: 10.1021/cr900112r

Nistelberger, H., Byrne, M., Coates, D., and Roberts, J. D. (2014). Strong phylogeographic structure in a millipede indicates pleistocene vicariance between populations on banded iron formations in semi-arid Australia. *PLoS ONE* 9:e93038. doi: 10.1371/journal.pone.0093038

Novaes, P., Imatomi, M., Varela, R. M., Molinillo, J. M., Lacret, R., Gaultier, S. C., et al. (2013). Allelopathic potential of *Rapanea umbellata* leaf extracts. *Chem. Biodivers.* 10, 1539–1548. doi: 10.1002/cbdv.201200367

Nunes, J. (2009). *Florística, Estrutura e Relações Solo-Vegetação em Gradiente Fitofisionómico Sobre Canga, na Serra sul, Flona de Carajás*. Universidade Federal de Viçosa, Viçosa, 101.

Oliveira, S. H. F., Negreiros, D., Fernandes, G. W., Brabosa, N. P. U., Rocha, R., and Almeida-Cortez, J. C. (2009). Seedling growth of the invader *Calotropis procera* in ironstone rupestrian field and seasonally dry forest soils. *Neotrop. Biol. Conserv.* 4, 69–76. doi: 10.4013/nbc.2009.42.01

Pifano, D. S., Valente, A. S., de Souza Almeida, H., Alves de Melo, P. H., Montianelli de Castro, R., and van den Berg, E. (2010). Caracterização florística e fitofisionómica da Serra do Condado, Minas Gerais, Brasil. *Biota Neotrop.* 10, 55–71. doi: 10.1590/S1676-0632010000100005

Porto, M. L., and da Silva, M. F. F. (1989). Tipos de vegetação metalófila em áreas de Serra de Carajás e de Minas Gerais, Brasil. *Acta. Bot. Bras.* 3, 13–21. doi: 10.1590/S0102-33061989000200002

Pradeep, B. V., Tejaswini, M., Nishal, P., Pardhu, G., Shylaja, S., and Kumar, K. Ch. (2013). Phytochemical screening and antimicrobial activities of plant extract of *Lantana camara*. *J. Environ. Biol.* 34, 645–649.

Reganold, J. P., Andrews, P. K., Reeve, J. R., Carpenter-Boggs, L., Schadt, C. W., Alldredge, J. R., et al. (2010). Fruit and soil quality of organic and conventional strawberry agroecosystems. *PLoS ONE* 5:e12346. doi: 10.1371/journal.pone.0012346

Rezende, L. A. L., Dias, L. E., Assis, I. R., and Braga, R. (2013). Rehabilitation of ironstones outcrops degraded by mining activity in Minas Gerais state of Brazil. *J. Am. Soc. Mining and Reclamation* 2, 151–159.

Ribeiro, A., Piló-Veloso, D., Romanha, A. J., and Zani, C. L. (1997). Trypanocidal flavonoids from *Trixis vauthieri*. *J. Nat. Prod.* 60, 836–841. doi: 10.1021/np970196p

Sarria, A. L., Soares, M. S., Matos, A. P., Fernandes, J. B., Vieira, P. C., and da Silva, M. F. (2011). Effect of triterpenoids and limonoids isolated from *Cabralea canjerana* and *Carapa guianensis* (Meliaceae) against *Spodoptera frugiperda* (J. E. Smith). *Z. Naturforsch. C* 66, 245–250. doi: 10.5560/ZNC.2011.66c0245

Silva, M. F. F. (1992). Distribuição de metais pesados na vegetação metalófila de Carajás. *Act. Bot. Bras.* 6, 107–121.

Silva, W. (2013). *Gradiente Vegetacional e Pedológico em Complexo Rupestre de Quartzito no Quadrilátero Ferrífero, Minas Gerais, Brasil*. University Federal of Viçosa, Viçosa, 79.

Silvério, M. S., Del-Vechio-Vieira, G., Pinto, M. A., Alves, M. S., and Sousa, O. V. (2013). Chemical composition and biological activities of essential oils of *Eremanthus erythropappus* (DC) McLeisch (Asteraceae). *Molecules.* 18, 9785–9796. doi: 10.3390/molecules18099785

Skirycz, A., Jozefczuk, S., Stobiecki, M., Muth, D., Zanor, M. I., Witt, I., et al. (2007). Transcription factor AtDOF4;2 affects phenylpropanoid metabolism in *Arabidopsis thaliana*. *New Phytol.* 175, 425–438. doi: 10.1111/j.1469-8137.2007.02129.x

Skirycz, A., Vandenbroucke, K., Clauw, P., Maleux, K., De Meyer, B., Dhondt, S., et al. (2011). Survival and growth of *Arabidopsis* plants given limited water are not equal. *Nat. Biotechnol.* 29, 212–214. doi: 10.1038/nbt.1800

Soberón, J. R., Sgariglia, M. A., Sampietro, D. A., Quiroga, E. N., Sierra, M. G., and Vattuone, M. A. (2010). Purification and identification of antibacterial phenolics from *Tripodanthus acutifolius* leaves. *J. Appl. Microbiol.* 108:1757–1768. doi: 10.1111/j.1365-2672.2009.04579.x

Souza, A. B., Martins, C. H., Souza, M. G., Furtado, N. A., Héleno, V. C., de Sousa, J. P., et al. (2011). Antimicrobial activity of terpenoids from *Copaifera langsdorffii* Desf. against cariogenic bacteria. *Phytother. Res.* 25, 215–220. doi: 10.1002/ptr.3244

Stouffer, D. B., Sales-Pardo, M., Sirer, M. I., and Bascompte, J. (2012). Evolutionary conservation of species roles in food webs. *Science* 335 1489–1492. doi: 10.1126/science.1216556

Taba, P. (2013). Metals and movement disorders. *Curr. Opin. Neurol.* 26, 435–441. doi: 10.1097/WCO.0b013e3283629beb

Vale Sustainability Report. (2012). Available at: <http://www.vale.com/EN/aboutvale/sustainability/links/LinksDownloadsDocuments/2012-sustainability-report.pdf>

Vale Sustainability Report. (2013). Available at: <http://www.vale.com/EN/aboutvale/sustainability/links/LinksDownloadsDocuments/2013-sustainability-report.pdf>

van der Ent, A., Baker, A. J. M., Reeves, R. D., Pollard, A. J., and Schat, H. (2013). Hyperaccumulator of metals and metalloid traces: facts and fiction. *Plant Soil* 362, 319–334. doi: 10.1007/s11104-012-1287-3

Viana, P. L., and Lombardi, J. A. (2007). Florística e caracterização dos campos rupestres sobre canga na Serra de Calçada, Minas Gerais, Brazil. *Rodriguesia* 58, 159–177.

Vincent, R. C., and Meguro, M. (2008). Plant soil relationships in ferruginous rocky soil vegetation. *Revista Brazil. Bot.* 31, 377–388.

Whiting, S. N., Reeves, R. D., Richards, D., Johnson, M. S., Cooke, J. A., Malaisse, F., et al. (2004). Research priorities for conservation of metallophyte biodiversity and their potential for restoration and site remediation. *Restor. Ecol.* 12, 106–116. doi: 10.1111/j.1061-2971.2004.00367.x

Zimmermann-Franco, D. C., Bolutari, E. B., Polonini, H. C., do Carmo, A. M., Chaves, M. D., and Raposo, N. R. (2013). Antifungal activity of *Copaifera langsdorffii* desf oleoresin against dermatophytes. *Molecules* 18, 12561–12570. doi: 10.3390/molecules181012561

Conflict of Interest Statement: 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. ITV-DS Institute is funded by mining company Vale, which has mining activities in the regions of canga occurrence. This review is related to one of the ITV roles, which is support of canga characterization and restoration affords. Considering Frontiers is a scientific journal all authors kept to the objectivity criteria.

Received: 16 June 2014; **accepted:** 03 November 2014; **published online:** 24 November 2014.

Citation: Skirycz A, Castilho A, Chaparro C, Carvalho N, Tzotzos G and Siqueira JO (2014) Canga biodiversity, a matter of mining. *Front. Plant Sci.* 5:653. doi: 10.3389/fpls.2014.00653

This article was submitted to Plant Genetics and Genomics, a section of the journal Frontiers in Plant Science.

Copyright © 2014 Skirycz, Castilho, Chaparro, Carvalho, Tzotzos and Siqueira. 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) or licensor 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.

Oil palm natural diversity and the potential for yield improvement

Edson Barcelos^{1*}, **Sara de Almeida Rios**¹, **Raimundo N. V. Cunha**¹, **Ricardo Lopes**¹,
Sérgio Y. Motoike², **Elena Babiyuchuk**³, **Aleksandra Skirycz**³ and **Sergei Kushnir**³

¹ Embrapa Amazonia Ocidental, Empresa Brasileira de Pesquisa Agropecuária, Manaus, Brazil, ² Department of Phytotechnology, Federal University of Viçosa, Viçosa, Brazil, ³ Department of Sustainable Development, Vale Institute of Technology, Belém, Brazil

OPEN ACCESS

Edited by:

Ann E. Stapleton,
University of North Carolina
Wilmington, USA

Reviewed by:

Mingsheng Chen,
Chinese Academy of Sciences, China
R. H. V. Corley,
Independent, UK

***Correspondence:**

Edson Barcelos,
Embrapa Amazonia Ocidental,
Empresa Brasileira de Pesquisa
Agropecuária, Rodovia AM 010,
Km 29, Manaus,
Amazonas 69011-970, Brazil
edson.barcelos@embrapa.br

Specialty section:

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

Received: 30 September 2014

Accepted: 09 March 2015

Published: 27 March 2015

Citation:

Barcelos E, Rios SA, Cunha RNV,
Lopes R, Motoike SY, Babiyuchuk E,
Skirycz A and Kushnir S (2015)
Oil palm natural diversity and the
potential for yield improvement.
Front. Plant Sci. 6:190.
doi: 10.3389/fpls.2015.00190

African oil palm has the highest productivity amongst cultivated oleaginous crops. Species can constitute a single crop capable to fulfill the growing global demand for vegetable oils, which is estimated to reach 240 million tons by 2050. Two types of vegetable oil are extracted from the palm fruit on commercial scale. The crude palm oil and kernel palm oil have different fatty acid profiles, which increases versatility of the crop in industrial applications. Plantations of the current varieties have economic life-span around 25–30 years and produce fruits around the year. Thus, predictable annual palm oil supply enables marketing plans and adjustments in line with the economic forecasts. Oil palm cultivation is one of the most profitable land uses in the humid tropics. Oil palm fruits are the richest plant source of pro-vitamin A and vitamin E. Hence, crop both alleviates poverty, and could provide a simple practical solution to eliminate global pro-vitamin A deficiency. Oil palm is a perennial, evergreen tree adapted to cultivation in biodiversity rich equatorial land areas. The growing demand for the palm oil threatens the future of the rain forests and has a large negative impact on biodiversity. Plant science faces three major challenges to make oil palm the key element of building the future sustainable world. The global average yield of 3.5 tons of oil per hectare (t) should be raised to the full yield potential estimated at 11–18t. The tree architecture must be changed to lower labor intensity and improve mechanization of the harvest. Oil composition should be tailored to the evolving needs of the food, oleochemical and fuel industries. The release of the oil palm reference genome sequence in 2013 was the key step toward this goal. The molecular bases of agronomically important traits can be and are beginning to be understood at the single base pair resolution, enabling gene-centered breeding and engineering of this remarkable crop.

Keywords: oil palm, *E. guineensis*, *E. oleifera*, germplasm, hybrid, breeding

Introduction

No human activity has altered the face of the planet more than agriculture (Foley et al., 2005) that is one of the principal causes of biodiversity loss (Green et al., 2005). The increase of the global agricultural production is thought to happen in tropical countries where cropland expanded by approximately 48,000 km² per year from 1999 to 2008 (Phalan et al., 2013). Cropland development is the most controversial in tropics, because they support the high species richness and endemism, and have large projected increases in demand for food from human populations growing in size

and wealth (Laurance et al., 2014). Analysis of crop distribution and expansion in 128 tropical countries showed that overall expansion of annual crops has been more rapid and more widespread than expansion of perennial crops, and has occurred across much of South America, Africa, and tropical Asia. Crops that expanded most during 1999–2008 were soybeans, maize, paddy rice, and sorghum, in that order (Phalan et al., 2013). Soybean expansion is further recognized as a major cause of biodiversity loss in the Brazilian Cerrado savannas (Fearnside, 2001).

Individual crops differ in their biodiversity impacts, depending on how and where they are cultivated. Coffee covers a relatively small area, but tends to replace habitats of particularly high biodiversity value (Phalan et al., 2013). Oil palm fruit is only fifth on a list of biodiversity threats (Phalan et al., 2013), nevertheless “Few developments generate as much controversy as the rapid expansion of oil palm into forest-rich developing countries such as Indonesia” (Sheil et al., 2009). Why some crops have received relatively little attention from conservationists is a matter of debate, yet the negative impacts of the South-East Asian oil palm industry on biodiversity, and on orangutans in particular, have been well documented and publicized (Fitzherbert et al., 2008). In regard of the great apes habitat destruction, similar concerns are also expressed about the wildlife habitat conversion by the oil palm plantations in Africa (Wich et al., 2014), where about two million hectares are likely to be converted for oil palm cultivation (Murphy, 2014). Species diversity, density and biomass of invertebrate communities is estimated to suffer at least 45% decreases from land-use transformation of tropical forests to oil palm plantations (Barnes et al., 2014). Furthermore, as a substitute for reforestation, the native biodiversity of oil palm plantations is far lower than that of rubber tree plantations, which are the primary current threat to the rain forests in Cambodia (Fitzherbert et al., 2008). On the other hand, oil palm was found most sustainable with respect to the maintenance of soil quality, net energy production and greenhouse gas emissions, when biodiversity loss due to oil palm expansion was analyzed in relation to alternative crops for oil or energy, such as soybean, rapeseed, corn or sugar cane (de Vries et al., 2010). The peatland deforestation for oil palm cultivation in West Kalimantan, Indonesia have a large negative impact on greenhouse gas emissions (Carlson et al., 2012). However, global analysis of oil palm cultivation suggests that crop may encourage forest reversion and lower global emissions (Villoria et al., 2013), mainly because oil palm plantations store more carbon than alternative agricultural land uses (Sayer et al., 2012).

Deforestation and peatland degradation can be avoided when degraded lands, such as *Imperata cylindrica* grassland in Indonesia (Wicke et al., 2011) and cattle pastures in Amazon (Godar et al., 2014; Villela et al., 2014) are used for oil palm cultivation; a solution embraced by environmentalist and policy makers (World Resources Institute¹). Even Greenpeace admits that “good palm oil” is acceptable if policy makers: (1) put an end to deforestation; (2) introduce peatland restoration policies; (3) support smallholder farms and (4) involve local communities in palm oil

¹<http://www.wri.org>

business². To reduce the environmental footprint of oil palm, The Roundtable for Sustainable Palm Oil (RSPO) has been established in 2004³. The RSPO is a non-profit association that brings together palm oil producers, processors and traders, consumer goods manufacturers, retailers, banks and investors, as well as environmental and social non-governmental organizations (NGOs) to develop and implement a global standard for sustainable palm oil in order to produce Certified Sustainable Palm Oil (CSPO). On the other hand, smallholder farmers have difficulties to meet CSPO criteria (Murphy, 2014).

Why this global integration is so important? Oil palm cultivation is one of the most profitable land uses in the humid tropics (Sayer et al., 2012). In the state of Pará, Brazil for example, the average annual monetary return on investment is US\$ 2000 per hectare (Villela et al., 2014). With further 32 million hectares of degraded land suitable for oil palm cultivation (Ramalho-Filho et al., 2010) crop has a potential to evolve into a multibillion dollar business in Brazil. The crop is often considered as an industrial crop, but in many areas it is a valuable smallholder crop (Feintrenie et al., 2010). Globally, three million smallholders live from oil palm cultivation. The share of palm oil production by small, family-owned, estates is 30% worldwide and reaches 80% in Nigeria, Africa’s largest producer (Morcillo et al., 2013). Government policies in Malaysia, Indonesia and Brazil favor smallholder involvement in the oil palm industry. Indonesia has a target of 40% production coming from smallholders, who supply the oil mills. In Indonesia, 25 million people livelihood depends one way or another on oil palm production (Murphy, 2014). Thus, oil palm cultivation alleviates poverty and with right governmental policies could transform livelihood of millions of people (Sayer et al., 2012).

Oil palm (*Elaeis guineensis*, Jacq.) is by far the most productive oil crop and alone is capable to fulfill the large and growing world demand for vegetable oils that is estimated to reach 240 million tons by 2050 (Corley, 2009). Per hectare of cropland, oil palm plantations give 3–8 times more oil than any other temperate or tropical oil crop. In 2012 for instance, 56.2 million tons of palm oil were produced on 17.24 million hectares. Only 23.6 million tons of oil were extracted from rapeseed grown on 36.4 million hectares⁴. On November 2014, palm oil was valued as a vegetable oil with the lowest production costs by the international commodities markets⁵, e.g., US\$ 700 and US\$ 850 per metric ton of oil palm and rapeseed oil, respectively. The oil palm cultivation however, is labor intensive. As the labor costs increase, overseas workers are often providing cheap labor force. Half a million Indonesian workers are now being recruited to Malaysian oil palm plantations. Manpower shortfall resulted in 15% losses of fruits in East Malaysia, whilst the lack of good field training is a further contributing factor for production losses (Murphy, 2014). It is a challenge for palm breeders to alter tree architecture in such a way as to lower the labor intensity of the crop and to facilitate mechanization of the harvest.

²<http://www.greenpeace.org>

³<http://www.rspo.org>

⁴<http://faostat3.fao.org/home/E>

⁵<http://www.indexmundi.com>

Oil palm is a C3 photosynthesis, evergreen tropical perennial tree (Corley and Tinker, 2003). Adult palms planted at optimal density of 130–150 trees per hectare are on a relative steady state in terms of canopy development and have a large leaf area index between 4 and 5, leading to a light interception efficiency close to one (Pallas et al., 2013). Unlike other studied angiosperms, oil palm does not regulate photosynthesis to adjust source–sink imbalances (Legros et al., 2009), instead photosynthates are converted into a reserve pool of non-structural carbohydrates (NSC) mainly located in the tree trunk as glucose and starch. The main physiological function of transitory NSC storage is to balance sink and source fluctuations within a day to seasons. The NSC reserve pool in oil palm is so large that it can theoretically sustain tree growth for 7 months, and most importantly to maintain fruit growth and energetically costly oil biosynthesis at fruit maturation, regardless of cloud cover or periodical suboptimal growth conditions (Legros et al., 2009). In contrast, intercepted solar radiation during seed filling is the rate limiting in an annual oil crop sunflower, and determines weight per seed and oil concentration (Aguirreza et al., 2003). Thus, unusual characteristics of source–sink interactions, combined with high efficiency of solar radiation interception and continuous year-round fruit production are likely ecophysiological traits that determine superior productivity of oil palm (Basri Wahid et al., 2005).

Plant scientists commonly argue that finding solutions for increasing crop yield potential, e.g., doubling yield by improving photosynthesis efficiency, and closing the yield gap will satisfy food demand by the growing human population that is estimated to reach 9–10 billion by the year 2050. The challenge for oil palm planters will be to close the yield gap between the average plantation output at present 3.5t, compared to some best known varieties that in favorable agro-climatic conditions produce up to 9–12t (Murphy, 2009). However, the intensification of land use appears to result in further biodiversity loss, the so-called trade-offs between biodiversity value and yield (Phalan et al., 2011, 2014). High-yield oil palm expansion spares land at the expense of forests in the Peruvian Amazon (Gutiérrez-Vélez et al., 2011). Thus, increase in yield potential of oil palm crop is a necessary, but not sufficient requirement for the sustainable future of tropical forests.

The aim of this review is to acquaint the reader with natural diversity of oil palm species and how it can be used to increase productivity of the future oil palm plantations. We will begin by covering basic facts about oil palm biology and cultivation, followed by a brief review of germplasm and genomic resources, and how these have been used for gene discovery and breeding purposes. The potential of other oleaginous palms to lower environmental impact of oil palm cultivation is emphasized.

Oil Palm Biogeography, Biology and Cultivation

The genus *Elaeis* of the monocotyledonous palm family *Arecaceae* was formally introduced into botanical classification in 1763 by Nicholas Joseph Jacquin, who described *Elaeis guineensis*, known as African oil palm. The Greek word “ελαῖον”—oil, transliterated “elaion,” gave the genus name (Hartley, 1977). The genus

comprises two taxonomically well-defined species, the second is American oil palm, *E. oleifera*. The well-known, phylogenetically closest relative of oil palms are coconut palms, *Cocos nucifera*, which are also vegetable oil producing crop.

African Oil Palm

The world-wide grown crop is African oil palm naturally abundant in all the African rain forests. Both climate, and humans shaped modern biogeographic distribution. The late Holocene phase of dramatic forest decline, around 2500 years ago was favorable for the expansion of this sun-loving, pioneering species (Maley and Chepstow-Lusty, 2001). The oleaginous properties of the fruits were important in the subsistence economy in Africa for the past 5000 years (Sowunmi, 1999).

Analysis of the species natural genetic diversity suggests that wild populations could be separated into three groups located at the extreme west of Africa, equatorial Africa and on Madagascar Island. The highest allelic diversity was found among Nigerian palm populations, indicating the possible center of origin (Bakoumé et al., 2015). Semi-wild feral populations of African oil palm found in a Brazilian state Bahia are very similar to palms from Nigeria and were most likely established during the period of slave trade (Bakoumé et al., 2015).

Thus, the majority of wild oil palm populations inhabit tropical lowlands with the average annual rainfall of about 1780–2280 mm and temperature ranging from 24 to 30°C. Accordingly, varieties standard in cultivation are usually sensitive to water deficit. The atmospheric humidity also strongly influences oil palm photosynthetic capacity. Low air humidity restricts stomatal opening and CO₂ uptake (Smith, 1989). Another ecophysiological characteristic that limits the latitude and altitude ranges of cultivation is cold sensitivity. The growth of common oil palm varieties is suppressed at ambient temperatures below 15°C (Corley and Tinker, 2003). Oil palm can be cultivated on a broad range of soils (Corley and Tinker, 2003).

African oil palm trees can reach 15–18 meters in height, up to 30 meters in a dense forest. It is believed that some palm groves are more than 200 years old (Corley and Tinker, 2003). The leaves could be 8 meters in length. It takes about 2 years for the first leaf primordia to reach the fully expanded stage. To achieve maximal yield on commercial plantations, the leaf length is a critical trait that determines tree planting density.

Oil palms are monoecious species that produce unisexual male and female inflorescences in an alternating cycle. Such “temporal dioecism” results in allogamous reproduction by cross-pollination (Adam et al., 2011). Inflorescences are enclosed during their development by spathe, a large bract, which is ruptured just before flower maturity is reached. Both genetic and environmental factors influence inflorescence sex determination. Reduced photosynthesis due to defoliation, or high density planting, for example, promotes male inflorescence development. This observation was critical for the development of industrial scale seed production by controlled pollination (Durand-Gasselin et al., 1999). Breeding for higher productivity result in varieties that produce larger number of female inflorescences and shortage of pollen. Thus, pollination efficiency has a large impact on yield of the crop.

Whilst wind pollination could occur, maximal pollination efficiency depends on insects. The main pollinators are weevils, a type of beetles, of the genus *Elaeidobius* spp., in particular *E. kamerunicus*. Weevils complete the entire life cycle by feeding on the palm male flowers and male inflorescence tissues. The aniseed scent that is the same between female and male inflorescences, is attributed to the emission of methyl chavicol (Lajis et al., 1985), that attracts *E. kamerunicus* (Hussein et al., 1989). In a search for the male inflorescences, weevils visit female flowers, depositing pollen grains by accident (Tandon et al., 2001). To achieve maximal pollination efficiency, *E. kamerunicus* were introduced on plantations both in South-East Asia, and Latin America (Corley and Tinker, 2003).

The oil palm fruit is a sessile drupe. Fruits grow in large bunches and mature in 5–6 months after pollination. Oil palm accessions show great variation in fruit shape and size (Corley and Tinker, 2003). The pericarp of the oil palm fruit is subdivided into the outer layer exocarp, fleshy mesocarp, and endocarp that in oil palm, is called shell. Shell encases the seed or kernel, i.e., embryo and endosperm. The crude palm oil and kernel palm oil are extracted from mesocarp and kernel, respectively.

In wild type palms, endocarp thickness varies from 2 to 8 mm in between different accessions (Corley and Tinker, 2003). Endocarp development depends on the major effect *SHELL* gene. *SHELL* mutant alleles with co-dominant monogenic inheritance characterize the so-called *pisifera* palms (*sh/sh*) that produce shell-less fruits (Beirnaert and Vanderwegen, 1941). Wild type *Sh/Sh* *dura* palms develop fruits with thick endocarp. Intraspecific heterozygous (*Sh/sh*) hybrids, known as *tenera* palms, have thinner shells surrounded by a distinct fiber ring (Beirnaert and Vanderwegen, 1941). The shell thickness has major effect on oil content, with *teneras* having 30% more mesocarp and respectively 30% greater oil content in bunches than *duras* (Corley and Tinker, 2003). Owing to their higher oil yields, *tenera* palms were selected already by the pre-colonial cultures in West Africa (Devuyst, 1953).

African oil palms have different coloring of exocarp, producing either *nigrescens* or *virescens* fruit types. Type *nigrescens* accumulate large amounts of anthocyanins, which accounts for the deep violet to black color at the fruit apex (Figure 1A). When unripe, *virescens* fruits are green turning orange due to accumulation of carotenoids and chlorophyll degradation in relation to ripening. Five spontaneous dominant mutant alleles in a *VIRESSENS* gene abolish anthocyanin synthesis, which explains the *virescens* fruit type (Singh et al., 2014). Fruit phenotype *nigrescens* is a likely wild type. The occurrence of *virescens* palms is usually less than 1%, however, in some Congolese oil palm populations up to 50% of trees will produce *virescens* fruits. Artificial selection by local communities has driven the persistence of the newly arising mutations (Zeven, 1972).

In most of angiosperms, flowers and immature fruitlets are naturally thinned by organ abscission in response to nutritional status. This phenomenon is negligible in oil palm that only shed ripe fruits (Roongsathian et al., 2012). It is plausible that low abscission of flowers and immature fruitlets contributes to the exceptional oil palm productivity. Ripe fruit shedding is a main indicator whether bunch is ready to be harvested (Corley and

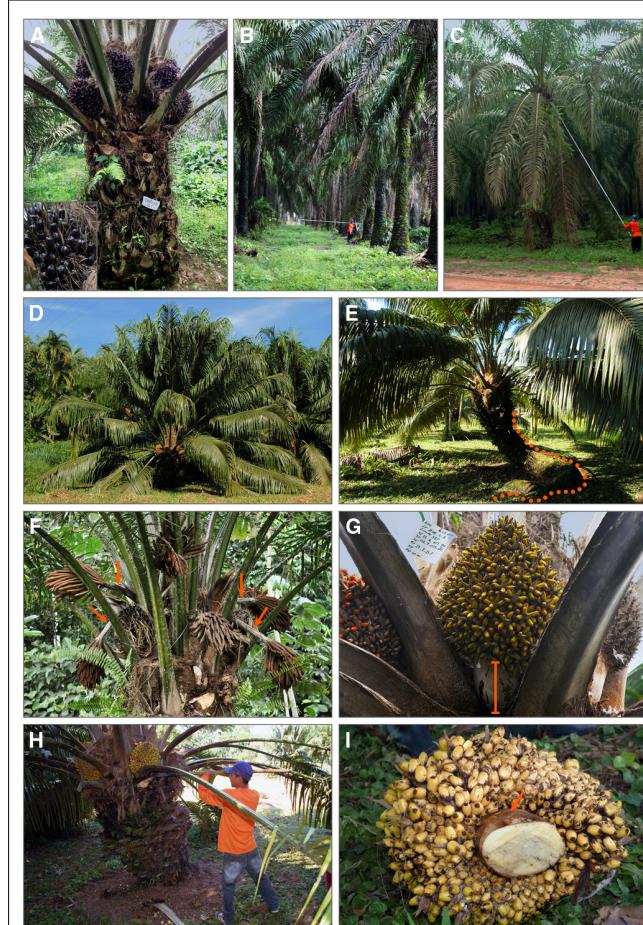


FIGURE 1 | *E. guineensis* and *E. oleifera*. (A) Commercial African *tenera* oil palm from a cross *dura* Deli × *pisifera* Nigeria. The tree is 5-years-old and has *nigrescens* fruit type (inset). (B) 26-years-old plantation of *tenera* palms (Deli × Ghana). Trees are 7–8 meters tall. (C) Fruit bunch harvest. A worker is using a knife to cut off bunches from a tree on a plantation shown in (B). Bunch ripeness is assessed by the presence on the ground of the shed-off fruits. (D) *E. oleifera*. The wild tree that is more than 30-years-old (Manicoré, Amazonas, Brazil). (E) Walking palm. The same tree as in (D) photographed at different angle to illustrate procumbent trunk that is outlined by the dotted line. (F) African oil palm male inflorescences. This individual has particularly long male inflorescence stalks that are pointed with arrows. (G) Fruit bunch stalk. The 3-months-old fruit bunch of the American oil palm is shown. Bar indicates short stalk. (H) Collection of *E. oleifera* bunch. Harvesting oil palm requires skilled workers, even when trees are not very tall. (I) Thickness of the bunch stalk. The cut-off bunch from a tree in (H). Cutting through the bunch stalk (orange arrow) composed of a very fibrous tissue, requires physical strength.

Tinker, 2003), on the other hand, shedding is one of the causes of losses at harvesting (Osborne et al., 1992).

The stalk of fruit bunches is short and thick in oil palms (Figures 1G,I). The stalks of male inflorescences are longer (Figure 1F). Cutting fruit bunches off the trees is a laborious process (Figures 1B,C,H), thus short and thick stalks are the traits that limit harvest mechanization (Le Guen et al., 1990).

To achieve synchronous germination of commercially produced seeds, combined temperature and humidity treatments

are required to break the dormancy. Each germinated seed is maintained in a pre-nursery for 4–5 months till a plantlet reaches a four-leaf stage, after which young palms are grown for about a year in a nursery before they are transferred to the field. Establishment of leguminous cover prior planting prevents soil erosion and surface run-off, improves soil structure and palm root development, increases the response to mineral fertilizer, and reduces the danger of micronutrient deficiencies (Corley and Tinker, 2003).

American Oil Palm

Elaeis oleifera (Kunth, Cortés) is known as the American oil palm. Species is native to and broadly dispersed in Central America and northern regions of South America. Small and dense *E. oleifera* populations grow along the riverbanks, tolerating well both shade, and flooding, indicating a broader environmental adaptability compared to the African oil palm (Corley and Tinker, 2003). Judging by the higher morphological variation of the trees, the region occupied by Colombia, Suriname, and North-West Brazil is thought to be the species center of origin (Meunier, 1975; Ooi et al., 1981). In the Amazon River Basin that is considered a center of secondary diversification (Meunier, 1975; Barcelos et al., 2002), many *E. oleifera* populations are found on Amazonian Dark Earths, Terra Preta de Índio in Portuguese. Amazonian Dark Earths were formed in the past by pre-Columbian populations and are highly sustained fertile soils supported by microbial communities that differ from those extant in adjacent soils (Lima et al., 2014). In spite of this association of palms with human habitation, there are no historical indications of artificial selection for improved yield that in *E. oleifera* remains significantly lower compared to the African oil palms. Oil to bunch ratio of *E. oleifera* is about 5%, as compared to 25% in *E. guineensis teneras* (Barcelos, 1998b).

A distinguishing feature of *E. oleifera* is a much shorter, often procumbent trunk, a trait from which species are also known as a walking palm (Figures 1D,E). After procumbence, the basal part of the plant dies whilst adventitious roots sprouting from the part in contact with soil allow plant to restart growth. The high proportion of parthenocarpic fruits that may constitute up to 90% of the total is another striking characteristic of the *E. oleifera* fruit bunches as compared to the African species. Parthenocarpic fruits often abort, contributing to poor yield. Immature fruits are green turning orange at maturity, which resembles *virescens* African oil palms (Figures 2A,B). *E. oleifera* leaves have a different from *E. guineensis* positioning of leaflets (Corley and Tinker, 2003). *E. oleifera* pollination depends on insects. However, the profile of volatiles emitted by the inflorescences at anthesis is different and species do not synthesize methyl chavicol (Gomes, 2011).

Interspecies Oil Palm Hybrids

African and American oil palm species are sexually compatible (Harden and Tan, 1969). F1 hybrids show vegetative vigor and mid-parent stem growth increment (Corley and Tinker, 2003). *E. oleifera* leaf morphology and parthenocarpic fruit development (Figure 2D) behave as dominant traits. F1 hybrids from crosses between some *E. oleifera* accessions and *nigrescens* *E. guineensis*



FIGURE 2 | *E. oleifera* × *E. guineensis* interspecific hybrids at EMBRAPA breeding station Rio Urubu (Amazonas, Brazil).

(A) Controlled pollination. Worker is removing the spathe on elite *E. oleifera* tree that is three meters tall. He wears safety gear to climb the tree, protective mask and gloves. In commercial seed production, female inflorescences are bagged 1 week before anthesis. **(B)** Bunch ripening. Three bunches at different times after pollination are visible. As fruit mature, the fruits change color from green to orange. The youngest female inflorescences were not used in pollination, those bunches are still enclosed in spathe, which is an *E. oleifera* species-specific characteristic. **(C)** F1 interspecific hybrid on commercial plantation. Tree is 4 years old. It has about twice shorter stem than African oil palm trees of the same age. Orange mature bunches are easily spotted within the tree crown. **(D)** Parthenocarpic fruits in mature bunch of the F1 interspecific hybrid. Insert: a few fruits were transversely cut. The fruit to the right has normal seed. White colored endosperm tissue of the kernel is surrounded by the black endocarp, mesocarp is orange. Parthenocarpic fruits have residual black endocarp and no kernel. **(E)** Female inflorescence of F1 hybrid at anthesis. Inflorescence is still enclosed in spathe, which is similar to the parent *E. oleifera*. **(F)** Female flowers. The same inflorescence as in **(E)**. Spathe was removed to show female flowers at anthesis. Three-lobed flower stigmas are yellowish in color and receptive for pollen. **(G)** Male inflorescence of F1 hybrid at anthesis. The overall morphology and flower identity are normal. **(H)** Andromorphic inflorescences. This 15-years-old tree developed five fully andromorphic inflorescences pointed with blue arrows and a single male inflorescence (behind the leaf rachises to the right). Orange arrow points a fruit bunch. **(I)** Partial andromorphy. Individual andromorphic inflorescence can show different proportions of male flowers (orange arrow) and female flowers (blue arrow). **(J)** Fruitlets from andromorphic inflorescence. Some fruitlets abort after anthesis, died dry flower pistils are black (orange arrow), other develop parthenocarpically (blue arrow). Andromorphic inflorescences in **(H)** are full of developing fruitlets.

parents, have *virescens* phenotype of fruits that are bright orange at maturity (Figures 2C,D).

To achieve reasonable yield from F1 hybrids, assisted pollination is required, i.e., one worker to pollinate 10–20 hectares of a plantation. In relation to the accession of the *E. oleifera* and *E. guineensis* parents, a number of developmental abnormalities could contribute to lower F1 hybrid fertility, including lower pollen yield, poor pollen germination, poor anther dehiscence (Corley and Tinker, 2003); lower emission of volatiles by the inflorescences at anthesis (Gomes, 2011), which is a likely cause of poor attractiveness for *E. kamerunicus* (Tan, 1985). The spathe encasing female inflorescence at anthesis could present a mechanical barrier for pollinating insects (Figures 2B,E,F). Of interest for developmental biologists are the andromorphic inflorescences, a trait that reduce pollen production and characterize some F1 hybrid combinations (Figures 2G–J). Whether allopolyploidization will alleviate, or exacerbate fertility problems of the interspecific oil palm hybrids is unknown.

It is possible that some of the F1 hybrid developmental abnormalities and dominance-recessiveness gene interactions can be attributed to the non-additive gene expression that is typical for interspecific hybrids (Chen, 2013). In a view of oil palm genomics development, it is tempting to consider ion beam deletion mutagenesis (Ishikawa et al., 2012) as an experimental tool to identify the molecular basis of the traits that distinguish oil palm species and their hybrids.

Natural Variation for Oil Palm Improvement

The large scale establishment of Congolese and South-East Asian commercial plantations in 1910–1920s, was quickly followed by the research on the crop improvement by selection and breeding (Corley and Tinker, 2003). Oil palm breeding was influenced by maize breeding that relies on development of inbred parental lines to produce homogeneous F1 hybrids. Accordingly, reciprocal recurrent selection and family-individual selections methods were commonly chosen by oil palm breeders for the development of parental lines that are used in commercial F1 hybrid seed production. Ten percent productivity gains per decade were reported by the French and Malaysian breeding programs (Corley and Tinker, 2003).

The exploitation of the superior oil content of the *teneras* began in 1930s on Congolese plantations, which led Beirnaert to explain *dura*, *pisifera*, and *tenera* phenotypes by the co-dominant monogenic inheritance of the *shell* (*sh*) mutant allele (Beirnaert and Vanderweyen, 1941). On Malaysian plantations, the cultivation of *teneras* (*Sh/Sh*) produced by controlled pollination of *duras* (*Sh/Sh*) with pollen of *pisifera* palms (*sh/sh*) took off in 1956. Most of commercial seeds today are intraspecific *dura* × *pisifera* (D × P) hybrids (Corley and Tinker, 2003).

Oil palm breeders use “breeding populations of restricted origin” (BPRO) that can be traced back to distinct, often small groups of wild or unimproved ancestral palms (Rosenquist, 1986). For example, Deli *duras* are used today as the mothers for almost all commercial *teneras* seed production. The Deli *dura* palms can be traced to four individuals planted in the Bogor Botanical Gardens (Java, Indonesia) in 1848 (Kushairi and Rajanaidu, 2000;

Cochard et al., 2009). Most commonly used *pisiferas* descend from limited number of origins, as well. A single Django *tenera* palm from Congo, gave rise to the AVROS (Algemeene Vereniging van Rubber Planters ter Oostkust van Sumatra, now Indonesian Oil Palm Research Institute—IOPRI⁶) *pisiferas* widely used for seed production in Indonesia, Malaysia, Papua New Guinea, and Costa Rica (Rajanaidu et al., 2000; Corley and Tinker, 2003). Thus, commercial *teneras* have a narrow genetic base due to the restricted number of ancestral progenitors (Hardon, 1968; Ooi and Rajanaidu, 1979).

Many breeders realized the need for new material of sufficient genetic variability for future progress. Thus, the major task of the oil palm research institutions was adequate characterization of the natural diversity either by prospection of *in situ* natural populations, or by the establishment of *ex situ* germplasm collections. Beginning in 1950s, tens of thousands of palms from *in situ* natural populations have been screened at low cost to isolate just a few elite individuals that were then used in breeding programs. This strategy was largely adopted by the Institut de Recherches pour les Huiles et Oléagineux (IHRO), presently CIRAD⁷ (Meunier, 1969). Since 1970s, to capture broader spectra of natural variation, and to ensure protection of the trees from destruction, Palm Oil Research Institute of Malaysia (PORIM), presently Malaysian Palm Oil Board (MPOB)⁸ engaged in random sampling of palms from wild populations to establish *ex situ* germplasm collection of 1467 accessions (Rajanaidu, 1994). This strategy is more expensive than *in situ* prospection. MPOB Nigerian collection alone occupies 200 hectares. However, *ex situ* approach allows more thorough characterization of the traits of interest, and provides guarantees for preservation of traits that might become of interest in the future (Rajanaidu et al., 2000; Paterson et al., 2013). The fate of *in situ* germplasm is unpredictable, they can be either destroyed, or replanted. In the survey organized by the Food and Agriculture Organization of the United Nations, 29 participating institutions reported a total of 21103 oil palm accessions (F.A.O, 2010)⁹.

Phenotypic screens of the *E. guineensis* germplasm collections conducted by the main oil palm research centers MPOB (Malaysia), CIRAD (France), IOPRI (Indonesia), and the Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA, Brazil) revealed a significant phenotypic diversity for the valuable agronomical characteristics, such as: (1) leaf petiole, rachis length, i.e., breeding for the so-called compact palms; (2) increment in the growth in height, i.e., breeding for shorter palms; (3) bunch number, weight and production, i.e., oil yield; (4) fresh fruit bunch and crude palm oil yield; (5) total and vegetative dry matter production; (6) fruit and kernel size, i.e., fatty acid (FA) profile; (7) fruit shell thickness; (8) Fusarium wilt disease tolerance; (9) FA composition and iodine value; (10) carotene and vitamin E contents; (11) lipase activity; (12) *in vitro* regeneration potential; (13) drought and cold tolerance (Rajanaidu et al., 2000; Corley and Tinker, 2003). Based on the phenotypic data, selected accessions have

⁶<http://iopri.org/>

⁷<http://www.cirad.fr/>

⁸<http://www.mpob.gov.my/>

⁹<http://www.fao.org/docrep/013/i1500e/i1500e00.htm>

been subsequently used to develop improved *dura* and *pisifera* fruit type parental palm varieties.

Oil palm breeders became interested in *E. oleifera* agronomic potential at the beginning of the last century. In 1920s *E. oleifera* was introduced in Africa and in 1950s to Asia. However, it is only in the last 30–40 years that *E. oleifera* natural populations have been thoroughly sampled to establish *ex situ* germplasm collections in Malaysia, Ivory Coast, Costa Rica and Brazil (Meunier, 1975; Escobar, 1981; Ooi et al., 1981; Rajanaidu, 1986; Barcelos et al., 1999, 2002). FAO Database registered 506 accessions, of which 244 are maintained by the EMBRAPA on a breeding research station¹⁰ located in the municipality of Rio Preto da Eva, state Amazonas, Brazil.

American oil palm is a source of many economically valuable traits, of which most important are (1) slow height increment, which simplify harvest (Corley and Tinker, 2003); (2) higher proportion of desaturated FAs in palm oil (Montoya et al., 2014); (3) lower lipase activity in mature fruit mesocarp, extending a period between harvest and fruit processing (Sambanthamurthi et al., 1995; Cadena et al., 2013); (4) higher vitamins A and E contents, improving oil nutritional value (Rajanaidu et al., 2000) and (5) broader environmental adaptability (Barcelos, 1986, 1998b). In addition, American oil palm is also more resistant to several diseases (Corley and Tinker, 2003), including bud-rot caused by *Phytophthora palmivora* and *Fusarium* wilt (Barcelos, 1986).

Hybrids between *E. guineensis* × *E. oleifera* excited much interest, because of the slower growth and higher desaturation of palm oil (Corley and Tinker, 2003). Interest in interspecific hybrids further increased with a recognition of their resistance to fatal yellowing disease that is a major threat to oil palm cultivation in Latin America, a discovery that led to the first commercial plantations in 1980s. In the study published in 1995, Amblard et al. (1995) analyzed 429 hybrid progenies obtained by crossing *E. oleifera* and *E. guineensis* of different origin. Bunch and oil production in the best inter-origin combinations reached 85 and 78% of the average values for the commercial oil palm cultivars, respectively. Some selected hybrids have oil productivity as high as commercial *teneras*, but only with assisted pollination because of the serious fertility problems. Seeds of high yielding interspecific hybrids are produced by CIRAD/PalmElit¹¹ (COARI hybrids) and EMBRAPA/Dendê do Pará S.A.¹² (MANICORÉ hybrids), using wild *E. oleifera* palms indigenous to the Coari and Manicoré municipalities in the Amazon river basin.

Genomics for Oil Palm Improvement

Oil palm is a diploid ($2n = 32$) with an estimated genome size of 1.8-gigabases (Gb). A total of 1.535 Gb of the *E. guineensis* (AVROS, *pisifera* fruit form) reference genome assembly were released to public in 2013 (Singh et al., 2013b) and is freely available¹³. For comparative purposes, the genome the American oil

palm was sequenced. Most of the reference genome is represented by segmental duplications, and not triplications, indicating that oil palm is a paleotetraploid. Analysis of conserved gene order revealed that the duplications were retained in *E. oleifera*, so that segmental duplications pre-dated the divergence of the African and American oil palms. On the other hand, 57% of the 1.8-Gb *E. guineensis* genome comprises repetitive elements of which 47% were uncharacterized previously, with 73% absent from *E. oleifera* genome, indicating extensive molecular speciation that might account for the fertility problems of interspecific oil palm hybrids.

Genome sequence and transcriptome data from 30 tissue types were used to predict at least 34,802 genes (Singh et al., 2013b). De novo assembly from RNA-seq data resulted in 51,452 oil palm unigenes (Lei et al., 2014). To characterize the genic regions in a greater detail, the methylation filtered libraries of the African and American oil palm species were sequenced (Low et al., 2014). Sequence analysis revealed single nucleotide polymorphisms (SNP) at densities 2.30 and 2.83 per 100 bp for *E. guineensis* and *E. oleifera*, respectively.

For a perennial tree that flowers only 2–3 years after seed germination, breeding oil palm requires 10–19 years per cycle of phenotypic selection (Wong and Bernardo, 2008). Molecular breeding uses genetic markers linked to the traits of choice for earliest pre-selection of desired phenotypes and has a potential to greatly shorten the breeding cycle, reducing costs. The data of (Low et al., 2014) were used to generate a final set of 4,451 SNPs that were selected for developing a customized oil palm specific SNP array (OPSNP3) printed on the Infinium HD iSelect BeadChips platform (Ting et al., 2014). Genotyping across 199 palms from two separate mapping F1 hybrid populations, e.g., *E. oleifera* × *E. guineensis* interspecific cross and a *dura* × *pisifera* intraspecific cross took less than 3 months (Ting et al., 2014) and greatly improved marker density and genome coverage in comparison to the first reference maps based on AFLP and SSR markers (Barcelos, 1998a; Barcelos et al., 2002; Singh et al., 2009; Billotte et al., 2010; Ting et al., 2013). Refined genetic maps combined with careful phenotyping of trees are likely to facilitate mapping and identification of molecular bases of both monogenic, and quantitative trait loci (QTL) that underpin major agricultural traits of interest (Yang et al., 2014).

Increasing Oil Palm Productivity

The immediate impact of the oil palm genome sequence was identification of the *SHELL* gene (Singh et al., 2013a) that was shown to encode a MADS-box transcription factor homologous to the *Arabidopsis* ovule identity and seed development regulator *SEEDSTICK*. Two different amino-acid substitution mutations in a dimerization and DNA-binding domain of the *SHELL* protein occur in the *sh^{MPOB}* and *sh^{AVROS}* spontaneous mutant alleles. Mutant proteins are likely to act as *trans*-dominant negative isoforms, which explains the co-dominant phenotype in *tenera* palms. Molecular markers for *SHELL* gene alleles could be used to distinguish *dura*, *tenera*, and *pisifera* plants in the nursery long before they are planted in a field. Nursery stage screening can eliminate erroneous planting of *dura* palms and

¹⁰<https://www.embrapa.br/en/amazonia-ocidental>

¹¹<http://www.palmelit.com/en/>

¹²<https://www.embrapa.br/en/amazonia-ocidental>

¹³<http://genomsawit.mpop.gov.my/genomsawit/>

control the precision of hybrid seed production. Marker-assisted introgression of the *SHELL* gene alleles on different genetic backgrounds could accelerate construction of new *dura* and *pisifera* palms.

It will be important to understand whether *sh^{MPOB}*, *sh^{AVROS}* alleles or similar *trans*-dominant alleles constructed by protein engineering will show a dosage effect that could further increase the oil yield in palms, which have genotype *Sh/sh/sh*, for example. The effect of the *SHELL* gene variation on the mesocarp yield of other commercially useful palm species, such as date (*Phoenix dactylifera*), açaí (*Euterpe oleracea*), peach (*Bactris gasipaes*) palms, can be tested.

Further yield enhancements were brought by breeding *dura* and *pisifera* parents for the higher ratio of female inflorescences; bunch weight; oil to bunch ratio; oil recovery and earlier flowering. Donors of such traits are breeding accessions of Deli palms that have high yield; AVROS lines characterized by precocity, high yields and growth vigor; Ekona palms that have high oil to bunch ratio and earlier flowering Yangambi palms (Alvarado et al., 2010).

Best *dura* × *pisifera* combination have 28–32% oil to bunch ratio and can produce annually up to 10t, e.g., (i) CIRAD Deli × Yangambi (PalmElit¹⁴); (ii) Evolution (Alvarado et al., 2010)¹⁵, crosses of Deli *dura* with composite *pisifera* carrying traits introgressed from several oil palm populations.

Potential yield of hypothetical oil palm genotypes that combine physiologically plausible attributes is being estimated at about 18.5t (Corley, 1998), which is almost a double of the best varieties on the oil palm seed market. Site yield potential varies amongst individuals of the same descent, in part due to residual genetic variation in parental lines. Exceptionally high yields of 12–13t could be anticipated if some well performing individual trees could be multiplied (Sharma and Tan, 1997). Oil palms do not branch unless terminal single vegetative shoot apical meristem is damaged. Thus, to produce planting material from high yield elite individuals, the only practical way of vegetative propagation is *in vitro* clonal propagation, micropropagation (Corley and Tinker, 2003). Oil palm micropropagation remains an inefficient, lengthy process, however. Most genotypes are recalcitrant in tissue culture, which requires empirical tests of numerous media formulas. Introgression into elite varieties the superior somatic embryogenesis capacity known for some palm accessions is considered a partial solution to the problem (Ting et al., 2013). Inducible versions of genes that promote somatic embryogenesis is a worthwhile approach (Heidmann et al., 2011) that has not been tested with oil palm.

The alternative to micropropagation of oil palm can be a reverse breeding, which is a plant breeding technique to produce parental lines for any heterozygous plant (Dirks et al., 2009). The suppression of meiotic crossovers and transmission of non-recombinant chromosomes to haploid gametes is a key to reverse breeding. Gametes are subsequently regenerated as doubled-haploid offspring among which the parental lines are selected (Wijnker et al., 2014). Thanks to the knowledge of the oil palm gene space (Low et al., 2014), meiosis regulators, such as DMC1orthologs

(Wijnker et al., 2012) can be identified and then controlled. The frequencies of spontaneous haploids in seed progeny are very low, 1,100 haploids among 60 million seedlings (Dunwell et al., 2010). Cultures of oil palm microspores have not yielded doubled-haploids so far (Corley and Tinker, 2003). The totipotency of the male gametophyte is thought to be negatively regulated by a histone deacetylase-dependent mechanism, which is affected by the stress treatments, such as cold or heat shock that are used to induce haploid embryo development in culture (Li et al., 2014). Two percent of oil palm microspores exposed to the low temperature and starvation stress initiated cell division and formed embryoids (Indrianto et al., 2014). It is likely that inhibitors of histone deacetylases, trichostatin A for example (Li et al., 2014), will further increase the efficiency of oil palm microspore reprogramming to somatic embryogenesis. Genotyping doubled-haploids using SNP arrays (Ting et al., 2014) could then enable reverse breeding of oil palms.

Yield Gap Caused by the Diseases

Two major diseases threaten oil palm industry. In South-East Asia, the basal stem rot disease that is caused by the white rot fungi of the genus *Ganoderma* spp. is the major problem (Paterson, 2007; Rees et al., 2009). Nearly 60% of plantations in Malaysia reported the diseased trees. The basal stem rot is lethal, infected plants stop producing fruit and eventually die. The average tree mortality rate of 3.7% is equivalent to losses of US\$ 570 million per year (Mohammed et al., 2014). White rot fungi are characterized as facultative saprophytes, which are generally difficult to control. There are no good sources of natural genetic disease resistance neither amongst African, nor American oil palm accessions (Durand-Gasselin et al., 2005). The research efforts have focused on a more detailed understanding of the molecular defense responses in those plant-pathogen interactions, with a hope to find practical solutions to control the disease (Ho and Tan, 2014). To make cellulose available, white rot fungi are capable of degrading lignin to carbon dioxide and water (Paterson, 2007). Thus, understanding lignin biosynthesis in oil palm is of interest, whilst lignin structure modification by breeding could result in genetic resistance. Candidate genes for breeding basal stem rot resistance are being looked for amongst transcripts and proteins that alter their expression patterns upon infection (Ho and Tan, 2014). Current practical solutions are biological control with *Trichoderma* spp. fungi, palm endophytes and implementation of correct agronomical and phytosanitary practices (Mohammed et al., 2014).

The mysterious and devastating disease known as a fatal yellowing (transliterated from Portuguese “amarelecimento fatal”) or lethal bud rot (“pudrición de cogollo,” Spanish) is considered a major problem for the oil palm industry in Latin America (Chinchilla, 2008). Entire estates in Panama, Colombia, Suriname, Brazil, and Ecuador were destroyed by the disease (De Franqueville, 2003). Fatal yellowing has variable symptoms, which causes considerable confusion in a research field (Corley and Tinker, 2003). In spite of numerous efforts testing candidate fungi, bacteria, phytoplasma and viroids, there is no conclusive evidence that a phytopathogen is the primary cause of the disease.

¹⁴<http://www.palmelit.com/>

¹⁵<http://www.asd-cr.com/>

Microorganisms are thought to play rather an opportunistic role in the development of the disease that is primed by certain environmental factors. Remarkably, the root system development is altered even before affected palms show symptoms in the shoot (Chinchilla, 2008). Morphological and histological study showed that contrary to healthy-looking palms, diseased palms from Ecuador and Brazil did not have roots with soft and white tips, the so-called fine root system. Only a few meristematic cells could be detected in the apical shoot and root meristems, indicating cell cycle arrest (Kastlein et al., 1990). This finding can be validated by using reporters of the entry into the M-phase of the cell cycle. Chemical screening to promote re-activation of the cell cycle in affected roots could result in a treatment of the fatal yellowing disease.

Elaeis oleifera shows resistance to fatal yellowing. The trait is dominant in interspecific *E. oleifera* × *E. guineensis* F1 hybrids. In spite of the lower productivity and the need for manual pollination, F1 hybrid seeds are produced on commercial scale for planting. Otherwise, we are not aware of any systematic efforts to introgress *E. oleifera* fatal yellowing genetic resistance onto African oil palm genetic background. Most probably because there is no causality link between (a)biotic factors and disease development, which makes the screening procedures unpredictable. On a contrary, ASD Costa Rica released for sale the seeds of an unusual interspecific hybrid variety AMAZON¹⁶, which is rather an introgression of the *E. guineensis* productivity traits onto American oil palm genetic background.

Palm Oil Composition and Content

Along with coconut oil, crude palm oil, and particularly kernel palm oil, are some of the few highly saturated vegetable fats. On average, crude palm oil contains 44% palmitic acid (C16:0), 5% stearic acid (C18:0) and traces of myristic acid (C14:0), which together constitute half of FAs found in triacylglycerols (TAG) synthesized by the *E. guineensis* fruit mesocarp (Sambanthamurthi et al., 2000). TAG unsaturated FAs are represented by 40% of oleic acid (C18:1), 10% linoleic acid (C18:2) and traces of linolenic acid (C18:3). Food industries consume eighty percent of palm oil, also as a replacement for *trans*-FAs. Oleochemical industry manufacture soaps, detergents, lubricants, solvents, bioplastics and biodiesel from the remaining twenty percent of palm oils.

Dietary FAs play significant roles in the cause and prevention of cardiovascular disease. *Trans*-FAs from partially hydrogenated vegetable oils have well-established adverse effects and should be eliminated from the human diet (Michas et al., 2014). Palm oil may be an unhealthy fat, because of its high saturated FA content. Meta-analysis of 51 dietary intervention studies showed both favorable, and unfavorable changes in coronary heart disease and cardiovascular disease risk markers when palm oil was substituted for the primary dietary fats, whereas only favorable changes occurred when palm oil was substituted for *trans*-FAs (Fattore et al., 2014).

Higher degree of FA unsaturation is therefore a desirable characteristic to alter in palm oil. Iodine index, commonly used as unsaturation measure, varies from approximately 50–60% in *E. guineensis*, highest values measured for La Mé variety (Montoya et al., 2014), but was found to be anywhere between 70 and 80% in *E. oleifera* (Chavez and Sterling, 1991). Amongst the *E. oleifera* accessions, the unsaturated FA content ranges from 47 to 69% for C18:1, 2 to 19% for C18:2, and 0.1 to 1.2% for C18:3. Interspecies *E. oleifera* × *E. guineensis* hybrids planted in Latin America have a mid-parent phenotype with iodine index varying from 58 to 71% (Ong et al., 1981). Nineteen QTLs controlling FA composition were identified in the interspecies pseudo-backcross populations (Montoya et al., 2014). Importantly for breeding purposes, work of Montoya et al. (2014) indicates that FA composition is not linked to biomass yield traits. Mapping intra-gene SNPs in candidate genes related to the oleic acid C18:1 biosynthesis, supported several QTLs underlying acyl-ACP thioesterase type A (FATA) and $\Delta 9$ stearoyl-ACP desaturase (SAD) (Montoya et al., 2013). Genome sequence analysis identified the oil palm gene repertoire playing a role in FA biosynthesis, TAG assembly, carbon fluxes, fruit ripening and regulators of these processes (Singh et al., 2013b). In combination with the high-density genetic map (Ting et al., 2014) and further intra-gene SNP characterization, introgression of high oleic acid content from *E. oleifera* into varieties of *E. guineensis* is becoming a reality within a reach.

An alternative approach to increase oleic acid content at expense of palmitic acid relies on genetic engineering of key enzymes for palmitic acid synthesis, β -ketoacyl-ACP synthase II (KAS II) or palmitoyl-ACP thioesterase (Sambanthamurthi et al., 2009). KAS II activity was shown to be positively correlated with unsaturated FA content across palms from PORIM germplasm in Malaysia (Sambanthamurthi et al., 2009). Reducing *Arabidopsis* KAS II levels was found sufficient to convert its oilseed composition to that resembling palm-like tropical oil (Pidkowich et al., 2007). Mesocarp specific over-expression of KAS II gene and anti-sense RNA suppression of palmitoyl-ACP thioesterase have been undertaken and currently await evaluation (Sambanthamurthi et al., 2009).

The freshly pressed unrefined palm oil is also known as red palm oil due to its deep orangey-red color. Large amounts of carotenoids, predominantly α - and β -carotene, in a range 180–2500 $\mu\text{g g}^{-1}$ mesocarp dry weight are measured in African oil palm populations (Rajanaidu et al., 2000; Tranbarger et al., 2011). Even higher contents up to 4000 $\mu\text{g g}^{-1}$ mesocarp characterizes American oil palm accessions, whereas F1 interspecies hybrids have a mid-parent values of carotenoid contents (Sambanthamurthi et al., 2000). In terms of retinol equivalents (RE), standard batches of red palm oil have seventeen times more of β -carotene than carrots. A few grams, i.e., 1.5–6.5 table spoons, of red palm oil provides approximately 600 RE of β -carotene¹⁷, which is sufficient to meet daily vitamin A requirements in humans and to prevent childhood blindness from

¹⁶<http://www.asd-cr.com/>

¹⁷<https://hungermath.wordpress.com/2013/01/02/red-palm-oil-as-a-source-of-beta-carotene/>

vitamin A deficiency (Burri, 2012). Indeed, worldwide interventions studies demonstrated usefulness of oil palm supplementation to improve vitamin A status; a deficiency commonly experienced by poor communities in Asia, Africa and South America (Rice and Burns, 2010). Red palm oil is also rich in tocotrienol, which is an unsaturated form of natural vitamin E. Tocotrienols have health benefits due to antioxidative, antihypercholesterolemic, and antiangiogenic effects on disease prevention (Wong and Radhakrishnan, 2012). *E. guineensis* population having high vitamin E content are available as well (Kushairi et al., 2004).

Oil content, composition and oil accumulating cells types are major traits of economic interest (Durrett et al., 2008). To decipher molecular mechanisms that underpin those traits, oil palm fruit development represents an excellent experimental model to apply “omics” trait dissection. Correlation analysis of the transcriptome and metabolome data has been performed on mesocarp samples harvested at multiple time-points during fruit development (Bourgis et al., 2011; Tranbarger et al., 2011). For comparative purposes and to advance understanding of the carbon partitioning between storage carbohydrates and TAG, similar data sets were generated in date palm (*Phoenix dactylifera*), a closely related palm species that accumulates almost exclusively sugars rather than oil in fruit mesocarp (Bourgis et al., 2011). Similar conclusions were drawn by both studies (Bourgis et al., 2011; Tranbarger et al., 2011). The transcript abundance of the FA biosynthetic machinery was remarkably coordinated with oil deposition in mesocarp tissues during fruit maturation. In contrast, TAG assembly pathway enzymes showed very low or the lack of up-regulation during fruit maturation, indicating that TAG assembly is not rate limiting for oil accumulation. The comparative co-expression analysis with transcriptomes of *Arabidopsis*, corn and date palm further implicated the oil palm APETALA2 (AP2)/ETHYLENE RESPONSE FACTOR family transcription factor EgWRI1-1 in regulation of FA accumulation. EgWRI1-1 is homologous to WRINKLED1, a transcriptional regulator of glycolysis and FA synthesis in *Arabidopsis* embryos (Cernac and Benning, 2004). Interspecies genetic complementation indicated that palm and *Arabidopsis* genes could be functional orthologs (Ma et al., 2013).

Oil content and composition differs between oil palm fruit mesocarp, endosperm and embryo. At 5 months after pollination, dry mass of endosperm contained 50% of oil in which lauric acid (C12:0) was predominant FA. The major FAs of mesocarp oil were palmitic acid (C16:0) and oleic (C18:1) acids. The oil palm embryo also stored up to 27% of oil, which contained 25% of linoleic acid (C18:2) (Dussert et al., 2013). To understand the mechanisms behind such differences in oil content and FA composition, transcriptome and lipid profiles were compared during development of oil palm fruit. Accumulation of lauric acid in endosperm relied on up-regulation of a acyl-acyl carrier protein thioesterase and TAG assembly enzymes isoforms (Dussert et al., 2013). Three paralogs of WRINKLED1 were proposed as candidate regulators determining different lipid profiles. In agreement with (Tranbarger et al., 2011), *EgWRI1-1* was found to operate in mesocarp. *EgWRI1-2* and *EgWRI1-3* were predominantly expressed in endosperm. Interestingly,

embryo did not express either of *EgWRI1* paralogs (Dussert et al., 2013).

To provide new breeding material, targeted approaches, such as Ecotilling (Till et al., 2006) can be applied to screen oil palm germplasm collections for the loss or gain of function alleles in the identified subsets of lipid biosynthetic and regulatory genes. There are very few reports on induced mutagenesis in oil palm (Corley and Tinker, 2003). We are not aware of any chemically mutagenized oil palm populations that can enable standard TILLING (Till et al., 2006). Due to the 4-years-long seed-to-seed cycle, the practicality of such oil palm population of a few thousands trees that will live for the next 100 years appears to be doubtful. This view may change, pending the progress with production of doubled haploids from oil palm microspores (Indrianto et al., 2014). Mutagenizing microspores may result in a “tilling” population of trees homozygous at all genetic loci the analysis of which will add supportive evidences for the “omics” data.

Oil Palm Tree Architecture

Harvesting oil palm is expensive in manual labor, difficult task, compared with the ease of combine-harvesting arable crops (Corley and Tinker, 2003). The radical changes to the oil palm tree architecture are needed to enable development of harvesting machines. The major utility in harvesting mechanization were found to be palm height and bunch stalk length and thickness (Le Guen et al., 1990). A number of problems arise as palms age on a plantation. Fruit harvest is complicated when oil palm trees are taller than two-three meters. Cutting bunch stalk becomes physically challenging. Bunch fall bruises fruits many of which detach, demanding additional labor effort to collect loose fruits from the ground. Bruising activates TAG hydrolysis that lower oil quality. It is more difficult to assess the bunch ripeness. Though oil palm is a long lived species, replanting is thought to be required for plantations 20–25 years of age (Corley and Tinker, 2003).

In relation to the environment and genetic makeup, African oil palm accessions have height increment of 45–75 cm a year. Annual height increment of *E. oleifera* may be only 5–10 cm. Interspecific F1 hybrids between *E. guineensis* and *E. oleifera* have a mid-parent growth phenotype of 15–25 cm annual height increment (Corley and Tinker, 2003), indicating a useful gene introgression source.

Intraspecific variation enabled Malaysian breeding programs to develop PORIM series of dwarfish palms with a yield potential of 7t and annual height increment of 40 cm (Rajanaidu, 1994; Rajanaidu et al., 2000). Hybrids from crosses of Bamenda and Ekona *E. guineensis* accessions developed by ASD Costa Rica¹⁸ are slower growing (45–50 cm/year) and are also known for overall high cold and drought tolerance.

To find QTLs that control the tree growth rates, two breeding populations of oil palm were used for linkage mapping (Lee et al., 2015). For selected genotypes of the *dura* and *pisifera* parents, the heights of the 6-year-old *tenera* palms in F1 populations were distributed from 71.0 to 180 cm with an average of 137.6 cm.

¹⁸<http://www.asd-cr.com/>

The QTL positioned on a linkage group 5 explained 51.0% of the phenotypic variation, suggesting that it should play a major role in height variation of selected palm genotypes. Oil palm genome sequence indicated that QTL is located more precisely within 65.6 kb region that includes eight genes, of which the gene encoding asparagine synthase-related protein is thought to be responsible for the tree height variation amongst analyzed *teneras* (Lee et al., 2015). Along with glutamine synthases, asparagine synthases have an important role in nitrogen assimilation and allocation within the plant. Interestingly, the ectopically expressed pine glutamine synthase accelerates the growth of the poplar trees (Man et al., 2011).

In *E. oleifera* × *E. guineensis* F1 hybrid population, a wild palm was discovered that in addition to short trunk, had relatively short leaves due to spontaneous heritable change in leaf length. Derived breeding program resulted in commercial COMPACT¹⁹ varieties sold as clones. Compared to 7–8 meters long leaves of standard *tenera* hybrids, COMPACT palm leaves are 6.5 meters long, which allows a very high density planting, 180–200 trees per hectare. COMPACT palms have < 40 cm/year height increment. Micropropagated clones are rather expensive. Fortunately, the leaf length trait appears to be semi-dominant. Hybrids between COMPACT palms and standard *E. guineensis* lines, such as Deli, Ghana and Nigeria, have 6.6–6.9 meters long leaves and can be planted at density of 170 trees per hectare, which is still higher than industry standard of 138–143 palms/hectare (Corley and Tinker, 2003). As compared to clones, seeds of such hybrids are more affordable for smallholder farmers, who can find such genetic material as an opportunity to increase production and make better use of scarce land resources (Alvarado et al., 2010).

The increases in wheat and rice yields during the “Green Revolution,” were enabled by the introgression of dwarfing traits into the plants (Hedden, 2003). The “Green Revolution” genes showed the central role of gibberellin (GA) in the control of plant stature. Wheat *Reduced height* (*Rht*) genes interfere with the GA signal transduction pathway. The rice *semidwarf1* (*sd1*) gene impair the GA biosynthesis. *Arabidopsis* *GA5* gene is the ortholog of rice “Green Revolution” gene *SD1* (Barboza et al., 2013). Importantly, semidwarf individuals found in natural *Arabidopsis thaliana* populations were 21 different independent loss-of-function mutations at *GA5*. Semidwarfness had no obvious general tradeoff affecting *Arabidopsis* plant performance traits (Barboza et al., 2013). Semidwarfism transgenes modifying GA, promoting root growth and enhancing morphological diversity, have been tested in hybrid poplar trees (Elias et al., 2012). Analysis of mutants in cereals further implicated brassinosteroids in the control of plant architecture (Dockter and Hansson, 2015). These findings have direct implications to the gene-centered analysis of oil palm natural variation and tree architecture engineering.

Preventing (Post)Harvest Losses

The chemical properties of oils used in commerce are extremely important. The hydrolysis of TAG and release of free FAs has

a strong impact on the quality of commodity oil, because free FA content above five percent is thought to be unfit for human consumption (Ebongue et al., 2008). Oil palm mesocarp contains a highly active lipase that within five minutes can bring the free FA content to 30% in crushed tissue. The biological function of lipase in palm fruit mesocarp is uncertain. Importantly, TAG hydrolysis does not occur in undamaged fruits. It is critical to reduce fruit bruising before they reach the oil mill where the first step of post-processing is high pressure steam sterilization to inactivate both palm, and microbial lipases (Corley and Tinker, 2003).

To gain flexibility for post-harvest fruit processing and extended ripening for increased yields, elite oil palm lines with a low lipase (LL) were selected. Oil pressed from LL fruits had substantially less free FAs than standard genotypes (Ebongue et al., 2008). *E. guineensis* *LIPASE1*, *EgLIP1* gene associated with the LL trait has been identified, allowing marker-based introgression of the LL trait into any elite oil palm genotypes (Morcillo et al., 2013). Approximately 30% of oil palm cultivation worldwide, and up to 80% in Africa, is in smallholder farms. To extract the best quality oil, farmers have only limited period of time to deliver their produce to the oil extraction mills. Commercialization of the LL trees is estimated to generate economic gain of almost billion US dollars per year (Morcillo et al., 2013).

Elaeis oleifera has naturally very LL activity in fruit mesocarp, whilst interspecific *E. oleifera* × *E. guineensis* hybrids are promising crosses with less lipase activity (Cadena et al., 2013). Overall it appears that several *E. oleifera* traits, e.g., *virescens* fruits, LL, lack of methyl chavicol biosynthesis that could have been interpreted as loss-of-function mutations, behave at least as co-dominant or dominant alleles in interspecies hybrids. Molecular identification of genes controlling such traits (Morcillo et al., 2013; Singh et al., 2014) will help to reach mechanistic understanding of interspecies genome interactions.

Shedding of ripe fruits from the bunches before they reach the factories is an important source of harvest losses (Osborne et al., 1992). The development of the fruit abscission zone takes place at the base of subtending fruit. It is a two-stage process involving primary and adjacent abscission zones. Abscission zones contain very low amount of methylated pectin and high levels of polygalacturonase (PG) activity that is involved in the depolymerisation of the cell wall pectin homogalacturonan (Henderson et al., 2001). The oil palm PG family comprise at least 14 genes, of which ethylene-inducible *EgPG4* is the most highly expressed in the fruit base (Roongsattham et al., 2012). Altogether, we may anticipate breeding of palms with delayed, if not abolished, fruit shedding. The question becomes how to assess the fruit bunch ripeness for such palms in practice?

The majority of commercial *teneras* have *nigrescens*, i.e., anthocyanin colored, fruit exocarp. To determine that bunches on *nigrescens* palms are ripe, harvesters rely on the presence of detached fruits on the ground (Corley and Tinker, 2003; Singh et al., 2014). It is thought that scoring the profound change in color from green to bright orange upon ripening in *virescens* fruits could be an alternative solution to assess bunch ripeness in the field

¹⁹<http://www.asd-cr.com/>

(Sambanthamurthi et al., 2009). The *E. guineensis* *VIRESCENS* (*VIR*) gene is a R2R3-MYB transcription factor (Singh et al., 2014). The dominant-negative *virescens* phenotype is explained by the expression of the *VIR* protein isoforms truncated at carboxyl-terminus, which is likely to function as transcription activation domain. Interestingly, *E. oleifera* naturally has *virescens* fruit phenotype. Singh et al. (2014) cannot identify *VIR* homolog in a current *E. oleifera* draft genome assembly, indicating that *virescens* phenotype of *E. oleifera* fruit could be explained by a natural deletion mutation.

Breeding for Expanded Cultivation Range

Standard cultivation range of oil palm commercial varieties lays within 20° of the equator (Corley and Tinker, 2003). The cultivation range, as well as the breeding challenges, are expected to evolve due to the climate change (Paterson et al., 2013). Expansion of cultivation range to sub-tropical regions (Lei et al., 2014) could lower the negative impact of the crop on tropical biodiversity.

Cold tolerance was observed in natural oil palm groves situated at 1000–2000 meters above sea level in the Bamenda Highlands of Cameroon and in Kigoma District, Tanzania (Blaak and Stirling, 1996). A breeding program by ASD de Costa Rica²⁰ and FAO²¹, using Bamenda and Kigoma germplasms, led to new cold tolerant commercial varieties that in addition showed precocity when planted at sea level, producing fruit at 2 years after planting (Chapman et al., 2003). With a goal to gain molecular understanding of cold stress response and to expand African oil palm cultivation to sub-tropical regions, including Hainan province located in the southern China, the cold stress response in oil palm was analyzed by deep RNA sequencing (Lei et al., 2014). Work revealed 51,452 expressed sequences from *E. guineensis*. Transcriptome data analysis resulted in discovery of 5791 gene-based simple sequence repeats (SSRs) markers of which 916 distinguished genes differentially expressed in response to cold stress (Xiao et al., 2014).

Breeding African oil palm for stress tolerance, especially drought tolerance, has been found to be challenging (Corley and Tinker, 2003). The alternative to lower the environmental footprint of the crop, whilst meeting the growing demand for vegetable oils, is to domesticate other oleaginous palm species that have different profile of ecophysiological adaptations.

To address the problem of fatal yellowing disease that restricts oil palm cultivation in Latin America, breeders of the ASD Costa Rica introduced a new hybrid variety AMAZON. The mother trees are *E. oleifera*, originating from wild palms indigenous to the Manaus region (Amazonas state, Brazil). The *pisifera* parents were selected from the progeny of *E. oleifera* × *E. guineensis* interspecies hybrid backcrossed to *E. guineensis*. In the AMAZON hybrid, *E. guineensis* genome is smaller than haploid in size. This work is a pioneering step toward domestication of *E. oleifera* through gene introgression for higher yield.

Acrocomia aculeata known as macaw palm, or macaúba in Portuguese, is particularly interesting for the development of a new oleaginous crop both because of high productivity potential, simpler harvest, and ability to grow in arid sub-tropical areas (Pires et al., 2013). Whilst collecting *E. oleifera* accessions in Latin America, Malaysian breeders added to their *ex situ* germplasm collection several other oleaginous palms, such as *Oenocarpus* spp., *Bactris gasipaes* (Corley and Tinker, 2003). It is beyond the doubt that the recent advances in understanding the oil palm genome, physiology, identification of key genes controlling productivity and TAG biosynthesis will accelerate the domestication breeding programs.

Concluding Remarks

In this review we focused on the oil palm genetic diversity and how modern genomics tools could contribute both to the basic understanding of the physiology, metabolism and development of this remarkable crop, and to accelerate breeding of the high yielding varieties with a tailored oil composition.

Many of the discussed traits could be engineered using genetic modification of the crop (Murphy, 2014). Of particular interest is the success with regeneration of plants from protoplast cultures that were shown to be the superior starting material for PEG-mediated DNA transfection and microinjection (Masani et al., 2014). This system could be an excellent recipient to implement genome editing technology (Joung and Sander, 2013; Shan et al., 2013). For instance, instead of time consuming gene introgression, *EgLIP1* gene can be destroyed in high oil yield *tenera* elite palm, or its *dura* and *pisifera* parents. High oleic acid content and the ability to synthesize polyunsaturated fatty acids (PUFA) could be similarly engineered using simpler *Agrobacterium*-mediated genetic modification (Murphy, 2014). Those traits in combination with high pro-vitamin A and vitamin E content could result in new palm varieties for the extraction of the virgin red palm oil of unprecedented nutritional quality serving millions, if not billions of people from impoverished countries. As of today, palm oil is the only non-GMO oil on a global market, which some activists believe increases the palm oil value.

Plant breeders will continue with the efforts to increase primary crop productivity, however, the immediate challenges in closing the yield gap lay in providing smallholder farmers with the access to the best planting material, with balanced application of fertilizers and overall corrects agronomical practices. The negative impact of the crop on biodiversity is undeniable; the high yielding varieties are available to spare the land. It is the socio-economical drivers, governmental decision and environmental activists views that make oil palms the “Palms of Controversies” (Rival and Levang, 2014).

Acknowledgments

We thank reviewers for constructive criticism and suggestions that allowed us to improve the quality of the text. We apologize for the colleagues whose valuable contribution to the oil palm breeding and improvement was not referred to due to the space limitation.

²⁰<http://www.asd-cr.com>

²¹<http://www.fao.org>

References

Adam, H., Collin, M., Richaud, F., Beule, T., Cros, D., Omore, A., et al. (2011). Environmental regulation of sex determination in oil palm: current knowledge and insights from other species. *Ann. Bot.* 108, 1529–1537. doi: 10.1093/aob/mcr151

Aguirrezaabal, L. A., Lavaud, Y., Dosio, G. A., Izquierdo, N. G., Andrade, E. H., and González, L. M. (2003). Intercepted solar radiation during seed filling determines sunflower weight per seed and oil concentration. *Crop Sci.* 43, 152–161. doi: 10.2135/cropsci2003.1520

Alvarado, A., Escobar, R., and Peralta, F. (2010). ASD's oil palm breeding program and its contributions to the oil palm industry. *ASD Oil Palm Papers* 34, 1–16.

Amblard, P., Noiret, J.-N., Kouamé, B., Potier, F., and Adon, B. (1995). Performances comparées des hybrides interspécifiques et du matériel commercial *E. guineensis*. *Oléagineux Corps Gras Lipides* 2, 335–340.

Bakoumé, C., Wickneswari, R., Siju, S., Rajanaidu, N., Kushairi, A., and Billotte, N. (2015). Genetic diversity of the world's largest oil palm (*Elaeis guineensis* Jacq.) field genebank accessions using microsatellite markers. *Genet. Resour. Crop Evol.* 62, 349–360. doi: 10.1007/s10722-014-0156-8

Barboza, L., Effgen, S., Alonso-Blanco, C., Kooke, R., Keurentjes, J. J. B., Koornneef, M., et al. (2013). *Arabidopsis* semidwarfs evolved from independent mutations in GA20ox1, ortholog to green revolution dwarf alleles in rice and barley. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15818–15823. doi: 10.1073/pnas.1314979110

Barcelos, E. (1986). *Características genético-ecológicas de populações naturais de caiauá (Elaeis oleifera (H.B.K.) Cortés) na Amazônia Brasileira*. Manaus, AM: Instituto Nacional de Pesquisas da Amazônia, Fundação Universidade do Amazonas.

Barcelos, E. (1998a). *Etude de la diversité géaétique du genre Elaeis (E. oleifera (Kunth) Cortés et E. guineensis Jacq.) par marqueurs moléculaires (RFLP et AFLP)*. Ph.D. thesis, Université Montpellier, Montpellier.

Barcelos, E. (1998b). *Etude de la diversité génétique du genre Elaeis (E. oleifera (Kunth) Cortés et E. guineensis Jacq.) par marqueurs moléculaires (RFLP et AFLP)*. Ph.D. thesis, Ecole Nationale Supérieure Agronomique, Montpellier.

Barcelos, E., Amblard, P., Berthaud, J., and Seguin, M. (2002). Genetic diversity and relationship in American and African oil palm as revealed by RFLP and AFLP molecular markers. *Pesqui. Agropecu. Brasileira* 37, 1105–1114. doi: 10.1590/S0100-204X2002000800008

Barcelos, E., Second, G., Kahn, F., Amblard, P., Lebrun, P., and Seguin, M. (1999). Molecular markers applied to the analysis of genetic diversity and to the biogeography of *Elaeis* (Palmae). *Mem. N. Y. Bot. Gard. N. Y.* 83, 191–201.

Barnes, A. D., Jochum, M., Mumme, S., Haneda, N. F., Farajallah, A., Widarto, T. H., et al. (2014). Consequences of tropical land use for multitrophic biodiversity and ecosystem functioning. *Nat. Commun.* 5, 5351. doi: 10.1038/ncomms6351

Basri Wahid, M., Abdullah, S. N. A., and Henson, I. (2005). Oil palm—achievements and potential. *Plant Prod. Sci.* 8, 288–297. doi: 10.1626/pp.8.288

Beirnaert, A. D. F., and Vanderwegen, R. (1941). *Contribution à l'étude génétique et biométrique des variétés d'Elaeis guineensis Jacquin*. Bruxelles: I.N.E.A.C.

Billotte, N., Jourjon, M. F., Marseillac, N., Berger, A., Flori, A., Asmady, H., et al. (2010). QTL detection by multi-parent linkage mapping in oil palm (*Elaeis guineensis* Jacq.). *Theor. Appl. Genet.* 120, 1673–1687. doi: 10.1007/s00122-010-1284-y

Blaak, G., and Stirling, F. (1996). The prospects of extending oil palm cultivation to higher elevations through using cold-tolerant plant material. *Planter* 72, 645–648, 650, 652.

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

Burri, B. J. (2012). Evaluating global barriers to the use of red palm oil as an intervention food to prevent vitamin A deficiency. *Compr. Rev. Food Sci. Food Saf.* 11, 221–232. doi: 10.1111/j.1541-4337.2011.00181.x

Cadena, T., Prada, F., Perea, A., and Romero, H. M. (2013). Lipase activity, mesocarp oil content, and iodine value in oil palm fruits of *Elaeis guineensis*, *Elaeis oleifera*, and the interspecific hybrid OxG (*E. oleifera* × *E. guineensis*). *J. Sci. Food Agric.* 93, 674–680. doi: 10.1002/jsfa.5940

Carlson, K. M., Curran, L. M., Ratnasari, D., Pittman, A. M., Soares-Filho, B. S., Asner, G. P., et al. (2012). Committed carbon emissions, deforestation, and community land conversion from oil palm plantation expansion in West Kalimantan, Indonesia. *Proc. Natl. Acad. Sci. U.S.A.* 109, 7559–7564. doi: 10.1073/pnas.1200452109

Cernac, A., and Benning, C. (2004). WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in *Arabidopsis*. *Plant J.* 40, 575–585. doi: 10.1111/j.1365-313X.2004.02235.x

Chapman, K., Escobar, R., and Griffee, P. (2003). Cold tolerant or altitude adapted oil palm hybrid development initiatives in Asia/Pacific region. *AU. J. T.* 6, 134–138.

Chavez, C., and Sterling, F. (1991). Variation in the total of unsaturated fatty acids in oil extracted from different oil palm germplasm. *ASD Oil Palm Papers* 3, 5–8.

Chen, Z. J. (2013). Genomic and epigenetic insights into the molecular bases of heterosis. *Nat. Rev. Genet.* 14, 471–482. doi: 10.1038/nrg3503

Chinchilla, C. (2008). The many faces of spear rots in oil palm: the need for an integrated management approach. *ASD Oil Palm Papers* 32, 1–25.

Cochard, B., Adon, B., Rekima, S., Billotte, N., De Chenon, R. D., Koutou, A., et al. (2009). Geographic and genetic structure of African oil palm diversity suggests new approaches to breeding. *Tree Genet. Genomes* 5, 493–504. doi: 10.1007/s11295-009-0203-3

Corley, R. H. V. (1998). "What is the upper limit to oil extraction ratio?," in *Proceedings of International Conference on Oil and Kernel Production in Oil Palm—A Global Perspective*, eds N. Rajanaidu, I. E. Henson, and B. S. Jalani (Kuala Lumpur: Palm Oil Research Institute of Malaysia), 256–269.

Corley, R. H. V. (2009). How much palm oil do we need? *Environ. Sci. Policy* 12, 134–139. doi: 10.1016/j.envsci.2008.10.011

Corley, R. H. V., and Tinker, P. B. (2003). *The Oil Palm*. Oxford: John Wiley & Sons. doi: 10.1002/9780470750971

De Franquerville, H. (2003). Oil palm bud rot in Latin America. *Exp. Agric.* 39, 225–240. doi: 10.1017/S0014479703001315

de Vries, S. C., Van De Ven, G. W., Van Ittersum, M. K., and Giller, K. E. (2010). Resource use efficiency and environmental performance of nine major biofuel crops, processed by first-generation conversion techniques. *Biomass Bioenergy* 34, 588–601. doi: 10.1016/j.biombioe.2010.01.001

Devuyst, A. (1953). Selection of the oil palm (*Elaeis guineensis*) in Africa. *Nature* 172, 685–686. doi: 10.1038/172685a0

Dirks, R., Van Dun, K., De Snoo, C. B., Van Den Berg, M., Lelivelt, C. L., Voermans, W., et al. (2009). Reverse breeding: a novel breeding approach based on engineered meiosis. *Plant Biotechnol. J.* 7, 837–845. doi: 10.1111/j.1467-7652.2009.00450.x

Dockter, C., and Hansson, M. (2015). Improving barley culm robustness for secured crop yield in a changing climate. *J. Exp. Bot.* doi: 10.1093/jxb/eru521 [Epub ahead of print].

Dunwell, J. M., Wilkinson, M. J., Nelson, S., Wening, S., Sitorus, A. C., Mienanti, D., et al. (2010). Production of haploids and doubled haploids in oil palm. *BMC Plant Biol.* 10:218. doi: 10.1186/1471-2229-10-218

Durand-Gasselin, T., Asmady, H., Flori, A., Jacquemard, J. C., Hayun, Z., Breton, F., et al. (2005). Possible sources of genetic resistance in oil palm (*Elaeis guineensis* Jacq.) to basal stem rot caused by *Ganoderma boninense*—prospects for future breeding. *Mycopathologia* 159, 93–100. doi: 10.1007/s11046-004-4429-1

Durand-Gasselin, T., Noiret, J., Kouamé, K., Cochard, B., and Adon, B. (1999). Availability of quality pollen for improved oil palm (*Elaeis guineensis* Jacq.) seed production. *Plant. Res. Dév.* 6, 264–276.

Durrett, T. P., Benning, C., and Ohlrogge, J. (2008). Plant triacylglycerols as feedstocks for the production of biofuels. *Plant J.* 54, 593–607. doi: 10.1111/j.1365-313X.2008.03442.x

Dussert, S., Guerin, C., Andersson, M., Joët, T., Tranbarger, T. J., Pizot, M., et al. (2013). Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. *Plant Physiol.* 162, 1337–1358. doi: 10.1104/pp.113.220525

Ebongue, G. F. N., Konna, P., Nouy, B., Zok, S., Carrière, F., Zollo, P.-H. A., et al. (2008). Identification of oil palm breeding lines producing oils with low acid values. *Eur. J. Lipid Sci. Technol.* 110, 505–509. doi: 10.1002/ejlt.200700263

Elias, A. A., Busov, V. B., Kosola, K. R., Ma, C., Etherington, E., Shevchenko, O., et al. (2012). Green revolution trees: semidwarfism transgenes modify gibberellins, promote root growth, enhance morphological diversity, and reduce competitiveness in hybrid poplar. *Plant Physiol.* 160, 1130–1144. doi: 10.1104/pp.112.200741

Escobar, R. (1981). "Preliminary results of the collection and evaluation of the American oil palm *Elaeis oleifera* (H.B.K.) Cortes in Costa Rica, Panama and Colombia," in *The Oil Palm in Agriculture in the Eighties*, eds E. Pushparajah and P. Chew (Kuala Lumpur: Incorporated Society of Planters), 17–20.

F.A.O. (2010). *The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture*. Rome: Food and Agriculture Organization of the United Nations.

Fattore, E., Bosetti, C., Brightenti, F., Agostoni, C., and Fattore, G. (2014). Palm oil and blood lipid-related markers of cardiovascular disease: a systematic review and meta-analysis of dietary intervention trials. *Am. J. Clin. Nutr.* 113:081190. doi: 10.3945/ajcn.113.081190

Fearnside, P. M. (2001). Soybean cultivation as a threat to the environment in Brazil. *Environ. Conserv.* 28, 23–38. doi: 10.1017/S0376892901000030

Feintrenie, L., Chong, W. K., and Levang, P. (2010). Why do farmers prefer oil palm? Lessons learnt from Bungo district, Indonesia. *Small Scale Forestry* 9, 379–396. doi: 10.1007/s11842-010-9122-2

Fitzherbert, E. B., Struebig, M. J., Morel, A., Danielsen, F., Bruhl, C. A., Donald, P. F., et al. (2008). How will oil palm expansion affect biodiversity? *Trends Ecol. Evol.* 23, 538–545. doi: 10.1016/j.tree.2008.06.012

Foley, J. A., Defries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., et al. (2005). Global consequences of land use. *Science* 309, 570–574. doi: 10.1126/science.1111772

Godar, J., Gardner, T. A., Tizado, E. J., and Pacheco, P. (2014). Actor-specific contributions to the deforestation slowdown in the Brazilian Amazon. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15591–15596. doi: 10.1073/pnas.1322825111

Gomes, S. M. (2011). *Polinizadores e semiquímicos de dendzeiro híbrido (Elaeis oleifera (H.B.K.) Cortes X Elaeis guineensis Jacq.)*. Ph.D. thesis, Universidade Federal de Vicosa, Viçosa.

Green, R. E., Cornell, S. J., Scharlemann, J. P., and Balmford, A. (2005). Farming and the fate of wild nature. *Science* 307, 550–555. doi: 10.1126/science.1106049

Gutiérrez-Vélez, V. H., Defries, R., Pinedo-Vásquez, M., Uriarte, M., Padoch, C., Baethgen, W., et al. (2011). High-yield oil palm expansion spares land at the expense of forests in the Peruvian Amazon. *Environ. Res. Lett.* 6, 044029. doi: 10.1088/1748-9326/6/4/044029

Hardon, J. (1968). Breeding and selection of the oil palm in Malaya. *Oléagineux* 23, 85–90.

Hardon, J., and Tan, G. (1969). Interspecific hybrids in the genus *Elaeis* I. crossability, cytogenetics, and fertility of F1 hybrids of *E. guineensis* (× *E. oleifera*). *Euphytica* 18, 372–379. doi: 10.1007/BF00397784

Hartley, C. W. S. (1977). *The Oil Palm: (Elaeis Guineensis Jacq.)*. London: Longman.

Hedden, P. (2003). The genes of the Green Revolution. *Trends Genet.* 19, 5–9. doi: 10.1016/S0168-9525(02)00009-4

Heidmann, I., De Lange, B., Lambalk, J., Angenent, G. C., and Boutilier, K. (2011). Efficient sweet pepper transformation mediated by the BABY BOOM transcription factor. *Plant Cell Rep.* 30, 1107–1115. doi: 10.1007/s00299-011-1018-x

Henderson, J., Davies, H. A., Heyes, S. J., and Osborne, D. J. (2001). The study of a monocotyledon abscission zone using microscopic, chemical, enzymatic and solid state 13 C CP/MAS NMR analyses. *Phytochemistry* 56, 131–139. doi: 10.1016/S0031-9422(00)00447-7

Ho, C.-L., and Tan, Y.-C. (2014). Molecular defense response of oil palm to *Ganoderma* infection. *Phytochemistry* doi: 10.1016/j.phytochem.2014.10.016 [Epub ahead of print].

Hussein, M., Lajis, N., Kinson, A., and Teo, C. (1989). Laboratory and field evaluation on the attractancy of *Elaeidobius kamerunicus* Faust to 4-allylanisole. *Porum Bull.* 18, 20–26.

Indrianto, A., Mariani, T. S., and Sari, D. A. (2014). Induction of embryogenic microspore in oil palm (*Elaeis guineensis* Jacq) by starvation and temperature stress. *Asian J. Appl. Sci.* 2, 668–677.

Ishikawa, S., Ishimaru, Y., Igura, M., Kuramata, M., Abe, T., Senoura, T., et al. (2012). Ion-beam irradiation, gene identification, and marker-assisted breeding in the development of low-cadmium rice. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19166–19171. doi: 10.1073/pnas.1211132109

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

Kastelein, P., Van Slokke, W., and De Leeuw, G. (1990). Symptomatological and histopathological observations on oil palms from Brazil and Ecuador affected by fatal yellowing. *Netherlands J. Plant Pathol.* 96, 113–117. doi: 10.1007/BF02005135

Kushairi, A., and Rajanaidu, N. (2000). *Breeding Population, Seed Production and Nursery Management*. Kuala Lumpur: Malaysian Palm Oil Board.

Kushairi, A., Rajanaidu, N., Sundram, K., and Maizura, I. (2004). *PS8: High Vitamin E Breeding Population*. Bangi: MPOB Information Series.

Lajis, M. N., Hussein, M. Y., and Toia, R. F. (1985). Extraction and identification of the main compound present in *Elaeis guineensis* flower volatiles. *Pertanika* 8, 105–108.

Laurance, W. F., Sayer, J., and Cassman, K. G. (2014). Agricultural expansion and its impacts on tropical nature. *Trends Ecol. Evol.* 29, 107–116. doi: 10.1016/j.tree.2013.12.001

Lee, M., Xia, J. H., Zou, Z., Ye, J., Rahmadsyah, Alfiko, Y., Jin, J., et al. (2015). A consensus linkage map of oil palm and a major QTL for stem height. *Sci. Rep.* 5, 8232. doi: 10.1038/srep08232

Le Guen, V., Ouattara, S., and Jacquemard, J. C. (1990). Oil palm selection with a view to easier harvesting. Initial results. *Oléagineux* 45, 523–531.

Legros, S., Miallet-Serra, I., Clement-Vidal, A., Caliman, J. P., Siregar, F. A., Fabre, D., et al. (2009). Role of transitory carbon reserves during adjustment to climate variability and source-sink imbalances in oil palm (*Elaeis guineensis*). *Tree Physiol.* 29, 1199–1211. doi: 10.1093/treephys/tp9057

Lei, X., Xiao, Y., Xia, W., Mason, A. S., Yang, Y., Ma, Z., et al. (2014). RNA-seq analysis of oil palm under cold stress reveals a different C-repeat binding factor (CBF) mediated gene expression pattern in *Elaeis guineensis* compared to other species. *PLoS ONE* 9:e114482. doi: 10.1371/journal.pone.0114482

Li, H., Soriano, M., Cordewener, J., Muino, J. M., Riksen, T., Fukuoka, H., et al. (2014). The histone deacetylase inhibitor trichostatin A promotes totipotency in the male gametophyte. *Plant Cell* 26, 195–209. doi: 10.1105/tpc.113.116491

Lima, A. B., Cannavan, F. S., Navarrete, A. A., Teixeira, W. G., Kuramae, E. E., and Tsai, S. M. (2014). Amazonian Dark Earth and plant species from the Amazon region contribute to shape rhizosphere bacterial communities. *Microb. Ecol.* [Epub ahead of print].

Low, E. T., Rosli, R., Jayanthi, N., Mohd-Amin, A. H., Azizi, N., Chan, K. L., et al. (2014). Analyses of hypomethylated oil palm gene space. *PLoS ONE* 9:e86728. doi: 10.1371/journal.pone.0086728

Ma, W., Kong, Q., Arondel, V., Kilaru, A., Bates, P. D., Thrower, N. A., et al. (2013). WRINKLED1, a ubiquitous regulator in oil accumulating tissues from *Arabidopsis* embryos to oil palm mesocarp. *PLoS ONE* 8:e68887. doi: 10.1371/journal.pone.0068887

Maley, J., and Chepstow-Lusty, A. (2001). *Elaeis guineensis* Jacq. (oil palm) fluctuations in central Africa during the late Holocene: climate or human driving forces for this pioneering species? *Veg. Hist. Archaeobot.* 10, 117–120. doi: 10.1007/PL00006920

Man, H., Pollmann, S., Weiler, E. W., and Kirby, E. G. (2011). Increased glutamine in leaves of poplar transgenic with pine GS1a caused greater anthranilate synthetase (α-subunit (ASA1) transcript and protein abundances: an auxin-related mechanism for enhanced growth in GS transgenics? *J. Exp. Bot.* 62, 4423–4431. doi: 10.1093/jxb/err026

Masani, M. Y. A., Noll, G. A., Parveez, G. K. A., Sambanthamurthi, R., and Prüfer, D. (2014). Efficient transformation of oil palm protoplasts by PEG-mediated transfection and DNA microinjection. *PLoS ONE* 9:e96831. doi: 10.1371/journal.pone.0096831

Meunier, J. (1969). Etude des populations naturelles d'*Elaeis guineensis* en Côte d'Ivoire. *Oléagineux* 24, 195–201.

Meunier, J. (1975). Le palmier à huile américain *Elaeis melanococca*. *Oléagineux* 30, 51–61.

Michas, G., Micha, R., and Zampelas, A. (2014). Dietary fats and cardiovascular disease: putting together the pieces of a complicated puzzle. *Atherosclerosis* 234, 320–328. doi: 10.1016/j.atherosclerosis.2014.03.013

Mohammed, C. L., Rimbawato, A., and Page, D. E. (2014). Management of basidiomycete root- and stem-rot diseases in oil palm, rubber and tropical hardwood plantation crops. *Forest Pathol.* 44, 428–446. doi: 10.1111/efp.12140

Montoya, C., Cochard, B., Flori, A., Cros, D., Lopes, R., Cuellar, T., et al. (2014). Genetic architecture of palm oil fatty acid composition in cultivated oil palm (*Elaeis guineensis* Jacq.) compared to its wild relative *E. oleifera* (H.B.K) Cortes. *PLoS ONE* 9:e95412. doi: 10.1371/journal.pone.0095412

Montoya, C., Lopes, R., Flori, A., Cros, D., Cuellar, T., Summo, M., et al. (2013). Quantitative trait loci (QTLs) analysis of palm oil fatty acid composition in an interspecific pseudo-backcross from *Elaeis oleifera* (HBK) Cortés and oil palm (*Elaeis guineensis* Jacq.). *Tree Genet. Genomes* 9, 1207–1225. doi: 10.1007/s11295-013-0629-5

Morcillo, F., Cros, D., Billotte, N., Ngando-Ebongue, G. F., Domonhédó, H., Pizot, M., et al. (2013). Improving palm oil quality through identification and mapping of the lipase gene causing oil deterioration. *Nat. Commun.* 4, 2160. doi: 10.1038/ncomms3160

Murphy, D. J. (2009). Oil palm: future prospects for yield and quality improvements. *Lipid Technol.* 21, 257–260. doi: 10.1002/lite.200900067

Murphy, D. J. (2014). The future of oil palm as a major global crop: opportunities and challenges. *J. Oil Palm Res.* 26, 1–24.

Ong, S. H., Chuah, C. C., and Sow, H. P. (1981). The co-dominance theory: genetic interpretations of analyses of mesocarp oils from elaeis in *Elaeis guineensis*, *Elaeis oleifera*, and their hybrids. *J. Am. Oil Chemists Soc.* 58, 1032–1038. doi: 10.1007/BF02679320

Ooi, S., Barcelos, E., Muller, A., and Nascimento, J. (1981). Oil palm genetic resources—native *E. oleifera* populations in Brazil offer promising sources. *Pesqui. Agropecu. Brasileira* 16, 385–395.

Ooi, S., and Rajanaidu, N. (1979). Establishment of oil palm genetic resources—theoretical and practical considerations. *Malaysian Appl. Biol.* 8, 15–28.

Osborne, D. J., Henderson, J., and Corley, R. H. V. (1992). Controlling fruit-shedding in the oil palm. *Endeavour* 16, 173–177. doi: 10.1016/0160-9327(92)90044-P

Pallas, B., Clement-Vidal, A., Rebollo, M. C., Soulie, J. C., and Luquet, D. (2013). Using plant growth modeling to analyze C source-sink relations under drought: inter- and intraspecific comparison. *Front. Plant Sci.* 4:437. doi: 10.3389/fpls.2013.00437

Paterson, R. (2007). *Ganoderma* disease of oil palm—a white rot perspective necessary for integrated control. *Crop Protect.* 26, 1369–1376. doi: 10.1016/j.cropro.2006.11.009

Paterson, R. R. M., Sariah, M., and Lima, N. (2013). How will climate change affect oil palm fungal diseases? *Crop Protect.* 46, 113–120. doi: 10.1016/j.cropro.2012.12.023

Phalan, B., Bertzky, M., Butchart, S. H., Donald, P. F., Scharlemann, J. P., Stattersfield, A. J., et al. (2013). Crop expansion and conservation priorities in tropical countries. *PLoS ONE* 8:e51759. doi: 10.1371/journal.pone.0051759

Phalan, B., Green, R., and Balmford, A. (2014). Closing yield gaps: perils and possibilities for biodiversity conservation. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20120285. doi: 10.1098/rstb.2012.0285

Phalan, B., Onial, M., Balmford, A., and Green, R. E. (2011). Reconciling food production and biodiversity conservation: land sharing and land sparing compared. *Science* 333, 1289–1291. doi: 10.1126/science.1208742

Pidkowich, M. S., Nguyen, H. T., Heilmann, I., Ischebeck, T., and Shanklin, J. (2007). Modulating seed β -ketoacyl-acyl carrier protein synthase II level converts the composition of a temperate seed oil to that of a palm-like tropical oil. *Proc. Natl. Acad. Sci. U.S.A.* 104, 4742–4747. doi: 10.1073/pnas.0611141104

Pires, T. P., Souza, E. S., Kuki, K. N., and Motoike, S. Y. (2013). Ecophysiological traits of the macaw palm: a contribution towards the domestication of a novel oil crop. *Ind. Crops Prod.* 44, 200–201. doi: 10.1016/j.indcrop.2012.09.029

Rajanaidu, N. (1986). “*Elaeis oleifera* collection in Central and South America,” in *Proceedings of the International Workshop “Oil palm germplasm and utilisation”*. (Kuala Lumpur: Palm Oil Research Institute of Malaysia).

Rajanaidu, N. (1994). *PORIM Oil Palm Genebank: Collection, Evaluation, Utilization, and Conservation of Oil Palm Genetic Resources*. Kuala Lumpur: Palm Oil Research Institute of Malaysia.

Rajanaidu, N., Kushairi, A., Rafii, M., Din, M., Maizura, I., and Jalani, B. (2000). “Oil palm breeding and genetic resources,” in *Advances in Oil Palm Research*, eds Y. Basiron, B. S. Jalani, and K. W. Chan (Kuala Lumpur: Malaysian Palm Oil Board), 171–227.

Ramalho-Filho, A., Da Motta, P., Freitas, P., and Teixeira, W. (2010). *Zoneamento Agroecológico, Produção E Manejo Para a Cultura da Palma de Óleo na Amazônia*. Rio de Janeiro: EMBRAPA.

Rees, R. W., Flood, J., Hasan, Y., Potter, U., and Cooper, R. M. (2009). Basal stem rot of oil palm (*Elaeis guineensis*); mode of root infection and lower stem invasion by *Ganoderma boninense*. *Plant Pathol.* 58, 982–989. doi: 10.1111/j.1365-3059.2009.02100.x

Rice, A. L., and Burns, J. B. (2010). Moving from efficacy to effectiveness: red palm oil's role in preventing vitamin A deficiency. *J. Am. Coll. Nutr.* 29, 302S–313S. doi: 10.1080/07315724.2010.10719845

Rival, A., and Levang, P. (2014). *Palms of Controversies: Oil Palm and Development Challenges*. Bogor: Center for International Forestry Research. doi: 10.17528/cifor/004860

Roongsattham, P., Morcillo, F., Jantasuriyarat, C., Pizot, M., Moussu, S., Jayaweera, D., et al. (2012). Temporal and spatial expression of polygalacturonase gene family members reveals divergent regulation during fleshy fruit ripening and abscission in the monocot species oil palm. *BMC Plant Biol.* 12:150. doi: 10.1186/1471-2229-12-150

Rosenquist, E. A. (1986). “The genetic base of oil palm breeding populations,” in *Proceedings of the International Workshop “Oil Palm Germplasm and Utilisation”*. (Bangi: PORIM—Palm Oil Research Institute of Malaysia).

Sambanthamurthi, R., Cheang, O., and Parman, S. (1995). “Factors affecting lipase activity in the oil palm (*Elaeis guineensis*) mesocarp,” in *Plant Lipid Metabolism*, eds J.-C. Kader and P. Mazliak (Berlin: Springer), 555–557.

Sambanthamurthi, R., Singh, R., Kadir, A. P. G., Abdullah, M. O., and Kushairi, A. (2009). “Opportunities for the oil palm via breeding and biotechnology,” in *Breeding Plantation Tree Crops: Tropical Species*, eds S. Mohan Jain and P. M. Priyadarshan (New York: Springer), 377–421. doi: 10.1007/978-0-387-71201-7

Sambanthamurthi, R., Sundram, K., and Tan, Y. (2000). Chemistry and biochemistry of palm oil. *Prog. Lipid Res.* 39, 507–558. doi: 10.1016/S0163-7827(00)00015-1

Sayer, J., Ghazoul, J., Nelson, P., and Klintuni Boedihartono, A. (2012). Oil palm expansion transforms tropical landscapes and livelihoods. *Global Food Secur.* 1, 114–119. doi: 10.1016/j.gfs.2012.10.003

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

Sharma, M., and Tan, Y. (1997). Oil palm breeding programme and the performance of D \times P planting materials at United Plantations Berhad. *Planter* 73, 591–610.

Sheil, D., Casson, A., Meijaard, E., Van Noordwijk, M., Gaskell, J., Sunderland-Groves, J., et al. (2009). *The Impacts and Opportunities of Oil Palm in Southeast Asia: What do we know and what do we need to know?*. Bogor: Center for International Forestry Research. doi: 10.17528/cifor/002792

Singh, R., Low, E.-T. L., Ooi, L. C.-L., Ong-Abdullah, M., Ting, N.-C., Nagappan, J., et al. (2013a). The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. *Nature* 500, 340–344. doi: 10.1038/nature12356

Singh, R., Ong-Abdullah, M., Low, E. T., Manaf, M. A., Rosli, R., Nookiah, R., et al. (2013b). Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* 500, 335–339. doi: 10.1038/nature12309

Singh, R., Low, E. T., Ooi, L. C., Ong-Abdullah, M., Nookiah, R., Ting, N. C., et al. (2014). The oil palm VIRESCENS gene controls fruit colour and encodes a R2R3-MYB. *Nat. Commun.* 5, 4106. doi: 10.1038/ncomms5106

Singh, R., Tan, S. G., Panandam, J. M., Rahman, R. A., Ooi, L. C., Low, E. T., et al. (2009). Mapping quantitative trait loci (QTLs) for fatty acid composition in an interspecific cross of oil palm. *BMC Plant Biol.* 9:114. doi: 10.1186/1471-2229-9-114

Smith, B. (1989). The effects of soil water and atmospheric vapour pressure deficit on stomatal behaviour and photosynthesis in the oil palm. *J. Exp. Bot.* 40, 647–651. doi: 10.1093/jxb/40.6.647

Sowunmi, M. A. (1999). The significance of the oil palm (*Elaeis guineensis* Jacq.) in the late Holocene environments of west and west central Africa: a further consideration. *Veg. Hist. Archaeobot.* 8, 199–210. doi: 10.1007/BF02342720

Tan, Y. (1985). “Weevil pollination in the *Elaeis oleifera* \times *Elaeis guineensis* hybrid,” in *Proceedings of the Symposium “Impact of the Pollinating Weevil on the Malaysian Oil Palm Industry”*. (Kuala Lumpur: Palm Oil Research Institute of Malaysia).

Tandon, R., Manohara, T. N., Nijalingappa, B. H. M., and Shivanna, K. R. (2001). Pollination and pollen-pistil interaction in oil palm, *Elaeis guineensis*. *Ann. Bot.* 87, 831–838. doi: 10.1006/anbo.2001.1421

Till, B. J., Zerr, T., Comai, L., and Henikoff, S. (2006). A protocol for TILLING and Ecotilling in plants and animals. *Nat. Protocols* 1, 2465–2477. doi: 10.1038/nprot.2006.329

Ting, N. C., Jansen, J., Mayes, S., Massawe, F., Sambanthamurthi, R., Ooi, L. C., et al. (2014). High density SNP and SSR-based genetic maps of two independent oil palm hybrids. *BMC Genomics* 15:309. doi: 10.1186/1471-2164-15-309

Ting, N. C., Jansen, J., Nagappan, J., Ishak, Z., Chin, C. W., Tan, S. G., et al. (2013). Identification of QTLs associated with callogenesis and embryogenesis in oil palm using genetic linkage maps improved with SSR markers. *PLoS ONE* 8:e53076. doi: 10.1371/journal.pone.0053076

Tranbarger, T. J., Dussert, S., Joët, T., Argout, X., Summo, M., Champion, A., et al. (2011). Regulatory mechanisms underlying oil palm fruit mesocarp maturation, ripening, and functional specialization in lipid and carotenoid metabolism. *Plant Physiol.* 156, 564–584. doi: 10.1104/pp.111.175141

Villela, A. A., Jaccoud, D., Rosa, L. P., and Freitas, M. V. (2014). Status and prospects of oil palm in the Brazilian Amazon. *Biomass Bioenergy* 67, 270–278. doi: 10.1016/j.biombioe.2014.05.005

Villoria, N. B., Golub, A., Byerlee, D., and Stevenson, J. (2013). Will yield improvements on the forest frontier reduce greenhouse gas emissions? A global analysis of oil palm. *Am. J. Agric. Econ.* 95, 1301–1308. doi: 10.1093/ajae/aat034

Wich, S. A., Garcia-Ulloa, J., Kuhl, H. S., Humle, T., Lee, J. S., and Koh, L. P. (2014). Will oil palm's homecoming spell doom for Africa's great apes? *Curr. Biol.* 24, 1659–1663. doi: 10.1016/j.cub.2014.05.077

Wicke, B., Sikkema, R., Dornburg, V., and Faaij, A. (2011). Exploring land use changes and the role of palm oil production in Indonesia and Malaysia. *Land Use Policy* 28, 193–206. doi: 10.1016/j.landusepol.2010.06.001

Wijnker, E., Deurhof, L., Van De Belt, J., De Snoo, C. B., Blankestijn, H., Becker, F., et al. (2014). Hybrid recreation by reverse breeding in *Arabidopsis thaliana*. *Nat. Protocols* 9, 761–772. doi: 10.1038/nprot.2014.049

Wijnker, E., Van Dun, K., De Snoo, C. B., Lelivelt, C. L., Keurentjes, J. J., Naharudin, N. S., et al. (2012). Reverse breeding in *Arabidopsis thaliana* generates homozygous parental lines from a heterozygous plant. *Nat. Genet.* 44, 467–470. doi: 10.1038/ng.2203

Wong, C. K., and Bernardo, R. (2008). Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. *Theor. Appl. Genet.* 116, 815–824. doi: 10.1007/s00122-008-0715-5

Wong, R. S., and Radhakrishnan, A. K. (2012). Tocotrienol research: past into present. *Nutr. Rev.* 70, 483–490. doi: 10.1111/j.1753-4887.2012.00512.x

Xiao, Y., Zhou, L., Xia, W., Mason, A. S., Yang, Y., Ma, Z., et al. (2014). Exploiting transcriptome data for the development and characterization of gene-based SSR markers related to cold tolerance in oil palm (*Elaeis guineensis*). *BMC Plant Biol.* 14:384. doi: 10.1186/s12870-014-0384-2

Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., et al. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5, 5087. doi: 10.1038/ncomms6087

Zeven, A. C. (1972). The partial and complete domestication of the oil palm (*Elaeis guineensis*). *Econ. Bot.* 26, 274–279. doi: 10.1007/BF02861041

Conflict of Interest Statement: 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.

Copyright © 2015 Barcelos, Rios, Cunha, Lopes, Motoike, Babiychuk, Skirycz and Kushnir. 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) or licensor 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.

Genetic diversity in tef [*Eragrostis tef* (Zucc.) Trotter]

Kebebew Assefa¹, Gina Cannarozzi², Dejene Girma^{2,3}, Rizqah Kamies⁴, Solomon Chanyalew¹, Sonia Plaza-Wüthrich², Regula Blösch², Abiel Rindisbacher², Suhail Rafudeen⁴ and Zerihun Tadele^{2*}

¹ National Tef Research Program, Debre Zeit Agricultural Research Center, Ethiopian Institute of Agricultural Research, Debre Zeit, Ethiopia, ² Crop Breeding and Genomics, Institute of Plant Sciences, Department of Biology, University of Bern, Bern, Switzerland, ³ National Agricultural Biotechnology Laboratory, Holetta Agricultural Research Center, Ethiopian Institute of Agricultural Research, Holetta, Ethiopia, ⁴ Plant Stress Laboratory, Department of Molecular and Cell Biology, University of Cape Town, Cape Town, South Africa

OPEN ACCESS

Edited by:

Joanna Marie-France Cross,
Inonu University, Turkey

Reviewed by:

Sergio Lanteri,
University of Turin, Italy
Rodomiro Ortiz,
Swedish University of Agricultural
Sciences, Sweden

***Correspondence:**

Zerihun Tadele,
Crop Breeding and Genomics,
Institute of Plant Sciences,
Department of Biology, University of
Bern, Altenbergrain 21, 3013 Bern,
Switzerland
zerihun.tadele@ips.unibe.ch

Specialty section:

This article was submitted to Plant
Genetics and Genomics, a section of
the journal *Frontiers in Plant Science*

Received: 29 December 2014

Accepted: 05 March 2015

Published: 26 March 2015

Citation:

Assefa K, Cannarozzi G, Girma D,
Kamies R, Chanyalew S,
Plaza-Wüthrich S, Blösch R,
Rindisbacher A, Rafudeen S and
Tadele Z (2015) Genetic diversity in tef
[*Eragrostis tef* (Zucc.) Trotter].
Front. Plant Sci. 6:177.
doi: 10.3389/fpls.2015.00177

Tef [*Eragrostis tef* (Zucc.) Trotter] is a cereal crop resilient to adverse climatic and soil conditions, and possessing desirable storage properties. Although tef provides high quality food and grows under marginal conditions unsuitable for other cereals, it is considered to be an orphan crop because it has benefited little from genetic improvement. Hence, unlike other cereals such as maize and wheat, the productivity of tef is extremely low. In spite of the low productivity, tef is widely cultivated by over six million small-scale farmers in Ethiopia where it is annually grown on more than three million hectares of land, accounting for over 30% of the total cereal acreage. Tef, a tetraploid with 40 chromosomes ($2n = 4x = 40$), belongs to the family Poaceae and, together with finger millet (*Eleusine coracana* Gaertn.), to the subfamily Chloridoideae. It was originated and domesticated in Ethiopia. There are about 350 *Eragrostis* species of which *E. tef* is the only species cultivated for human consumption. At the present time, the gene bank in Ethiopia holds over five thousand tef accessions collected from geographical regions diverse in terms of climate and elevation. These germplasm accessions appear to have huge variability with regard to key agronomic and nutritional traits. In order to properly utilize the variability in developing new tef cultivars, various techniques have been implemented to catalog the extent and unravel the patterns of genetic diversity. In this review, we show some recent initiatives investigating the diversity of tef using genomics, transcriptomics and proteomics and discuss the prospect of these efforts in providing molecular resources that can aid modern tef breeding.

Keywords: *Eragrostis tef*, diversity, genomics, proteomics, tef, transcriptomics, variability

Introduction

Tef [*Eragrostis tef* (Zucc.) Trotter] is the major food crop in Ethiopia where it is annually cultivated on more than three million hectares of land (CSA, 2014). Compared to other cereals, tef is more tolerant to extreme environmental conditions especially to water-logging. It is unique in its ability to grow and yield on poorly drained Vertisols which most cereals cannot tolerate. Unlike other cereals, the seeds of tef can be easily stored under local storage conditions without losing viability since the grains are resistant to attack by storage pests (Ketema, 1997). Tef grain is also a rich source of protein and nutrients and has additional health benefits including that the seeds are free from

gluten (Spaenij-Dekking et al., 2005). According to a recent study, the bio-available iron content was significantly higher in tef bread than in wheat bread (Alaunyte et al., 2012). In general, tef provides quality food and grows under marginal conditions, many of which are poorly suited to other cereals. However, tef is considered to be an orphan crop since it is only of regional importance and has until recently not been the focus of crop improvement (Naylor et al., 2004; Assefa, 2014).

Despite its versatility in adapting to extreme environmental conditions, the productivity of tef is very low with the national average standing at 1.5 t/ha (CSA, 2014). Tef's major yield limiting factors are lack of cultivars tolerant to lodging, drought, and pests (Assefa et al., 2011). Lodging is the permanent displacement of the stem from the upright position. Tef possesses tall, weak stems that easily succumb to lodging caused by wind or rain. In addition, lodging hinders the use of high input husbandry since the application of increased amounts of nitrogen fertilizer to boost the yield results in severe lodging. When this occurs, both the yield and the quality of the grain and the straw are severely reduced and both manual and mechanical harvesting are impeded. Various attempts have been made by the research community to develop lodging-resistant tef cultivars (Assefa et al., 2011; Tadele and Assefa, 2012) but presently no cultivar with reasonable lodging resistance has been obtained.

The analysis of genetic relationships amongst tef varieties is an important component of improvement programs because it provides information about the genetic diversity of the crop and sets a platform for stratified sampling of breeding populations. Tef represents a unique biodiversity component in the agriculture and food security of millions of farmers in Ethiopia. The conservation, characterization, and utilization of the existing tef genetic diversity are becoming increasingly important in view of the evolving needs and manifold challenges of small-scale farmers in Ethiopia. This is primarily because tef has remarkable genetic traits useful for most Ethiopian farmers to utilize for coping with erratic climatic conditions, generation of household income, and fulfilling concerns of nutritional needs. Moreover, the conservation and utilization of the tef genetic resources offer a reliable basis for enhancing food security and developing crop diversification in the moisture stress and challenging agro-ecological areas of the country.

Here, we present an overview of the results of the major studies made on tef diversity and recent initiatives underway to better understand the diversity at molecular level and utilize these diversities in improving the crop using modern genetic and genomic tools.

Taxonomy and Accessions of Tef

Tef belongs to the Poaceae or Grass family as do all economically important cereals. It is closely related to finger millet (*Eleusine coracana* Gaerth.) as both are in the subfamily Chloridoideae. The genus *Eragrostis* comprises about 350 species from which only tef is cultivated for human consumption. Unlike wheat, barley and rice, which are all C₃ plants,

tef (along with maize and sorghum) is a C₄ plant which efficiently utilizes carbon dioxide during photosynthesis. This can be seen by tef's Kranz-type leaf anatomy with vascular centers surrounded by bundle sheath cells containing a high number of chloroplasts and by the low CO₂ compensation point of the leaves, also typical of C₄ as opposed to C₃ species (Kebede et al., 1989).

Tea is an allotetraploid (2n = 4x = 40). Over the past few decades the ancestry of tef has been investigated using morphological and cytogenetic methods (Jones et al., 1978), biochemical methods (Bekele and Lester, 1981), and phylogenetic analysis using ribosomal DNA and transcription factor genes (Espelund et al., 2000) or nuclear and plastid genes (Ingram and Doyle, 2003). It has been suggested that *Eragrostis pilosa* is closely related to tef while *E. heteromera* and *E. ciliaris* are more distantly related (Ingram and Doyle, 2003). Similar conclusions were reached using biochemical methods (Bekele and Lester, 1981). The close relationship between tef and *E. pilosa* is also evidenced by the successful hybridization of these two species (Tefera et al., 2003a). This hybridization generated viable offspring and ultimately resulted in the release in 2009 of a variety called *Simada* (DZ-Cr-285 RIL295) from the inter-specific hybrid of tef [DZ-01-2785 × *E. pilosa* (line 30-5); MoA, 2013]. However, since *E. pilosa*, like tef, is a tetraploid, the diploid ancestors of tef remain unknown.

Ethiopia is the origin and center of diversity for tef (Vavilov, 1951), harboring landraces with a wide array of phenotypic diversity, and also wild progenitors and related wild species. Charring experiments suggest that the domestication history of tef might be different from that of barley and wheat since in some cases tef might not survive the high temperatures tolerated by other cereals (D'Andrea, 2008).

As in any crop improvement program, tef breeding also relies mainly upon the germplasm resources existing in the genetic stock. Diverse types of accessions are available in the country, and collection, evaluation, and utilization of tef germplasm by national and international groups began in Ethiopia in the late 1950s. However, organized collection at the national level was made after the establishment of the Plant Genetic Resources Center of Ethiopia (PGRC/E) in 1976. After several changes in its name and mandate, the institute responsible for germplasm collection and maintenance as well as distribution is currently called the Ethiopian Institute of Biodiversity (EIB). The institute with only 1067 tef accessions in Demissie (1991) has reached to 5169 accessions in 2011 (Tesema, 2013). This fourfold increase in the collection size in just two decades shows the presence of both a wide diversity of germplasm in the country and also the commitment of institutes and individuals to collect and preserve these germplasm for future use.

Characterization of the accessions according to their properties such as morphology is important in order to provide information to interested researchers or other sectors of society. The first and most comprehensive detailed morphological descriptions for 35 tef cultivars were given based on phenology, plant vigor, shoot and root related traits, panicle form, spikelet

size, growth habit, and lemma and caryopsis color (Ebba, 1975; **Table 1**).

Phenotypic Diversity in Tef

Tef is highly diverse and variable in terms of morphological and agronomic characters. The distribution of the crop in different agro-ecological zones coupled with the selection by farmers on the basis of their preferred traits has resulted in a number of varieties with unique characters. Genetic diversity analysis of tef accessions facilitates the development of improved varieties with high productivity and yield stability. In view of this fact, efforts have been made to assess

and quantify the extent and pattern of genetic diversity in the tef germplasm collections using different approaches (**Table 2**).

Diversity in Natural Populations

The first studies on phenotypic diversity in tef germplasm used 124 single panicles collected from the major tef producing areas in Ethiopia as a source of seed. The germplasm accessions showed significant variability for plant height, panicle length, maturity, seed color, seed yield, lodging, and panicle form (Mengesha et al., 1965). As shown in **Figure 1**, at least four distinct panicle forms are present in tef accessions, namely very-compact, semi-compact, fairly loose, and very-loose.

TABLE 1 | Selected properties of 35 tef ecotypes (cultivars) characterized (Ebba, 1975).

No	Ecotype (cultivar) name	Seed color	Plant height (cm)	Panicle		Days to		No. of spikelets per panicle	No. of florets per spikelet
				Length (cm)	Form	Heading	Maturity		
1	Ada	yWh	80	31	S-comp	45–50	95–115	320	8.3
2	Addisie	yWh	80	30	V-comp	45–50	95–110	310	6.0
3	Adoensis	mBr	70	30	V-loose	45–50	90–95	440	6.5
4	Alba	yWh	85	45	F-loose	45–50	95–120	525	10.0
5	Balami	yWh	88	36	V-loose	40–45	90–110	424	8.0
6	Beten	yWh	70	30	V-loose	40–45	85–95	220	7.0
7	Bunninye	mBr	34	16	V-loose	35–40	75–85	90	6.0
8	Burssa	yWh	58	20	S-comp	45–50	85–90	210	7.4
9	Curati	poW	88	40	S-comp	50–60	95–120	600	6.5
10	Dabbi	mBr	70	30	V-loose	40–45	80–95	295	6.5
11	Denekye	mBr	60	20	S-comp	45–50	90–115	165	9.6
12	Dschanger	mBr	75	30	F-loose	40–45	90–110	210	6.2
13	Enatitie	yWh	70	30	V-loose	40–45	90–100	270	6.2
14	Fesho	Br	50	20	V-loose	38–45	75–85	135	6.2
15	Gea-Lamie	Br	30	15	V-loose	25–30	60–70	60	6.8
16	Gofarie	yWh	78	28	S-comp	45–50	90–100	285	7.0
17	Gommadie	yWh	75	25	S-comp	45–50	90–100	290	8.9
18	Gorradie	yWh	90	40	V-comp	50–55	95–120	356	6.8
19	Hamrawe Murri	yWh	75	30	V-comp	50–55	90–100	310	6.9
20	Hatalla	yWh	90	38	V-loose	50–55	90–115	420	6.7
21	Janno	yWh	75	30	F-loose	45–50	85–105	335	8.5
22	Karadebi	Br	55	22	V-loose	40–45	85–90	160	6.7
23	Kaye Agachem	lBr	77	30	V-comp	45–50	90–110	300	6.8
24	Kaye Murri	yWh	80	30	V-comp	45–50	90–105	280	6.5
25	Manya	yWh	75	35	F-loose	40–45	90–110	350	9.0
26	Murri	yWh	95	38	V-comp	50–55	105–120	330	5.0
27	Purpurea	Br	85	38	F-loose	45–50	90–105	285	7.7
28	Rosea	yWh	75	30	F-loose	45–50	90–100	215	10.7
29	Rubicunda	yWh	85	35	F-loose	45–50	90–115	290	8.4
30	Shawa Gemerra	Br	35	16	F-loose	30–35	60–75	60	12
31	Trotteriana	Br	70	25	V-comp	50–55	90–95	210	8.0
32	Tullu Nasy	poW	42	17	V-loose	35–40	60–70	115	6.3
33	Variegata	lBr	70	32	F-loose	45–50	90–100	160	10.9
34	Viridis	poW	75	35	F-loose	45–50	85–95	275	6.7
35	Zuccagniana	Br	65	27	V-comp	45–50	90–100	200	6.4

Panicle form: V-loose, very loose; F-loose, fairly loose; S-comp, semi-compact; V-comp, very compact. Seed color: Br, brown; mBr, medium brown; lBr, light brown; poW, purple orange white; yWh, yellow white.

TABLE 2 | Studies made on phenotypical and morphological diversity in tef.

Germplasm or genotypes		Sites (No)	Reference
Type	Number		
Natural accession	124	1	Mengesha et al. (1965)
Hybrids	559	1	Berhe et al. (1989a,b,c)
Natural accession	21		Bekele (1996)
Natural accession	225		Ketema (1993)
RIL (key Murri \times Fesho)	165	3	Tefera et al. (2003b)
RIL (key Murri \times <i>E. pilosa</i>)	200	3	Tefera et al. (2003a)
<i>F</i> ₂ (12 crosses)	12	1	Tefera et al. (2008)
Natural accession	320	2	Assefa et al. (1999, 2000)
Natural accession	120	4	Assefa et al. (2001b)
Natural accession	1080	1	Assefa et al. (2001a)
Natural accession	3000	1	Assefa et al. (2002a,b, 2003b)
Natural accession	3600	1	Kefyalew et al. (2000)
Natural accession	144	2	Adnew et al. (2005)
RIL (196 \times 974)	196	2	Chanyalew et al. (2006)
RIL (196 \times 2356)	190	2	Chanyalew et al. (2009)
RIL (Kay Murri \times <i>E. pilosa</i>)	94	3	Yu et al. (2007)
Natural accession	37	1	Ayalew et al. (2011)
Natural accession	15	1	Plaza-Wüthrich et al. (2013)
Natural accession	81	1	Lule and Mengistu (2014)

RIL: recombinant inbred lines.

Later, studies involving 2255 tef lines collected from different parts of the country showed high variation for flag leaf area, single plant grain yield, and straw yield (Ketema, 1993). The analyses of 9885 accessions collected from 14 former provinces of Ethiopia showed simple coefficient of variation (SCV) estimates ranging from 32% for primary panicle branches to 217% for grain yield/plant (Bekele, 1996). While using SCV, the extent of variation among traits is not affected by the magnitudes of values and units of measurement. Since SCV does not efficiently

measure diversity among traits, phenotypic (PCV) and genotypic (GCV) coefficients of variation, which are based on partitioning of the total variance into components of genetic and non-genetic factors, are now more extensively used. Accordingly, various breeders have applied these two indices in evaluating the tef germplasm (Tefera et al., 1990; Hundera et al., 1999; Assefa et al., 2000; Chanyalew et al., 2009). Most of these studies revealed significant to highly significant differences among the genotypes for most of the traits examined, and this variability would serve as a basis for the improvement of the crop. Because the magnitude of genetic variation is better assessed from GCV, breeders usually focus on traits with high GCV estimates. High GCV values were reported for tiller number, panicle weight, grain yield per panicle, plant biomass, and grain yield (Assefa et al., 1999, 2001b; Balcha et al., 2003; Tefera et al., 2008; Chanyalew et al., 2009). This wide genetic variation indicates much potential for improving the crop through direct selection and/or hybridization.

Characters with huge variability include: days to panicle emergence (25–81), days to maturity (50–140), number of grains/plant (9,000–90,000), plant height (20–156 cm), number of tillers/plant (5–35), panicle type (from very loose to very compact), flag leaf area (2–26 cm²), culm diameter (1.2–5 mm; Ketema, 1993; Assefa et al., 2001a,b). Variability in tef germplasm for culm internode diameter is a key factor in the identification of tef lines with improved lodging resistance.

Soon after the discovery of breeding techniques for tef (Berhe, 1975), several studies were made to investigate the inheritance of key agronomic traits and their contributions to tef breeding. The initial studies dealt with investigations of the inheritance of lemma color, seed color, panicle form in *F*₂ and *F*₃ populations of crosses involving genotypes with contrasting phenotypes (Berhe et al., 1989a,b,c). Subsequent studies were made by several other researchers (Tefera et al., 2003a,b, 2008; Chanyalew et al., 2006; Yu et al., 2006; Zeid et al., 2011).

Diversity due to Agro-Ecology

Significant clinal diversity was reported in tef germplasm populations collected from different altitudinal zones for traits such as days to maturity, number of culm nodes, first and second basal culm internode diameter, and harvest index (Assefa et al., 2001b). Likewise, significant altitude-based diversity in tef germplasm



FIGURE 1 | Diversity in the form of tef panicles. (A) Very compact, (B) semi-compact, (C) fairly loose, (D) very loose.

populations was found for traits such as main shoot culm node number, days to maturity, diameters of the first and second lowest primary shoot culm internodes, and harvest index (Assefa et al., 2002a). However, no significant differences for qualitative traits (such as lemma, seed and anther colors and panicle form) were reported among the altitudinal zones (Kefyalew et al., 2000). On the other hand, for the trait days to maturity, 36 heterogeneous tef populations had lower diversity levels for accessions collected between altitudes of 1800 and 2400 m, while high diversity was noted for accessions obtained below 1800 m above sea level (Assefa et al., 2000).

Evaluations of 70 accessions of tef collected from different regions of Ethiopia showed significant variations within populations, among populations within regions, and among regions in most of the phenotypic traits (Tadesse, 1993). On the other hand, studies based on evaluations of 3600 tef lines representing 36 populations collected from the Central and Northern Regions of Ethiopia revealed significant regional diversity for seed color and days to maturity (Kefyalew et al., 2000). Furthermore, other studies showed significant regional diversity for lemma color, number of culm internodes, and counts of basal and middle spikelet florets in tef germplasm populations from different parts of the country (Assefa et al., 2002b).

An experiment at two locations using 144 accessions collected from different regions of Ethiopia showed that accessions from the same origin clustered into different classes and those from different origins also clustered into the same group (Adnew et al., 2005). Other studies further confirmed that the level of genetic diversity is higher in tef germplasm within a region than between regions, and as a result, accessions that had originated from the same region and altitude were grouped into distinct and distant clusters (Assefa et al., 2001b).

On the other hand, no significant differences were obtained among diverse altitude zones for parameters like days to panicle emergence, culm and panicle length, number of panicle branches, counts of fertile florets/spikelet, and shoot biomass (Assefa et al., 2001a,b). Moreover, diversity studies using 33 accessions collected from North-Western Ethiopia and four improved varieties (Ayalew et al., 2011) and selected tef genotypes (Plaza-Wüthrich et al., 2013) revealed considerable variations among the genotypes for many of the traits assessed.

However, this genetic variability is rapidly declining as farmers are quickly adopting improved cultivars and using them instead of landraces. In order to reduce the expected genetic erosion, the EIB has made rescue collections from different agro-ecological zones.

Molecular Diversity in Tef

In the past, efforts have been made to characterize and analyze the diversity levels in cultivars of tef and its relatives based on approaches other than morphological or phenotypic data (Table 3). Before high-throughput sequencing provided copious amounts of molecular data, chromatography, flow cytometry,

TABLE 3 | Studies made on molecular (genotypic) diversity in tef.

Germplasm or genotypes		Technique	Reference
Type	Number		
Natural accessions	11	biochemical	Bekele and Lester (1981)
Natural accessions	37	Protein markers	Bekele et al. (1995)
Natural accessions	47	AFLP	Bai et al. (1999b)
RIL (Kaye Murri x Fesho)	85	AFLP	Bai et al. (1999a)
Natural accessions	47	RAPD	Bai et al. (2000)
Natural accessions	14	AFLP	Ayale et al. (1999)
Natural accessions	6	Diverse*	Espelund et al. (2000)
RIL (Kaye Murri x <i>E. pilosa</i>)	116	RFLP	Zhang et al. (2001)
Natural accessions	92	ISSR	Assefa et al. (2003a)
RIL (tef x <i>E. pilosa</i>)	124	AFLP, EST, ISSR, SSR	Chanyalew et al. (2007)
RIL (Kaye Murri x <i>E. pilosa</i>)	94	Diverse markers**	Yu et al. (2006)
Natural + improved	326	SSR markers	Zeid et al. (2012)
Natural accessions	31	Haplotype analysis and LD	Smith et al. (2012)
Natural accessions	20	SSR	Cannarozzi et al. (2014)

*chloroplast DNA, 18S r, VP1 DNA.

**AFLP, EST-SSR, SNP/INDEL, IFLP, ISSR.

gel electrophoresis, and polymorphism assays were used for the molecular characterization of genetic diversity.

Proteins as a Marker

Early work using differences in protein content to classify and distinguish different accessions of tef employed the chromatography and electrophoresis of proteins involved in traits of interest such as seed storage proteins. Studies on the relatedness between *Eragrostis* species and tef accessions using chromatography of leaf phenolics and electrophoresis of seed proteins as biochemical markers showed complex patterns of variation amongst tef cultivars (Bekele and Lester, 1981). Similarly, polymorphisms among tef seed storage proteins (albumin, globulin, and prolammin) were found based on SDS-PAGE (Bekele et al., 1995). The study was able to classify 37 cultivars into seven groups, and suggested that the polymorphisms in albumins and globulins could be exploited to identify genotypes with desirable nutritional qualities.

Genomics

Finding and exploiting DNA sequence variation within a genome is of utmost importance for crop genetics and breeding (Varshney et al., 2009). Over the last three decades, different methods have been developed to detect and quantify the genetic diversity of tef. The first techniques employed were flow cytometry, sequencing of single genes or regions and genotyping using AFLP, RAPD, RFLP, inter-simple sequence repeat (ISSR), and simple sequence repeat (SSR) markers, and these have all shed light on the structure of allelic diversity within selected tef germplasm collections (Girma et al., 2014). As shown in

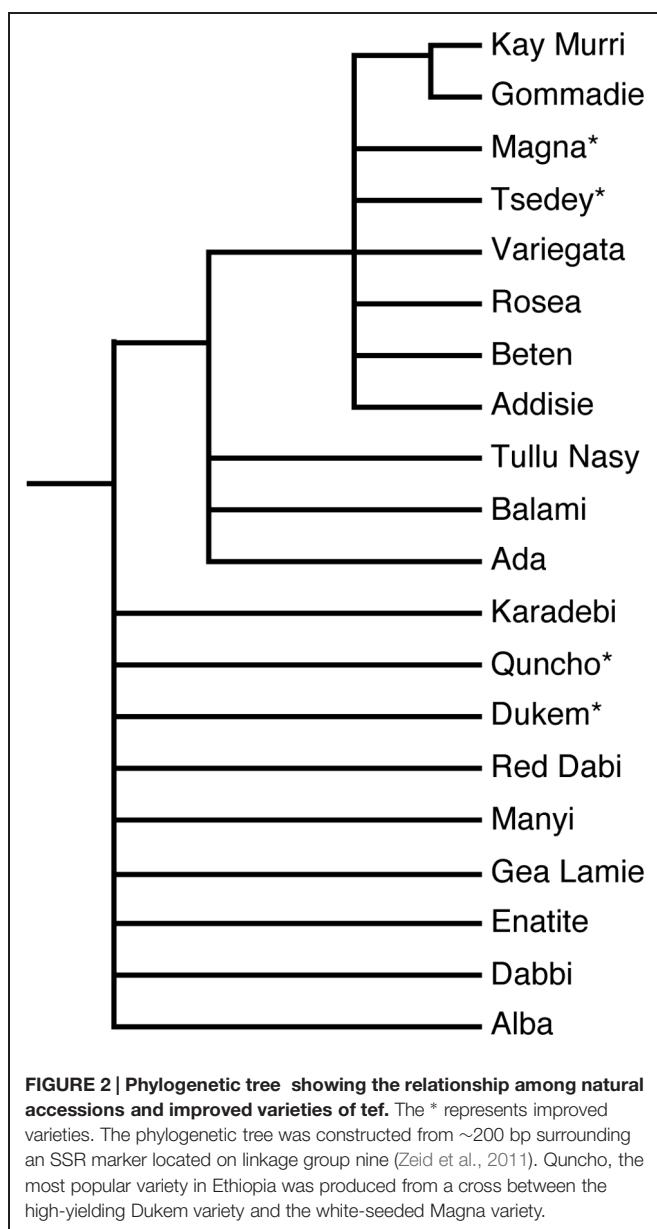


Figure 2, an SSR marker was used successfully to study the relationships among diverse tef genotypes, including natural accessions and improved varieties. However, only a small part of the diversity has been studied, and many of the essential questions still remain unanswered. Currently high-throughput single nucleotide polymorphism (SNP) genotyping is one of the methods that has been used to detect and exploit the genetic diversity of several crops. Genetic diversity analysis in some of the agriculturally important food crops such as sorghum (Nelson et al., 2011) and (Morris et al., 2013), barley (Close et al., 2009), rice (Thomson et al., 2012), bread wheat and emmer wheat (Akhunov et al., 2009), durum wheat (Trebbi et al., 2011), and maize (Yan et al., 2010) have been carried out with SNP genotyping methods employing next generation sequencing technologies.

Genome Size and Ploidy Determination

The genomic content of tef was first studied using flow cytometry, a popular method for ploidy screening and genome size estimation (Dolezel and Bartos, 2005). In the first measurement using four tef cultivars, the genome size was found to be between 714 and 733 Mbp (Ayele et al., 1996), relatively small for a grass (Table 4). The small genome size of tef made it a good candidate for genetic mapping and later genome sequencing. In addition, 32 of the first 35 tef ecotypes characterized (Ebba, 1975) as well as three commercial varieties were tested for ploidy level; all were tetraploid. In a another study with 10 released varieties of tef, following optimization of the flow cytometry conditions, the resulting genome size estimates were between 648 and 926 Mbp (Hundera et al., 2000; Table 4).

Sequence-Based Diversity

Around the same time, sequencing of single genes and small genomic regions was also employed to measure diversity and genetic relationships. Sequence analysis of non-coding regions of chloroplast DNA, 18S rDNA, and the transcription factor VP1 did not show significant intra-specific variation among six tef cultivars (Espelund et al., 2000). In addition, two rht1 (reduced height) gene homologs and three sd1 (semi-dwarf) genes were later sequenced for 31 accessions of tef (Smith et al., 2012). A low level of nucleotide diversity was observed and the genetic diversity could be organized into 2–4 haplotypes, a relatively small number.

Molecular Markers

Molecular markers are short sections of DNA that differ between varieties, and thus can be used for identification of a germplasm by a specific pattern of polymorphisms, to assess diversity and to determine relationships. Genetic relationships among accessions of *E. tef*, *E. pilosa*, and *E. curvula* which were collected from Ethiopia and USA were assessed based on AFLP (Bai et al., 1999b; Ayele and Nguyen, 2000) and RAPD markers (Bai et al., 2000). These analyses depicted relatively low levels (18%) of polymorphism within *E. tef*, and high similarity between *E. tef* and *E. pilosa*. The Jaccard similarity coefficient (size of the intersection of two sets divided by the size of the union) among two tef populations ranged from 84 to 96% for RAPD and from 73 to 99% for AFLP markers, indicating very close similarity among accessions. On the other hand, ISSRs analysis on 92 tef genotypes from seven regions plus improved varieties showed higher diversity among tef cultivars with Jaccard similarity coefficients ranging from 26 to 86% (Assefa et al., 2003a). A comparison of AFLP, EST-SSR, ISSR, and SSR markers for polymorphisms in tef recombinant inbred lines concluded that EST-SSR and ISSR makers had as much polymorphism as AFLP markers (Chanyalew et al., 2007).

Assessment of genetic diversity and relationships among 326 tef accessions, 13 wild relatives, and four commercial varieties from the United States based on 39 SSR markers, 26 of which were flanking QTL intervals for stem strength related traits, yield and lodging index showed genetic similarity (GS) estimates of between 0.20 and 0.99 among tef accessions (Zeid et al., 2012), and this contrasted with the narrow genetic background suggested in the earlier studies described above. A large base of

TABLE 4 | Variations in 2C DNA content and genome size among tef genotypes.

Genotype	Seed color	Panicle form	2C DNA content (pg)	1C genome size (Mbp)	Reference
Burssa	White	Semi compact	1.48 ± 0.02	714	Ayele et al. (1996)
Fesho	Brown	Very loose	1.51 ± 0.03	729	Ayele et al. (1996)
Key Murri	White	Very compact	1.50 ± 0.04	724	Ayele et al. (1996)
Trotteriana	Brown	Very compact	1.52 ± 0.03	733	Ayele et al. (1996)
DZ-01-354	Pale white	Very loose	1.34 ± 0.67	648	Hundera et al. (2000)
DZ-01-974	White	Very loose	1.49 ± 0.75	719	Hundera et al. (2000)
DZ-01-787	White	Very loose	1.49 ± 0.75	719	Hundera et al. (2000)
DZ-01-99	Brown	Very loose	1.65 ± 0.83	798	Hundera et al. (2000)
DZ-01-196	Very white	Fairly loose	1.68 ± 0.84	811	Hundera et al. (2000)
DZ-Cr-37	White	Very loose	1.61 ± 0.80	772	Hundera et al. (2000)
DZ-Cr-82	White	Very loose	1.92 ± 0.96	926	Hundera et al. (2000)
DZ-Cr-44	White	Very loose	1.78 ± 0.89	859	Hundera et al. (2000)
DZ-Cr-255	White	Very loose	1.52 ± 0.77	754	Hundera et al. (2000)
DZ-Cr-358	White	Very loose	1.38 ± 0.69	666	Hundera et al. (2000)

genetic diversity is indispensable for successful breeding programs. However, the diversity in tef has never been sufficient to produce the desired improvement in lodging resistance. Given the complexity of lodging and its component traits such as plant height, and culm internode length and diameter, alternative approaches including genetic transformation in line with marker-assisted selection should be considered for improving the malignant lodging syndrome in tef.

The afore-mentioned study of Zeid et al. (2012) also revealed 27 cases where accessions were identical to one or more of the other accessions. According to the authors, the high GS estimates from previous studies (Ayele et al., 1999; Bai et al., 1999a, 2000; Ayele and Nguyen, 2000) using the same plant material (lan-draces), was a marker dependent issue rather than due to low polymorphism in tef as previously suggested. An SSR marker used to construct a phylogenetic tree for 16 natural accessions and four improved varieties of tef showed the relationship among these genotypes (Cannarozzi et al., 2014). A multiple sequence alignment of approximately 200 base pairs was variable at 32 sites of which 25 were informative for determining evolutionary relationships.

Genetic Mapping

Genetic maps show the position of the molecular markers and QTLs relative to each other in terms of recombination frequency, and are used to find genes responsible for traits of interest. The first genetic map of tef was produced with an intra-specific cross between the 'Kaye Murri' and 'Fesho' cultivars and contained 211 AFLP markers in 25 linkage groups (Bai et al., 1999a). The low number of polymorphisms found between the two varieties of tef impeded its use in breeding. The same group later produced an RFLP linkage map using 116 RILs from the cross of 'Kaye Murri' with *E. pilosa* (Zhang et al., 2001). This inter-specific cross produced far more polymorphisms; however, the level of polymorphism was still smaller than that of other grasses.

The group of Sorrells has been instrumental in identifying QTLs associated with yield related traits and producing genetic

maps of tef using RILs from a cross between 'Kaye Murri' and *E. pilosa* with a variety of markers (Chanyalew et al., 2005; Yu et al., 2006; Zeid et al., 2011). Clusters of QTLs controlling yield and plant architecture were identified, thereby forming useful targets for applied breeding.

High-Throughput Genomics

During the last 5 years, tef genomics research has moved from analysis of a handful of genetic polymorphisms, toward whole genome sequencing and genome-wide polymorphism search. The genome and the transcriptome of the tef genotype Tsedey (DZ-Cr-37) were sequenced by the Tef Improvement Project at the Institute of Plant Sciences, University of Bern (Cannarozzi et al., 2014). Genome sequencing has many applications in tef improvement. First and foremost, primer sequences can be identified without resorting to other genomes or degenerate primers. This is especially important for the isolation of homeologous copies of each sub-genome for techniques such as Targeting Induced Local Lesions IN Genome (TILLING) which require genome specific primers. The genome has already been used to discover genetic markers such as SNPs and SSRs useful for marker-assisted breeding, for the construction of high density genetic maps and for linkage disequilibrium studies on diverse germplasm. Possession of the genomic sequences allows an understanding of the molecular basis of the mechanisms of tef's many desirable properties such as its tolerance to many abiotic and biotic stresses. The genes obtained from these analyses could be then transferred to other economically important crops.

Transcriptomics

To date, the transcriptome from only one tef improved variety has been sequenced (Cannarozzi et al., 2014), precluding comparison of transcriptomes between varieties or accessions. For the Tsedey improved variety (DZ-Cr-37), a normalized transcriptome library was prepared and sequenced from roots and shoots of tef seedlings resulting in a transcriptome with 27756 gene clusters and 38333 transcripts. In addition, a second non-normalized

library was obtained from various tef tissues subjected to drought and water-logging, resulting in a similar number of gene clusters.

An RNA-Seq study of two different varieties of quinoa (Raney et al., 2014), one representing valley ecotypes and another one representing high plains ecotypes, under different watering conditions was recently conducted. It was found that 27 putative gene products were differentially expressed based on variety \times treatment interaction. These included significant differences in root tissue in response to increasing water stress. A similar strategy could be employed for tef varieties to discover the QTLs responsible for specific accessions' traits.

Proteomics

Proteomics has emerged as an indispensable tool to analyze the whole or specific protein complement present in a particular tissue, organ, cell, or organelle (Agrawal et al., 2005; Benkeblia, 2011). In recent years, plant proteome analysis has evolved into high-throughput techniques resulting in the generation of high quality data with the continuous improvements made in sample preparation, protein separation, mass spectrometry, and protein search algorithms (Thelen, 2007; Benkeblia, 2011).

The application of proteomic studies has led to the discovery of a number of important proteins, and has facilitated attempts to explore their importance in improving plant yield and tolerance to environmental stresses (Salekdeh and Komatsu, 2007; Mochida and Shinozaki, 2010; Benkeblia, 2011). Similarly, to take advantage of the diversity among tef lines, proteomic approaches can be narrowed and refined to investigate which proteins are characteristic of specific lines or play important roles in a selected tef line. The corresponding genes of these proteins of interest can then be isolated and characterized from the tef genome provided it is comprehensively annotated. The particular phenotype conferred by the protein(s) of interest can then be introduced or enhanced in other tef lines using genetic and transgenic approaches to improve crop productivity. This functional genomics approach has been proposed as a standard 'omic' strategy for the improvement of many crop species (Agrawal and Rakwal, 2006).

To date, there has been no published proteomic study on tef with respect to protein profiling or comparative proteomics, while numerous such studies have been done on maize (Mohammed, 2005; Zhu et al., 2006; Prinsi et al., 2009), wheat (Jiang et al., 2012; Budak et al., 2013) and rice (Agrawal and Rakwal, 2006; Kim et al., 2014) using both gel-based and gel-free (mass spectrometry) techniques. Recently, proteomic profiling of the tef drought response has been undertaken, and should contribute valuable information on the key biological processes affected by water loss in tef (Kamies et al., 2014).

A key constraint affecting tef yield is salinity in the lowland and Rift Valley areas of Ethiopia, especially in the Awash valley and lower plains (Asfaw and Dano, 2011). The effects of increased salinity on tef yield and yield components were investigated by screening 15 lowland tef genotypes (10 accessions and 5 varieties) at different salinity levels. They found grain yield per main panicle to be the most affected by increased salinity, and although there were differences in genetic variation between tef

varieties and accessions, salt tolerance was observed in accession 237186 and variety DZ-Cr-37 (Tsedey) genotypes (Asfaw and Dano, 2011). This particular variety of tef, thus requires further proteomic and metabolomic investigation in order to elucidate the mechanisms of salt tolerance in tef and for identification of salt tolerant markers.

A comparative proteomics approach could be employed to investigate the cell wall proteome in both the tef stem and root tissues. A similar comparative proteomic study was done on maize primary and lateral roots whereby proteins involved in cell wall metabolism, cell elongation, lignin metabolism, defense, and citrate cycle were identified (Liu et al., 2006; Zhu et al., 2006). Such a study can be done on tef to identify and characterize stress-related cell wall proteins.

It is important to note that future tef improvements using the 'omics' tools should be conducted on one standardized consensus tef variety to allow for ease of comparison across functional genomic studies and to facilitate interpretation of data. Many studies have been conducted on the improved variety DZ-Cr-37 mostly because it is grown in areas which receive low rainfall (especially terminal drought-prone areas), and has been proposed to have a degree of drought tolerance in addition to being widely adaptable to differing climates (Ayele et al., 2001; Assefa et al., 2003a, 2011; Admas and Belay, 2011; Cannarozzi et al., 2014). Furthermore, since the genome and transcriptome information of this variety is available (Cannarozzi et al., 2014), it provides a platform for different proteomic strategies such as sub-cellular proteomics or phospho-proteomics to investigate stresses associated with tef. As stated earlier, proteomics is a functional tool that can provide insight to phenotypes of interest, and is largely dependent on the level of clarity and surety provided by the databases generated and the level of annotations made to the sequences. Since tef genome sequencing has been conducted and database annotation is in its infancy, proteo-bioinformatic approaches are somewhat limited, which in time will be remedied as more and more protein sequences are curated.

Tef Diversity in Key Traits

Grain Yield and Shoot Biomass

Development of varieties with high grain yield has been one of the top priorities of the National Tef Improvement Program in Ethiopia (Assefa et al., 2011). This varietal development process depends on the variability available within the gene pool. Over the past three decades, several studies (Assefa et al., 1999, 2000, 2001b, 2003b; Teklu and Tefera, 2005) were conducted to assess this variability, and tests both at research stations and on-farm yield trials were carried out at various locations. Over 30 improved varieties have been developed pushing the national average tef yield from 0.7 t/ha in 1994 to 1.5 t/ha in 2013 (CSA, 2014) hinting that the yield potential in tef can be further exploited. Variability in shoot biomass has also been studied in the majority of the above-mentioned studies, and a wide range (4–105 g/plant) was reported, suggesting the presence of high variability for this trait within the tef gene pool.

Seed Size and Seed Coat Characteristics

Despite the importance of seed size in terms of both agronomy and productivity, there exists only one study on the variability of seed size in tef. Using two improved tef genotypes, sieve-graded larger tef seeds had an increased seed yield, but it was concluded that this increase did not justify seed grading in tef (Belay et al., 2009). Seed coat characteristics in tef have received little research attention. The only study reported in literature showed slime cell differences in two tef genotypes and a wild *Eragrostis* species (Kreitschitz et al., 2009). The authors reported the presence of slime cells, a type of modified epidermal cells covering the fruit of the genotypes under investigation, and that such cells could play an adaptive role for tef plants growing in dry areas.

Physiology and Agronomy Related Traits

Due to a growing interest in utilizing tef as a gluten-free alternative to rice, there is corresponding interest in producing tef at a larger scale in some western countries. However, as a short day tropical cereal, growing tef in the temperate regions during the summer when the days get longer poses a big challenge. In order to investigate the ability of tef to flower in response to changes in the photoperiod, the effect of the relative lengths of day and night using four tef cultivars were studied. Two of the four cultivars had a stronger photoperiod response; panicle initiation as well as development and outgrowth of the panicle were influenced by photoperiod (van Delden et al., 2012).

Nitrogen-Use Efficiency

Nitrogen use efficiency (NUE), defined as the ratio of grain yield to supplied N, is a key parameter for evaluating a crop cultivar, and it is composed of N uptake efficiency and N physiological use efficiency (de Macale and Vlek, 2004). Breeding for NUE in tef could play a considerable role in reducing the amount of nitrogen fertilizer applied without affecting yield significantly. The NUE of tef is very low, ranging from 16 to 34% (Tulema et al., 2005). In the last decade, some authors looked at the genetic variation in NUE of tef (Tulema et al., 2005; Balcha et al., 2006; Habtegebrail et al., 2007). We suggest that further comparisons of nitrogen-use efficiency within the tef gene pool are important to evaluate their performance under limited nitrogen supply.

Osmotic Adjustment and Root Depth

Water deficit and salinity are among the abiotic production constraints limiting survival, growth, and productivity of tef. However, it is likely that there exists variability within the tef germplasm pool, and certain tef genotypes could adopt some strategies such as osmotic adjustment to resist these constraints. Systematic sampling of 54 tef genotypes from the entire gene pool showed a significant genotype effect on osmotic adjustment and root depth, irrespective of the area from where the genotypes were collected (Ayele et al., 2001).

Stress Related Traits

Drought Tolerance

The production areas of tef range from the cool highlands to the dry lowlands that are often associated with moisture deficit during critical stages of plant development. Studies investigating the

effect of moisture deficit on the performance of tef plants range from variability in key characters and response studies (Degu et al., 2008; Mengistu, 2009; Ginbot and Farrant, 2011; Shiferaw et al., 2012) to mapping QTLs related to economically important traits under water deficit conditions (Degu, 2010). In general, the majority of the studies have shown that there is genetic variability among the genotypes investigated suggesting that the tef gene pool harbors moisture stress tolerant genotypes that could be screened through efficient tools such as molecular markers.

Salinity and Acidity Tolerance

Due to the anticipated changes in the climate and expansion of farmlands in the rift valley areas, studying and documenting the effect of such growing conditions on tef production and productivity is worthwhile. Earlier, a few of such studies have been published including one which showed the presence of broad intra-specific variability among the ten tef accessions studied for salinity tolerance (Asfaw and Dano, 2011), and one which showed the presence of genetic variability for tolerance to soil acidity and aluminum toxicity in selected tef genotypes (Abate et al., 2013).

Nutrition, Health, and Consumers' Preference Related Traits

Seed Color Consumers' Preference

The Ethiopian Standards Agency recognizes four classes of tef grain mainly based on color of the seed (QSAE, 2001). These are very white, white, brown and mixed (commonly known as *Sergegna*). Oftentimes, farmers produce brown-seeded types for home consumption and white types for sale. Assessment of the diversity patterns of the seed color in tef with respect to growing regions and altitude zones revealed that the majority of tef collections from the north and northwestern part of Ethiopia were white-seeded as compared to those from the southern part of the country which were brown-seeded (Assefa et al., 2002b).

Nutritional Quality and Physico-Chemical Properties of Tef Seed

Knowledge of the physical properties of tef seed can be useful for agronomy, storage, marketing, and other socio-cultural purposes. A handful of studies have been carried out on the starch and protein contents of tef seed. Starch is the principal carbohydrate of all cereals, and represents, from 56% (oats) to 80% (maize) of the grain dry matter (Eliasson and Larsson, 1993). The starch characteristics of tef seed have been extensively studied (Bultosa et al., 2002, 2008; Bultosa and Taylor, 2003, 2004; Bultosa, 2007). The scientific study of tef grain protein and more specifically the amino acid composition extends back for over 50 years. Previous, reports indicated that tef seed contains a good balance of the essential amino acids, except lysine (Jansen et al., 1962). Three decades later, investigations of the polymorphism of seed albumin, globulin, and prolamin fractions showed the existence of considerable polymorphism in the studied protein fractions among the 37 tef cultivars investigated (Bekle et al., 1995). At the same time, Tatham et al. (1996) purified and characterized prolamins of tef. According to this study, the tef protein is made up of 9–14% prolamins and these are similar to prolamins of maize and sorghum.

This value is in a similar range to the previous results (3–15%; Bekele et al., 1995). However, according to a recent report, the prolamin content of three tef genotypes studied reached as much as 40% (Adebawale et al., 2011). In these studies, there is a discrepancy between the number of genotypes used and the methods employed. Clearly variability exists within the tef gene pool and a comprehensive study with more genotypes and modern tools to characterize and document the seed protein fractions is necessary. More recently, studies on tef seeds have changed course and three studies by Gebremariam et al. (2013a,b,c) investigated the malt quality attributes, while another by Boka et al. (2013) assessed the antioxidant properties of differentially processed tef grain.

As a potential alternative gluten-free food source for celiac patients, tef has been studied along with wheat, oat, rye, barley, rice, maize, and triticale (Spaenij-Dekking et al., 2005). This study showed that the tef cultivars evaluated contained no gluten or gluten homologs. This is the first scientific evidence for the absence of gluten in tef flour. Recently, this has been supported by results from the genome sequence initiative (Cannarozzi et al., 2014).

References

Abate, E., Hussein, S., Laing, M., and Mengistu, F. (2013). Quantitative responses of tef [*Eragrostis tef* (Zucc.) Trotter] and weeping love grass [*Eragrostis curvula* (Schrad.) Nees] varieties to acid soil. *Aust. J. Crop Sci.* 7, 1854–1860.

Adebawale, A. A., Emmambux, M. N., Beukes, M., and Taylor, J. R. N. (2011). Fractionation and characterization of teff proteins. *J. Cereal Sci.* 54, 380–386. doi: 10.1016/j.jcs.2011.08.002

Admas, S., and Belay, G. (2011). Drought-resistance traits variability in *Eragrostis tef* X *Eragrostis pilosa* recombinant inbred lines. *African J. Agric. Res.* 6, 3755–3761.

Adnew, T., Ketema, S., Tefera, H., and Sridhara, H. (2005). Genetic diversity in tef [*Eragrostis tef* (Zucc.) Trotter] germplasm. *Genet. Res. Crop Evol.* 52, 891–902. doi: 10.1007/s10722-003-6099-0

Agrawal, G. K., and Rakwal, R. (2006). Rice proteomics: a cornerstone for cereal food crop proteomes. *Mass Spectrom. Rev.* 25, 1–53. doi: 10.1002/mas.20056

Agrawal, G. K., Yonekura, M., Iwahashi, Y., Iwahashi, H., and Rakwal, R. (2005). System, trends and perspectives of proteomics in dicot plants Part I: technologies in proteome establishment. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 815, 109–123. doi: 10.1016/j.jchromb.2004.11.024

Akhunov, E., Nicolet, C., and Dvorak, J. (2009). Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theor. Appl. Genet.* 119, 507–517. doi: 10.1007/s00122-009-1059-5

Alaunyte, I., Stojceska, V., Plunkett, A., Ainsworth, P., and Derbyshire, E. (2012). Improving the quality of nutrient-rich Teff (*Eragrostis tef*) breads by combination of enzymes in straight dough and sourdough breadmaking. *J. Cereal Sci.* 55, 22–30. doi: 10.1016/j.jcs.2011.09.005

Asfaw, K. G., and Dano, F. I. (2011). Effects of salinity on yield and yield components of tef [*Eragrostis tef* (Zucc.) Trotter] accessions and varieties. *Curr. Res. J. Biol. Sci.* 3, 289–299.

Assefa, A. (2014). *The dire need to support ‘orphan crop’ research [Online]. Sci Dev Net*. Available at: <http://www.scidev.net/global/agriculture/opinion/the-dire-need-to-support-orphan-crop-research.html?hts0021=&stay=full> [accessed 16 February, 2015].

Assefa, K., Ketema, S., Tefera, H., Kefyalew, T., and Hundera, F. (2000). Trait diversity, heritability and genetic advance in selected germplasm Lines of tef [*Eragrostis tef* (Zucc.) Trotter]. *Hereditas* 133, 29–37. doi: 10.1111/j.1601-5223.2000.t01-1-00029.x

Assefa, K., Ketema, S., Tefera, H., Nguyen, H. T., Blum, A., Ayele, M., et al. (1999). Diversity among germplasm lines of the Ethiopian cereal tef [*Eragrostis tef* (Zucc.) Trotter]. *Euphytica* 106, 87–97. doi: 10.1023/A:1003582431039

Assefa, K., Merker, A., and Tefera, H. (2002a). Qualitative trait variation in Tef [*Eragrostis tef* (Zucc.) Trotter] germplasm from Western and Southern Ethiopia. *Euphytica* 127, 399–410. doi: 10.1023/A:1020318903145

Assefa, K., Tefera, H., and Merker, A. (2002b). Variation and inter-relationships of quantitative traits in tef (*Eragrostis tef* (Zucc.) Trotter) germplasm from western and southern Ethiopia. *Hereditas* 136, 116–125. doi: 10.1034/j.1601-5223.2002.1360205.x

Assefa, K., Merker, A., and Tefera, H. (2003a). Inter simple sequence repeat (ISSR) analysis of genetic diversity in tef [*Eragrostis tef* (Zucc.) Trotter]. *Hereditas* 139, 174–183. doi: 10.1111/j.1601-5223.2003.01800.x

Assefa, K., Merker, A., and Tefera, H. (2003b). Multivariate analysis of diversity of tef (*Eragrostis tef* (Zucc.) Trotter) germplasm from western and southern Ethiopia. *Hereditas* 138, 228–236. doi: 10.1034/j.1601-5223.2003.01638.x

Assefa, K., Tefera, H., Merker, A., Kefyalew, T., and Hundera, F. (2001a). Quantitative trait diversity in tef [*Eragrostis tef* (Zucc.) Trotter] germplasm from Central and Northern Ethiopia. *Genet. Res. Crop Evol.* 48, 53–61. doi: 10.1023/A:1011245918356

Assefa, K., Tefera, H., Merker, A., Kefyalew, T., and Hundera, F. (2001b). Variability, heritability and genetic advance in pheno-morphic and agronomic traits of tef [*Eragrostis tef* (Zucc.) Trotter] germplasm from eight regions of Ethiopia. *Hereditas* 134, 103–113. doi: 10.1111/j.1601-5223.2001.00103.x

Assefa, K., Yu, J. K., Zeid, M., Belay, G., Tefera, H., and Sorrells, M. E. (2011). Breeding tef [*Eragrostis tef* (Zucc.) trotter]: conventional and molecular approaches. *Plant Breed.* 130, 1–9. doi: 10.1111/j.1439-0523.2010.01782.x

Ayalew, H., Genet, T., Dessalegn, T., and Wondale, L. (2011). Multivariate diversity, heritability and genetic advance in tef landraces in Ethiopia. *African Crop Sci. J.* 19, 201–212.

Ayele, M., Blum, A., and Nguyen, H. T. (2001). Diversity for osmotic adjustment and root depth in TEF [*Eragrostis tef* (Zucc.) Trotter]. *Euphytica* 121, 237–249. doi: 10.1023/A:1012099914738

Ayele, M., Dolezel, J., Vanduren, M., Brunner, H., and Zapataarias, F. J. (1996). Flow cytometric analysis of nuclear genome of the Ethiopian cereal Tef [*Eragrostis tef* (Zucc.) Trotter]. *Genetica* 98, 211–215. doi: 10.1007/BF00121369

Conclusion

The broad spectrum of trait diversity in tef implies great opportunities for genetic improvement through either direct selection or intra-specific hybridization between parental lines with desirable traits. In addition, statistical tools such as correlation analysis can be used to aid selection of candidates in breeding programs. Additionally several mutagenized populations have been developed to supplement the natural diversity present in tef. As some studies reviewed here, used only few or selected tef genotypes, they may not be representative of the existing diversity in tef accessions. Future research is required to explore diversity in different traits of agronomic and nutritional importance. Concerted efforts of all stakeholders in research, development and funding are required to promote the research and development of vital crops such as tef in order to promote food and nutrition security.

Acknowledgments

Research and development work in Tadele’s Lab is supported by the Syngenta Foundation for Sustainable Agriculture, SystemsX and the University of Bern.

Ayele, M., and Nguyen, H. T. (2000). Evaluation of amplified fragment length polymorphism markers in tef, *Eragrostis tef* (Zucc.) Trotter, and related species. *Plant Breed.* 119, 403–409. doi: 10.1046/j.1439-0523.2000.00512.x

Ayele, M., Tefera, H., Assefa, K., and Nguyen, H. T. (1999). Genetic characterization of *Eragrostis* species using AFLP and morphological traits. *Hereditas* 130, 33–40. doi: 10.1111/j.1601-5223.1999.00033.x

Bai, G. H., Ayele, M., Tefera, H., and Nguyen, H. T. (2000). Genetic diversity in tef [*Eragrostis tef* (Zucc.) Trotter] and its relatives as revealed by Random Amplified Polymorphic DNAs. *Euphytica* 112, 15–22. doi: 10.1023/A:1003802207158

Bai, G., Tefera, H., Ayele, M., and Nguyen, H. T. (1999a). A genetic linkage map of tef [*Eragrostis tef* (Zucc.) Trotter] based on amplified fragment length polymorphism. *Theor. Appl. Genet.* 99, 599–604. doi: 10.1007/s001220051274

Bai, G. H., Ayele, M., Tefera, H., and Nguyen, H. T. (1999b). Amplified fragment length polymorphism analysis of tef [*Eragrostis tef* (Zucc.) trotter]. *Crop Sci.* 39, 819–824. doi: 10.2135/cropsci1999.0011183X003900030034x

Balcha, A., Gretzmacher, R., and Vollmann, J. (2003). Estimation of genetic parameters for grain yield and yield related traits in tef. *J. Genet. Breed.* 57, 251–257.

Balcha, A., Gretzmacher, R., and Vollmann, J. (2006). Genetic variation in nitrogen-use efficiency of tef. *J. Plant Nutr. Soil Sci.* 169, 704–710. doi: 10.1002/jpln.200625102

Bekele, E. (1996). Morphological analysis of *Eragrostis tef*: detection for regional patterns of variation. *SINET: Eth. J. Sci.* 19, 117–140.

Bekele, E., Fido, R. J., Tatham, A. S., and Shewry, P. R. (1995). Heterogeneity and polymorphism of seed proteins in tef (*Eragrostis tef*). *Hereditas* 122, 67–72. doi: 10.1111/j.1601-5223.1995.00067.x

Bekele, E., and Lester, R. N. (1981). Biochemical assessment of the relationships of *Eragrostis tef* (Zucc.) trotter with some wild *Eragrostis* species (Gramineae). *Ann. Bot.* 48, 717–725.

Belay, G., Zemedi, A., Assefa, K., Metaferia, G., and Tefera, H. (2009). Seed size effect on grain weight and agronomic performance of tef [*Eragrostis tef* (Zucc.) Trotter]. *African J. Agric. Res.* 4, 836–839.

Benkebia, N. (2011). *Sustainable Agriculture and New Biotechnologies*. Florida: CRC Press. doi: 10.1201/b10977

Berhe, T. (1975). *A Break-Through in Tef Breeding Techniques*. Rome: FAO.

Berhe, T., Nelson, L. A., Morris, M. R., and Schmidt, J. W. (1989a). Inheritance of phenotypic traits in tef.1. Lemma Color. *J. Heredity* 80, 62–65.

Berhe, T., Nelson, L. A., Morris, M. R., and Schmidt, J. W. (1989b). Inheritance of phenotypic traits in tef.2. seed color. *J. Heredity* 80, 65–67.

Berhe, T., Nelson, L. A., Morris, M. R., and Schmidt, J. W. (1989c). Inheritance of phenotypic traits in tef.3. panicle form. *J. Heredity* 80, 67–70.

Boka, B., Woldegiorgis, A. Z., and Haki, G. D. (2013). Antioxidant properties of Ethiopian traditional bread (injera) as affected by processing techniques and tef grain [*Eragrostis tef* (Zucc.) Trotter] varieties. *Can. Chem. Trans.* 1, 7–24. doi: 10.13179/chanchemtrans.2013.01.01.0012

Budak, H., Kantar, M., and Kurtoglu, K. Y. (2013). Drought Tolerance in Modern and Wild Wheat. *Sci. World J.* 2013, 16. doi: 10.1155/2013/548246

Bultosa, G. (2007). Physicochemical characteristics of grain and flour in 13 tef [*Eragrostis tef* (Zucc.) Trotter] grain varieties. *J. Appl. Sci. Res.* 3, 2042–2051.

Bultosa, G., Hall, A. N., and Taylor, J. R. N. (2002). Physico-chemical characterization of grain tef [*Eragrostis tef* (Zucc.) Trotter] starch. *Starch Starke* 54, 461–468. doi: 10.1002/1521-379X(200210)54:10<461::AID-STAR461>3.0.CO;2-U

Bultosa, G., Hamaker, B. R., and Bemiller, J. N. (2008). An SEC-MALLS study of molecular features of water-soluble amylopectin and amylose of Tef [*Eragrostis tef* (Zucc.) Trotter] starches. *Starch Starke* 60, 8–22. doi: 10.1002/star.200700642

Bultosa, G., and Taylor, J. R. N. (2003). Chemical and physical characterisation of grain Tef [*Eragrostis tef* (Zucc.) Trotter] starch granule composition. *Starch Starke* 55, 304–312. doi: 10.1002/star.200390065

Bultosa, G., and Taylor, J. R. N. (2004). Paste and gel properties and in vitro digestibility of Tef [*Eragrostis tef* (Zucc.) trotter] starch. *Starch Starke* 56, 20–28. doi: 10.1002/star.200200191

Cannarozzi, G., Plaza-Wuthrich, S., Esfeld, K., Larti, S., Wilson, Y. S., Girma, D., et al. (2014). Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (*Eragrostis tef*). *BMC Genomics* 15:581. doi: 10.1186/1471-2164-15-581

Chanyalew, C., Tefera, H., Harjit-Singh, and Sorrells, M. E. (2007). Comparison of AFLP, EST-SSR, ISSR and SSR markers for polymorphism among recombinant inbred lines of tef (*Eragrostis tef*). *J. Genet. Breed.* 61, 27–34.

Chanyalew, C., Tefera, H., Zelleke, H., and Singh, H. (2006). Correlation and path coefficient analysis of yield related traits in recombinant inbred lines of tef (*Eragrostis tef*). *J. Genet. Breed.* 60, 209–216.

Chanyalew, S., Singh, H., Tefera, H., and Sorrel, M. (2005). Molecular genetic map and QTL analysis of agronomic traits based on a *Eragrostis tef* x *Eragrostis pilosa* recombinant inbred population. *J. Genet. Breed.* 59, 53–66.

Chanyalew, S., Tefera, H., and Harjit Singh. (2009). Genetic variability, heritability and trait relationships in recombinant inbred lines of tef [*Eragrostis tef* (Zucc.) Trotter]. *J. Agric. Biol. Sci.* 5, 474–479.

Close, T. J., Bhat, P. R., Lonardi, S., Wu, Y. H., Rostoks, N., Ramsay, L., et al. (2009). Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics* 10:582. doi: 10.1186/1471-2164-10-582

CSA. (2014). *Agricultural Sample Survey for 2013/14*. Statistical Bulletin 532. Addis Ababa, Ethiopia: Central Statistical Agency.

D'Andrea, A. (2008). T'ef (*Eragrostis tef*) in ancient agricultural systems of Highland Ethiopia. *Econ. Bot.* 62, 547–566. doi: 10.1007/s12231-008-9053-4

de Macalel, M. A. R., and Vlek, P. L. G. (2004). The role of *Azolla* cover in improving the nitrogen use efficiency of lowland rice. *Plant Soil* 263, 311–321. doi: 10.1023/B:PLSO.00000447742.67467.50

Degu, H. D. (2010). Mapping QTLs related to plant height and root development of *Eragrostis tef* under drought. *J. Agric. Sci.* 2, 62–72.

Degu, H. D., Ohta, M., and Fujimura, T. (2008). Drought tolerance of *Eragrostis tef* and development of roots. *Int. J. Plant Sci.* 169, 768–775. doi: 10.1086/588064

Demissie, A. (1991). *A decade of germplasm exploration and collecting activities by the Plant Genetic Resources Centre/Ethiopia*. Cambridge: Cambridge University Press.

Dolezel, J., and Bartos, J. (2005). Plant DNA flow cytometry and estimation of nuclear genome size. *Ann. Bot.* 95, 99–110. doi: 10.1093/aob/mci005

Ebba, T. (1975). *Tef Cultivars: Morphology and Classification*. Dire Dawa: University, College of Agriculture.

Eliasson, A. C., and Larsson, K. (1993). *Cereals in Breadmaking*. New York: Marcel Dekker Inc.

Espelund, M., Bekele, E., Holst-Jensen, A., Jakobsen, K. S., and Nordal, I. (2000). A molecular genetic analysis of *Eragrostis tef* (Zucc.) Trotter: non-coding regions of chloroplast DNA, 18S rDNA and the transcription factor VP1. *Hereditas* 132, 193–202. doi: 10.1111/j.1601-5223.2000.00193.x

Gebremariam, M. M., Zarnkow, M., and Becker, T. (2013a). Effect of drying temperature and time on alpha-amylase, beta-amylase, limit dextrinase activities and dimethyl sulphide level of teff (*Eragrostis tef*) malt. *Food Bioprocess Technol.* 6, 3462–3472. doi: 10.1007/s11947-012-1025-0

Gebremariam, M. M., Zarnkow, M., and Becker, T. (2013b). Effect of teff (*Eragrostis tef*) variety and storage on malt quality attributes. *J. Inst. Brew.* 119, 64–70. doi: 10.1002/jib.65

Gebremariam, M. M., Zarnkow, M., and Becker, T. (2013c). Thermal stability of starch degrading enzymes of teff (*Eragrostis tef*) malt during isothermal mashing. *Process Biochem.* 48, 1928–1932. doi: 10.1016/j.procbio.2013.08.019

Ginbot, Z. G., and Farrant, J. M. (2011). Physiological response of selected *Eragrostis* species to water-deficit stress. *African J. Biotechnol.* 10, 10405–10417.

Girma, D., Assefa, K., Chanyalew, S., Cannarozzi, G., Kuhlemeier, C., and Tadele, Z. (2014). The origins and progress of genomics research on Tef (*Eragrostis tef*). *Plant Biotechnol. J.* 12, 534–540. doi: 10.1111/pbi.12199

Habtegebral, K., Sing, B. R., and Haile, M. (2007). Impact of tillage and nitrogen fertilization on yield, nitrogen use efficiency of tef (*Eragrostis tef* (Zucc.) Trotter) and soil properties. *Soil Tillage Res.* 94, 55–63. doi: 10.1016/j.still.2006.07.002

Hundera, F., Arumuganathan, K., and Baenziger, P. (2000). Determination of relative nuclear DNA content of tef [*Eragrostis tef* (Zucc.) Trotter] using flow cytometry. *J. Genet. Breed.* 54, 165–168.

Hundera, F., Tefera, H., Assefa, K., Tefera, T., and Kefyalew, T. (1999). Genetic variability and correlation of morpho-agronomic characters in tef landraces. *Trop. Sci.* 39, 140–146.

Ingram, A. L., and Doyle, J. J. (2003). The origin and evolution of *Eragrostis tef* (Poaceae) and related polyploids: evidence from nuclear waxy and plastid rps16. *Am. J. Bot.* 90, 116–122. doi: 10.3732/ajb.90.1.116

Jansen, G. R., Dimaio, L. R., and Hause, N. L. (1962). Amino acid composition and lysine supplementation of teff. *Agric. Food Chem.* 10, 62–64. doi: 10.1021/jf60119a021

Jiang, S. S., Liang, X. N., Li, X., Wang, S. L., Lv, D. W., Ma, C. Y., et al. (2012). Wheat drought-responsive grain proteome analysis by linear and nonlinear 2-DE and MALDI-TOF mass spectrometry. *Int. J. Mol. Sci.* 13, 16065–16083. doi: 10.3390/ijms131216065

Jones, B. M. G., Ponti, J., Tavassoli, A., and Dixon, P. A. (1978). Relationships of Ethiopian cereal tef (*Eragrostis tef* (Zucc.) Trotter) - evidence from morphology and chromosome-number. *Ann. Bot.* 42, 1369–1373.

Kamies, R., Tadele, Z., Cannarozzi, G., Farrant, J. M., and Rafudeen, M. S. (2014). A proteomic approach to investigate the response of *Eragrostis tef* to drought, in *Proceedings of the 1st INPPO World Congress on Plant Proteomics: Methodology to Biology*. Hamburg: Universität Hamburg.

Kebede, H., Johnson, R. C., and Ferris, D. M. (1989). Photosynthetic response of *Eragrostis tef* to temperature. *Physiol. Plant.* 77, 262–266. doi: 10.1111/j.1399-3054.1989.tb04979.x

Kefyalew, T., Tefera, H., Assefa, K., and Ayele, M. (2000). Phenotypic diversity for qualitative and phenologic characters in germplasm collections of tef (*Eragrostis tef*). *Genet. Resour. Crop Evol.* 47, 73–80. doi: 10.1023/A:1008781022038

Ketema, S. (1993). *Tef (*Eragrostis tef*): Breeding, Agronomy, Genetic Resources, Utilization, and Role in Ethiopian Agriculture*. Addis Ababa: Institute of Agricultural Research.

Ketema, S. (1997). *Tef, *Eragrostis tef* (Zucc.) Trotter*. Rome: Institute of Plant Genetics and Crop Plant Research, Gatersleben/International Plant Genetic Resources Institute.

Kim, S. T., Kim, S. G., Agrawal, G. K., Kikuchi, S., and Rakwal, R. (2014). Rice proteomics: a model system for crop improvement and food security. *Proteomics* 14, 593–610. doi: 10.1002/pmic.201300388

Kreitschitz, A., Tadele, Z., and Gola, E. M. (2009). Slime cells on the surface of *Eragrostis* seeds maintain a level of moisture around the grain to enhance germination. *Seed Sci. Res.* 19, 27–35. doi: 10.1017/S0960258508186287

Liu, Y., Lamkemeyer, T., Jakob, A., Mi, G. H., Zhang, F. S., Nordheim, A., et al. (2006). Comparative proteome analyses of maize (*Zea mays* L.) primary roots prior to lateral root initiation reveal differential protein expression in the lateral root initiation mutant rum1. *Proteomics* 6, 4300–4308. doi: 10.1002/pmic.200600145

Lule, D., and Mengistu, G. (2014). Correlation and path coefficient analysis of quantitative traits in tef [*Eragrostis tef* (Zucc.) Trotter] germplasm accessions from different regions of Ethiopia. *American J. Res. Commun.* 2, 194–204.

Mengesha, M. H., Picket, R. C., and Davis, R. L. (1965). Genetic variability and interrelationship of characters in tef, *Eragrostis tef* (Zucc.) Trotter. *Crop Sci.* 5, 155–157. doi: 10.2135/cropsci1965.0011183X000500020016x

Mengistu, D. K. (2009). The influence of soil water deficit imposed during various developmental phases on physiological processes of tef (*Eragrostis tef*). *Agric. Ecosyst. Environ.* 132, 283–289. doi: 10.1016/j.agee.2009.04.013

MoA. (2013). *Crop Variety Register Issue No. 15*. Addis Ababa: Ministry of Agriculture, Animal and Plant Health Regulatory Directorate.

Mochida, K., and Shinozaki, K. (2010). Genomics and bioinformatics resources for crop improvement. *Plant Cell Physiol.* 51, 497–523. doi: 10.1093/pcp/pcq027

Mohammed, A. (2005). Two-dimensional electrophoresis of soluble proteins and profile of some isozymes isolated from maize plant in response to NaCl. *Res. J. Agric. Biol. Sci.* 1, 38–44.

Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., et al. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci. U.S.A.* 110, 453–458. doi: 10.1073/pnas.1215985110

Naylor, R. L., Falcon, W. P., Goodman, R. M., Jahn, M. M., Sengooba, T., Tefera, H., et al. (2004). Biotechnology in the developing world: a case for increased investments in orphan crops. *Food Policy* 29, 15–44. doi: 10.1016/j.foodpol.2004.01.002

Nelson, J. C., Wang, S. C., Wu, Y. Y., Li, X. R., Antony, G., White, F. F., et al. (2011). Single-nucleotide polymorphism discovery by high-throughput sequencing in sorghum. *BMC Genomics* 12:352. doi: 10.1186/1471-2164-12-352

Plaza-Wüthrich, S., Cannarozzi, G., and Tadele, Z. (2013). Genetic and phenotypic diversity in selected genotypes of tef [*Eragrostis tef* (Zucc.)] Trotter. *African J. Agric. Res.* 8, 1041–1049. doi: 10.5897/AJAR2012.6682

Prinsi, B., Negri, A. S., Pesaresi, P., Cocucci, M., and Espen, L. (2009). Evaluation of protein pattern changes in roots and leaves of *Zea mays* plants in response to nitrate availability by two-dimensional gel electrophoresis analysis. *BMC Plant Biol.* 9:113. doi: 10.1186/1471-2229-9-113

QSAE. (2001). *Teff specification [Online]*. Available: <https://law.resource.org/pub/et/ibr/et.671.2001.pdf> [accessed December 23, 2014].

Raney, J. A., Reynolds, D. J., Elzinga, D. B., Page, J., Udall, J. A., Jellen, E. N., et al. (2014). Transcriptome analysis of drought induced stress in *Chenopodium quinoa*. *Am. J. Plant Sci.* 5, 338–357. doi: 10.4236/ajps.2014.53047

Salekdeh, G. H., and Komatsu, S. (2007). Crop proteomics: aim at sustainable agriculture of tomorrow. *Proteomics* 7, 2976–2996. doi: 10.1002/pmic.200700181

Shiferaw, W., Balcha, A., and Mohammed, H. (2012). Evaluation of drought tolerance indices in Tef [*Eragrostis tef* (Zucc.)Trotter]. *African J. Agric. Res.* 7, 3433–3438.

Smith, S. M., Yuan, Y. N., Doust, A. N., and Bennetzen, J. L. (2012). Haplotype analysis and linkage disequilibrium at five loci in *Eragrostis tef*. *G3-Genes Genomes Genet.* 2, 407–419.

Spaenij-Dekking, L., Kooy-Winkelhaar, Y., and Koning, F. (2005). The Ethiopian cereal tef in celiac disease. *N. Engl. J. Med.* 353, 1748–1749. doi: 10.1056/NEJM051492

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

Tadesse, D. (1993). Study on genetic variation of landraces of tef [*Eragrostis tef* (Zucc.) Trotter] in Ethiopia. *Genet. Resour. Crop Evol.* 40, 101–104. doi: 10.1007/BF00052640

Tatham, A. S., Fido, R. J., Moore, C. M., Kasarda, D. D., Kuzmicky, D. D., Keen, J. N., et al. (1996). Characterisation of the major prolamins of tef (*Eragrostis tef*) and finger millet (*Eleusine coracana*). *J. Cereal Sci.* 24, 65–71. doi: 10.1006/jcrs.1996.0038

Tefera, H., Assefa, K., and Belay, G. (2003a). Evaluation of interspecific recombinant inbred lines of *Eragrostis tef* x *E pilosa*. *J. Genet. Breed.* 57, 21–30.

Tefera, H., Assefa, K., Hundera, F., Kefyalew, T., and Tefera, T. (2003b). Heritability and genetic advance in recombinant inbred lines of tef (*Eragrostis tef*). *Euphytica* 131, 91–96. doi: 10.1023/A:1023009720870

Tefera, H., Belay, G., and Assefa, K. (2008). Genetic variation in F(2) populations and their potential in the improvement of grain yield in tef (*Eragrostis tef*). *Euphytica* 164, 105–111. doi: 10.1007/s10681-008-9655-z

Tefera, H., Ketema, S., and Tesemma, T. (1990). Variability, heritability and genetic advance in tef (*Eragrostis tef* (Zucc.) Trotter) Cultivars. *Tropical Agric.* 67, 317–320.

Teklu, Y., and Tefera, H. (2005). Genetic improvement in grain yield potential and associated agronomic traits of tef (*Eragrostis tef*). *Euphytica* 141, 247–254. doi: 10.1007/s10681-005-7094-7

Tesema, A. (2013). “Genetic diversity of tef in Ethiopia,” in *Achievements and Prospects of Tef Improvement*, eds A. Assefa, S. Chanyalew, and A. Tadele (Bern: EIAR-University of Bern), 15–20.

Thelen, J. J. (2007). *Introduction to Proteomics: A Brief Historical Perspective on Contemporary Approaches*. Berlin: Springer-Verlag.

Thomson, M. J., Zhao, K. Y., Wright, M., McNally, K. L., Rey, J., Tung, C. W., et al. (2012). High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. *Mol. Breed.* 29, 875–886. doi: 10.1007/s11032-011-9663-x

Trebbi, D., Maccaferri, M., De Heer, P., Sorensen, A., Giuliani, S., Salvi, S., et al. (2011). High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum* Desf.). *Theor. Appl. Genet.* 123, 555–569. doi: 10.1007/s00122-011-1607-7

Tulema, B., Zapata, F., Aune, J., and Sitaula, B. (2005). N fertilisation, soil type and cultivars effects on N use efficiency in tef [*Eragrostis tef* (Zucc.) Trotter]. *Nutrient Cycling Agroecosyst.* 71, 203–211. doi: 10.1007/s10705-004-5083-1

van Delden, S. H., Vos, J., Stomph, T. J., Brouwer, G., and Struik, P. C. (2012). Photoperiodism in *Eragrostis tef*: analysis of ontogeny and morphology in response to photoperiod. *Eur. J. Agron.* 37, 105–114. doi: 10.1016/j.eja.2011.11.001

Varshney, R. K., Nayak, S. N., May, G. D., and Jackson, S. A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* 27, 522–530. doi: 10.1016/j.tibtech.2009.05.006

Vavilov, I. (1951). *The Origin, Variation, Immunity and Breeding of Cultivated Plants. Translated From the Russian.* ed. K. S. Chester (New York: Ronald Press Co).

Yan, J. B., Yang, X. H., Shah, T., Sanchez-Villeda, H., Li, J. S., Warburton, M., et al. (2010). High-throughput SNP genotyping with the GoldenGate assay in maize. *Mol. Breed.* 25, 441–451. doi: 10.1007/s11032-009-9343-2

Yu, J. K., Graznak, E., Bresghello, F., Tefera, H., and Sorrells, M. E. (2007). QTL mapping of agronomic traits in tef [*Eragrostis tef* (Zucc) Trotter]. *BMC Plant Biol.* 7:30. doi: 10.1186/1471-2229-7-30

Yu, J. K., Kantety, R. V., Graznak, E., Benschoter, D., Tefera, H., and Sorrells, M. E. (2006). A genetic linkage map for tef [*Eragrostis tef* (Zucc.) Trotter]. *TAG. Theor. Appl. Genet.* 113, 1093–1102. doi: 10.1007/s00122-006-0369-0

Zeid, M., Assefa, K., Haddis, A., Chanyalew, S., and Sorrells, M. E. (2012). Genetic diversity in tef (*Eragrostis tef*) germplasm using SSR markers. *Field Crops Res.* 127, 64–70. doi: 10.1016/j.fcr.2011.10.013

Zeid, M., Belay, G., Mulkey, S., Poland, J., and Sorrells, M. E. (2011). QTL mapping for yield and lodging resistance in an enhanced SSR-based map for tef. *Theor. Appl. Genet.* 122, 77–93. doi: 10.1007/s00122-010-1424-4

Zhang, D., Ayele, M., Tefera, H., and Nguyen, H. T. (2001). RFLP linkage map of the Ethiopian cereal tef [*Eragrostis tef* (Zucc) Trotter]. *Theor. Appl. Genet.* 102, 957–964. doi: 10.1007/s001220000486

Zhu, J. M., Chen, S. X., Alvarez, S., Asirvatham, V. S., Schachtman, D. P., Wu, Y. J., et al. (2006). Cell wall proteome in the maize primary root elongation zone. I. Extraction and identification of water-soluble and lightly ionically bound proteins. *Plant Physiol.* 140, 311–325. doi: 10.1104/pp.105.070219

Conflict of Interest Statement: 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.

Copyright © 2015 Assefa, Cannarozzi, Girma, Kamies, Chanyalew, Plaza-Wüthrich, Blösch, Rindisbacher, Rafudeen and Tadele. 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) or licensor 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.

Genetic diversity and genomic resources available for the small millet crops to accelerate a New Green Revolution

Travis L. Goron and Manish N. Raizada *

Department of Plant Agriculture, University of Guelph, Guelph, ON, Canada

OPEN ACCESS

Edited by:

Joanna Marie-France Cross,
 Inonu University, Turkey

Reviewed by:

Dayong Li,
 Chinese Academy of Sciences, China
 Velu Govindan,
 CIMMYT, Mexico
 Anil Kumar,
 G B Pant University of Agriculture and
 Technology, India

*Correspondence:

Manish N. Raizada,
 Department of Plant Agriculture,
 University of Guelph, 50 Stone Road
 East, Guelph, ON N1G 2W1, Canada
 raizada@uoguelph.ca

Specialty section:

This article was submitted to Plant Genetics and Genomics, a section of the journal *Frontiers in Plant Science*

Received: 21 December 2014

Accepted: 27 February 2015

Published: 24 March 2015

Citation:

Goron TL and Raizada MN (2015) Genetic diversity and genomic resources available for the small millet crops to accelerate a New Green Revolution. *Front. Plant Sci.* 6:157. doi: 10.3389/fpls.2015.00157

Small millets are nutrient-rich food sources traditionally grown and consumed by subsistence farmers in Asia and Africa. They include finger millet (*Eleusine coracana*), foxtail millet (*Setaria italica*), kodo millet (*Paspalum scrobiculatum*), proso millet (*Panicum miliaceum*), barnyard millet (*Echinochloa* spp.), and little millet (*Panicum sumatrense*). Local farmers value the small millets for their nutritional and health benefits, tolerance to extreme stress including drought, and ability to grow under low nutrient input conditions, ideal in an era of climate change and steadily depleting natural resources. Little scientific attention has been paid to these crops, hence they have been termed “orphan cereals.” Despite this challenge, an advantageous quality of the small millets is that they continue to be grown in remote regions of the world which has preserved their biodiversity, providing breeders with unique alleles for crop improvement. The purpose of this review, first, is to highlight the diverse traits of each small millet species that are valued by farmers and consumers which hold potential for selection, improvement or mechanistic study. For each species, the germplasm, genetic and genomic resources available will then be described as potential tools to exploit this biodiversity. The review will conclude with noting current trends and gaps in the literature and make recommendations on how to better preserve and utilize diversity within these species to accelerate a New Green Revolution for subsistence farmers in Asia and Africa.

Keywords: finger millet, kodo millet, foxtail millet, barnyard millet, proso millet, little millet, New Green Revolution, biodiversity

Small Millets—Valuable Crops Neglected by the Green Revolution

The “Green Revolution” represents a period of massive agricultural advancement, and is often credited with saving over a billion people from starvation in the developing world (Borlaug, 2000; Evenson and Gollin, 2003). The initial focus of the Revolution was the promotion of semi-dwarf varieties of major cereal grain crops especially rice, wheat, and maize. Such modern varieties were also methodically bred to deal with environmental stresses, and in many cases produced yields several times higher than local cultivars. A highly cited example is the global success of “miracle rice” in the 1960s (De Datta et al., 1968). When faced with potential mass famine, the

Abbreviations: EST, expressed-sequence tag; RFLP, restriction fragment length polymorphism; AFLP, amplified fragment length polymorphism; SSR, simple sequence repeat; WUE, water use efficiency; NUE, nitrogen use efficiency.

Punjab region of India collaborated with international advisors to introduce IR8, a semi-dwarf rice modern variety. IR8 was found to produce up to 10 times the yield of traditionally grown varieties (De Datta et al., 1968) and helped to transform India's food production from deficit to surplus; national rice production tripled accompanied by a dramatic drop in price. IR8 and its progenitors as well as other modern varieties of cereals were further exported to other regions of the world with similar results especially in Latin America and Asia (Evenson and Gollin, 2003).

However, there are regions of the world that did not experience a Green Revolution. Sub-Saharan Africa experienced a lag in the benefits of modern varieties although efforts were made for their introduction and establishment (Ejeta, 2010). Reasons for the failure are complex. Many commentators point to institutional and political difficulties that may have hindered dissemination of new technology (Ejeta, 2010). However, it is also important to consider the agroeconomic complexities of the region, where a mixture of species less common elsewhere in the world are traditionally grown (Evenson and Gollin, 2003). A wide range of climatic zones and unique farming practices with a spectrum of soil types also created a challenge. In the early part of the Green Revolution, breeding generally consisted of modifying pre-existing genetic resources of wheat, maize, and rice in which research had already been conducted by developed nations. These varieties would be further bred to incorporate additional traits to increase yields. The strategy was not applicable to many African crops where essentially no formal work existed for researchers to build upon. In fact, it has been suggested that some African farmers faced increased hardship in response to the Green Revolution as a result of a global drop in food prices caused by its massive success elsewhere (Evenson and Gollin, 2003).

More optimistically, in the later years of the Green Revolution, research broadened to include less common food crops and began to close the gap in yield increases due to modern varieties. Locally administered organizations, such as the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), established research programs that included farmers in the dialog to strategically build a bank of genetic resources for traditionally grown species better suited to local climates and cropping systems. One group of such species is collectively known as the small millets and includes six cereal crops: finger millet (*Eleusine coracana*), foxtail millet (*Setaria italica*), kodo millet (*Paspalum scrobiculatum*), proso millet (*Panicum miliaceum*), barnyard millet (*Echinochloa* spp.), and little millet (*Panicum sumatrense*). Though all six cereals share a similar superficial classification (small grained cereals), they differ vastly in their phylogenies and continue to be grown in some of the most remote farms on Earth—thus isolation has maintained a wealth of agricultural and functional diversity. Their uses vary from animal fodder to human consumption, in which the small seeds can be ground into flour, cooked as porridge, or alternately fermented into enriched foods or alcoholic products. Where they are traditionally grown (Figure 1), small millets are highly valued for their diverse benefits and in many instances are considered nutritionally superior to other carbohydrate sources like rice and wheat (Hegde et al., 2005). Additionally, many of the small millets require very little fertilizer input as compared to more



FIGURE 1 | Depictions of small millet cultivation. **(A)** A typical subsistence small millet farm in India where the crops are grown under low input conditions and valued for their high stress tolerance. Source: M. Raizada. **(B)** Finger millet seed heads nearing maturity at the University of Guelph in Canada. The seed heads resemble the fingers of a human hand. Source: T. Goron. **(C)** Finger millet growing in a terraced field on a smallholder farm in Nepal. Source: M. Raizada. **(D)** Drudgery associated with transporting grain in the rural areas of Nepal. Source: M. Thilakarathna.

intensive grain cropping monocultures. Many reports also exist regarding their high degree of pest resistance and long-term storability, both traits which make the cultivation of small millets good insurance against famine and crop failure (Tsehay et al., 2006; Reddy et al., 2011).

Although previously neglected, the value of small millets in modern agricultural stability has begun to be identified. Much work has been accomplished toward the development of modern varieties with the goal of better directing existing diversity toward agricultural challenges of the new millennium. The purpose of this review is to highlight the diverse traits of each crop that are valued by farmers and consumers (e.g., nutritional quality) that have potential for selection, improvement or mechanistic study, along with other phenotypes of interest, then to describe the germplasm, genetic and genomic resources available as potential tools to exploit this biodiversity. The review will conclude with noting current trends and gaps in the literature and make recommendations on how to better preserve and utilize diversity within these species to accelerate a New Green Revolution.

Diversity of the Small Millets

Finger Millet (*Eleusine coracana*)

Finger millet was domesticated in western Uganda and the Ethiopian highlands (Figure 2) at least 5000 years ago before introduction to India approximately 3000 years ago (Dida et al., 2008). It is called finger millet, because the inflorescence resembles the fingers of a human hand (Figure 1). The morphology of the inflorescence can be used to differentiate between the two subspecies, *africana* and *coracana* (Dida and Devos, 2006). Each subspecies can be further divided into several races. Finger millet

is an allotetraploid. Genomic donors of the "A" genome are most likely *Eleusine indica* and *Eleusine trisachya* (Liu et al., 2014). The "B" genome has yet to be uncovered, and may have been contributed by an extinct ancestor (Liu et al., 2014). It is cultivated on 1.8 million ha in India, and also fills a substantial niche in eastern Africa (Table 1) (Dida and Devos, 2006). Kenyan farmers receive a high price for the grain, often twice that of maize and sorghum (Dida and Devos, 2006). The crop is highly valued in part due to its nutritional content, being especially calcium rich. Finger millet also contains methionine and tryptophan, amino acids which are often absent in starch-based diets of some subsistence farmers (Bhatt et al., 2011). Health benefits have been investigated, including anti-cancer and anti-diabetic activity, arising, respectively, from the grain's polyphenol content (anti-oxidant activity) and high fiber (which promotes slow digestion and hence stability

of blood sugar) (Chandrasekara and Shahidi, 2011a; Devi et al., 2014). The species will produce 5 tons/ha under optimum conditions (Dida and Devos, 2006) and requires very little nitrogen fertilization, with some reports indicating the most economic rate of application may be between 20 and 60 kg/ha (Hegde and Gowda, 1986; Pradhan et al., 2011). The plant is highly tolerant to drought and salt stress, though a wide diversity of stress resistance has been reported across genotypes (Uma et al., 1995; Bhatt et al., 2011). Unlike many crops consumed by subsistence farmers, finger millet has maintained high socio-economic importance in the Indian and African semi-arid tropics (Benin et al., 2004; Gull et al., 2014) and has received a level of investigation unattained by some of its cousins.

ICRISAT conserves 6804 finger millet germplasm accessions originating from 25 different countries. Other organizations

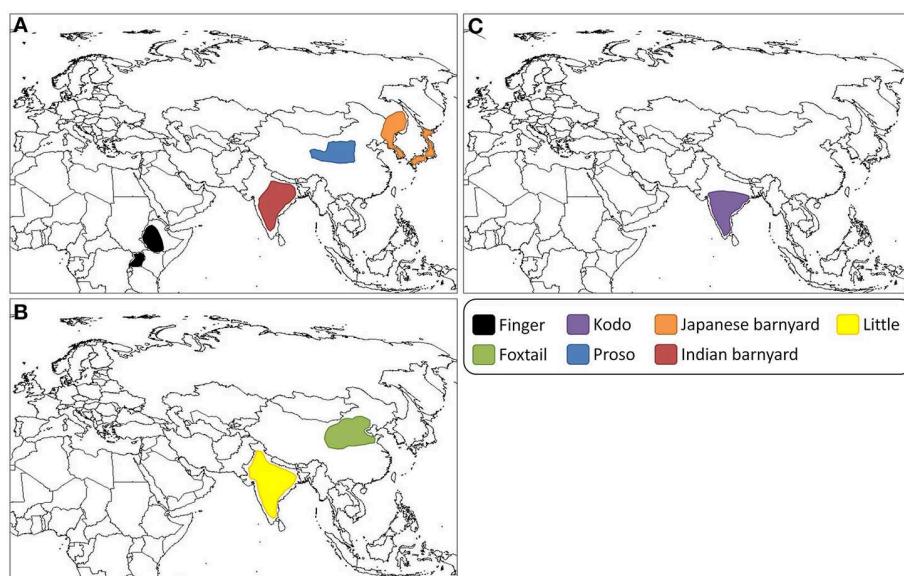


FIGURE 2 | Predicted geographic centers of domestication of the small millets. (A) Finger millet is predicted to have been domesticated in Uganda and the Ethiopian Highlands (Dida et al., 2008). Proso millet was likely domesticated on the Loess Plateau, China (M'Ribu and Hilu, 1994; Hu et al., 2008, 2009). Japanese barnyard millet was likely domesticated in Japan or Eastern Asia (Yabuno, 1962). It has been suggested that Indian

barnyard millet was domesticated at multiple sites across its current cultivation range in India (de Wet et al., 1983c). (B) Predicted sites of domestication of foxtail millet and little millet, respectively, on the North China Plain (Yang et al., 2012) and in India (de Wet et al., 1983a). (C) Kodo millet may have been domesticated at multiple sites across its current range of cultivation in India (de Wet et al., 1983b).

TABLE 1 | Areas where small millets are cultivated in significant quantities for human consumption.

Common name	Species name	Regions of cultivation	References
Finger millet	<i>Eleusine coracana</i>	India, Nepal, China, Myanmar, Sri Lanka, Kenya, Uganda, Eritrea, Sudan, Zimbabwe, Zambia, Malawi, Madagascar, Rwanda, Burundi	Dwivedi et al., 2012
Foxtail millet	<i>Setaria italica</i>	China (dry northern regions), India, Nepal, Korea, Japan	Dwivedi et al., 2012
Kodo millet	<i>Paspalum scrobiculatum</i>	India	Dwivedi et al., 2012
Proso millet	<i>Panicum miliaceum</i>	India, China, Nepal, western Myanmar, Sri Lanka, Pakistan, and South East Asian countries	Hu et al., 2008; Nirmalakumari et al., 2008
Japanese barnyard millet	<i>Echinochloa esculenta</i>	Japan, Korea, Northeastern China	Yabuno, 1987
Indian barnyard millet	<i>Echinochloa frumentacea</i>	Pakistan, India, Nepal, and central Africa	Yabuno, 1987
Little millet	<i>Panicum sumatrense</i>	India, Sri Lanka, Pakistan, Myanmar, and other South East Asian countries	Hiremath et al., 1990

manage germplasm banks of their own, the largest of which are summarized in **Table 2**. From these large collections, ICRISAT and other institutions group all genotypes according to region of origin or other parameters (Brown, 1989; Diwan et al., 1995; Hu et al., 2000; Wang et al., 2007). A subset of each group is selected that is representative of the genetic diversity of the crop: this group is termed the “core collection” and typically consists of ~10% of all available accessions. Core collections facilitate breeding by providing an efficient means to screen for desired traits from a large pool of genotypes. Mini-core collections, that represent ~1% of the total accessions, can be used by these institutions to further streamline the available genetic diversity.

The morphological diversity present within finger millet is immense. For example, a range of seed colors can be produced which are correlated with protein and calcium content (Vadi-voo et al., 1998). Landraces with different attributes (e.g., time to maturity, bird tolerance, drought tolerance, disease tolerance) are valued by farmers based on local agricultural complexities that reflect their productivity across multiple agroeconomic zones (Tsehay et al., 2006). For example, in the Ethiopian highlands, three high-yield landraces were identified and further developed into the commercial lines Tadesse, Padet, and Boneya (Aduguna, 2007). During a severe drought, Tadesse finger millet was the only cereal that remained productive. Farmers received double the price for the grain as compared to maize (Aduguna, 2007). This study illustrates what can be accomplished if germplasm banks are properly utilized for the selection of desirable traits.

The degree of morphological differences in finger millet requires that even core collections to be quite large; specialized tools will be needed to simplify characterization of functional diversity. Molecular markers represent one class of such tools, including restriction fragment length polymorphisms (RFLP), amplified fragment length polymorphisms (AFLP), expressed-sequenced tags (EST), and simple sequence repeats (SSR). Very few are reported for finger millet but more are beginning to appear in the literature. Molecular markers have been utilized in attempts to characterize calcium dynamics (Yadav et al., 2014b), disease resistance (Babu et al., 2014d), and in the association mapping of various agronomic traits as well as tryptophan accumulation (Babu et al., 2014a,b). Marker-assisted research has suggested that there was little sequence diversity in finger millet populations (Muza et al., 1995; Salimath et al., 1995; Yadav et al., 2014b), but this would be surprising given the geographic diversity in which finger millet is grown. Molecular markers have enabled linkage maps of the genome to be assembled (Dida et al., 2007). While progress has recently increased, the availability of a published genomic sequence would accelerate the development of markers to assist with genotype classification and breeding. In March 2014, the Bio-resources Innovations Network for Eastern Africa Development (Bio-Innovate) announced a finger millet sequencing project (**Table 3**); the initial genome assembly has been completed and the full sequence is expected by the end of 2014¹.

Research illuminating the finger millet transcriptome is beginning to appear. As the crop is valued for its high calcium content, studies have characterized calcium sensing and accumulation mechanisms across genotypes differing in their grain calcium content with the use of transcriptome high-throughput sequencing (Kumar et al., 2014b; Singh et al., 2014). A similar transcriptome analysis has been conducted on salinity responsiveness (Rahman et al., 2014). To investigate mechanisms behind the crop’s impressively high nitrogen utilization efficiency (NUE), the behavior of transcription factors Dof1 and Dof2 have been analyzed. It was found that in the roots of a high-protein variety, the *EcDof1/EcDof2* ratio was greater than that of a low protein variety, indicating a higher activation of N uptake and assimilation genes (Gupta et al., 2014a). The authors suggest that this ratio may in the future be utilized to screen other genotypes for high NUE.

Homologs of genes known to be agronomically important in major cereals, such as the transcripts described above, may assist with targeted breeding efforts in crops that are less characterized. Specifically, sequence variants of these genes may be used to develop orthologous molecular markers; those variants that correlate with desired traits may be used to screen accessions and subsequently assist in marker-assisted breeding efforts. This strategy may represent a way forward in the small millets. For example, finger millet researchers have isolated orthologs of genes known to be involved in grain amino acid composition (*Opaque 2*) and calcium content (calcium transporters, calmodulin) (Reddy et al., 2011; Nirgude et al., 2014). The researchers then associated SSR polymorphisms within these genes to characterize accessions that differed in their protein and calcium content, thus creating a targeted, cost-effective crop improvement strategy. A similar strategy to improve finger millet seed calcium content was also reported independently that focused on orthologs of calcium-binding proteins (CBPs) with extensive characterization of a seed dominant calmodulin (Kumar et al., 2014a,c). A parallel strategy has been suggested for disease resistance in finger millet based on the initial isolation of disease resistance receptors (Reddy et al., 2011; Babu et al., 2014c).

Progress has also occurred with respect to transgenic protocols for finger millet utilizing *Agrobacterium* and callus cell bombardment (Kothari et al., 2005; Ceasar and Ignacimuthu, 2009, 2011; Sharma et al., 2011; Jagga-Chugh et al., 2012; Plaza-Wüthrich and Tadele, 2012). Such techniques have allowed finger millet plants to be improved for drought and salinity tolerance (Ramegowda et al., 2012; Anjaneyulu et al., 2014; Hema et al., 2014), zinc accumulation (Cakmak, 2008; Ramegowda et al., 2013), and disease resistance (Latha et al., 2005).

Foxtail Millet (*Setaria italica*)

Named for the bushy, tail-like appearance of its immature panicles, foxtail millet has received a promising amount of research attention. Domesticated in China (**Figure 2**) approximately 8700 years ago, foxtail millet is considered one of the world’s oldest crops and ranks second in total world millet production, providing six million tons of grain for people throughout areas in southern Europe and Asia (Li and Wu, 1996; Yang et al., 2012). It is one of the main food crops in regions of the dry north of

¹<http://bioinnovate-africa.org/about-us/news/item/162-finger-millet-genomics-project-to-provide-researchers-with-better-tools-for-variety-production>

TABLE 2 | Significant germplasm collections of the small millets.

Common name	Institution	Headquarters	Number of accessions
Finger millet	<ul style="list-style-type: none"> National Bureau of Plant Genetic Resources (NBPGR) International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) All India Coordinated Minor Millet Project (AICMMP) Kenya Agricultural Research Institute (KARI) Institute of Biodiversity Conservation (IBC) USDA Agricultural Research Service (USDA-ARS) Serere Agricultural and Animal Production Research Institute (SAARI) SADC Plant Genetic Resource Centre Central Plant Breeding and Biotechnology Division, Nepal Agricultural Research Council (CPBBD) National Center for Genetic Resources Preservation National Institute of Agrobiological Sciences (NIAS) Mt. Makulu Central Research Station Institute of Crop Germplasm Resources, Chinese Academy of Agricultural Sciences (ICGR-CAAS) 	New Delhi, India Patancheru, India Bangalore, India Muguga, Kenya Addis Ababa, Ethiopia Griffin, USA Soroti, Uganda Lusaka, Zambia Kathmandu, Nepal Fort Collins, USA Kannondai, Japan Chilanga, Zambia Beijing, China	9522 (Dwivedi et al., 2012) 6804 ^a 6257 (Dwivedi et al., 2012) 2875 (Dwivedi et al., 2012) 2156 (Dwivedi et al., 2012) 1452 ^b 1231 (Dwivedi et al., 2012) 1037 (Dwivedi et al., 2012) 869 (Dwivedi et al., 2012) 702 (Dwivedi et al., 2012) 565 (Dwivedi et al., 2012) 390 (Dwivedi et al., 2012) 300 (Dwivedi et al., 2012)
Foxtail millet	<ul style="list-style-type: none"> Chinese National Genebank (CNGB) National Bureau of Plant Genetic Resources (NBPGR) ORSTOM-MONTP All India Coordinated Minor Millet Project (AICMMP) International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) National Institute of Agrobiological Sciences (NIAS) North Central Regional Plant Introduction Station, USDA-ARS Biologie Végétale Appliquée, Institut Louis Pasteur (IUT) Kenya Agricultural Research Institute (KARI) USDA Agricultural Research Service (USDA-ARS) Estación de Iguala, Instituto Nacional de Investigaciones Agrícolas (INIA) 	Shenzhen, China New Delhi, India Montpellier, France Bangalore, India Patancheru, India Kannondai, Japan Ames, USA I'Argonne-Strasbourg, France Muguga, Kenya Griffin, USA Iguala, Mexico	26,670 (Wang et al., 2012) 4330 (Dwivedi et al., 2012) 3500 (Dwivedi et al., 2012) 2512 (Dwivedi et al., 2012) 1535 ^c 1299 ^d 1000 (Dwivedi et al., 2012) 850 (Dwivedi et al., 2012) 772 (Dwivedi et al., 2012) 762 ^e 350 (Dwivedi et al., 2012)
Kodo millet	<ul style="list-style-type: none"> National Bureau of Plant Genetic Resources (NBPGR) All India Coordinated Minor Millet Project (AICMMP) International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) USDA Agricultural Research Service (USDA-ARS) 	New Delhi, India Bangalore, India Patancheru, India Griffin, USA	2170 (Dwivedi et al., 2012) 1111 (Dwivedi et al., 2012) 656 (Upadhyaya et al., 2014) 336 ^f
Proso millet	<ul style="list-style-type: none"> N.I. Vavilov All-Russian Scientific Research Institute of Plant Industry Institute of Crop Germplasm Resources, Chinese Academy of Agricultural Sciences (ICGR-CAAS) Ustymivka Experimental Station of Plant Production Yuryev Plant Production Institute UAAS International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) Botanical Garden of the Plant Breeding and Acclimatization Institute USDA Agricultural Research Service (USDA-ARS) North Central Reg. Plant Introd. Station, USDA-ARS Estación de Iguala, Instituto Nacional de Investigaciones Agrícolas (INIA) National Institute of Agrobiological Sciences (NIAS) 	St. Petersburg, Russian Federation Beijing, China S. Ustymivka, Ukraine Kharkiv, Ukraine Patancheru, India Bydgoszcz, Poland Griffin, USA Ames, USA Iguala, Mexico Kannondai, Japan	8778 (Dwivedi et al., 2012) 6517 (Dwivedi et al., 2012) 3976 (Dwivedi et al., 2012) 1046 (Dwivedi et al., 2012) 842 ^g 721 (Dwivedi et al., 2012) 719 ^h 713 (Dwivedi et al., 2012) 400 (Dwivedi et al., 2012) 302 ⁱ

(Continued)

TABLE 2 | Continued

Common name	Institution	Headquarters	Number of accessions
Barnyard millet (both species)	• International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)	Patancheru, India	743 ^j
Japanese barnyard millet	• National Institute of Agrobiological Sciences (NIAS)	Kannondai, Japan	159 ^k
Indian barnyard millet	• USDA Agricultural Research Service (USDA-ARS)	Griffin, USA	232 ^l
Little millet	• All India Coordinated Minor Millet Project (AICMMP) • International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) • USDA Agricultural Research Service (USDA-ARS)	Bangalore, India Patancheru, India Griffin, USA	544 (Dwivedi et al., 2012) 466 ^m 212 ⁿ

^a<http://www.icrisat.org/crop-fingermillet.htm>^b<http://www.ars-grin.gov/npgs/index.html>^c<http://www.icrisat.org/crop-foxtailmillet.htm>^dhttp://www.gene.affrc.go.jp/index_en.php^e<http://www.ars-grin.gov/npgs/index.html>^f<http://www.ars-grin.gov/npgs/index.html>^g<http://www.icrisat.org/crop-prosomillet.htm>^h<http://www.ars-grin.gov/npgs/index.html>ⁱhttp://www.gene.affrc.go.jp/index_en.php^j<http://www.icrisat.org/crop-barnyardmillet.htm>^khttp://www.gene.affrc.go.jp/index_en.php^l<http://www.ars-grin.gov/npgs/index.html>^m<http://www.icrisat.org/crop-littlemillet.htm>ⁿ<http://www.ars-grin.gov/npgs/index.html>**TABLE 3 | Small millet genomic resources and features.**

Common name	Ploidy	Chromosome number	Genome size estimate (pg, 2C)	ESTs available from NCBI ^a	Genome sequence availability
Finger millet	Tetraploid	$2n = 4x = 36$ (Bisht and Mukai, 2001)	3.34–3.87 (Mysore and Baird, 1997)	1982	Sequencing in progress ^b
Foxtail millet	Diploid	$2n = 2x = 18$ (Wanous, 1990)	1.02–1.04 (D’Ennequin et al., 1998)	66,051	Two reference genomes (Bennetzen et al., 2012; Zhang et al., 2012)
Kodo millet	Tetraploid	$2n = 4x = 40$ (Burton, 1940)	1.91–1.98 (Jarret et al., 1995)	29	N/A
Proso millet	Tetraploid	$2n = 4x = 36$ (Baltensperger, 1996)	2.08 (Kubešová et al., 2010)	211	N/A
Japanese barnyard millet	Hexaploid	$2n = 6x = 36$ (de Wet et al., 1983c)	N/A	0 (74 in closely-related <i>Echinochloa crus-galli</i>)	N/A
Indian barnyard millet	Hexaploid	$2n = 6x = 36$ (Wanous, 1990)	2.7 (Abrahamson et al., 1973)	0	N/A
Little millet	Tetraploid	$2n = 4x = 36$ (Wanous, 1990)	N/A	0	N/A

^a<http://www.ncbi.nlm.nih.gov/> ^b<http://bioinnovate-africa.org/about-us/news/item/162-finger-millet-genomics-project-to-provide-researchers-with-better-tools-for-variety-production>

China (Wang et al., 2012). Foxtail millet is cultivated to a limited extent in North America for silage, birdseed, and as a cover crop. It is quick to mature, able to produce seed in 75–90 days, and sometimes grown as a “catch-crop” in between the plantings of other species (Baltensperger, 2002). Herbicide-resistant lines of foxtail millet have been identified and studied in detail (Zhu et al., 2006). Additionally, the plant is quite drought resistant and tolerant to salt stress (Jayaraman et al., 2008). The cultivar “Prasad” has been identified as being particularly salt-tolerant, perhaps due to an effective antioxidant mechanism mediated by polyamine accumulation (Sudhakar et al., 2015).

As opposed to finger millet which was the result of a single domestication event (Dida et al., 2008), the history of foxtail millet is more complex. Sequence diversity of 250 Chinese genotypes was found to be quite high, averaging 20.9 alleles per locus when

examined with 77 SSRs (Wang et al., 2012). Alleles clustered into two main geographic diversity centers, indicating the possibility of two domestication events within China; more work is needed to confirm this hypothesis (Wang et al., 2012). Additionally, it has been suggested that foxtail millet was independently domesticated in Europe based on archeological evidence (Jusuf and Pernes, 1985; Hunt et al., 2008; Hirano et al., 2011).

Foxtail millet is closely related to the hardy weed *Setaria viridis*, which is assumed to be its progenitor. *S. viridis*, or green foxtail, often exists in close proximity to its cultivated cousin and is problematic throughout Eurasia and North America with many reports of herbicide resistance (Morrison et al., 1989; Marles et al., 1993; Heap, 1997). Some evidence suggests genetic clustering across foxtail species is dictated primarily by region and not taxonomy, implying that interspecific hybridization between

S. viridis and modern *S. italica* is common (Li et al., 1942; Jusuf and Pernes, 1985). Indeed, deliberate crosses between these species have resulted in resistance to a variety of herbicides (Darmency and Pernes, 1985, 1989; Wang et al., 1996; Wang and Darmency, 1997). However, agronomic traits in many of the crosses were closer to the weedy variety of *Setaria*; hybrids displayed seed shedding, spindly shoot tissue, and low yield as well as the fertility losses associated with hybridization. These reports highlight the possibility of using interspecific hybridization to study different agronomically valuable traits from wild millet relatives in a domesticated genetic background for future breeding applications.

After its domestication in China, foxtail millet spread throughout Asia, Europe, and eventually to North America (Jusuf and Pernes, 1985). Its large range has resulted in three different races, each with multiple subraces. *Moharia* is common in Europe, Russia, and the Middle East. *Maxima* can be found in Eastern China, Georgia, Japan, Korea, Nepal, northern India, and the USA where it was introduced for the purposes of animal feed. *Indica* predominates in southern India and Sri Lanka (Table 1) (Jusuf and Pernes, 1985).

An interesting feature of modern foxtail millet diversity is the global distribution of two phenotypically different varieties—the waxy and non-waxy grain type (Van et al., 2008). Waxiness in cereal grains is caused by lowered levels of amylose in the grain endosperm, which gives the grain a sticky texture when cooked (Van et al., 2008). Geographical occurrence of these two groups of foxtail millet varieties coincides with the ethnological preferences of local human populations. In East and South East Asia, some local communities are known to prefer sticky cereals (e.g., glutinous rice) driven by the use of chopsticks by these cultures—it is in these regions that the waxy millet phenotype can be found (Van et al., 2008). The non-waxy grain phenotype is more widespread, cultivated throughout Eurasia and parts of Africa (Kawase et al., 2005). Control of the phenotype is due to transposable-element (TE) insertion events interrupting amylase production, and foxtail millet has been suggested as a model for studying TE-mediated evolution (Kawase et al., 2005).

Like finger millet, there is an abundance of foxtail millet germplasm available to the scientific community (Table 2). Due to its importance in China, the Chinese National Genebank (CNGB) appears to maintain the largest collection by far, totalling 26,670 accessions as of 2012 (Wang et al., 2012). ICRISAT holds germplasm from 26 countries, and genebanks in Japan (National Institute of Agrobiological Sciences, NIAS) and the USA (USDA, Plant Genetic Resources Conservation Unit, PGRCU) ensure access to a wide range of foxtail millet diversity. Some core and mini-core collections have been assembled (Upadhyaya et al., 2008, 2011). However, considering the wide range of foxtail millet cultivation and the diversity of accessions, many more core collections should be generated, especially in China (Li et al., 1998) to facilitate breeding efforts. Diverse foxtail millet landraces may provide valuable alleles to assist in these breeding efforts. For example, landraces from the north of China are typically well-adapted to cold weather with short growing seasons, and are highly sensitive to light and temperature changes while those from southern regions grow better in high temperatures

and humidity (Wang et al., 2012), demonstrating the types of useful alleles that may exist for this crop.

Foxtail millet has enjoyed more genetic characterization than the other small millets. Recently there has been a push to utilize the species as a model system for biofuel grasses. It is closely related to the bioenergy crops switchgrass (*Panicum virgatum*), napier grass (*Pennisetum purpureum*), and pearl millet (*Pennisetum glaucum*) (Doust et al., 2009). Foxtail millet has several characteristics that are valued in a model system—a small genome (~490 Mbp), small plant size, and a quick generation time, unusual for C4 grasses. As a result, two full reference sequences have been compiled using genotypes Yugul and Zhang Gu (Bennetzen et al., 2012; Zhang et al., 2012). In these studies, the authors also created high-density linkage maps with another foxtail millet line and green foxtail, and examined the evolution and mechanisms of C4 photosynthesis in detail (Bennetzen et al., 2012; Zhang et al., 2012).

Instigated by the newly available sequence data, research in foxtail millet molecular genomics continues to rapidly progress. Many genetic markers have been reported and utilized in foxtail millet to generate maps, analyze DNA polymorphisms, evolutionary origin(s), and relatedness to other cereals for future crop improvement efforts (Wang et al., 1998; Schontz and Rether, 1999; Jia et al., 2009; Yadav et al., 2014a). A large library of markers consisting of intron-length polymorphisms (ILPs) has been generated, in part enabled by an abundance of EST data which can be used to generate flanking primers (Muthamilarasan et al., 2014). Initial work toward marker-based, high-throughput genotype identification has been accomplished (Gupta et al., 2012; Pandey et al., 2013). For example, an allele-specific single nucleotide polymorphism (SNP) coding for a dehydration responsive element binding (DREB) gene was shown to associate with stress tolerance (Lata et al., 2011). The SNP has potential in marker-assisted breeding selection, and was validated in a foxtail millet core collection in which the allele was found to account for 27% of total variation of stress-induced lipid peroxidation (Lata and Prasad, 2013). In an association mapping study, eight SSR markers were found to correlate with nine different agronomic traits (Gupta et al., 2014b). ESTs and peptides have been identified which are differentially expressed between salt tolerant and non-tolerant cultivars (Veeranagamallaiah et al., 2008; Puranik et al., 2011). A genome-wide transcriptome has been generated after exposure to drought stress, in which regulatory roles of small interfering RNAs and non-coding RNAs were described (Qi et al., 2013). From this study, 2824 annotated genes were identified with drought-responsive expression patterns. Such comprehensive studies should be extended to other stress pathways for better characterization of available foxtail millet germplasm. The data might also be used to design useful millet microarrays. Using the reference genomes described above, research groups have begun to re-sequence genotypes of foxtail millet and identify vast libraries of SNPs and other markers (Bai et al., 2013; Jia et al., 2013). This information has been used to classify landraces according to flowering time, yield attributes, waxy character, and other agronomically important traits (Jia et al., 2013; Bai et al., 2013). The re-sequencing of diverse foxtail millet germplasm should continue as a strategy to aid marker-assisted breeding

efforts. Much work has also been accomplished in the behavior of transcription factors in foxtail millet under a variety of stressful conditions, details of which have been conveniently compiled in the database “FmTFDb” (Bonthala et al., 2014). The availability of this data is expected to greatly accelerate functional genomics in all small millet species.

Lastly, transgenic protocols have been developed for foxtail millet, with both *Agrobacterium* (Wang et al., 2011) and callus bombardment methods reported (Kothari et al., 2005; Ceasar and Ignacimuthu, 2009; Plaza-Wüthrich and Tadele, 2012), enabling some potentially useful molecular analyses. In one study, a pollen-specific gene has been altered to impair anther function by a co-suppression mechanism (Qin et al., 2008) which might be adapted for the development of male-sterile plants, valuable in breeding foxtail millet hybrid varieties.

Kodo Millet (*Paspalum scrobiculatum*)

Kodo millet was domesticated roughly 3000 years ago in India (Figure 2), the only country today where it is harvested as a grain in significant quantities, mainly on the Deccan plateau (Table 1) (de Wet et al., 1983b). The grain contains a diverse range of high-quality protein (Geervani and Eggum, 1989; Kulkarni and Naik, 2000), and has high anti-oxidant activity (anti-cancer) even when compared to other millets (Hegde and Chandra, 2005; Hegde et al., 2005; Chandrasekara and Shahidi, 2011b). Like finger millet, kodo is rich in fiber and hence may be useful for diabetics (Geervani and Eggum, 1989). It is drought tolerant and can be grown in a variety of poor soil types from gravelly to clay (de Wet et al., 1983b; M'Ribu and Hilu, 1996). Most genotypes take 4 months to mature (de Wet et al., 1983b). Like foxtail millet, a weedy counterpart of kodo exists and is problematic throughout old-world farming systems especially in damp areas (de Wet et al., 1983b; Becker and Johnson, 2001). It is believed that kodo was probably first harvested as a weed alongside other cereals like rice, perhaps leading to multiple domestication events of the millet across its current range (de Wet et al., 1983b). This practice continues in parts of Africa where the weed is also sometimes harvested during famine (de Wet et al., 1983b; Neumann et al., 1996; Ogie-Odia et al., 2010). In Africa, kodo is referred to as black rice or bird's grass (M'Ribu and Hilu, 1996). Limited molecular marker analysis has shown that kodo millet genotypes cluster by African vs. Indian origin (M'Ribu and Hilu, 1996).

Kodo millet is divided into three races (*regularis*, *irregularis*, and *variabilis*) based on panicle morphology (de Wet et al., 1983b). In southern India, there are small (*karu varagu*) and large seeded (*peru varagu*) varieties recognized, often grown together in the same field (de Wet et al., 1983b). General morphological variability is high, with large variance reported in many phenotypic parameters such as time before flowering, tiller number, and yield (Subramanian et al., 2010; Upadhyaya et al., 2014).

Kodo millet is a crop that might be described as incompletely domesticated, with some authors calling the cereal “pseudo-cultivated” (de Wet, 1992; Blench, 1997). As such, systematic breeding of kodo millet remains neglected but limited efforts have shown promise. Various metrics of plant productivity including dry fodder yield, plant height, and grain yield have revealed good heritability; improvement of these traits has been observed

through breeding, with four highly productive genotypes thus far identified (Upadhyaya et al., 2014). Pathogen resistance has been noted as a good breeding target, in particular resistance to smut (*Sorosporium paspali* and *Ustilago* spp.) and rust (*Puccinia substriata* Ellis and Barth), which are both major hindrances of kodo yield (Upadhyaya et al., 2014). Another potential target for breeding may be resistance to the fungi *Aspergillus flavus* and *Aspergillus tamari* which produce cyclopiazonic acid that can cause sleepiness, tremors, and giddiness in those that consume infected grain, known as “kodua poisoning” (Rao and Husain, 1985). Grain lodging can occur before harvest, therefore an earlier maturity time might also be targeted (de Wet et al., 1983b). It is also interesting that some cultivated landraces have maintained the perennial nature of their wild ancestor and continue to initiate culms following the maturity of older shoots (de Wet et al., 1983b). If this regeneration trait can be encouraged through breeding and hybridization, it may reduce fertilization inputs and labor.

Unfortunately, no genetic or molecular maps of the kodo millet genome appear to be available (Dwivedi et al., 2012), likely because of the problem of persistent cross-hybridization with its wild relatives. Molecular markers for kodo millet are few, but have been utilized in characterizing diversity and phylogeny (M'Ribu and Hilu, 1996; Kushwaha et al., 2015). There has been some preliminary work in miRNA target site prediction using ESTs from kodo (Babu et al., 2013). In this study, target genes were found to be involved in carbohydrate metabolism, cellular transport, and as structural proteins, but a severe lack of kodo DNA information limited this study; the closely-related rice genomic sequence was used for binding-site prediction. With respect to transgene methodology for kodo, the media conditions for callus regeneration protocols have been investigated; regenerated plantlets were successfully grown to maturity in soil (Ceasar and Ignacimuthu, 2010).

ICRISAT conserves 656 accessions of kodo millet, and a core collection has been established that reflects the phenotypic diversity of the entire collection (Upadhyaya et al., 2014). Some universities also maintain large kodo millet seed banks, a good example being the University of Agricultural Sciences in Bangalore (Ceasar and Ignacimuthu, 2010). As the crop is not significant outside of India, there are few reports of other banks with substantial numbers of accessions (Table 2). However, some organizations do keep collections for the purposes of studying the species as a weed as noted above; the US Department of Agriculture has 336 accessions in their National Plant Germplasm System (GRIN)². While seed of African origin does exist in some of these sources, it is rare. Better coverage and ecological exploration of the African continent would help to reveal and preserve diversity of valuable traits which might otherwise be missed by international scientists.

Proso Millet (*Panicum miliaceum*)

Proso millet, also called broomcorn and common millet, was domesticated in Neolithic China as early as 10,000 years ago (Figure 2) (Lu et al., 2009). The sequence diversity within proso

²<http://www.ars-grin.gov/>

provides evidence for a single site of domestication in the Chinese Loess Plateau (M'Ribu and Hilu, 1994; Hu et al., 2008, 2009). Proso millet expanded across Eurasia and was introduced to North America in the 1700s where it is now primarily used for animal fodder and birdseed (Bagdi et al., 2011). Proso is the true millet referenced in classical European and Middle Eastern sources, referred to by ancient Romans as "*milium*" (Smith, 1977). Archeological evidence of proso in Eastern Europe dating to 8000 years ago raises the possibility of a secondary independent domestication event, but additional study is needed to confirm this observation (Hunt et al., 2008, 2011). Proso millet was important in the diets of humans across Eurasia prior to the introduction of wheat, barley and potatoes (Kalinova and Moudry, 2006). Today it is only consumed in significant quantities in India (where it is known as *pani varagu* in Tamil), Nepal, western Myanmar, Sri Lanka, Pakistan, and South East Asian countries (Nirmalakumari et al., 2008). A weedy variety is widespread, which is likely the result of field escape and not due to the spread of the wild ancestor (McCanny and Cavers, 1988). Recent molecular analysis using chromosomal *in situ* hybridization has implicated *Panicum capillare* or a close relative as one of the genetic ancestors of proso (Hunt et al., 2014).

The benefits of consuming proso include its high protein content which ranges from 11.3 to 17% of grain dry matter (Kalinova and Moudry, 2006). Genotypic diversity in protein content and amino acid profile has been observed (Kalinova and Moudry, 2006). Like other small millets, the applicability of the grain in preventing cancer, heart disease, and managing liver disease and diabetes has been investigated with promising results (Nishizawa and Fudamoto, 1995; Nishizawa et al., 2002; Park et al., 2014; Zhang et al., 2014). There may be additional untapped phytochemical value as indicated by a wide range of genotype-specific grain colors (Zhang et al., 2014).

Proso millet is well-adapted to dry sandy soils, and might be the earliest dryland-farming crop in East Asia (Baltensperger, 2002; Lu et al., 2009). It may have the lowest water requirement of any cereal, able to produce harvestable grain with only 330–350 mm of annual rainfall (Baltensperger, 2002; Seghatoleslami et al., 2008; Hunt et al., 2011). Proso millet matures quickly within 60–90 days, a feature that contributes to its drought resistance and also makes it a good catch-crop (Baltensperger, 2002; Hunt et al., 2014). Genotype has been shown to affect drought tolerance by influencing harvest-index, yield, and water use efficiency (WUE) (Seghatoleslami et al., 2008). In the latter study, a hybrid genotype outperformed local varieties, validating the potential in breeding highly WUE proso millet. Preliminary work in characterizing proso miRNAs has been accomplished with the goal of understanding mechanisms responsible for the cereal's impressive drought resistance (Wu et al., 2012). Despite its drought tolerance, proso is best adapted to temperate latitudes unlike other small millets. It grows further north than any other millet up to a latitude of 54°N, and at elevations as high as 3500 m (Baltensperger, 2002). Substantial salinity tolerance has been reported in proso but with significant varietal diversity, with some especially tolerant varieties reported (Sabir et al., 2011; Liu et al., 2015). A higher sodium concentration in roots compared to

shoots has been suggested as a biomarker for future breeding efforts (Sabir et al., 2011; Liu et al., 2015).

Cultivated proso millet is divided into five races (Reddy et al., 2007). Race *miliaceum* resembles wild proso with large, open inflorescences and sub-erect branches with few subdivisions. *Patentissimum* is very similar to *miliaceum* with narrow, diffuse panicle branches. These two races are found across the entire Eurasian range of proso, and are considered primitive. *Contractum*, *compactum*, and *ovatum* have more compact inflorescences which are drooped, cylindrical, and curved, respectively (Reddy et al., 2007). ICRISAT holds 842 accessions from all five races (Table 2) (Reddy et al., 2007). The diversity of this collection has been characterized in terms of flowering time, plant height, panicle exertion, and inflorescence length (Reddy et al., 2007). Other significant collections of proso are summarized in Table 2. Perhaps the largest collection of proso is held by the N.I. Vavilov All-Russian Scientific Research Institute of Plant Industry in St. Petersburg, with roughly 8778 accessions as of 2012 (Dwivedi et al., 2012). Aside from ICRISAT (Upadhyaya et al., 2014), few proso millet core collections appear to exist for breeding purposes. Preliminary diversity clustering based on agronomic traits was performed on the Chinese collection for the purpose of SSR-based characterization (Hu et al., 2009). Perhaps the Chinese subset of 118 landraces could be repurposed and slightly modified to become a true core collection. Explant regeneration techniques have been published for proso, allowing transgenic work to be explored in the future (Plaza-Wüthrich and Tadele, 2012).

The genetic sequence diversity of proso has been examined to a limited degree. The sequence diversity is moderate to high (Karam et al., 2006; Cho et al., 2010; Hunt et al., 2011), perhaps due to continuing hybridization with wild varieties (Colosi and Schaal, 1997). Molecular markers in proso have often been derived from the available sequence data of related species including switchgrass, rice, wheat, barley and oat (Hu et al., 2009; Rajput et al., 2014). AFLP markers have shown promise in grouping proso based on biotype, but were insufficient in differentiating between wild and cultivated varieties (Karam et al., 2004). To the best of our knowledge, no genetic or molecular maps of the proso millet genome are available (Dwivedi et al., 2012).

Like kodo millet, waxy varieties of proso grain exist and are preferred in some areas of Asia because of their glutinous nature—again to facilitate consumption with chopsticks (Graybosch and Baltensperger, 2009). Clustering by geographical sequence diversity corresponds with this regional preference (Hu et al., 2008). Like other glutinous cereals, waxy types of proso have no detectable amylose in the seed endosperm, due to a mutation in the *Waxy* gene (Hunt et al., 2010). Molecular markers have been developed to identify these waxy genotypes and breed glutinous varieties that are highly valued by consumers (Araki et al., 2012). Proso has been compared to maize in its ethanol production ability, and fermentation efficiency was found to be the highest in waxy varieties (Rose and Santra, 2013). The authors suggest that encouraging the fermentation of proso millet could help stabilize its price in the USA where it is already grown for birdseed and fodder. Finally, proso millet has been utilized as a model organism for C4 carbon metabolism, specifically in the study of aspartate aminotransferase and malate translocation

which both contribute to the higher efficiency of C4 photosynthesis (Taniguchi et al., 1995; Taniguchi and Sugiyama, 1996, 1997; Sentoku et al., 2000).

Barnyard Millet (*Echinochloa* spp.)

Although sometimes referred to as a single taxonomic group, barnyard millet is composed of two separate species belonging to the genus *Echinochloa*. *Echinochloa esculenta* (syn. *Echinochloa utilis*, *Echinochloa crusgalli*) is cultivated in Japan, Korea, and the northeastern part of China while *Echinochloa frumentacea* (syn. *Echinochloa colona*) is found in Pakistan, India, Nepal, and central Africa (Table 1) (Yabuno, 1987; Wanous, 1990). Both species have overlapping morphological traits that make differentiation problematic. Visual identification is only possible based on the presence or absence of an awn and subtle differences in spikelet and glume morphology (de Wet et al., 1983c). Consequently, the common names Japanese and Indian barnyard millet have been suggested to simplify research and investigation of their phylogeny (Yabuno, 1987). Despite having such strong phenotypic similarities, cytology and marker work have shown the two millets to be genetically distinct; F₁ hybrids of the two species are sterile (Yabuno, 1962; Hilu, 1994). Both species are known for their fast maturity, high storability, and the ability to grow on poor soil (Yabuno, 1987). ICRISAT currently holds 743 accessions of these barnyard millets from nine countries, with a core collection of 89 varieties recently established (Upadhyaya et al., 2014). Other significant collections can be found at NIAS and the USDA (Hilu, 1994). Sequence data and genetic map availability for both millets are generally low (Dwivedi et al., 2012). Initial transgenic work has been reported on the Japanese variety, but callus regeneration protocols have been reported for both species (Gupta et al., 2001; Kothari et al., 2005).

In addition to the two cultivated species, research has also been conducted on 20–30 wild *Echinochloa* barnyard millet relatives, some of which have agriculturally interesting traits including rice-mimicry and perennial growth habit. Hybridization within the genus is rampant, and is thought to have contributed to the evolution and current diversity of barnyard millets (Hilu, 1994; Yamaguchi et al., 2005).

Japanese Barnyard Millet (*Echinochloa esculenta*)

Japanese barnyard millet originated in eastern Asia (Figure 2) from its wild counterpart *E. crus-galli*, “barnyard grass” (Yabuno, 1987; Hilu, 1994). It can be differentiated from the Indian species by its larger, awned spikelets with glumes that appear papery instead of membranous (de Wet et al., 1983c). It is tolerant to cold and was historically grown in areas where the climate or land did not suit rice production, particularly in the north of Japan (Yabuno, 1987). In Japan, folklore states that barnyard millet originated from the dead body of a god. Along with proso millet, it makes up part of the “Gokoku,” a general term for five staple grains (Yabuno, 1987). Japanese barnyard millet has been found in the coffins of 800-year-old mummies from the Iwate prefecture, and documents from the 1700s list different cultivars organized by maturity time (Yabuno, 1987). Its historical importance might be attributed to the relief it provided in times of rice crop failure. However, Japanese barnyard millet production

has sharply decreased in the last century due to the introduction of cold-tolerant rice varieties and better irrigation practices (Yabuno, 1987). Nevertheless, today it remains the most common millet consumed in Japan, with reported health benefits common to many of the small millets such as its ability to lower plasma glucose concentration, insulin, adiponectin and tumor necrosis factor- α when fed to diabetic mice (Nishizawa et al., 2009). The protein content of Japanese barnyard millet is twice as high as that of rice (Yabuno, 1987). Across genotypes there is diversity in the levels of proteins and healthy lipids, with one genotype suggested as having particularly beneficial antioxidant activity (Kim et al., 2011).

Unlike other small millets consumed in East Asian countries such as foxtail and proso, barnyard millet has no glutinous variety. However, some landraces have been identified which contain very low levels of amylose due to a deletion in one of three waxy genes. One such landrace, “Noge-Hie,” was treated with γ -radiation resulting in progeny lacking the Waxy (Wx) protein (Hoshino et al., 2010). The trait was stably inherited, and this new glutinous variety (“Chojuromochi” in Japan) might be useful for increasing demand for millet products among Japanese consumers.

The morphological and physiological diversity of Japanese barnyard millet is suggested to be high (Nozawa et al., 2006). Flowering time, inflorescence shape, and spikelet pigmentation, among other features, vary across landraces. The species can be grouped into the races *utilis* and *intermedia* (Upadhyaya et al., 2014). Molecular diversity studies for Japanese barnyard millet have begun using the non-coding regions of chloroplast DNA as well as nuclear molecular markers (RAPDs, SSRs) and isozymes, although these studies appear to be limited in their sample number (Hilu, 1994; Nakayama et al., 1999; Yamaguchi et al., 2005; Nozawa et al., 2006). Though DNA sequence information in Japanese barnyard millet is otherwise lacking, studies performed on the closely related barnyard grass (*E. crus-galli*) have generated important sequence information. For example, extensive transcriptomic profiling and annotation have been performed on herbicide resistant varieties of barnyard grass resulting in 74 ESTs, which might be adapted to the study of the cultivated relative (Li et al., 2013; Yang et al., 2013).

Indian Barnyard Millet (*E. frumentacea*)

Indian barnyard millet, or sawa, was domesticated in India (Figure 2) across its current range from its wild counterpart *E. colona*, “jungle rice” (Yabuno, 1987; Hilu, 1994). In India, this millet is either harvested as a weed along with a main crop or is grown in a mixture with finger millet and foxtail millet (Gupta et al., 2009b). It is generally cultivated on hilly slopes in tribal areas where few other agricultural options exist and is indispensable in the northwest Himalayan region (Gupta et al., 2009b). Quick maturity makes the species well-adapted to regions with little rainfall (Channappagoudar et al., 2008). Indian barnyard millet contains antifeedants which are present at concentrations higher than in rice, and it displays resistance to the feeding activity of brown planthopper (Kim et al., 2008). In central Africa it is fermented to make beer or used for food, and has been found in the intestines of pre-dynastic Egyptian mummies (de Wet et al.,

1983c). When fed to diabetic humans, significant reductions of blood glucose levels and LDL cholesterol have been reported (Ugare et al., 2014).

Significant phenotypic variation is observed in Indian barnyard millet. Four morphological races (*laxa*, *robusta*, *intermedia*, and *stolonifera*) were recognized by de Wet in 1983 based on the lengths of flag leaves, peduncles, inflorescences, racemes, as well as plant height and basal tiller number. Race *laxa* is endemic to the Sikkim Himalayas and only available in a few collections (de Wet et al., 1983c). More recently, a variety of morphological parameters were examined, and principle component analysis (PCA) indicated three morphotypes corresponding to races *robusta*, *intermedia*, and *stolonifera*; *laxa* was absent suggesting that efforts must be made to collect more of this race (Gupta et al., 2009b). The authors saw high variability in grain yield, straw yield, and number of productive tillers. They report that the number of racemes, flag leaf width, and internode length showed high correlation with grain yield and should be considered by breeders when performing selections, and promising donor genotypes of these and other traits have been reported (Channappagoudar et al., 2008; Gupta et al., 2009b). Variation across genotypes in photosynthesis and related traits such as transpiration and stomatal conductance has also been observed (Subrahmanyam and Rathore, 1999). Grain smut (*Ustilago panici-frumentacei*) is a major hindrance of yield, but progress has been made in advanced breeding lines which display low susceptibility when compared to other accessions in which high variability remains (Gupta et al., 2009a).

An early study (Hilu, 1994) using RAPD markers suggested that the sequence diversity of Indian barnyard millet is significantly higher than the Japanese species, perhaps because of multiple domestication events in different locations across India (Hilu, 1994). Variation of markers was 44%, which is high when considering the inbreeding nature of the crop (Hilu, 1994). However, more comprehensive studies are needed that utilize a greater number of molecular markers and genotypes. Similarly, DNA sequence analyses are lacking in Indian barnyard millet.

Little Millet (*Panicum sumatrense*)

Also called *sama*, little millet is cultivated to a limited extent in India, Sri Lanka, Pakistan, Myanmar, and other South East Asian countries (Table 1) (Hiremath et al., 1990). In India it is important to tribes of the Eastern Ghat mountains and grown in combination with other millets (Hiremath et al., 1990). Little millet is a domesticated form of the weedy species *Panicum psilopodium* (de Wet et al., 1983a). The chromosomes of hybrids of *Panicum sumatrense* and *P. psilopodium* pair almost perfectly with only a single quadrivalent, indicating that divergence between the two species may have initially occurred through a single reciprocal translocation (Hiremath et al., 1990). Hybrid plants are fertile and vigorous with non-shattering spikelets, and thus introgression of genes between the two species is common (Hiremath et al., 1990). This hybridization ability combined with its wide range of cultivation across India suggests that little millet was domesticated independently several times, although exact dates remain undetermined (de Wet et al., 1983a). Little millet is comparable to other cereals in terms of fiber, fat, carbohydrates, and protein, and rich in phytochemicals including phenolic acids,

flavonoids, tannins, and phytate (Pradeep and Guha, 2011). Like many other small millets, it is drought, pest and salt tolerant (Sivakumar et al., 2006b; Bhaskaran and Panneerselvam, 2013; Ajithkumar and Panneerselvam, 2014). The time to maturity for most cultivars is about 90 days (de Wet et al., 1983a).

Little millet is divided into two races based on panicle morphology, *nana* and *robusta*. Race *nana* matures faster and produces less biomass than *robusta* (de Wet et al., 1983a). In a tribal area of the Indian Kolli hills, diversity among locally grown landraces of little millet was found to be high for all morphological traits measured both within and between landraces despite a small sampling area (Arunachalam et al., 2005). High diversity, heritability and genetic advancement was observed in terms of yield and productive tillers in a collection of 109 landraces, meaning that the crop might be a good candidate for varietal development (Nirmalakumari et al., 2010). A different collection of 460 accessions of little millet held by ICRISAT displayed genetic variation for most of the traits examined (Upadhyaya et al., 2014). A core collection of 56 genotypes was identified which was representative of the entire seed bank. Increased heritable lodging resistance has been introduced to a population of little millet with γ -ray mutational breeding (Nirmalakumari et al., 2007).

The molecular biology of little millet has been explored to a limited extent. As part of a study to identify seven millet species based on their chloroplast DNA, the *trnS-psbC* gene region was characterized and subjected to RFLP analysis (Parani et al., 2001). This study showed that it was possible to distinguish all the millet species when the enzymes *Hae*III and *Msp*I were used in combination. To investigate mechanisms behind little millet's high prolamine content, a zein-like storage protein was isolated and sequenced (Sivakumar et al., 2006a). Furthermore, α -amylase from little millet has been isolated and characterized in terms of biomass and optimum pH (Usha et al., 2011). To the best of our knowledge, no protocols for callus regeneration or transgenic technology have been published. Little millet is perhaps the least studied of the small millet species and there is much that requires investigation, including the establishment of a genetic map and sequenced genome.

Trends, Gaps and Recommendations on How to Foster Diversity within Orphaned Small Millets for the New Green Revolution

The World Summit on Food Security has set a target of 70% more food production by 2050, requiring annual increases of 44 million tons, 38% above current annual increases (Tester and Langridge, 2010). Climate change will cause additional difficulties as many regions are becoming drier with increasingly severe weather patterns (Dai, 2011), and fossil-fuel based nitrogen use is increasingly restricted by legislation intended to slow climate change (Tester and Langridge, 2010). The small millets have the potential to meet these challenges, given their drought tolerance and ability to grow under low input conditions, along with other health-promoting traits valued by humans. Unfortunately, the small millets suffer from low yields (only 0.8 tons grain per hectare) (Plaza-Wüthrich and Tadele, 2012). For the small millets to succeed, priority traits for breeding will need to include

improving yield under stress conditions (low input, salt, drought, pests, pathogens). Fortunately, an attractive feature of the small millets is that they continue to be cultivated in remote areas which has preserved their biodiversity, giving breeders potential access to unique genes for crop improvement. Due to limited resources, however, current efforts thus far have concentrated primarily on characterizing and reporting the extensive diversity present in seed banks, with few genetic and genomic tools available to exploit this biodiversity for crop improvement. A further challenge in some species (e.g., foxtail millet) is persistent cross-hybridization with wild relatives. Improved varieties of small millets could play a role in the “New Green Revolution”—a term coined to reflect novel strategies which will be required to deal with complex challenges in developing nations including increasing population and ever-diminishing arable land (Den Herder et al., 2010).

Exploiting Diversity within Seed Banks

Diversity is the basis of crop improvement. As described in this review, the small millets possess considerable morphological and genetic sequence variation that can be used by breeders to generate improved varieties. Seed banks across the globe conserve collections of small millets as shown in **Table 2**, but a challenge is that less diverse germplasm is available for species that are cultivated in a limited geographic region. For example, little millet, which is mainly grown in the Eastern Ghats of India, is represented by a collection of only 466 accessions (Upadhyaya et al., 2014). By contrast, ICRISAT currently holds 6804 accessions of finger millet, a crop widely grown on 1.8 million ha throughout India with extensive cultivation in Eurasia and Africa³. Core collections follow the same patterns, with several reported for finger millet but only one for little millet (Upadhyaya et al., 2014). It is essential that core collections be established for all of the millets, however, especially at larger seed banks, to facilitate efficient trait selection. As modern small millet cultivation for human consumption typically occurs in poor nations (with some exceptions), the seed bank infrastructure and associated reporting in the scientific literature and in online databases is sparse and difficult for breeders from foreign nations to access. Furthermore, trait descriptions for each accession are often not reported. Improved funding, coordination, communication and sharing of genetic resources are needed to overcome these problems.

Harvesting Genes from the Wild

Though interspecific hybrids between some cultivated and wild millets can be problematic, the wild relatives of the small millets may serve as donors of useful genes for crop improvement (e.g., herbicide resistance). To enable breeding, the hybridization ability of Indian and Japanese barnyard millet (Yabuno, 1962; de Wet et al., 1983c) may thus serve as an advantage. However, full realization of this breeding potential may require embryo rescue techniques to bring weak F₁ progeny to adulthood (Plaza-Wüthrich and Tadele, 2012) and better access by breeding programs to wild germplasm (Hajjar and Hodgkin, 2007). Today, the

wild germplasm is sometimes studied only from a weed science perspective (Peterson and Nalewaja, 1992; Dilday et al., 2001).

Combining Traditional Knowledge of Diversity with Modern Techniques

Small millets are often grown in remote regions of the world, and hence significant traditional knowledge of millet diversity persists that can serve as a valuable resource for crop improvement. Isolated farming communities often cultivate dozens of locally known millet landraces that are valued for a wide variety of traits (e.g., short duration to combat delayed rains as the result of climate change). Farmers use a complex system to classify their landraces, and in some instances this classification is considered more informative than scientific phylogeny (Rengalakshmi, 2005). On the opposite end of the technological spectrum, research using simple DNA barcoding *in lieu* of larger numbers of molecular markers is being attempted to classify the small millets down to the landrace level (Newmaster et al., 2013). A unique opportunity in the small millets is combining traditional knowledge with molecular techniques to characterize diversity for the purposes of crop improvement.

The Need for Complete Linkage Maps, Molecular Markers and Genome Sequences

As described above, in some species, markers including RFLPs, AFLPs, ESTs, and SSRs have been linked to beneficial traits including stress tolerance (Lata et al., 2011). Other, less conventional selective biomarkers have been suggested including differing ratios of transcription factors under stress (Gupta et al., 2014a). However, several small millets lack molecular and genetic markers (e.g., little millet and kodo millet) and no robust linkage maps appear to exist (Dwivedi et al., 2012). Genome and EST sequencing efforts will assist in the development of molecular markers in these species, along with using reference genomes (e.g., from major cereal relatives) to identify orthologous markers. Currently, only the foxtail millet genome has been sequenced and published (Bennetzen et al., 2012; Zhang et al., 2012).

Advances in Transgene Research and Molecular Mechanisms

As noted in this review, detailed protocols for callus regeneration and transgene protocols have been published for all small millet species except little millet (Kothari et al., 2005; Ceasar and Ignacimuthu, 2009; Plaza-Wüthrich and Tadele, 2012). Since small millet women farmers toil in the drudgery of removing weeds manually (Rengalakshmi, 2005), an attractive transgene trait may be glyphosate herbicide resistance (RoundupReady).

As the small millets are respected by traditional farmers for their extreme abiotic and biotic stress resistance, an understanding of the molecular mechanisms underlying these traits may lead to agronomic improvement of related major cereals. Unfortunately millet diversity remains largely unexplored at the level of molecular mechanism, with the exception of a limited number of studies noted earlier. One especially attractive target will be to understand the ability of barnyard millet to grow under extremely low nitrogen conditions.

³<http://www.icrisat.org/crop-fingermillet.htm>



FIGURE 3 | Indigenous technologies and practices of modern small millet farmers. (A) A typical granary in the Eastern Ghats of India used for small millet storage. (B) A woman farmer in Northern India holds a basket used for separating millet grain from chaff. She stands beside a manual millstone used for grinding millet grain into flour. Source: M. Raizada.

Socio-Economic Constraints

Despite the promise of the small millets, various socio-economic constraints have limited their consumption and hence contributed to a loss of cultivated diversity:

First, a major reason why the small millets are declining in production is that these crops are typically labor-intensive; women are often responsible for manual post-harvest processing, grain threshing and milling (Rengalakshmi, 2005) (Figure 3). To overcome this obstacle, inexpensive machinery is needed.

As noted above, a second challenge to greater adoption of small millets is their comparatively low yield (Plaza-Wüthrich and Tadele, 2012) as a result of the lack of scientific attention. However, the benefits of adding millet to the cropping system may outweigh the drawbacks of low yield (e.g., to combat local protein deficiency or crop failure in stressful environments) (Plaza-Wüthrich and Tadele, 2012). Furthermore, the small millets can be grown in very stressful environments, where major cereals may fail.

Third, family-farm-level diversity is heavily affected by community access to seed which may be limited by current rural seed systems (Nagarajan et al., 2007). However, the presence of local seed markets has been found to increase millet diversity indicating that such markets may serve as good points of introduction for improved varieties.

Finally, agricultural policies in different nations have negatively impacted the cultivation and research of small millets. Production in many areas is becoming displaced by mainstream cereals: in Kenya, the focus has been placed on the cultivation of maize instead of finger millet (Dida et al., 2008), while in Northern Japan, cold-tolerant rice has almost completely replaced barnyard millet (Yabuno, 1987). Reduced cultivation of

these millets in financially-rich countries like Japan is problematic, because it may decrease global research funding for these crops. However, recent reports revealing medicinal and nutritional benefits of these species (absence of gluten, cancer inhibition, control of blood-glucose and cholesterol) might catalyze consumer interest and hence funding in the developed world (Hegde et al., 2005; Nishizawa et al., 2009; Kim et al., 2011; Zhang et al., 2014). Nevertheless, landraces from these areas should be preserved in seed banks to ensure their conservation.

Given these socio-economic constraints, millets must not be blindly advocated in the developing world in biodiversity strategies. Prior to their introduction, multi-disciplinary surveys must be undertaken with local farmers concerning their nutrition, seed availability, economy, climate, and other crops in the cropping system.

Conclusions

Modern agriculture is characterized by dominance of a few crop species with a trend toward genetic homogenization as a result of the global exchange of alleles via breeding. In contrast, traditional farmer landraces of the small millets continue to be cultivated under relative genetic isolation, and hence provide living examples of genetic and phenotypic biodiversity in contemporary agriculture. The small millets are valued by traditional farmers for their nutritional content and health promoting properties, ability to grow under low input conditions and tolerance to extreme environmental stress, especially drought. In a world facing limiting natural resources and climate change, these crops thus hold tremendous potential as valuable instruments in the toolkit of the New Green Revolution. It is hoped that germplasm resources combined with modern genomic tools can help to accelerate exploitation of this biodiversity.

Author Contributions

Both TG and MR conceived of the manuscript. TG wrote the manuscript and MR edited the manuscript. Both authors read and approved the final manuscript.

Acknowledgments

We thank Dr. Malinda Thilakarathna (University of Guelph, Raizada Lab) for providing photos of millet cropping systems, and Dr. Kirit Patel (Canadian Mennonite University) for inspiring this review. TG received partial scholarship support from the Queen Elizabeth II Graduate Scholarship in Science and Technology and additional support from a grant to MR from the International Development Research Centre (IDRC) and the Canadian Department of Foreign Affairs, Trade and Development (DFATD) as part of the CIFSDF program.

References

Abrahamson, S., Bender, M. A., Conger, A. D., and Wolff, S. (1973). Uniformity of radiation-induced mutation rates among different species. *Nature* 245, 460–462. doi: 10.1038/245460a0

Aduguna, A. (2007). The role of introduced sorghum and millets in Ethiopian agriculture. *SAT J.* 3, 1–4.

Ajithkumar, I. P., and Panneerselvam, R. (2014). ROS scavenging system, osmotic maintenance, pigment and growth status of *Panicum sumatrense* roth. Under drought stress. *Cell Biochem. Biophys.* 68, 587–595. doi: 10.1007/s12013-013-9746-x

Anjaneyulu, E., Reddy, P. S., Sunita, M. S., Kishor, P. B., and Meriga, B. (2014). Salt tolerance and activity of antioxidative enzymes of transgenic finger millet over-expressing a vacuolar H⁺-pyrophosphatase gene (SbVPPase) from *Sorghum bicolor*. *J. Plant Physiol.* 171, 789–798. doi: 10.1016/j.jplph.2014.02.001

Araki, M., Numaoka, A., Kawase, M., and Fukunaga, K. (2012). Origin of waxy common millet, *Panicum miliaceum* L. in Japan. *Genet. Res. Crop Evol.* 59, 1303–1308. doi: 10.1007/s10722-011-9755-9

Arunachalam, V., Rengalakshmi, R., and Raj, M. S. K. (2005). Ecological stability of genetic diversity among landraces of little millet (*Panicum sumatrense*) in south India. *Genet. Res. Crop Evol.* 52, 15–19. doi: 10.1007/s10722-005-6693-4

Babu, B. K., Agrawal, P. K., Pandey, D., Jaiswal, J. P., and Kumar, A. (2014a). Association mapping of agro-morphological characters among the global collection of finger millet genotypes using genomic SSR markers. *Mol. Biol. Rep.* 41, 5287–5297. doi: 10.1007/s11033-014-3400-6

Babu, B. K., Agrawal, P. K., Pandey, D., and Kumar, A. (2014b). Comparative genomics and association mapping approaches for *opaque2* modifier genes in finger millet accessions using genic, genomic and candidate gene-based simple sequence repeat markers. *Mol. Breed.* 34, 1261–1279. doi: 10.1007/s11032-014-0115-2

Babu, B. K., Dinesh, P., Agrawal, P. K., Sood, S., Chandrashekara, C., Bhatt, J. C., et al. (2014c). Comparative genomics and association mapping approaches for blast resistant genes in finger millet using SSRs. *PLoS ONE* 9:e99182. doi: 10.1371/journal.pone.0099182

Babu, B. K., Pandey, D., Agrawal, P. K., Sood, S., and Kumar, A. (2014d). *In silico* mining, type and frequency analysis of genic microsatellites of finger millet (*Eleusine coracana* (L.) Gaertn.): a comparative genomic analysis of NBS-LRR regions of finger millet with rice. *Mol. Biol. Rep.* 41, 3081–3090. doi: 10.1007/s11033-014-3168-8

Babu, R. N., Jyothi, M. N., Sharadamma, N., Sahu, S., Rai, D. V., and Devaraj, V. R. (2013). Computational identification of conserved micro RNAs from kodo millet (*Paspalum scrobiculatum*). *Afr. Crop Sci. J.* 21, 75–83.

Bagdi, A., Baláz, G., Schmidt, J., Szatmári, M., Schoenlechner, R., Berghofer, E., et al. (2011). Protein characterization and nutrient composition of Hungarian proso millet varieties and the effect of decortication. *Acta Aliment.* 40, 128–141. doi: 10.1556/AAlim.40.2011.1.15

Bai, H., Cao, Y., Quan, J., Dong, L., Li, Z., Zhu, Y., et al. (2013). Identifying the genome-wide sequence variations and developing new molecular markers for genetics research by re-sequencing a landrace cultivar of foxtail millet. *PLoS ONE* 8:e73514. doi: 10.1371/journal.pone.0073514

Baltensperger, D. D. (1996). “Foxtail and proso millet,” in *Trends in New Crops and New Uses*, eds J. Janick and A. Whipkey (Alexandria: ASHS Press), 182–190.

Baltensperger, D. D. (2002). “Progress with proso, pearl and other millets,” in *Trends in New Crops and New Uses*, eds J. Janick and A. Whipkey (Alexandria: ASHS Press), 100–103.

Becker, M., and Johnson, D. E. (2001). Cropping intensity effects on upland rice yield and sustainability in West Africa. *Nutr. Cycl. Agroecosyst.* 59, 107–117. doi: 10.1023/A:1017551529813

Benin, S., Smale, M., Pender, J., Gebrernedhin, B., and Ehui, S. (2004). The economic determinants of cereal crop diversity on farms in the Ethiopian highlands. *Agric. Econ.* 31, 197–208. doi: 10.1111/j.1574-0862.2004.tb00257.x

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

Bhaskaran, J., and Panneerselvam, R. (2013). Accelerated reactive oxygen scavenging system and membrane integrity of two *Panicum* species varying in salt tolerance. *Cell Biochem. Biophys.* 67, 885–892. doi: 10.1007/s12013-013-9576-x

Bhatt, D., Negi, M., Sharma, P., Saxena, S. C., Dobriyal, A. K., and Arora, S. (2011). Responses to drought induced oxidative stress in five finger millet varieties differing in their geographical distribution. *Physiol. Mol. Biol. Plants* 17, 347–353. doi: 10.1007/s12298-011-0084-4

Bisht, M. S., and Mukai, Y. (2001). Genomic *in situ* hybridization identifies genome donor of finger millet (*Eleusine coracana*). *Theor. Appl. Genet.* 102, 825–832. doi: 10.1007/s001220000497

Blench, R. (1997). Neglected species, livelihoods and biodiversity in difficult areas: how should the public sector respond? *Nat. Resour. Perspect.* 23, 1–10.

Bonthala, V. S., Muthamilarasan, M., Roy, R., and Prasad, M. (2014). FmTFDb: a foxtail millet transcription factors database for expediting functional genomics in millets. *Mol. Biol. Rep.* 41, 6343–6348. doi: 10.1007/s11033-014-3574-y

Borlaug, N. E. (2000). Ending world hunger. The promise of biotechnology and the threat of antiscience zealotry. *Plant Physiol.* 124, 487–490. doi: 10.1104/pp.124.2.487

Brown, A. H. D. (1989). Core collections: a practical approach to genetic resources management. *Genome* 31, 818–824. doi: 10.1139/g89-144

Burton, G. W. (1940). A cytological study of some species in the genus *Paspalum*. *J. Agric. Res.* 60, 193–198.

Cakmak, I. (2008). Enrichment of cereal grains with zinc: Agronomic or genetic biofortification? *Plant Soil* 302, 1–17. doi: 10.1007/s11104-007-9466-3

Cesar, S. A., and Ignacimuthu, S. (2009). Genetic engineering of millets: current status and future prospects. *Biotechnol. Lett.* 31, 779–788. doi: 10.1007/s10529-009-9933-4

Cesar, S. A., and Ignacimuthu, S. (2010). Effects of cytokinins, carbohydrates and amino acids on induction and maturation of somatic embryos in kodo millet (*Paspalum scrobiculatum* Linn.). *Plant Cell Tissue Organ Cult.* 102, 153–162. doi: 10.1007/s11240-010-9716-6

Cesar, 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

Chandrasekara, A., and Shahidi, F. (2011a). Antiproliferative potential and DNA scission inhibitory activity of phenolics from whole millet grains. *J. Funct. Foods* 3, 159–170. doi: 10.1016/j.jff.2011.03.008

Chandrasekara, A., and Shahidi, F. (2011b). Determination of antioxidant activity in free and hydrolyzed fractions of millet grains and characterization of their phenolic profiles by HPLC-DAD-ESI-MS. *J. Funct. Foods* 3, 144–158. doi: 10.1016/j.jff.2011.03.007

Channappagoudar, B. B., Hiremath, S. M., Biradar, N. R., Koti, R. V., and Bharamagoudar, T. D. (2008). Influence of morpho-physiological and biochemical traits on the productivity of barnyard millet. *Karnataka J. Agric. Sci.* 20, 477–480.

Cho, Y.-I., Chung, J.-W., Lee, G.-A., Ma, K.-H., Dixit, A., Gwag, J.-G., et al. (2010). Development and characterization of twenty-five new polymorphic microsatellite markers in proso millet (*Panicum miliaceum* L.). *Genes Genomics* 32, 267–273. doi: 10.1007/s13258-010-0007-8

Colosi, J. C., and Schaal, B. A. (1997). Wild proso millet (*Panicum miliaceum*) is genetically variable and distinct from crop varieties of proso millet. *Weed Sci.* 45, 509–518.

D’Ennequin, M. L. T., Panaud, O., Brown, S., Siljak-Yakovlev, S., and Sarr, A. (1998). First evaluation of nuclear DNA content in *Setaria* genus by flow cytometry. *J. Hered.* 86, 556–559.

Dai, A. (2011). Drought under global warming: a review. *Wiley Interdiscip. Rev. Clim. Chang.* 2, 45–65. doi: 10.1002/wcc.81

Darmency, H., and Pernes, J. (1985). Use of wild *Setaria viridis* (L.) Beauv. to improve triazine resistance in cultivated *S. italica* (L.) by hybridization. *Weed Res.* 25, 175–179. doi: 10.1111/j.1365-3180.1985.tb00633.x

Darmency, H., and Pernes, J. (1989). Agronomic performance of a triazine resistant foxtail millet (*Setaria italica* (L.) Beauv.). *Weed Res.* 29, 147–150. doi: 10.1111/j.1365-3180.1989.tb00853.x

De Datta, S. K., Tauro, A. C., and Balaong, S. N. (1968). Effect of plant type and nitrogen level on the growth characteristics and grain yield of Indica rice in the tropics. *Agron. J.* 60, 663–647. doi: 10.2134/agronj1968.00021962006000060017x

Den Herder, G., Van Isterdael, G., Beeckman, T., and De Smet, I. (2010). The roots of a new green revolution. *Trends Plant Sci.* 15, 600–607. doi: 10.1016/j.tplants.2010.08.009

Devi, P. B., Vijayabharathi, R., Sathyabama, S., Malleshi, N. G., and Priyadarisini, V. B. (2014). Health benefits of finger millet (*Eleusine coracana* L.) polyphenols and dietary fiber: a review. *J. Food Sci. Technol.* 51, 1021–1040. doi: 10.1007/s13197-011-0584-9

de Wet, J. M. J. (1992). "The three phases of cereal domestication," in *Grass Evolution and Domestication*, ed G. P. Chapman (Cambridge: Cambridge University Press), 176–191.

de Wet, J. M. J., Prasada Rao, K. E., and Brink, D. E. (1983a). Systematics and domestication of *Panicum sumatrense* (Gramineae). *J. d'agriculture Tradit. Bot. appliquée* 30, 159–168.

de Wet, J. M. J., Rao, K. E. P., Mengesha, M. H., and Brink, D. E. (1983b). Diversity in kodo millet, *Paspalum scrobiculatum*. *Econ. Bot.* 37, 159–163. doi: 10.1007/BF02858779

de Wet, J. M. J., Rao, K. E. P., Mengesha, M. H., and Brink, D. E. (1983c). Domestication of sawa millet. *Econ. Bot.* 37, 283–291. doi: 10.1007/BF02858883

Dida, M. M., and Devos, K. M. (2006). "Finger millet," in *Cereals and Millets*, ed C. Kole (New York, NY: Springer), 333–343.

Dida, M. M., Srinivasachary, Ramakrishnan, S., Bennetzen, J. L., Gale, M. D., and Devos, K. M. (2007). The genetic map of finger millet, *Eleusine coracana*. *Theor. Appl. Genet.* 114, 321–332. doi: 10.1007/s00122-006-0435-7

Dida, M. M., Wanyera, N., Harrison Dunn, M. L., Bennetzen, J. L., and Devos, K. M. (2008). Population structure and diversity in finger millet (*Eleusine coracana*) germplasm. *Trop. Plant Biol.* 1, 131–141. doi: 10.1007/s12042-008-9012-3

Dilday, R. H., Mattice, J. D., Moldenhauer, K. A., and Yan, W. (2001). Allelopathic potential in rice germplasm against ducksalad, redstem and barnyard grass. *J. Crop Prot.* 4, 287–301. doi: 10.1300/J144v04n02_11

Diwan, N., McIntosh, M. S., and Bauchan, G. R. (1995). Methods of developing a core collection of annual *Medicago* species. *Theor. Appl. Genet.* 90, 755–761. doi: 10.1007/BF00222008

Doust, A. N., Kellogg, E. A., Devos, K. M., and Bennetzen, J. L. (2009). Foxtail millet: a sequence-driven grass model system. *Plant Physiol.* 149, 137–141. doi: 10.1104/pp.108.129627

Dwivedi, S., Upadhyaya, H., Senthilvel, S., Hash, C., Fukunaga, K., Diao, X., et al. (2012). "Millets: genetic and genomic resources," in *Plant Breeding Reviews*, ed J. Janick (Hoboken, NJ: John Wiley and Sons, Inc.), 247–374.

Ejeta, G. (2010). African green revolution needn't be a mirage. *Science* 327, 831–832. doi: 10.1126/science.1187152

Evenson, R. E., and Gollin, D. (2003). Assessing the impact of the green revolution, 1960 to 2000. *Science* 300, 758–762. doi: 10.1126/science.1078710

Geervani, P., and Eggum, B. O. (1989). Nutrient composition and protein quality of minor millets. *Plant Foods Hum. Nutr.* 39, 201–208. doi: 10.1007/BF01091900

Graybosch, R. A., and Baltensperger, D. D. (2009). Evaluation of the waxy endosperm trait in proso millet (*Panicum miliaceum*). *Plant Breed.* 128, 70–73. doi: 10.1111/j.1439-0523.2008.01511.x

Gull, A., Jan, R., Nayik, G. A., Prasad, K., and Kumar, P. (2014). Significance of finger millet in nutrition, health and value added products: a review. *J. Environ. Sci. Comput. Sci. Eng. Technol.* 3, 1601–1608.

Gupta, A., Joshi, D., Mahajan, V., and Gupta, H. S. (2009a). Screening barnyard millet germplasm against grain smut (*Ustilago panici-frumentacei* Brefeld). *Plant Genet. Resour.* 8, 52–54. doi: 10.1017/S1479262109990141

Gupta, A., Mahajan, V., Kumar, M., and Gupta, H. S. (2009b). Biodiversity in the barnyard millet (*Echinochloa frumentacea* Link, Poaceae) germplasm in India. *Genet. Resour. Crop Evol.* 56, 883–889. doi: 10.1007/s10722-009-9462-y

Gupta, P., Raghuvanshi, S., and Tyagi, A. K. (2001). Assessment of the efficiency of various gene promoters via biolistics in leaf and regenerating seed callus of millets, *Eleusine coracana* and *Echinochloa crusgalli*. *Plant Biotechnol.* 18, 275–282. doi: 10.5511/plantbiotechnology.18.275

Gupta, S., Gupta, S. M., Gupta, A. K., Gaur, V. S., and Kumar, A. (2014a). Fluctuation of Dof1/Dof2 expression ratio under the influence of varying nitrogen and light conditions: involvement in differential regulation of nitrogen metabolism in two genotypes of finger millet (*Eleusine coracana* L.). *Gene* 546, 327–335. doi: 10.1016/j.gene.2014.05.057

Gupta, S., Kumari, K., Muthamilarasan, M., Parida, S. K., and Prasad, M. (2014b). Population structure and association mapping of yield contributing agronomic traits in foxtail millet. *Plant Cell Rep.* 33, 881–893. doi: 10.1007/s00299-014-1564-0

Gupta, S., Kumari, K., Sahu, P. P., Vidapu, S., and Prasad, M. (2012). Sequence-based novel genomic microsatellite markers for robust genotyping purposes in foxtail millet [*Setaria italica* (L.) P. Beauv]. *Plant Cell Rep.* 31, 323–337. doi: 10.1007/s00299-011-1168-x

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

Heap, I. M. (1997). The occurrence of herbicide-resistant weeds worldwide. *Pestic. Sci.* 51, 235–243.

Hegde, B. R., and Gowda, L. (1986). Cropping systems and production technology for small millets in India. *Proc. First Int. Small Millets Workshop* 209–236.

Hegde, P. S., and Chandra, T. S. (2005). ESR spectroscopic study reveals higher free radical quenching potential in kodo millet (*Paspalum scrobiculatum*) compared to other millets. *Food Chem.* 92, 177–182. doi: 10.1016/j.foodchem.2004.08.002

Hegde, P. S., Rajasekaran, N. S., and Chandra, T. S. (2005). Effects of the antioxidant properties of millet species on oxidative stress and glycemic status in alloxan-induced rats. *Nutr. Res.* 25, 1109–1120. doi: 10.1016/j.nutres.2005.09.020

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

Hilu, K. W. (1994). Evidence from RAPD markers in the evolution of *Echinochloa* millets (Poaceae). *Plant Syst. Evol.* 189, 247–257. doi: 10.1007/BF00939730

Hirano, R., Naito, K., Fukunaga, K., Watanabe, K. N., Ohsawa, R., and Kawase, M. (2011). Genetic structure of landraces in foxtail millet (*Setaria italica* (L.) P. Beauv.) revealed with transposon display and interpretation to crop evolution of foxtail millet. *Genome* 54, 498–506. doi: 10.1139/g11-015

Hiremath, S. C., Patil, G. N. V., and Salimath, S. S. (1990). Genome homology and origin of *Panicum sumatrense* (Gramineae). *Cytologia (Tokyo)* 55, 315–319. doi: 10.1508/cytologia.55.315

Hoshino, T., Nakamura, T., Seimiya, Y., Kamada, T., Ishikawa, G., Ogasawara, A., et al. (2010). Production of a fully waxy line and analysis of waxy genes in the allohexaploid crop, Japanese barnyard millet. *Plant Breed.* 129, 349–355. doi: 10.1111/j.1439-0523.2009.01668.x

Hu, J., Zhu, J., and Xu, H. M. (2000). Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theor. Appl. Genet.* 101, 264–268. doi: 10.1007/s001220051478

Hu, X., Wang, J., Lu, P., and Zhang, H. (2009). Assessment of genetic diversity in broomcorn millet (*Panicum miliaceum* L.) using SSR markers. *J. Genet. Genomics* 36, 491–500. doi: 10.1016/S1673-8527(08)60139-3

Hu, Y. G., Zhu, J., Liu, F., Zhang, Z., Chai, Y., and Weining, S. (2008). Genetic diversity among Chinese landraces and cultivars of broomcorn millet (*Panicum miliaceum*) revealed by the polymerase chain reaction. *Ann. Appl. Biol.* 153, 357–364. doi: 10.1111/j.1744-7348.2008.00263.x

Hunt, H. V., Badakshi, F., Romanova, O., Howe, C. J., Jones, M. K., and Heslop-Harrison, J. S. P. (2014). Reticulate evolution in *Panicum* (Poaceae): the origin of tetraploid broomcorn millet, *P. miliaceum*. *J. Exp. Bot.* 65, 3165–3175. doi: 10.1093/jxb/eru161

Hunt, H. V., Campana, M. G., Lawes, M. C., Park, Y.-J., Bower, M. A., Howe, C., et al. (2011). Genetic diversity and phylogeography of broomcorn millet (*Panicum miliaceum* L.) across Eurasia. *Mol. Ecol.* 20, 4756–4771. doi: 10.1111/j.1365-294X.2011.05318.x

Hunt, H. V., Denyer, K., Packman, L. C., Jones, M. K., and Howe, C. J. (2010). Molecular basis of the waxy endosperm starch phenotype in broomcorn millet (*Panicum miliaceum* L.). *Mol. Biol. Evol.* 27, 1478–1494. doi: 10.1093/molbev/msq040

Hunt, H. V., Vander Linden, M., Liu, X., Motuzaite-Matuzeviciute, G., Colledge, S., and Jones, M. K. (2008). Millets across Eurasia: chronology and context of early records of the genera *Panicum* and *Setaria* from archaeological sites in the Old World. *Veg. Hist. Archaeobot.* 17, S5–S18. doi: 10.1007/s00334-008-0187-1

Jagga-Chugh, S., Kachhwaha, S., Sharma, M., Kothari-Chajer, A., and Kothari, S. L. (2012). Optimization of factors influencing microprojectile bombardment-mediated genetic transformation of seed-derived callus and regeneration of transgenic plants in *Eleusine coracana* (L.) Gaertn. *Plant Cell Tissue Organ Cult.* 109, 401–410. doi: 10.1007/s11240-011-0104-7

Jarret, R. L., Ozias-Akins, P., Phatak, S., Nadimpalli, R., Duncan, R., and Hiliard, S. (1995). DNA contents in *Paspalum* spp. determined by flow cytometry. *Genet. Resour. Crop Evol.* 42, 237–242. doi: 10.1007/BF02431258

Jayaraman, A., Puranik, S., Rai, N. K., Vidapu, S., Sahu, P. P., Lata, C., et al. (2008). cDNA-AFLP analysis reveals differential gene expression in response to salt stress in foxtail millet (*Setaria italica* L.). *Mol. Biotechnol.* 40, 241–251. doi: 10.1007/s12033-008-9081-4

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

Jia, X., Zhang, Z., Liu, Y., Zhang, C., Shi, Y., Song, Y., et al. (2009). Development and genetic mapping of SSR markers in foxtail millet (*Setaria italica* (L.) P. Beauv.). *Theor. Appl. Genet.* 118, 821–829. doi: 10.1007/s00122-008-0942-9

Jusuf, M., and Pernes, J. (1985). Genetic variability of foxtail millet (*Setaria italica* P. Beauv.): electrophoretic study of five isoenzyme systems. *Theor. Appl. Genet.* 71, 385–391. doi: 10.1007/BF00251177

Kalinova, J., and Moudry, J. (2006). Content and quality of protein in proso millet (*Panicum miliaceum* L.) varieties. *Plant Foods Hum. Nutr.* 61, 45–49. doi: 10.1007/s11130-006-0013-9

Karam, D., Westra, P., Niessen, S. J., Ward, S. M., and Figueiredo, J. E. F. (2006). Assessment of silver-stained AFLP markers for studying DNA polymorphism in proso millet (*Panicum miliaceum* L.). *Rev. Bras. Bot.* 26, 609–615. doi: 10.1590/S0100-84042006000400011

Karam, D., Westra, P., Nissen, S. J., Ward, S. M., and Figueiredo, J. E. F. (2004). Genetic diversity among proso millet (*Panicum miliaceum*) biotypes assessed by AFLP technique. *Planta Daninha* 22, 167–174. doi: 10.1590/S0100-83582004000200001

Kawase, M., Fukunaga, K., and Kato, K. (2005). Diverse origins of waxy foxtail millet crops in East and Southeast Asia mediated by multiple transposable element insertions. *Mol. Genet. Genomics* 274, 131–140. doi: 10.1007/s00438-005-0013-8

Kim, C.-S., Alamgir, K. M., Matsumoto, S., Tebayashi, S., and Koh, H.-S. (2008). Antifeedants of Indian barnyard millet, *Echinochloa frumentacea* Link, against brown planthopper, *Nilaparvata lugens* (Stål). *Z. Naturforsch.* 63, 755–760.

Kim, J. Y., Jang, K. C., Park, B.-R., Han, S.-I., Choi, K.-J., Kim, S.-Y., et al. (2011). Physicochemical and antioxidative properties of selected barnyard millet (*Echinochloa utilis*) species in Korea. *Food Sci. Biotechnol.* 20, 461–469. doi: 10.1007/s10068-011-0064-z

Kothari, S. L., Kumar, S., Vishnoi, R. K., Kothari, A., and Watanabe, K. N. (2005). Applications of biotechnology for improvement of millet crops: review of progress and future prospects. *Plant Biotechnol.* 22, 81–88. doi: 10.5511/plantbiotechnology.22.81

Kuběšová, M., Moravcová, L., Suda, J., Jarosík, V., and Pyšek, P. (2010). Naturalized plants have smaller genomes than their non-invading relatives: a flow cytometric analysis of the Czech alien flora. *Preslia* 82, 81–96.

Kulkarni, L. R., and Naik, R. K. (2000). Nutritive value, protein quality and organoleptic quality of kodo millet (*Paspalum scrobiculatum*). *Karnataka J. Agric. Sci.* 13, 125–129.

Kumar, A., Gaur, V. S., Goel, A., and Gupta, A. K. (2014a). *De novo* assembly and characterization of developing spikes transcriptome of finger millet (*Eleusine coracana*): a minor crop having nutraceutical properties. *Plant Mol. Biol. Rep.* doi: 10.1007/s11105-014-0802-5

Kumar, A., Kanwal, P., Gupta, A. K., Singh, B. R., and Gaur, V. S. (2014b). A full-length Dof1 transcription factor of finger millet and its response to a circadian cycle. *Plant Mol. Biol. Rep.* 32, 419–427. doi: 10.1007/s11105-013-0653-5

Kumar, A., Mirza, N., Charan, T., Sharma, N., and Gaur, V. S. (2014c). Isolation, characterization and immunolocalization of a seed dominant CaM from finger millet (*Eleusine coracana* L. Gartn.) for studying its functional role in differential accumulation of calcium in developing grains. *Appl. Biochem. Biotechnol.* 172, 2955–2973. doi: 10.1007/s12010-013-0714-0

Kushwaha, H., Jillo, K. W., Singh, V. K., Kumar, A., and Yadav, D. (2015). Assessment of genetic diversity among cereals and millets based on PCR amplification using Dof (DNA binding with One Finger) transcription factor gene-specific primers. *Plant Syst. Evol.* 301, 833–840. doi: 10.1007/s00606-014-1095-8

Lata, C., Bhutty, S., Bahadur, R. P., Majee, M., and Prasad, M. (2011). Association of an SNP in a novel DREB2-like gene SiDREB2 with stress tolerance in foxtail millet [*Setaria italica* (L.)]. *J. Exp. Bot.* 62, 3387–3401. doi: 10.1093/jxb/err016

Lata, C., and Prasad, M. (2013). Validation of an allele-specific marker associated with dehydration stress tolerance in a core set of foxtail millet accessions. *Plant Breed.* 132, 496–499. doi: 10.1111/j.1439-0523.2012.01983.x

Latha, A. M., Rao, K. V., and Reddy, V. D. (2005). Production of transgenic plants resistant to leaf blast disease in finger millet (*Eleusine coracana* (L.) Gaertn.). *Plant Sci.* 169, 657–667. doi: 10.1016/j.plantsci.2005.05.009

Li, C. H., Pao, W. K., and Li, H. W. (1942). Interspecific crosses in *Setaria*. *J. Hered.* 33, 351–355.

Li, G., Wu, S., Cai, L., Wang, Q., Zhao, X., and Wu, C. (2013). Identification and mRNA expression profile of glutamate receptor-like gene in quinolone-resistant and susceptible *Echinochloa crus-galli*. *Gene* 531, 489–495. doi: 10.1016/j.gene.2013.09.013

Li, Y., Wang, J., Cao, Y., Gao, W., Fang, J., and Lou, X. (1998). The use of genetic resources in crop improvement: lessons from China. *Genet. Resour. Crop Evol.* 45, 181–186. doi: 10.1023/A:1008691532378

Li, Y., and Wu, S. (1996). Traditional maintenance and multiplication of foxtail millet (*Setaria italica* (L.) P. Beauv.) landraces in China. *Euphytica* 87, 33–38. doi: 10.1007/BF00022961

Liu, M., Qiao, Z., Zhang, S., Wang, Y., and Lu, P. (2015). Response of broomcorn millet (*Panicum miliaceum* L.) genotypes from semi arid regions of China to salt stress. *Crop J.* 3, 57–66. doi: 10.1016/j.cj.2014.08.006

Liu, Q., Jiang, B., Wen, J., and Peterson, P. M. (2014). Low-copy nuclear gene and McGISH resolves polyploid history of *Eleusine coracana* and morphological character evolution in *Eleusine*. *Turk. J. Bot.* 38, 1–12. doi: 10.3906/bot-1305-12

Lu, H., Zhang, J., Liu, K., Wu, N., Li, Y., Zhou, K., et al. (2009). Earliest domestication of common millet (*Panicum miliaceum*) in East Asia extended to 10,000 years ago. *PNAS* 106, 7367–7372. doi: 10.1073/pnas.0900158106

M'Ribou, H. K., and Hilu, K. W. (1994). Detection of interspecific and intraspecific variation in *Panicum* millets through random amplified polymorphic DNA. *Theor. Appl. Genet.* 88, 412–416.

M'Ribou, H. K., and Hilu, K. W. (1996). Application of random amplified polymorphic DNA to study genetic diversity in *Paspalum scrobiculatum* L. (Kodo millet, Poaceae). *Genet. Resour. Crop Evol.* 43, 203–210.

Marles, M. A. S., Devine, M. D., and Hall, J. C. (1993). Herbicide resistance in *Setaria viridis* conferred by a less sensitive form of acetyl coenzyme a carboxylase. *Pesticide Biochem. Physiol.* 46, 7–14. doi: 10.1006/pest.1993.1031

McCann, S. J., and Cavers, P. B. (1988). Spread of proso millet (*Panicum miliaceum* L.) in Ontario, Canada. II. Dispersal by combines. *Weed Res.* 28, 67–72. doi: 10.1111/j.1365-3180.1988.tb0788.x

Morrison, I. N., Todd, B. G., and Nawolsky, K. M. (1989). Confirmation of trifluralin-resistant green foxtail (*Setaria viridis*) in Manitoba. *Weed Technol.* 3, 544–551.

Muthamilarasan, M., Venkata Suresh, B., Pandey, G., Kumari, K., Parida, S. K., and Prasad, M. (2014). Development of 5123 intron-length polymorphic markers for large-scale genotyping applications in foxtail millet. *DNA Res.* 21, 41–52. doi: 10.1093/dnare/dst039

Muza, F. R., Lee, D. J., Andrews, D. J., and Gupta, S. C. (1995). Mitochondrial DNA variation in finger millet (*Eleusine coracana* L. Gaertn.). *Euphytica* 81, 199–205. doi: 10.1007/BF00025434

Mysore, K. S., and Baird, V. (1997). Nuclear DNA content in species of *Eleusine* (Gramineae): a critical re-evaluation using laser flow cytometry. *Plant Syst. Evol.* 207, 1–11. doi: 10.1007/BF00985206

Nagarajan, L., Smale, M., and Glewwe, P. (2007). Determinants of millet diversity at the household-farm and village-community levels in the drylands of India: the role of local seed systems. *Agric. Econ.* 36, 157–167. doi: 10.1111/j.1574-0862.2007.00195.x

Nakayama, Y., Umemoto, S., and Yamaguchi, H. (1999). Identification of polyploid groups in the genus *Echinochloa* by isozyme analysis. *J. Weed Sci. Technol.* 44, 205–217. doi: 10.3719/weed.44.205

Neumann, K., Ballouche, A., and Klee, M. (1996). The emergence of plant food production in the West African Sahel: new evidence from northeast Nigeria and northern Burkina Faso. *Asp. Afr. Archaeol.* 441–448.

Newmaster, S. G., Ragupathy, S., Dhivya, S., Jijo, C. J., Sathishkumar, R., and Patel, K. (2013). Genomic valorization of the fine scale classification of small millet landraces in southern India. *Genome* 56, 123–127. doi: 10.1139/gen-2012-0183

Nirgude, M., Babu, B. K., Shambhavi, Y., Singh, U. M., Upadhyaya, H. D., and Kumar, A. (2014). Development and molecular characterization of genic

molecular markers for grain protein and calcium content in finger millet (*Eleusine coracana* (L.) Gaertn.). *Mol. Biol. Rep.* 41, 1189–1200. doi: 10.1007/s11033-013-2825-7

Nirmalakumari, A., Arulselvi, S., Ganapathy, S., Soufmanian, J., Senthil, N., and Devan, P. (2007). Gamma ray induced variation for lodging resistance and its associated characters in little millet (*Panicum sumatrense* Roth Ex-roem and schult.). *Madras Agric. J.* 94, 151–155.

Nirmalakumari, A., Salini, K., and Veerabadhiran, P. (2010). Morphological characterization and evaluation of little millet (*Panicum sumatrense* Roth ex. Roem. and Schultz) germplasm. *Electron. J. Plant Breed.* 1, 148–155.

Nirmalakumari, A., Sumathi, P., Joel, A. J., Kumaravadivel, N., Senthil, N., Devan, P., et al. (2008). A high yielding and early maturing panivaragu variety CO (PV) 5. *Madras Agric. J.* 95, 1–6.

Nishizawa, N., and Fudamoto, Y. (1995). The elevation of plasma concentration of high-density lipoprotein cholesterol in mice fed with protein from proso millet. *Biosci. Biotechnol. Biochem.* 52, 333–335. doi: 10.1271/bbb.59.333

Nishizawa, N., Sato, D., Ito, Y., Nagasawa, T., Hatakeyama, Y., Choi, M. R., et al. (2002). Effects of dietary protein of proso millet on liver injury induced by D-galactosamine in rats. *Biosci. Biotechnol. Biochem.* 66, 92–96. doi: 10.1271/bbb.66.92

Nishizawa, N., Togawa, T., Park, K.-O., Sato, D., Miyakoshi, Y., Inagaki, K., et al. (2009). Dietary Japanese millet protein ameliorates plasma levels of adiponectin, glucose, and lipids in type 2 diabetic mice. *Biosci. Biotechnol. Biochem.* 73, 351–360. doi: 10.1271/bbb.80589

Nozawa, S., Takahashi, M., Nakai, H., and Sato, Y.-I. (2006). Difference in SSR variations between Japanese barnyard millet (*Echinochloa esculenta*) and its wild relative *E. crus-galli*. *Breed. Sci.* 56, 335–340. doi: 10.1270/jssbs.56.335

Ogie-Odia, E. A., Mokwenye, A. I., Kekere, O., and Timothy, O. (2010). Comparative vegetative and foliar epidermal features of three *Paspalum* L. species in Edostate, Nigeria. *Ozean J. Appl. Sci.* 3, 29–38.

Pandey, G., Misra, G., Kumar, K., Gupta, S., Parida, S. K., Chattopadhyay, D., et al. (2013). Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet (*Setaria italica* (L.)). *DNA Res.* 20, 197–207. doi: 10.1093/dnarese/dst002

Parani, M., Rajesh, K., Lakshmi, M., Parducci, L., Szmidt, A. E., and Parida, A. (2001). Species identification in seven small millet species using polymerase chain reaction – restriction fragment length polymorphism of *trnS-psbC* gene region. *Genome* 44, 495–499. doi: 10.1139/g01-023

Park, K.-O., Ito, Y., Nagasawa, T., Choi, M.-R., and Nishizawa, N. (2014). Effects of dietary Korean proso-millet protein on plasma adiponectin, HDL cholesterol, insulin levels, and gene expression in obese type 2 diabetic mice. *Biosci. Biotechnol. Biochem.* 72, 2918–2925. doi: 10.1271/bbb.80395

Peterson, D. E., and Nalewaja, J. D. (1992). Environment influences green foxtail (*Setaria viridis*) competition with wheat (*Triticum aestivum*). *Weed Technol.* 6, 607–610.

Plaza-Wüthrich, S., and Tadele, Z. (2012). Millet improvement through regeneration and transformation. *Biotechnol. Mol. Biol. Rev.* 7, 48–61. doi: 10.5897/BMBR12.001

Pradeep, S. R., and Guha, M. (2011). Effect of processing methods on the nutraceutical and antioxidant properties of little millet (*Panicum sumatrense*) extracts. *Food Chem.* 126, 1643–1647. doi: 10.1016/j.foodchem.2010.12.047

Pradhan, A., Thakur, A., Patel, S., and Mishra, N. (2011). Effect of different nitrogen levels on kodo and finger millet under rainfed conditions. *Res. J. Agric. Sci.* 2, 136–138.

Puranik, S., Jha, S., Srivastava, P. S., Sreenivasulu, N., and Prasad, M. (2011). Comparative transcriptome analysis of contrasting foxtail millet cultivars in response to short-term salinity stress. *J. Plant Physiol.* 168, 280–287. doi: 10.1016/j.jplph.2010.07.005

Qi, X., Xie, S., Liu, Y., Yi, F., and Yu, J. (2013). Genome-wide annotation of genes and noncoding RNAs of foxtail millet in response to simulated drought stress by deep sequencing. *Plant Mol. Biol.* 83, 459–473. doi: 10.1007/s11103-013-0104-6

Qin, F. F., Zhao, Q., Ao, G. M., and Yu, J. J. (2008). Co-suppression of *Si401*, a maize pollen specific *Zm401* homologous gene, results in aberrant anther development in foxtail millet. *Euphytica* 163, 103–111. doi: 10.1007/s10681-007-9610-4

Rahman, H., Jagadeeshselvam, N., Valarmathi, R., Sachin, B., Sasikala, R., Senthil, N., et al. (2014). Transcriptome analysis of salinity responsiveness in contrasting genotypes of finger millet (*Eleusine coracana* L.) through RNA-sequencing. *Plant Mol. Biol.* 85, 485–503. doi: 10.1007/s11103-014-0199-4

Rajput, S. G., Plyler-harveson, T., and Santra, D. K. (2014). Development and characterization of SSR markers in proso millet based on switchgrass genomics. *Am. J. Plant Sci.* 5, 175–186. doi: 10.4236/ajps.2014.51023

Ramegowda, V., Senthil-Kumar, M., Nataraja, K. N., Reddy, M. K., Mysore, K. S., and Udayakumar, M. (2012). Expression of a finger millet transcription factor, EcNAC1, in tobacco confers abiotic stress-tolerance. *PLoS ONE* 7:e40397. doi: 10.1371/journal.pone.0040397

Ramegowda, V., Venkategowda, R., Jagadish, P., Govind, G., Hanumanthareddy, R.-R., Makarla, U., et al. (2013). Expression of a rice Zn transporter, OsZIP1, increases Zn concentration in tobacco and finger millet transgenic plants. *Plant Biotechnol. Rep.* 7, 309–319. doi: 10.1007/s11816-012-0264-x

Rao, B. L., and Husain, A. (1985). Presence of cyclopiazonic acid in kodo millet (*Paspalum scrobiculatum*) causing “kodua poisoning” in man and its production by associated fungi. *Mycopathologia* 89, 177–180. doi: 10.1007/BF00447028

Reddy, I. N. B. L., Reddy, D. S., Narasu, M. L., and Sivaramakrishnan, S. (2011). Characterization of disease resistance gene homologues isolated from finger millet (*Eleusine coracana* L. Gaertn.). *Mol. Breed.* 27, 315–328. doi: 10.1007/s11032-010-9433-1

Reddy, V. G., Upadhyaya, H. D., and Gowda, C. L. L. (2007). Morphological characterization of world's proso millet germplasm. *SATJ.* 3, 1–4.

Rengalakshmi, R. (2005). Folk biological classification of minor millet species in Kolli hills, India. *BioOne* 25, 59–70. doi: 10.2993/0278-0771(2005)25[59:FBCOMM]2.0.CO;2

Rose, D. J., and Santra, D. K. (2013). Proso millet (*Panicum miliaceum* L.) fermentation for fuel ethanol production. *Ind. Crops Prod.* 43, 602–605. doi: 10.1016/j.indcrop.2012.08.010

Sabir, P., Ashraf, M., and Akram, N. A. (2011). Accession variation for salt tolerance in proso millet (*Panicum miliaceum* L.) using leaf proline content and activities of some key antioxidant enzymes. *J. Agron. Crop Sci.* 197, 340–347. doi: 10.1111/j.1439-037X.2011.00471.x

Salimath, S. S., Oliveira, A. C. D., Godwin, I. D., and Bennetzen, J. L. (1995). Assessment of genome origins and genetic diversity in the genus *Eleusine* with DNA markers. *Genome* 38, 757–763. doi: 10.1139/g95-096

Schontz, D., and Rether, B. (1999). Genetic variability in foxtail millet, *Setaria italica* (L.) P. Beauv.: identification and classification of lines with RAPD markers. *Plant Breed.* 118, 190–192. doi: 10.1046/j.1439-0523.1999.118002190.x

Seghatoleslami, M. J., Kafi, M., and Majidi, E. (2008). Effect of drought stress at different growth stages on yield and water use efficiency of five proso millet (*Panicum miliaceum* L.) genotypes. *Pakistan J. Bot.* 40, 1427–1432.

Sentoku, N., Taniguchi, M., Sugiyama, T., Ishimaru, K., Ohsugi, R., Takaiwa, F., et al. (2000). Analysis of the transgenic tobacco plants expressing *Panicum miliaceum* aspartate aminotransferase genes. *Plant Cell Rep.* 19, 598–603. doi: 10.1007/s002990050779

Sharma, M., Kothari-Chajer, A., Jagga-Chugh, S., and Kothari, S. L. (2011). Factors influencing *Agrobacterium tumefaciens*-mediated genetic transformation of *Eleusine coracana* (L.) Gaertn. *Plant Cell Tissue Organ Cult.* 105, 93–104. doi: 10.1007/s11240-010-9846-x

Singh, U. M., Chandra, M., Shankhdhar, S. C., and Kumar, A. (2014). Transcriptome wide identification and validation of calcium sensor gene family in the developing spikes of finger millet genotypes for elucidating its role in grain calcium accumulation. *PLoS ONE* 9:e103963. doi: 10.1371/journal.pone.0103963

Sivakumar, S., Franco, O. L., Thayumanavan, B., Murad, A. M., Manickam, A., Mohan, M., et al. (2006a). Cloning and structural analysis of an Indian little millet (*Panicum sumatrense*) zein-like storage protein: Implications for molecular assembly. *Biochemistry* 71, 1183–1191. doi: 10.1134/S0006297906110034

Sivakumar, S., Mohan, M., Franco, O. L., and Thayumanavan, B. (2006b). Inhibition of insect pest α -amylases by little and finger millet inhibitors. *Pestic. Biochem. Physiol.* 85, 155–160. doi: 10.1016/j.pestbp.2005.11.008

Smith, P. M. (1977). “Minor crops,” in *Evolution of Crop Plants*, ed N. W. Simmonds (London; New York: Longman), 301–324.

Subrahmanyam, D., and Rathore, V. S. (1999). Variation in photosynthetic traits in barnyard millet (*Echinochloa frumentaceae*) genotypes. *J. Agron. Crop Sci.* 183, 199–203. doi: 10.1046/j.1439-037X.1999.00341.x

Subramanian, A., Nirmalakumari, A., and Veerabhadiran, P. (2010). Trait based selection of superior kodo millet (*Paspalum scrobiculatum* L.) genotypes. *Electron. J. Plant Breed.* 1, 852–855.

Sudhakar, C., Veeranagamallaiah, G., Nareshkumar, A., Sudhakarbabu, O., Sivakumar, M., Pandurangaiah, M., et al. (2015). Polyamine metabolism influences antioxidant defense mechanism in foxtail millet (*Setaria italica* L.) cultivars with different salinity tolerance. *Plant Cell Rep.* 34, 141–156. doi: 10.1007/s00299-014-1695-3

Taniguchi, M., Kobe, A., Kato, M., and Sugiyama, T. (1995). Aspartate aminotransferase isozymes in *Panicum miliaceum* L., an NAD-malic enzyme-type C4 plant: comparison of enzymatic properties, primary structures, and expression patterns. *Arch. Biochem. Biophys.* 318, 295–306. doi: 10.1006/abbi.1995.1233

Taniguchi, M., and Sugiyama, T. (1996). Isolation, characterization and expression of cDNA clones encoding a mitochondrial malate translocator from *Panicum miliaceum* L. *Plant Mol. Biol.* 30, 51–64. doi: 10.1007/BF00017802

Taniguchi, M., and Sugiyama, T. (1997). The expression of 2-oxoglutarate/malate translocator in the bundle-sheath mitochondria of *Panicum miliaceum*, a NAD-malic enzyme-type C4 plant, is regulated by light and development. *Plant Physiol.* 114, 285–293.

Tester, M., and Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science* 327, 818–822. doi: 10.1126/science.1183700

Tsehaye, Y., Berg, T., Tsegaye, B., and Tanto, T. (2006). Farmers' management of finger millet (*Eleusine coracana* L.) diversity in Tigray, Ethiopia and implications for on-farm conservation. *Biodivers. Conserv.* 15, 4289–4308. doi: 10.1007/s10531-005-3581-3

Ugare, R., Chimmad, B., Naik, R., Bharati, P., and Itagi, S. (2014). Glycemic index and significance of barnyard millet (*Echinochloa frumentacea*) in type II diabetics. *J. Food Sci. Technol.* 51, 392–395. doi: 10.1007/s13197-011-0516-8

Uma, S., Prasad, T. G., and Kumar, M. U. (1995). Genetic variability in recovery growth and synthesis of stress proteins in response to polyethylene glycol and salt stress in finger millet. *Ann. Bot.* 76, 43–49. doi: 10.1006/anbo.1995.1076

Upadhyaya, H. D., Dwivedi, S. L., Singh, S. K., Singh, S., Vetriventhan, M., and Sharma, S. (2014). Forming core collections in barnyard, kodo, and little millets using morphoagronomic descriptors. *Crop Sci.* 54, 1–10. doi: 10.2135/cropsci2014.03.0221

Upadhyaya, H. D., Pundir, R. P. S., Gowda, C. L. L., Gopal Reddy, V., and Singh, S. (2008). Establishing a core collection of foxtail millet to enhance the utilization of germplasm of an underutilized crop. *Plant Genet. Res.* 7, 177–184. doi: 10.1017/S1479262108178042

Upadhyaya, H. D., Ravishankar, C. R., Narasimhudu, Y., Sarma, N. D. R. K., Singh, S. K., Varshney, S. K., et al. (2011). Identification of trait-specific germplasm and developing a mini core collection for efficient use of foxtail millet genetic resources in crop improvement. *Field Crop. Res.* 124, 459–467. doi: 10.1016/j.fcr.2011.08.004

Usha, B., Krishna Veni, G., Muni Kumar, D., and Hemalatha, K. P. J. (2011). Partial characterization of α -amylase from germinating little millets (*Panicum sumatrense*). *J. Phytol.* 3, 1–8.

Vadivoo, A. S., Joseph, R., and Ganesan, N. M. (1998). Genetic variability and diversity for protein and calcium contents in finger millet (*Eleusine coracana* (L.) Gaertn) in relation to grain color. *Plant Foods Hum. Nutr.* 52, 353–364. doi: 10.1023/A:1008074002390

Van, K., Onoda, S., Kim, M. Y., Kim, K. D., and Lee, S. H. (2008). Allelic variation of the *Waxy* gene in foxtail millet (*Setaria italica* (L.) P. Beauv.) by single nucleotide polymorphisms. *Mol. Genet. Genomics* 279, 255–266. doi: 10.1007/s00438-007-0310-5

Veeranagamallaiah, G., Jyothsnakumari, G., Thippeswamy, M., Reddy, P. C. O., Surabhi, G.-K., Sriranganayakulu, G., et al. (2008). Proteomic analysis of salt stress responses in foxtail millet (*Setaria italica* L. cv. Prasad) seedlings. *Plant Sci.* 175, 631–641. doi: 10.1016/j.plantsci.2008.06.017

Wang, C., Jia, G., Zhi, H., Niu, Z., Chai, Y., Li, W., et al. (2012). Genetic diversity and population structure of Chinese foxtail millet [*Setaria italica* (L.) Beauv.] landraces. *G3* 2, 769–777. doi: 10.1534/g3.112.002907

Wang, J. C., Hu, J., Xu, H. M., and Zhang, S. (2007). A strategy on constructing core collections by least distance stepwise sampling. *Theor. Appl. Genet.* 115, 1–8. doi: 10.1007/s00122-007-0533-1

Wang, M., Pan, Y., Li, C., Liu, C., Zhao, Q., Ao, G.-M., et al. (2011). Culturing of immature inflorescences and Agrobacterium-mediated transformation of foxtail millet (*Setaria italica*). *Afr. J. Biotechnol.* 10, 16466–16479. doi: 10.5897/ajb10.2330

Wang, T., and Darmency, H. (1997). Inheritance of sethoxydim resistance in foxtail millet, *Setaria italica* (L.). *Euphytica* 94, 69–73. doi: 10.1023/A:1002989725995

Wang, T., Fleury, A., Ma, J., and Darmency, H. (1996). Genetic control of dinitroaniline resistance in foxtail millet (*Setaria italica*). *J. Hered.* 87, 423–426. doi: 10.1093/oxfordjournals.jhered.a023031

Wang, Z. M., Devos, K. M., Liu, C. J., Wang, R. Q., and Gale, M. D. (1998). Construction of RFLP-based maps of foxtail millet, *Setaria italica* (L.) P. Beauv. *Theor. Appl. Genet.* 96, 31–36. doi: 10.1007/s001220050705

Wanous, M. K. (1990). Origin, taxonomy and ploidy of the millets and minor cereals. *Plant Var. Seeds* 3, 99–112.

Wu, Y., Du, J., Wang, X., Fang, X., Shan, W., and Liang, Z. (2012). Computational prediction and experimental verification of miRNAs in *Panicum miliaceum* L. *Sci. China* 55, 807–817. doi: 10.1007/s11427-012-4367-y

Yabuno, T. (1962). Cytotaxonomic studies on the two cultivated species and the wild relatives in the genus *Echinochloa*. *Cytologia* 27, 296–305. doi: 10.1508/cytologia.27.296

Yabuno, T. (1987). Japanese barnyard millet (*Echinochloa utilis*, Poaceae) in Japan. *Econ. Bot.* 41, 484–493. doi: 10.1007/BF02908141

Yadav, C. B., Muthamilarasan, M., Pandey, G., Khan, Y., and Prasad, M. (2014a). Development of novel microRNA-based genetic markers in foxtail millet for genotyping applications in related grass species. *Mol. Breed.* 34, 2219–2224. doi: 10.1007/s11032-014-0137-9

Yadav, S., Gaur, V. S., Jaiswal, J. P., and Kumar, A. (2014b). Simple sequence repeat (SSR) analysis in relation to calcium transport and signaling genes reveals transferability among grasses and a conserved behavior within finger millet genotypes. *Plant Syst. Evol.* 300, 1561–1568. doi: 10.1007/s00606-014-0982-3

Yamaguchi, H., Utano, A. Y. A., Yasuda, K., Yano, A., and Soejima, A. (2005). A molecular phylogeny of wild and cultivated *Echinochloa* in East Asia inferred from non-coding region sequences of trn T-L-F. *Weed Biol. Manag.* 5, 210–218. doi: 10.1111/j.1445-6664.2005.00185.x

Yang, X., Wan, Z., Perry, L., Lu, H., Wang, Q., Zhao, C., et al. (2012). Early millet use in northern China. *PNAS* 109, 3726–3730. doi: 10.1073/pnas.1115430109

Yang, X., Yu, X.-Y., and Li, Y.-F. (2013). *De novo* assembly and characterization of the Barnyardgrass (*Echinochloa crus-galli*) transcriptome using next-generation pyrosequencing. *PLoS ONE* 8:e69168. doi: 10.1371/journal.pone.0069168

Zhang, G., Liu, X., Quan, Z., Cheng, S., Xu, X., Pan, S., et al. (2012). Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat. Biotechnol.* 30, 549–554. doi: 10.1038/nbt.2195

Zhang, L., Liu, R., and Niu, W. (2014). Phytochemical and antiproliferative activity of proso millet. *PLoS ONE* 9:e104058. doi: 10.1371/journal.pone.0104058

Zhu, X. L., Zhang, L., Chen, Q., Wan, J., and Yang, G. F. (2006). Interactions of aryl-oxypyhenoxypyropionic acids with sensitive and resistant acetyl-coenzyme A carboxylase by homology modeling and molecular dynamic simulations. *J. Chem. Inf. Model.* 46, 1819–1826. doi: 10.1021/ci0600307

Conflict of Interest Statement: 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.

Copyright © 2015 Goron and Raizada. 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) or licensor 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.



Comparative analysis of fruit aroma patterns in the domesticated wild strawberries “Profumata di Tortona” (*F. moschata*) and “Regina delle Valli” (*F. vesca*)

Alfredo S. Negri¹, Domenico Allegra², Laura Simoni², Fabio Rusconi², Chiara Tonelli³, Luca Espen^{1*} and Massimo Galbiati^{2,3*}

¹ Dipartimento di Scienze Agrarie e Ambientali - Produzione, Territorio, Agroenergia, Università degli Studi di Milano, Milan, Italy

² Plant Model System Platform, Fondazione Filarete, Milan, Italy

³ Department of Life Sciences, Università degli Studi di Milano, Milan, Italy

Edited by:

Aleksandra Skirycz, Instituto Tecnológico Vale Desenvolvimento Sustentável/Vale Institute of Technology Sustainable Development, Brazil

Reviewed by:

Camila Caldana, Brazilian Bioethanol Science and Technology Laboratory (CTBE) - Centro Nacional de Pesquisa em Energia e Materiais/Associação Brasileira de Tecnologia de Luz Sincrotron, Brazil
Detlef Ulrich, Julius Kühn-Institute, Germany

*Correspondence:

Luca Espen, Dipartimento di Scienze Agrarie e Ambientali - Produzione, Territorio, Agroenergia, Università degli Studi di Milano, Via Celoria 2, Milan 20133, Italy
e-mail: luca.espen@unimi.it;

Massimo Galbiati, Department of Life Sciences, Università degli Studi di Milano, Via Celoria 26, Milan 20133, Italy
e-mail: massimo.galbiati@unimi.it

Strawberry is one of the most valued fruit worldwide. Modern cultivated varieties (*Fragaria × ananassa*) exhibit large fruits, with intense color and prolonged shell life. Yet, these valuable traits were attained at the cost of the intensity and the variety of the aroma of the berry, two characteristics highly appreciated by consumers. Wild species display smaller fruits and reduced yield compared with cultivated varieties but they accumulate broader and augmented blends of volatile compounds. Because of the large diversity and strength of aromas occurring in natural and domesticated populations, plant breeders regard wild strawberries as important donors of novel scented molecules. Here we report a comprehensive metabolic map of the aroma of the wild strawberry Profumata di Tortona (PdT), an ancient clone of *F. moschata*, considered as one of the most fragrant strawberry types of all. Comparison with the more renowned woodland strawberry Regina delle Valli (RdV), an aromatic cultivar of *F. vesca*, revealed a significant enrichment in the total level of esters, alcohols and furanones and a reduction in the content of ketones in the aroma of PdT berries. Among esters, particularly relevant was the enhanced accumulation of methyl anthranilate, responsible for the intensive sweetish impression of wild strawberries. Interestingly, increased ester accumulation in PdT fruits correlated with enhanced expression of the *Strawberry Alcohol Acyltransferase* (SAAT) gene, a key regulator of flavor biogenesis in ripening berries. We also detected a remarkable 900-fold increase in the level of mesifurane, the furanone conferring the typical caramel notes to most wild species.

Keywords: strawberry, *Fragaria moschata*, *Fragaria vesca*, gas chromatography-mass spectrometry, aroma, volatiles

INTRODUCTION

Garden strawberries (*Fragaria × ananassa*) are among the most appreciated fruits and represent a valuable economic crop with a global annual production that exceeds 4.5 Mt (FAOSTAT, 2014). Berries of high yield modern varieties are characterized by large size, attractive color and prolonged shell life (Hancock, 1999). Yet, the sensory quality of traded strawberries is often criticized, as they lack flavor and fragrance. Sensory perceptions originate from the combination of sweetness, texture and aroma (Christensen, 1983). Among these features, aroma remains the most valued quality indicator for consumers worldwide (Azodanlou et al., 2003; Colquhon et al., 2012). Just as in other fruits, strawberry's aroma is a complex blend of volatile organic compounds (VOCs). These compounds only represent 0.001 to 0.01% of the berry fresh weight but have a major effect on its flavor and fragrance (Buttery, 1983). As many as 360 volatiles have been identified in ripe strawberries; these include esters, aldehydes, ketones, alcohols, terpenes and furanones (Menager et al., 2004; Jetti et al.,

2007). Individual compounds, although often present in minute quantities, may have a significant impact on the aroma. The reduced fragrance of most garden strawberries derives from the relatively limited accumulation of esters molecules, frequently combined with an excess of lactones, which usually cause a disproportionate peach note.

As opposite to garden varieties, wild strawberries are renowned for their intense flavor and fragrance. Most spontaneous species bear small fruits which accumulate higher levels and wider assortments of volatile molecules, compared with cultivated varieties (Honkanen and Hirvi, 1990). The ample natural variation occurring among the wild ancestors of garden strawberries provides a valuable source of novel volatile compounds for breeding new commercial strawberries with improved aroma properties (Ulrich and Hoberg, 2000a). Over 20 wild species are found within the *Fragaria* genus, of which the diploid woodland strawberry (*F. vesca*) is the most common (Rousseau-Gueutin et al., 2009). Among other species, musk strawberries

(*F. moschata*) are recognized for their distinguished and extraordinary strong aroma. Native to highland areas from France to Siberia, musk strawberries were widely cultivated in Europe to the mid-1900, when they were replaced by firmer, higher yielding and more remunerative *F. × ananassa* cultivars (Darrow, 1966).

Today, only few musk strawberries survive in farm plantings, although on a very small scale. Noteworthy is the Italian clone Profumata di Tortona (PdT), regarded as one of the most fragrant strawberry types of all (Urruty et al., 2002). PdT is a dioecious strawberry having the male and female reproductive organs in separate flowers on separate plants. Berries, distinguishable for the intense red color of the peel and the whitish flesh, possess a delightful sour-sweet, slightly astringent flavor, with green, caramel and clove-like notes (Pet'ka et al., 2012). Hallmark of PdT is its peculiar floral, spicy aroma, with hints of honey, musk and wine. The fragrance of this strawberry is so intense that a few ripe berries can perfume an entire room with a penetrating mango-like, tropical scent. Differently from most cultivated strawberries, the harvesting season for PdT is extremely limited. Berries are only available for a period of 10–15 days, coinciding with the second half of June. Currently, the commercial cultivation of PdT is restricted to the municipality of Tortona, in the Piedmont region in Northern Italy. Remarkably, the first historical evidence of musk strawberries in this area dates back to year 1411 (Bergaglio, 2007). Cultivation lingered into the early 1960s, when strawberry fields succumbed to urban development. Most recently, there has been a renewed interest for the Profumata di Tortona, considered a delicacy both for fresh consumption and gourmet preparations.

Diversity of volatile patterns in woodland strawberries in comparison to cultivated garden varieties has been extensively investigated (Drawert et al., 1973; Ulrich et al., 1997). Recent surveys of aroma profiles across 16 *F. vesca* accessions and five *F. × ananassa* cultivars, identified significant differences in the accumulation of individual esters, ketones and terpinoids between the two strawberries (Ulrich and Olbricht, 2013, 2014). In particular, small esters, including ethyl hexanoate, methyl butanoate and methyl hexanoate, were found in higher amounts in garden strawberries compared with woodland accessions. Conversely, the key ester methyl antranilate (MA) was more abundant in *F. vesca*. Similarly, ketones (e.g., 2-pentanone, 2-heptanone, and 2-nonenone) and terpinoids (e.g., myrtenal, myrtenil acetate, α -terpineol) occurred at higher levels in wild berries, with the exception of the monoterpene linalool, which was more abundant in garden strawberries (Ulrich and Olbricht, 2013, 2014).

Detailed profiling of the aroma composition in musk strawberries has only been reported for few spontaneous populations (Ulrich et al., 2007; Pet'ka et al., 2012). Urruty and colleagues performed a first partial assessment of the volatile compounds produced by ripe PdT berries (Urruty et al., 2002). These authors determined the abundance of 23 preselected VOCs in the aroma of two *F. moschata* clones (Capron Royal and PdT) compared with 15 garden varieties. Selected compounds, representing major constituents of the strawberry aroma, included esters (e.g. methyl hexanoate, MA), monoterpenes (e.g., linalool, nerolidol), ketones (e.g., 2-pentanone, 2-heptanone), aldehydes (e.g., 2-hexenal), lactones and furanones (e.g., γ -decalactone, mesifurane). Among

the volatiles analyzed, Capron Royal and PdT displayed lower levels of small esters as methyl hexanoate, compared with cultivated strawberries. In contrast, both *F. moschata* clones revealed exceptionally high levels of MA, which was barely detectable in most garden varieties (Urruty et al., 2002). Despite the relevance of these findings, a more comprehensive analysis of the aroma profile of PdT is required to fully uncover the volatile composition of ripe PdT berries and gain more insights into its extraordinary aromatic properties.

Here we report a re-assessment of the VOCs composition of PdT berries, based on a non-targeted Solid-Phase Micro-Extraction/Gas chromatography-Mass Spectrophotometry (SPME/GC-MS) approach (Ulrich and Hoberg, 2000b). We compared the aroma of PdT with that of the woodland strawberry Regina delle Valli (RdV). The latter represents a widely cultivated cultivar of *F. vesca*, renowned for its intense and pleasant aroma. Most importantly, the aroma composition of this strawberry has not been investigated in previous studies. In total, we identified 131 VOCs in the headspace of the two strawberries, which provide a comprehensive picture of the aroma patterns of PdT and RdV berries. As a whole, our results contribute to shed new light on the natural variation occurring in the aroma of wild strawberry species.

MATERIALS AND METHODS

PLANT MATERIAL

Plants of Profumata di Tortona were provided by “Consorzio per la valorizzazione e la tutela della Fragola Profumata di Tortona,” Tortona, Italy. Regina delle Valli plants were purchased from Azienda Agricola Ortomio, Forlì, Italy (<http://www.ortomio.it/>). Both strawberries were grown in the production area of Tortona (Italy) under commercial conditions, accordingly to the standards adopted by the “Consorzio per la valorizzazione e la tutela della Fragola Profumata di Tortona.” Fruits for the analysis were harvested with the assistance of local producers, to ensure selection of uniform, healthy and fully ripe berries. Five randomized samples, composed of 15 individual ripe fruits each, were collected the early morning in a single harvest. The extremely reduced harvesting season of PdT did not justify the adoption of multiple harvests. Berries employed in our analysis represent a faithful sample of the commercial fruits that are normally available to consumers.

SPME/GC-MS SAMPLE PREPARATION

Harvested fruits samples were immediately frozen at -20°C and stored at -80°C . Prior to the analysis fruits were powdered in liquid nitrogen and 1 g of fresh weight for each sample was incubated at 30°C for 5 min. Following addition of 300 μL of a NaCl saturated solution, 900 μL of the homogenized mixture were transferred to a 10 mL screw cap headspace vial. Three technical replicas were performed for each sample.

AUTOMATED SPME/GC-MS

Volatiles were sampled by SPME with a 2 cm \times 50/30-micron DVB/Carboxen/PDMS Stable Flex fiber (Sigma, Milano, Italy). Extraction and desorption of the volatiles were performed automatically by a CombiPAL autosampler (CTC Analytics,

Zwingen, Switzerland) as described (Zorrilla-Fontanesi et al., 2012). Chromatography was performed on a DB-5 ms (30 m × 0.25 mm × 1 mm) column (Sigma, Milano, Italy) with Helium at a constant flow of 1.2 mL/min, accordingly to Zorrilla-Fontanesi et al. (2012). Mass spectra were recorded in scan mode in the 35 to 220 mass-to-charge ratio range by a 5975B mass spectrometer (Agilent Technologies, Cernusco sul Naviglio, Italy) (ionization energy 70 eV; scanning speed 7 scans/s). The Enhanced ChemStation software (Agilent Technologies, Cernusco sul Naviglio, Italy) was used for recording and processing of chromatograms and spectra. Three technical replicas were conducted for each sample.

COMPOUND IDENTIFICATION AND RELATIVE QUANTIFICATION

Compound abundance was determined using the software MET-IDEA that directly extracts ion intensities exploiting a list of ions coupled with their relative retention time values (Broeckling et al., 2006). The ion list was built as following. For each biological replicate, a randomly chosen chromatogram was analyzed by Automated Mass Spectral Deconvolution and Identification System (AMDIS; <http://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata:amdis>) comparing the deconvoluted spectra with entries at the National Institute of Standards and Technology MS library (NIST08) (Type of Analysis, Simple; Resolution, Medium; Sensitivity, Medium; Shape Requirement, Low; Component Width, 12; Minimum Match Factor, 70). The 113 entries with a match score larger than 800 on the NIST Search software were extracted and added to a .msl file containing the 1537 plant derived metabolites of the VOC BinBase library (Skogerson et al., 2011) to assemble the library used for the analysis of the 10 randomly-chosen chromatograms. During this second cycle of AMDIS analysis, the Minimum match Factor was set to 90 and the resulting .fin files were joined in a single ion list, manually eliminating overlapping entries. The resulting ion list was employed for Met-IDEA analysis over all the 30 chromatograms. Peak areas of selected specific ions were integrated for each compound. The relative content (R.C.) of each tentatively identified metabolite (expressed as percentage) was calculated as the ratio between each peak area and the sum of all the peak areas present in the chromatogram, multiplied by 100 [R.C. = (Area_{peak} / \sum Area_{peak}) × 100]. CAS numbers and flavoring descriptors were retrieved from the web-based Chemical Search Engine (<http://www.chemindustry.com/apps/chemicals>) and from the online edition of the “Specifications for Flavorings” database (<http://www.fao.org/ag/agn/jecfa-flav/>), respectively.

STATISTICAL ANALYSIS

Differences in volatiles accumulation between the two cultivars were investigated through the Soft Independent Modeling of Class Analogies (SIMCA) provided by the software Unscrambler (Camo Process AS, Oslo, Norway), through the construction of a Principal Component Analysis (PCA) model for each cultivar. Samples were projected on the orthogonal system constituted by the two models to assess their object-to-model distance and to judge their membership to one of the two classes. The capability of variable k in discriminating between model PdT and RdV (fitting samples from model PdT onto model RdV) was described by the Discrimination Power (DiscrPower), computed as:

$$\text{DiscrPower} = \frac{S_{\text{PdT}}(\text{RdV}, k)^2 + S_{\text{RdV}}(\text{RdV}, k)^2}{S_{\text{RdV}}(\text{RdV}, k)^2 + S_{\text{PdT}}(\text{PdT}, k)^2}$$

The significance of the differences in volatile levels in the two strawberries was tested through a *t*-Student test considering as significant variations with $p < 0.01$.

QUANTITATIVE PCR (qPCR) ANALYSIS

RNA was isolated from Small Green, Turning and Red fruits according to Schultz et al. (1994). Reverse transcription, and qPCR analysis were performed as previously described (Galbiati et al., 2011). SAAT expression was analyzed using primers SAAT-F1 (5'-TTGGATGGGGAGGACATCAT-3') and SAAT -R1 (5'-CACCCACGCTTCAATTCCAGTA-3'). Gene expression was normalized using the ACTIN gene (GenBank: JN616288.1), amplified with primers ACT-F1 (5'-ATGTTGCCCTTGACTACGAACAA-3') and ACT-R1 (5'-TGGCCGTCGGAAAGCTCATA-3'). Primers efficiency was first assessed on both genomic DNA and cDNAs derived from PdT and RdV fruits, to avoid differences in amplification efficiency in the two genotypes. Changes in SAAT gene expression were calculated relative to ACTIN using the $\Delta\Delta Ct$ method (Livak and Schmittgen, 2001).

RESULTS

PdT AND RdV BERRIES DISPLAY DISTINCT VOCs PATTERNS

Fully ripe berries of *F. moschata*, clone Profumata di Tortona (PdT) and *F. vesca*, cv Regina delle Valli (RdV) were analyzed by SPME/GC-MS. In total, through the construction of a non-redundant ion list collecting information from the AMDIS analysis of 10 different chromatograms, 131 VOCs were tentatively identified in the headspace of the two strawberries. GC-MS data obtained from individual biological and technical replicas were analyzed using the Soft Independent Modeling of Class Analogies (SIMCA) based on the models relative to the two genotypes built with the Principal Component Analysis (PCA) (Svante and Michael, 1977). The SIMCA Cooman's plot showed full differentiation of PdT and RdV, based on their aroma components. All the samples grouped in the relative membership class, as determined by a significance level of 5% (Figure 1). The model assigned the highest discriminating power to the ketones heptan-2-one and nonan-2-one (Table 1). A third ketone molecule, undecan-2-one, also scored among the most significant volatiles for the discrimination between the two strawberries. Additional molecules with high discriminating power included several esters, such as hexyl butanoate, methyl benzoate, ethyl hexanoate, pinocarvyl acetate, the furanone γ -hexalactone, and the terpenes pinocarveol and myrtenyl acetate (Table 1).

COMPARATIVE ANALYSIS OF AROMA PATTERNS IN PdT AND RdV BERRIES

Among the 131 volatiles identified in the aroma of the two strawberries, 47 were classified as esters, which are long known to represent the most abundant VOCs found in ripe strawberries (Latrasse, 1991) (Table 2). Terpenoids included 20 mono- and 4 sesqui-terpenes (Table 3). Amongst the remaining aroma constituents we identified 5 alcohols, 9 aldehydes (Table 4), 6 ketones,

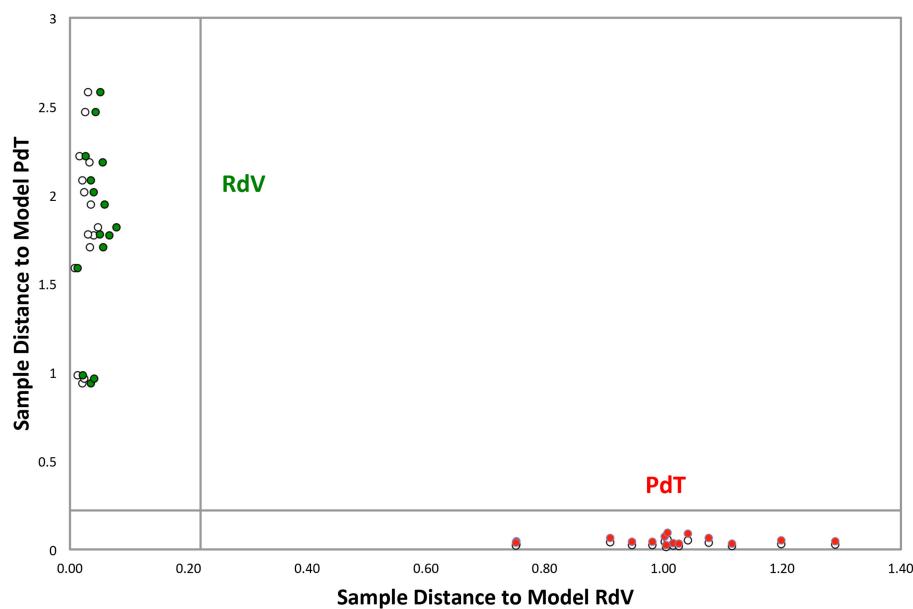


FIGURE 1 | Cooman's plot of the classification analysis between PdT and RdV. Green and red circles represent calibration samples belonging to RdV and PdT, respectively. White circles represent test set samples of the two

classes. Graph inner lines represent the significance level of 5%. The analysis employed 131 peak signals. Among these, 95 peaks were tentatively identified by MS library search.

Table 1 | Volatiles with the highest discriminating power identified by the SIMCA model.

Compound	Discriminating power	Compound	Discriminating power
heptan-2-one	317.183	[(Z)-hex-3-enyl] acetate	61.793
nonan-2-one	201.421	(1Z,4Z,7Z)-1,5,9,9-tetramethylcycloundeca-1,4,7-triene	60.002
hexyl butanoate	106.238	alpha-lonyl acetate	55.384
methyl benzoate	87.871	4-(2,6,6-trimethylcyclohex-2-en-1-yl)butan-2-yl acetate	55.077
2-methylbutanoic acid	81.037	methyl 2-methylbutanoate	54.337
ethyl hexanoate	75.867	methyl decanoate	54.171
pinocarvyl acetate	75.321	decyl acetate	51.343
octyl acetate	73.294	4-acetoxybutyl acetate	51.341
undecan-2-one	69.701	pinocarveol	49.644
γ -Hexalactone	65.968	myrtenyl acetate	48.208
butyl butanoate	65.954	4-acetoxy pentan-2-yl acetate	48.201
methyl tiglate	65.668	hexyl acetate	47.006

and 4 lactones (Table 5). Additional compounds comprised a single fatty acid (hexanoic acid), a single alkane hydrocarbon (tetradecane), and 36 volatiles of unknown chemical identity.

The relative abundance of individual chemical classes significantly differed between the two strawberries (Supplementary Figure 1). Quantitatively, esters were the predominant volatiles in berries of PdT, accounting for nearly 50% of the aroma. In comparison, their relative amount was drastically reduced in the aroma of RdV, only representing 25% of the total volatiles. Terpenes were the second most abundant class of compounds found in PdT (24%). A comparable level of total terpenes was identified in RdV (21%). Conversely, the two strawberries disclosed a marked difference in the relative level of ketones, which were severely reduced in PdT (4%), while highly abundant in

RdV (27%). We also observed a disparity in the relative abundance of alcohols, which were far more copious in PdT compared with RdV, representing 7 and 2% of the total volatile molecules, respectively. No substantial differences were detected in the relative amount of aldehydes and lactones. Aldehydes accounted for 6 and 9% of total volatiles in PdT and RdV, respectively, whilst lactones were the least represented molecules, only covering 1.4 and 0.4% of the aroma of PdT and RdV, respectively. Finally, compounds of uncertain identity were slightly more abundant in RdV (14%) compared with PdT (8%) (Supplementary Figure 1).

Analysis of individual constituents of the aroma identified the monoterpene myrtenyl acetate as the most abundant volatile in both PdT and RdV (Table 3). This finding is in line with a

Table 2 | Ester molecules identified in the headspace of PdT and RdV berries by GC-MS.

IUPAC name	Ester compound	Synonym	Quantification Ion	Retention time	CAS	Descriptor	Relative Content (%)		Fold change	p-value
							PdT	RdV		
octyl acetate	43.1	19.8167	112-14-1	Fruity, orange-like, jasmine-like odor	12.687	4.644	2.73	2.152E-07		
4-acetoxybutyl acetate	43.1	19.878	628-67-1	Ethereal, fruity, leafy, green odor	11.564	3.645	3.17	1.374E-06		
hexyl formate	56.1	9.3341	629-33-4	Grape-like or orange aroma	4.555	1.292	3.53	1.017E-04		
methyl 2-aminobenzoate	119	23.4659	134-20-3		3.395	0.373	9.10	1.983E-05		
hexyl acetate	43.1	13.9064	88230-35-7	Fruity odor	2.717	3.766	0.72	7.554E-02		
2-methylbutanoic acid	74	9.76	600-07-7		1.805	0.015	120.33	1.221E-03		
(7,7-dimethyl-4-methylidene-3-bicyclo[3.1.1]heptanyl) acetate	92.1	17.8485	33045-02-2		1.348	0.591	2.28	6.287E-08		
[(Z)-hex-2-enyl] acetate	67.1	13.9831	56922-75-9	Powerful green note	1.087	1.018	1.07	3.949E-01		
[(Z)-hex-3-enyl] acetate	67.1	13.7052	3681-71-8	Fruity, green aroma	0.968	0.414	2.34	1.535E-03		
(E)-hex-2-enyl butanoate	43.1	19.3505	53398-83-7	Sweet, fruity, apple-like odor	0.752	0.784	0.96	4.126E-01		
methyl 2-methylbutanoate	85.1	6.5488	868-57-5		0.751	0.005	150.20	1.577E-08		
phenyl/methyl acetate	108	18.485	140-11-4	Sweet, fruity, floral (Jasmine) odor	0.677	2.765	0.24	1.914E-04		
methyl octanoate	74	17.3418	67762-39-4	Winey, fruity, orange odor	0.549	0.330	1.66	4.148E-02		
decyl acetate	55.1	25.0756	112-17-4	Pear, floral, orange-rose odor	0.337	0.472	0.71	4.107E-03		
tridecan-2-yl acetate	55.1	30.2373	not available		0.300	0.022	13.64	2.966E-09		
heptan-2-yl acetate	43.1	14.8247	5921-82-4		0.299	0.238	1.26	2.395E-01		
methyl benzoate	105	16.5338	93-58-3	Pungent, heavy, floral odor with fruity undertones	0.247	0.174	1.42	7.030E-02		
octyl 3-methylbutanoate	103.1	25.7912	7786-58-5	Apple-pineapple odor	0.235	0.015	15.67	8.846E-07		
methyl (E)-2-methylbut-2-enoate	114	9.2297	6622-76-0		0.215	0.017	12.65	4.559E-07		
[(E)-hex-2-enyl] propanoate	57.1	16.8525	53398-80-4	Fruity, ripe apple-pear aroma	0.159	0.073	2.18	4.585E-03		
(2,4-ditert-butylphenyl) 5-hydroxypentanoate	191.1	27.6179	not available		0.135	0.218	0.62	4.427E-02		
3-hydroxy-2,2,4-trimethylpentyl 2-methylpropanoate	89	24.2546	80525-37-7		0.123	0.167	0.74	8.474E-02		
hexyl butanoate	89.1	19.2782	2639-63-6	Fruity, apricot odor	0.111	0.444	0.25	1.173E-03		
butyl butanoate	71	13.3589	109-21-7	Fruity, pineapple-like odor	0.089	0.853	0.10	3.059E-04		
methyl decanoate	74	22.9111	110-42-9		0.073	0.181	0.40	3.490E-04		
[2,2,4-trimethyl-1-(2-methylpropanoyloxy) pentan-3-yl] 2-methylpropanoate	71.1	29.6078	6846-50-0		0.072	0.085	0.85	2.162E-01		

(Continued)

Table 2 | Continued

IUPAC name	Ester compound	Synonym	Quantification Ion	Retention time	CAS	Descriptor	Relative Content (%)		Fold change	p-value
							PdT	Rw		
2-phenylethyl acetate		104	21.1851	103-45-7	Very sweet, rosy-fruity honey-like odor	0.071	0.037	1.92	1.962E-05	
methyl dodecanoate		74	27.8978	111-82-0	Fatty, floral, winey odor	0.054	0.008	6.75	5.111E-04	
methyl (E)-oct-2-enoate		55.1	18.6982	7367-81-9	Fruity, green aroma	0.052	0.019	2.74	7.342E-04	
(Z)-hex-3-enyl butanoate		67.1	19.1306	16491-36-4	Green, fruity, buttery odor	0.048	0.032	1.50	8.847E-03	
[(E)-3-phenylprop-2-enyl] acetate		161	22.5682	103-54-8	sweet, balsamic, floral odor	0.048	0.002	24.00	1.191E-06	
ethyl hexanoate		117.1	24.4915	8068-81-3	Wine-like odor	0.047	0.953	0.05	4.466E-02	
(1-hydroxy-2,4,4-trimethylpentan-3-yl) 2-methylpropanoate		71.1	23.7336	74367-33-2		0.045	0.075	0.60	4.326E-02	
octyl hexanoate		99.1	29.2171	4887-30-3		0.038	0.016	2.38	1.072E-04	
hexyl hexanoate		117.1	24.4915	6379-65-0	Herbaceous odor	0.034	0.029	1.17	1.677E-01	
4-(2,6,6-trimethylcyclohex-2-en-1-yl)but-3-en-2-yl acetate		180.1	27.2825	94224-42-7		0.030	0.007	4.29	1.685E-06	
ethyl octanoate		88	19.4368	106-32-1	Wine, brandy, fruity floral odor	0.029	0.776	0.04	4.239E-02	
4-acetyl/oxypentan-2-yl acetate		103	18.1022	7371-86-0		0.029	0.003	9.67	1.321E-06	
(4-propan-2-ylphenyl)methyl acetate		107	25.4881	59230-57-8		0.027	0.007	3.86	1.471E-15	
ethyl decanoate		88	24.7276	110-38-3	Oily brandy-like odor	0.023	0.590	0.04	3.889E-02	
nonyl acetate		97.1	22.5408	143-13-5	Floral, fruity odor	0.021	0.008	2.63	2.589E-11	
7,7-dimethyl-2-bicyclo[3.1.1]heptanylmethyl acetate		80	24.5669	29021-36-1		0.017	0.017	1.00	3.528E-01	
dodecyl acetate		70.1	29.8252	70808-58-1	Waxy, citrus-rose odor	0.015	0.002	7.50	2.659E-09	
methyl (E)-3-phenylprop-2-enoate [Methyl (E)-cinnamate]		161	22.5682	1754-62-7	Fruity balsamic odor	0.015	0.003	5.00	8.742E-07	
4-(2,6,6-trimethylcyclohex-2-en-1-yl)butan-2-yl acetate		136.1	29.0433	not available		0.014	0.001	14.00	4.276E-10	
1-O-cyclobutyl 2-O-(2-methylpropyl) benzene-1,2-dicarboxylate		149	35.347	not available		0.013	0.025	0.52	8.367E-02	
nonan-2-yl butanoate		126.1	25.0083	not available		0.001	0.032	0.03	3.438E-04	
Sub total						45.921	25.243	1.82		

Table 3 | Terpene molecules identified in the headspace of PdT and RdV berries by GC-MS.

IUPAC name	Terpene compound	Synonym	Quantification Ion	Retention time	CAS	Descriptor	Relative Content (%)		fold change	<i>p</i> -value
							PdT	RdV		
7,7-dimethyl-4-bicyclo[3.1.1]hept-3-enyl)methyl acetate	myrtenyl acetate	91.1	23.0091	36203-31-3	Pleasant, refreshing, sweet-herbaceous	20.819	15.394	1.35	2.068E-03	
7,7-dimethylbicyclo[3.1.1]hept-3-ene-4-carbaldehyde	myrtenal	107	19.4862	57526-63-3	Refreshing, spicy-herbaceous odor	0.765	0.503	1.52	4.039E-06	
4-prop-1-en-2-ylcyclohexen-1-yl)methyl acetate	perillyl acetate	91.1	25.7741	5111-96-3	Warm-herbaceous spicy odor	0.464	0.246	1.89	1.492E-07	
4-methyl-1-propan-2-ylcyclohex-3-en-1-ol	4-carvomenthol	71.1	18.9548	562-74-3	Warm-peppery, mildly earthy, musty-woody odor	0.357	0.507	0.70	1.156E-03	
(5-methylidene-6-prop-1-en-2-yl)cyclohex-3-en-1-yl) acetate		117	32.0095	not available		0.223	0.003	74.33	2.140E-10	
1-methyl-2-propan-2-ylbenzene	o-cymene	119.1	14.3651	95660-61-0	Citrusy aroma reminiscent of lemon	0.166	0.502	0.33	1.280E-04	
(4-methyl-1-propan-2-yl-4-bicyclo[3.1.0]hexanyl) acetate		59.1	19.3391	not available		0.150	0.492	0.30	2.085E-05	
4,7,7-trimethylbicyclo[3.1.1]hept-3-ene	α -pinene	94.1	11.4057	80-56-8	Warm, resinous, pine-like aroma	0.114	0.534	0.21	6.476E-03	
1,1,3a-trimethyl-7-methyldiene-1a,2,3,4,5,6,7a,7b-octahydrocyclopropo[al]naphthalene	1H-Cyclopropalnaphthalene	204.2	28.3446	not available		0.108	0.001	108.00	3.190E-03	
1-methyl-4-prop-1-en-2-ylcyclohexene	limonene	93.1	14.4629	95327-98-3	fresh, light, sweet, citrusy aroma	0.094	0.180	0.52	2.519E-04	
3,7-dimethyl-1,6-dien-3-ol	linalool	71.1	16.5999	8008-26-2	pleasant, floral odor	0.063	1.182	0.05	1.523E-06	
4-methylidene-1-propan-2-ylcyclohexane	sabinene	93.1	12.7934	3337-41-5	terpineol odor	0.042	0.135	0.31	6.104E-03	
4-methyl-1-propan-2-ylbicyclo[3.1.0]hex-3-ene	α -thuiene	92	11.4178	75715-79-6		0.041	0.194	0.21	5.281E-03	
4E,6E-2,6-dimethylocta-2,4,6-trien-3,7,7-trimethylbicyclo[4.1.0]hept-4-ene	alloocimene	121.1	16.3138	673-84-7	Warm herbaceous aroma	0.030	0.149	0.20	1.899E-04	
(1S,2S,5R)-2,7,7-trimethylbicyclo[3.1.1]heptan-3-one	4-carene	136.1	16.3028	not available	fruity aroma	0.023	0.090	0.26	1.892E-04	
4-(1E)-buta-1,3-dienyl-3,5,5-trimethylcyclohexene	isopinocamphone	95.1	18.8561	15358-88-0	cedar camphor aroma	0.023	0.028	0.82	3.825E-02	
megastigma-3,7(E)-9-triene		105	20.0655	not available		0.018	0.001	18.00	6.869E-06	
pinocarveol		92.1	17.8485	5947-36-4	Warm, woody-balsamic aroma	0.016	0.090	0.18	2.921E-09	

(Continued)

Table 3 | Continued

IUPAC name	Terpene compound	Synonym	Quantification	Retention time	CAS	Descriptor	Relative Content (%)		fold change	p-value
							PdT	RdV		
1-methyl-4-propan-2-yl/cyclohexa-1,4-diene	γ -terpinene		93.1	14.4665	99-85-4	Refreshing, herbaceous-citrusy aroma	0.015	0.085	0.18	2.741E-09
3,7-dimethyloct-6-en-1-ol 4-methyl-2,3,4,5,6,7-hexahydro-1H-indene	α -citronelol		69.1 136.1	20.313 23.228	106-22-9 not available	rose-like aroma	0.006 0.049	0.481 0.097	0.01 0.51	1.305E-08 1.557E-06
4,7-dimethyl-1-propan-2-yl-1,2,4a,5,6,8a-hexahydronaphthalene (1Z,4Z,7Z)-1,5,9,9-tetramethylcycloundeca-1,4,7-triene	zizanene		204.1	27.4861	483-75-0		0.012	0.004	3.00	1.294E-03
1-ethenyl-1-methyl-2,4-diprop-1-en-2-yl/cyclohexane			80.1	26.4092	not available		0.002	0.016	0.13	9.112E-08
			176.1	24.8102	515-13-9		0.000	0.001	0.02	4.165E-04
							23.601	20.915	1.13	
		Sub total								

previous work, which demonstrated that this molecule dominates the terpenoid profile of wild strawberries (Aharoni et al., 2004). Interestingly, the level of myrtenyl acetate was significantly augmented in PdT (20.1%) compared with RdV (15.4%). The relative quantities of two esters molecules were also preeminent in the aroma of PdT, namely octyl acetate (12.7%) and 4-acetoxybutyl acetate (11.6%) (Table 2). Together with myrtenyl acetate these two compounds constituted 45% of the total volatiles found in this strawberry. Both esters were also detected in the aroma of RdV, even though their relative abundance was drastically reduced compared with PdT (4.6 and 3.6%, respectively). Two ketones, 2-heptanone, and 2-nonenone, were the most abundant molecules, after myrtenyl acetate, in the aroma of RdV, accounting for 12.7 and 11.9% of the total volatiles, respectively (Table 5). Conversely, the amount of these two molecules was significantly reduced in PdT (0.7 and 0.2%, respectively).

Additional abundant components of the aroma of PdT included, 1-hexanol (5%) (Table 4) and several other esters, as hexyl formate (4.5%), methyl anthranilate (MA) (3.4%) and hexyl acetate (2.7%) (Table 2). On average, the content of these volatiles was significantly higher in the aroma of PdT compared with RdV. The only aldehyde found in relatively high amounts in both strawberries was 2-hexenal (Table 4). Its content was significantly greater in RdV (5.4%) compared with PdT (2.7%). The most abundant ketone in the aroma of PdT was 2-tridecanone, reaching 2.3% of the total volatiles (Table 5). As opposite to the other ketones, the level of this molecule was significantly reduced in RdV (0.6%) compared with PdT.

Among less abundant molecules (<1% of total volatiles), major differences between the two strawberries were observed within terpenes, esters, and furanones. In particular, the levels of α -pinene, a monterpene specifically identified in the aroma of *F. vesca* (Aharoni et al., 2004), linalool, known to dominate the terpenoid profile of cultivated strawberry (Aharoni et al., 2004) and α -citronellol, were significantly reduced in the aroma of PdT in comparison with RdV (Table 3). Contrariwise, the level of megastigma-3,7(E),9-triene, the major terpene found in the essential oil of some Eucalyptus species (El-Mageed et al., 2011), was 18 times higher in PdT compared with RdV. We also observed a 74- and 108-fold increase in the accumulation of the terpenes 3-cyclohexen-1-ol,5-methylene-6-(1-methylethyl)-(9CI) and 1H-Cyclopropa[a]naphthalene, in the aroma of RdV compared with PdT (Table 3). Among minor esters, the most striking differences were detected for 2-methylbutanoic acid and methyl 2-methylbutanoate, whose content was 120- and 150-times more abundant in berries from PdT relatively to RdV, respectively (Table 2). Additional esters over-represented in the aroma of RdV included [(E)-3-phenylprop-2-enyl] acetate (24-fold), octyl 3-methylbutanoate (16-fold), tridecan-2-yl acetate (14-fold) and methyl tiglate (13-fold) (Table 2).

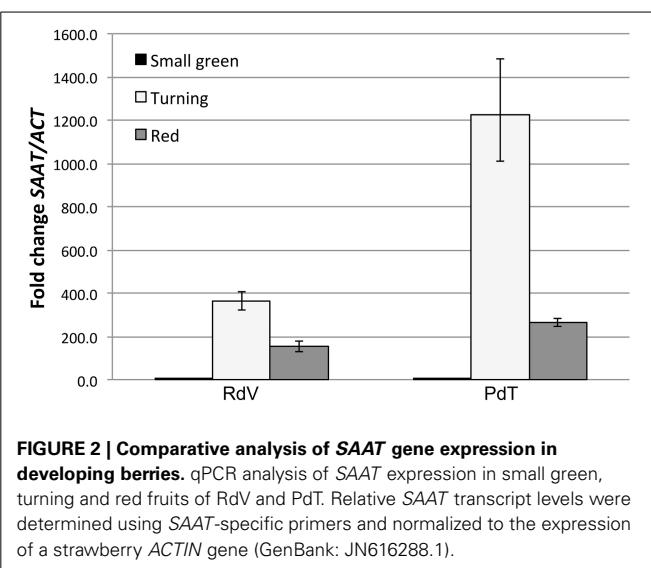
Even if present in lower amount compared to other volatiles, lactones and furanones, were more copious in the aroma of PdT compared to RdV. Remarkably, mesifurane, the typical furanone conferring sweet caramel notes to wild strawberries, was nearly 900 times more abundant in PdT compared with RdV (Table 5).

Table 4 | Alcohol and aldehyde molecules identified in the headspace of PdT and RdV berries by GC-MS.

IUPAC name	Compound	Quantification ion	Retention time	CAS	Descriptor	Relative Content (%)		fold change	p-value
						PdT	RdV		
ALCOHOLS									
hexan-1-ol	1-hexanol	56.1	9.3234	25917-35-5	Ethereal, fruity, alcoholic, sweet green aroma	4.990	1.386	3.60	1.084E-04
cyclohexanol		82.1	9.2826	108-93-0	Menthol-like camphorous odor	0.696	0.272	2.56	9.449E-04
octan-1-ol		56.1	15.7268	72-69-5	Sharp fatty-citrus odor	0.661	0.242	2.73	2.226E-04
tridecan-2-ol		45.1	27.3706	67989-40-6		0.163	0.035	4.66	1.890E-05
decan-1-ol		55.1	21.523	85566-12-7	Floral, waxy, fruity odor	0.008	0.084	0.10	5.441E-05
Sub-total						6.518	2.019	3.23	
ALDEHYDES									
(E)-hex-2-enal	2-hexenal	98.1	8.7726	73543-95-0	Strong fruity, green, vegetable-like aroma	2.672	5.426	0.49	1.123E-06
(2E,4E)-hexa-2,4-dienal	2,4-hexadienal	81	10.7482	80466-34-8	Sweet-green aroma	0.741	1.681	0.44	5.828E-08
hexanal		57.1	7.5617	9012-63-9	Fatty-green, grassy odor	0.699	1.039	0.67	1.826E-03
(E)-oct-2-enal	trans-2-octen-1-al	83	15.3431	2548-87-0	Fatty, green aroma	0.676	0.042	16.10	1.114E-06
nonanal		57.1	16.7129	75718-12-6	Fruity odor	0.673	0.444	1.52	4.207E-03
2,2-dimethylhexanal		57	12.9171	not available		0.149	0.178	0.84	1.226E-01
2-phenylacetalddehyde	benzenecacetaldehyde	91	14.984	122-78-1	Very powerful and penetrating pungent green floral and sweet odor of hyacinth type	0.069	0.062	1.11	2.931E-01
(E)-non-2-ena	2-nonenal, (E)-	43.1	18.4043	30551-15-6	Powerful, penetrating fatty, violet aroma	0.039	0.191	0.20	4.341E-03
decanal		70.1	19.7005	75718-12-6	Fatty, floral-orange odor	0.023	0.021	1.10	2.257E-01
Sub total						5.741	9.084	0.63	

Table 5 | Ketone and lactone molecules identified in the headspace of PdT and RdV berries by GC-MS.

IUPAC name	Compound/d	Quantification	Retention time	CAS	Descriptor	Relative Content (%)		fold change	p-value
						RdT	RdV		
KETONES									
tridecan-2-one	2-tridecanone	58.1	27.1904	593-08-8	Milky, herbaceous, slightly spicy odor	2.277	0.644	3.54	4.828E-09
heptan-2-one	2-heptanone	58.1	9.9628	29308-56-3	Fruity, spicy odor	0.662	12.694	0.05	7.379E-07
undecan-2-one	2-undecanone	58.1	22.0932	112-12-9	Citrus, fatty, rue-like odor	0.510	1.182	0.43	4.140E-04
nonan-2-one	2-nonanone	43.1	16.3408	821-55-6	Fruity, floral, fatty, herbaceous odor	0.240	11.953	0.02	8.818E-08
pentadecan-2-one	2-pentadecanone	71.1	31.7836	2345-28-0		0.036	0.043	0.84	1.232E-01
1-phenylethanalone	acetophenone	105	15.7516	98-86-2		0.031	0.099	0.31	7.649E-05
Sub total						3.756	26.615	0.14	
LACTONES AND FURANONES									
4-methoxy-2,5-dimethylfuran-3-one	mesifuran (DMF)	142	15.3621	4077-47-8	Sweet, carmellic, burnt sugar, aroma	0.869	0.001	869.00	1.145E-06
5-hexyloxolan-2-one	γ -decalactone	85	26.5909	706-14-9	Fruity, peach-like odor	0.391	0.376	1.04	4.634E-01
5-pentyloxolan-2-one	γ -nonalactone	85	21.22	82373-92-0	Coconut-like odor	0.155	0.032	4.84	1.750E-07
5-ethylloxolan-2-one	γ -hexalactone	85	15.231	695-06-7	Herbaceous, sweet odor	0.030	0.022	1.36	1.957E-01
Sub total						1.445	0.431	3.35	

**FIGURE 2 | Comparative analysis of SAAT gene expression in developing berries.** qPCR analysis of SAAT expression in small green, turning and red fruits of RdV and PdT. Relative SAAT transcript levels were determined using SAAT-specific primers and normalized to the expression of a strawberry ACTIN gene (GenBank: JN616288.1).

ANALYSIS OF SAAT EXPRESSION IN DEVELOPING BERRIES

In strawberries only very few genes have been directly associated with aroma biogenesis in ripening fruits. Among them, the *Strawberry Alcohol Acyltransferase* gene (SAAT), controlling a key step in esters biosynthesis (Aharoni et al., 2000), is of particular interest. The enzyme AAT catalyzes the transfer of an acyl moiety from acyl-CoA onto specific alcohols, resulting in the production of ester molecules (Harada et al., 1985). Intriguingly, octyl-acetate, the most abundant ester in both PdT and RdV aroma, has been demonstrated to be a genuine AAT product (Aharoni et al., 2000). The activation of SAAT expression during berry development has been positively correlated with the on-set of esters accumulation. In *F. × ananassa* SAAT expression is induced at early stage during fruit ripening, it peaks in correspondence of the turning stage, and it is rapidly down-regulated in red fruits (Aharoni et al., 2000). We compared the expression profile of the SAAT gene in developing berries from PdT and RdV (Supplementary Figure 2) to unravel potential differences in the level of gene expression and/or in the kinetic of SAAT activation between the two strawberries.

As shown in Figure 2, both RdV and PdT accumulated comparable low levels of SAAT transcripts in small green fruits. At the turning stage, both strawberries displayed very strong activation of SAAT expression. Interestingly, at this stage, the degree of gene activation was significantly enhanced in PdT compared with RdV (*t*-test, $p < 0.01$). This finding is conceivable with the increased accumulation of esters observed in PdT relatively to RdV fruits. Finally, in red fruits SAAT expression was down-regulated to the same extent in both berries (Figure 2).

DISCUSSION

The vast array of wild species found in the genus *Fragaria* provide an exceptionally large and conveniently located germplasm, which can serve to breeding novel quality traits into garden strawberries (Hancock and Luby, 1993). Key to the successful exploitation of the natural variation occurring in the *Fragaria* genus is the detailed characterization of the aroma profile of individual

species, ecotypes and clones. This study unraveled the volatile composition of two domesticated wild strawberries: *F. moschata* clone Profumata di Tortona and *F. vesca* cv. Regina delle Valli. Both strawberries are regarded as highly aromatic, although the scent of PdT is far more intense and persistent (Urruty et al., 2002).

Sampling procedures are key to the analysis of aroma composition in strawberries, as significant changes in VOCs profiles can occur among harvest dates within a single season (Schwieterman et al., 2014). Different strategies employing multiple harvests or the laborious harvest of all the available fruits throughout the season have been proposed to overcome this obstacle (Ulrich and Olbricht, 2013; Schwieterman et al., 2014). Our results rely on the analysis of biological replicas from a single harvest. In contrast to perpetual flowering accession characterized by prolonged harvesting seasons, the seasonal flowering PdT clone, only bears fruit for less than 2 weeks. Environmental changes and differences in plant physiology, known to affect fruit quality over months (Schwieterman et al., 2014), are unlike to influence the aroma composition of PdT berries within a period of 15 days.

We identified 131 volatiles in ripe berries of PdT and RdV, a number exceeding the aroma compounds usually found in commercial strawberries. A recent survey of the chemical diversity of the aromas of 35 different garden varieties recognized no more than 80 VOCs even in the most fragrant commercial genotypes (Schwieterman et al., 2014). Comparative analysis of aroma patterns revealed the significant enrichment in esters and alcohols along with the severe reduction in ketones in berries from PdT compared with RdV. Conversely, the two strawberries disclosed comparable levels of terpenes, aldehydes and lactones (Supplementary Figure 1). Over 130 different ester molecules have been identified in strawberries (Latrasse, 1991). In *F. × ananassa*, the chemical composition of ester volatiles is usually dominated by methyl and ethyl esters even though the abundance of each form varies with cultivar (Forney et al., 2000). Methyl esters were the prevalent form in the aroma of both PdT and RdV (22 molecules), whereas ethyl esters were poorly represented (3 molecules) (Table 2).

Even if it is generally difficult to establish a direct correlation between individual aroma constituents and specific sensory impressions, the fragrance of wild strawberries largely depend upon the relatively high amounts of methyl anthranilate (MA) (Ulrich et al., 1997). The intensive sweetish-flowery impression of this ester is at the basis of the definition of strawberry aroma chemo-types, which are essentially subdivided into MA-containing and MA-free types (Ulrich et al., 1997). Our analysis uncovered a nine-fold increase in the level of MA in berries from PdT compared with RdV, thus emphasizing the role of this key ester in determining the unique fragrance of musk strawberry. This conclusion is corroborated by previous studies, which reported an exceedingly higher concentration of MA in *F. moschata* relatively to *F. vesca* (Urruty et al., 2002). It is also interesting to note that, MA and γ -decalactone can directly inhibit the *in vitro* growth of relevant strawberry pathogens, thus implying that these volatiles may contribute to the healthiness of the berry at harvest (Chambers et al., 2013).

We also detected significant differences in the accumulation of low abundant esters between the two strawberries, as for instance methyl 2-methylbutanoate, which was 150 times more abundant in PdT compared with RdV (Table 2). Along with other esters of butanoic acid this molecule, conferring a sweet and fruity impact to the aroma, is found in higher amounts in garden strawberries compared with woodland accessions (Ulrich and Olbricht, 2013). We also detected a 16-fold increase in the relative level of octyl 3-methylbutanoate, in the headspace of PdT compared with RdV. This molecule, conferring an apple-pineapple odor, is present in the most flavored commercial varieties but undetected in the least flavorful. Evidence indicates that, octyl 3-methylbutanoate is an important component of strawberry aroma, potentially enhancing the perceived sweetness intensity, independently of individual sugars (Schwieterman et al., 2014).

Ester accumulation in ripening strawberries is directly associated with the expression of the *SAAT* gene, encoding a fruit-specific ALCOHOL ACYLTRANSFERASE (AAT) (Aharoni et al., 2000). It is intriguingly to speculate that the enhanced accumulation of ester molecules in the aroma of PdT, results from the hyper activation of *SAAT* expression observed in turning fruits from PdT, as compared with RdV (Figure 2). Further support to this hypothesis comes from the observation that genuine products of the *SAAT* enzyme, including octyl acetate, [(Z)-hex-3-enyl] acetate, 2-phenylethyl acetate and nonyl acetate (Aharoni et al., 2000), were found at higher levels in the headspace of PdT compared with RdV (Table 2).

The implication for the increased alcohol accumulation in PdT, as compared with RdV, on the quality of the berry is questionable, as these molecules have normally little impact on the aroma (Larsen and Watkins, 1995). Nevertheless, Schwieterman and colleagues recently reported a direct effect of the level of 1-hexanol on sweetness and flavor intensity in different garden cultivars (Schwieterman et al., 2014). Notably, we found a significant 4-fold increase in the relative amount of 1-hexanol in PdT compared with RdV, possibly suggesting a positive role for this molecule in determining the unique flavor of musk strawberries.

PdT and RdV displayed a similar terpene profile, largely dominated by myrtenyl acetate, by far the most abundant molecule found in the aroma of both strawberries. Higher concentrations of myrtenyl acetate are normally found in woodland strawberries compared with garden varieties (Ulrich and Olbricht, 2013). We observed a moderate, although significant, increased level of myrtenyl acetate in PdT compared with RdV (Table 3). This is in accord with previous comparative analysis of other *F. moschata* clones with *F. vesca* accessions (Ulrich et al., 1997). Major differences in the accumulation of low abundant terpenes involved linalool and 1H-Cyclopropa[a]naphthalene. The former, conferring pleasant flowery, citrus-like notes, represents the predominant monoterpene found in cultivated strawberries (Aharoni et al., 2004; Ulrich and Olbricht, 2014). Its relative concentration was 19 fold higher in RdV compared with PdT (Table 3). Conversely, 1H-Cyclopropa[a]naphthalene was 108 times more abundant in PdT in comparison with RdV. Interestingly, this molecule is among the major constituent of some agarwood oils, highly appreciated for their unique and intense fragrance and for their therapeutic properties (Takemoto et al., 2008).

To our knowledge, this volatile compound has never been reported in previous analyses of strawberry aromas and could represent a novel target to enhancing the fragrance of traded strawberries.

Even if present in relative low amounts, furanones are considered as dominating components of strawberry aroma. In particular furaneol (DHF) and its methyl ether mesifurane (DMF), contribute to the typical caramel-like, sweet, floral and fruity aroma of the berry (Jetti et al., 2007). Interestingly, PdT strawberries revealed an enhanced accumulation of total furanones compared with RdT. In particular, we detected a nearly 900-fold increase in the relative amount of DMF compared with RdV (Table 5). Augmented levels of DMF in *F. moschata* accessions relatively to *F. vesca* have been previously reported, yet not to this very large extent (Ulrich et al., 1997). The identification of the genetic bases for such an increased mesifurane production is beyond the scope of this work. Yet it is important to note that, PdT may represent a valuable breeding material to enhance the DMF content of garden varieties. Our analysis did not reveal detectable amount of DHF neither in PdT nor in RdV berries. This is in contrast with a preceding work, which identified this furanone in different *F. vesca* and *F. moschata* genotypes (Urruty et al., 2002). One possible explanation for such a discrepancy could reside in differences in the analytic methods employed in previous studies and ours. It is interesting to note, that DHF accumulation has been negatively correlated to the quality of the berry, as DHF-type strawberries are generally characterized by medium to poor flavor (Ulrich et al., 1997). The lack of DHF in PdT and RdV berries could alternatively be correlated to the organoleptic excellence of their fruits.

As a whole, our data provide a comprehensive metabolic map of PdT, the most fragrant strawberry of all. Despite the fact that *F. moschata* is not a direct ancestor of commercial garden strawberries, the aroma profile of PdT could assist the exploitation of this ancient clone as breeding material to enhance fruit quality in commercial strawberries. The successful introgression of wild-derived traits into cultivated garden varieties largely depends upon the possibility to generate viable and fertile interspecific hybrids. Evidently, crosses between octoploid *F. × ananassa* and species at lower ploidy level, including hexaploid PdT, are rather difficult. Yet, breeders have successfully performed inter-ploid crosses between cultivated strawberries with *F. vesca* and *F. moschata*, which produced viable hybrids with partial seed set (Luby et al., 1991). Synthetic octoploids containing varying levels of *F. moschata* have also been generated using colchicine-induced artificial doubling of chromosome number (Evans, 1982a,b). Finally, advancements in genetic engineering of cultivated strawberries have opened unprecedented possibilities for the breeding of new varieties with desirable traits. Genetic transformation has been successfully employed to enhance strawberry resistance to pests, herbicides, diseases, environmental stresses as well as fruit quality (reviewed in Qin et al., 2008). Further developments are expected in which metabolomic data, as those provided in this study, combined with genome-wide transcriptomic analysis and next generation genome-sequencing strategies will allow the identification of suitable molecular targets for engineering of volatile biosynthesis in strawberries. In this perspective, the genome of

Profumata di Tortona will prove an invaluable source of genetic material.

ACKNOWLEDGMENTS

The authors thank the “Consorzio per la valorizzazione e la tutela della Fragola Profumata di Tortona” (Tortona, Italy) and “Progetto Derthona” (Tortona, Italy) for providing the plant material along with useful technical and historical information on the clone “Profumata di Tortona”; and “Agrodinamica s.r.l.” (Tortona, Italy) for technical support with field activities.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpls.2015.00056/abstract>

REFERENCES

Aharoni, A., Giri, A. P., Verstappen, F. W., Bertea, C. M., Sevenier, R., Sun, Z., et al. (2004). Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. *Plant Cell* 16, 3110–3131. doi: 10.1105/tpc.104.023895

Aharoni, A., Keizer, L. C., Bouwmeester, H. J., Sun, Z., Alvarez-Huerta, M., Verhoeven, H. A., et al. (2000). Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *Plant Cell* 12, 647–662. doi: 10.1105/tpc.12.5.647

Azodanlou, R., Darbellay, C., Luisier, J. L., Villettaz, J. C., and Amado, R. (2003). Quality assessment of strawberries (*Fragaria* species). *J. Agric. Food Chem.* 51, 715–721. doi: 10.1021/jf0200467

Bergaglio, A. (2007). *Il Posto Delle Fragole*. Tortona: Amministrazione Comunale di Tortona.

Broeckling, C. D., Reddy, I. R., Duran, A. L., Zhao, X., and Sumner, L. W. (2006). MET-IDEA: data extraction tool for mass spectrometry-based metabolomics. *Anal. Chem.* 78, 4334–4341. doi: 10.1021/ac0521596

Buttery, R. (1983). “Vegetable and fruit flavors,” in *Flavor Research: Recent Advances*, eds R. Teranishi, R. Flath, and H. Sugiyama (New York, NY: Marcel Dekker).

Chambers, A. H., Evans, S. A., and Folta, K. M. (2013). Methyl anthranilate and gamma-decalactone inhibit strawberry pathogen growth and achene germination. *J. Agric. Food Chem.* 61, 12625–12633. doi: 10.1021/jf404255a

Christensen, C. (1983). Effect of color on aroma, flavor and texture judgment of food. *J. Food Sci.* 48, 787–790. doi: 10.1111/j.1365-2621

Colquhon, T., Levin, L., Moskowitz, H., Whiteker, V., Clark, D., and Folta, K. (2012). Framing the perfect strawberry: an exercise in consumer-assisted selection of fruit crops. *J. Berry Res.* 2, 45–61. doi: 10.3233/JBR-2011-027

Darrow, G. (1966). *The Strawberry: History, Breeding and Physiology*. New York, NY: Holt, Rinehart and Winston.

Drawert, F., Tressl, R., Staudt, G., and Koppler, H. (1973). Gaschromatographical-massspectrometrical differentiation of aroma substances from strawberries. *J. Biosci. C* 28, 488–493.

El-Mageed, A. A., Osman, A. K., Tawfik, A. Q., and Mohammed, H. A. (2011). Chemical composition of the essential oils of four *Eucalyptus* species (Myrtaceae) from Egypt. *Res. J. Phytochem.* 5, 115–122. doi: 10.3923/rjphyto.2011.115.122

Evans, W. (1982a). Guoeph SO1 synthetic octoploid strawberry breeding clone. *Hort Sci.* 17, 833–834.

Evans, W. (1982b). Guoeph SO2 synthetic octoploid strawberry breeding clone. *Hort Sci.* 17, 834.

FAOSTAT. (2014). Available online at: Available: <http://faostat3.fao.org/faostat-gateway/go/to/browse/Q/QC/E> (Accessed September 15, 2014).

Forney, C., Kalt, W., and Jordan, M. (2000). The composition of strawberry aroma is influenced by cultivar, maturity and storage. *Hort Sci.* 35, 1022–1026.

Galbiati, M., Matus, J. T., Francia, P., Rusconi, F., Canon, P., Medina, C., et al. (2011). The grapevine guard cell-related VvMYB60 transcription factor is involved in the regulation of stomatal activity and is differentially expressed in response to ABA and osmotic stress. *BMC Plant Biol.* 11:142. doi: 10.1186/1471-2229-11-142

Hancock, J., and Luby, J. (1993). Genetic resources at our doorstep: the wild strawberries. *Bioscience* 43, 141–147. doi: 10.2307/1312017

Hancock, J. (1999). *Strawberries*. Oxford, UK: CAB International.

Harada, M., Ueda, Y., and Iwata, T. (1985). Purification and some properties of alcohol acetyltransferase from banana fruit. *Plant Cell Physiol.* 26, 1067–1074.

Honkanen, E., and Hirvi, Y. (1990). “The flavour of berries,” in *Food Flavours*, eds I. Morton and A. Macleod (Amsterdam: Elsevier Scientific Publications), 125–193.

Jetti, R. R., Yang, E., Kurnianta, A., Finn, C., and Qian, M. C. (2007). Quantification of selected aroma-active compounds in strawberries by headspace solid-phase microextraction gas chromatography and correlation with sensory descriptive analysis. *J. Food Sci.* 72, S487–S496. doi: 10.1111/j.1750-3841.2007.00445.x

Larsen, M., and Watkins, C. B. (1995). Firmness and concentrations of acetaldehyde, ethyl acetate and ethanol in strawberries stored in controlled and modified atmospheres. *Postharvest Biol. Technol.* 5, 39–50. doi: 10.1016/0925-5214(94)00012

Latrasse, A. (1991). “Fruits III,” in *Volatile Compounds in Food and Beverages*, ed H. Maarse (New York; Basel, Hong Kong: Marcel Dekker Inc.), 334–340.

Livak, K., and Schmittgen, T. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001

Luby, J., Hancock, J. F., and Cameron, J. C. (1991). “Expansion of the strawberry germplasm base in North America,” in *The Strawberry into the 21st Century*, eds A. Dale and J. Luby (Portland; Oregon: Timber Press), 65–75.

Menager, I., Jost, M., and Aubert, C. (2004). Changes in physicochemical characteristics and volatile constituents of strawberry (Cv. *Cigaline*) during maturation. *J. Agric. Food Chem.* 52, 1248–1254. doi: 10.1021/f0350919

Pet'ka, J., Leitner, E., and Parameswaran, B. (2012). Musk strawberries: the flavour of a formerly famous fruit reassessed. *Flavour Fragr. J.* 27, 273–279. doi: 10.1002/ff.3095

Qin, Y., Silva, J. T. D., Zhang, L., and Zhang, S. (2008). Transgenic strawberry: state of the art for improved traits. *Biotech. Adv.* 26, 219–232. doi: 10.1016/j.biotechadv.2007.12.004

Rousseau-Guetin, M., Gaston, A., Ainouche, A., Ainouche, M., Olbricht, K., Staudt, G., et al. (2009). Tracking the evolutionary history of polyploid strawberries (*Fragaria L.*): new insights from phylogenetic analyses of low-copy nuclear genes. *Mol. Phylog. Evol.* 51, 515–530. doi: 10.1016/j.ympev.2008.12.024

Schultz, D., Craig, R., Cox-Foster, D., Mumma, R., and Medford, J. (1994). RNA isolation from recalcitrant plant tissues. *Plant Mol. Biol. Rep.* 12, 310–316. doi: 10.1007/BF02669273

Schiwietzman, M. L., Colquhoun, T. A., Jaworski, E. A., Bartoshuk, L. M., Gilbert, J. L., Tieman, D. M., et al. (2014). Strawberry flavor: diverse chemical compositions, a seasonal influence, and effects on sensory perception. *PLoS ONE* 9:e88446. doi: 10.1371/journal.pone.0088446

Skogerson, K., Wohlgemuth, G., Barupal, D. K., and Fiehn, O. (2011). The volatile compound BinBase mass spectral database. *BMC Bioinform.* 12:321. doi: 10.1186/1471-2105-12-321

Svante, W., and Michael, S. (1977). SIMCA: a method for analyzing chemical data in terms of similarity and analogy. *ACS Symp. Ser.* 52, 243–282. doi: 10.1021/bk-1977-0052.ch012

Takemoto, H., Ito, M., Shiraki, T., Yagura, T., and Honda, G. (2008). Sedative effects of vapor inhalation of agarwood oil and spikenard extract and identification of their active components. *J. Nat. Med.* 62, 41–46. doi: 10.1007/s11418-007-0177-0

Ulrich, D., and Hoberg, E. (2000a). “Flavour analysis in plant breeding research on strawberries,” in *Frontiers of Flavour Sciences*, eds P. Schieberle and K.-H. Engel. (Garching: Deutsche Forschungsanstalt Lebensmittelchemie), 161–163.

Ulrich, D., and Hoberg, E. (2000b). Rapid methods in plant aroma analysis-mass spectrometric sensor measurements on strawberries. *Acta Hortic.* 538, 443–446. Available online at: http://www.actahort.org/books/538/538_77.htm

Ulrich, D., Hoberg, E., Rapp, A., and Kecke, S. (1997). Analysis of strawberry flavour discrimination of aroma types by quantification of volatile compounds. *Z. Lebensm.-Unters. Forsch. A* 205, 218–223. doi: 10.1007/s002170 050154

Ulrich, D., Komes, D., Olbricht, K., and Horberg, E. (2007). Diversity of aroma patterns in wild and cultivated *Fragaria* accessions. *Genet. Resour. Crop Evol.* 54, 1185–1196. doi: 10.1007/s10722-006-9009-4

Ulrich, D., and Olbricht, K. (2013). Diversity of volatile patterns in sixteen *Fragaria vesca* L. accessions in comparison to cultivars of *Fragaria x ananassa*. *J. App. Bot. Food Quality* 86, 37–46. doi: 10.5073/JABFQ.2013.086.006

Ulrich, D., and Olbricht, K. (2014). Diversity of metabolite patterns and sensory characters in wild and cultivated strawberries. *J. Berry Res.* 4, 11–17. doi: 10.3233/JBR-140067

Urruty, L., Giraudel, J. L., Lek, S., Roudeillac, P., and Montury, M. (2002). Assessment of strawberry aroma through SPME/GC and ANN methods. Classification and discrimination of varieties. *J. Agric. Food Chem.* 50, 3129–3136. doi: 10.1021/jf0116799

Zorrilla-Fontanesi, Y., Rambla, J. L., Cabeza, A., Medina, J. J., Sanchez-Sevilla, J. F., Valpuesta, V., et al. (2012). Genetic analysis of strawberry fruit aroma and identification of O-methyltransferase FaOMT as the locus controlling natural variation in mesifuran content. *Plant Physiol.* 159, 851–870. doi: 10.1104/pp.111.188318

Conflict of Interest Statement: 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.

Received: 27 October 2014; accepted: 22 January 2015; published online: 11 February 2015.

Citation: Negri AS, Allegra D, Simoni L, Rusconi F, Tonelli C, Espen L and Galbiati M (2015) Comparative analysis of fruit aroma patterns in the domesticated wild strawberries “Profumata di Tortona” (*F. moschata*) and “Regina delle Valli” (*F. vesca*). *Front. Plant Sci.* 6:56. doi: 10.3389/fpls.2015.00056

This article was submitted to Plant Genetics and Genomics, a section of the journal Frontiers in Plant Science.

Copyright © 2015 Negri, Allegra, Simoni, Rusconi, Tonelli, Espen and Galbiati. 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) or licensor 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.

Extensive sequence variation in rice blast resistance gene *Pi54* makes it broad spectrum in nature

Shallu Thakur^{1,2}, Pankaj K. Singh¹, Alok Das^{1†}, R. Rathour³, M. Variar⁴, S. K. Prashanthi⁵, A. K. Singh⁶, U. D. Singh⁶, Duni Chand², N. K. Singh¹ and Tilak R. Sharma^{1*}

OPEN ACCESS

Edited by:

Joanna Marie-France Cross,
 Inönü University, Turkey

Reviewed by:

Mehboob-ur- Rahman,
 National Institute for Biotechnology
 and Genetic Engineering, Pakistan
 Swarup Kumar Parida,
 National Institute of Plant Genome
 Research, India

*Correspondence:

Tilak R. Sharma,
 National Research Centre on Plant
 Biotechnology, Pusa Campus, LBS
 Building, New Delhi-110012, India
 trsharma@hrcpb.org;
 trsharma1965@gmail.com

†Present Address:

Alok Das,
 Indian Institute of Pulses Research,
 Kanpur, India

Specialty section:

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

Received: 30 January 2015

Accepted: 30 April 2015

Published: 21 May 2015

Citation:

Thakur S, Singh PK, Das A, Rathour R, Variar M, Prashanthi SK, Singh AK, Singh UD, Chand D, Singh NK and Sharma TR (2015) Extensive sequence variation in rice blast resistance gene *Pi54* makes it broad spectrum in nature. *Front. Plant Sci.* 6:345. doi: 10.3389/fpls.2015.00345

Rice blast resistant gene, *Pi54* cloned from rice line, Tetep, is effective against diverse isolates of *Magnaporthe oryzae*. In this study, we prospected the allelic variants of the dominant blast resistance gene from a set of 92 rice lines to determine the nucleotide diversity, pattern of its molecular evolution, phylogenetic relationships and evolutionary dynamics, and to develop allele specific markers. High quality sequences were generated for homologs of *Pi54* gene. Using comparative sequence analysis, InDels of variable sizes in all the alleles were observed. Profiling of the selected sites of SNP (Single Nucleotide Polymorphism) and amino acids (N sites ≥ 10) exhibited constant frequency distribution of mutational and substitutional sites between the resistance and susceptible rice lines, respectively. A total of 50 new haplotypes based on the nucleotide polymorphism was also identified. A unique haplotype (H_3) was found to be linked to all the resistant alleles isolated from *indica* rice lines. Unique leucine zipper and tyrosine sulfation sites were identified in the predicted *Pi54* proteins. Selection signals were observed in entire coding sequence of resistance alleles, as compared to LRR domains for susceptible alleles. This is a maiden report of extensive variability of *Pi54* alleles in different landraces and cultivated varieties, possibly, attributing broad-spectrum resistance to *Magnaporthe oryzae*. The sequence variation in two consensus region: 163 and 144 bp were used for the development of allele specific DNA markers. Validated markers can be used for the selection and identification of better allele(s) and their introgression in commercial rice cultivars employing marker assisted selection.

Keywords: allele mining, *Pi54*, rice landraces, polymorphism, blast resistance genes, allele specific markers

Introduction

Blast disease caused by the fungus, *Magnaporthe oryzae* is one of the most widespread and devastating diseases of rice. Management of rice blast through host resistance is a promising component of the Integrated Disease Management (IDM) programme. Till date, about 101 major rice blast resistance (*R*) genes have been identified, and 20 of them cloned and characterized (Sharma et al., 2012). Numerous *R*-genes identified, cloned and characterized are categorized in

eight classes based on their amino acid motif organization (Sharma et al., 2014). Majority of loci associated with rice blast disease resistance have been reported on chromosome 11 of rice based on genome wide association studies (Wang et al., 2014). Although several blast resistance loci have been identified but only few of them have been employed in breeding for blast management in India (Singh et al., 2011). Further limited success has been realized in durable resistance breeding programmes due to variability of pathogen across locations. Harnessing rice diversity adapted in farmers' fields over the years appears promising alternative to look for resistance source.

Exploring the genetic variants from germplasm (wild and cultivated) is currently being envisaged in many crop species. One of the most widely used methods to identify variants employs polymerase chain reaction (PCR) based techniques to amplify homologs (possible alleles) from the gene pool, known as *allele mining*. Recently, allele mining for blast resistance has been reported from wild and cultivated species of rice (Yang et al., 2007; Geng et al., 2008; Huang et al., 2008). Studies of *Pi-ta* gene in rice lines including wild (AA and CC genome) and cultivated species indicated consensus conserved sequence before divergence (Wang et al., 2008). In another study, *Pi-ta* orthologs from 26 accessions (*Oryza rufipogon*, *O. sativa*, *O. meridionalis*, and *O. officinalis*), collected from 10 different countries highlighted dimorphic pattern of nucleotide polymorphism and low nucleotide diversity at the LRD region (Yoshida and Miyashita, 2009). In similar lines, the allelic variants and flanking sequences of *Pi-ta* have been studied in 159 geographically diverse accessions of *Oryza* species (AA genome) (Lee et al., 2009). The *Pi-ta* alleles also have been studied extensively in Indian landraces (Thakur et al., 2013b). Other blast resistance loci like *Pid3*, *Pi9* and *Piz(t)* has been explored to study the nucleotide polymorphism and evolutionary pressure (Shang et al., 2009; Liu et al., 2011; Thakur et al., 2013a). However, such detailed analysis is lacking for the important blast resistance gene, *Pi54* that confers broad spectrum resistance to blast disease (Sharma et al., 2010). The *Pi54* gene located on chromosome 11 having unique zinc finger domain, besides LRR domain (Sharma et al., 2005a,b; Gupta et al., 2012). Functional complementation indicated that this gene provides stable and high level of resistance against geographically diverse strains of *M. oryzae*, collected from different parts of India (Rai et al., 2011). The gene possibly triggers up-regulation of defense response genes (callose, laccase, PAL, and peroxidase), transcription factors (NAC6, Dof Zinc finger, MAD box, bZIP, and WRKY) that fortify cell wall/plasmodesmata leading to hypersensitive response, and affecting resistance reaction (Gupta et al., 2011). Currently, the gene is being used in enhanced blast resistant breeding programme (Singh et al., 2011). An ortholog of *Pi54* gene from wild species of rice has also been recently cloned and functionally validated (Das et al., 2012). However, the allelic variants of *Pi54* gene have not been characterized from rice landraces, that are believed to have co-evolved with pathogen, and hence represents better "evo-devo" perspective of resistance reaction.

Till date, cultivated and wild species of rice have been employed for prospecting novel variants of blast resistance genes. Landraces too represents unmatched genetic potential for rice

improvement. The local landraces or local rice varieties are genetically diverse, balanced population and are in equilibrium with the environment and pathogens, in contrast to the rice varieties. Unlike high yielding varieties, the landraces are endowed with tremendous genetic variability, as they are not subjected to subtle selection over a long period of time. Probably, it helps landraces to adapt in wide agro-ecological niches with unmatched qualitative traits, medicinal properties and important genetic resources for resistance to pests and diseases. Owing to their specific domination in geo-graphical niches, landraces have genes of resistance to biotic stresses, which have not been widely utilized or incorporated into modern varieties (Ram et al., 2007). The landraces grown in rice blast "hot-spots" of the Indian sub-continent has remained largely unexplored. Molecular markers linked to major *R*-genes represent an important tool for marker assisted selection (MAS) (Costanzo and Jia, 2010). Variations in terms of Single Nucleotide Polymorphisms (SNPs) and insertion-deletions (InDels) covering the entire genic segment can be compared among genotypes to identify functional markers to aid the selection process. Markers associated with two cloned blast *R* genes (*Pi-b* and *Pi-ta*) as well as a PCR-based SNP markers for *Piz* locus and *Pik* locus (Hayashi et al., 2004; Jia et al., 2009; Zhai et al., 2011) are to mention a few. Conventional breeding with MAS would therefore, benefit from the development of new *R* gene specific markers, which would allow pyramiding of multiple genes in adapted germplasm toward realizing broader spectrum disease resistance. Molecular population genetic analysis of local landraces and cultivated varieties might provide insight on the selection forces maintaining resistance and preventing evolution of new specificities in natural pathogen populations. Therefore, this study was conducted with objectives, (i) analysis of variants of *Pi54* alleles from the cultivated varieties and Indian landraces of rice collected from different eco-geographical regions (ii) structural analysis of *Pi54* alleles to understand molecular evolution at the loci, and (iii) development of allele specific functional markers for use in marker assisted selection.

Materials and Methods

Plant Material and Fungal Culture

A set of 92 rice lines (landraces and cultivated varieties) were selected from different geographic locations of India for prospecting of *Pi54* alleles. The diagnostic isolate of *M. oryzae* (Mo-nwi-37-1) was used for the phenotypic evaluation of all the rice lines (Rai et al., 2011; Rathour et al., unpublished data).

Preparation of Fungal Culture

Fungal culture of Mo-nwi-37-1 was maintained on Oat Meal Agar (HiMedia, India) medium in pre-sterilized petriplates (90 mm diameter). For sporulation, the culture was multiplied in Mathur's medium (Dextrose 8 g/L, Magnesium sulfate 2.5 g/L, Potassium phosphate 2.75 g/L, Neo-Peptone 2.5 g/L, Yeast Extract 2.0 g/L, and agar 16 g/L). The culture plates were maintained at 22°C for 12–16 days under constant illumination with white fluorescent light (55 μF/Em/s). For the preparation of fungal spores, 5 ml of 0.2% gelatine solution was added to individual plate on agar surface and gently rubbed with scrapper to separate conidia from the conidiophores. The spore

concentration was brought to approximately 10^5 spores/ml. The seedlings were sprayed with spore suspension of about 1 ml per plant at 2–3 leaf stage.

Inoculation of Rice Lines with Diagnostic *M. oryzae* Isolate

Rice lines were grown in plastic pots (12 inch dia.) containing sterilized potting mixture in the rice blast testing facility, NRCPB, IARI, New Delhi. Rice lines, Tetep, and Taipei 309 were used as positive and negative controls, respectively. Physical parameters were set for 16 h/8 h light-dark photoperiod. The day and night temperatures were maintained at 25°C and 21°C, respectively, with relative humidity (RH) of more than 90%. All the seedlings assessed in the experiment were sprayed simultaneously with *M. oryzae* spore suspension of 10^5 spores/ml. Disease reaction was recorded after 7 days of inoculation using 0–5 disease assessment scale (Bonmann et al., 1986). Where, 0 = No evidence of infection; 1 = Brown specks smaller than 0.5 mm in diameter, no sporulation; 2 = Brown specks about 0.5–1.0 mm in diameter, no sporulation; 3 = Roundish to elliptical lesions, 1–3 mm in diameter, gray center surrounded by brown margins, lesions capable of sporulation; 4 = Typical spindle shaped blast lesions capable of sporulation, 3 mm or longer; 5 = lesions as in 4 but about half of 1–2 leaf blades killed by coalescence of lesions. Reaction types 0, 1, 2, and 3 were considered resistant, while 4 and 5 considered as susceptible.

PCR Amplification and Sequencing

Genomic DNA was extracted from fresh leaves of selected rice lines using the modified Cetyltrimethyl Ammonium Bromide (CTAB) method of DNA isolation (Murray and Thompson, 1980). For PCR amplification, nucleotide sequence of the blast resistance gene *Pi54* (Loc_Os11g42010) was retrieved from NCBI database (www.ncbi.nlm.nih.gov/). Overlapping oligos Pi54_F1 (CAATATAGCTGGAAATTTCAGAGG) and Pi54_R1 (AGATAATGTGTTGCTGGCTGTC); Pi54_F2 (CATGAA CAGAGCACTGATGACATA) and Pi54_R2 (GGATAACAA GCACTGAGCCATATC); Pi54_F3 (CCGTTCTGACCATAG AAATTATCG) and Pi54_R3 (GTGCAATTACATAAGCTA GACCTTG) were designed using Primer 3 software (Rozen and Skaletsky, 2000) to amplify 1.5 kb region using primer walking technique. PCR was performed with genomic DNA isolated from rice landraces and cultivated varieties using Pfu polymerase (FINNZYMES OY, Keilaranta, Espoo, Finland) with the following thermal cycling conditions: initial DNA denaturation at 95°C for 2 min followed by 30 cycles of 95°C for 30 s, 58°C (Pi54_F1 and Pi54_R1; Pi54_F2 and Pi54_R2) or 60°C (Pi54_F3 and Pi54_R3) for 30 s, 72°C for 1 min, final elongation at 72°C for 10 min and hold at 4°C. The PCR derived amplification products were used as template for determining DNA sequences using Sanger's dideoxy method of sequencing.

Sequencing of PCR Amplicon

The purified PCR amplicon was sequenced according to manufacturer's instruction directly by using modified Sanger's dideoxy terminator cycle sequencing chemistry on an automated capillary-based DNA sequencer (ABI 3730xl DNA Analyzer)

in both forward and reverse direction twice using amplified product specific primers. The PCR products were run in a cycle sequencing reaction with thermal cycling conditions as 30 cycles of denaturation (95°C for 20 s), annealing (60°C for 20 s), and extension (60°C for 4 min) followed by hold at 4°C. The purified sequencing products were resolved on a capillary-based automated DNA sequencer (ABI 3730xl DNA Analyzer). Full length sequence reads were obtained by assembly of multiple reads of each fragment using Phred/Phrap and Consed software (Ewing and Green, 1998). Each fragment was sequenced at least four times and high quality (Phred 20) consensus sequence was used for data analysis.

Analysis of Sequenced *Pi54* Alleles

The sequenced data were aligned using ClustalW 2.0 (Larkin et al., 2007) at their default alignment parameters and manually corrected by MEGA 4.0. Gene coding regions were predicted with FGENESH (Solovyev et al., 2006) using the original *Pi54* (Tetep) sequence as a reference. The LRR domain was identified as described earlier (Sharma et al., 2005a). Motif was identified using motif scan software (<http://hits.isb-sib.ch/cgi-bin/PFSCAN>) and SMART tool (<http://smart.embl-heidelberg.de/>). Phylogenetic analysis was performed with MEGA 4.0 using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1000 replicates was used to represent the evolutionary history of the taxa analyzed (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and were depicted in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). Analysis of overall transition: transversion ratio, variable, and parsimony informative positions were calculated using MEGA 4.0 software (Tamura et al., 2007).

The DnaSP 5.10 software was used for the analysis of nucleotide polymorphism (Rozas et al., 2003). The aligned DNA sequences were imported into the DnaSP software to calculate S (number of polymorphic or segregating sites), π (nucleotide diversity), θ (Theta from S, Theta-W), and D (Tajima's D), and to draw the sliding window of nucleotide diversity (π). Haplotype networks were constructed for each potential SNP's sites by statistical parsimony with the software TCS 1.21 (Templeton et al., 1992; Clement et al., 2000). The networks were assembled based on an absolute distance matrix between haplotypes, i.e., the number of mutations separating each haplotypes, with a parsimony probability of 95%. Haplotype diversity (Hd) was analyzed between disease resistance and susceptible phenotypes of *Oryza* species. The DnaSP 5.10 program (Rozas et al., 2003) was used for the analysis of haplotype diversity Hd (Nei, 1987).

Development of Functional Markers

The PCR-based co-dominant and dominant STS (sequence tagged site) markers targeting consensus InDels of 144 and

163 bp, respectively were designed. In case of dominant marker, the forward primer (DPi54_163F) was designed flanking to the insertion (163 bp) and reverse primer (DPi54_163R) was designed from the sequence of insertion. The forward (CPi54_144F) and reverse (CPi54_144R) primers in case of co-dominant markers were designed from the flanking regions of 144 bp insertion. Primer pairs [CPi54_144F (AAGTACTTCAT GATCTATTCTACTGG) and CPi54_144R (CCGTTCTGACC ATAGAAATTATCG)]; DPi54_163F (ACCATGACTAGCTATG AAAAATCT) and DPi54_163R (AGAATAGATCATGAACTA CTTGAAAC)] were designed by using Primer 3 software (Rozen and Skaletsky, 2000). PCR amplification was carried out on programmable Thermal Cycler (BioRad, Washington DC, USA) using the following temperature profile: initial DNA denaturation, 95°C for 2 min; followed by 35 cycles of denaturation, 94°C for 20 s; annealing, 55°C (DPi54_163F and DPi54_163R) or 58°C (CPi54_144F and CPi54_144R) for 30 s; extension, 72°C for 1 min; and final extension at 72°C for 10 min and then hold at 4°C using Taq DNA polymerase (Vivantis, USA). The PCR amplified product was resolved in 2% agarose gel using 1.0X TAE buffer.

Results

Phenotypic Evaluation of Rice Landraces

The rice lines (92) used in present study were grown in contained condition and the 15 days-old-seedlings were challenged with representative *M. oryzae* isolate, Mo-nwi-37-1. After a week of inoculation, all the rice lines were grouped into resistant and susceptible categories based on their reaction to *M. oryzae*. Out of 92 rice lines, 72 were found resistant and the rest 20 susceptible, based on disease assessment scale (Bonmann et al., 1986) (Table S1). These lines were used for the allele mining studies of *Pi54* gene.

Nucleotide Polymorphism

To determine the nucleotide diversity at the *Pi54* locus in rice lines, 1.5 kb long fragments were amplified from 92 rice lines and sequenced using 3 different overlapping primer combinations (Figure 1). All the fragments were sequenced and high quality (>Phred 20) assembled sequence of each allele has been deposited in the EMBL database (Table S1). For sequence analysis, the *Pi54* alleles were grouped into three different categories: (i) phenotypes: resistant and susceptible (ii) landraces and cultivated varieties; (iii) *indica* and *japonica* types (Table 1). Nucleotide variations were high in the genic region of *Pi54* allele. A total of 197 SNPs, large insertions of 38, 49, 144, and 163 bps and single base pair deletions were identified in the *Pi54* alleles.

We calculated percentage of mutational change with respect to the reference *Pi54* gene and compared within and between disease resistance and susceptible alleles of *Oryza* species. All mutational changes were scored for the specific positions only. The number of mutations per site was found to be equal to and greater than 10 (i.e., $N_{\text{mut}} \geq 10$). Overall, 40 mutational sites were identified across the alignment, out of which 27 are transitions (ti) and 12 transversion (tv) (Table S2).

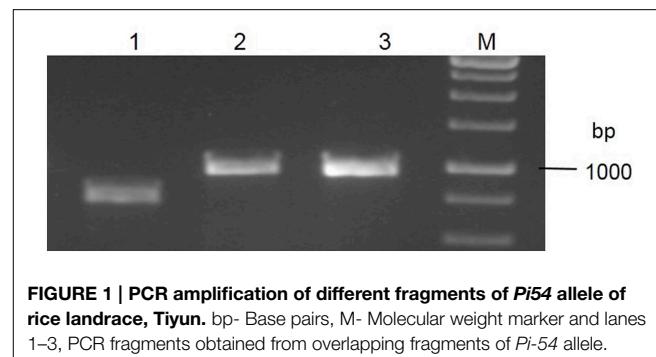


FIGURE 1 | PCR amplification of different fragments of *Pi54* allele of rice landrace, Tiyun. bp- Base pairs, M- Molecular weight marker and lanes 1–3, PCR fragments obtained from overlapping fragments of *Pi54* allele.

TABLE 1 | Different groups and rice accessions used in mining for blast resistance gene *Pi54*.

Group	Species	No. of accessions
Cultivar	<i>O. sativa indica</i>	48
	<i>O. sativa japonica</i>	3
Landrace	<i>O. sativa indica</i>	32
	<i>O. sativa japonica</i>	3
	<i>O. sativa aus</i>	2
Breeding line	<i>O. sativa indica</i>	2
Hybrid	<i>O. sativa indica</i>	1
Unknown	<i>O. sativa indica</i>	1

Further, mutational profiling of disease resistance and susceptible phenotypes of *Oryza* species was constructed in all alleles mined from 92 rice landraces. It was found that 61% of mutational sites have mutations at one site, 18% of the sites have mutations at two-nine sites and rest of the 21% sites have mutation at 10 or more than 10 sites (Figure 2A). Most of the mutational sites exhibited a constant frequency distribution between the resistance and susceptible groups (Figure 2B).

The *Pi54* alleles of landraces harbor substantially higher polymorphism as compared to the *Pi54* alleles of cultivated species, because of heterogeneous nature of rice landraces used in present study, which might have accumulated more mutations during the course of evolution. The nucleotide diversity was high in the *Pi54* alleles of resistant lines compared to the *Pi54* alleles of susceptible lines (Table 2). The *Pi54* alleles of *indica* and *japonica* species were almost equally diverse at nucleotide level. Higher nucleotide variation at the resistant loci is possibly due to interaction with highly avirulent and frequently mutating avirulent strains of *M. oryzae*. Overall, a total of 198 polymorphic sites excluding InDels, were identified in the 1.5 Kb region of *Pi54* alleles. Maximum (225) polymorphic sites were identified in the *Pi54* alleles isolated from Indian landraces; however, very less diversity was obtained in the alleles cloned from *japonica* species (Table 2). Average pair wise nucleotide diversity (π) and Watterson's nucleotide diversity estimator (θ_w) over the *Pi54* alleles in susceptible rice lines ($\pi = 0.0208$ and $\theta_w = 0.01916$) was lowest among all other *Pi54* alleles included in the present study. Within the groups, the nucleotide diversity was lowest in *Pi54* alleles of cultivated varieties ($\pi = 0.02254$

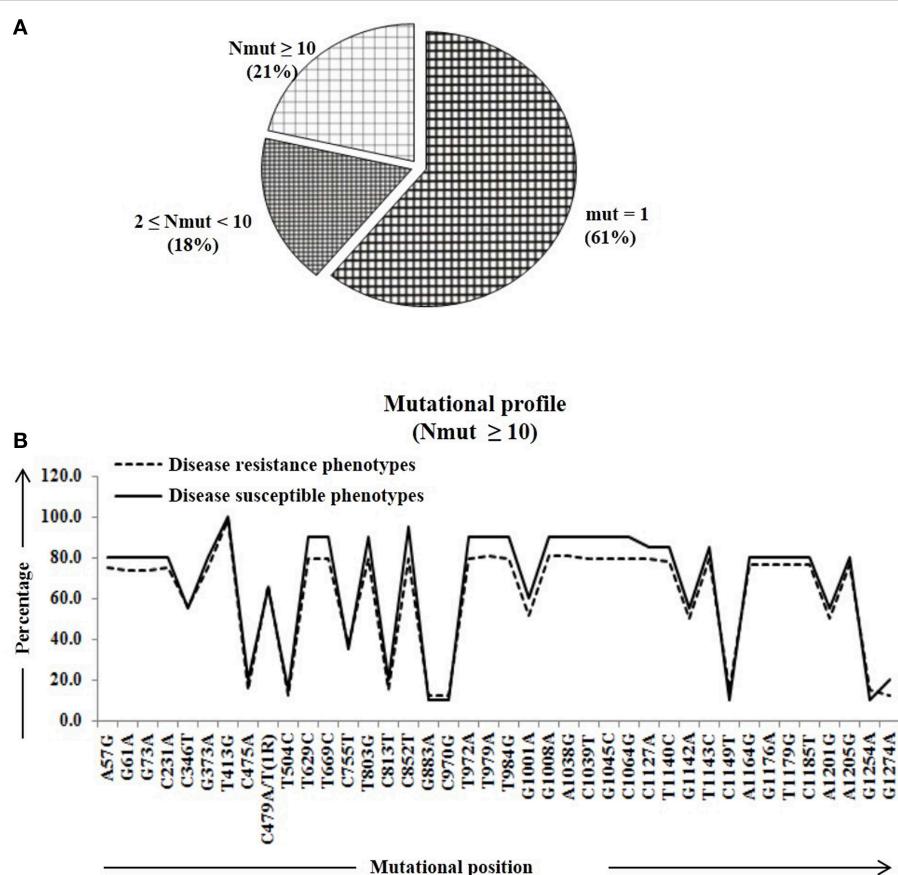


FIGURE 2 | The number of amino acid substitutions in the predicted 64 *Pi54* protein. **(A)** Distribution of amino acid substitutions across 64 predicted proteins. **(B)** Substitutional profile of amino acids of disease resistant and susceptible phenotype.

and $\theta_w = 0.02102$) compared to landraces ($\pi = 0.03417$ and $\theta_w = 0.03877$). Among the *Pi54* alleles of *indica* and *japonica* species, diversity was low in the alleles of *japonica* species (Table 2). The LRR domain harbors substantial diversity, explaining the selection pressure it underwent. Higher diversity was observed in LRR domain of *japonica* species as compared to others (Figure 3).

Phylogenetic Relationship between the Alleles

Phylogenetic tree constructed based on *Pi54* allelic sequences obtained from 92 accessions resulted in two major clusters (Figure 4). Both the clusters were further divided into separate sub-clusters but species specific clustering was not obtained. Similarly, separate phylogenetic tree was also constructed for the *Pi54* alleles derived from resistant and susceptible lines. All the resistant *Pi45* orthologs were grouped into three clusters (Figure 4A). Cluster I and II were further divided into two sub-clusters, i.e., I_A, I_B, and II_A, II_B. The sub-cluster I_A and cluster III consisted of all the resistance alleles belonging to *indica* group whereas all other sub-clusters included alleles of *indica* as well as *japonica* groups. The *Pi54* alleles from susceptible lines were grouped into two major distinct clusters (Figure 4B).

Pattern of Molecular Evolution

A haplotype network was constructed to identify mutational changes, based on potential SNPs of all *Pi54* alleles isolated from 92 rice lines. We identified fifty new haplotypes from the nucleotide polymorphism of the *Pi54* alleles. To determine the linkage among these haplotypes, a haplotype network was constructed (Figure 5). In this network, 50 haplotypes were clustered in five major haplogroup (major haplogroup contain three or more *Pi54* alleles) and the rest as minor haplogroups. We identified resistant phenotype specific haplotype (H-3) and rest of the major haplotypes contained alleles from resistant and susceptible lines as well. Furthermore, haplotype, H-4 was the only mixed haplotype consisted of the *Pi54* alleles of *indica*, *japonica* and *aus* rice accessions. Similar haplotype network has been constructed for resistant as well susceptible *Pi54* alleles. In case of resistance alleles, total number of identified haplotypes was 41 which clustered into five major haplogroups (Figure S1). All the major haplogroups consisted of *Pi54* alleles isolated from *indica* rice lines. Total number of 14 haplotypes identified in *Pi54* alleles was from susceptible rice lines and were, clustered into four major haplogroups (Figure S2). The major haplogroup of these rice lines consisted of *Pi54* alleles of *Indica* rice lines except for H1 which is a mixed cluster of alleles from both

TABLE 2 | Nucleotide polymorphism of the *Pi54* alleles mined from different accessions of rice.

Group	Elements	Position (nt)	S	π	$P(\pi)$	Θ	$P(\Theta)$	Tajima's D	Ka/Ks
All	Whole sequence	[1–4194]	198	0.0226	0.00146	0.02972	0.00801	-1.07559	<1
	CDS	[1338–3585]	163	0.02541	0.0018	0.03289	0.00893	-1.03898	<1
	LRR	[3324–3462]	15	0.01073	0.00218	0.0235	0.00847	-1.70363	>1
<i>O. sativa indica</i>	Whole sequence	[1–4194]	179	0.02245	0.00131	0.02717	0.00753	-0.87382	<1
	CDS	[1338–3585]	140	0.02431	0.00149	0.029	0.00811	-0.86041	<1
	LRR	[3324–3462]	9	0.00839	0.00123	0.01451	0.00607	-1.34456	>1
<i>O. sativa japonica</i>	Whole sequence	[1–4194]	77	0.0238	0.00487	0.02285	0.01094	-0.06202	<1
	CDS	[1338–3585]	68	0.02899	0.00632	0.02815	0.01352	-0.08869	<1
	LRR	[3324–3462]	9	0.02815	0.01044	0.0292	0.01615	-0.21398	>1
Cultivar	Whole sequence	[1–4194]	123	0.02254	0.00144	0.02102	0.00646	-0.08589	<1
	CDS	[1338–3585]	106	0.02576	0.00176	0.02431	0.00752	-0.14919	<1
	LRR	[3324–3462]	7	0.00883	0.00153	0.01259	0.00587	-1.13434	>1
Landrace	Whole sequence	[1–4194]	225	0.03417	0.00467	0.03877	0.01246	-0.94232	<1
	CDS	[1338–3585]	141	0.02451	0.00371	0.03502	0.01139	-1.36704	<1
	LRR	[3324–3462]	13	0.01247	0.00461	0.02524	0.01036	-1.72346	>1
Disease resistance phenotypes	Whole sequence	[1–4194]	192	0.0237	0.00171	0.03004	0.00848	-0.9818	<1
	CDS	[1338–3585]	158	0.0267	0.00215	0.03366	0.00957	-0.96141	<1
	LRR	[3324–3462]	15	0.01213	0.00267	0.02481	0.00917	-1.67675	>1
Disease susceptible phenotypes	Whole sequence	[1–4194]	87	0.0208	0.00286	0.01916	0.00728	0.27313	<1
	CDS	[1338–3585]	71	0.02208	0.00327	0.02104	0.00806	0.15424	<1
	LRR	[3324–3462]	3	0.0057	0.00174	0.00699	0.00454	-0.56505	>1

S, Number of Polymorphic or Segregating sites; π , Nucleotide diversity; Θ , Watterson's nucleotide diversity estimator based on silent sites. Tajima's D calculation based on the difference between the number of segregating sites and the average number of nucleotide differences. Jukes–Cantor (JC) corrected synonymous differences per synonymous site (Ks) and non-synonymous differences per non-synonymous site (Ka). For Tajima's D, no value is significant at 0.10 level ($P > 0.10$) and 0.10 $> P > 0.05$ (for bold).

indica and *japonica* lines. Statistically high haplotypes diversity (0.935/0.019) was observed within the studied data set of 92 *Pi54* alleles of *Oryza* species. In case of susceptible and resistant alleles, high haplotypes diversity (0.958/0.028) was observed in disease susceptible *Pi54* alleles of *Oryza* species than resistance (0.935/0.019) alleles (Table 3).

In present study, sequence analysis indicated variable number (0–3) of Open Reading Frames (ORFs) in the *Pi54* locus, with exception of 28 rice lines where no ORF was detected. Absence of ORF in these sequences might be due to reshuffling or recombination events of the locus resulting in absence of start codon or pseudogenized allele, which has lost its function in due course of time. In 40 rice lines, single exon was predicted. However, two and three exons were also predicted in the allelic sequences of 21 and 10 rice lines, respectively (Table S3). This may be due to the creation of new splice sites during the course of evolution. Variation in the number of ORF might have generated based on the selection pressure, it underwent during the evolutionary process. Various insertions have also been identified in the ORFs. The presence of insertion implicates its differential role in regulating disease resistance.

Percentage of substitutional change was calculated at protein level in the *Pi54* gene and compared within and between

disease resistance and susceptible phenotypes of *Oryza* species. The amino acid substitutions were scored for the specific positions. The number of substitution per site was found to be equal to and greater than 10 (i.e., $N_{mut} \geq 10$). Overall, 23 substitutional sites were identified across the alignment of 64 *Pi54* proteins (Table S4). The number of mutations per mutational site was calculated in all the aligned 64 predicted sequences implicating 66% of mutational sites having mutation in one site, 19% of the sites having mutations in two to nine sites and rest of the 15% sites having mutation in 10 or more than 10 sites (Figure 6A). The mutational profiling of disease resistance and susceptible phenotypes of *Oryza* species indicated that most of the mutational sites were showing a constant frequency distribution between the resistance and susceptible groups (Figure 6B).

The nucleotide sequences of all the *Pi54* alleles were translated and the predicted proteins ranged between 73 and 486 amino acid residues having many predicted functional domains. The Zinc-finger domain (ZnF) was predicted in all the *Pi54* proteins except for a few (Figure S3). The sequence of predicted ZnF domain was highly conserved (100% similarity) in all the *Pi54* proteins (Figure S3). Important motifs identified in the translated sequences of the *Pi54* alleles are N-glycosylation sites,

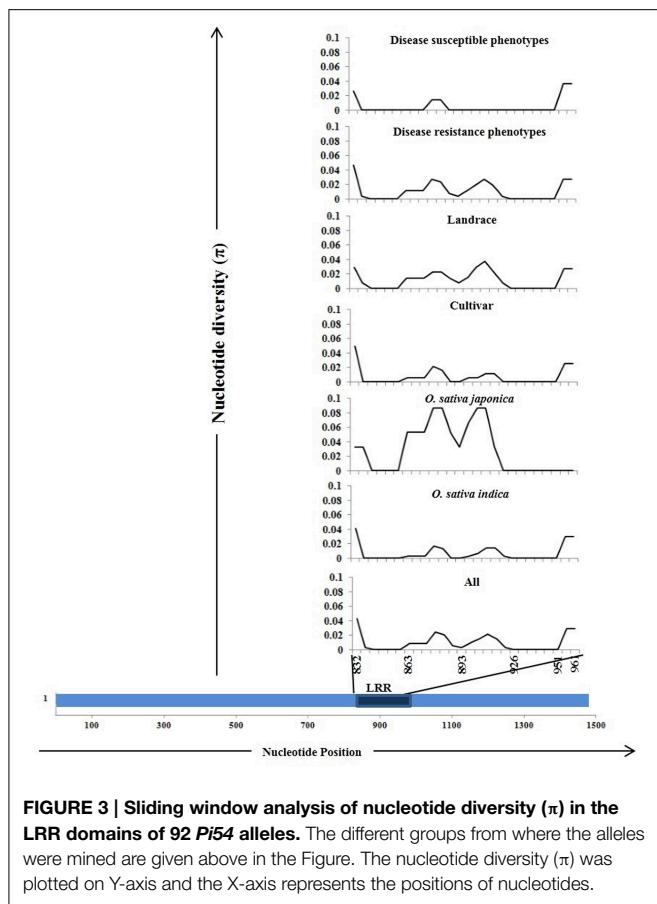


FIGURE 3 | Sliding window analysis of nucleotide diversity (π) in the LRR domains of 92 *Pi54* alleles. The different groups from where the alleles were mined are given above in the Figure. The nucleotide diversity (π) was plotted on Y-axis and the X-axis represents the positions of nucleotides.

phosphorylation (kinase C phosphorylation site, casein kinase II phosphorylation site, tyrosine kinase phosphorylation site), tyrosine sulfation site, N-myristoylation site, and leucine zipper (Table S5). These sites were present in variable numbers in all the *Pi54* alleles except for tyrosine kinase phosphorylation site and leucine zipper. The presence of phosphorylation sites in *Pi54* alleles, indicate their involvement in signal transduction by activating further downstream genes. The presence of N-myristoylation sites in the predicted proteins play important role in membrane anchoring whereas N-glycosylation sites has significant role in protein targeting. Unique leucine zipper was identified in 54 and 73% of the resistant and susceptible *Pi54* proteins, respectively. They are usually found as a part of DNA-binding domain in many transcription factors, and are therefore involved in regulating gene expression. Similarly, tyrosine sulfation sites were also present in 18 and 21% of the resistant and susceptible *Pi54* protein, respectively, which plays important role in strengthening the protein-protein interaction. From above results, it can be concluded that unique Leucine zipper and single tyrosine sulfation sites identified in some of the *Pi54* predicted protein sequences which was absent in the reference *Pi54* protein.

To evaluate the phylogenetic relationship amongst the predicted *Pi54* proteins, neighbor-joining trees were constructed using the LRR regions (Figure S4). All the predicted *Pi54*

proteins were grouped into three separate clusters of mixed type. This is in contrast to species-specific groups obtained from NBS and LRR domains of *Pi54* alleles (Liu et al., 2011).

Analysis of Evolutionary Dynamics

To test the evolutionary selection dynamics of the *Pi54* alleles in 92 *Oryza* accessions, we evaluated the extent of neutral selection with D statistics (Tajima's D test) (Tajima, 1989). In the present study, *Pi54* alleles have been subjected to positive selection [Tajima's D = -1.07559] for all the alleles and deviating from the model of neutrality (Table 2). It is noteworthy that *Pi54* alleles of different groups, such as landraces and cultivated varieties, *indica* and *japonica* species, and blast resistant lines have been subjected to positive selection whereas the balancing selection operates in *Pi54* alleles of susceptible rice lines [Tajima's D = 0.27313]. This might be due to the variable selection pressure acting on the locus, or diverse sample size used in present analysis. Further, coding region of 92 *Pi54* alleles and *Pi54* alleles of different groups have been analyzed. The value of Tajima's D was negative in the entire coding region (-1.03898) and LRR domain (-1.70363) indicating purifying selection in the CDS and LRR regions (Table 2). The ratio of synonymous (k_s) and non-synonymous (k_a) divergence in whole sequence as well as coding region and parts of the coding region (LRR domains) was calculated in all the 92 *Pi54* alleles and separately for the *Pi54* alleles from different groups (Table 2). The value of k_a/k_s ratio was used as a criterion for the presence or absence of positive selection for amino acid substitutions. The value of k_a/k_s in the whole sequence, coding region of the *Pi54* alleles was less than one, indicating low level of polymorphism in these regions, in contrast to high level of polymorphism in LRR domain of all *Pi54* alleles as the ratio of k_a/k_s was greater than one (Table 2). In the LRR region of the *Pi54* alleles of different groups, the value of k_a/k_s is greater than one, which indicates that positive directional selection might have favored amino acid substitution in this region. In the present study, LRR region of *Pi54* alleles was quite variable and might have role in different recognition specificities, possibly making it more durable.

Development of Allele Specific Markers

Allele specific DNA markers are important for the introgression of resistant alleles in cultivated rice varieties using marker assisted selection (MAS) strategy. In this study, we developed dominant and co-dominant STS markers based on the large DNA insertions in the allelic sequences of *Pi54* alleles. The dominant markers (DPi54_163F and DPi54_163R) were specifically designed to amplify a fragment of 278 bp in resistant *Pi54* alleles and absence of band in the susceptible alleles (Figure 7A). By using this marker, we were able to distinguish 22 rice lines having resistant *Pi54* allele (presence of 278 bp amplification product) and 15 rice lines having susceptible *Pi54* alleles (Table 4). Similarly, co-dominant marker (CPi54_144F and CPi54_144R) was also tested in a set of rice lines used in present study. PCR amplification with co-dominant markers amplifies fragments of 557 and 313 bp (Figure 7B). The 557 bp fragments were amplified in 10 susceptible rice lines, and a 313 bp band was present in 28

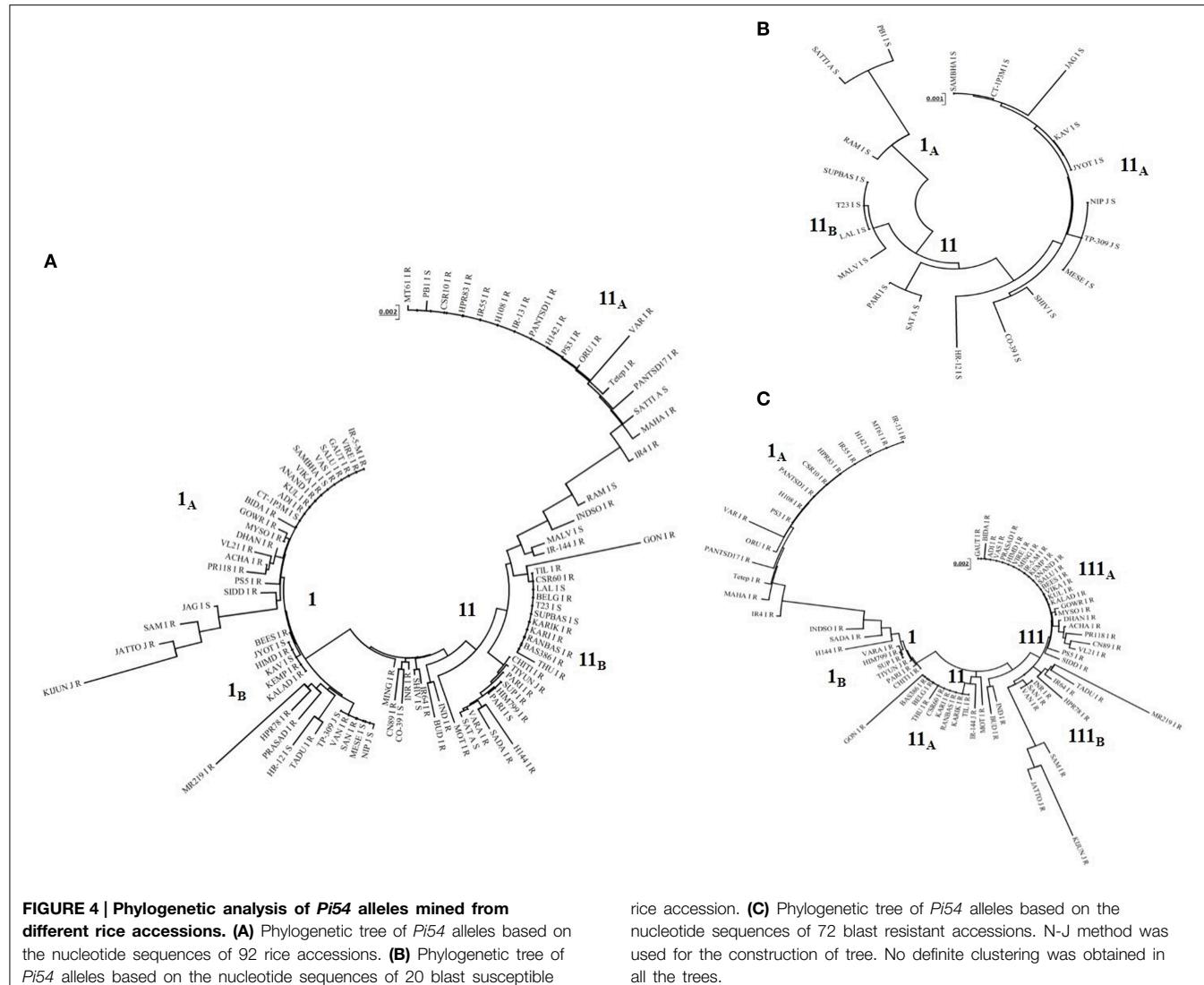


FIGURE 4 | Phylogenetic analysis of *Pi54* alleles mined from different rice accessions. **(A)** Phylogenetic tree of *Pi54* alleles based on the nucleotide sequences of 92 rice accessions. **(B)** Phylogenetic tree of *Pi54* alleles based on the nucleotide sequences of 20 blast susceptible

resistant rice lines. Eight rice lines showed presence of both the alleles (Table 5).

Discussion

Breeding efforts have capitalized only a fraction of the genetic diversity available to us. Food availability needs to be increased in face of intensifying demand, climate change, soil degradation, land, and water shortages. Farmers are savors of seeds of crop species, primitive varieties (local domesticates called *landraces*), wild relatives of crop species (McCouch, 2013). The biodiversity present within the farmer adapted land races must be mined to discover novel sources of resistance to pests and diseases. Chromosome 11 of rice as reported, has the most associated disease resistance loci and the highest frequency of copy number variations (CNVs). Genes in most of CNVs were reported to be associated with resistance phenotype (Yu et al., 2011; Wang et al., 2014). The allelic variants of the dominant blast resistance gene, *Pi54* located on chromosome 11 were prospected and variations

in terms of SNPs and InDels were documented. These variations possibly might play an important role in the durability of *Pi54* gene against *M. oryzae* population. In earlier studies, InDels and SNPs have shown to play a pivotal role in *R*-gene evolution through selection (Shen et al., 2006). The presence of 5 Mb region (super locus) physically linked to *Pi-ta* gene impart resistance phenotype (Jia and Martin, 2008; Lee et al., 2009). The higher frequency of SNP observed in present study might be due to the combined analysis of both landraces and cultivated species. Similarly, higher variation was also observed between *O. sativa* and *O. rufipogon* in 26 kb region of DNA sequence spanning 22 loci (Rakshit et al., 2007). Further in all the *Pi54* alleles, transitions were more frequent than transversions. This complies with the common composition of any type of DNA, where transitions have been reported to occur at higher frequencies than transversions (Brown et al., 1982; Gojobori et al., 1982; Curtis and Clegg, 1984; Wakeley, 1996). This is in consistent with the earlier genome wide SNP discovery studies in multiple rice genotypes (Huang et al., 2009; McNally et al., 2009; Yamamoto et al.,

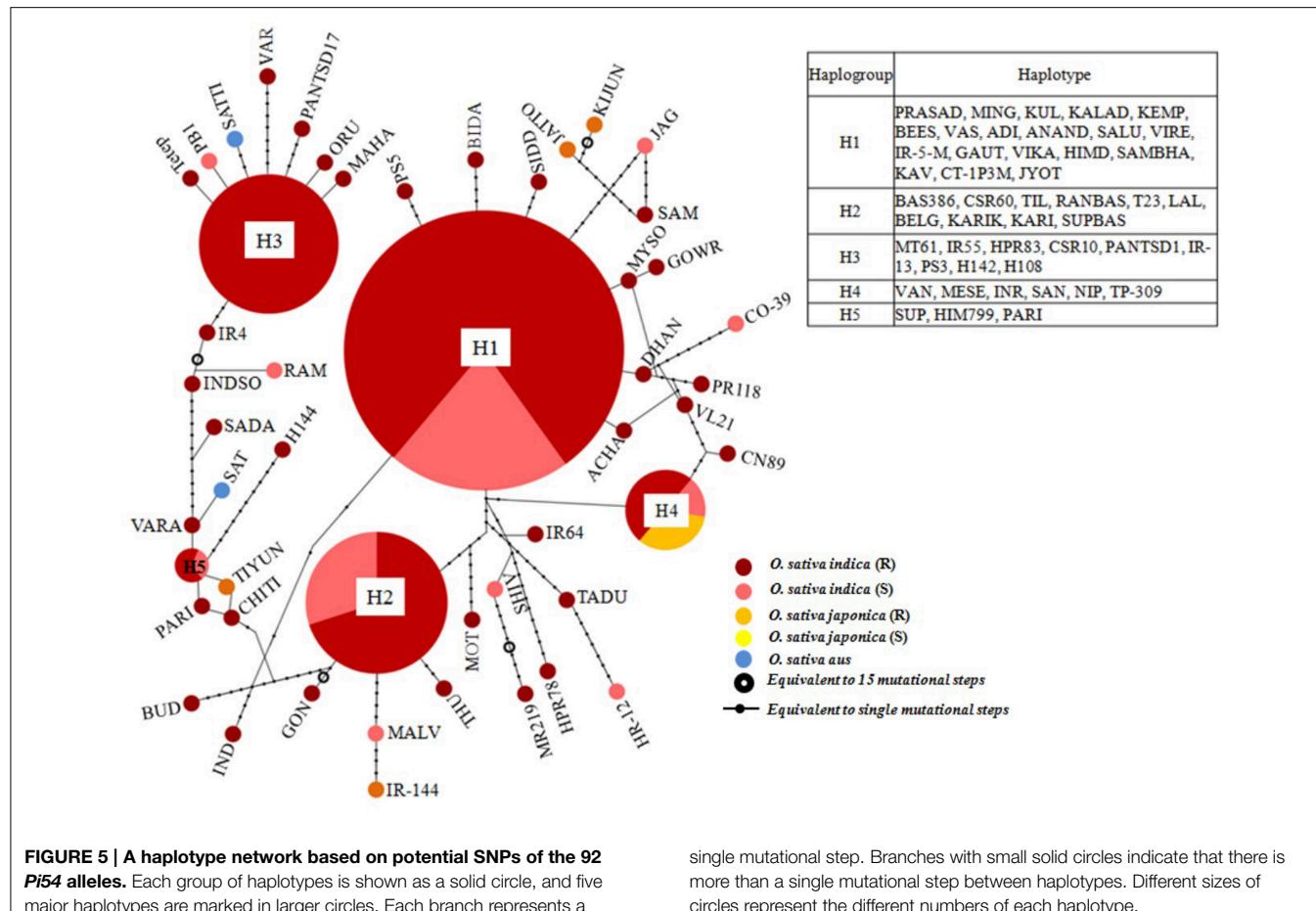


FIGURE 5 | A haplotype network based on potential SNPs of the 92 *Pi54* alleles. Each group of haplotypes is shown as a solid circle, and five major haplotypes are marked in larger circles. Each branch represents a

single mutational step. Branches with small solid circles indicate that there is more than a single mutational step between haplotypes. Different sizes of circles represent the different numbers of each haplotype.

TABLE 3 | Summary statistics of haplotype diversity in *Pi54* alleles.

Group	Hn	Hd	V(Hd)/SD
All	50	0.935	0.00027/0.016
Disease resistance phenotypes	41	0.935	0.00036/0.019
Disease susceptible phenotypes	14	0.958	0.00079/0.028

Hn, Number of Haplotypes; Hd, Haplotype diversity; V(Hd), Variance of Haplotype diversity; SD, Standard deviation of Haplotype diversity.

2010; Thakur et al., 2014). Relatively higher (70%) frequency of transition substitutions between *indica* and *japonica* was observed in earlier studies (Feltus et al., 2004; Shen et al., 2004; International Rice Genome Sequencing Project, 2005). The present study indicates that *Pi54* alleles belong to type II category (intermediate diversified), similar to other blast resistance gene *Pi* 9 (Yang et al., 2008; Liu et al., 2011). It is increasingly believed that percent polymorphism is directly correlated to evolutionary change (Shen et al., 2006; Yang et al., 2008). Our results suggest that intermediate level of polymorphism in the *Pi54* alleles may be due to the mixed evolutionary pressure experienced by the loci during co-evolution of rice blast pathogen. Since this gene has not been transferred to cultivated varieties and might have less pressure from pathogen side.

In present study, 198 polymorphic sites were identified in the 1.5 Kb region of *Pi54* alleles. The nucleotide diversity was high in the *Pi54* alleles of resistant lines compared to the susceptible alleles. Previously, it was reported that *R*-genes experience both high and low levels of sequence diversity depending upon the locus (Yang et al., 2008). Nucleotide diversity (0.024) in *A. thaliana* was higher than the average π (0.008) in 334 randomly distributed genomic regions due to nucleotide difference between resistant and susceptible alleles indicating that these alleles have been maintained for long period of time under natural conditions (Schmid et al., 2005). In barley, the frequency and distribution of the nucleotide diversity ranged from 0.0021 to 0.0189 for the genes associated with grain germination (Russell et al., 2004). In another study, the pattern of diversity observed was lowest in cultivated species as compared to other *Oryza* species (Lee et al., 2009). However, the variation at the flanking regions of the *Pi-ta* gene was highest in *O. rufipogon* (0.00355) followed by cultivated species of *Oryza* (Lee et al., 2011). Studies of many *R* gene loci, such as *Rpp5* in *Arabidopsis thaliana* and *Rp1* in *Zea mays* (Noel et al., 1999; Sun et al., 2001) have revealed a high level of polymorphism between the alleles. In resistance landraces, the expression of *Os11g0225100* locus was higher compared to susceptible. Even after inoculation, the resistance level in the resistant landrace increased, while it has no change

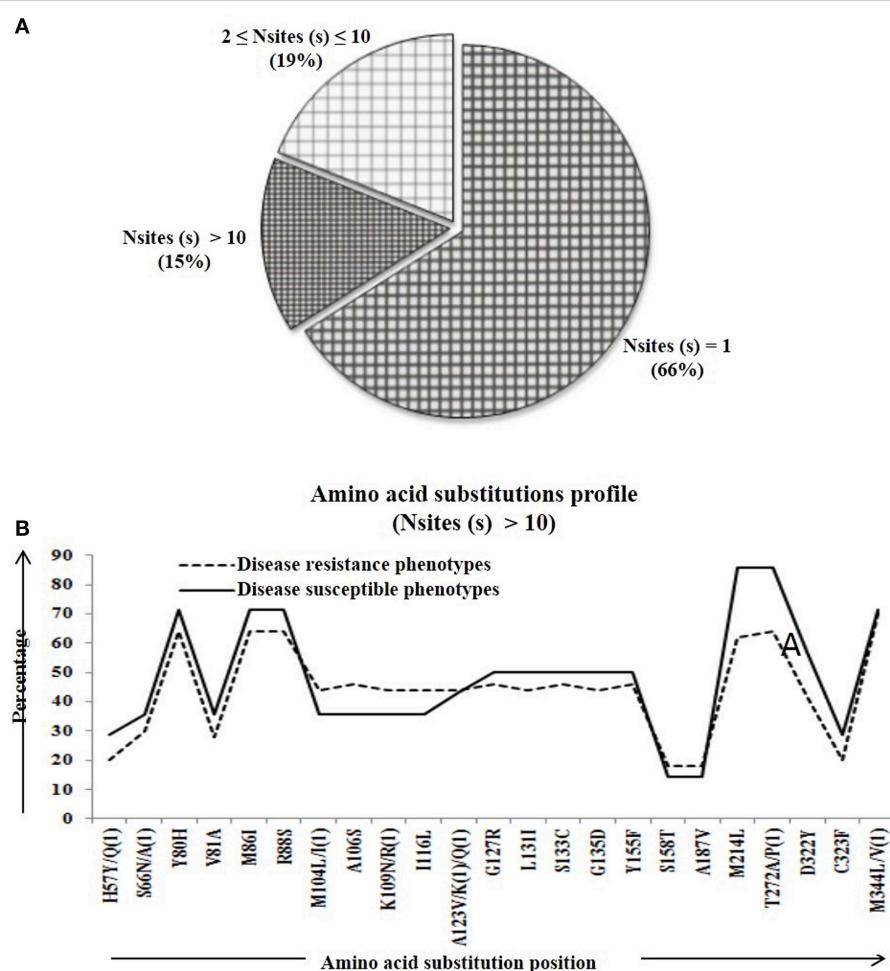


FIGURE 6 | The number of amino acid substitutions in the predicted 64 *Pi54* protein. **(A)** Distribution of amino acid substitutions across 64 predicted proteins. **(B)** Substitutional profile of amino acids of disease resistant and susceptible phenotype.

in the susceptible landrace. This high diversity is interpreted as evidence for the fast evolution of these *R* gene loci.

Phylogenetic analyses of the *Pi54* alleles from susceptible lines were grouped into two major clusters, while all the resistant *Pi54* orthologs were grouped into three clusters. In all the clusters and sub-clusters mixed type of grouping was obtained. However, landraces are found in all the sub-clusters corroborating the claim of higher variability and heterogeneity. This is in contrast to *Pi9* alleles wherein cultivated rice along with its ancestors clustered into one group and African cultivated rice along with its ancestors grouped into separate cluster suggesting that different selection pressure has occurred in two groups during domestication and/or natural selection (Liu et al., 2011).

Important motifs identified in the translated sequences of the *Pi54* alleles, i.e., N-glycosylation, phosphorylation, tyrosine sulfation, N-myristoylation, and leucine zipper. These sites were present in variable numbers in all the *Pi54* alleles except for tyrosine kinase phosphorylation site and leucine zipper. The Zn-finger domain protein is reported responsive to wounding,

stress hormones, cold, salt, submergence, heavy metals and desiccation (Vij and Tyagi, 2008). The presence of motif sites in the translated sequence indicate its role in downstream signaling of defense response genes (callose, laccase, PAL, and peroxidase), transcription factors (NAC6, Dof Zinc finger, MAD box, bZIP, and WRKY) that fortify cell wall/plasmodesmata leading to hypersensitive response, and affecting resistance reaction (Gupta et al., 2011).

In the study, fifty new haplotypes were identified from the nucleotide polymorphism of the *Pi54* alleles. Interestingly, we identified one haplotype which is resistant specific (H_3). Small number of haplotypes was detected previously within the gene pool of cultivated (*H. vulgare*) barley. In *Hordeum* species, the total number of haplotypes identified (46) in *H. spontaneum* almost double from those detected in *H. vulgare* (Russell et al., 2004). Similarly, 16 haplotypes were identified from nucleotide polymorphism of fiftyone *Pi-ta* alleles (Huang et al., 2008). In another study, 53 *Pi-ta* haplotypes were identified from the nucleotide polymorphism of 229 rice accessions belonging to

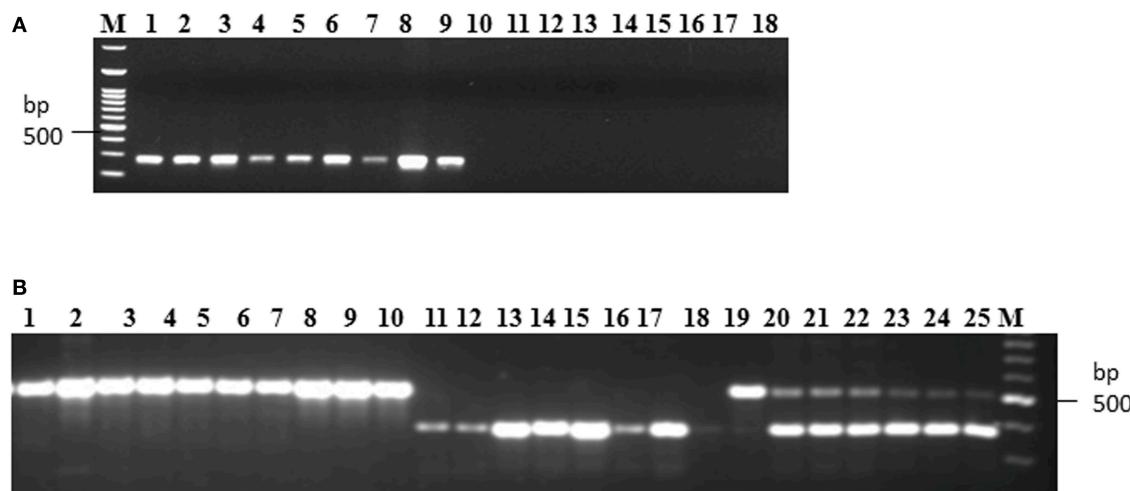


FIGURE 7 | PCR amplification of rice lines with the dominant and co-dominant PCR marker. (A) PCR amplification with the dominant marker (D163_F and D163_R) in 18 rice lines. Lanes 1–9: presence of band in rice lines (Acharmati, Aditya, Ananda, Beesginsali, Bidarlocal-2, Dhanaprasad, Gautam, Himdhan, and Pusa sugandh 5); lanes 10–18: absence of band in rice lines (HR-12, TP-309, Co-39, Nipponbare, Shiva, Satti, Pusa basmati-1, Ram Jawain-100, and Sathia-2) **(B)** PCR amplification with the co-dominant

marker (C144_F and C144_R) in 26 rice lines. bp- base pair; M- 100 bp molecular marker. Lanes 1–10: presence of susceptible band (CO-39, HR-12, CT-10006-7-2M-5-1P3M, Jyoti, Kavali kannu, Mesebatta, Nipponbare, Samba mahsuri, Shiva, Taipei-309); lanes 11–18: presence of resistant band (Tetep, Budda, Chiti zhini, CSR 10, Gonrra bhog, Himalya 799, Pusa Sugandh 3, Ranbir basmati, Sadabahar); lanes 19–25: presence of both the bands (Belgaum basmati, CSR-60, HLR-142, Indrayani, IRAT-144, Kariya).

TABLE 4 | List of rice accessions distinguishable based on dominant marker.

Homozygous RR	Homozygous rr
Acharmati	Pusa sugandh 5
Aditya	Salumpikit
Ananda	Samleshwari
Beesginsali	Siddasala
Bidarlocal-2	V L Dhan 21
Dhanaprasad	Vasanesanna batta
Gautam	Vikash
Himdhan	Virendra
IRBB-4	Ram Jawain-100
IRBLB-BIRBLB-5-M	Sathia-2
Kala Dhan	Lalan kanda
Kulanji pille	Superbasmati
Mingola	Parijat
Mysore mallige	Malviya dhan
	T23

TABLE 5 | List of rice accessions distinguishable based on co-dominant marker.

Homozygous (RR)	Homozygous (rr)	Heterozygous (Rr)
Budda	Mahamaya	CO-39
Chiti zhini	MTU-1061	CT-10006-7-2M-5-1P3M
CSR 10	Pant sankar dhan 1	HR-12
Gonrra bhog	Pant sugandh dhan 17	Jyoti
Himalya 799	Tetep	Kavali kannu
HLR-108	Pusa Sugandh 3	Mesebatta
HLR-144	Ranbir basmati	Nipponbare
HPR 2083	Sadabahar	Samba mahsuri
HPR-2178	Suphala	Shiva
Indira sona	Thule atte	Taipei-309
IRBB 55	Tilak chandan	
IRBB-13	Tiyun	
IRBB-4	Varalu	
Kari kantiga	Varun dhan	

seven *Oryza* species. These findings highlighted the importance of analysis and utilization of haplotypes from landraces and related wild species for crop improvement (Lee et al., 2009).

Balancing and purifying selection have been observed for the evolution of *R*-genes. The value of Tajima's D was negative in the entire coding region (-1.03898) and LRR domain (-1.70363) indicating purifying selection in the CDS and LRR regions. Similar, values were also reported for *Pi-ta* alleles of *O. rufipogon* and *Pi9* alleles of five *Oryza* species (AA genome) (Lee et al., 2009;

Liu et al., 2011). Our analyses implicate selection pressure high on both CDS and LRR domains of resistant lines, in contrast to the LRR domain of susceptible rice lines. Elucidation of the selection mechanism acting on the domains could help to design new crop improvement strategies for the future.

In the LRR region of the *Pi54* alleles of different groups, the value of $k_a/k_s > 1.0$ implying positive directional selection might have favored amino acid substitution in the region. Similar results were also obtained in *Piz(t)* alleles of Indian landraces

of diverse locations (Thakur et al., 2013b). In contrast, the LRR regions encoded by *Pi-km1* and *Pi-km2* blast resistance genes were highly conserved (Ashikawa et al., 2010). Importantly, LRRs have direct interacting roles with effector proteins (Young and Innes, 2006). Most isolated *R*-genes encode proteins possessing LRR domain, of which the majority also contains a NBS domain (Sharma et al., 2012). Higher level of polymorphism in the LRR region as obtained in case of *Pi54* gene is thought to be involved in the recognition of effector proteins, and consequently the evolutionary pressure on the host by virulent *M. oryzae* races results in high variability in LRR domain. The LRR regions of many *Arabidopsis* *R* genes have k_a/k_s ratios >1 , suggesting that these *R* genes have evolved under positive selection pressure (Bergelson et al., 2001). Two basic strategies have evolved for an *R* protein to recognize a pathogen effector (which is also called avirulence (Avr) factor): direct physical interaction and indirect interaction via association with other host proteins targeted by the Avr factor (Xiao et al., 2008). It has also been reported that variation for disease resistance is maintained by frequency-dependent selection, even though there is a fitness cost associated with the maintenance of *R* genes in the absence of their matching Avr (Stahl et al., 1999). Flax *L* genes and their matching Avr genes in flax rust undergo strong diversifying selection, suggesting direct interaction (Dodds et al., 2006). However, in case of *A. thaliana* indirect interaction between *RPM1* and *AvrB*, in which the RIN4 protein acts as a target for binding to *AvrB*, and the *AvrB*-induced phosphorylation of RIN4 then activates *RPM1* has been reported as balancing selection (Mackey et al., 2002).

Development of allele specific functional markers holds the key for marker assisted selection. These functional markers based on genic *Pi54* InDels can be applicable in MAS for blast resistance breeding programme. Absence of diagnostic bands in resistant/susceptible cultivars might be because of genetic recombination during meiotic cell cycle. However, the present study extends the repertoire of functional markers toward screening of genotypes. Similar functional InDel-based marker has been reported to be developed for blast resistance gene *Pikm* (Costanzo and Jia, 2010). The sequence of nine blast resistance genes was used for the development of functional markers based on InDels (Hayashi et al., 2006). However, allele specific markers are not known for blast resistance gene *Pi54*, hence the markers developed in this study would be of great significance to the breeders. Increasing efforts to clone more resistance genes worldwide will accelerate the development of more dominant resistance gene based markers for molecular breeding, thereby accelerating introduction of durable, broad-spectrum blast resistant genes into widely adapted high yielding rice cultivars.

From the above discussion, we conclude that the nucleotide variation was high in the LRR domain of all the *Pi54* alleles cloned and characterized in this study. In disease resistant alleles, selection pressure was high in LRR domain and CDS region whereas in susceptible counterpart selection pressure exerted

only in the LRR domain. It was also evident that LRR domain of *Pi54* alleles was diversified because of high selection pressure. The co-dominant and dominant functional markers developed in the present study can be used in marker-assisted breeding programs aimed at improvement of blast resistance in elite rice cultivars. The diversity information based on genetic structure is an extremely important pre-breeding material in selecting parents for intra- and inter- group crosses to broaden the genetic base of modern rice cultivars. This study helps understand the extent of variability present in the landraces and cultivated varieties of rice that can be employed in future for selection of better alleles and their utilization in resistant breeding programmes.

Author Contributions

Conceived and designed the experiments: TRS. Performed the experiments and analyzed data: ST. Contributed materials/analysis tools: RR, MV, SP, DC, AD, NS, AS, UD, and PS. Wrote the paper: TRS, ST, and AD.

Acknowledgments

TRS is thankful to the National Agricultural Innovation Project (NAIP) (C4/C1071), ICAR, for financial support. TRS is thankful to the Department of Science and Technology, Govt. of India for JC Bose National Fellowship. The authors are thankful to the Officer in Charge, National Phytotron Facility, Indian Agricultural Research Institute, New Delhi, for providing basic facilities for growing and maintaining Indian local landraces.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.00345/abstract>

Figure S1 | Haplotype network based on 187 potential SNPs of the *Pi54* resistant alleles. Each group of haplotypes is shown as a solid circle, and five major haplotypes are marked in larger circles. Each branch represents a single mutational step. Branches with small solid circles indicate that there is more than a single mutational step between haplotypes. Different sizes of circles represent the different numbers of each haplotype.

Figure S2 | Haplotype network based on 97 potential SNPs of the *Pi54* susceptible alleles. Each group of haplotypes is shown as a solid circle, and four major haplotypes are marked in larger circles. Each branch represents a single mutational step. Branches with small solid circles indicate that there is more than a single mutational step between haplotypes. Different sizes of circles represent the different numbers of each haplotype.

Figure S3 | Analysis of ZnF domain of 61 *Pi54* proteins. Arrows indicate that the ZnF domain was not predicted inspite of the sequence of the ZnF in the *Pi54* proteins. Single amino acid substitution in two *Pi54* proteins has been encircled.

Figure S4 | Phylogenetic tree of LRR domain of 64 *Pi54* protein. Neighbor-Joining method was used for the construction of tree. Definite clustering was not obtained in all the trees.

References

Ashikawa, I., Wu, J., Matsumoto, T., and Ishikawa, R. (2010). Haplotype diversity and molecular evolution of the rice *Pikm* locus for blast resistance. *J. Gen. Plant Pathol.* 76, 37–42. doi: 10.1007/s10327-009-0213-x

Bergelson, J., Kreitman, M., Stahl, E. A., and Tian, D. (2001). Evolutionary dynamics of plant *R*-genes. *Science* 292, 2281–2285. doi: 10.1126/science.1061337

Bonmann, J. M., Dedios, T. V., and Khin, M. M. (1986). Physiologic specialization 718 of *Pyricularia oryzae* in Philippines. *Plant Dis.* 70, 767–769. doi: 10.1094/PD-70-767

Brown, W. M., Prager, E. M., Wang, A., and Wilson, A. C. (1982). Mitochondrial DNA sequences of primates, tempo and mode of evolution. *J. Mol. Evol.* 18, 225–239. doi: 10.1007/BF01734101

Clement, M., Posada, D., and Crubéziel, K. (2000). TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9, 1657–1660. doi: 10.1046/j.1365-294x.2000.01020.x

Costanzo, S., and Jia, Y. (2010). Sequence variation at the rice blast resistance gene *Pi-km* locus: implications for the development of allele specific markers. *Plant Sci.* 178, 523–530. doi: 10.1016/j.plantsci.2010.02.014

Curtis, S. E., and Clegg, M. T. (1984). Molecular evolution of chloroplast DNA sequences. *Mol. Biol. Evol.* 1, 291–301.

Das, A., Soubam, D., Singh, P. K., Thakur, S., Singh, N. K., and Sharma, T. R. (2012). A novel blast resistance gene, *Pi54rh* cloned from wild species of rice, *Oryza rhizomatis* confers broad spectrum resistance to *Magnaporthe oryzae*. *Funct. Integr. Genomics* 12, 215–228. doi: 10.1007/s10142-012-0284-1

Dodds, P. N., Lawrence, G. J., Catanzariti, A., Teh, T., Wang, C., Ayliffe, M. A., et al. (2006). Direct protein interaction underlies gene-for- gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8888–8893. doi: 10.1073/pnas.0602577103

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using Phred II error probabilities. *Genome Res.* 8, 186–194.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.2307/2408678

Feltus, A., Wan, J., Schulze, S. R., Estill, J. C., Jiang, N., and Paterson, A. H. (2004). An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. *Genome Res.* 14, 1812–1819. doi: 10.1101/gr.247940

Geng, X. S., Yang, M. Z., Huang, X. Q., Cheng, Z. Q., Fu, J., Sun, T., et al. (2008). Cloning and analyzing of rice blast resistance gene *Pi-ta*⁺ allele from Jinghong erect type of common wild rice (*Oryza rufipogon* Griff) in Yunnan. *Yi Chuan* 30, 109–114. doi: 10.3724/SP.J.1005.2008.00109

Gojobori, T., Li, W. H., and Graur, D. (1982). Patterns of nucleotide substitution in pseudogenes and functional genes. *J. Mol. Evol.* 18, 360–369. doi: 10.1007/BF01733904

Gupta, S. K., Rai, A. K., Kanwar, S. S., Chand, D., Singh, N. K., and Sharma, T. R. (2011). The single functional blast resistance gene *Pi54* activates a complex defense mechanism in rice. *J. Exp. Bot.* 63, 757–772. doi: 10.1093/jxb/err297

Gupta, S. K., Rai, A. K., Kanwar, S. S., and Sharma, T. R. (2012). Comparative analysis of zinc finger proteins involved in plant disease resistance. *PLoS ONE* 7:e42578. doi: 10.1371/journal.pone.0042578

Hayashi, K., Hashimoto, N., Daigen, M., and Ashikawa, I. (2004). Development of PCR-based SNP markers for rice blast resistance genes at the *Piz* locus. *Theor. Appl. Genet.* 108, 1212–1220. doi: 10.1007/s00122-003-1553-0

Hayashi, K., Yoshida, H., and Ashikawa, I. (2006). Development of PCR-based allele-specific and InDel marker sets for nine rice blast resistance genes. *Theor. Appl. Genet.* 113, 251–260. doi: 10.1007/s00122-006-0290-6

Huang, C. L., Hwang, S. Y., Chiang, Y. C., and Lin, T. P. (2008). Molecular evolution of the *Pi-ta* gene resistant to rice blast in wild rice (*Oryza rufipogon*). *Genetics* 179, 1527–1538. doi: 10.1534/genetics.108.089805

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

International Rice Genome Sequencing Project. (2005). The map-based sequence of the rice genome. *Nature* 436, 793–800. doi: 10.1038/nature03895

Jia, Y., Liu, G., Costanzo, S., Lee, S., and Dai, Y. (2009). Current progress on genetic interactions of rice with rice blast and sheath blight fungi. *Front. Agric. China* 3:6. doi: 10.1007/s11703-009-0062-6

Jia, Y., and Martin, R. (2008). Identification of a new locus *Ptr(t)* required for rice blast resistance gene *Pi-ta*-mediated resistance. *Mol. Plant Microbe Interact.* 21, 396–403. doi: 10.1094/MPMI-21-4-0396

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Lee, S., Costanzo, S., Jia, Y., Olsen, K. M., and Caicedo, A. L. (2009). Evolutionary dynamics of the genomic region around the blast resistance gene *Pi-ta* in AA genome *Oryza* species. *Genetics* 183, 1315–1325. doi: 10.1534/genetics.109.108266

Lee, S., Jia, Y., Jia, M., Gealy, D. R., and Olsen, K. M. (2011). Molecular evolution of the rice blast resistance gene *Pi-ta* in invasive weedy rice in the USA. *PLoS ONE* 6:e26260. doi: 10.1371/journal.pone.0026260

Liu, J., Hu, Y., Ning, Y., Jiang, N., Wu, J., Xiao, Y., et al. (2011). Genetic variation and evolution of the *Pi9* blast resistance locus in the AA genome *Oryza* species. *J. Plant Biol.* 54, 294–302. doi: 10.1007/s12374-011-9166-7

Mackey, D., Holt, B. F., Wiig, A., and Dangl, J. L. (2002). RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell* 108, 743–754. doi: 10.1016/S0092-8674(02)00661-X

McCouch, S. (2013). Feeding the future. *Nature* 499, 23–24. doi: 10.1038/499023a

McNally, K. L., Childs, K. L., Bohert, R., Davidson, R. M., Zhao, K., Ulat, B. J., et al. (2009). Genome-wide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12273–12278. doi: 10.1073/pnas.0900992106

Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucl. Acids. Res.* 8, 4321–4325. doi: 10.1093/nar/8.19.4321

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York, NY: Columbia University Press.

Noel, L., Moores, T. L., van der Biezen, E. A., Parniske, M., Daniels, M. J., Parker, J. E., et al. (1999). Pronounced intra specific haplotype divergence at the *RPP5* complex disease resistance locus of *Arabidopsis*. *Plant Cell* 11, 2099–2111. doi: 10.1101/tpc.11.11.2099

Rai, A. K., Kumar, S. P., Gupta, S. K., Gautam, N., Singh, N. K., and Sharma, T. R. (2011). Functional complementation of rice blast resistance gene *Pi-k^h* (*Pi54*) conferring resistance to diverse strains of *Magnaporthe oryzae*. *J. Plant Biochem. Biotech.* 20, 55–65. doi: 10.1007/s13562-010-0026-1

Rakshit, S., Rakshit, A., Matsumura, H., Takahashi, Y., Hasegawa, Y., Ito, A., et al. (2007). Large scale DNA polymorphism study of *Oryza sativa* and *O. rufipogon* reveals the origin and divergence of Asian rice. *Theor. Appl. Genet.* 114, 731–743. doi: 10.1007/s00122-006-0473-1

Ram, S. G., Thiruvengadam, V., and Vinod, K. K. (2007). Genetic diversity among cultivars, land races and wild relatives of rice as revealed by microsatellite markers. *J. Appl. Genet.* 48, 337–345. doi: 10.1007/BF03195230

Rozas, J., Sánchez-DelBarrio, J. C., Meseguer, X., and Rozas, R. (2003). DnaSP: DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19, 2496–2497. doi: 10.1093/bioinformatics/btg359

Rozen, S., and Skaletsky, H. (2000). Primer 3 on the WWW for general users for biologist programmers. *Methods Mol. Biol.* 132, 365–386.

Russell, J., Booth, A., Fuller, J., Harrower, B., Hedley, P., Machray, G., et al. (2004). A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. *Genome* 47, 389–398. doi: 10.1139/g03-125

Saitou, N., and Nei, M. (1987). The neighbor-joining method: anew method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.

Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., et al. (2005). A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* 37, 501–506. doi: 10.1038/ng1543

Shang, J., Tao, Y., Chen, X., Zou, Y., Lei, C., Wang, J., et al. (2009). Identification of a new rice blast Resistance gene, *Pid3*, by genome wide comparison of paired Nucleotide-Binding Site-Leucine-Rich Repeat genes and their pseudogene alleles between the two sequenced rice genomes. *Genetics* 182, 1303–1311. doi: 10.1534/genetics.109.102871

Sharma, T. R., Das, A., Thakur, S., and Jalali, S. (2014). Recent understanding on structure, function and evolution of plant disease resistance genes. *Proc. Indian Nat. Sci. Acad.* 1, 83–93. doi: 10.16943/ptinsa/2014/v80i1/55088

Sharma, T. R., Madhav, M. S., Singh, B. K., Shanker, P., Jana, T. K., Dalal, V., et al. (2005a). High-resolution mapping, cloning, and molecular characterization of

the *Pi-k^h* gene of rice, which confers resistance to *Magnaporthe grisea*. *Mol. Genet. Genomics* 274, 569–578. doi: 10.1007/s00438-005-0035-2

Sharma, T. R., Rai, A. K., Gupta, S. K., and Singh, N. K. (2010). Broad-spectrum Blast Resistance gene *Pi-k^h* cloned from rice line Tetep designated *Pi54*. *J. Plant Biochem. Biotechnol.* 191, 87–89. doi: 10.1007/BF03323441

Sharma, T. R., Rai, A. K., Gupta, S. K., Vijayan, J., Devanna, B. N., and Ray, S. (2012). Rice blast management through host-plant resistance: retrospect and prospects. *Agri. Res.* 1, 37–52. doi: 10.1007/s40003-011-0003-5

Sharma, T. R., Shanker, P., Singh, B. K., Jana, T. K., Madhav, M. S., Gaikwad, K., et al. (2005b). Molecular mapping of rice blast resistance gene *Pi-k^h* in rice variety Tetep. *J. Plant Biochem. Biotechnol.* 14, 127–133. doi: 10.1007/BF03263240

Shen, J. H., Araki, L., Chen, J. Q., and Chen Tian, D. (2006). Unique evolutionary mechanism in *R*-genes under the presence/absence polymorphism in *Arabidopsis thaliana*. *Genetics* 172, 1243–1250. doi: 10.1534/genetics.105.047290

Shen, Y. J., Jiang, H., Jin, J. P., Zhang, Z. B., Xi, B., He, Y. Y., et al. (2004). Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiol.* 135, 1198–1205. doi: 10.1104/pp.103.038463

Singh, A. K., Gopalakrishnan, S., Singh, V. P., Prabhu, K. V., Mohapatra, T., Singh, N. K., et al. (2011). Marker assisted selection: a paradigm shift in Basmati breeding. *Indian J. Genet.* 71, 1–9.

Solovyev, V., Kosarev, P., Seledsov, I., and Vorobyev, D. (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* 7(Suppl. 1), S10. doi: 10.1186/gb-2006-7-s1-s10

Stahl, E. A., Dwyer, G., Mauricio, R., Kreitman, M., and Beregelson, J. (1999). Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* 400, 667–671. doi: 10.1038/23260

Sun, Q., Collins, N. C., Ayliffe, M., Smith, S. M., Drake, J., Pryor, T., et al. (2001). Recombination between paralogues at the *Rp1* rust resistance locus in maize. *Genetics* 158, 423–438.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599. doi: 10.1093/molbev/msm092

Tamura, K., Nei, M., and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11030–11035. doi: 10.1073/pnas.0404206101

Templeton, A. R., Crandall, A. K., and Sing, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data III Cladogram estimation. *Genetics* 132, 619–633.

Thakur, S., Gupta, Y. K., Singh, P. K., Rathour, R., Variar, M., Prashanthi, S. K., et al. (2013a). Molecular diversity in rice blast resistance gene *Pi-ta* makes it highly effective against dynamic population of *Magnaporthe oryzae*. *Funct. Integr. Genomics* 13, 309–322. doi: 10.1007/s10142-013-0325-4

Thakur, S., Singh, P. K., Rathour, R., Variar, M., Prashanthi, S. K., Gopalakrishnan, S., et al. (2014). Genotyping and development of single-nucleotide polymorphism (SNP) markers associated with blast resistance genes in rice using GoldenGate assay. *Mol. breeding* 34, 1449–1463. doi: 10.1007/s11032-014-0129-9

Thakur, S., Singh, P. K., Rathour, R., Variar, M., Prashanthi, S. K., Singh, A. K., et al. (2013b). Positive selection pressure on rice blast resistance allele *Pi-ta* makes it divergent in Indian land races. *J. Plant Interact.* 8, 34–44. doi: 10.1080/17429145.2012.721523

Vij, S., and Tyagi, A. K. (2008). A20/AN1 zinc-finger domain-containing proteins in plants and animals represent common elements in stress response. *Funct. Integr. Genomics* 8, 301–307. doi: 10.1007/s10142-008-0078-7

Wakeley, J. (1996). The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Tree* 11, 158–163. doi: 10.1016/0169-5347(96)10009-4

Wang, C., Yang, Y., Yuan, X., Xu, Q., Feng, Y., Yu, H., et al. (2014). Genome-wide association study of blast resistance in indica rice. *BMC Plant Biol.* 14:6. doi: 10.1186/s12870-014-0311-6

Wang, X., Jia, Y., Shu, Q. Y., and Wu, D. (2008). Haplotype diversity at the *Pi-ta* locus in cultivated rice and its wild relatives. *Phytopathology* 98, 1305–1311. doi: 10.1094/PHYTO-98-12-1305

Xiao, S., Wang, W., and Yang, X. (2008). “Evolution of resistance genes in plants,” in *Innate Immunity of Plants, Animals and Human, 1st Edn.*, ed H. Heine (Berlin; Hidelberg: Springer), 1–25.

Yamamoto, T., Nagasaki, H., Yonemaru, J., Ebana, K., Nakajima, M., Shibaya, T., et al. (2010). Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. *BMC Genomics* 11:267. doi: 10.1186/1471-2164-11-267

Yang, M. Z., Cheng, Z. Q., Chen, S. N., Qian, J., Xu, L. L., and Huang, X. Q. (2007). A rice blast resistance genetic resource from wild rice in Yunnan, China. *J. Plant Physiol. Mol. Biol.* 33, 589–595.

Yang, S., Gu, T., Pan, C., Feng, Z., Ding, J., Hang, Y., et al. (2008). Genetic variation of NBS-LRR class resistance genes in rice lines. *Theor. Appl. Genet.* 16, 165–177. doi: 10.1007/s00122-007-0656-4

Yoshida, K., and Miyashita, N. T. (2009). DNA polymorphism in the blast disease resistance gene *Pi-ta* of the wild rice *Oryza rufipogon* and its related species. *Genes Genet. Syst.* 84, 121–136. doi: 10.1266/ggs.84.121

Young, B. J. D., and Innes, R. W. (2006). Plant NBS-LRR proteins in pathogen sensing and host defense. *Nat. Immunol.* 7, 1243–1249. doi: 10.1038/ni1410

Yu, P., Wang, C., Xu, Q., Feng, Y., Yuan, X., Yu, H., et al. (2011). Detection of copy number variations in rice using array-based comparative genomic hybridization. *BMC Genomics* 12:372. doi: 10.1186/1471-2164-12-372

Zhai, C., Lin, F., Dong, Z., He, X., Yuan, B., Zeng, X., et al. (2011). The isolation and characterization of *Pik*, a rice blast resistance gene which emerged after rice domestication. *New Phytol.* 189, 321–334. doi: 10.1111/j.1469-8137.2010.03462.x

Conflict of Interest Statement: 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.

Copyright © 2015 Thakur, Singh, Das, Rathour, Variar, Prashanthi, Singh, Singh, Chand, Singh 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) or licensor 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.

Analysis of genetic variation and diversity of Rice stripe virus populations through high-throughput sequencing

OPEN ACCESS

Edited by:

Joanna Marie-France Cross,
Inonu, Turkey

Reviewed by:

Abdelaleim Ismail ElSayed,
Zagazig University, Egypt
Angelantonio Minafra,
CNR Institute for Sustainable Plant
Protection, Italy
Shahin S. Ali,
United States Department of
Agriculture, USA

*Correspondence:

Xueping Zhou,
State Key Laboratory of Rice Biology,
Institute of Biotechnology, Zhejiang
University, Hangzhou 310058,
People's Republic of China
zzhou@zju.edu.cn;
Rongxiang Fang,
State Key Laboratory of Plant
Genomics, Institute of Microbiology,
Chinese Academy of Sciences,
Beijing 100101, People's Republic of
China
fangrx@im.ac.cn

[†]These authors have contributed
equally to this work.

Specialty section:

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

Received: 31 October 2014

Accepted: 05 March 2015

Published: 24 March 2015

Citation:

Huang L, Li Z, Wu J, Xu Y, Yang X,
Fan L, Fang R and Zhou X (2015)
Analysis of genetic variation and
diversity of Rice stripe virus
populations through high-throughput
sequencing.
Front. Plant Sci. 6:176.
doi: 10.3389/fpls.2015.00176

Lingzhe Huang^{1†}, Zefeng Li^{2†}, Jianxiang Wu¹, Yi Xu¹, Xiuling Yang¹, Longjiang Fan²,
Rongxiang Fang^{3*} and Xueping Zhou^{1*}

¹ State Key Laboratory of Rice Biology, Institute of Biotechnology, Zhejiang University, Hangzhou, People's Republic of China,

² Institute of Crop Science and Institute of Bioinformatics, Zhejiang University, Hangzhou, People's Republic of China, ³ State
Key Laboratory of Plant Genomics, Institute of Microbiology, Chinese Academy of Sciences, Beijing, People's Republic of
China

Plant RNA viruses often generate diverse populations in their host plants through error-prone replication and recombination. Recent studies on the genetic diversity of plant RNA viruses in various host plants have provided valuable information about RNA virus evolution and emergence of new diseases caused by RNA viruses. We analyzed and compared the genetic diversity of Rice stripe virus (RSV) populations in *Oryza sativa* (a natural host of RSV) and compared it with that of the RSV populations generated in an infection of *Nicotiana benthamiana*, an experimental host of RSV, using the high-throughput sequencing technology. From infected *O. sativa* and *N. benthamiana* plants, a total of 341 and 1675 site substitutions were identified in the RSV genome, respectively, and the average substitution ratio in these sites was 1.47 and 7.05 %, respectively, indicating that the RSV populations from infected *N. benthamiana* plant are more diverse than those from infected *O. sativa* plant. Our result gives a direct evidence that virus might allow higher genetic diversity for host adaptation.

Keywords: Rice stripe virus, population, diversity, deep sequencing

Introduction

RNA viruses are known for their high evolutionary potential due mainly to their error-prone replication, and fast rate of genome replication (Domingo and Holland, 1997; Roossinck, 1997; Elena and Sanjuán, 2007; Lauring and Andino, 2010; Sanjuán et al., 2010). In addition, template-switching of viral RNA-dependent RNA polymerase (RdRp) during virus replication, known as RNA recombination, also functions to accelerate virus evolution (Nagy and Simon, 1997; Garcia-Arenal et al., 2001). Using these strategies, viruses are capable of bringing beneficial mutations into and removing deleterious mutations from their genomes (Worobey and Holmes, 1999). Mutation in the viral genomic sequence may lead to changes of virus host range, disease symptoms, and emergence of new viruses in nature. Previous reports have indicated that mutant viruses often appeared soon after the RNA virus infected its host plant and to form diverse populations in infected cells, and these studies were analyzed through sequencing random selected clones derived from an infected plant (Domingo and Holland, 1997; Roossinck, 1997; Schneider and Roossinck, 2000; Garcia-Arenal et al., 2001, 2003).

Rice stripe virus (RSV) is one of the most damaging rice pathogens in China and was firstly identified in rice field in China in 1963 (Lin et al., 1990). It is now identified in rice fields in 16 provinces of China and caused significant yield losses (Xiong et al., 2008; Wei et al., 2009). RSV is the type member of the genus *Tenuivirus* and infects mainly rice and a few other species in the family *Poaceae* including maize, oat and wheat. In field, RSV is transmitted efficiently by small brown plant hopper (*Laodelphax striatellus*) in a persistent manner. RSV can propagate within the insect vector and transmitted to its progenies (Koganezawa et al., 1975; Falk and Tsai, 1998). Under the laboratory conditions, RSV can be transmitted to an experimental host, *Nicotiana benthamiana*, through mechanical inoculation (Kong et al., 2014).

Rice stripe virus is a single-stranded RNA virus containing four genome segments, designated as RNA 1, 2, 3, and 4 (Figure 1; Zhu et al., 1991, 1992; Takahashi et al., 1993; Toriyama et al., 1994). RNA 1 is in negative sense and encodes a putative protein of approximately 337 kDa. This protein was suggested to be part of the RSV RdRp complex and is associated with the RSV filamentous ribonucleoprotein (Toriyama et al., 1994). The other three RNA segments of RSV are all ambisense and each RNA segment contains two ORFs, one is in the viral sense (vRNA) and the other in the viral complementary-sense (vcRNA; Ramirez and Haenni, 1994). The vRNA 2 ORF encodes a membrane-associated protein of 22.8 kDa and the vcRNA 2 ORF encodes a poly-glycoprotein of 94 kDa. Functions of these proteins are still unknown (Takahashi et al., 1993; Falk and Tsai, 1998). The vRNA 3 ORF encodes a 23.9-kDa suppressor of RNA silencing (Xiong et al., 2009) and the vcRNA 3 ORF encodes a nucleocapsid protein of 35 kDa (Kakutani et al., 1991; Zhu et al., 1991). The vRNA 4 ORF encodes a major non-capsid protein of 21.5 kDa that functions in disease symptom development (Kong et al., 2014). The vcRNA 4 ORF encodes a protein of 32.5 kDa and is required for RSV cell-to-cell movement in infected plant (Xiong et al., 2008).

Determination of the complexity of RNA virus population in their host plants requires a rapid, reliable, and cost-effective sequencing method. Because viral RNAs with specific mutations may accumulate to lower levels than the most frequent viral RNAs

in infected cells, traditional small-scale sequencing used in many earlier studies might fail to detect these non-conserved or lowly accumulated viral RNAs. Although the evolution of RSV over long periods of time and in areas covering large geographical districts in China has been studied (Wei et al., 2009), the population diversity of RSV within a plant is still unknown. In this study, we used the high-throughput sequencing technology (also known as the second generation sequencing technology) to investigate the diverse population of RSV in an infected *Oryza sativa* plant (a natural host) and compare it with that observed in an infected *N. benthamiana* plant (an experimental host) and we found that the RSV populations from the infected *N. benthamiana* plant are more diverse than those from the infected *O. sativa* plant. Our results give a direct evidence that virus might allow higher genetic diversity for host adaptation.

Materials and Methods

Virus Source, Host Plants, and Virus Inoculation

Infected *O. sativa* plants with typical RSV symptoms were collected from rice fields in Jiangsu Province, China. A single plant with only RSV infection was identified through RT-PCR using specific primers and ELISA using specific antibodies as described previously (Wang et al., 2004) and leaves of this plant were then stored at -80°C till use. *N. benthamiana* seedlings were grown inside an insect-free room with a constant temperature at 25°C and a 16 h light supply. Leaves from the RSV-infected *O. sativa* plant were ground in 0.1 M phosphate buffer, pH 7.0, and the crude extract was mechanically inoculated to leaves of *N. benthamiana* seedlings at six- to seven-leaf stages. Young systemic leaves of one *N. benthamiana* plant showing yellow vein symptoms were harvested at 30 days post inoculation (dpi). The harvested fresh *N. benthamiana* leaves and the frozen RSV-infected *O. sativa* leaves were used for the viral population analysis.

Viral cDNA Library Construction and Sequencing

Total RNA was extracted from 1g leaf tissues of the frozen *O. sativa* and fresh *N. benthamiana* using TRIzol reagent following manufacturer's instructions (Invitrogen, Carlsbad, CA, USA), respectively, and cDNA libraries of RSV were then constructed. First strand cDNA was synthesized using the SuperScript III reverse transcriptase (Invitrogen, Carlsbad, CA, USA), primers used for first strand cDNA synthesis were designed according to the highly conserved 3' terminal regions of the four RSV RNAs: 5'-acacaaagtccagggaaaaca-3' for RNA 1, 5'-acacaaagtctgggtaaaacttctt-3' for RNA 2, 5'-acacaaagtccctgggaaaatag-3' for RNA 3, and 5'-acacaaagtccaggcattgt-3' for RNA 4. Second strand cDNAs were synthesized using DNA polymerase I and RNase H. Ends of double strand cDNAs were repaired using T4 DNA polymerase and Klenow DNA polymerase. cDNA pair-end libraries were prepared using standard Illumina protocols and then the libraries were sequenced using Illumina Genome Analyzer II instrument (Illumina, San Diego, CA, USA).

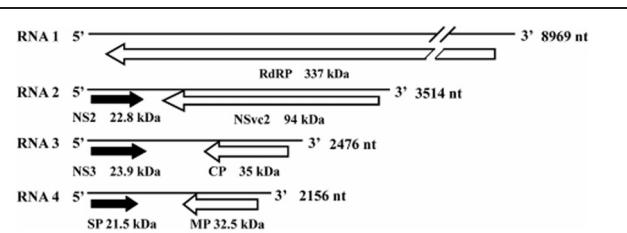


FIGURE 1 | Genome organization of Rice stripe virus (RSV). ORFs are indicated as black arrows on the viral-sense RNAs and white arrows on the viral complementary-sense RNAs. Direction of arrowheads indicates the direction of translation. RdRp, RNA-dependent RNA polymerase; CP, coat protein; SP, disease-specific protein; MP, movement protein; NS2 and NS3, non-structural proteins on the viral-sense RNA 2 and 3; NSvc2, non-structural protein on the viral complementary-sense RNA 2.

Bioinformatics Analysis

To minimize artificial mistakes from sequencing errors, short reads were pre-processed using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) by removing the low quality reads (below Q30) and adaptors. The reads representing each sample were mapped onto the reference sequences of RSV genome segments (NC_003753.1, NC_003754.1, NC_003755.1, and NC_003776.1) using the MAQ program (Mapping and Assembly with Qualities, Version 0.7.1) with default parameters (Li et al., 2008).

Results

High-throughput Sequencing of RSV Genome from Infected *O. sativa* and *N. benthamiana* Plants

The cDNA libraries derived from RSV-infected *O. sativa* plant and that prepared from RSV-infected *N. benthamiana* plant were sequenced, using the Illumina/Solexa platform. The sequences have been submitted to Sequence Read Archive (SRA) with accession number SRP051574. A total of 27,965,556 reads were obtained and 6.67% of them were mapped to the RSV genomic sequence. The average sequencing depths of the RSV genome segments in *O. sativa* were from 953.8 (RNA2) to 2129.1 (RNA4) and those in *N. benthamiana* were from 3260.2 (RNA1) to 7330.5 (RNA3). The number of mapped bases and the sequencing depth of each RSV RNA segment are shown in **Table 1**. Our results indicate that the data we obtained are sufficient to represent the RSV populations in a single infected plant.

Genetic Diversity of RSV Populations in *O. sativa*

With the high-throughput sequencing results, a “two-dimensional view” of RSV populations is drawn. This includes average nucleotide diversity (number of sites with substitution/number of total sites of corresponding sequence) and substitution ratio within each site (number of substituted nucleotide observed/number of total nucleotide sequenced). Substitutions determined in the RSV populations from the infected *O. sativa* plant located at 123 different sites in RNA 1 (1.37%), 22 sites in RNA 2 (0.63%), 76 sites in RNA 3 (3.07%), and 60 sites in RNA 4 (2.78%). Among the four RSV RNAs, substitutions occurred more densely in RNA 4. There are 13 sites with substitution ratio above 20% in RSV RNAs and all of them

TABLE 1 | Number of mapped bases and sequencing depth of RSV genome segments.

Genome segment	Mapped bases		Sequencing depth	
	<i>Oryza sativa</i>	<i>Nicotiana benthamiana</i>	<i>Oryza sativa</i>	<i>Nicotiana benthamiana</i>
RNA1	10374012	29240682	1156.7	3260.2
RNA2	3351675	25596930	953.8	7284.3
RNA3	7833087	18150381	3163.6	7330.5
RNA4	4590294	11924373	2129.1	5530.8

were found in RNA 4, except one at the 3570th site in RNA 1 (**Figure 2A**). Average nucleotide diversities in the NS3, CP, SP, and MP ORFs were found to be much higher than those in the RdRp, NS2, NSvc2 ORFs and non-coding regions (**Figure 3A**). Although the average nucleotide diversities in the NS3 and CP ORFs were similar to those in the SP and MP ORFs, the average substitution ratios in these two ORFs were much lower (**Figure 3B**). Interestingly, the average substitution ratio (5.64%) in non-coding regions was higher than any ratios found in the ORFs analyzed. The average nucleotide diversity in non-coding regions was, however, only 0.89%, lower than all ORFs analyzed except the NS2 (**Figure 3**).

Transitions (i.e., A↔G and C↔T) were detected at 238 sites and transversions (i.e., A ↔ C, A ↔ T, G ↔ C, and G ↔ T) were found at 103 sites. The ratio of transition to transversion was about 2.3 (**Table 2**). A previous report also indicated that transitions occurred on an average of 2.5 times more frequently than transversions (Lemey et al., 2009). Our result shows that there is no significant difference for substitution tendency among the four transitions (**Table 2**). At the transition or transversion sites, the average ratio of each substitution varied without clear preferences, e.g., the average ratio of A → T transversion (2.79%) was much higher than A → G transition (0.80%). The average ratio of C → A transversion (2.12%) was similar to C → T transition (2.19%; **Table 2**).

Non-synonymous substitutions were also found in 71 codons located within 6 of the 7 ORFs analyzed. These include 25 codons in RdRp, 7 in NSvc2, 9 in NS3, 12 in CP, 8 in SP, and 10 in MP. No non-synonymous substitution was detected

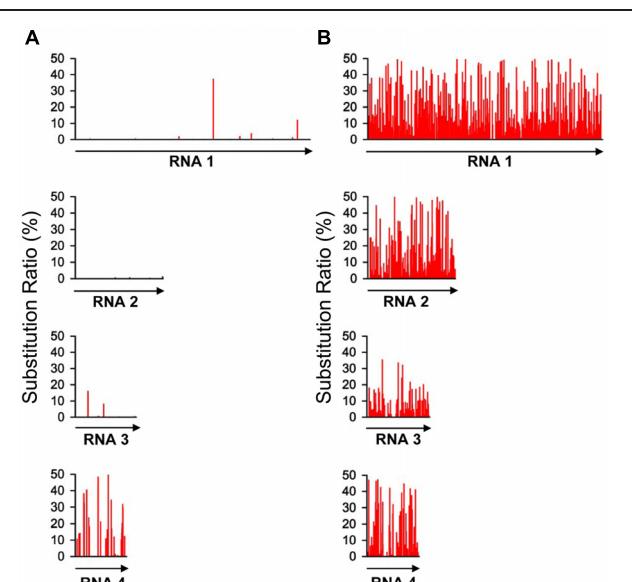


FIGURE 2 | Substitution patterns in the RSV genome occurred in *Oryza sativa* (A) and *Nicotiana benthamiana* (B). The substitution pattern in each of the four RSV RNA segments is presented. The vertical coordinates represent the substitution ratios derived from the master sequence of each populations at each site (number of substituted nucleotide observed/number of total nucleotide sequenced).

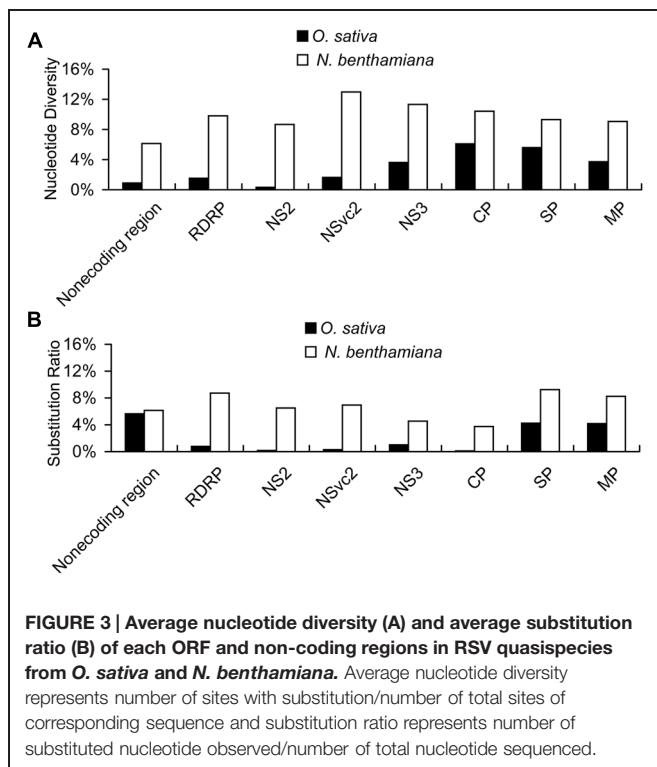


FIGURE 3 | Average nucleotide diversity (A) and average substitution ratio (B) of each ORF and non-coding regions in RSV quasispecies from *O. sativa* and *N. benthamiana*. Average nucleotide diversity represents number of sites with substitution/number of total sites of corresponding sequence and substitution ratio represents number of substituted nucleotide observed/number of total nucleotide sequenced.

in NS2 ORF. Only 4 out of the 71 codons had substitution ratio above 10%. These include the 150th codon in the NS3 ORF (16.19%, Pro → Leu), the 8th codon in the SP ORF (10.92%, Val → Ile), and the 12th and the 200th codons in the MP ORF (12.62 and 34.45%, Ser → Asn and Ile → Val, respectively; **Figure 4A**). The sites of synonymous substitutions were 3.5 times higher than those of non-synonymous substitutions. The average substitution ratio of synonymous substitution

TABLE 2 | Substitution tendency among RSV quasispecies in different hosts.

Substitution type	Substitution number		Substitution ratio	
	<i>O. sativa</i>	<i>N. benthamiana</i>	<i>O. sativa</i>	<i>N. benthamiana</i>
A→	G	61	345	0.80% 7.67%
	C	16	34	0.24% 2.63%
	T	16	90	2.79% 8.11%
T→	C	68	353	1.56% 7.69%
	A	21	90	0.21% 6.54%
	G	9	36	0.21% 7.08%
G→	A	58	326	1.97% 5.57%
	C	3	2	0.08% 2.62%
	T	18	40	0.77% 6.42%
C→	T	51	299	2.19% 7.84%
	A	16	57	2.12% 6.87%
	G	4	3	0.35% 3.52%

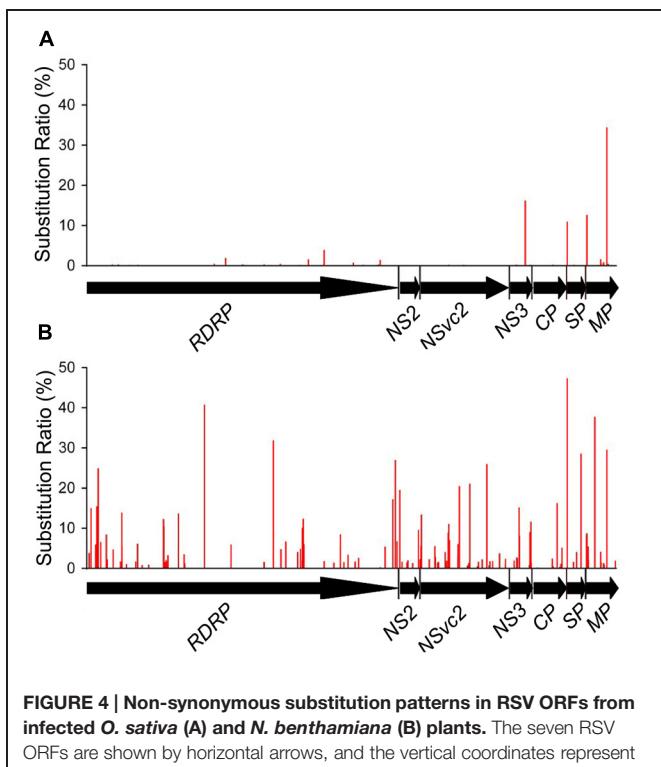


FIGURE 4 | Non-synonymous substitution patterns in RSV ORFs from infected *O. sativa* (A) and *N. benthamiana* (B) plants. The seven RSV ORFs are shown by horizontal arrows, and the vertical coordinates represent the ratios of substitutions derived from the master sequence of each populations at each site (number of substituted nucleotide observed/number of total nucleotide sequenced).

was 1.64 times higher than that of non-synonymous substitutions.

Genetic Diversity of RSV Populations in *N. benthamiana*

To investigate the role of host in RSV evolution, the virus was mechanically inoculated in *N. benthamiana*, an experimental host of RSV. In infected *N. benthamiana* plant, substitutions were found at 1675 sites in the RSV populations, approximately five times more than those found in the RSV populations from infected *O. sativa* plant. The average nucleotide diversity and substitution ratio were also very different between these two groups of RSV populations. For RSV populations from *N. benthamiana*, substitutions were detected in 874 sites in RNA 1 (9.74%), 368 sites in RNA 2 (10.47%), 204 sites in RNA 3 (8.24%), and 149 sites in RNA 4 (6.91%). The substitutions occurred more densely in RNA 1 and 2. There were 206 substitution sites with the substitution ratios above 20%, and more than half of them were found in RNA 1 (**Figure 2B**).

Our results also indicate that the average nucleotide diversity in each ORF and non-coding regions is higher for RSV populations from *N. benthamiana*, especially in the RdRp, NS2, and NSvc2 ORFs and the non-coding region. Consequently the differences of average nucleotide diversities among the ORFs and non-coding regions are decreased. The highest average nucleotide diversity is in the NSvc2 ORF (12.97%) and the lowest average nucleotide diversity is in non-coding regions (6.14%;

TABLE 3 | Substitution tendency of the master sequence after RSV was transmitted from *O. sativa* to *N. benthamiana*.

Original base	No. of substitutions at mutated base			
	A	C	G	T
A	—	2	27	11
C	3	—	0	45
G	34	0	—	1
T	7	21	5	—

Figure 3A). The average substitution ratios are also increased in all ORFs and non-coding regions analyzed. The highest average substitution ratio is found in the *SP* ORF (9.25%) and the lowest in the *CP* ORF (3.75%). Interestingly, the average substitution ratio is not increased in non-coding regions (**Figure 3B**).

A total of 1323 transition sites and 352 transversion sites were detected in the sequences of RSV populations from *N. benthamiana*. The transitions occurred on an average of 3.76 times more frequently than the transversions, but no significant difference in frequency was observed for the four transition groups (**Table 2**). The average substitution ratios of the four transition groups varied without a clear preference. The highest average substitution ratio was for transversion A → T (8.11%) found in RSV populations from *O. sativa* (**Table 2**).

Non-synonymous substitutions were found in 147 codons distributed in the seven ORFs. These include 60 non-synonymous substituted codons in *RdRp*, 6 in *NS2*, 35 in *NSvc2*, 16 in *NS3*, 11 in *CP*, 6 in *SP*, and 13 in *MP* ORF. Twenty-four of the 147 codons had substitution ratios above 10%, and two of them were detected at the eighth codon in the *SP* ORF and the 200th codon in the *MP* ORF, respectively. It is worth to mention that the RSV populations from *O. sativa* also had high substitution ratios at these two codons. There are 24 unique non-synonymous substitutions in the RSV populations from *N. benthamiana*, likely to be the mutations caused by the interaction between RSV and *N. benthamiana* (**Figure 4B**). The average substitution ratios of the synonymous and non-synonymous substitutions were about five times of those found in the RSV populations from *O. sativa*.

Population Structures of RSV in Different Host Plants

The master sequences of RSV populations from the infected *O. sativa* and *N. benthamiana* plants were identified after mapping the reads from the high-throughput sequencing data on the reference sequences of RSV genome segments. There are 156 different (0.9% of genome nucleotides) substitutions between

the two master sequences. The master sequence from infected *O. sativa* plant is composed of 39.09% GC nucleotides while that from the infected *N. benthamiana* plant is 38.92%. When the substitution tendency bias was analyzed, more transitions than transversions (approximately fourfold) were observed, and there were remarkably more G/C to A/T substitutions than the reverse substitutions (**Table 3**). After RSV was transmitted from *O. sativa* to *N. benthamiana*, the G/C ↔ A/T substitutions varied among the regions in the RSV genome (**Table 4**). For example the G/C ↔ A/T substitutions were similar in the non-coding regions of the two master sequences. In the coding regions, the G/C → A/T substitutions seemed to occur more often than the reverse substitutions (**Table 4**). As reported previously, the rice genome is GC rich and rice coding sequences contain 45–75% G and C, whereas the tobacco genome is GC poor and its coding sequences contain 40–60% G and C (Salinas et al., 1988; Matassi et al., 1989; Carels and Bernardi, 2000). The decreased GC level after the virus was transmitted from *O. sativa* to *N. benthamiana* suggests that RSV may adjust its codon usage to ensure its efficient protein translation in a new host.

Discussion

Since the first introduction of quasispecies hypothesis by Eigen and Schuster (1977), this population-based framework has been used extensively to study RNA virus evolution. Based on the complete genomic sequences or partial genomic sequences of RSV isolates, RSV isolates in China were mainly divided into two subtypes: subtype II from Yunnan province and subtype I from rest parts of China, RSV isolates in subtype II were more genetically diverse than those in subtype I, and the mean genetic distance of RSV genes between two subtypes, within subtype I and subtype II ranged from 0.0529 to 0.0865, 0.0123 to 0.0256, and 0.0183 to 0.0387, respectively (Wei et al., 2009; Huang et al., 2013). In this study, we used the high-throughput sequencing technology to analyze and profile RSV populations from infected *O. sativa* or *N. benthamiana* plants. Two parameters, average nucleotide diversity and average substitution ratio at single sites, were obtained through high-throughput sequencing and used to evaluate genetic diversities in RSV populations from an infected *O. sativa* or *N. benthamiana* plant. For the whole RSV genome, the average nucleotide diversity and average substitution ratio were 1.99% and 1.47% from *O. sativa*, and 9.79% and 7.05% from *N. benthamiana*, respectively. RSV isolates in China were reported to have low genetic diversity because of its narrow host range (Wei et al., 2009). Our study find the diversity of RSV

TABLE 4 | Number of A/T ↔ G/C substitutions in the master sequence after RSV was transmitted from *O. sativa* to *N. benthamiana*.

Transition type	Transition number							
	Non-coding	<i>RdRp</i>	<i>NS2</i>	<i>NSvc2</i>	<i>NS3</i>	<i>CP</i>	<i>SP</i>	<i>MP</i>
A/T → G/C	14	26	1	4	4	1	1	4
G/C → A/T	14	45	2	12	3	4	3	1

populations increased after RSV was transmitted from *O. sativa* to *N. benthamiana*. *N. benthamiana* might provide quite different select pressure from *O. sativa*, and this observation also gave a direct evidence that virus might allow higher genetic diversity for host adaptation (Fargette et al., 2006).

The two parameters are independent of each other regardless of the origin of the two virus populations. For example, the average nucleotide diversity in the *CP* ORF was the highest in the RSV populations from the infected *O. sativa* plant, whereas the average substitution ratio in this ORF was much lower than some of the other ORFs (Figure 3). Also transitions were much higher than transversions in the RSV populations from both infected *O. sativa* and *N. benthamiana* plants (Table 4), but the average substitution ratio did not have difference (Figure 2). These results indicate that these two parameters are likely driven by different factors.

The genetic diversity in RNA virus populations was considered to be controlled by the interactions between host and viral factors (Schneider and Roossinck, 2001). So there were at least two steps during RNA virus evolution: substitutions emerged via replication errors and the selection of the original and substituted bases by host. During the second step, the accumulation of the less adaptive one might be repressed by host–virus interaction. So if host–virus interaction is in favor of the substituted base, it will accumulate more than the original base and eventually replace the original one. Otherwise, the substituted base will accumulate at lower level or even be eliminated. So substitution ratio reveals the destiny of each substitutions under host selection. Thus the sites with high substitution ratios may be essential in virus evolution in corresponding host. In this study, sites with high substitution ratio in *O. sativa* and *N. benthamiana* were documented. Further work is necessary to confirm their function in virus evolution and host–virus interaction.

References

Carels, N., and Bernardi, G. (2000). Two classes of genes in plants. *Genetics* 154, 1819–1825.

Domingo, E., and Holland, J. J. (1997). RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* 51, 151–178. doi: 10.1146/annurev.micro.51.1.151

Eigen, M., and Schuster, P. (1977). The hypercycle. a principle of natural self-organization. Part A: emergence of the hypercycle. *Naturwissenschaften* 64, 541–565. doi: 10.1007/BF00450633

Elena, S. F., and Sanjuán, R. (2007). Virus evolution: Insights from an experimental approach. *Annu. Rev. Ecol. Evol. Syst.* 38, 27–52. doi: 10.1146/annurev.ecolsys.38.091206.095637

Falk, B. W., and Tsai, J. H. (1998). Biology and molecular biology of viruses in the genus *Tenuivirus*. *Annu. Rev. Phytopathol.* 36, 139–163. doi: 10.1146/annurev.phyto.36.1.139

Fargette, D., Konate, G., Fauquet, C., Muller, E., Peterschmitt, M., and Thresh, J. M. (2006). Molecular ecology and emergence of tropical plant viruses. *Annu. Rev. Phytopathol.* 44, 235–260. doi: 10.1146/annurev.phyto.44.120705.104644

Garcia-Arenal, F., Fraile, A., and Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annu. Rev. Phytopathol.* 39, 157–186. doi: 10.1146/annurev.phyto.39.1.157

Garcia-Arenal, F., Fraile, A., and Malpica, J. M. (2003). Variation and evolution of plant virus populations. *Int. Microbiol.* 6, 225–232. doi: 10.1007/s10123-003-0142-z

Huang, L. Z., Rao, L. X., Zhou, X. P., and Wu, J. X. (2013). Genetic variability and evolution of rice stripe virus. *J. Zhejiang Univ. Sci. B* 14, 875–885. doi: 10.1631/jzus.B1300014

The observation that the average substitution ratio of non-coding regions was not changed after RSV was transmitted from *O. sativa* to *N. benthamiana* indicates that the RSV non-coding sequences may not play an important role in RSV–host plant interactions. On the other hand both synonymous and non-synonymous substitution ratios were remarkably increased after RSV was transmitted from *O. sativa* to *N. benthamiana*. This observation suggests that both synonymous and non-synonymous substitutions are also essential in RSV–host plant interactions. Possible scenarios that may be used to explain our hypothesis are that the non-synonymous substitutions may alter features of RSV protein(s) to better interact with specific host factors during its infection, e.g., replication and movement in plant, while synonymous substitutions alter the codon composition to accord with host codon usage bias.

Conclusion

We find the RSV populations from infected *N. benthamiana* plant is more diverse than that obtained from rice and the high-throughput sequencing technology is a powerful technology to investigate population genetic diversity of plant RNA viruses.

Acknowledgments

This research was supported by the National High Technology Research and Development Program of China (2012AA101505), the earmarked fund for Modern Agro-industry Technology Research System (CARS-01-19) and the National Natural Science of China (31272015).

Kakutani, T., Hayano, Y., Hayashi, T., and Minobe, Y. (1991). Ambisense segment 3 of rice stripe virus: the first instance of a virus containing two ambisense segments. *J. Gen. Virol.* 72, 465–468. doi: 10.1099/0022-1317-72-2-465

Koganezawa, H., Doi, Y., and Yora, K. (1975). Purification of Rice stripe virus. *Ann. Phytopathol. Soc. Jpn.* 41, 148–154. doi: 10.3186/jjphytopath.41.148

Kong, L., Wu, J., Lu, L., Xu, Y., and Zhou, X. (2014). Interaction between Rice stripe virus disease-specific protein and host PsbP enhances virus symptoms. *Mol. Plant* 7, 691–708. doi: 10.1093/mp/sst158

Lauring, A. S., and Andino, R. (2010). Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog.* 6:e1001005. doi: 10.1371/journal.ppat.1001005

Lemey, P., Salemi, M., and Vandamme, A. M. (2009). *The Phylogenetic Handbook: a Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511819049

Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 18, 1851–1858. doi: 10.1101/gr.078212.108

Lin, Q. Y., Xie, L. H., Zhou, Z. J., Xie, L. Y., and Wu, Z. J. (1990). Studies on rice stripe: i. distribution of and losses caused by the disease. *J. Fujian Agric. Coll.* 19, 421–425.

Matassi, G., Montero, L. M., Salinas, J., and Bernardi, G. (1989). The isochore organization and the compositional distribution of homologous coding sequences in the nuclear genome of plants. *Nucleic Acids Res.* 17, 5273–5290. doi: 10.1093/nar/17.13.5273

Nagy, P. D., and Simon, A. E. (1997). New insights into the mechanisms of RNA recombination. *Virology* 235, 1–9. doi: 10.1006/viro.1997.8681

Ramirez, B. C., and Haenni, A. L. (1994). Molecular biology of *Tenuivirus*es, a remarkable group of plant viruses. *J. Gen. Virol.* 75, 467–475. doi: 10.1099/0022-1317-75-3-467

Roossinck, M. J. (1997). Mechanisms of plant virus evolution. *Annu. Rev. Phytopathol.* 35, 191–209. doi: 10.1146/annurev.phyto.35.1.191

Salinas, J., Matassi, G., Montero, L. M., and Bernardi, G. (1988). Compositional compartmentalization and compositional patterns in the nuclear genomes of plants. *Nucleic Acids Res.* 16, 4269–4285. doi: 10.1093/nar/16.10.4269

Sanjuan, R., Nebot, M. R., Chirico, N., Mansky, L. M., and Belshaw, R. (2010). Viral mutation rates. *J. Virol.* 84, 9733–9748. doi: 10.1128/JVI.00694-10

Schneider, W. L., and Roossinck, M. J. (2000). Evolutionarily related *Sindbis*-like plant viruses maintain different levels of population diversity in a common host. *J. Virol.* 74, 3130–3134. doi: 10.1128/JVI.74.7.3130-3134.2000

Schneider, W. L., and Roossinck, M. J. (2001). Genetic diversity in RNA virus quasispecies is controlled by host-virus interactions. *J. Virol.* 75, 6566–6571. doi: 10.1128/JVI.75.14.6566-6571.2001

Takahashi, M., Toriyama, S., Hamamatsu, C., and Ishihama, A. (1993). Nucleotide sequence and possible ambisense coding strategy of rice stripe virus RNA segment 2. *J. Gen. Virol.* 74, 769–773. doi: 10.1099/0022-1317-74-4-769

Toriyama, S., Takahashi, M., Sano, Y., Shimizu, T., and Ishihama, A. (1994). Nucleotide sequence of RNA 1, the largest genomic segment of rice stripe virus, the prototype of the *Tenuivirus*es. *J. Gen. Virol.* 75, 3569–3579. doi: 10.1099/0022-1317-75-12-3569

Wang, G., Zhou, Y., Chen, Z., and Zhou, X. (2004). Production of monoclonal antibodies to Rice stripe virus and application in virus detection. *Acta Phytopathol. Sin.* 34, 302–306.

Wei, T. Y., Yang, J. G., Liao, F. R., Gao, F. L., Lu, L. M., Zhang, X. T., et al. (2009). Genetic diversity and population structure of rice stripe virus in China. *J. Gen. Virol.* 90, 1025–1034. doi: 10.1099/vir.0.006858-0

Worobey, M., and Holmes, E. C. (1999). Evolutionary aspects of recombination in RNA viruses. *J. Gen. Virol.* 80, 2535–2543.

Xiong, R., Wu, J., Zhou, Y., and Zhou, X. (2008). Identification of a movement protein of the *Tenuivirus* rice stripe virus. *J. Virol.* 82, 12304–12311. doi: 10.1128/JVI.01696-08

Xiong, R., Wu, J., Zhou, Y., and Zhou, X. (2009). Characterization and subcellular localization of an RNA silencing suppressor encoded by rice stripe *Tenuivirus*. *Virology* 387, 29–40. doi: 10.1016/j.virol.2009.01.045

Zhu, Y., Hayakawa, T., and Toriyama, S. (1992). Complete nucleotide sequence of RNA 4 of rice stripe virus isolate T, and comparison with another isolate and with maize stripe virus. *J. Gen. Virol.* 73, 1309–1312. doi: 10.1099/0022-1317-73-5-1309

Zhu, Y., Hayakawa, T., Toriyama, S., and Takahashi, M. (1991). Complete nucleotide sequence of RNA 3 of rice stripe virus: an ambisense coding strategy. *J. Gen. Virol.* 72, 763–767. doi: 10.1099/0022-1317-72-4-763

Conflict of Interest Statement: 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.

Copyright © 2015 Huang, Li, Wu, Xu, Yang, Fan, Fang and Zhou. 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) or licensor 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 Genome-wide Combinatorial Strategy Dissects Complex Genetic Architecture of Seed Coat Color in Chickpea

OPEN ACCESS

Edited by:

Maria J. Monteros,
The Samuel Roberts Noble
Foundation, USA

Reviewed by:

Steven B. Cannon,
United States Department of
Agriculture - Agricultural Research
Service, USA
Jianping Wang,
University of Florida, USA
Bharadwaj Chellapilla,
Indian Agricultural Research Institute,
India

***Correspondence:**

Swarup K. Parida
swarup@nigr.ac.in;
swarupdbt@gmail.com

[†]These authors have contributed
equally to this work.

Specialty section:

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

Received: 21 February 2015

Accepted: 26 October 2015

Published: 17 November 2015

Citation:

Bajaj D, Das S, Upadhyaya HD,
Ranjan R, Badoni S, Kumar V,
Tripathi S, Gowda CLL, Sharma S,
Singh S, Tyagi AK and Parida SK
(2015) A Genome-wide Combinatorial
Strategy Dissects Complex Genetic
Architecture of Seed Coat Color in
Chickpea. *Front. Plant Sci.* 6:979.
doi: 10.3389/fpls.2015.00979

Deepak Bajaj^{1†}, Shouvik Das^{1†}, Hari D. Upadhyaya^{2†}, Rajeev Ranjan¹, Saurabh Badoni¹,
Vinod Kumar³, Shailesh Tripathi⁴, C. L. Laxmipathi Gowda², Shivali Sharma²,
Sube Singh², Akhilesh K. Tyagi¹ and Swarup K. Parida^{1*}

¹ National Institute of Plant Genome Research, New Delhi, India, ² International Crops Research Institute for the Semi-Arid Tropics, Telangana, India, ³ National Research Centre on Plant Biotechnology, New Delhi, India, ⁴ Division of Genetics, Indian Agricultural Research Institute, New Delhi, India

The study identified 9045 high-quality SNPs employing both genome-wide GBS- and candidate gene-based SNP genotyping assays in 172, including 93 cultivated (*desi* and *kabuli*) and 79 wild chickpea accessions. The GWAS in a structured population of 93 sequenced accessions detected 15 major genomic loci exhibiting significant association with seed coat color. Five seed color-associated major genomic loci underlying robust QTLs mapped on a high-density intra-specific genetic linkage map were validated by QTL mapping. The integration of association and QTL mapping with gene haplotype-specific LD mapping and transcript profiling identified novel allelic variants (non-synonymous SNPs) and haplotypes in a MATE secondary transporter gene regulating light/yellow brown and beige seed coat color differentiation in chickpea. The down-regulation and decreased transcript expression of beige seed coat color-associated MATE gene haplotype was correlated with reduced proanthocyanidins accumulation in the mature seed coats of beige than light/yellow brown seed colored *desi* and *kabuli* accessions for their coloration/pigmentation. This seed color-regulating MATE gene revealed strong purifying selection pressure primarily in LB/YB seed colored *desi* and wild *Cicer reticulatum* accessions compared with the BE seed colored *kabuli* accessions. The functionally relevant molecular tags identified have potential to decipher the complex transcriptional regulatory gene function of seed coat coloration and for understanding the selective sweep-based seed color trait evolutionary pattern in cultivated and wild accessions during chickpea domestication. The genome-wide integrated approach employed will expedite marker-assisted genetic enhancement for developing cultivars with desirable seed coat color types in chickpea.

Keywords: chickpea, GBS, GWAS, haplotype, QTL, seed coat color, SNP, wild accession

INTRODUCTION

Chickpea (*Cicer arietinum* L.) is the third largest produced food legume crop globally that serves as a vital human dietary source of protein enriched with essential amino acids (Kumar et al., 2011; Gaur et al., 2012; Varshney et al., 2013). The seeds are defining characteristics of chickpea aside from having high economic value. The physical appearance of seeds such as light colored and large sized seeds of chickpea has always been a trait of consumer preference and trade value, besides an important target quality component and adaptation traits (Cassells and Caddick, 2000; Graham et al., 2001; Hossain et al., 2010). The seed coat color is an important agronomic trait in chickpea that varies widely among core and mini-core germplasm collections, cultivated *desi* and *kabuli* types, landraces and wild species accessions (Upadhyaya and Ortiz, 2001; Upadhyaya et al., 2002, 2006). Considering the agronomic importance and wider level of existing phenotypic variation of seed coat color in chickpea, it is essential to identify the underlying heritable forces and potential genes/QTLs (quantitative trait loci) influencing this large trait variation in *desi* and *kabuli* chickpea. Based on defined agromorphological descriptors, the large-scale phenotyping of core, mini-core and reference core chickpea accessions have classified and characterized these accessions into 24 diverse classes of seed coat color types (like brown, light brown, green, yellow-brown and creamy) (Upadhyaya and Ortiz, 2001; Upadhyaya et al., 2002, 2008). This revealed the predominance of yellow/light brown and beige seed coat color in *desi* and *kabuli* chickpea.

Tremendous efforts have been made for studying the genetic inheritance pattern as well as identification and molecular cloning of multiple genes harboring the major QTLs regulating seed coat color trait in diverse crop plants, including rice, soybean, sesame, *Brassica* and grape (Furukawa et al., 2006; Liu et al., 2006; Sweeney et al., 2006, 2007; Yang et al., 2010; Rahman et al., 2011; Huang et al., 2012; Zhang et al., 2013). The formation of different seed coat color is governed by organ/tissue-specific accumulation of plant secondary metabolites like flavonoid compounds, including anthocyanin, flavonols, isoflavones, flavonol glucosides, phlobaphenes, proanthocyanidin, leucoanthocyanidin, and proanthocyanidin in the seed coat (Furukawa et al., 2006). A complex biochemical pathway like phenylpropanoid (flavonoid and anthocyanin biosynthetic pathway) pathway is known to be involved in synthesis of these secondary metabolites in the specific tissues of crop plants (Dixon et al., 2002). The advances in structural and functional genomics have greatly catalyzed isolation of most of the transcriptionally regulated loci/genes that encode enzymes of this important biosynthetic pathway governing seed coat color in rice, maize, *Arabidopsis*, *Brassica*, *Medicago*, common bean, pea *Antirrhinum* and *Petunia* (Ludwig et al., 1989; Mol et al., 1998; Hu et al., 2000; Forkmann and Martens, 2001; Sakamoto et al., 2001; Winkel-Shirley, 2001; Saitoh et al., 2004; Xie and Dixon, 2005; Furukawa et al., 2006; Lepiniec et al., 2006; Quattrocchio et al., 2006; Zhao and Dixon, 2009; Golam Masum Akond et al., 2011; Li et al., 2012; Ferraro et al., 2014; Liu et al., 2014). However, the genetic analysis of seed color trait has mostly been restricted to estimation of heritability and molecular mapping

of only a few seed coat color-associated QTLs because of its quantitative and polygenic nature of inheritance (epistatic and pleiotropic effects) both in *desi* and *kabuli* chickpea (Hossain et al., 2010). However, the markers associated with these QTLs controlling seed color are yet to be validated across diverse genetic backgrounds and environments for their deployment in marker-assisted breeding program to develop improved chickpea cultivars with desirable seed coat color.

A combinatorial approach of SNP (single nucleotide polymorphism) and SSR (simple sequence repeat) marker-based genome-wide association study (GWAS), candidate gene-based association analysis and traditional bi-parental QTL mapping in conjunction with differential gene expression profiling and gene-based molecular haplotyping/LD (linkage disequilibrium) mapping is a widely recognized strategy for dissecting the complex yield and quality component quantitative traits in different crop plants, including chickpea (Konishi et al., 2006; Sweeney et al., 2007; Tian et al., 2009; Mao et al., 2010; Li et al., 2011; Kharabian-Masouleh et al., 2012; Parida et al., 2012; Kujur et al., 2013, 2014; Negrão et al., 2013; Saxena et al., 2014a; Zuo and Li, 2014; Bajaj et al., 2015). Natural variation and artificial selection in many cloned genes (e.g., red pericarp color-regulating *Rc* gene in rice) underlying the QTLs regulating seed coat color during domestication have established seed color as a common target trait for both domestication and artificial breeding in crop plants (Sweeney et al., 2007; Hossain et al., 2011; Meyer et al., 2012). Henceforth, the above-said integrated strategy can be deployed for identification of functionally relevant novel molecular tags (markers, genes, QTLs, and alleles) controlling seed coat color and unraveling the molecular genetic basis of such natural trait variation in various cultivated and wild accessions of chickpea adapted to diverse agro-climatic regions. The resultant findings would assist us to genetically select *desi* and *kabuli* accessions for diverse desirable seed coat color, driving genomics-assisted breeding and ultimately development of tailored chickpea cultivars with preferred seed color.

In view of the aforementioned prospects, our study integrated GWAS and candidate gene-based association analysis with traditional QTL mapping, differential gene expression profiling, gene haplotype-specific LD mapping and proanthocyanidin (PA) quantitation assay to identify novel potential genomic loci (genes/gene-associated targets) and alleles/haplotypes regulating seed coat color and understand their haplotype-based evolutionary/domestication patterns in cultivated and wild chickpea.

MATERIALS AND METHODS

Discovery, Genotyping and Annotation of Genome-wide SNPs

For large-scale discovery and high-throughput genotyping of SNPs at a genome-wide scale, 93 phenotypically and genotypically diverse *desi* (39 accessions) and *kabuli* (54) chickpea accessions (representing diverse eco-geographical regions of 21 countries of the world) were selected (Table S1) from the available chickpea germplasm collections (16991,

including 211 minicore germplasm lines, Upadhyaya and Ortiz, 2001; Upadhyaya et al., 2001, 2008). In addition, 79 diverse *Cicer* accessions belonging to five annual wild species, namely *C. reticulatum* (14), *C. echinospermum* (8), *C. judaicum* (22), *C. bijugum* (19), and *C. pinnatifidum* (15) and one perennial species *C. microphyllum* were chosen for haplotype-based evolutionary study (Table S1). The genomic DNA from the young leaf samples of 172 cultivated and wild chickpea accessions were isolated using QIAGEN DNeasy 96 Plant Kit (QIAGEN, USA).

The genomic DNA of accessions was digested with *Ape*KI, ligated to adapters-carrying unique barcodes and 2 × 96-plex GBS libraries were constructed. Each of two 96-plex GBS libraries were pooled together and sequenced (100-bp single end) individually using Illumina two-lane flow cell of HiSeq2000 NGS (next-generation sequencing) platform following Elshire et al. (2011) and Spindel et al. (2013). The unique barcodes attached to each of 172 accessions were used as reference for demultiplexing of their high-quality sequence reads. The decoded high-quality FASTQ sequences were mapped to reference *kabuli* draft chickpea genome (Varshney et al., 2013) employing Bowtie v2.1.0 (Langmead and Salzberg, 2012). The sequence alignment map files generated from *kabuli* genome were analyzed using the reference-based GBS pipeline/genotyping approach of STACKS v1.0 (<http://creskolab.uoregon.edu/stacks>) for identifying the high-quality SNPs (SNP base-quality, 20 supported by minimum sequence read-depth, 10) from 172 accessions (as per Kujur et al., 2015). The *kabuli* genome annotation (Varshney et al., 2013) was utilized as an anchor for structural and functional annotation of GBS-based SNPs in different coding (synonymous and non-synonymous SNPs) and non-coding sequence components of genes and genomes (chromosomes/pseudomolecules and scaffolds) of *kabuli* chickpea.

Mining and Genotyping of Candidate Gene-based SNPs

For large-scale mining and high-throughput genotyping of candidate gene-based SNPs in chickpea, a selected set of 28 cloned genes/QTLS known to be involved in regulation of seed pericarp color in related dicot species, *Arabidopsis thaliana*, *Medicago*, and soybean (Lepiniec et al., 2006; Zhao and Dixon, 2009; Yang et al., 2010) were acquired. This includes a tandem array of three seed color candidate co-orthologous *tt12*-MATE family transporter repeated genes located at the selective sweep target region of *kabuli* chromosome 4 (reported previously by Varshney et al., 2013). The coding sequences (CDS) of these known/candidate seed color-regulating genes were BLAST searched against the CDS of *kabuli* genes to identify the best possible gene orthologs in chickpea. The CDS (functional domains) and 1000-bp upstream regulatory regions (URRs) of identified true *kabuli* chickpea gene orthologs (*E*-value: 0 and bit score ≥ 500) were targeted to design (Batch Primer3, <http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi>) the forward and reverse primers with expected amplification product size of 500–700 bp. For mining the gene-based SNPs, primarily 24 *desi*, *kabuli*, and wild chickpea accessions were selected randomly from 172 accessions (utilized for genome-wide GBS-based SNP genotyping) (Table S1). The gene-based primers were

PCR amplified using the genomic DNA of all these 24 accessions and resolved in 1.5% agarose gel. The amplified PCR products were sequenced, the high-quality sequences aligned and SNPs were detected in diverse sequence components of orthologous chickpea genes among accessions following Saxena et al. (2014a). The gene-derived coding and URR-SNPs discovered were further genotyped in the genomic DNA of 172 cultivated and wild chickpea accessions employing Sequenom MALDI-TOF (matrix-assisted laser desorption ionization-time of flight) MassARRAY (<http://www.sequenom.com>) as per Saxena et al. (2014a). According to allele-specific mass differences between extension products, the genotyping information of different coding and URR-SNPs among accessions were recorded.

Estimation of Polymorphism and Diversity Statistics

To measure the PIC (polymorphism information content) and genetic diversity coefficient (Nei's genetic distance; Nei et al., 1983), and construct an unrooted neighbor-joining (NJ)-based phylogenetic tree (with 1000 bootstrap replicates) among chickpea accessions, the high-quality SNP genotyping data were analyzed with PowerMarker v3.51 (Liu and Muse, 2005), MEGA v5.0 (Tamura et al., 2011) and TASSEL v5.0 (<http://www.maizegenetics.net>). A 100-kb non-overlapping sliding window approach of TASSEL v5.0 was implemented to estimate the various nucleotide diversity measures ($\theta\pi$, $\theta\omega$, and Tajima's D) following Xu et al. (2011) and Varshney et al. (2013).

Determination of Population Genetic Structure and LD Patterns

A model-based program STRUCTURE v2.3.4 was utilized to determine the population genetic structure among chickpea accessions adopting the methods of Kujur et al. (2013, 2014) and Saxena et al. (2014a). The high quality genome-wide and candidate gene-based genotyping data of SNPs physically mapped on eight *kabuli* chromosomes were analyzed using a command line interface of PLINK and the full-matrix approach of TASSELv5.0 (Saxena et al., 2014a; Kumar et al., 2015) to evaluate the genome-wide LD patterns (r^2 , frequency correlation among pair of alleles across a pair of SNP loci) and LD decay (by plotting average r^2 against 50 and 20 kb uniform physical intervals across chromosomes) in chromosomes and population of chickpea.

Phenotyping for Seed Coat Color

One hundred seventy-two chickpea accessions were grown in the field [following randomized complete block design (RCBD)] during crop growing season for two consecutive years (2011 and 2012) at two diverse geographical locations (New Delhi; latitude/longitude: $28.4^\circ\text{N}/77.1^\circ\text{E}$ and Hyderabad; $17.1^\circ\text{N}/78.9^\circ\text{E}$) of India. The seeds of all accessions were phenotyped and classified/characterized into different seed color types by visual estimation of seed coat/pericarp color of mature seeds (stored no longer than 5 months) representing each accession with at least three replications (Upadhyaya

and Ortiz, 2001; Upadhyaya et al., 2002, 2006; Hossain et al., 2011). Based on these large-scale replicated multi-location/years field phenotyping, the 172 chickpea accessions were broadly categorized/characterized in accordance with 24 predominant seed coat color types (BL: Black, B: Brown, LB: Light brown, DB: Dark brown, RB: Reddish brown, GB: Grayish brown, SB: Salmon brown, OB: Orange brown, GR: Gray, BB: Brown beige, Y: Yellow, LY: Light yellow, YB: Yellow brown, OY: Orange yellow, O: Orange, YE: Yellow beige, I: Ivory, G: Green, LG: Light green, BR: Brown reddish, M: Variegated, BM: Black brown mosaic, LO: Light orange and BE: Beige) (Figure S1). The frequency distribution and broad-sense heritability (H^2) of seed coat color were measured using SPSSv17.0 (<http://www.spss.com/statistics>) as per Kujur et al. (2013, 2014) and Saxena et al. (2014a).

Trait Association Mapping

The genome-wide and candidate gene-based SNP genotyping information of 93 cultivated *desi* and *kabuli* accessions was integrated with their seed coat color phenotyping, ancestry coefficient (Q matrix obtained from population structure) and relative kinship matrix (K) data using GLM (Q model)- and MLM (Q+K model)-based approaches of TASSELv5.0 (Kujur et al., 2013, 2014; Kumar et al., 2015). Furthermore, mixed-model (P+K, K, and Q+K) association (EMMA) and P3D/compressed mixed linear model (CMLM) interfaces of GAPIT (Lipka et al., 2012) relying upon principal component analysis (PCA) were employed for GWAS. The relative distribution of observed and expected $-\log_{10}(P)$ -value for each SNP marker-trait association were compared individually utilizing quantile-quantile plot of GAPIT. According to false discovery rate (FDR cut-off ≤ 0.05) (Benjamini and Hochberg, 1995), the adjusted P -value threshold of significance was corrected for multiple comparisons. By integrating all four model-based outputs of TASSEL and GAPIT, the potential SNP loci in the target genomic (gene) regions revealing significant association with seed coat color trait variation at highest R^2 (degree of SNP marker-trait association) and lowest FDR adjusted P -values (threshold $P < 1 \times 10^{-6}$) were identified.

QTL Mapping

For validation of seed color-associated genomic loci through QTL mapping, an intra-specific 190 F_7 RIL mapping population (ICC 12299 \times ICC 8261) was developed. The chickpea accessions ICC 12299 (originated from Nepal) and ICC 8261 (originated from Turkey) are LB/YB and BE seed coat colored *desi* and *kabuli* landraces, respectively. A selected set of 384 SNPs showing polymorphism between mapping parental accessions (ICC 12299 and ICC 8261) and 96 previously reported SSR markers (Winter et al., 1999, 2000; Hossain et al., 2011; Thudi et al., 2011) physically/genetically mapped on LGs (linkage groups)/chromosomes (considered as anchors) were genotyped using MALDI-TOF mass array SNP genotyping assay and fluorescent dye-labeled automated fragment analyser (Kujur et al., 2013, 2014; Saxena et al., 2014a,b; Bajaj et al., 2015). The genotyping data of parental polymorphic SNP and SSR markers mapped on LGs (chromosomes) of an intra-specific

genetic map and replicated multi-location seed coat color field phenotyping information (following aforesaid phenotyping methods of 172 accessions) of 190 RILmapping individuals and parental accessions were analyzed using the composite interval mapping (CIM) function of MapQTL v6.0 (Van Ooijen, 2009, as per Saxena et al., 2014a). The phenotypic variation explained (PVE) by QTLs (R^2 %) was measured at significant LOD (logarithm of odds threshold > 5.0) to identify and map the major genomic loci underlying robust seed color QTLs in chickpea.

Differential Expression Profiling

To infer differential gene regulatory function, the expression profiling of seed color-associated genes validated by both GWAS/candidate gene-based association analysis and QTL mapping was performed. For expression analysis, 11 contrasting chickpea accessions as well as parental accessions and two homozygous individuals of a RIL mapping population (ICC 12299 \times ICC 8261) representing two predominant seed coat color types, LB/YB (*desi*: ICCX810800, *desi*: ICC 12299, *desi*: ICCV 10, *desi*: ICC 4958, *desi*: ICC 15061, *kabuli*: ICC 6204, and *kabuli*: Annigeri) and BE (*kabuli*: ICC 20268, *kabuli*: ICC 8261, *kabuli*: Phule G0515, and *desi*: ICC 4926) were selected (Table S1). The vegetative leaf tissues, including seeds without seed coats and seed coats scrapped from the immature (15–25 days after podding/DAP) and mature (26–36 DAP) seeds (30–40 seeds per accession in replicates) of all these accessions and mapping parents/individuals were used for RNA isolation. The isolated RNA was amplified with the gene-specific primers using the semi-quantitative and quantitative RT-PCR assays as per Bajaj et al. (2015). Three independent biological replicates of each sample and two technical replicates of each biological replicate with no template and primer as control were included in the quantitative RT-PCR assay. A house-keeping gene elongation factor 1-alpha ($EF1\alpha$) was utilized as internal control in RT-PCR assay for normalization of expression values across various tissues and developmental stages of accessions and mapping parents/individuals. The significant difference of gene expression in immature and mature seed coats as compared to vegetative leaf tissues and seeds without seed coats (considered as control) of each accessions and mapping parents/individuals was determined and compared among each other to construct a heat map using TIGR MultiExperiment Viewer (MeV, <http://www.tm4.org/mev>).

Molecular Haplotyping

For gene-based marker haplotyping/LD mapping, the 2 kb URR, exons, introns and 1 kb DRR (downstream regulatory region) of one strong seed color-associated gene (validated by GWAS, QTL mapping, and differential expression profiling) amplified from 172, including 93 cultivated and 79 wild chickpea accessions (Table S1) were cloned and sequenced. For these, the gene amplified PCR products were purified (QIAquick PCR Purification Kit, QIAGEN, USA), ligated with pGEM-T Easy Vector (Promega, USA) and transformed into competent DH5 α *E. coli* strain by electroporation (BIO-RAD, USA) to screen blue/white colonies. The plasmids isolated from 20 random positive white clones carrying the desired inserts were sequenced

in both forward and reverse directions twice on a capillary-based Automated DNA Sequencer (Applied Biosystems, ABI 3730xl DNA Analyzer, USA) using the BigDye Terminator v3.1 sequencing kit and M13 forward and reverse primers. The trace files were base called separately for all the 10 clones of individual accessions and/primers, checked for quality using *phred* and then assembled into contigs using *phrap*. The high-quality sequences generated for the gene were aligned and compared among 172 chickpea accessions using the CLUSTALW multiple sequence alignment tool in MEGA 6.0 (Tamura et al., 2013) to discover the SNP and SSR allelic variants among these accessions. The SNP-SSR marker-based haplotypes were constituted in a gene to determine the haplotype-based LD patterns among accessions following Kujur et al. (2013, 2014) and Saxena et al. (2014a). To evaluate the trait-association potential and determine the haplotype-based evolutionary pattern of the gene, the marker-based gene haplotype-derived genotyping information was correlated with seed color field phenotyping data of accessions used. To assess the potential of seed color-regulatory haplotypes constituted in a gene, differential expression profiling in the seed coats dissected from mature seeds of aforesaid six contrasting chickpea accessions as well as mapping parents and two homozygous RIL individuals representing LB/YB (ICCX-810800, ICC 12299, and Annigeri)- and BE (ICC 20268, ICC 8261, and ICC 4926)-specific haplotype groups was performed using their respective gene haplotype-based primers (following aforementioned methods).

Estimation of Proanthocyanidin (PA)

The seed coats (~50 mg) (three independent biological replicates per accession) from matured seeds of aforementioned six contrasting chickpea accessions, mapping parents and two homozygous RIL individuals representing LB/YB and BE seed coat color haplotypes were scraped. These seed coat tissues were used for soluble (with 0.2% DMACA) and insoluble (with butanol-HCl) PA extraction and quantification following the methods of Pang et al. (2008) and Liu et al. (2014).

RESULTS AND DISCUSSION

Large-scale Discovery, High-throughput Genotyping and Annotation of Genome-wide GBS- and Candidate Gene-derived SNPs

The sequencing of 2×96 -plex *ApeKI* GBS libraries generated on an average of 200.7 million high-quality sequence reads that are evenly distributed (mean: 2.1 million reads per accession) across 172, including 93 cultivated and 79 wild chickpea accessions (Table S1). Notably, on an average 87% of these sequence reads were mapped to unique physical positions on *kabuli* reference genome. The sequencing data generated from 93 cultivated (*desi* and *kabuli*) and 79 wild chickpea accessions in this study have been submitted to a freely accessible NCBI-short read archive (SRA) database (www.ncbi.nlm.nih.gov/sra) with accession numbers SRX845396 (<http://www.ncbi.nlm.nih.gov/gquery/?term=SRX845396>) and SRX971856, respectively. In

total, 8673 high-quality SNPs (with read-depth 10, SNP base quality ≥ 20 , <1% missing data and ~2% heterozygosity in each accession) differentiating 172 accessions were discovered using *kabuli* reference genome-based GBS assay (Table S2). This highlights the advantages of GBS assay in rapid large-scale mining and high-throughput genotyping of accurate and high-quality SNPs simultaneously at a genome-wide scale in chickpea. The successful applications of GBS assay particularly in construction of high-density intra- and inter-specific genetic linkage maps, high-resolution GWAS and fine mapping/positional cloning of genes/QTLs governing important agronomic traits have been demonstrated recently in diverse legume crops, including chickpea (Sonah et al., 2013, 2014; Deokar et al., 2014; Jaganathan et al., 2014; Liu et al., 2014). In this context, GBS-based genome-wide SNPs identified by us have potential to be utilized for genomics-assisted breeding applications in chickpea.

A number of diverse known cloned genes governing complex transcriptional regulatory anthocyanin and flavonoid biosynthetic pathways to accumulate specific secondary metabolites (majorly anthocyanin and proanthocyanidin) in the seed coat for its coloration/pigmentation have been identified and characterized in *Arabidopsis*, soybean and *Medicago* (Focks et al., 1999; Debeaujon et al., 2001; Sagasser et al., 2002; Saitoh et al., 2004; Baxter et al., 2005; Xie and Dixon, 2005; Furukawa et al., 2006; Lepiniec et al., 2006; Liu et al., 2006, 2014; Quattrocchio et al., 2006; Sweeney et al., 2006, 2007; Zhao and Dixon, 2009; Erdmann et al., 2010; Yang et al., 2010; Golam Masum Akond et al., 2011; Rahman et al., 2011; Bowerman et al., 2012; Chen et al., 2012, 2014; Huang et al., 2012; Li et al., 2012; Routaboul et al., 2012; Nguyen et al., 2013; Pourcel et al., 2013; Wang et al., 2013; Zhang et al., 2013; Ichino et al., 2014; Pesch et al., 2014; Wroblewski et al., 2014; Xu et al., 2014). Recently, a tandem array of three seed color-regulating candidate co-orthologous *tt12*-MATE family transporter repeated genes of *M. truncatula* primarily affecting proanthocyanidin biosynthesis under strong purifying selection pressure has been identified at the selective sweep target region of *kabuli* chromosome 4 (Varshney et al., 2013). Besides, the detailed genetic regulation of seed color in chickpea is poorly understood vis-a-vis other legumes. Therefore, the seed color-related known cloned and candidate gene orthologs of legumes (*Medicago* and soybean) and *Arabidopsis* were targeted in our study for genetic association mapping of seed color traits in chickpea. To access the potential of aforesaid known/candidate genes for seed color trait association in chickpea, genetic association mapping was performed by high-throughput genotyping of novel SNP allelic variants mined from diverse sequence components of these genes in 93 cultivated and 79 wild accessions. The PCR amplicon-based sequencing of 28 seed color-related known/candidate *kabuli* gene (CDS and 1 kb-URRs) orthologs in 24 selected chickpea accessions and further comparison of their high-quality gene amplicon sequences detected numerous coding and non-coding SNP allelic variants. The subsequent large-scale validation and high-throughput genotyping of these mined SNP alleles in 172 cultivated and wild accessions by MALDI-TOF mass array successfully identified 372 genic SNPs (233 coding

non-synonymous/synonymous and 139 URR SNPs) with an average frequency of 2.82 SNPs/kb (varied from 1.58 to 3.65 SNPs/kb) (**Table 1**).

The genome-wide GBS- and candidate gene-based SNP genotyping overall identified 9045 SNPs showing polymorphism among 172 cultivated and wild chickpea accessions. The 7488 and 1557 SNPs of these, were physically mapped on eight chromosomes and scaffolds of *kabuli* genome, respectively (**Figure 1A**). A maximum number of 1628 SNPs (21.7%) were mapped on *kabuli* chromosome 4, whereas minimum of 321 SNPs (4.3%) were mapped on chromosome 8. The structural annotation of 9045 SNPs identified 5903 (65.3%) and 3142 (34.7%) SNPs in the 1859 genes and intergenic regions of *kabuli* genome (**Figure 1B**). The abundance of gene-based SNPs was observed in the CDS (2385 SNPs, 40.4%), followed by introns (1576, 26.7%) and least (928, 15.7%) in the URRs. In total, 1205 (50.5%) and 1180 (49.5%) coding SNPs in the 756 and 743 genes exhibited synonymous and non-synonymous substitutions, respectively (**Figure 1B**). The relative distribution of GBS-based SNPs, including non-synonymous SNPs physically mapped on eight *kabuli* chromosomes showing polymorphism among 93 cultivated (*desi* and *kabuli*) and 79 wild accessions are depicted individually in a Circos circular ideogram (**Figure 1C**). The functional annotation of 1859 SNP-carrying genes exhibited their maximum correspondence to growth, development and metabolism-related proteins (43%), followed by transcription factors (22%) and signal transduction proteins (11%). The structurally and functionally annotated genome-wide and gene-derived SNPs physically mapped on *kabuli* chromosomes and novel SNP allelic variants mined from diverse seed color-regulating known/cloned and candidate genes can be utilized for various large-scale genotyping applications particularly in complex seed color quantitative trait dissection of chickpea.

Polymorphism and Molecular Diversity Potential of SNPs

The identified 9045 genome-wide GBS- and candidate gene-based SNPs revealed polymorphism among 172 cultivated and wild chickpea accessions with a higher PIC (0.13–0.41 with a mean 0.32) and nucleotide diversity ($\theta\pi$: 2.14, $\theta\omega$: 2.19, and Tajima's D: -1.59) potential (Table S3). The intra-specific polymorphic and nucleotide diversity potential detected by SNPs within wild and cultivated (*desi* and *kabuli*) chickpea species was lower as compared to that of inter-specific polymorphism and nucleotide diversity among species (Table S3).

The molecular diversity potential detected by 9045 SNPs among 172 *desi*, *kabuli* and wild chickpea accessions exhibited a broader range of genetic distances varying from 0.15 to 0.83 with a mean of 0.52. A lower genetic diversity level (mean genetic distance: 0.32) was observed among 93 *desi* and *kabuli* chickpea accessions than that of 79 wild chickpea accessions (0.47). The unrooted neighbor-joining phylogenetic tree construction using 9045 genome-wide and candidate gene-based SNPs depicted a distinct differentiation among 172 cultivated and wild chickpea accessions and further clustered these accessions into two major groups (cultivated and wild) as expected (Figure S2). A relatively higher intra- and inter-specific polymorphism and nucleotide

diversity potential as well as broader allelic (functional) diversity detected by 9045 SNPs among 172 cultivated (*desi* and *kabuli*) and wild chickpea accessions is much higher compared to that of earlier similar documentation (Nayak et al., 2010; Gujaria et al., 2011; Roorkiwal et al., 2013; Varshney et al., 2013). This suggests the utility of our developed genome-wide and gene-based SNP markers for faster screening of preferable diverse accessions/inter-specific hybrids in cross (introgression)-breeding program for chickpea varietal improvement.

Association Analysis for Seed Coat Color Trait

For GWAS and candidate gene-based association mapping, 8837 SNPs showing polymorphism among 93 *desi* and *kabuli* chickpea accessions based on genome-wide GBS- and candidate gene-based SNP genotyping were utilized. The neighbor-joining phylogenetic tree, high-resolution population genetic structure and PCA differentiated all 93 chickpea accessions from each other and clustered into two distinct populations; POP I and POP II representing mostly the *kabuli* and *desi* chickpea accessions with BE and LB/YB seed coat color, respectively (**Figures 2A–C**). The implication of chromosomal LD patterns, including LD decay in GWAS to identify potential genomic loci associated with diverse agronomic traits have been well demonstrated in many crop plants, including chickpea (Zhao et al., 2011; Thudi et al., 2014). The determination of LD patterns in a population of 93 accessions using 7488 SNPs (physically mapped across eight *kabuli* chromosomes) demonstrated a higher LD estimate (r^2 : 0.68) and extended LD decay (r^2 decreased half of its maximum value) approximately at 350–400 kb physical distance in *kabuli* chromosomes (**Figure 2D**). This extensive LD estimates and extended LD decay in the domesticated self-pollinating crop plant like chickpea in contrast to other domesticated self-pollinated crops like rice and soybean (Hyten et al., 2007; Mather et al., 2007; McNally et al., 2009; Huang et al., 2010; Lam et al., 2010; Zhao et al., 2011) is quite expected. This could be due to extensive contribution of four sequential bottlenecks during the domestication of chickpea (Lev-Yadun et al., 2000; Abbo et al., 2003; Berger et al., 2005; Burger et al., 2008; Toker, 2009; Jain et al., 2013; Kujur et al., 2013; Varshney et al., 2013; Saxena et al., 2014b) leading to reduction of genetic diversity in cultivated chickpea than that of other domesticated selfing plant species. The longer LD in chickpea chromosomes requires fewer markers to saturate the genome resulting in low resolution and thus indicating low significant association between the genetic markers and genes controlling the phenotypes in chickpea. In the present context due to insufficient and non-uniform marker coverage on a larger genome of chickpea with low intra-specific polymorphism, the GWAS integrated with candidate gene-based association analysis will be of much relevance in efficient quantitative dissection of complex traits in chickpea (Neale and Savolainen, 2004). This strategy of trait association mapping has been successfully implemented in a domesticated as well as larger genome self-pollinated crop species like barley, where extended LD decay was observed at genome level (Haseneyer et al., 2010; Varshney et al., 2012). Henceforth, the combinatorial approach of GWAS

TABLE 1 | Twenty-eight seed color-related known/candidate gene orthologs of chickpea selected for genetic association mapping.

Crop plants	Gene identity	Characterized known cloned genes			<i>Kabuli</i> chickpea gene homologs				
		Gene accession IDs	References	Gene identity	Chromosomes	Sequenced gene amplicon size (bp)	Coding and upstream regulatory SNPs discovered	Average SNP frequency (SNP/kb)	
Soybean	AR (Anthocyanidin reductase)	Glyma08g06650	Yang et al., 2010	Ca05336	Ca_Kabuli_Chro06	3793	12	3.16	
Soybean	ANS (Leucoanthocyanidin dioxygenase)	Glyma01g42350		Ca02395	Ca_Kabuli_Chro08	2179	6	2.75	
Soybean	DFR1 (Dihydroflavonol 4-reductase 1)	Glyma14g07940		Ca10786	Ca_Kabuli_Chro04	5497	15	2.73	
Soybean	DFR2 (Dihydroflavonol 4-reductase 2)	Glyma17g37060		Ca10787	Ca_Kabuli_Chro04	4719	13	2.75	
Soybean	Myb TF (Myeloblastoma Transcription Factor)	Glyma13g09980		Ca11427	Ca_Kabuli_Chro05	1816	5	2.75	
Soybean	F3H1 (Flavanone 3-hydroxylase 1)	Glyma16g23880		Ca09972	Ca_Kabuli_Chro07	3822	13	3.40	
Soybean	F3H2 (Flavanone 3-hydroxylase 2)	Glyma02g05450		Ca09973	Ca_Kabuli_Chro07	3872	10	2.58	
Arabidopsis	tt5-DFR (Dihydroflavonol 4-reductase)	AT5G42800	Wang et al., 2013	Ca10786	Ca_Kabuli_Chro04	5497	20	3.64	
Arabidopsis	tt4-CHS (Chalcone Synthase)	AT5G13930	Wroblewski et al., 2014	Ca08294	Ca_Kabuli_Chro03	3010	11	3.65	
Arabidopsis	tt5-CHI (Chalcone isomerase)	AT3G55120	Pourcelet et al., 2013	Ca03250	Ca_Kabuli_Chro07	11604	37	3.19	
Arabidopsis	tt7-F3'H (Flavonoid 3'-hydroxylase)	AT5G07990	Routabou et al., 2012	Ca13196	Ca_Kabuli_Chro07	1443	4	2.77	
Arabidopsis	tt10-PPO (Polyphenol oxidase)	AT5G48100	Zhang et al., 2012	Ca23340	Ca_Kabuli_Chro03	3106	10	3.22	
Arabidopsis	tt12-MATE secondary transporter (Multi antimicrobial extrusion)	AT3G59030	Debeaujon et al., 2001	Ca10158	Ca_Kabuli_Chro02	3676	9	2.45	
Arabidopsis	tt15-GT (Glucosyl transferase)	AT7G43620	Focks et al., 1999	Ca04376	Ca_Kabuli_Chro04	6919	21	3.04	
Arabidopsis	tt18-LDOX (Leucoanthocyanidin dioxygenase)	AT4G22880	Liu et al., 2014	Ca02395	Ca_Kabuli_Chro08	2179	4	1.84	
Arabidopsis	tt19-GST (Glutathione S-transferase)	AT5G17220	Xu et al., 2014	Ca03442	Ca_Kabuli_Chro04	1903	3	1.58	
Arabidopsis	ban-ANR (Anthocyanin reductase)	AT7G61720	Bowerman et al., 2012	Ca05336	Ca_Kabuli_Chro06	3793	9	2.37	
Arabidopsis	ana10 (Autoninhibited H ⁺ ATPase isoform 10)	AT7G17260	Baxter et al., 2005	Ca01370	Ca_Kabuli_Chro03	10262	35	3.41	
Arabidopsis	tt1-TF WIP (Wiskott-Aldrich syndrome protein type zinc finger)	AT7G34790	Sagasser et al., 2002	Ca27078	Ca_Kabuli_Scaffold-2269	2324	5	2.15	
Arabidopsis	tt2-TF AtMYB123 (Myeloblastoma transcription factor)	AT5G35550	Chen et al., 2012	Ca12762	Ca_Kabuli_Chro07	2226	5	2.25	
Arabidopsis	tt8-TF AtbHLH042 (Basic helix-loop-helix)	AT4G09820	Chen et al., 2014	Ca14089	Ca_Kabuli_Chro01	2283	6	2.63	
Arabidopsis	tt16-TF MADS AtAGL32 (Agamous like-32)	AT5G223260	Erdmann et al., 2010	Ca15760	Ca_Kabuli_Chro07	3027	8	2.64	
Arabidopsis	ttg1-WD40	AT5G24520	Nguyen et al., 2013	Ca04058	Ca_Kabuli_Chro05	2028	4	1.97	
Arabidopsis	ttg2-TF WRKY	AT2G37260	Pesch et al., 2014	Ca17273	Ca_Kabuli_Chro02	4380	16	3.65	
Arabidopsis	tt9-Unknown	AT3G22430	Ichino et al., 2014	Ca22116	Ca_Kabuli_Chro01	16647	47	2.82	
Medicago	MATE secondary transporter (TRANSPARENT TESTA)	MTR1g108990	Zhao and Dixon, 2009	Ca05556 ^a	Ca_Kabuli_Chro04	3936	10	2.54	
Medicago	MATE secondary transporter (TRANSPARENT TESTA)	MTR1g108840		Ca05557 ^a	Ca_Kabuli_Chro04	5315	19	3.57	
Medicago	MATE secondary transporter (TRANSPARENT TESTA)	MTR1g108810		Ca05558 ^a	Ca_Kabuli_Chro04	4540	15	3.30	

^a Tandem array of three seed color candidate co-orthologous *tt12*-MATE family transporter repeated genes located at the selective sweep target region of *kabuli* chromosome 4 reported by Varshney et al. (2013).

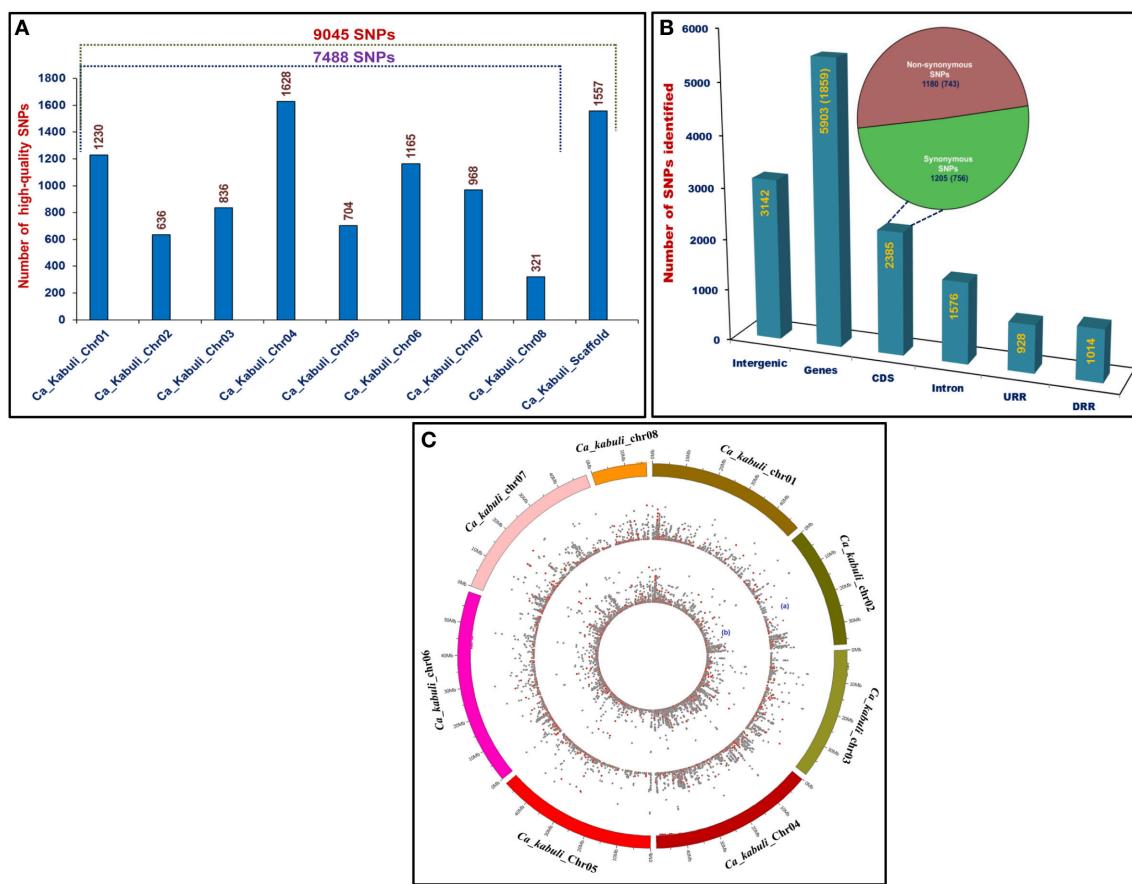


FIGURE 1 | (A) Frequency distribution of 9045 SNPs discovered employing reference *kabuli* genome (eight chromosomes and scaffolds)-based GBS and candidate gene-derived SNP genotyping assays. **(B)** Structural annotation of SNPs in the diverse coding (synonymous and non-synonymous) and non-coding (intron, URR, and DRR) sequence components of genes and intergenic regions of *kabuli* genome. The gene annotation information of *kabuli* genome (Varshney et al., 2013) was considered as reference to infer CDS (coding sequences)/exons, URR (upstream regulatory region) and DRR (downstream regulatory region) sequence components of genes. **(C)** A Circos circular ideogram depicting the relative distribution of 9045 SNPs, including non-synonymous SNPs (marked with red dots) physically mapped on eight *kabuli* chromosomes. The outermost circles illustrate the eight *kabuli* chromosomes coded with different colors, while the two inner circles "a" and "b" represent the distribution of SNPs mined from 93 cultivated (*desi*) and 79 wild chickpea accessions, respectively.

and candidate gene-based association mapping would strengthen the possibility of identifying numerous high-resolution trait-associated valid genomic loci at a genome-wide scale in chickpea.

Keeping that in view, the GWAS and candidate gene-based association analysis was performed by integrating 8837 SNP genotyping data of 93 accessions with their replicated multi-location/years seed coat color field phenotyping information. The normal frequency distribution along with broader seed color phenotypic variation (Figure S3A) and 75% broad-sense heritability (H^2) among 93 accessions belonging to a population was evident. The accessions exclusively exhibiting consistent phenotypic expression of seed coat color across two geographical locations/years (supported by high H^2) were utilized for subsequent SNP marker-trait association. The integration of outcomes from four model-based approaches [TASSEL (GLM and MLM) and GAPIT (EMMA and CMLM)] with a quantile-quantile plot of the expected and observed $-\log_{10} P$ -values at FDR cut-off ≤ 0.05 in GWAS and candidate gene-based association analysis identified 15 genomic loci

showing significant association with seed coat color at a $P \leq 10^{-6}$ (Figures 3A,B, Table 2). This includes eight genome-wide GBS-based and seven candidate gene-derived SNPs. Fourteen seed color-associated genomic loci were physically mapped on seven *kabuli* chromosomes. The rest one SNP locus was represented from scaffold region of *kabuli* genome. One of 15 seed color-associated genomic SNP loci was derived from the intergenic region, whereas rest 14 SNP loci were represented from diverse coding (seven, including six non-synonymous SNPs) as well as non-coding intronic (two SNPs) and URR (five) sequence components of 13 *kabuli* genes (Table 2).

The identified 15 significant SNPs explained (R^2) an average of 28% (ranging 20–43%) seed coat color phenotypic variation in a population of 93 chickpea accessions. The total phenotypic variation for seed coat color explained by the 15 SNPs was 61%. The significant association of 15 multiple SNP loci in more than one genomic region (genes) with seed coat color gave clues for complex genetic inheritance pattern/regulation of this target quantitative trait in chickpea. This association was evident from

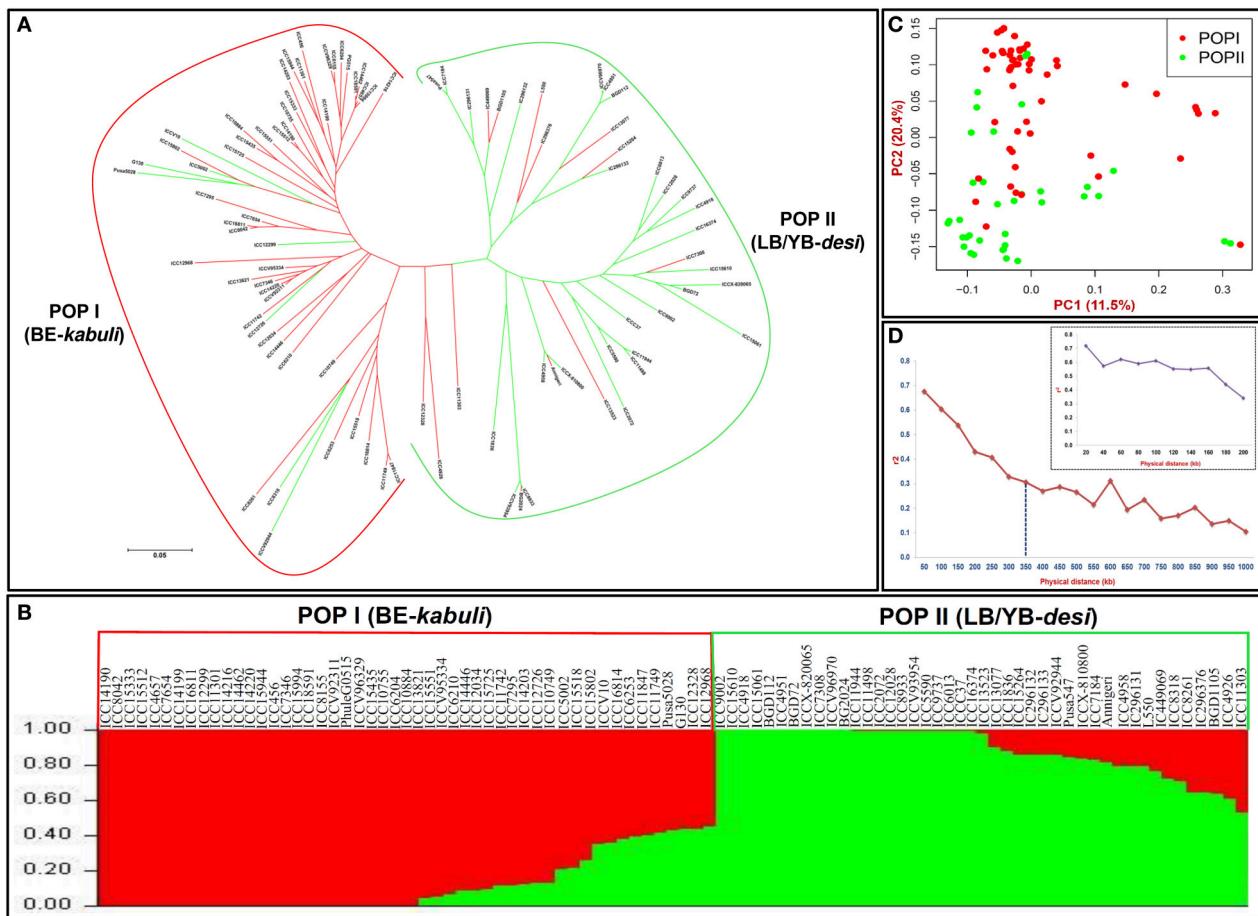


FIGURE 2 | (A) Unrooted phylogenetic tree (Nei's genetic distance). **(B)** Population genetic structure (optimal population number $K = 2$ with two diverse color) and **(C)** Principal component analysis (PCA) using 8837 genome-wide GBS- and candidate gene-based SNPs assigned 93 BE (beige) and LB/YB (light/yellow brown) seed coat color representing *kabuli* and *desi* chickpea accessions mostly into two major populations-POP I and POP II, respectively. In population structure, the accessions represented by vertical bars along the horizontal axis were classified into K color segments based on their estimated membership fraction in each K cluster. In PCA, the PC1 and PC2 explained 7.1 and 17.4% of the total variance, respectively. **(D)** LD decay (mean r^2) measured in a population of 93 cultivated *desi* and *kabuli* chickpea accessions using 7488 SNPs physically mapped on eight *kabuli* chromosomes. The plotted curved line denotes the mean r^2 -values among SNP loci spaced with uniform 50 kb physical intervals from 0 to 1000 kb across chromosomes. The plotted line in uppermost indicates the mean r^2 -values among SNPs spaced with uniform 20 kb physical intervals from 0 to 200 kb on chromosomes.

high chromosomal and population-specific LD estimates with preferable LD decay of trait-associated linked/unlinked multiple SNP loci and further by the observed genetic heterogeneity of seed color traits across two populations. We observed no significant differences concerning the association potential of SNP loci with seed color in different populations despite diverse genetic architecture of seed coat color traits in two populations and along entire population. Therefore, seed coat color-associated genomic loci (genes) identified by us employing both GWAS and candidate gene-based association mapping could have potential for establishing rapid marker-trait linkages and identifying genes/QTLs regulating seed color in chickpea. Seven (three non-synonymous coding and four URR SNPs) SNP loci in the six different known cloned and candidate seed color genes (*tt12*, *tt4*, *tt3*, *tt19*, *ban-ANR*, and *tt18*) had significant association ($P: 1.3 \times 10^{-6}$ to 1.5×10^{-8} with $R^2: 20\text{--}35\%$) with seed coat color trait in chickpea (Table 2). Interestingly,

significant association of non-synonymous and regulatory SNPs identified in the different dormancy and stress-related known/candidate genes (encoding inositol monophosphate and disease resistance proteins) with seed coat color was observed in chickpea. This is consistent with previous studies that reported strong correlation/interactions among genes/QTLs regulating all three important agronomic traits, including seed coat pigmentation, seed dormancy/germination ability and stress tolerance in legumes (Harborne and Williams, 2000; Furukawa et al., 2006; Lepiniec et al., 2006; Kovinich et al., 2012; Saxena et al., 2013; Smýkal et al., 2014). The transcriptional regulation of numerous genes involved in anthocyanin and flavonoid biosynthetic pathways lead to accumulation of diverse major secondary metabolites in specific plant organs, which are known to play most important role in collective alteration of seed coat pigmentation/coloration, dormancy and defense (stress) response in crop plants

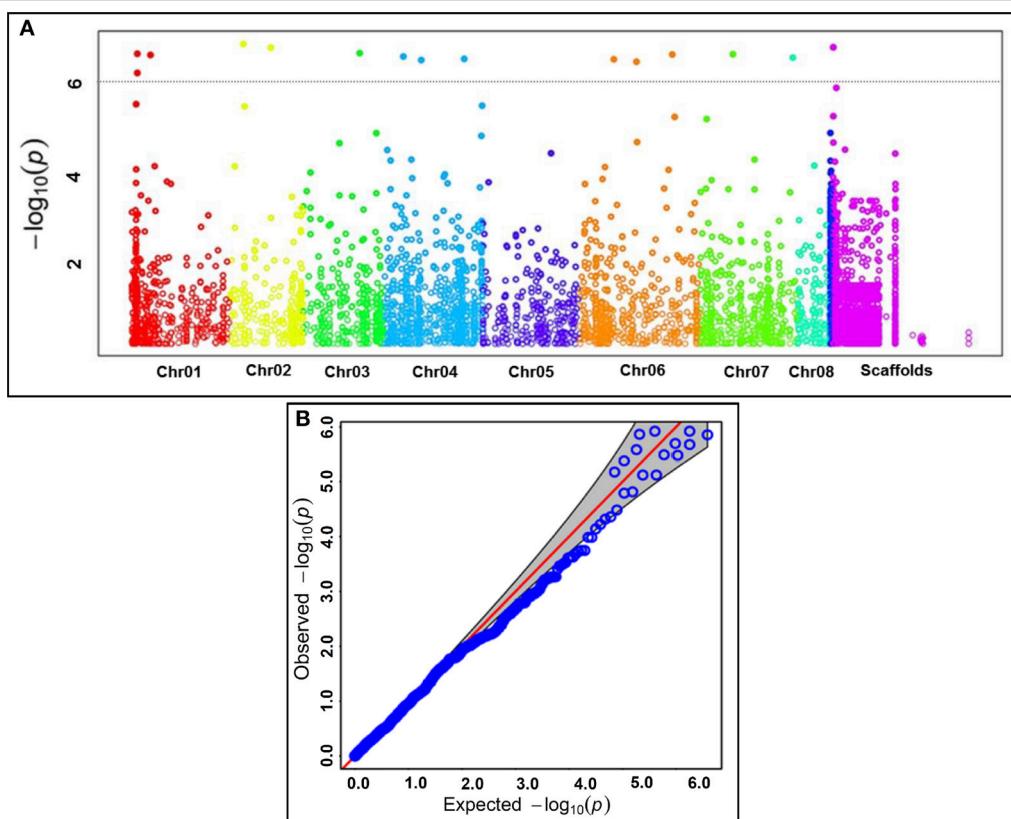


FIGURE 3 | (A) GWAS-derived Manhattan plot showing significant P -values (measured integrating GLM, MLM, EMMA, and CMLM) associated with seed coat color trait using 9045 genome-wide GBS- and candidate gene-based SNPs. The x-axis represents the relative density of reference genome-based SNPs physically mapped on eight chromosomes and scaffolds of *kabuli* genome. The y-axis indicates the $-\log_{10}(P)$ -value for significant association of SNPs with seed color trait. The SNPs showing significant association with seed color at cut-off $P < 1 \times 10^{-6}$ are marked with dotted lines. **(B)** Quantile-quantile plots illustrating the comparison between expected and observed $-\log_{10}(P)$ -values with FDR cut-off <0.05 to detect significant genomic loci (genes) associated with seed color trait in chickpea.

(Gore et al., 2002; Benitez et al., 2004; Senda et al., 2004; Isemura et al., 2007; Caldas and Blair, 2009; Diaz et al., 2010; Kongjaimun et al., 2012). Notably, one non-synonymous SNP locus (T/G) in the CDS of the MATE (multidrug and toxic compound extrusion) secondary transporter gene (Ca18123, mapped on *kabuli* chromosome 2) exhibiting strong association ($P: 1.7 \times 10^{-9}$ with $R^2: 43\%$) with seed color than that of other 14 seed color-associated SNPs was selected as a potential candidate for further analysis. The discovery of non-synonymous and regulatory SNP loci particularly in coding and URRs of known/candidate genes associated with seed color traits signifies their functional relevance in rapid seed color trait-regulatory gene identification and characterization in chickpea. The non-synonymous (amino acid substitutions) and regulatory SNPs in the genes have known to alter their transcriptional regulatory mechanism for controlling diverse traits of agronomic importance, including seed/grain size and weight in rice (Fan et al., 2006; Mao et al., 2010; Li et al., 2011; Zhang et al., 2012) and chickpea (Kujur et al., 2013, 2014; Bajaj et al., 2015; Saxena et al., 2014b).

Two candidate seed color-associated MATE secondary transporter genes [including one (Ca05557) reported earlier at

the selective sweep region of *kabuli* chromosome 4, (Varshney et al., 2013)] identified by use of both candidate gene-based association mapping and GWAS were selected (Table S4) to compare and deduce the nucleotide diversity measures as well as the direction and magnitude of natural selection acting between these genes. These genes revealed a similar trend of nucleotide diversity ($\theta\pi$ and $\theta\omega$) and Tajima's D-based selection pressure occurring on LB/YB and BE seed colored *desi* and *kabuli* chickpea accessions, respectively (Table S4). Interestingly, the degree of nucleotide diversity was reduced by 61% in the LB/YB seed colored cultivated *desi* accessions than that of *kabuli* chickpea accessions with BE seed color. The extreme reduction of nucleotide diversity level ($\theta\pi: 0.73\text{--}0.79$ and $\theta\omega: 0.77\text{--}0.82$) and strong purifying selection along with reduced Tajima's D (-2.89 to -2.63) specifically in LB/YB seed colored cultivated *desi* chickpea accessions compared with BE seed colored *kabuli* accessions ($\theta\pi: 1.93\text{--}1.96$, $\theta\omega: 1.95\text{--}1.98$ and D: $1.82\text{--}1.90$) was evident (Table S4). These findings provide definite evidence regarding signature of strong purifying selection in LB/YB seed colored *desi* accessions in contrast to BE seed colored *kabuli* chickpea accessions, which is in line with enduring purifying selection for seed coat/pericarp color traits in crop plants,

TABLE 2 | Fifteen seed coat color-regulating genomic loci (genes) identified by GWAS and candidate gene-based association mapping.

SNP IDs ^a	Kabuli chromosomes/scaffolds	SNPs physical positions (bp)	SNPs (Reference/alternate alleles)	Genes/gene-associated targets ^b	Structural annotation	Association analysis	
						P	R ² (%)
CWSNP151	1	1858509	(C/T)	Ca00233 (Inositol monophosphate)	CDS (Syn)	2.2×10^{-7}	25
CWSNP156	1	1859914	(G/A)	Ca00233 (Inositol monophosphate)	Intron	2.0×10^{-7}	27
CWSNP208	1	2053788	(A/G)	Ca00260 (Disease resistance protein)	CDS (NSyn) Threonine (ACG)-Alanine (GCG)	1.8×10^{-7}	22
CWSNP1275	2	7701199	(T/G)	Ca18123 (<i>tt12</i> -MATE secondary transporter)	CDS (NSyn) Valine (GTG)-Glycine (GGG)	1.7×10^{-9}	43
gSNP1	2	33782378	(C/A)	Ca10158 (<i>tt12</i> -MATE secondary transporter)	CDS (NSyn) Alanine (GCT)-Aspartic acid (GAT)	1.2×10^{-7}	20
gSNP2	3	26031490	(A/G)	Ca08294 (<i>tt4</i> -Chalcone synthase)	URR	1.5×10^{-7}	33
gSNP5	4	48021943	(G/T)	Ca10786 (<i>tt3</i> -Dihydroflavonol-4-reductase)	CDS (NSyn) Glycine (GGG)-Valine (GTG)	1.7×10^{-7}	29
gSNP3	4	7931791	(T/C)	Ca03442 (<i>tt19</i> -Glutathione-S-reductase)	URR	1.4×10^{-7}	30
gSNP4	4	15386163	(T/G)	Ca05557 (<i>tt12</i> -MATE secondary transporter)	CDS (NSyn) Serine (TCA)-Alanine (GCA)	1.3×10^{-6}	24
gSNP6	6	15514534	(C/G)	Ca05336 (ban-ANR-Anthocyanin reductase)	URR	2.1×10^{-8}	35
CWSNP5434	6	28325392	(C/T)	Ca25968 (Ribonuclease T2)	CDS (NSyn) Serine (TCA)-Leucine (TTA)	2.4×10^{-7}	27
CWSNP5762	6	56793021	(G/T)	–	Intergenic	1.5×10^{-7}	21
CWSNP6261	7	14281993	(G/T)	Ca23042 (domain of unknown function DUF231)	Intron	2.1×10^{-7}	38
gSNP7	8	2062481	(G/A)	Ca02395 (<i>tt18</i> -Leucoanthocyanidin dioxygenase)	URR	2.1×10^{-6}	26
CWSNP8591	Scaffold93	645442	(A/C)	Ca23085 (Integrase)	URR	1.9×10^{-7}	23

^aCWSNP (*cultivated wild SNP*) and gSNP (*gene-derived SNPs*).

CDS, coding sequence; NSyn, non-synonymous; Syn, synonymous; URR, upstream regulatory region.

^bDetails regarding SNPs are mentioned in the Table S2.

gSNP: gene-derived SNPs associated with seed coat color trait identified by candidate gene-based association mapping.

including chickpea (Sweeney et al., 2007; Varshney et al., 2013). Therefore, seed color in chickpea vis-à-vis other crop plants can be considered as a target trait for both domestication and artificial breeding (Sweeney et al., 2007; Hossain et al., 2011; Meyer et al., 2012).

Validation of Seed Color-associated Genomic Loci through QTL Mapping

We constructed a high-density intra-specific genetic linkage map (ICC 12299 \times ICC 8261) by integrating 415, including 382 SNP and 33 previously reported parental polymorphic SSR markers across eight chickpea LGs (LG1 to LG8). This genetic map covered a total map length of 1065.7 cM, with a mean inter-marker distance of 2.57 cM (Table S5). The most and least saturated genetic map was LG1 (average inter-marker distance: 1.92 cM) and LG5 (4.0 cM), respectively. The mean map density (average inter-marker distance: 2.57 cM) observed in our constructed intra-specific genetic linkage map

was higher/comparable with earlier documentation (Cho et al., 2002, 2004; Cobos et al., 2005; Radhika et al., 2007; Kottapalli et al., 2009; Gaur et al., 2011; Kujur et al., 2013, 2014; Sabbavarapu et al., 2013; Stephens et al., 2014; Varshney et al., 2014). Therefore, this constructed high-density intra-specific genetic linkage map have potential to be utilized as a reference for rapid targeted mapping of potential genes/QTLs governing diverse agronomic traits, including seed coat color in chickpea. The bi-directional transgressive segregation-based normal frequency distribution (Figure S3B), including a significant variation in seed coat color trait along with 78% H^2 in this developed mapping population (ICC 12299 \times ICC 8261) was observed. Two principal seed coat color (BE and LB/YB) groupings; primary (PC1) and secondary (PC2) components accounted for 52 and 48% of the total seed coat color variance, respectively in 190 RIL mapping individuals and parental accessions were evident (Figure S3C). This continuous color distribution inferred the quantitative genetic inheritance pattern of seed

coat color trait in the mapping population and thus the developed bi-parental mapping population could be utilized for complex seed color trait dissection through QTL mapping in chickpea.

The QTL mapping was performed by correlating the genotyping data of 415 mapped SNP and SSR markers with multilocation/years seed coat color field phenotyping data of the RIL mapping individuals and parental accessions. This analysis identified five major (LOD: 5.6–11.5) genomic regions harboring robust QTLs (validated across two locations/years) controlling seed coat color traits, which were mapped on five *kabuli* chromosomes (Table S6). The phenotypic variation explained (PVE) for seed color by individual QTL (R^2) varied from 20.1 to 38.7%. The PVE estimated for all five major QTLs was 39.4%. All these QTLs revealed additive gene effects (ranging -3.9 to -2.5), inferring the effective contribution of ICC 8261 alleles at these loci for BE seed coat color trait. The five SNPs in the genes showing tight linkage with all five major seed color-associated robust QTLs (*CaqSC2.1*, *CaqSC4.1*, *CaqSC4.2*, *CaqSC6.1*, and *CaqSC7.1*) had high seed color trait association potential based on our GWAS and candidate gene-based association analysis (Table 2; Table S6). Remarkably, strong association potential of one non-synonymous SNP (T/G) identified in the CDS of the MATE secondary transporter gene as compared to other genomic loci associated with seed coat color trait was ascertained both by QTL mapping (R^2 : 35.6–38.7%) and genetic association analysis (R^2 : 43%) (Table 2). Therefore, we selected MATE secondary transporter gene as a target candidate for seed coat color trait regulation by its further validation through differential expression profiling and molecular haplotyping in chickpea.

To determine the potential and novelty of seed coat color-regulating 15 genomic loci and five major QTLs detected by association and QTL mapping, respectively, the markers linked/flanking the seed color-associated known QTLs/genes (reported earlier in QTL mapping studies, Hossain et al., 2011) were selected for their validation in seed color-specific 93 diverse *desi* and *kabuli* chickpea accessions and mapping population under study. These comparative analyses revealed correspondence of one known QTL (*CaqSC2.1*) regulating seed coat color between past (SSR marker interval: TA194-TR19) and our present studies based on congruent flanking [CWSNP 1273 (25.2 cM)-CWSNP1277 (30.2 cM)]/linked marker (CWSNP 1275) physical/genetic positions on chromosome (LG) 2. Henceforth, 14 genomic loci and 4 QTLs regulating seed color identified by us employing high-resolution GWAS/candidate gene-based association and QTL mapping are novel and exhibited population-specific genetic inheritance pattern for seed color trait regulation in chickpea. Further validation of these functionally relevant molecular tags is required in diverse genetic backgrounds and/or through fine mapping/map-based positional cloning prior to their utilization in marker-assisted genetic improvement of chickpea for desirable seed color.

Validation of Seed Color-associated Genes through Differential Expression Profiling

To determine the regulatory pattern of genes for seed coat color, thirteen, including five seed color-associated genes validated

both by association analysis and QTL mapping were utilized for differential expression profiling. The RNA isolated from vegetative leaf tissues, seeds without seed coats and seed coats scrapped from the immature and mature seeds of 11 contrasting chickpea accessions as well as mapping parents and two homozygous RIL individuals representing two predominant seed coat color types, LB/YB (*desi*: ICCX810800, *desi*: ICC 12299, *desi*: ICCV10, *desi*: ICC 4958, *desi*: ICC 15061, *kabuli*: ICC 6204, and *kabuli*: Annigeri) and BE (*kabuli*: ICC 20268, *kabuli*: ICC 8261, *kabuli*: Phule G0515, and *desi*: ICC 4926) was amplified with these gene-specific primers using semi-quantitative and quantitative RT-PCR assays (Table S7). Five genes, including four seed color known cloned genes (*tt4*-chalcone synthase, *tt3*-dihydroflavonol-4-reductase, *ban*-ANR-anthocyanin reductase, and *tt19*-glutathione-s-reductase) showed seed coat-specific differential (>2.5 -fold up- and down-regulation, $P \leq 0.01$) expression [compared with leaves and seeds (without seed coats)] specifically in two seed developmental stages of LB/YB and BE seed colored *desi* and *kabuli* accessions (Figure S4). This suggests that differential regulation and accumulation of transcripts encoded by these four known genes might be essential in synthesis of proanthocyanidin and anthocyanin biosynthetic enzymes in chickpea seed coats for its coloration/pigmentation (Lepiniec et al., 2006; Zhao and Dixon, 2009; Zhao et al., 2010).

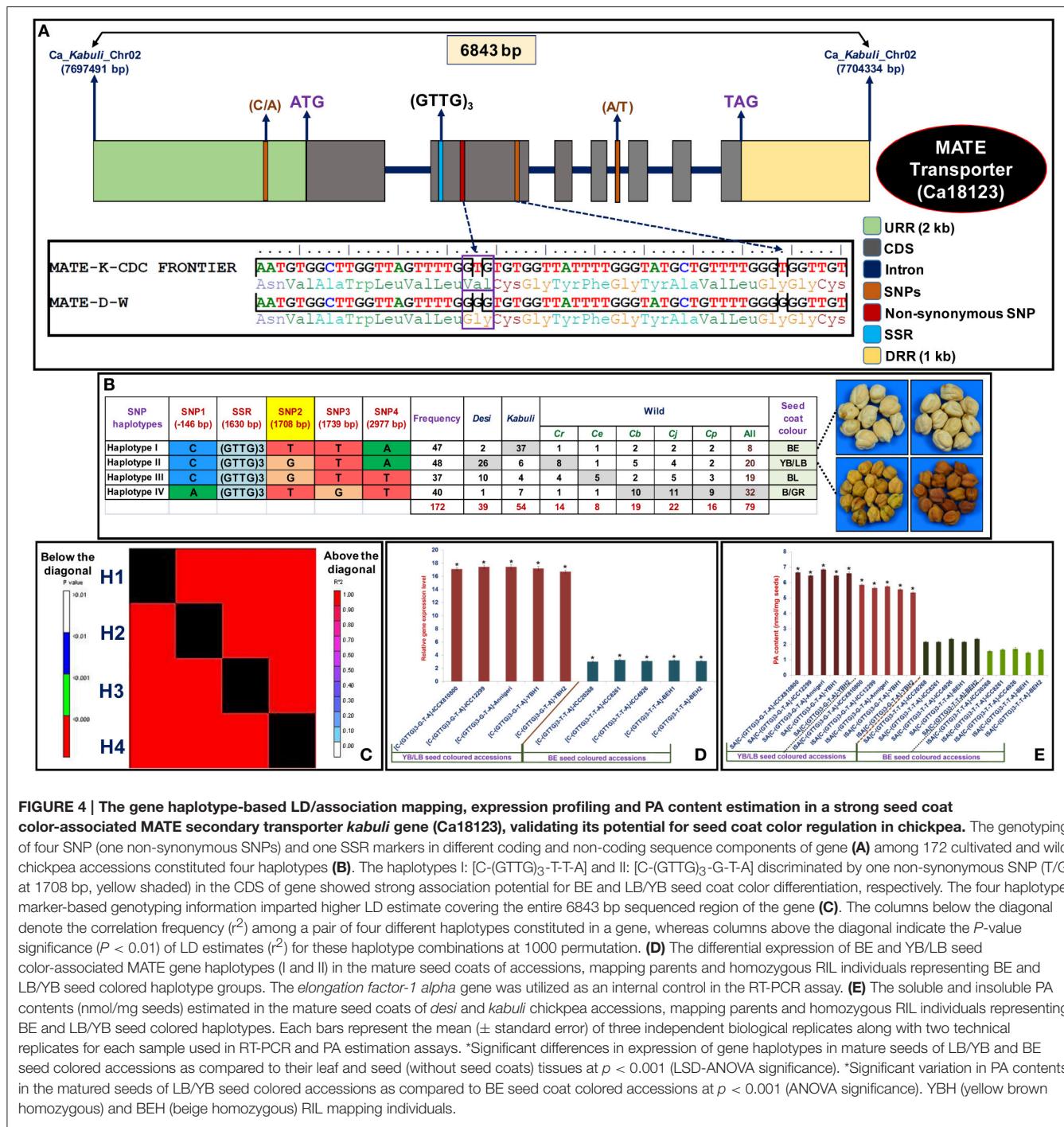
Remarkably, a seed coat-specific MATE secondary transporter gene (validated by both association and QTL mapping) showing strong association with seed coat color trait exhibited its higher differential down-regulation (~ 5 –7-folds, $P \leq 0.001$) in the mature/immature seed coats of BE seed colored chickpea accessions (ICC 20268, ICC 8261, Phule G0515, and ICC 4926) as compared to that of mature/immature seed coats of LB/YB seed colored accessions (ICCX810800, ICC 12299, ICCV10, ICC 4958, ICC 15061, ICC 6204, and Annigeri) (Figure S4). A pronounced down regulation (~ 13 -folds, $P \leq 0.001$) of this seed coat-specific gene in the mature seed coats of all four BE seed colored chickpea accessions in contrast to immature seed coats of all seven LB/YB seed colored accessions was observed. However, differential down-regulation (~ 3 -folds, $P \leq 0.001$) of MATE gene in the mature seed coats than the immature seed coats of the accessions within an individual LB/YB and BE seed coat colored group was evident (Figure S4). The reduction of chickpea MATE gene transcript level specifically in the mature seed coat parallels with differential expression pattern that observed previously using *TT12* and *MATE1* genes of *Arabidopsis* and *Medicago*, respectively (Marinova et al., 2007; Pang et al., 2008; Zhao and Dixon, 2009; Zhao et al., 2010). These outcomes over imply the functional significance of MATE gene in possible transcriptional regulation of seed color trait in chickpea.

Molecular Haplotyping of a Strong Seed Color-regulating Gene

For molecular haplotyping of a strong seed color-associated gene (validated by association analysis, QTL mapping and expression profiling), the 6843 bp cloned amplicon covering the entire 2 kb URR, 7 exons, 1 kb DRR and 6 intronic regions of MATE secondary transporter *kabuli* gene (Ca18123)

were sequenced and compared among 93 cultivated and 79 wild chickpea accessions. These analyses identified four SNPs, including one URR and two coding SNPs as well as one coding SSR in the MATE gene (**Figure 4A**, Table S8). This includes one missense non-synonymous coding SNPs (T/G) showing valine (GTG) to glycine (GGG) amino acid substitution in the gene. The gene-based haplotype analysis by integrating the genotyping data of four SNPs and one SSR among 172 accessions constituted four haplotypes (**Figure 4B**). The haplotype-specific

LD mapping and association analysis using four marker-based haplotypes identified in a *MATE* gene inferred strong association potential ($P: 2.3 \times 10^{-11}$ with $R^2: 45\%$) of the gene with seed coat color trait. Moreover, we observed a significant higher degree of LD ($r^2 > 0.85$ with $P < 1.1 \times 10^{-6}$) resolution across the entire 6843 bp sequenced region of this strong seed color-associated gene (**Figure 4C**). Remarkably, two specific haplotypes, haplotype I [C-(GTTG)₃-T-T-A] and haplotype II [C-(GTTG)₃-G-T-A] affected by one functional



non-synonymous coding SNPs (T/G) in the *MATE* gene had strong association potential for BE and YB/LB seed coat color differentiation, respectively (**Figure 4B**). The haplotypes I and II than other two haplotypes (III and IV) identified in the gene was represented most commonly by YB/LB and BE seed colored *desi* (66.7%) and *kabuli* (68.5%) accessions, respectively. A higher down-regulated expression (~5-fold) of BE seed color-specific gene haplotype I was observed in mature seed developmental stages of BE seed colored three contrasting chickpea accessions (ICC 20268, ICC 8261, and ICC 4926) as well as mapping parents and homozygous RIL individuals compared with three YB/LB seed colored accessions (ICCX 810800, ICC 12299, and Annigeri), mapping parents and homozygous mapping individuals (**Figure 4D**). To correlate the differential expression/regulatory pattern of *MATE* gene-encoding transcript with proanthocyanidins (PA) synthesis/accumulation (Lepiniec et al., 2006; Marinova et al., 2007; Zhao and Dixon, 2009; Zhao et al., 2010), we estimated the soluble and insoluble PA content (three replicates) in aforementioned mature seeds of six contrasting accessions as well as mapping parents and homozygous RIL individuals representing BE and YB/LB-specific haplotype groups I and II, respectively. This detected a significant reduction of both soluble and insoluble PA contents in the mature seed coats of BE seed colored accessions (soluble and insoluble PA: 2.1 and 1.6 nmol/mg seeds, respectively) than that of LB/YB (6.7 and 5.9 nmol/mg seeds) seed colored accessions (**Figure 4E**). Collectively, a strong association potential of *MATE* gene with seed color trait was ensured by integrating GWAS and candidate gene-based association analysis with QTL mapping, differential expression profiling, SNP-SSR marker-based high-resolution gene-specific LD mapping/haplotyping, haplotype-specific transcript profiling and PA estimation (soluble and insoluble) assay. These analyses identified novel natural allelic variant (T/G) showing non-synonymous amino acid substitution (valine to glycine) and potential haplotypes [C-(GTTG)₃-T-T-A] and [C-(GTTG)₃-G-T-A] in a *MATE* gene regulating BE and YB/LB seed coat color differentiation, respectively in *kabuli* and *desi* chickpea accessions. The implication of such combinatorial approach for rapid delineation of functionally relevant genes/QTLs controlling diverse agronomic traits, including starch biosynthesis, seed pericarp coloration/pigmentation, seed shattering and seed setting rate in rice as well as seed size/weight and pod number in chickpea have been well understood (Konishi et al., 2006; Sweeney et al., 2007; Kujur et al., 2013, 2014; Li et al., 2013; Saxena et al., 2014a; Bajaj et al., 2015). Therefore, natural allelic variants and haplotypes identified in a strong seed color-associated *MATE* gene by use of an integrated genomic approach can have potential to decipher its complex transcriptional regulatory function of seed coat coloration during seed development and subsequently for marker-assisted genetic enhancement of chickpea for preferred seed color.

We observed sharing of all four SNP-SSR marker-based haplotypes among 93 cultivated *desi* and *kabuli* as well as 79 wild accessions representing primary, secondary and tertiary gene pools (**Figure 4B**). The B/GR and BE seed color-associated haplotypes IV and I were represented most (40.5%) and least

(10.1%) in the wild chickpea accessions, respectively. However, the B/GR-associated haplotype IV was primarily represented by the wild chickpea accessions (38%) of secondary gene pool (*C. bijugum*, *C. judaicum*, and *C. pinnatifidum*) (**Figure 4B**). The wild accessions of *C. reticulatum* from primary gene pool had maximum YB/LB (10.1%) seed color-associated haplotype II. A maximum haplotypes sharing of BE seed color-associated haplotype I in cultivated *kabuli* accessions with YB/LB seed color-associated haplotype II in *desi* accessions following YB/LB seed color-associated haplotype II between *desi* and wild *C. reticulatum* accessions was evident (**Figure 4B**). The B/GR seed color-associated haplotype IV gave evidence for its putative recombination event during chickpea domestication. These marker haplotypes sharing analyses reflected single origin of "G" functional SNP allele primarily from *C. reticulatum*, and subsequently it was introgressed into most of the LB/YB seed colored *desi* accessions as compared to *kabuli* accessions with BE seed coat color. The nucleotide diversity ($\theta\pi$ and $\theta\omega$) and Tajima's D estimation (Table S3) further inferred the occurrence of very strong purifying selection (with reduced D) in favor of retention of LB/YB seed coat color-associated "G" SNP locus and haplotype II in the *MATE* secondary transporter gene specifically in *desi* and wild *C. reticulatum* accessions toward assortment of more preferential LB/YB seed coat color in chickpea. These findings may be a result of domestication bottlenecks coupled with artificial selection/modern breeding efforts that are constantly practiced during the genetic improvement program of chickpea accessions for diverse seed coat color characteristics of high consumer preference and trade value. These findings infer that the natural allelic/haplotype variation discovered in the *MATE* gene are possibly associated with seed coat color trait evolution and therefore, the seed color is expected to represent an important component of domestication trait in chickpea.

The chickpea *MATE* gene is a homologs of *Arabidopsis TT12* (transparent testa 12) and *Medicago MATE1* genes encodes a membrane protein of multidrug and toxic compound extrusion (MATE) secondary transporter (Debeaujon et al., 2001). The detailed phenotypic characterization of *tt12* and *mate1* mutants and complementation analysis of their corresponding genes in *Arabidopsis* and *Medicago*, respectively inferred that these tonoplast late biosynthetic genes mediate transportation of precursors for proanthocyanidin (PA) biosynthesis into the vacuoles of seed coat endothelial cells in developing immature seeds for their coloration/pigmentation (Lepiniec et al., 2006; Marinova et al., 2007; Zhao and Dixon, 2009; Zhao et al., 2010). The use of combinatorial approach of GWAS, QTL mapping, transcript profiling, gene-based marker haplotyping/LD mapping and PA estimation assay indicate that BE seed color-associated haplotype I [C-(GTTG)₃-T-T-A] identified in a *MATE* gene might be involved in down-regulation/lower expression and accumulation of its encoding transcript, which led to reduced synthesis of both soluble and insoluble PAs in pericarps of BE seed colored mature seeds than that of YB/LB seed colored mature seeds with haplotype II [C-(GTTG)₃-G-T-A]. These results are in agreement with earlier observations on lower accumulation of PAs

(soluble/insoluble) in the vacuole endothelial cells of mature seed coat during later stages of seed development and their involvement in differential seed coat/pericarp coloration in crop plants (Abrahams et al., 2003; Kitamura et al., 2004; Liu et al., 2014). However, detailed molecular characterization of *MATE* secondary transporter gene is required to decipher its underlying regulatory mechanism, evolutionary cues and biochemical interactions governing diverse seed coat coloration in *desi*, *kabuli* and wild accessions during chickpea domestication. Therefore, the strong seed color-regulatory *MATE* gene once functionally well characterized, can be utilized in marker-assisted seed color breeding for developing varieties with improved seed coat color of consumer preference and trade value in chickpea.

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, 1081–1087. doi: 10.1071/FP03084

Abrahams, S., Lee, E., Walker, A. R., Tanner, G. J., Larkin, P. J., and Ashton, A. R. (2003). The *Arabidopsis* *TDS4* gene encodes leucoanthocyanidin dioxygenase (LDOX) and is essential for proanthocyanidin synthesis and vacuole development. *Plant J.* 35, 624–636. doi: 10.1046/j.1365-313X.2003.01834.x

Bajaj, D., Saxena, M. S., Kujur, A., Das, S., Badoni, S., Tripathi, S., et al. (2015). Genome-wide conserved non-coding microsatellite (CNMS) marker-based integrative genetical genomics for quantitative dissection of seed weight in chickpea. *J. Exp. Bot.* 66, 1271–1290. doi: 10.1093/jxb/eru478

Baxter, I. R., Young, J. C., Armstrong, G., Foster, N., Bogenschutz, N., Cordova, T., et al. (2005). A plasma membrane H-ATPase is required for the formation of proanthocyanidins in the seed coat endothelium of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2649–2654. doi: 10.1073/pnas.0406377102

Benitez, E. R., Funatsuki, H., Kaneko, Y., Matsuzawa, Y., Bang, S. W., and Takahashi, R. (2004). Soybean maturity gene effects on seed coat pigmentation and cracking in response to low temperatures. *Crop Sci.* 44, 2038–2042. doi: 10.2135/cropsci2004.2038

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* 57, 289–300.

Berger, J. D., Buck, R., Henzell, J. M., and Turner, N. C. (2005). Evolution in the genus *Cicer* vernalisation response and low temperature pod set in chickpea (*C. arietinum* L.) and its annual wild relatives. *Aust. J. Agric. Res.* 56, 1191–1200. doi: 10.1071/AR05089

Bowerman, P. A., Ramirez, M. V., Price, M. B., Helm, R. F., and Winkel, B. S. (2012). Analysis of T-DNA alleles of flavonoid biosynthesis genes in *Arabidopsis* ecotype Columbia. *BMC Res. Notes* 5:485. doi: 10.1186/1756-0500-5-485

Burger, J. C., Champan, M. A., and Burke, J. M. (2008). Molecular insights into the evolution of crop plants. *Am. J. Bot.* 95, 113–122. doi: 10.3732/ajb.95.2.113

Caldas, G. V., and Blair, M. W. (2009). Inheritance of seed condensed tannins and their relationship with seed-coat color and pattern genes in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 119, 131–142. doi: 10.1007/s00122-009-1023-4

Cassells, J., and Caddick, L. (2000). Harvest timing critical for chickpea quality. *Farming Ahead* 107, 41–42.

Chen, M., Wang, Z., Zhu, Y., Li, Z., Hussain, N., Xuan, L., et al. (2012). The Effect of TRANSPARENT TESTA2 on seed fatty acid biosynthesis and tolerance to environmental stresses during young seedling establishment in *Arabidopsis*. *Plant Physiol.* 160, 1023–1036. doi: 10.1104/pp.112.202945

Chen, M., Xuan, L., Wang, Z., Zhou, L., Li, Z., Du, X., et al. (2014). TRANSPARENT TESTA8 inhibits seed fatty acid accumulation by targeting several seed development regulators in *Arabidopsis*. *Plant Physiol.* 165, 905–916. doi: 10.1104/pp.114.235507

Cho, S., Chen, W., and Muehlbauer, F. J. (2004). Pathotype-specific genetic factors in chickpea (*Cicer arietinum* L.) for quantitative resistance to *Ascochyta* blight. *Theor. Appl. Genet.* 109, 733–739. doi: 10.1007/s00122-004-1693-x

Cho, S., Kumar, J., Shultz, J. F., Anupama, K., Tefera, F., and Muehlbauer, F. J. (2002). Mapping genes for double podding and other morphological traits in chickpea. *Euphytica* 125, 285–292. doi: 10.1023/A:1020872009306

Cobos, M. J., Fernandez, M., Rubio, J., Kharrat, M., Moreno, M. T., Gil, J., et al. (2005). Linkage map of chickpea (*Cicer arietinum* L.) based on populations from *Kabuli* × *Desi* crosses: Location of genes for resistance to *Fusarium* wilt race 0. *Theor. Appl. Genet.* 110, 1347–1353. doi: 10.1007/s00122-005-1980-1

Debeaujon, I., Peeters, A. J. M., Léon-Kloosterziel, K. M., and Koornneef, M. (2001). The TRANSPARENT TESTA12 gene of *Arabidopsis* encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. *Plant Cell* 13, 853–872. doi: 10.1101/tpc.13.4.853

Deokar, A. A., Ramsay, L., Sharpe, A. G., Diapari, M., Sindhu, A., Bett, K., et al. (2014). Genome wide SNP identification in chickpea for use in development of a high density genetic map and improvement of chickpea reference genome assembly. *BMC Genomics* 15:708. doi: 10.1186/1471-2164-15-708

Diaz, A. M., Caldas, G. V., and Blair, M. W. (2010). Concentrations of condensed tannins and anthocyanins in common bean seed coats. *Food Res. Int.* 43, 595–601. doi: 10.1016/j.foodres.2009.07.014

Dixon, R. A., Achnina, L., Kota, P., Liu, C. J., Reddy, M. S. S., and Wang, L. (2002). The phenylpropanoid pathway and plant defence-a genomics perspective. *Mol. Plant Pathol.* 3, 371–390. doi: 10.1046/j.1364-3703.2002.00131.x

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379. doi: 10.1371/journal.pone.0019379

Erdmann, R., Gramzow, L., Melzer, R., Theissen, G., and Becker, A. (2010). GORDITA (*AGL63*) is a young paralog of the *Arabidopsis thaliana* B (sister) MADS box gene *ABS* (*TT16*) that has undergone neofunctionalization. *Plant J.* 63, 914–924. doi: 10.1111/j.1365-313X.2010.04290.x

Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., et al. (2006). GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* 112, 1164–1171. doi: 10.1007/s00122-006-0218-1

Ferraro, K., Jin, A. L., Nguyen, T. D., Reinecke, D. M., Ozga, J. A., and Ro, D. K. (2014). Characterization of proanthocyanidin metabolism in pea (*Pisum sativum*) seeds. *BMC Plant Biol.* 14:238. doi: 10.1186/s12870-014-0238-y

Focks, N., Sagasser, M., Weisshaar, B., and Benning, C. (1999). Characterization of *tt15*, a novel transparent testa mutant of *Arabidopsis thaliana* (L.) Heynh. *Planta* 208, 352–357. doi: 10.1007/s004250050569

Forkmann, G. and Martens, S. (2001). Metabolic engineering and applications of flavonoids. *Curr. Opin. Biotechnol.* 12, 155–160. doi: 10.1016/S0958-1669(00)00192-0

Furukawa, T., Maekawa, M., Oki, T., Suda, I., Iida, S., Shimada, H., et al. (2006). The *Rc* and *Rd* genes are involved in proanthocyanidin

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support for this research study provided by a research grant from the Department of Biotechnology (DBT), Government of India (102/IFD/SAN/2161/2013-14). SD and RR acknowledge the DBT and CSIR (Council of Scientific and Industrial Research) for Junior/Senior Research Fellowship awards.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.00979>

synthesis in rice pericarp. *Plant J.* 49, 91–102. doi: 10.1111/j.1365-313X.2006.02958.x

Gaur, R., Azam, S., Jeena, G., Khan, A. W., Choudhary, S., Jain, M., et al. (2012). High-throughput SNP discovery and genotyping for constructing a saturated linkage map of chickpea (*Cicer arietinum* L.). *DNA Res.* 19, 357–373. doi: 10.1093/dnares/dss018

Gaur, R., Sethy, N. K., Choudhary, S., Shokeen, B., Gupta, V., and Bhatia, S. (2011). Advancing the STMS genomic resources for defining new locations on the intraspecific genetic linkage map of chickpea (*Cicer arietinum* L.). *BMC Genomics* 12:117. doi: 10.1186/1471-2164-12-117

Golam Masum Akond, A. S. M., Khandaker, L., Berthold, J., Gates, L., Peters, K., Delong, H., et al. (2011). Anthocyanin, total polyphenols and antioxidant activity of common bean. *Am. J. Food. Technol.* 6, 385–394. doi: 10.3923/ajft.2011.385.394

Gore, M. A., Hayes, A. J., Jeong, S. C., Yue, Y. G., Buss, G. R., and Saghai Maroof, M. A. (2002). Mapping tightly linked genes controlling potyvirus infection at the *Rsv1* and *Rpv1* region in soybean. *Genome* 45, 592–599. doi: 10.1139/g02-009

Graham, J., Matassa, V., Panozzo, J., and Starick, N. (2001). “Genotype and environment interaction for whole grain color in chickpea,” in *4th European Conference on Grain Legumes* (France), 372–373.

Gujaria, N., Kumar, A., Dauthal, P., Dubey, A., Hiremath, P., Bhanu Prakash, A., et al. (2011). Development and use of genic molecular markers (GMMs) for construction of a transcript map of chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* 122, 1577–1589. doi: 10.1007/s00122-011-1556-1

Harborne, J. B., and Williams, C. A. (2000). Advances in flavonoid research since 1992. *Phytochemistry* 55, 481–504. doi: 10.1016/S0031-9422(00)00235-1

Haseneyer, G., Stracke, S., Piepho, H. P., Sauer, S., Geiger, H. H., and Graner, A. (2010). DNA polymorphisms and haplotype patterns of transcription factors involved in barley endosperm development are associated with key agronomic traits. *BMC Plant Biol.* 10:5. doi: 10.1186/1471-2229-10-5

Hossain, S., Ford, R., McNeil, D., Pittock, C., and Panozzo, J. F. (2010). Inheritance of seed size in chickpea (*Cicer arietinum* L.) and identification of QTL based on 100-seed weight and seed size index. *Aust. J. Crop. Sci.* 4, 126–135.

Hossain, S., Panozzo, J. F., Pittock, C., and Ford, R. (2011). Quantitative trait loci analysis of seed coat color components for selective breeding in chickpea (*Cicer arietinum* L.). *Can. J. Plant Sci.* 91, 49–55. doi: 10.4141/cjps10112

Hu, J., Reddy, V. S., and Wessler, S. R. (2000). The rice *R* gene family: two distinct subfamilies containing several miniature inverted-repeat transposable elements. *Plant Mol. Biol.* 42, 667–678. doi: 10.1023/A:1006355510883

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967. doi: 10.1038/ng.695

Huang, Z., Ban, Y., Yang, L., Zhang, Y., Li, H., Xiao, E., et al. (2012). Fine mapping of the yellow seed locus in *Brassica juncea* L. *Genome* 55, 8–14. doi: 10.1139/g11-072

Hyten, D. L., Choi, I. Y., Song, Q., Shoemaker, R. C., Nelson, R. L., Costa, J. M., et al. (2007). Highly variable patterns of linkage disequilibrium in multiple soybean populations. *Genetics* 175, 1937–1944. doi: 10.1534/genetics.106.069740

Ichino, T., Fuji, K., Ueda, H., Takahashi, H., Koumoto, Y., Takagi, J., et al. (2014). *GFS9/TT9* contributes to intracellular membrane trafficking and flavonoid accumulation in *Arabidopsis thaliana*. *Plant J.* 80, 410–423. doi: 10.1111/tpj.12637

Isemura, T., Kaga, A., Konishi, S., Ando, T., Tomooka, N., Han, O. K., et al. (2007). Genome dissection of traits related to domestication in azuki bean (*Vigna angularis*) and comparison with other warm-season legumes. *Ann. Bot.* 100, 1053–1071. doi: 10.1093/aob/mcm155

Jaganathan, D., Thudi, M., Kale, S., Azam, S., Roorkiwal, M., Gaur, P. M., et al. (2014). Genotyping-by-sequencing based intra-specific genetic map refines a “QTL-hotspot” region for drought tolerance in chickpea. *Mol. Genet. Genomics* 2, 559–571. doi: 10.1007/s00438-014-0932-3

Jain, M., Misra, G., Patel, R. K., Priya, P., Jhanwar, S., Khan, A. W., et al. (2013). A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *Plant J.* 74, 715–729. doi: 10.1111/tpj.12173

Kharabian-Masouleh, A., Waters, D. L. E., Reinke, R. F., Ward, R., and Henry, R. J. (2012). SNP in starch biosynthesis genes associated with nutritional and functional properties of rice. *Sci. Rep.* 2:557. doi: 10.1038/srep00557

Kitamura, S., Shikazono, N., and Tanaka, A. (2004). *TRANSPARENT TESTA 19* is involved in the accumulation of both anthocyanins and proanthocyanidins in *Arabidopsis*. *Plant J.* 37, 104–114. doi: 10.1046/j.1365-313X.2003.01943.x

Kongjaimun, A., Kaga, A., Tomooka, N., Somta, P., Vaughan, D. A., and Srinivas, P. (2012). The genetics of domestication of yardlong bean, *Vigna unguiculata* (L.) Walp. ssp. *unguiculata* cv. gr. *Sesquipedalis*. *Ann. Bot.* 109, 1185–1200. doi: 10.1093/aob/mcs048

Konishi, S., Izawa, T., Lin, S. Y., Ebana, K., Fukuta, Y., Sasaki, T., et al. (2006). An SNP caused loss of seed shattering during rice domestication. *Science* 312, 1392–1396. doi: 10.1126/science.1126410

Kottapalli, P., Gaur, P. M., Katiyar, S. K., Crouch, J. H., Buhariwalla, H. K., Pande, S., et al. (2009). Mapping and validation of QTLs for resistance to an Indian isolate of *Ascochyta* blight pathogen in chickpea. *Euphytica* 165, 79–88. doi: 10.1007/s10681-008-9762-x

Kovinich, N., Saleem, A., Arnason, J. T., and Miki, B. (2012). Identification of two anthocyanidin reductase genes and three red-brown soybean accessions with reduced anthocyanidin reductase 1 mRNA, activity, and seed coat proanthocyanidin amounts. *J. Agric. Food Chem.* 60, 574–584. doi: 10.1021/jf2033939

Kujur, A., Bajaj, D., Saxena, M. S., Tripathi, S., Upadhyaya, H. D., Gowda, C. L. L., et al. (2013). Functionally relevant microsatellite markers from chickpea transcription factor genes for efficient genotyping applications and trait association mapping. *DNA Res.* 20, 355–374. doi: 10.1093/dnares/dst015

Kujur, A., Bajaj, D., Saxena, M. S., Tripathi, S., Upadhyaya, H. D., Gowda, C. L. L., et al. (2014). An efficient and cost-effective approach for genic microsatellite marker-based large-scale trait association mapping: identification of candidate genes for seed weight in chickpea. *Mol. Breed.* 34, 241–265. doi: 10.1007/s11032-014-0033-3

Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015). A genome-wide SNP scan accelerates trait-regulatory genomic loci identification in chickpea. *Sci. Rep.* 5:11166. doi: 10.1038/srep11166

Kumar, A., Choudhary, A. K., Solanki, R. K., and Pratap, A. (2011). Towards marker-assisted selection in pulses: a review. *Plant Breed.* 130, 297–313. doi: 10.1111/j.1439-0523.2011.01851.x

Kumar, V., Singh, A., Amitha Mithra, S. V., Krishnamurthy, S. L., Parida, S. K., Jain, S., et al. (2015). Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). *DNA Res.* 2, 133–145. doi: 10.1093/dnares/dsu046

Lam, H. M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F. L., et al. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42, 1053–1059. doi: 10.1038/ng.715

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lepiniec, L., Debeaujon, I., Routaboul, J. M., Baudry, A., Pourcel, L., Nesi, N., et al. (2006). Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* 57, 405–430. doi: 10.1146/annurev.arplant.57.032905.105252

Lev-Yadun, S., Gopher, A., and Abbo, S. (2000). The cradle of agriculture. *Science* 288, 1602–1603. doi: 10.1126/science.288.5471.1602

Li, S., Zhao, B., Yuan, D., Duan, M., Qian, Q., Tang, L., et al. (2013). Rice zinc finger protein DST enhances grain production through controlling *Gn1a/OsCKX2* expression. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3167–3172. doi: 10.1073/pnas.1300359110

Li, X., Chen, L., Hong, M., Zhang, Y., Zu, F., Wen, J., et al. (2012). A large insertion in *bHLH* transcription factor *BrTT8* resulting in yellow seed coat in *Brassica rapa*. *PLoS ONE* 7:e44145. doi: 10.1371/journal.pone.0044145

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

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444

Liu, C., Jun, J. H., and Dixon, R. A. (2014). *MYB5* and *MYB14* play pivotal roles in seed coat polymer biosynthesis in *Medicago truncatula*. *Plant Physiol.* 165, 1424–1439. doi: 10.1104/pp.114.241877

Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282

Liu, L. Z., Meng, J. L., Lin, N., Chen, L., Tang, Z. L., Zhang, X. K., et al. (2006). QTL mapping of seed coat color for yellow seeded *Brassica napus*. *Acta Genet. Sinica* 33, 181–187. doi: 10.1016/S0379-4172(06)60037-1

Ludwig, S. R., Habera, L. F., Dellaporta, S. L., and Wessler, S. R. (1989). *Lc*, a member of the maize *R* gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcription activators and contains the *myc* homology region. *Proc. Natl. Acad. Sci. U.S.A.* 86, 7092–7096. doi: 10.1073/pnas.86.18.7092

Mao, H., Sun, S., Yao, J., Wang, C., Yu, S., Xu, C., et al. (2010). Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19579–19584. doi: 10.1073/pnas.1014419107

Marinova, K., Pourcel, L., Weder, B., Schwarz, M., Barron, D., Routaboul, J. M., et al. (2007). The *Arabidopsis* MATE transporter *TT12* acts as a vacuolar flavonoid/H⁺-antiporter active in proanthocyanidin- accumulating cells of the seed coat. *Plant Cell* 19, 2023–2038. doi: 10.1105/tpc.106.046029

Mather, K. A., Caicedo, A. L., Polato, N. R., Olsen, K. M., McCouch, S., and Purugganan, M. D. (2007). The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics* 177, 2223–2232. doi: 10.1534/genetics.107.079616

McNally, K. L., Childs, K. L., Bohnert, R., Davidson, R. M., Zhao, K., Ulat, V. J., et al. (2009). Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12273–12278. doi: 10.1073/pnas.0900992106

Meyer, R. S., DuVal, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48. doi: 10.1111/j.1469-8137.2012.04253.x

Mol, J., Grotewold, E., and Koes, R. (1998). How genes paint flowers and seeds. *Trends Plant Sci.* 3, 212–217. doi: 10.1016/S1360-1385(98)01242-4

Nayak, S. N., Zhu, H., Varghese, N., Datta, S., Choi, H. K., Horres, R., et al. (2010). Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with *Medicago truncatula* genome. *Theor. Appl. Genet.* 120, 1415–1441. doi: 10.1007/s00122-010-1265-1

Neale, D. B., and Savolainen, O. (2004). Association genetics of complex traits in conifers. *Trends Plant Sci.* 9, 325–330. doi: 10.1016/j.tplants.2004.05.006

Negrão, S., Almadaní, M. C., Pires, I. S., Abreu, I. A., Maroco, J., Courtois, B., et al. (2013). New allelic variants found in key rice salt-tolerance genes: an association study. *Plant Biotechnol. J.* 11, 87–100. doi: 10.1111/pbi.12010

Nei, M., Tajima, F., and Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J. Mol. Evol.* 19, 153–170. doi: 10.1007/BF02300753

Nguyen, H. N., Kim, J. H., Hyun, W. Y., Nguyen, N. T., Hong, S. W., and Lee, H. (2013). *TTG1*-mediated flavonols biosynthesis alleviates root growth inhibition in response to ABA. *Plant Cell Rep.* 32, 503–514. doi: 10.1007/s00299-012-1382-1

Pang, Y., Peel, G. J., Sharma, S. B., Tang, Y., and Dixon, R. A. (2008). A transcript profiling approach reveals an epicatechin-specific glucosyltransferase expressed in the seed coat of *Medicago truncatula*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14210–14215. doi: 10.1073/pnas.0805954105

Parida, S. K., Mukerji, M., Singh, A. K., Singh, N. K., and Mohapatra, T. (2012). SNPs in stress-responsive rice genes: validation, genotyping, functional relevance and population structure. *BMC Genomics* 13:426. doi: 10.1186/1471-2164-13-426

Pesch, M., Dartan, B., Birkenbihl, R., Somssich, I. E., and Hülkamp, M. (2014). *Arabidopsis TTG2* regulates TRY expression through enhancement of activator complex-triggered activation. *Plant Cell* 26, 4067–4083. doi: 10.1105/tpc.114.129379

Pourcel, L., Irani, N. G., Koo, A. J., Bohorquez-Restrepo, A., Howe, G. A., and Grotewold, E. (2013). A chemical complementation approach reveals genes and interactions of flavonoids with other pathways. *Plant J.* 74, 383–397. doi: 10.1111/tpj.12129

Quattrochio, F., Verweij, W., Kroon, A., Spelt, C., Mol, J., and Koes, R. (2006). *PH4* of *Petunia* is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *Plant Cell* 18, 1274–1291. doi: 10.1105/tpc.105.034041

Radhika, P., Gowda, S. J. M., Kadoo, N. Y., Mhase, L. B., Jamadagni, B. M., Sainani, M. N., et al. (2007). Development of an integrated intraspecific map of chickpea (*Cicer arietinum* L.) using two recombinant inbred line populations. *Theor. Appl. Genet.* 115, 209–216. doi: 10.1007/s00122-007-0556-7

Rahman, M., Li, G., Schroeder, D., and McVetty, P. B. E. (2011). Inheritance of seed coat color genes in *Brassica napus* (L.) and tagging the genes using SRAP, SCAR and SNP molecular markers. *Mol. Breed.* 26, 439–453. doi: 10.1007/s11032-009-9384-6

Roorkiwal, M., Sawargaonkar, S. L., Chitikineni, A., Thudi, M., Saxena, R. K., Upadhyaya, H. D., et al. (2013). Single nucleotide polymorphism genotyping for breeding and genetics applications in chickpea and pigeonpea using the BeadXpress platform. *Plant Genome* 6, 1–10. doi: 10.3835/plantgenome2013.05.0017

Routaboul, J. M., Dubos, C., Beck, G., Marquis, C., Bidzinski, P., Loudet, O., et al. (2012). Metabolite profiling and quantitative genetics of natural variation for flavonoids in *Arabidopsis*. *J. Exp. Bot.* 63, 3749–3764. doi: 10.1093/jxb/ers067

Sabbavarapu, M. M., Sharma, M., Chamarthi, S. K., Swapna, N., Rathore, A., Thudi, M., et al. (2013). Molecular mapping of QTLs for resistance to *Fusarium* wilt (race 1) and *Ascochyta* blight in chickpea (*Cicer arietinum* L.). *Euphytica* 193, 121–133. doi: 10.1007/s10681-013-0959-2

Sagasser, M., Lu, G. H., Hahlbrock, K., and Weisshaar, B. (2002). *A. thaliana* TRANSPARENT TESTA 1 is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins. *Genes Dev.* 16, 138–149. doi: 10.1101/gad.212702

Saitoh, K., Onishi, K., Mikami, I., Thidar, K., and Sano, Y. (2004). Allelic diversification at the C (*OsCI*) locus of wild and cultivated rice: nucleotide changes associated with phenotypes. *Genetics* 168, 997–1007. doi: 10.1534/genetics.103.018390

Sakamoto, W., Ohmori, T., Kageyama, K., Miyazaki, C., Saito, A., Murata, M., et al. (2001). The Purple leaf (*Pl*) locus of rice: the *Pl* (*w*) allele has a complex organization and includes two genes encoding basic helix-loop-helix proteins involved in anthocyanin biosynthesis. *Plant Cell Physiol.* 42, 982–991. doi: 10.1093/pcp/pce128

Saxena, M. S., Bajaj, D., Das, S., Kujur, A., Kumar, V., Singh, M., et al. (2014a). An integrated genomic approach for rapid delineation of candidate genes regulating agro-morphological traits in chickpea. *DNA Res.* 6, 695–710. doi: 10.1093/dnares/dsu031

Saxena, M. S., Bajaj, D., Kujur, A., Das, S., Badoni, S., Kumar, V., et al. (2014b). Natural allelic diversity, genetic structure and linkage disequilibrium pattern in wild chickpea. *PLoS ONE* 9:e107484. doi: 10.1371/journal.pone.0107484

Saxena, S. C., Salvi, P., Kaur, H., Verma, P., Petla, B. P., Rao, V., et al. (2013). Differentially expressed myo-inositol monophosphatase gene (*CaIMP*) in chickpea (*Cicer arietinum* L.) encodes a lithium-sensitive phosphatase enzyme with broad substrate specificity and improves seed germination and seedling growth under abiotic stresses. *J. Exp. Bot.* 64, 5623–5639. doi: 10.1093/jxb/ert336

Senda, M., Masuta, C., Ohnishi, S., Goto, K., Kasai, A., Sano, T., et al. (2004). Patterning of virus-infected soybean seed coat is associated with suppression of endogenous silencing of chalcone synthase genes. *Plant Cell* 16, 807–818. doi: 10.1105/tpc.019885

Smýkal, P., Vernoud, V., Blair, M. W., Soukup, A., and Thompson, R. D. (2014). The role of the testa during development and in establishment of dormancy of the legume seed. *Front. Plant Sci.* 5:351. doi: 10.3389/fpls.2014.00351

Sonah, H., Bastien, M., Iquiria, E., Tardivel, A., Légaré, G., Boyle, B., et al. (2013). An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS ONE* 8:e54603. doi: 10.1371/journal.pone.0054603

Sonah, H., O'Donoghue, L., Cober, E., Rajcan, I., and Belzile, F. (2014). Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnol. J.* 2, 211–221. doi: 10.1111/pbi.12249

Spindel, J., Wright, M., Chen, C., Cobb, J., Gage, J., Harrington, S., et al. (2013). Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. *Theor. Appl. Genet.* 126, 2699–2716. doi: 10.1007/s00122-013-2166-x

Stephens, A., Lombardi, M., Cogan, N. O. I., Foster, J. W., Hobson, K., Materne, M., et al. (2014). Genetic marker discovery, intraspecific linkage map construction and quantitative trait locus analysis of *Ascochyta* blight resistance in chickpea (*Cicer arietinum* L.). *Mol. Breed.* 33, 297–313. doi: 10.1007/s11032-013-9950-9

Sweeney, M. T., Thomson, M. J., Cho, Y. G., Park, Y. J., Williamson, S. H., Bustamante, C. D., et al. (2007). Global dissemination of a single

mutation conferring white pericarp in rice. *PLoS Genetics* 3:e133. doi: 10.1371/journal.pgen.0030133

Sweeney, M. T., Thomson, M. J., Pfeil, B. E., and McCouch, S. (2006). Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* 18, 283–294. doi: 10.1105/tpc.105.038430

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Thudi, M., Bohra, A., Nayak, S. N., Varghese, N., Shah, T. M., Penmetsa, R. V., et al. (2011). Novel SSR markers from BAC-end sequences, DARt arrays and a comprehensive genetic map with 1,291 marker loci for chickpea (*Cicer arietinum* L.). *PLoS ONE* 6:e27275. doi: 10.1371/journal.pone.0027275

Thudi, M., Upadhyaya, H. D., Rathore, A., Gaur, P. M., Krishnamurthy, L., Roorkiwal, M., et al. (2014). Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping approaches. *PLoS ONE* 9:e96758. doi: 10.1371/journal.pone.0096758

Tian, Z., Qian, Q., Liu, Q., Yan, M., Liu, X., Yan, C., et al. (2009). Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21760–21765. doi: 10.1073/pnas.0912396106

Toker, C. (2009). A note on the evolution of *kabuli* chickpeas as shown by induced mutations in *Cicer reticulatum* Ladizinsky. *Genet. Resour. Crop. Evol.* 56, 7–12. doi: 10.1007/s10722-008-9336-8

Upadhyaya, H. D., Bramel, P. J., and Singh, S. (2001). Development of a chickpea core subset using geographic distribution and quantitative traits. *Crop Sci.* 41, 206–210. doi: 10.2135/cropsci2001.411206x

Upadhyaya, H. D., Dwivedi, S. L., Baum, M., Varshney, R. K., Udupa, S. M., Gowda, C. L. L., et al. (2008). Genetic structure, diversity, and allelic richness in composite collection and reference set in chickpea (*Cicer arietinum* L.). *BMC Plant Biol.* 8:106. doi: 10.1186/1471-2229-8-106

Upadhyaya, H. D., Furman, B. J., Dwivedi, S. L., Udupa, S. M., Gowda, C. L. L., Baum, M., et al. (2006). Development of a composite collection for mining germplasm possessing allelic variation for beneficial traits in chickpea. *Plant Genet. Resour.* 4, 13–19. doi: 10.1079/PGR2005101

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. doi: 10.1007/s00122-001-0556-y

Upadhyaya, H. D., Ortiz, R., Bramel, P. J., and Singh, S. (2002). Phenotypic diversity for morphological and agronomic characteristics in chickpea core collection. *Euphytica* 123, 333–342. doi: 10.1023/A:1015088417487

Van Ooijen, J. W. (2009). *MapQTL 6: Software for the Mapping Of Quantitative Trait Loci in Experimental Populations of Diploid Species*. Wageningen: Kyazma BV.

Varshney, R. K., Paulo, M. J., Grando, S., van Eeuwijk, F. A., Keizer, L. C. P., Guo, P., et al. (2012). Genome wide association analyses for drought tolerance related traits in barley (*Hordeum vulgare* L.). *Field Crop Res.* 126, 171–180. doi: 10.1016/j.fcr.2011.10.008

Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* 31, 240–246. doi: 10.1038/nbt.2491

Varshney, R. K., Thudi, M., Nayak, S. N., Gaur, P. M., Kashiwagi, J., Krishnamurthy, L., et al. (2014). Genetic dissection of drought tolerance in chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* 127, 445–462. doi: 10.1007/s00122-013-2230-6

Wang, H., Fan, W., Li, H., Yang, J., Huang, J., and Zhang, P. (2013). Functional characterization of dihydroflavonol-4-reductase in anthocyanin biosynthesis of purple sweet potato underlies the direct evidence of anthocyanins function against abiotic stresses. *PLoS ONE* 8:e78484. doi: 10.1371/journal.pone.0078484

Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* 126, 485–493. doi: 10.1104/pp.126.2.485

Winter, P., Benko-Iseppon, A. M., Hüttel, B., Ratnaparkhe, M., Tullu, A., Sonnante, G., et al. (2000). A linkage map of chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from a *C. arietinum* × *C. reticulatum* cross: localization of resistance genes for *Fusarium* wilt races 4 and 5. *Theor. Appl. Genet.* 101, 1155–1163. doi: 10.1007/s001220051592

Winter, P., Pfaff, T., Udupa, S. M., Hüttel, B., Sharma, P. C., Sahi, S., et al. (1999). Characterization and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. *Mol. Gen. Genet.* 262, 90–101. doi: 10.1007/s004380051063

Wroblewski, T., Matvienko, M., Piskurewicz, U., Xu, H., Martineau, B., Wong, J., et al. (2014). Distinctive profiles of small RNA couple inverted repeat-induced post-transcriptional gene silencing with endogenous RNA silencing pathways in *Arabidopsis*. *RNA* 20, 1987–1999. doi: 10.1261/rna.046532.114

Xie, D. Y., and Dixon, R. A. (2005). Proanthocyanidin biosynthesis – still more questions than answers? *Phytochemistry* 66, 2127–2144. doi: 10.1016/j.phytochem.2005.01.008

Xu, W., Grain, D., Bobet, S., Le Gourrierec, J., Thévenin, J., Kelemen, Z., et al. (2014). Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in *Arabidopsis* seed. *New Phytol.* 202, 132–144. doi: 10.1111/nph.12620

Xu, X., Liu, X., Ge, S., Jensen, J. D., Hu, F., Li, X., et al. (2011). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat. Biotechnol.* 30, 105–111. doi: 10.1038/nbt.2050

Yang, K., Jeong, N., Moon, J. K., Lee, Y. H., Lee, S. H., Kim, H. M., et al. (2010). Genetic analysis of genes controlling natural variation of seed coat and flower colors in soybean. *J. Hered.* 101, 757–768. doi: 10.1093/jhered/esq078

Zhang, K., Lu, K., Qu, C., Liang, Y., Wang, R., Chai, Y., et al. (2013). Gene silencing of *BnTT10* family genes causes retarded pigmentation and lignin reduction in the seed coat of *Brassica napus*. *PLoS ONE* 8:e61247. doi: 10.1371/journal.pone.0061247

Zhang, X., Wang, J., Huang, J., Lan, H., Wang, C., Yin, C., et al. (2012). Rare allele of *OsPPKL1* associated with grain length causes extralarge grain and a significant yield increase in rice. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21534–21539. doi: 10.1073/pnas.1219776110

Zhao, J., and Dixon, R. A. (2009). MATE transporters facilitate vacuolar uptake of epicatechin 3'-O-glucoside for proanthocyanidin biosynthesis in *Medicago truncatula* and *Arabidopsis*. *Plant Cell* 21, 2323–2340. doi: 10.1101/tpc.109.067819

Zhao, J., Pang, Y., and Dixon, R. A. (2010). The mysteries of proanthocyanidin transport and polymerization. *Plant Physiol.* 153, 437–443. doi: 10.1104/pp.110.155432

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

Zuo, J., and Li, J. (2014). Molecular genetic dissection of quantitative trait loci regulating rice grain size. *Annu. Rev. Genet.* 45, 99–118. doi: 10.1146/annurev-genet-120213-092138

Conflict of Interest Statement: The reviewer Bharadwaj Chellapilla declares that, despite being affiliated with the same institute and having previously collaborated with the author Shailesh Tripathi, the review process was handled objectively. 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.

Copyright © 2015 Bajaj, Das, Upadhyaya, Ranjan, Badoni, Kumar, Tripathi, Gowda, Sharma, Singh, Tyagi and Parida. 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) or licensor 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.

Mitogen-activated protein kinase cascades in *Vitis vinifera*

Birsen Çakır^{1*} and Ozan Kılıçkaya²

¹ Department of Horticulture, Faculty of Agriculture, Ege University, Izmir, Turkey, ² Department of Pharmaceutical Biotechnology, Faculty of Pharmacy, Cumhuriyet University, Sivas, Turkey

OPEN ACCESS

Edited by:

Joanna Marie-France Cross,
İnönü University, Turkey

Reviewed by:

Matthew R. Willmann,
University of Pennsylvania, USA
Pao-Yang Chen,
Academia Sinica, Taiwan
Samia Daldoul,
Center of Biotechnology of Borj
Cedria, Tunisia

*Correspondence:

Birsen Çakır,
Department of Horticulture, Faculty of
Agriculture, Ege University,
Bornova/Izmir 35100, Turkey
birsencakir@hotmail.com

Specialty section:

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

Received: 08 February 2015

Accepted: 07 July 2015

Published: 22 July 2015

Citation:

Çakır B and Kılıçkaya O (2015)
Mitogen-activated protein kinase
cascades in *Vitis vinifera*.
Front. Plant Sci. 6:556.
doi: 10.3389/fpls.2015.00556

Protein phosphorylation is one of the most important mechanisms to control cellular functions in response to external and endogenous signals. Mitogen-activated protein kinases (MAPK) are universal signaling molecules in eukaryotes that mediate the intracellular transmission of extracellular signals resulting in the induction of appropriate cellular responses. MAPK cascades are composed of four protein kinase modules: MAPKKK kinases (MAPKKKKs), MAPKK kinases (MAPKKKs), MAPK kinases (MAPKKs), and MAPKs. In plants, MAPKs are activated in response to abiotic stresses, wounding, and hormones, and during plant pathogen interactions and cell division. In this report, we performed a complete inventory of MAPK cascades genes in *Vitis vinifera*, the whole genome of which has been sequenced. By comparison with MAPK, MAPK kinases, MAPK kinase kinases and MAPK kinase kinase kinase members of *Arabidopsis thaliana*, we revealed the existence of 14 MAPKs, 5 MAPKKs, 62 MAPKKKs, and 7 MAPKKKKs in *Vitis vinifera*. We identified orthologs of *V. vinifera* putative MAPKs in different species, and ESTs corresponding to members of MAPK cascades in various tissues. This work represents the first complete inventory of MAPK cascades in *V. vinifera* and could help elucidate the biological and physiological functions of these proteins in *V. vinifera*.

Keywords: MAP kinase, *Vitis vinifera*, signal transduction, protein phosphorylation

Introduction

Mitogen-activated protein kinase (MAPK) cascades are highly conserved modules of signal transduction in eukaryotes including yeast, animals, and plants. MAPK cascades play an important role in protein phosphorylation of signal transduction events (Rodriguez et al., 2010). MAPK cascades typically consist of three protein kinases, MAPK, MAPK kinase (MAPKK), and MAPK kinase kinase (MAPKKK), but sometimes include MAP3K kinase (MAP4K) that phosphorylate the corresponding downstream substrates (Jonak et al., 2002; Champion et al., 2004).

MAPK is activated via phosphorylation of conserved threonine (T) and tyrosine (Y) residues in the catalytic subdomain by its specific MAPKK, which is in turn activated by phosphorylation of two serine/threonine residues in a conserved S/T-X33-5-S/T motif by an upstream MAPKKK (Stulemeijer et al., 2007; Zaïdi et al., 2010; Huang et al., 2011). Upon activation, the MAPK could be translocated into the nucleus or cytoplasm to trigger the cellular responses through phosphorylation of downstream transcription factors or components of transcription machinery while some MAP kinases, like ERK3, are constitutively present in the nucleus and may function

Abbreviations: MAPK, mitogen-activated protein kinase; ORF, open reading frame.

in the nucleus (Lee et al., 2004; Pedley and Martin, 2005; Fiil et al., 2009; Nadarajah and Sidek, 2010). MAPKKK is usually activated by a G protein, but sometimes activation is mediated via an upstream MAP4K (Champion et al., 2004).

MAPK proteins contain 11 evolutionary conserved kinase domains that may be involved in substrate specificity or protein-protein interaction (Nadarajah and Sidek, 2010). MAPK cascade proteins have TEY or TDY phosphorylation motifs in the region between kinase domains VII and VIII (Group et al., 2002), which provides a protein-binding domain for the activation of MAPKs (Rohila and Yang, 2007).

In plants, MAPKs are involved in cellular responses to hormones, plant growth and development, regulation of the cell cycle, and responses to biotic and abiotic stresses (Jonak et al., 1993; Wilson et al., 1997; Zhang and Klessig, 1997; Bögref et al., 1999; Nishihama et al., 2001; Bergmann et al., 2004; Lukowitz et al., 2004; Katou et al., 2005; Meng et al., 2012).

A variety of genes encoding MAPKs have been cloned from *Arabidopsis*, rice, tobacco and barley, and oat (Hutty and Phillips, 1995; Knetsch et al., 1996; Mizoguchi et al., 1998; Nadarajah and Sidek, 2010; Zaïdi et al., 2010; Sun et al., 2014). The *Arabidopsis* genome contains 20 MAPK genes (Group et al., 2002; Jonak et al., 2002). MAPK genes such as AtMPK4 and AtMPK6, have been identified in *Arabidopsis* (Ichimura et al., 1998, 2000; Nadarajah and Sidek, 2010). It has been reported that MAPK genes are involved in biotic and abiotic stress responses (Mizoguchi et al., 1996; Ichimura et al., 2000; Asai et al., 2002; Nadarajah and Sidek, 2010). For example, OsMAPK3, OsMAPK6, and the MAPK kinase OsMKK4 are induced by a chitin elicitor in rice and the activated form of OsMKK4 induces cell death (Kishi-Kaboshi et al., 2010). Similarly, NtWIPK, OsMPK5, and AtMPK3 were activated by pathogens and abiotic stresses (Zhang and Klessig, 2001; Hamel et al., 2006; Rohila and Yang, 2007). AtMPK4 and AtMPK6 are activated by osmotic stress, low humidity, low temperature, and wounding (Ichimura et al., 2000; Teige et al., 2004). AtMPK3 and AtMPK6 are also regulated by biotic elicitors via AtMKK4/5 and AtMPK4 is a negative regulator of defense response (Asai et al., 2002). In addition, AtMPK3 and AtMPK6 are involved in the embryo, anther and inflorescence development and stomatal distribution on the leaf surface (Bergmann et al., 2004; Gray and Hetherington, 2004; Bush and Krysan, 2007).

MKks are activated by the phosphorylation on conserved serine and threonine residues in the S/T-X3-5-S/T motif and characterized by a putative MAPK-docking domain K/R-K/R-K/R-X1-6-L-X-L/V/S, and a kinase domain (Group et al., 2002). To date, many MAPKs have been identified from several plant species. All the identified MAPKK genes from *Arabidopsis*, rice and poplar contain 11 catalytic subdomains (Ichimura et al., 2002; Rao et al., 2010; Wang et al., 2014c). In *Arabidopsis*, MKK1 was activated by wounding and abiotic stress (Matsuoka et al., 2002). Alfalfa SIMKK mediates both salt and elicitor-induced signals (Kiegerl et al., 2000; Cardinale et al., 2002). NtMEK2 activates SIPK and WIPK resulting in cell death (Yang et al., 2001).

MAPKKKs form the largest class of MAPK cascade enzymes with 80 members classified into three subfamilies, MEKK,

Raf, and ZIK containing 21, 11, and 48 genes, respectively in *Arabidopsis* (Jonak et al., 2002). Plant MAPKKKs are characterized by different primary structures of their kinase domains, but are conserved within a single group (Champion et al., 2004). The MEKK subfamily comprises a conserved kinase domain of G(T/S)Px(W/Y/F)MAPEV (Jonak et al., 2002). The ZIK subfamily contains GTPEFMAPE(L/V)Y while the Raf subfamily has GTxx(W/Y)MAPE (Jonak et al., 2002). All the MAPKKK proteins have a kinase domain, and most of them have a serine/threonine protein kinase active site (Wang et al., 2015). In the RAF subfamily, most of the proteins have a long N-terminal regulatory domain and C-terminal kinase domain. By contrast, majority of the members in the ZIK subfamily have an N-terminal kinase domain (Wang et al., 2015). However, the MEKK subfamily has a less conserved protein structure with a kinase domain located either at the C- or N-terminal or in the central part of the protein (Wang et al., 2015). Homologs of MAPKKKs have been identified in plant species such as alfalfa, *Arabidopsis*, tobacco (Kovtun et al., 2000; Nishihama et al., 2001; Lukowitz et al., 2004; Nakagami et al., 2004). The MEKK subfamily contains NPK1, NbMAPKK α , NbMAPKK γ , NbMAPKK δ in tobacco (Jin et al., 2002; del Pozo et al., 2004; Liu et al., 2004; Melech-Bonfil and Sessa, 2010), MEKK1 in *Arabidopsis* (Asai et al., 2002), and SIMAPKK α and SIMAPKK δ in tomato (Oh et al., 2010; Sun et al., 2014). The second subfamily, Raf, includes *Arabidopsis* CTR1/raf1 (Kieber et al., 1993), EDR/Raf2 (Frye et al., 2001), and DSM1 in rice (Ning et al., 2010). In *Arabidopsis*, MEKK1 regulates defense responses against different pathogens including bacteria and fungi (Asai et al., 2002; Qiu et al., 2008; Galletti et al., 2011). In addition, AtEDR1, a Raf-like MAPKKK, regulates SA-inducible defense responses (Frye et al., 2001). The ZIK subfamily which contains 10 and 9 members in *Arabidopsis* and rice, respectively, are able to regulate flowering time and circadian rhythms (Wang et al., 2008; Kumar et al., 2011).

A putative phosphorylation domain T/S_{x5}T/S is found between domains VII and VIII in MAP4Ks, which is identical to the phosphorylation motif of MAPKs from plants (Jouannic et al., 1999; Ichimura et al., 2002). Both domains participate in peptide-substrate recognition (Champion et al., 2004). MAP4Ks can be linked to the plasma membrane through association with a small GTPase or lipid (Qi and Elion, 2005). They are directly activated by stimulated interaction with adaptor proteins (Qi and Elion, 2005). The MAP4Ks are divided into eight classes including PAK-related, Gck, Mst, Tao, Ste/PAK, Sok (Champion et al., 2004). The majority of MAP4Ks are from the large class of Ste20 protein kinases, which exhibit a highly diverse noncatalytic domain (Dan et al., 2001). The PAKs, which have a C-terminal catalytic domain, are separated from the GC Kinase-related polypeptides, which contain an N-terminal catalytic domain (Dan et al., 2001). Most of the MAP4Ks contain an N-terminal catalytic domain, but members of the STE20/PAK group have a C-terminal kinase domain and some plant MAP4Ks have their kinase domain in the middle of the sequences (Leprince et al., 1999). The *Arabidopsis* genome contains 10 putative MAP4Ks (Champion et al., 2004). A maize gene encoding MIK is a GCK-like kinase being a subfamily of MAP4K (Llompart et al., 2003),

which relates membrane-located receptors to MAP kinases (Dan et al., 2001). Some MAP4K are able to phosphorylate MEKK or Raf members whereas other MAP4Ks either phosphorylate MAPKKs or function as adaptors (Champion et al., 2004).

However, the functions of most MAPK genes in plants are still unknown. Although MAPK cascades are involved in signaling multiple defense responses, the role of *Vitis* MAPK cascades in response to biotic and abiotic stresses are not elucidated. In previous studies in grapevine, a few components of the MAPK gene family were isolated (Wang et al., 2014a). In addition, the gene family of MAPKKKs were identified and their expression profiles were analyzed in different organs in response to different stresses (Wang et al., 2014b). Interestingly, the expression of *VvMAP* kinase gene was induced by salinity and drought (Daldoul et al., 2012). However, the MAPKK and the MAPKKKK subfamilies have not yet been characterized. To explore the role of MAPK cascade proteins in biotic and abiotic stress responses in grapevine, the publicly available grapevine genome (Jaillon et al., 2007) was analyzed to identify all members of MAPK cascade proteins. Using these databases, we characterized all members of MAPK cascades of *V. vinifera* and performed a phylogenetic analysis in comparison with members of *Arabidopsis* MAPK cascade proteins.

Materials and Methods

Genome-wide Identification of MAPK Cascade Genes in Grapevine

The MAPK cascade protein sequences of *Arabidopsis thaliana* were used to search against the *V. vinifera* proteome 12× database (<http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/>) using a BLASTP analysis (<http://www.ncbi.nlm.nih.gov/blast>) (Altschul et al., 1990) with scores higher than 400 and an “E” value $> e-120$ (Çakır and Kılıçkaya, 2013). The sequences of *Arabidopsis* MAPK cascade proteins were obtained from the TAIR (<http://www.arabidopsis.org/>). MAPK domain (PS01351), ATP-binding domain (PS00107), protein kinase domain (PS50011), serine/threonine protein kinase active site (PS00108) were identified in the sequences of polypeptides corresponding to *V. vinifera* MAPK cascade proteins by the Conserved Domain Database (CDD) at NCBI (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) and PROSITE (<http://prosite.expasy.org/>) (Marchler-Bauer et al., 2009). In addition, the NCBI non-redundant protein database was screened with each sequence in order to independently validate the automatic annotation.

Multiple-sequence Alignment and Phylogenetic Tree Construction

Multiple-sequence alignments of the putative MAPK cascade proteins were aligned using CLUSTAL W and subjected to phylogenetic analysis by both the maximum parsimony and distance with neighbor-joining methods with 1000 bootstrap replicates (Saitou and Nei, 1987; Thompson et al., 1994). The phylogenetic tree was illustrated using MEGA5. Because similar results were obtained with both methods, only the single tree retrieved from the distance analysis is discussed in detail.

For MAPK cascade subfamilies from both *V. vinifera* and *A. thaliana*, multiple sequence alignment was performed using the multiple sequence comparison by log-expectation (MUSCLE) alignment tool (<http://www.ebi.ac.uk/Tools/msa/muscle/>) (Edgar, 2004). The phylogenetic analysis was performed using a neighbor-joining method with 1000 bootstrap replicates and visualized with MEGA5 software (Tamura et al., 2011). The protein theoretical molecular weight and isoelectric point were predicted using compute pI/MW (<http://au.expasy.org/tools/>).

Orthology Analysis and Database Search

Orthology analysis was performed using the PHOG web server (<http://phylofacts.berkeley.edu/orthologs/>) (Datta et al., 2009). The sequences of conserved domains with similarity over 70% and an “E” value of 0.0 were selected as queries. The selected sequences of conserved domains from different species were then used in a BLASTP search against the *V. vinifera* protein sequence database. The best hits were annotated as putative orthologous sequences (Moreno-Hagelsieb and Latimer, 2008).

Expressed sequence tags (ESTs) were identified by BLASTn of the *V. vinifera* expressed sequence tag (EST) database (<http://www.ncbi.nlm.nih.gov/dbEST>). Using the sequences of all of the MAPK cascade proteins as queries. The positive sequences were then confirmed by alignment with the query ORF.

Results and Discussion

Genome-wide Identification of MAPK Cascade Genes in *Vitis vinifera*

Vitis vinifera MAPK cascade sequences were mined from the grapevine genome proteome 12x database (Jaillon et al., 2007). We identified 88 ORFs encoding putative MAPK cascade proteins containing at least MAPK domain by BLAST searches of the grapevine genome proteome 12× database with the amino acid sequences of the MAPK cascade proteins from *A. thaliana* as queries (Table 1). The completed *Vitis* genome contains 14 MAPKs, 5 MAPKKs, 62 MAPKKKs, and 7 MAPKKKKs (Table 1).

Phylogenetic Analysis

All predicted MAPK cascade family sequences were aligned using ClustalW (Thompson et al., 1994). A rooted phylogenetic tree was constructed by alignment of full length amino acid sequences using the MEGA5 program and maximum parsimony and distance with neighbor-joining methods (Saitou and Nei, 1987) (Figure 1). One thousand bootstrap replicates were produced for each analysis.

Vitis MAPK cascade sequences can be divided into four subfamilies on the basis of the presence of conserved threonine and tyrosine residues in the motif TxY located in the activation loop (T-loop) between kinase subdomains VII and VIII. In addition, we identified MAPKKKK subfamily with 7 members in *Vitis* genome, which has the conserved amino acid motifs TFVGTPxWMAPEV as described (Jonak et al., 2002). The members of four subfamilies clustered more tightly with each other than with members of other subfamilies (Figure 1).

TABLE 1 | Detailed inventory of the *Vitis* MAPK cascade proteins.

Subfamily name	12X <i>Vitis vinifera</i> ID	NCBI ID	Chr	Str	Genomic location	Gene length in bp	CDS length in bp	Length of protein in AA	Number of Exon	Number of Intron	pl	mW (kDa)
VMPKs												
VMPK1	GSVVT01000784001	CB131754.3	12	+	124452-133238	8787	1518	505	10	9	9.34	57.43
VMPK2	GSVVT01005924001	CB135594.3	7	+	886169-898284	12116	1341	446	16	15	5.40	51.22
VMPK3	GSVVT01008408001	CB115552.3	17	+	2368190-2377747	9558	1806	601	11	10	6.89	67.98
VMPK4	GSVVT01009766001	CB119748.3	18	+	11125765-11129338	3574	588	195	4	3	5.44	22.50
VMPK5	GSVVT011749001	CB128902.3	1	-	4565334-4574753	9420	1842	613	11	10	8.68	70.46
VMPK6	GSVVT01014081001	CB120098.3	19	+	224299-234190	9892	1797	599	10	9	9.21	67.79
VMPK7	GSVVT01017873001	CB126170.3	5	-	42050509-4215917	10409	1692	563	10	9	8.59	64.03
VMPK8	GSVVT01018883001	CB117457.3	4	+	18974001-19005635	31635	2310	769	10	9	5.51	87.48
VMPK9	GSVVT01019406001	CB134380.3	2	-	380310-386888	6579	1128	375	6	5	5.86	42.80
VMPK10	GSVVT01022771001	CB137450.3	2	+	16326975-16335400	8426	1359	452	16	15	9.62	51.58
VMPK11	GSVVT01025091001	CB116237.3	6	+	4580755-4584961	4207	1116	371	6	5	4.94	42.53
VMPK12	GSVVT01025105001	CB116244.3	6	-	4432854-4436338	3485	990	329	6	5	5.52	38.17
VMPK13	GSVVT01026984001	CB140425.3	15	-	18821560-18826926	5367	1128	375	6	5	6.43	43.27
VMPK14	GSVVT01038192001	CB124707.3	5	+	24220239-24241107	20870	993	330	6	5	5.64	38.37
VMAPKKs												
VMAPKK1	GSVVT01008476001	CB115608.3	17	+	1537423-1538551	1129	675	224	3	2	6.38	24.66
VMAPKK2	GSVVT01015156001	CB12870.3	11	+	1417439-1424337	6899	1065	355	8	7	6.00	39.28
VMAPKK3	GSVVT010152283001	CB127984.3	11	+	2377698-2381398	3701	1065	355	8	7	6.02	39.98
VMAPKK4	GSVVT01016115001	CB125274.3	9	+	19257788-19265261	7474	1188	396	5	4	10.15	43.78
VMAPKK5	GSVVT01032414001	CB134873.3	14	-	27139003-27145873	6871	1557	519	9	8	5.56	57.61
VMAPKKKs												
VMAPKKK1	GSVVT01000047001	CB136768.3	14	+	3063647-3072319	8673	1992	664	17	16	5.36	72.89
VMAPKKK2	GSVVT01000256001	CB127711.3	7	+	20596048-20597073	1026	921	307	2	1	8.97	33.93
VMAPKKK3	GSVVT01001193001	CB128728.3	7	+	944892-950225	5334	1215	405	6	5	7.03	44.92
VMAPKKK4	GSVVT01001690001	CB135056.3	18	-	1426312-1329573	33262	1653	551	16	15	5.07	61.80
VMAPKKK5	GSVVT01002332001	CB135719.3	Un	+	34161697-341676722	5926	696	232	7	6	9.43	26.14
VMAPKKK6	GSVVT01004254001	CB118826.3	Un	+	37734319-37739476	5158	1158	386	10	9	9.51	43.07
VMAPKKK7	GSVVT01007446001	CB128553.3	Un	+	31988209-31995727	7519	2124	708	11	10	9.73	77.57
VMAPKKK8	GSVVT01007632001	CB11941.3	17	-	10966272-10980533	14262	1464	488	9	8	5.47	54.86
VMAPKKK9	GSVVT01007646001	CB11949.3	17	+	10874999-108777438	2440	1059	353	6	5	8.13	40.08
VMAPKKK10	GSVVT01007762001	CB115038.3	17	+	9308909-9314007	5100	909	303	3	2	7.97	34.37
VMAPKKK11	GSVVT01007775001	CB115048.3	17	-	9166429-9172256	5829	1050	350	6	5	7.02	38.56
VMAPKKK12	GSVVT01008413001	CB11555.3	17	-	2321687-2342403	20717	2697	899	16	15	5.24	99.09
VMAPKKK13	GSVVT01008728001	CB118907.3	18	-	1477098-1491666	14569	1569	523	16	15	6.67	59.46
VMAPKKK14	GSVVT01008938001	CB119081.3	18	-	3594893-3606331	11439	822	274	8	7	6.45	30.15
VMAPKKK15	GSVVT01009192001	CB119282.3	18	-	5939861-5949524	9664	2718	906	15	14	8.41	101.48
VMAPKKK16	GSVVT01009575001	CB119581.3	18	+	9549009-9561275	12267	1101	367	10	9	8.39	40.91
VMAPKKK17	GSVVT0112031001	CB127127.3	1	-	206896-2039042	32147	4191	1397	25	24	5.73	154.60

(Continued)

TABLE 1 | Continued

Subfamily name	12X <i>Vitis vinifera</i> ID	NCBI ID	GenBank ID	Chr	Str	Genomic location	Gene length in bp	CDS length in bp	Length of protein in AA	Number of Exon	Number of Intron	pI	mW (kDa)
ViMAPKKK18	GSVVT01012116001	CB127196.3	1	+	+	1303636-1315494	11859	3717	1239	11	10	5.32	136.73
ViMAPKKK19	GSVVT01012632001	CB123172.3	10	+	+	222774-228777	6004	1287	429	4	3	5.48	48.37
ViMAPKKK20	GSVVT01012686001	CB12211.3	10	+	+	641937-648890	6954	3369	1123	8	7	6.09	123.96
ViMAPKKK21	GSVVT01012895001	CB125598.3	11	+	+	6576324-6581862	5539	1470	490	9	8	5.40	54.29
ViMAPKKK22	GSVVT01015494001	CB128162.3	11	-		4194704-4202363	7650	3066	1022	7	6	8.36	114.09
ViMAPKKK23	GSVVT01017915001	CB128208.3	5	+	+	4619229-4633132	13904	2061	687	17	16	6.76	75.81
ViMAPKKK24	GSVVT01017968001	CB12245.3	5	+	+	5145293-5148759	3467	1632	544	11	10	5.43	63.24
ViMAPKKK25	GSVVT01018020001	CB12291.3	5	-		5521761-5528578	6818	1371	457	7	6	8.32	51.27
ViMAPKKK26	GSVVT01018052001	CB126318.3	5	+	+	5824109-5839208	15100	3354	1118	11	10	5.59	124.59
ViMAPKKK27	GSVVT01019010001	CB117559.3	4	+	+	17821355-17829046	7692	906	302	5	4	6.66	33.93
ViMAPKKK28	GSVVT01019630001	CB13567.3	2	+	+	2093246-2095958	6353	1791	597	16	15	6.59	68.4
ViMAPKKK29	GSVVT01019739001	CB13657.3	2	+	+	2874833-2882840	8008	2706	902	11	10	9.12	97.91
ViMAPKKK30	GSVVT01019821001	CB134722.3	2	+	+	3631707-3639546	7840	2088	696	15	14	6.37	78.14
ViMAPKKK31	GSVVT01020712001	CB121988.3	12	+	+	2837402-2887099	49698	2454	818	11	10	5.99	91.70
ViMAPKKK32	GSVVT0121854001	CB134208.3	14	-		6462432-6466040	3609	1821	607	10	9	5.22	69.40
ViMAPKKK33	GSVVT0121884001	CB13231.3	14	-		6026347-6048725	22379	2631	877	16	15	5.50	97.14
ViMAPKKK34	GSVVT01022098001	CB121399.3	7	+	+	16573422-16578723	5302	1287	429	8	7	5.31	47.06
ViMAPKKK35	GSVVT01022115001	CB121414.3	7	-		16707463-16710545	3083	1233	411	8	7	4.66	44.78
ViMAPKKK36	GSVVT01022116001	CB121415.3	7	+	+	16711209-16721450	10242	2469	823	7	6	8.72	89.41
ViMAPKKK37	GSVVT01023037001	CB123895.3	12	+	+	16524280-16552240	27961	1296	432	11	10	7.05	48.78
ViMAPKKK38	GSVVT01023048001	CB122901.3	12	+	+	16381443-16399280	178238	525	175	5	4	8.42	20.19
ViMAPKKK39	GSVVT01023216001	CB126680.3	12	+	+	21019776-21020867	1092	1092	364	1	0	8.88	39.99
ViMAPKKK40	GSVVT01023958001	CB137812.3	3	-		2141138-2162161	21024	1743	581	16	15	5.56	65.53
ViMAPKKK41	GSVVT01024578001	CB115829.3	6	-		8664971-8669192	4222	1896	632	7	6	6.83	71.74
ViMAPKKK42	GSVVT01026487001	CB137539.3	4	+	+	22814276-22820209	5934	1803	601	12	11	8.99	65.01
ViMAPKKK43	GSVVT01026546001	CB137576.3	4	+	+	21993962-21997860	3899	1701	567	10	9	5.52	64.16
ViMAPKKK44	GSVVT01027189001	CB140585.3	15	-		17151471-17154829	3359	1776	592	10	9	5.82	67.45
ViMAPKKK45	GSVVT01028897001	CB122687.3	16	-		17707492-17719297	11806	2679	893	11	10	9.43	95.93
ViMAPKKK46	GSVVT01029055001	CB13351.3	5	+	+	11545076-11551320	6345	2859	953	13	12	5.42	105.79
ViMAPKKK47	GSVVT01029147001	CB117788.3	11	+	+	19186659-19204657	17999	2937	979	11	10	5.40	108.81
ViMAPKKK48	GSVVT01029426001	CB133520.3	17	+	+	17077436-17089089	11654	555	185	4	3	5.62	21.36
ViMAPKKK49	GSVVT01030044001	CB123411.3	12	-		9089290-9097431	8142	2496	832	8	7	8.80	93.29
ViMAPKKK50	GSVVT01030194001	CB11047.3	8	+	+	10650328-10661181	10854	1014	338	6	5	8.85	37.96
ViMAPKKK51	GSVVT01030202001	CB118051.3	8	-		10519384-10524077	4694	1881	627	7	6	5.29	71.71
ViMAPKKK52	GSVVT01031721001	CB132391.3	3	-		3812818-3816504	3687	1194	398	8	7	8.32	43.57
ViMAPKKK53	GSVVT01032232001	CB12046.3	11	+	+	13477774-13485357	7584	447	149	3	2	5.90	16.90
ViMAPKKK54	GSVVT01032389001	CB13850.3	14	-		26886951-26893203	6253	1059	353	6	5	6.90	39.53
ViMAPKKK55	GSVVT01032487001	CB134936.3	14	+	+	27812667-27819608	6942	1083	361	6	5	6.34	40.52

(Continued)

TABLE 1 | Continued

Subfamily name	12X <i>Vitis vinifera</i> ID	NCBI GenBank ID	Chr	Str	Genomic location	Gene length in bp	CDS length in bp	Length of protein in AA	Number of Exon	Number of Intron	pI	mW (kDa)
VvMAPKK56	GSMVVT01033779001	CB130245.3	8	–	17884683-17902051	17369	2238	746	17	16	6.14	82.96
VvMAPKK57	GSMVVT01034710001	CB140217.3	13	–	8150739-8202754	52016	2277	759	15	14	7.89	85.25
VvMAPKK58	GSMVVT01034988001	CB122876.3	5	–	694778-701765	6988	2199	733	14	13	6.42	81.58
VvMAPKK59	GSMVVT01035409001	CB120668.3	4	+	1079723-1090460	10738	2631	877	15	14	5.26	97.28
VvMAPKK60	GSMVVT01036758001	CB124172.3	19	–	22924599-22935504	10906	3138	1046	10	9	5.23	115.67
VvMAPKK61	GSMVVT01037773001	CB126734.3	19	–	7730105-7732118	2014	894	298	2	1	6.17	34.22
VvMAPKK62	GSMVVT01038760001	CB132969.3	12	+	610174-641062	30889	1476	492	8	7	6.06	55.76
VvMAPKKKs												
VvMAP4K1	GSMVVT01012233001	CB127303.3	5	+	7076729-7078360	1632	1362	454	4	3	5.20	88.93
VvMAP4K2	GSMVVT01013739001	CB128527.3	9	–	18395010-18410644	15635	1704	568	15	14	9.31	41.78
VvMAP4K3	GSMVVT01014297001	CB12268.3	1	–	8012586-8021182	8597	1107	369	9	8	6.34	78.12
VvMAP4K4	GSMVVT01016074001	CB125246.3	14	–	27646939-27657107	10169	1803	601	16	15	5.58	62.83
VvMAP4K5	GSMVVT01019643001	CB13578.3	2	–	2150787-2160118	9332	2190	730	22	21	6.68	81.13
VvMAP4K6	GSMVVT01027718001	CB123577.3	1	–	342583-373752	31170	2430	810	19	18	5.80	50.65
VvMAP4K7	GSMVVT01032461001	CB134913.3	19	–	2416077-2441445	25369	2121	707	20	19	5.81	67.08

The identified open reading frames (ORFs) are classified into four subfamilies, MAPK, MAPKK, MAPKKK, and MAPKKKK. Columns 1–13 contain the protein acronym (Name), coding sequence (CDS), *Vitis* proteome 12× ID, GenBank ID, chromosome location (Chr), strand (str), gene length, number of introns and exons, protein length, estimates of molecular weight, and isoelectric point of the protein (pI) for each gene are given.

MAPKs

The phylogenetic analysis showed that the VvMAPKs were divided into five distinct groups, which is higher than previous reports (Kumar and Kirti, 2010; Nadarajah and Sidek, 2010). Group V MAPKs are found only in the grapevine genome among other plant species. All of identified ORFs encoding MAPK were named VvMPK1 through 14. Hyun et al. (2010) reported 12 MAPKs based on 8x sequence coverage in grapevine genome whereas we identified a total of 14 ORFs in *Vitis* 12x genome coverage (Hyun et al., 2010), which may be due to the errors corrected in 12x genome sequence coverage. The grapevine genome contains less MAPKs than *Arabidopsis* (20 MAPKs) (Ichimura et al., 2002) and rice (17 MAPKs) (Liu and Xue, 2007). Members of the *Vitis* MAPK subfamily show 20–86% identity to each other. Full length MAPK proteins ranged in size from 195 to 769 amino acids (Table 1). Variation in length of the entire MAPK gene is usually due to differences in the length of MAPK domain and/or, due to the number of introns. The difference in length among MAPK genes may indicate the presence or absence of motifs which could affect functional specificity.

VvMPK12, VvMPK14 belong to the group I, which contains well-characterized MAPK genes including *AtMPK3*, *AtMPK6* (Figure 2). It has been demonstrated that *AtMPK3*, *OsMPK5* were activated in response to pathogens and abiotic stresses (Zhang and Klessig, 2001; Hamel et al., 2006; Rohila and Yang, 2007). *OsMPK5* plays an important role for the resistance to blast disease (Song and Goodman, 2002; Huang et al., 2011). *AtMPK6* can be activated by various abiotic and biotic stresses (Ichimura et al., 2000; Yuasa et al., 2001; Feilner et al., 2005; Huang et al., 2011). Similarly, *PtrMAPK* is involved in resistance to both dehydration and cold (Huang et al., 2011).

Group II MAPKs are involved in both abiotic stresses and cell division in *Arabidopsis*. VvMPK13, VvMPK11, and VvMPK9 are clustered with Group II, which includes *AtMPK4*, *AtMPK5*, *AtMPK12*, and *AtMPK11*. *AtMPK4* and its upstream MAPKK AtMKK2 can be activated by biotic and abiotic stresses (Ichimura et al., 2000; Teige et al., 2004).

VvMPK4 and VvMPK8 belong to group III. *AtMPK1* in the group III is regulated by salt stress treatment (Mizoguchi et al., 1996). In addition, *AtMPK1* and *AtMPK2* are activated by ABA (Ortiz-Masia et al., 2007). The group III genes, such as rice BWMK1 and alfalfa TDY1, are activated by wounding and pathogens (Nowak et al., 1997; Lynch et al., 2001).

Group IV, which includes VvMPK1, VvMPK3, VvMPK5, VvMPK6, and VvMPK7 of the *Vitis* MAPKs, have the TDY motif in their T-loop and the absence of the C-terminal CD domain, which is consistently found in members of the other MAPK groups. VvMPK2 and VvMPK10 belonging to group V were separated from other groups.

The orthology analysis program identified one hundred-fourteen orthologs from various plant species for this subfamily (Table 2). The VvMPK3 amino acid sequence shows 83% similarity with *AtMPK9*, and VvMPK12 shows 84% similarity with *AtMPK3* from *A. thaliana*. The members of VvMAPK subfamily share between 75.8 and 91.8% similarity to the MAPK members from *Ricinus communis*,

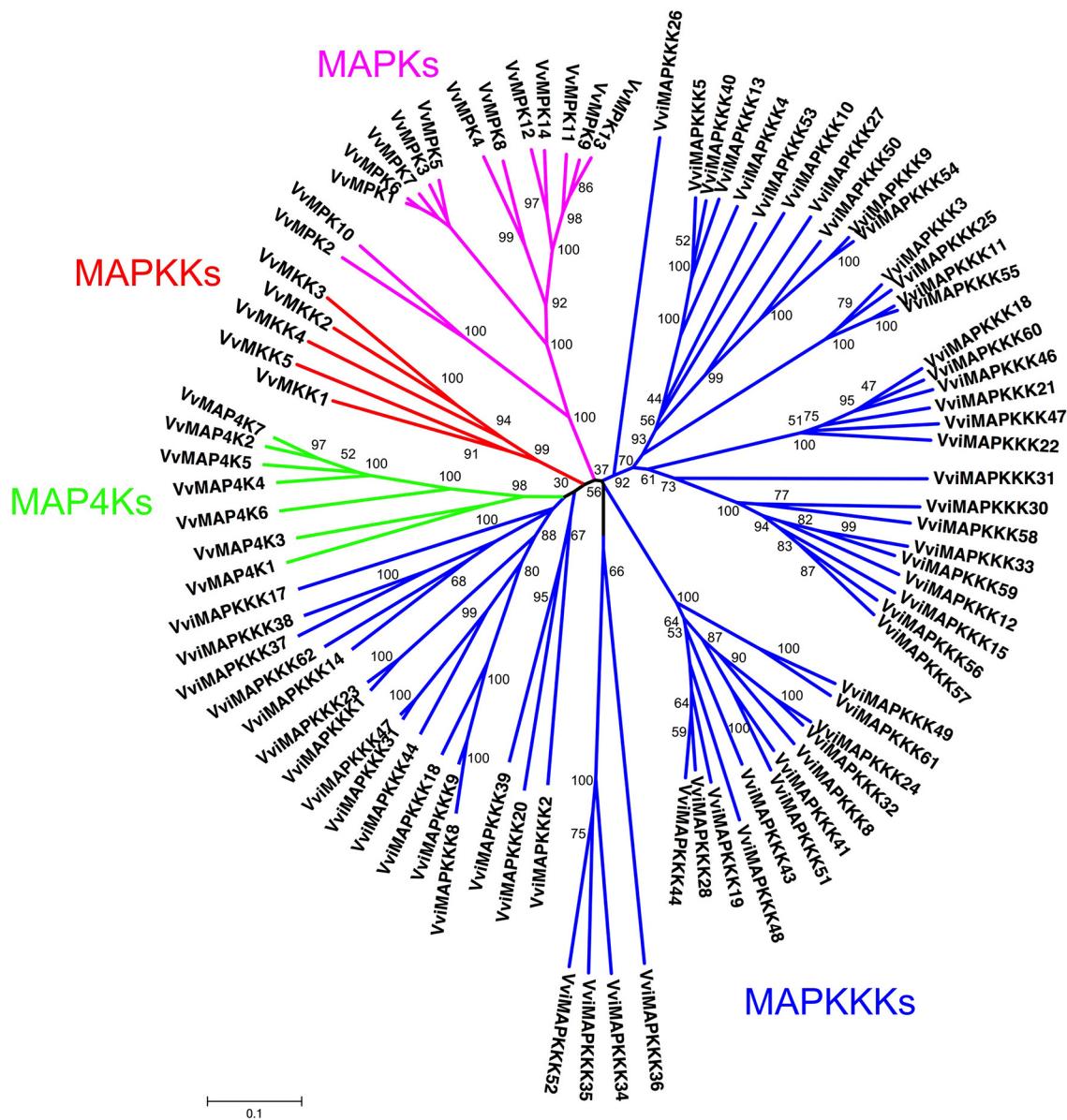


FIGURE 1 | Construction of phylogenetic tree of *Vitis* MAPK cascade proteins. The amino sequences of all *Vitis* ABC proteins were aligned using the ClustalW program and were subjected to phylogenetic analysis by the distance with neighbor-joining method. The reliabilities of each branch point, as assessed by the analysis of 1000

computer-generated trees (bootstrap replicates), were in excess of 90%, except for those discussed in the text. The abbreviations of MAPK cascade proteins are as follows: MAPK, Mitogen-activated Protein Kinase; MAPKK, MAPK Kinase; MAPKKK, MAPKK Kinase; MAPKKKK, MAPKKK Kinase as described in the text.

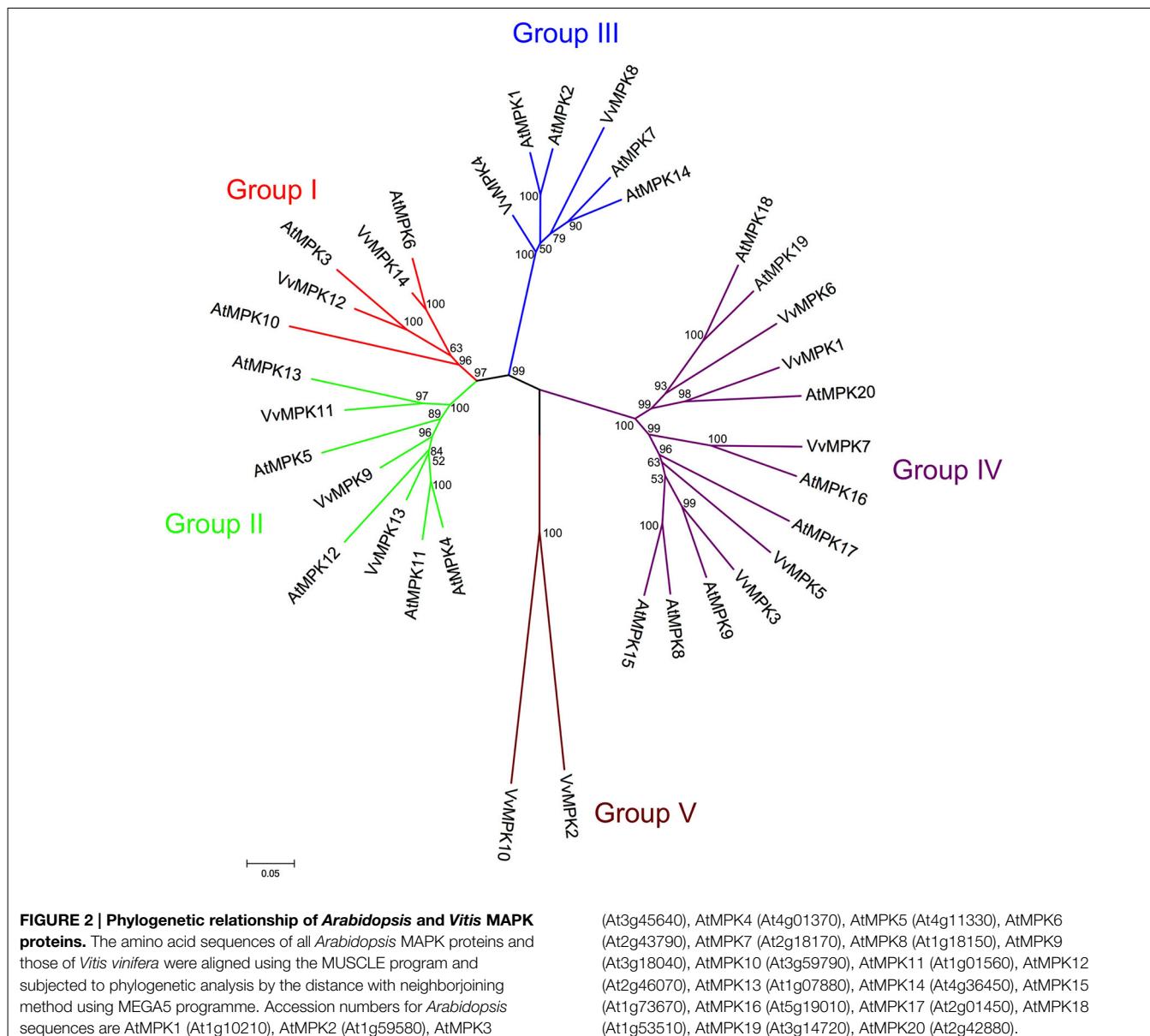
Oryza sativa, and *A. thaliana*. The phylogenetic analysis of *A. thaliana* and *V. vinifera* MAPK subfamilies confirmed the orthologs of VvMPK14/AtMPK6, VvMPK12/AtMPK3, VvMPK11/AtMPK13, VvMPK13/AtMPK12, VvMPK7/AtMPK16, and VvMPK3/AtMPK9 (Figure 2).

All of the 14 *Vitis* MAPK proteins are represented in the *Vitis* ESTs database (Supplementary Table 1) and are expressed in different tissues such as fruits, berries, buds, flowers, leaves, and roots. In addition, 12 VvMPK genes were isolated (Wang et al., 2014a). Expression analysis of VvMPK genes showed that

all VvMPK genes are expressed during grapevine growth and development, and in biotic and abiotic stresses (Wang et al., 2014a).

MAPKs

This subfamily consists of 10 members in *Arabidopsis* genome (Group et al., 2002), whereas *Vitis* genome contains 5 members of MAPKK subfamily. The full length VvMKK sequences range in size from 224 to 519 amino acids (Table 1). The members of the MAPKK subfamily in the *Vitis* genome share 29–40%



similarity with each other. By phylogenetic analysis, we also identified orthologs of *Vitis* MAPKs in *Arabidopsis* such as VvMKK5/AtMKK3 (78.6% similarity), VvMKK3/AtMKK6 (83.1% similarity), and VvMKK2/AtMKK2 (70.4% similarity) supported with significant bootstrap values. The phylogenetic analysis confirmed that VvMKK3 shares 83.3% similarity with its homolog from *Arabidopsis* on the basis of orthology analysis, (Figure 3, Table 2).

To date, none of the *Vitis* MAPKK homologs have been cloned or characterized. However, 98 ESTs were identified for this subfamily in different tissues in response to biotic or abiotic stresses (Supplementary Table 2). A role of MAPK kinase, MKK1 in abiotic stress signaling was previously demonstrated (Matsuoka et al., 2002). Analysis of MKK1 revealed that drought, salt stress, cold, wounding activated MKK1, which in turns

activates its downstream target MPK4 (Matsuoka et al., 2002). Tobacco NtMEK2 is functionally interchangeable with two *Arabidopsis* MAPKKs, AtMKK4, and AtMKK5 in activating the downstream MAPKs (Ren et al., 2002). MdMKK1 was reported to be downregulated by ABA (Wang et al., 2010). In *Arabidopsis*, AtMKK3 is upregulated in response to ABA (Hwa and Yang, 2008). Interestingly, AtMKK1/AtMKK2 play an important role in signaling in ROS homeostasis (Liu, 2012).

MAPKKKs

With 62 members, the MAPKKK subfamily represents the largest subfamily of *V. vinifera* MAPK cascade proteins, which is smaller than those of *Arabidopsis* (80 members) and rice (75 members) (Colcombet and Hirt, 2008; Rao et al., 2010). Recently, Wang et al. (2014b) identified 45 MAPKKK genes in grapevine 12x

TABLE 2 | Orthologs of *Vitis* MAPK cascade proteins identified in diverse plant species.

Subfamily name	<i>Vitis</i> proteome 12x ID	Species	%ID	UniprotKB ID
VvMPK1	GSVIVT01000784001	<i>Ricinus communis</i>	82.7	B9H811_POPTR
VvMPK2	GSVIVT01005924001	<i>Ricinus communis</i>	75.8	B9SYK7_RICCO
VvMPK3	GSVIVT01008408001	<i>Populus trichocarpa</i>	84.0	B9I2G2_POPTR
		<i>Brassica napus</i>	82.0	Q5XU40_BRANA
		<i>Arabidopsis thaliana</i>	81.8	MPK9_ARATH
		<i>Arabidopsis lyrata</i> subsp. <i>Lyrata</i>	81.2	D7L7Z0_ARALL
		<i>Oryza sativa</i> subsp. <i>Indica</i>	77.6	B3GCL0_ORYSI
		<i>Ricinus communis</i>	76.6	B9T7E7_RICCO
		<i>Zea mays</i>	75.7	B4F907_MAIZE
		<i>Oryza sativa</i> subsp. <i>Japonica</i>	75.6	MPK16_ORYSJ
		<i>Populus trichocarpa</i>	78.7	B9GGZ4_POPTR
		<i>Ricinus communis</i>	90.9	B9RSS7_RICCO
VvMPK6	GSVIVT01014081001	<i>Populus trichocarpa</i>	90.2	B9HY78_POPTR
		<i>Arabidopsis lyrata</i> subsp. <i>Lyrata</i>	84.7	D7LYJ6_ARALL
		<i>Arabidopsis thaliana</i>	84.5	MPK16_ARATH
		<i>Oryza sativa</i> subsp. <i>Japonica</i>	83.6	MPK15_ORYSJ
		<i>Oryza sativa</i> subsp. <i>Indica</i>	83.2	B3GCK9_ORYSI
		<i>Zea mays</i>	79.9	Q6TAR9_MAIZE
		<i>Triticum aestivum</i>	77.9	A9RAB1_WHEAT
		<i>Populus trichocarpa</i>	91.4	B9GWV0_POPTR
		<i>Papaver rhoeas</i>	89.3	Q683Y4_9MAGN
		<i>Ricinus communis</i>	87.4	B9RCG7_RICCO
VvMPK7	GSVIVT01017873001	<i>Solanum tuberosum</i>	84.8	Q8LT16_SOLTU
		<i>Sorghum bicolor</i>	82.7	C5YH06_SORBI
		<i>Medicago truncatula</i>	81.9	B7FK53_MEDTR
		<i>Populus trichocarpa</i>	75.7	B9GZV6_POPTR
		<i>Nicotiana attenuata</i>	86.6	A5H7H6_NICAT
		<i>Ricinus communis</i>	86.0	B9SW68_RICCO
		<i>Malus domestica</i>	85.3	D1MFM2_MALDO
		<i>Solanum lycopersicum</i>	83.5	E2GLN8_SOLLC
		<i>Nicotiana benthamiana</i>	83.5	B2NIC1_NICBE
		<i>Medicago sativa</i>	82.7	Q9ZP91_MEDSA
VvMPK10	GSVIVT01022771001	<i>Nicotiana tabacum</i>	82.7	NTF6_TOBAC
		<i>Solanum tuberosum</i>	82.7	Q8LT15_SOLTU
		<i>Arabidopsis lyrata</i> subsp. <i>Lyrata</i>	80.4	D7KHW6_ARALL
		<i>Arabidopsis thaliana</i>	79.8	MPK13_ARATH
		<i>Citrus sinensis</i>	90.3	A2IB54_CITSI
		<i>Populus trichocarpa</i>	90.0	B9HNK3_POPTR
		<i>Catharanthus roseus</i>	89.7	B8LFE0_CATRO
		<i>Cucumis sativus</i>	89.3	Q0R4I2_CUCSA
		<i>Solanum lycopersicum</i>	86.9	Q84MI4_SOLLC
		<i>Medicago truncatula</i>	86.8	B7FJD9_MEDTR
VvMPK11	GSVIVT01025091001	<i>Solanum peruvianum</i>	86.6	A8VJL7_SOLPE
		<i>Solanum tuberosum</i>	86.6	Q3V6C4_SOLTU
		<i>Nicotiana attenuata</i>	86.6	A5H2L1_NICAT
		<i>Ricinus communis</i>	86.4	B9T1Z7_RICCO
		<i>Capsicum annuum</i>	86.3	Q9LKZ2_CAPAN
		<i>Nicotiana benthamiana</i>	86.3	Q8H0B4_NICBE
		<i>Nicotiana tabacum</i>	86.0	Q8W406_TOBAC
		<i>Pisum sativum</i>	86.0	Q9M6S1_PEA

(Continued)

TABLE 2 | Continued

Subfamily name	Vitis proteome 12x ID	Species	%ID	UniprotKB ID
VvMPK13	GSVIVT01026984001	<i>Brassica napus</i>	86.0	Q5IV18_BRANA
		<i>Medicago sativa</i>	85.7	O24077_MEDSA
		<i>Petroselinum crispum</i>	85.7	O04694_PETCR
		<i>Glycine max</i>	85.4	Q5K6Q4_SOYBN
		<i>Arabidopsis thaliana</i>	84.2	MPK3_ARATH
		<i>Saccharum officinarum</i>	78.2	Q4QWQ7_SACOF
		<i>Oryza sativa</i> subsp. <i>Indica</i>	77.6	MPK5_ORYSI
		<i>Avena sativa</i>	77.3	Q43379_AVESA
		<i>Nicotiana attenuata</i>	90.7	A5H7H4_NICAT
		<i>Ricinus communis</i>	90.5	B9PDW5_RICCO
		<i>Glycine max</i>	89.9	C6TEP0_SOYBN
		<i>Populus trichocarpa</i>	89.2	B9GQC1_POPTR
		<i>Malus hupehensis</i>	89.2	B1N8Y5_9ROSA
		<i>Petroselinum crispum</i>	89.0	Q84XZ6_PETCR
		<i>Nicotiana tabacum</i>	88.5	Q3C254_TOBAC
		<i>Solanum lycopersicum</i>	88.2	D7R517_SOLLC
		<i>Thellungiella halophila</i>	87.4	E4MW58_THEHA
		<i>Brassica napus</i>	87.3	E3US78_BRANA
		<i>Arabidopsis thaliana</i>	87.2	MPK4_ARATH
		<i>Arabidopsis lyrata</i> subsp. <i>Lyrata</i>	86.9	D7M4W5_ARALL
		<i>Malus micromalus</i>	86.4	Q8GZR5_MALMI
		<i>Medicago sativa</i>	86.3	MMK2_MEDSA
		<i>Oryza sativa</i> subsp. <i>Indica</i>	83.6	A2Z9P1_ORYSI
		<i>Oryza sativa</i> subsp. <i>Japonica</i>	83.6	MPK6_ORYSJ
		<i>Zea mays</i>	83.2	B4FH09_MAIZE
		<i>Sorghum bicolor</i>	82.4	C5WUG0_SORBI
		<i>Physcomitrella patens</i> subsp. <i>patens</i>	80.9	A9S9Q8_PHYP
VvMPK14	GSVIVT01038192001	<i>Pinus taeda</i>	78.0	C7ENI4_PINTA
		<i>Populus trichocarpa</i>	95.7	B9HGK0_POPTR
		<i>Malus domestica</i>	95.2	D1MFM1_MALDO
		<i>Pisum sativum</i>	94.8	MAPK_PEA
		<i>Ricinus communis</i>	94.5	B9SFT4_RICCO
		<i>Medicago sativa</i>	94.2	MMK1_MEDSA
		<i>Glycine max</i>	94.2	Q5K6N6_SOYBN
		<i>Nicotiana tabacum</i>	93.3	NTF4_TOBAC
		<i>Solanum tuberosum</i>	93.0	Q8LT17_SOLTU
		<i>Nicotiana benthamiana</i>	93.0	B3IWK6_NICBE
		<i>Solanum lycopersicum</i>	93.0	Q84M15_SOLLC
		<i>Capsicum annuum</i>	92.7	Q9LKZ1_CAPAN
		<i>Solanum peruvianum</i>	92.7	B5B2H6_SOLPE
		<i>Nicotiana attenuata</i>	92.4	A5H2L0_NICAT
		<i>Arabidopsis thaliana</i>	91.8	MPK6_ARATH
		<i>Arabidopsis lyrata</i> subsp. <i>Lyrata</i>	91.8	D7LKI6_ARALL
		<i>Brassica napus</i>	91.5	E1B2J5_BRANA
		<i>Sorghum bicolor</i>	90.9	C5Z4D1_SORBI
		<i>Oryza sativa</i> subsp. <i>Japonica</i>	90.5	MPK1_ORYSJ
		<i>Zea mays</i>	90.5	B8QN51_MAIZE
		<i>Oryza sativa</i> subsp. <i>Indica</i>	90.5	B3GCK7_ORYSI
		<i>Triticum aestivum</i>	89.9	Q84XZ3_WHEAT
		<i>Pinus taeda</i>	87.5	C7ENI3_PINTA
		<i>Populus trichocarpa</i>	81.7	B9IKC3_POPTR

(Continued)

TABLE 2 | Continued

Subfamily name	Vitis proteome 12x ID	Species	%ID	UniprotKB ID
VvMKK3	GSVIVT01015283001	<i>Ricinus communis</i>	81.2	B9RK49_RICCO
		<i>Petroselinum crispum</i>	79.3	Q6QMT5_PETCR
		<i>Malus domestica</i>	77.6	D1MFM3_MALDO
		<i>Nicotiana tabacum</i>	76.8	Q9M6Q9_TOBAC
		<i>Solanum lycopersicum</i>	76.3	O48616_SOLLC
		<i>Arabidopsis thaliana</i>	75.6	C02ZL0_ARATH
		<i>Glycine max</i>	75.1	Q5JCLO_SOYBN
		<i>Ricinus communis</i>	89.0	B9RKG0_RICCO
		<i>Solanum lycopersicum</i>	88.7	Q66MH7_SOLLC
		<i>Nicotiana tabacum</i>	88.1	Q9AYN9_TOBAC
		<i>Nicotiana benthamiana</i>	87.3	B2NIC2_NICBE
		<i>Origanum onites</i>	86.1	A7U0S8_9LAMI
		<i>Arabidopsis thaliana</i>	83.3	M2K6_ARATH
		<i>Arabidopsis lyrata</i> subsp. <i>lyrata</i>	83.1	D7MLT9_ARALL
VvMKK5	GSVIVT01032414001	<i>Oryza sativa</i> subsp. <i>Japonica</i>	77.3	M2K1_ORYSJ
		<i>Oryza sativa</i> subsp. <i>Índica</i>	77.3	Q0Z7Z4_ORYSI
		<i>Zea mays</i>	77.1	O49975_MAIZE
		<i>Sorghum bicolor</i>	76.8	C5XIE1_SORBI
		<i>Ricinus communis</i>	86.3	B9S641_RICCO
VviMAPKK3	GSVIVT01001193001	<i>Populus trichocarpa</i>	84.6	B9GI57_POPTR
		<i>Nicotiana tabacum</i>	82.8	Q40542_TOBAC
		<i>Suaeda salsa</i>	79.0	Q8L8I2_SUASA
		<i>Arabidopsis thaliana</i>	78.6	O80396_ARATH
		<i>Populus trichocarpa</i>	85.5	B9GSK4_POPTR
VviMAPKK4	GSVIVT01001690001	<i>Ricinus communis</i>	77.8	B9SFH0_RICCO
		<i>Populus trichocarpa</i>	82.2	B9GTK7_POPTR
VviMAPKK8	GSVIVT01007637001	<i>Ricinus communis</i>	80.7	B9T446_RICCO
		<i>Populus trichocarpa</i>	80.5	B9IF6_POPTR
		<i>Glycine max</i>	79.3	B9RAT5_RICCO
		<i>Medicago sativa</i>	78.7	C0M0P4_SOYBN
		<i>Ricinus communis</i>	78.0	Q84RS1_MEDSA
VviMAPKK9	GSVIVT01007646001	<i>Ricinus communis</i>	88.3	B9RAU4_RICCO
		<i>Glycine max</i>	84.7	C6T9D3_SOYBN
		<i>Oryza sativa</i>	82.1	B8AEQ7_ORYSI
		<i>Zea mays</i>	82.1	C0P3M4_MAIZE
		<i>Oryza sativa</i> subsp. <i>Japonica</i>	82.1	Q6ZH81_ORYSJ
VviMAPKK10	GSVIVT01007762001	<i>Arabidopsis thaliana</i>	81.8	Q9FGS7_ARATH
		<i>Populus trichocarpa</i>	88.0	B9IEQ9_POPTR
VviMAPKK12	GSVIVT01008413001	<i>Ricinus communis</i>	84.1	B9RB33_RICCO
		<i>Populus trichocarpa</i>	80.8	B9IEA9_POPTR
VviMAPKK17	GSVIVT01012031001	<i>Ricinus communis</i>	79.8	B9S4I8_RICCO
		<i>Arabidopsis thaliana</i>	76.9	Q9LJD8_ARATH
VviMAPKK25	GSVIVT01018020001	<i>Populus trichocarpa</i>	84.0	B9HUS5_POPTR
		<i>Ricinus communis</i>	79.4	B9RTM1_RICCO
		<i>Arabidopsis thaliana</i>	75.1	Q9LUI6_ARATH
VviMAPKK29	GSVIVT01019739001	<i>Ricinus communis</i>	76.3	B9RCD5_RICCO
		<i>Populus trichocarpa</i>	76.0	B9GKG5_POPTR
VviMAPKK31	GSVIVT01020712001	<i>Ricinus communis</i>	84.0	B9SRD1_RICCO
		<i>Populus trichocarpa</i>	82.4	B9IA51_POPTR
VviMAPKK34	GSVIVT01022098001	<i>Populus trichocarpa</i>	82.7	B9HHA4_POPTR
VviMAPKK40	GSVIVT01023958001	<i>Populus trichocarpa</i>	82.2	B9H1M1_POPTR

(Continued)

TABLE 2 | Continued

Subfamily name	Vitis proteome 12x ID	Species	%ID	UniprotKB ID
VviMAPKK42	GSVIVT01026487001	<i>Ricinus communis</i>	79.8	B9S5G6_RICCO
VviMAPKK45	GSVIVT01028897001	<i>Ricinus communis</i>	77.0	B9SUR2_RICCO
		<i>Populus trichocarpa</i>	83.2	B9RIV9_RICCO
VviMAPKK50	GSVIVT01030194001	<i>Ricinus communis</i>	82.9	B9IDA8_POPTR
		<i>Populus trichocarpa</i>	85.4	B9T3P6_RICCO
		<i>Medicago truncatula</i>	84.3	B9IGR7_POPTR
		<i>Glycine max</i>	84.0	B7FKS6_MEDTR
		<i>Arabidopsis thaliana</i>	81.6	C6TMB8_SOYBN
VviMAPKK54	GSVIVT01032389001	<i>Populus trichocarpa</i>	80.9	Q8L6Y9_ARATH
		<i>Ricinus communis</i>	91.2	B9GI75_POPTR
		<i>Cucumis sativus</i>	89.5	B9S662_RICCO
		<i>Arabidopsis thaliana</i>	82.4	Q7XJ65_CUCSA
VviMAPKK55	GSVIVT01032487001	<i>Populus trichocarpa</i>	81.3	Q9LT56_ARATH
		<i>Ricinus communis</i>	89.2	B9IJN5_POPTR
		<i>Arabidopsis thaliana</i>	88.0	B9RB44_RICCO
		<i>Oryza sativa</i> subsp. <i>Japonica</i>	86.9	Q9SSA4_ARATH
		<i>Oryza sativa</i> subsp. <i>Indica</i>	84.6	Q6L5F3_ORYSJ
		<i>Zea mays</i>	84.6	A2Y7U2_ORYSI
VviMAPKK56	GSVIVT01033779001	<i>Populus trichocarpa</i>	83.3	B6U656_MAIZE
		<i>Prunus salinica</i>	80.9	B9IFS3_POPTR
		<i>Ricinus communis</i>	79.8	A9UAN3_9ROSA
		<i>Prunus persica</i>	79.8	B9SRG7_RICCO
		<i>Rosa hybrid</i> cultivar	78.9	C4PKQ3_PRUPE
		<i>Malus domestica</i>	78.6	Q93XL9_ROSHC
		<i>Solanum lycopersicum</i>	78.5	A2T3V2_MALDO
VviMAPKK58	GSVIVT01034988001	<i>Ricinus communis</i>	77.1	Q5YKK5_SOLLC
		<i>Populus trichocarpa</i>	78.3	B9RZR2_RICCO
VviMAPKK61	GSVIVT01037773001	<i>Ricinus communis</i>	78.0	B9HIN4_POPTR
		<i>Gossypium hirsutum</i>	80.4	B9SSS7_RICCO
		<i>Arabidopsis thaliana</i>	79.5	Q7Y236_GOSHI
VvMAP4K3	GSVIVT01014297001	<i>Ricinus communis</i>	76.8	WNK11_ARATH
		<i>Populus trichocarpa</i>	76.6	B9T3W4_RICCO
VvMAP4K5	GSVIVT01019643001	<i>Populus trichocarpa</i>	76.1	B9N1E7_POPTR
		<i>Ricinus communis</i>	79.8	B9GK86_POPTR
VvMAP4K7	GSVIVT01032461001	<i>Populus trichocarpa</i>	79.3	B9RYT1_RICCO
		<i>Carica papaya</i>	80.0	B9GHQ7_POPTR
		<i>Arabidopsis thaliana</i>	79.0	A7L4B0_CARPA
			75.6	Q9LER4_ARATH

Columns 1–4 contain the protein name, Vitis proteome 12x ID, GenBank ID, species, percentage identity (%ID), UniprotKB ID.

genome coverage (Wang et al., 2014b). The difference in the number of MAPKKK members in grapevine genome may be related to the “E” value > E-120 used in this report, which is more significant. In addition, domain scan using two different databases (PROSITE and CDD) can identify more sequences in the grapevine genome.

The members of the *Vitis* MAPKKK subfamily share 11–35% identity with each other and distributed on various chromosomes (from 2 to 18) (Table 1). The full length *Vitis* MAPKKK sequences range from 175 (VviMAPKK38) to 1397 (VviMAPKK17) amino acids. The phylogenetic analysis of

both *Vitis* and *Arabidopsis* MAPKK sequences shows that this subfamily is categorized into three main groups with bootstrap values up to 93% (Figure 4).

The first group contains MAPKKs whose kinase domains have similarity to MEKK subfamily members (Figure 4) (Jonak et al., 2002). A second group includes Raf subfamily members while a third group presents ZIK subfamily members (Figure 4) (Jonak et al., 2002). In total, there are 21 VviMAPKKs in the MEKK subfamily, while there are 12 in the ZIK subfamily and 29 in the Raf subfamily among the 62 members in the *Vitis* genome.

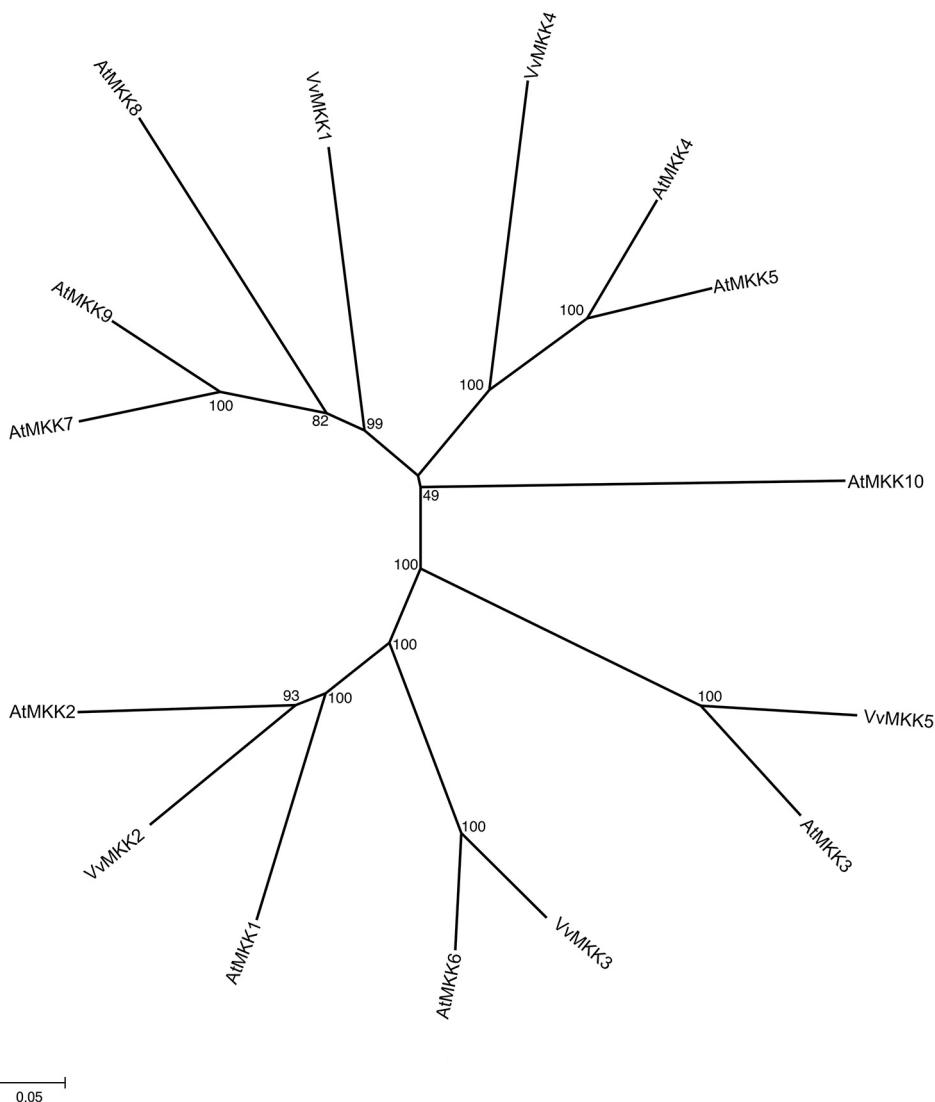


FIGURE 3 | Phylogenetic tree of MAPKK protein sequences from *Arabidopsis* and *Vitis vinifera*. The amino acid sequences of all *Arabidopsis* MAPKK proteins and those of *Vitis vinifera* were aligned using the MUSCLE program and subjected to phylogenetic analysis by the distance with neighborjoining method using MEGA5

programme. Accession numbers for *Arabidopsis* sequences are AtMKK1 (At4g26070), AtMKK2 (At4g29810), AtMKK3 (At5g40440), AtMKK4 (At1g51660), AtMKK5 (At3g21220), AtMKK6 (At5g56580), AtMKK7 (At1g18350), AtMKK8 (At3g06230), AtMKK9 (At1g73500), AtMKK10 (At1g32320).

Analysis of conserved domain of VviMAPKKs identified a long regulatory domain in the N-terminal region and a kinase domain in the C-terminal region in most of VviMAPKKs. It is suggested that the long regulatory domain in the N-terminal region of the Raf subfamily may be involved in protein-protein interactions and regulate or specify their kinase activity (Jouannic et al., 1999). Twenty members of the *Vitis* MAPKK subfamily share 75.1–89.2% similarity with their orthologs from different plant species (Table 2).

We identified at least 640 ESTs for 59 of the *Vitis* MAPKKs (Supplementary Table 3) indicating that MAPKK subfamily is transcriptionally active. Expression profile of VviMAPKK genes suggested that some of them are involved in response to biotic and abiotic stresses in different tissues and organs (Wang et al.,

2014b). In support of a role for some *Vitis* MAPKKs, AtMEKK1 expression is enhanced by drought, salt, stress (Mizoguchi et al., 1996). Recently, it was reported that AtMKK1/MKK2 and AtMEKK1 were able to negatively regulate programmed cell death (PCD) as well as immune responses (Kong et al., 2012). In tobacco, NPK1-MEK1-Ntf6 are also involved in resistance to tobacco mosaic virus (TMV) (Jin et al., 2002; Liu et al., 2004). In addition, AtEDR1, a Raf-like MAPKK could regulate SA-inducible defense responses negatively (Frye et al., 2001).

MAPKKKs

In non-plants, MAPKKs are activated either through phosphorylation by MAPKK kinase (MAPKKK or MAP4K) (Posas and Saito, 1997; Sells et al., 1997) or by G protein and G

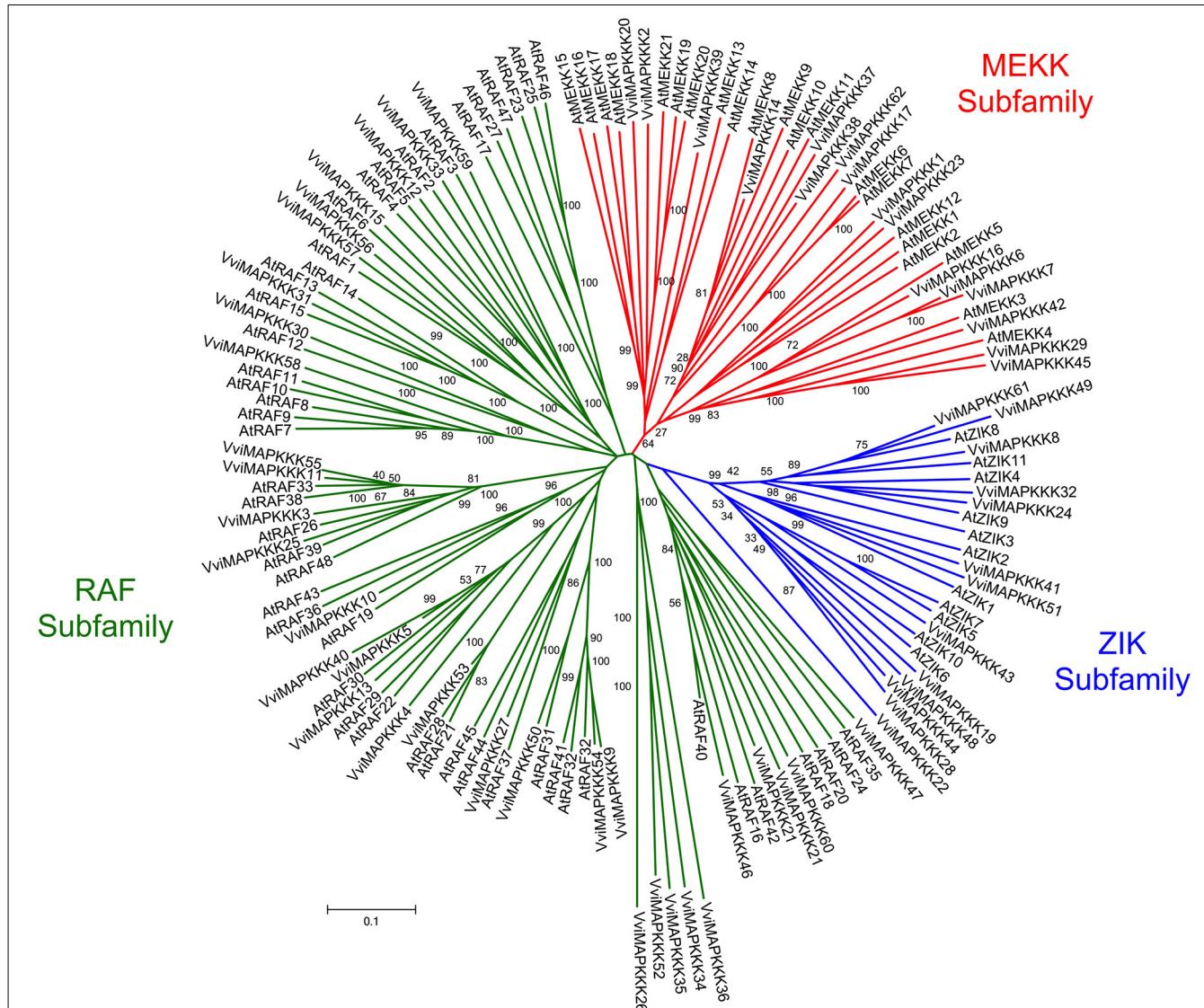


FIGURE 4 | Phylogenetic tree of MAPKK protein sequences from *Arabidopsis* and *Vitis vinifera*. The amino acid sequences of all *Arabidopsis* MAPKKK proteins and those of *Vitis vinifera* were aligned using the MUSCLE program and subjected to phylogenetic analysis by the distance with neighborjoining method using MEGA5 programme. MAPKKK forms the largest group of MAPK cascades with 62 members classified into three subfamilies, MEKK, Raf, and ZIK containing 21, 29, and 12 genes, respectively in *Vitis* genome. Accession numbers for *Arabidopsis* sequences are AtMEKK1 (At1g09000), AtMEKK2 (At1g54960), AtMEKK3 (At1g53570), AtMEKK4 (At1g63700), AtMEKK5 (At5g66850), AtMEKK6 (At3g07980), AtMEKK7 (At3g13530), AtMEKK8 (At4g08500), AtMEKK9 (At4g08480), AtMEKK10 (At4g08470), AtMEKK11 (At4g12020), AtMEKK12 (At3g06030), AtMEKK13 (At1g07150), AtMEKK14 (At2g30040), AtMEKK15 (At5g55090), AtMEKK16 (At4g26890), AtMEKK17 (At2g32510), AtMEKK18 (At1g05100), AtMEKK19 (At5g67080), AtMEKK20 (At3g50310), AtMEKK21 (At4g36950), AtRAF1 (At5g03730), AtRAF2 (At1g08720), AtRAF3 (At5g11850), AtRAF4 (At1g18160), AtRAF5 (At1g73660), AtRAF6

(At4g24480), AtRAF7 (At3g06620), AtRAF8 (At3g06630), AtRAF9 (At3g06640), AtRAF10 (At5g49470), AtRAF11 (At1g67890), AtRAF12 (At4g23050), AtRAF13 (At2g31010), AtRAF14 (At2g42630), AtRAF15 (At3g58640), AtRAF16 (At1g04700), AtRAF17 (At1g14000), AtRAF18 (At1g16270), AtRAF19 (At1g62400), AtRAF20 (At1g79570), AtRAF21 (At2g17700), AtRAF22 (At2g24360), AtRAF23 (At2g31800), AtRAF24 (At2g35050), AtRAF25 (At2g43850), AtRAF26 (At4g14780), AtRAF27 (At4g18950), AtRAF28 (At4g31170), AtRAF29 (At4g35780), AtRAF30 (At4g38470), AtRAF31 (At5g01850), AtRAF32 (At5g40540), AtRAF33 (At5g50000), AtRAF34 (At5g50180), AtRAF35 (At5g57610), AtRAF36 (At5g58950), AtRAF37 (At5g66710), AtRAF38 (At3g01490), AtRAF39 (At3g22750), AtRAF40 (At3g24720), AtRAF41 (At3g27560), AtRAF42 (At3g46920), AtRAF43 (At3g46930), AtRAF44 (At3g50720), AtRAF45 (At3g50730), AtRAF46 (At3g59830), AtRAF47 (At3g58760), AtRAF48 (At3g63260), AtZIK1 (At3g51630), AtZIK2 (At5g58350), AtZIK3 (At3g22420), AtZIK4 (At3g04910), AtZIK5 (At3g18750), AtZIK6 (At5g41990), AtZIK7 (At1g49160), AtZIK8 (At5g55560), AtZIK9 (At5g28080), AtZIK10 (At1g64630), AtZIK11 (At3g48260).

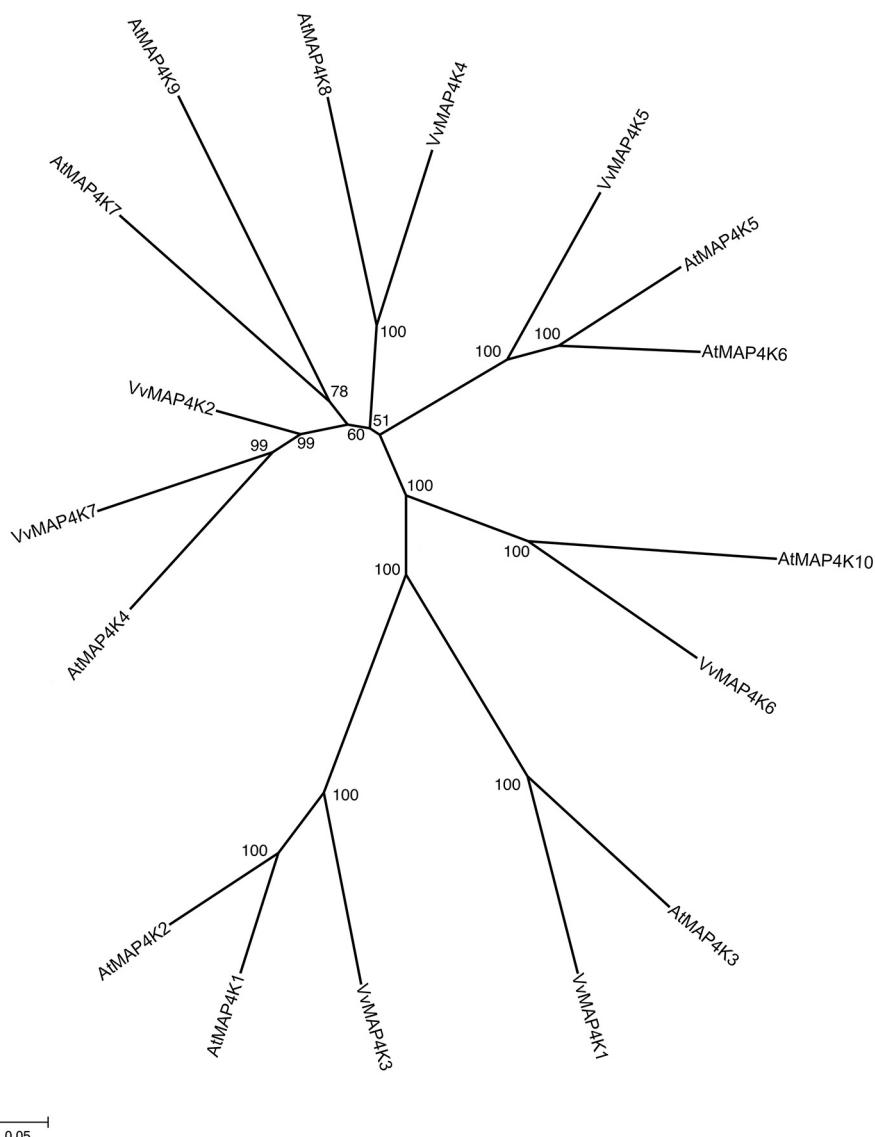


FIGURE 5 | Phylogenetic tree of MAPKKKK protein sequences from *Arabidopsis* and *Vitis vinifera*. The amino acid sequences of all *Arabidopsis* MAPKKKK proteins and those of *Vitis vinifera* were aligned using the MUSCLE program and subjected to phylogenetic analysis by the distance with neighborjoining method using MEGA5 programme.

Accession numbers for *Arabidopsis* sequences are AtMAP4K1 (At1g53165), AtMAP4K2 (At3g15220), AtMAP4K3 (At1g69220), AtMAP4K4 (At5g14720), AtMAP4K5 (At4g24100), AtMAP4K6 (At4g10730), AtMAP4K7 (At1g70430), AtMAP4K8 (At1g79640), AtMAP4K9 (At1g23700), AtMAP4K10 (At4g14480).

protein-coupled receptors (Fanger et al., 1997; Sugden and Clerk, 1997).

Several MAP4Ks have been identified in plant genomes based on phylogenetic analyses of their kinase domain. A MAP4K, named MIK, was characterized from the *Zea mays* (Wang et al., 2014d). Recently, a new MAP4K from GCK-II subfamily named ScMAP4K1, which play important roles in ovule, seed, and fruit development was characterized (Major et al., 2009).

In fully sequenced genomes, like *Arabidopsis* and rice at least 10 protein kinases can be phylogenetically classified as MAP4K (Champion et al., 2004). Little is known about

the roles of MAP4Ks in plants. Seven ORFs showing strong similarity with the 10 *Arabidopsis* MAP4Ks were identified in *Vitis* genome (Figure 5) and shared 18–74% similarity with each other. They have been named VvMAP4K1 through 7 (Table 1). The phylogenetic analysis of *V. vinifera* and *A. thaliana* MAP4Ks proteins identified several orthologs in the two species such as VvMAP4K4/AtMAP4K8 (70% similarity), VvMAP4K1/AtMAP4K3 (66% similarity), VvMAP4K7/AtMAP4K4 (68% similarity), and VvMAP4K6/AtMAP4K10 (64% similarity) (Figure 5).

In addition, we identified several orthologs from different species for 3 VvMAP4Ks (Table 2). Among 7 ORFs encoding

Vitis MAP4Ks, all of them are transcriptionally active (Supplementary Table 4), but none of them has been cloned and characterized.

Conclusions

This report represents the first complete genome-wide analysis of MAPK cascade proteins in grapevine. The identification of *Vitis* MAPK cascade proteins and their comparative analysis with the *Arabidopsis* MAPK cascade proteins indicates that MAPK cascade genes have been conserved during evolution. In this report, we annotated 90 ORFs encoding MAPK cascade proteins in *V. vinifera* using a bioinformatics approach. Taken as a whole, our data provide significant insights into future biological and physiological analysis of MAPK cascades from *V. vinifera*.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Asai, T., Tena, G., Plotnikova, J., Willmann, M. R., Chiu, W. L., Gomez-Gomez, L., et al. (2002). MAP kinase signalling cascade in *Arabidopsis* innate immunity. *Nature* 415, 977–983. doi: 10.1038/415977a

Bergmann, D. C., Lukowitz, W., and Somerville, C. R. (2004). Stomatal development and pattern controlled by a MAPKK kinase. *Science* 304, 1494–1497. doi: 10.1126/science.1096014

Bögre, L., Calderini, O., Binarova, P., Mattauch, M., Till, S., Kiegerl, S., et al. (1999). A MAP kinase is activated late in plant mitosis and becomes localized to the plane of cell division. *Plant Cell* 11, 101–113. doi: 10.1105/tpc.11.1.101

Bush, S. M., and Krysan, P. J. (2007). Mutational evidence that the *Arabidopsis* MAP kinase MPK6 is involved in anther, inflorescence, and embryo development. *J. Exp. Bot.* 58, 2181–2191. doi: 10.1093/jxb/erm092

Çakir, B., and Kılıçkaya, O. (2013). Whole-genome survey of the putative ATP-binding cassette transporter family genes in *Vitis vinifera*. *PLoS ONE* 8:e78860. doi: 10.1371/journal.pone.0078860

Cardinale, F., Meskiene, I., Ouaked, F., and Hirt, H. (2002). Convergence and divergence of stress-induced mitogen-activated protein kinase signaling pathways at the level of two distinct mitogen-activated protein kinase kinases. *Plant Cell* 14, 703–711. doi: 10.1105/tpc.010256

Champion, A., Picaud, A., and Henry, Y. (2004). Reassessing the MAP3K and MAP4K relationships. *Trends Plant Sci.* 9, 123–129. doi: 10.1016/j.tplants.2004.01.005

Colcombet, J., and Hirt, H. (2008). *Arabidopsis* MAPKs: a complex signalling network involved in multiple biological processes. *Biochem. J.* 413, 217–226. doi: 10.1042/BJ20080625

Daldoul, S., Hoefer, M. II, and Mliki, A. (2012). Osmotic Stress Induces the Expression of VvMAP Kinase Gene in Grapevine (*Vitis vinifera* L.). *J. Bot.* 11, 1–4. doi: 10.1155/2012/737035

Dan, I., Watanabe, N. M., and Kusumi, A. (2001). The Ste20 group kinases as regulators of MAP kinase cascades. *Trends Cell Biol.* 11, 220–230. doi: 10.1016/S0962-8924(01)01980-8

Datta, R. S., Meacham, C., Samad, B., Neyer, C., and Sjolander, K. (2009). Berkeley PHOG: PhyloFacts orthology group prediction web server. *Nucleic Acids Res.* 37, W84–W89. doi: 10.1093/nar/gkp373

del Pozo, O., Pedley, K. F., and Martin, G. B. (2004). MAPKKK α is a positive regulator of cell death associated with both plant immunity and disease. *EMBO J.* 23, 3072–3082. doi: 10.1038/sj.emboj.7600283

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Fanger, G. R., Johnson, N. L., and Johnson, G. L. (1997). MEK kinases are regulated by EGF and selectively interact with Rac/Cdc42. *Embo J.* 16, 4961–4972. doi: 10.1093/embj/16.16.4961

Feilner, T., Hultschig, C., Lee, J., Meyer, S., Immink, R. G. H., Koenig, A., et al. (2005). High throughput identification of potential *Arabidopsis* mitogen-activated protein kinases substrates. *Mol. Cell. Proteomics* 4, 1558–1568. doi: 10.1074/mcp.M500007-MCP200

Fiil, B. K., Petersen, K., Petersen, M., and Mundy, J. (2009). Gene regulation by MAP kinase cascades. *Cur. Opin. Plant Biol.* 12, 615–621. doi: 10.1016/j.pbi.2009.07.017

Frye, C. A., Tang, D., and Innes, R. W. (2001). Negative regulation of defense responses in plants by a conserved MAPKK kinase. *Proc. Natl. Acad. Sci. U.S.A.* 98, 373–378. doi: 10.1073/pnas.98.1.373

Galletti, R., Ferrari, S., and De Lorenzo, G. (2011). *Arabidopsis* MPK3 and MPK6 play different roles in basal and oligogalacturonide- or flagellin-induced resistance against *Botrytis cinerea*. *Plant Physiol.* 157, 804–814. doi: 10.1104/pp.111.174003

Gray, J. E., and Hetherington, A. M. (2004). Plant development: YODA the stomatal switch. *Curr. Biol.* 14, R488–R490. doi: 10.1016/j.cub.2004.06.019

Group, M., Ichimura, K., Shinozaki, K., Tena, G., Sheen, J., Henry, Y., Champion, A., et al. (2002). Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends Plant Sci.* 7, 301–308. doi: 10.1016/S1360-1385(02)02302-6

Hamel, L. P., Nicole, M. C., Sritubtim, S., Morency, M. J., Ellis, M., Ehling, J., et al. (2006). Ancient signals: comparative genomics of plant MAPK and MAPKK gene families. *Trends Plant Sci.* 11, 192–198. doi: 10.1016/j.tplants.2006.02.007

Huang, X., Luo, T., Fu, X., Fan, Q., and Liu, J. (2011). Cloning and molecular characterization of a mitogen-activated protein kinase gene from *Poncirus trifoliata* whose ectopic expression confers dehydration/drought tolerance in transgenic tobacco. *J. Exp. Bot.* 62, 5191–5206. doi: 10.1093/jxb/err229

Hutty, A. K., and Phillips, A. L. (1995). Gibberellin-regulated expression in oat aleurone cells of two kinases that show homology to MAP kinase and a ribosomal protein kinase. *Plant Mol. Biol.* 27, 1043–1052. doi: 10.1007/bf00037031

Hwa, C.-M., and Yang, X.-C. (2008). The AtMKK3 pathway mediates ABA and salt signaling in *Arabidopsis*. *Acta Physiol. Plantarum* 30, 277–286. doi: 10.1007/s11738-007-0117-3

Hyun, T. K., Kim, J. S., Kwon, S. Y., and Kim, S. H. (2010). Comparative genomic analysis of mitogen activated protein kinase gene family in grapevine. *Genes Genomics* 32, 275–281. doi: 10.1007/s13258-010-0010-0

Ichimura, K., Mizoguchi, T., Irie, K., Morris, P., Giraudat, J., Matsumoto, K., et al. (1998). Isolation of ATMEKK1 (a MAP kinase kinase kinase)-interacting proteins and analysis of a MAP kinase cascade in *Arabidopsis*. *Biochem. Biophys. Res. Commun.* 253, 532–543. doi: 10.1006/bbrc.1998.9796

Ichimura, K., Mizoguchi, T., Yoshida, R., Yuasa, T., and Shinozaki, K. (2000). Various abiotic stresses rapidly activate *Arabidopsis* MAP kinases ATMPK4 and ATMPK6. *Plant J.* 24, 655–665. doi: 10.1046/j.1365-313x.2000.00913.x

Author Contributions

BC conceived and designed all research. OK performed the bioinformatic analyses. BC analyzed data and wrote the article.

Acknowledgments

This work was funded by the Department of Horticulture, Ege University, Turkey.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.00556>

Ichimura, K., Shinozaki, K., Tena, G., Sheen, J., Henry, Y., Champion, A., et al. (2002). Mapk G: Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends Plant Sci.* 7, 301–308. doi: 10.1016/S1360-1385(02)02302-6

Jaillon, O., Aury, J. M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–467. doi: 10.1038/nature06148

Jin, H., Axtell, M. J., Dahlbeck, D., Ekwenna, O., Zhang, S., Staskawicz, B., et al. (2002). NPK1, and MEKK1-like mitogen-activated protein kinase kinase kinase, regulates innate immunity and development in plants. *Dev. Cell* 3, 291–297. doi: 10.1016/S1534-5807(02)00205-8

Jonak, C., Ökrész, L., Bögö, L., and Hirt, H. (2002). Complexity, Cross Talk and Integration of Plant MAP Kinase Signalling. *Curr. Opin. Plant Biol.* 5, 415–424. doi: 10.1016/S1369-5266(02)00285-6

Jonak, C., Pay, A., Bogre, L., Hirt, H., and Heberle-Bors, E. (1993). The plant homologue of MAP kinase is expressed in a cell cycle-dependent and organ-specific manner. *Plant J.* 3, 611–617. doi: 10.1046/j.1365-313X.1993.03040611.x

Jouannic, S., Hamal, A., Leprince, A. S., Tregear, J. W., Kreis, M., and Henry, Y. (1999). Characterisation of novel plant genes encoding MEKK/STE11 and RAF-related protein kinases. *Gene* 229, 171–181. doi: 10.1016/S0378-1119(99)00012-8

Katou, S., Yoshioka, H., Kawakita, K., Rowland, O., Jones, J. D., Mori, H., et al. (2005). Involvement of PPS3 phosphorylated by elicitor-responsive mitogen-activated protein kinases in the regulation of plant cell death. *Plant Physiol.* 139, 1914–1926. doi: 10.1104/pp.105.066795

Kieber, J. J., Rothenberg, M., Roman, G., Feldmann, K. A., and Ecker, J. R. (1993). CTR1, a negative regulator of the ethylene response pathway in *Arabidopsis*, encodes a member of the raf family of protein kinases. *Cell* 72, 427–441. doi: 10.1016/0092-8674(93)90119-B

Kiegerl, S., Cardinale, F., Siligan, C., Gross, A., Baudouin, E., Liwosz, A., et al. (2000). SIMKK, a mitogen-activated protein kinase (MAPK) kinase, is a specific activator of the salt stress-induced MAPK, SIMK. *Plant Cell* 12, 2247–2258. doi: 10.1105/tpc.12.11.2247

Kishi-Kaboshi, M., Okada, K., Kurimoto, L., Murakami, S., Umezawa, T., Shibuya, N., et al. (2010). A rice fungal MAMP-responsive MAPK cascade regulates metabolic flow to antimicrobial metabolite synthesis. *Plant J.* 63, 599–612. doi: 10.1111/j.1365-313X.2010.04264.x

Knetsch, M. L. W., Wang, M., Snaar-Jagalska, B. E., and Heimovaara-Dijkstra, S. (1996). Abscisic acid induces mitogen-activated protein kinase activation in barley aleuron protoplasts. *Plant Cell* 8, 1061–1067. doi: 10.1105/tpc.8.6.1061

Kong, Q., Qu, N., Gao, M., Zhang, Z., Ding, X., Yang, F., et al. (2012). The MEKK1-MKK1/MKK2-MPK4 kinase cascade negatively regulates immunity mediated by a mitogen-activated protein kinase kinase kinase in *Arabidopsis*. *Plant Cell* 24, 2225–2236. doi: 10.1105/tpc.112.097253

Kovtun, Y., Chiu, W. L., Tena, G., and Sheen, J. (2000). Functional analysis of oxidative stress-activated mitogen-activated protein kinase cascade in plants. *Proc. Natl. Acad. Sci. U.S.A.* 97, 2940–2945. doi: 10.1073/pnas.97.6.2940

Kumar, K., Rao, K. P., Biswas, D. K., and Sinha, A. K. (2011). Rice WNK1 is regulated by abiotic stress and involved in internal circadian rhythm. *Plant Signal. Behav.* 6, 316–320. doi: 10.4161/psb.6.3.13063

Kumar, K. R., and Kirti, P. B. (2010). A mitogen-activated protein kinase, AhMPK6 from peanut localizes to the nucleus and also induces defense responses upon transient expression in tobacco. *Plant Physiol. Biochem.* 48, 481–486. doi: 10.1016/j.plaphy.2010.03.010

Lee, J., Rudd, J. J., Macioszek, V. K., and Scheel, D. (2004). Dynamic changes in the localization of MAPK cascade components controlling pathogenesis-related (PR) gene expression during innate immunity in parsley. *J. Biol. Chem.* 279, 22440–22448. doi: 10.1074/jbc.M401099200

Leprince, A., Jouannic, S., Hamal, A., Kreis, M., and Henry, Y. (1999). Molecular characterisation of plant cDNAs BnMAP4K alpha 1 and BnMAP4K alpha 2 belonging to the GCK/SPS1 subfamily of MAP kinase kinase kinase kinase. *Biochim. Biophys. Acta* 1444, 1–13. doi: 10.1016/S0167-4781(98)00246-2

Liu, Q., and Xue, Q. (2007). Computational identification and phylogenetic analysis of the MAPK gene family in *Oryza sativa*. *Plant Physiol. Biochem.* 45, 6–14. doi: 10.1016/j.plaphy.2006.12.011

Liu, Y. (2012). Roles of mitogen-activated protein kinase cascades in ABA signaling. *Plant Cell Rep.* 31, 1–12. doi: 10.1007/s00299-011-1130-y

Liu, Y., Schiff, M., and Dinesh-Kumar, S. P. (2004). Involvement of MEK1 MAPKK, NTF6 MAPK, WRKY/MYB transcription factors, COI1 and CTR1 in N-mediated resistance to tobacco mosaic virus. *Plant J.* 38, 800–809. doi: 10.1111/j.1365-313X.2004.02085.x

Llompart, B., Castells, E., Río, A., Roca, R., Ferrando, A., Stiefel, V., et al. (2003). The direct activation of MIK, a germinal center kinase (GCK)-like kinase, by MARK, a maize atypical receptor kinase, suggests a new mechanism for signaling through kinase-dead receptors. *J. Biol. Chem.* 28, 48105–48111. doi: 10.1074/jbc.M307482200

Lukowitz, W., Roeder, A., Parmenter, D., and Somerville, C. (2004). A MAPKK kinase gene regulates extra-embryonic cell fate in *Arabidopsis*. *Cell* 116, 109–119. doi: 10.1016/S0092-8674(03)01067-5

Lynch, M., O'Hely, M., Walsh, B., and Force, A. (2001). The probability of preservation of a newly arisen gene duplicate. *Genetics* 159, 1789–1804.

Major, G., Daigle, C., Lafleur, E., Caron, S., and Matton, D. (2009). Characterization of ScMAP4K1, a MAP kinase kinase kinase involved in ovule, seed and fruit development in *Solanum chacoense* Bitt. *Curr. Topics Plant Biol.* 10, 27–46.

Marchler-Bauer, A., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., et al. (2009). CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res.* 37, D205–D210. doi: 10.1093/nar/gkn845

Matsuoka, D., Nanmori, T., Sato, K., Fukami, Y., Kikkawa, U., and Yasuda, T. (2002). Activation of AtMEK1, an *Arabidopsis* mitogen-activated protein kinase kinase, *in vitro* and *in vivo*: analysis of active mutants expressed in *E. coli* and generation of the active form in stress response in seedlings. *Plant J.* 29, 637–647. doi: 10.1046/j.0960-7412.2001.01246.x

Melech-Bonfil, S., and Sessa, G. (2010). Tomato MAPKKepsilon is a positive regulator of cell-death signaling networks associated with plant immunity. *Plant J.* 64, 379–391. doi: 10.1111/j.1365-313X.2010.04333.x

Meng, X., Wang, H., He, Y., Liu, Y., Walker, J. C., Torii, K. U., et al. (2012). A MAPK cascade downstream of ERECTA receptor-like protein kinase regulates *Arabidopsis* inflorescence architecture by promoting localized cell proliferation. *Plant Cell* 24, 4948–4960. doi: 10.1105/tpc.112.104695

Mizoguchi, T., Ichimura, K., Irie, K., Morris, P., Giraudat, J., Matsumoto, K., et al. (1998). Identification of a possible MAP kinase cascade in *Arabidopsis thaliana* based on pairwise yeast two-hybrid analysis and functional complementation tests of yeast mutants. *FEBS Lett.* 437, 56–60. doi: 10.1016/S0014-5793(98)01197-1

Mizoguchi, T., Irie, K., Hirayama, T., Hayashida, N., Yamaguchi-Shinozaki, K., Matsumoto, K., et al. (1996). A gene encoding a mitogen-activated protein kinase kinase kinase is induced simultaneously with genes for a mitogen-activated protein kinase and an S6 ribosomal protein kinase by touch, cold, and water stress in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 93, 765–769. doi: 10.1073/pnas.93.2.765

Moreno-Hagelsieb, G., and Latimer, K. (2008). Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* 24, 319–324. doi: 10.1093/bioinformatics/btm585

Nadarajah, K., and Sidek, H. (2010). The green MAPKs. *Asian J. Plant Sci.* 9, 1–10. doi: 10.3923/ajpls.2010.1.10

Nakagami, H., Kiegerl, S., and Hirt, H. (2004). OMTK1, a novel MAPKKK, channels oxidative stress signaling through direct MAPK interaction. *J. Biol. Chem.* 279, 26959–26966. doi: 10.1074/jbc.M312662200

Ning, J., Li, X., Hicks, L. M., and Xiong, L. (2010). A raf-like MAPKK gene DSM1 mediates drought resistance through reactive oxygen species scavenging in rice. *Plant Physiol.* 152, 876–890. doi: 10.1104/pp.109.149856

Nishihama, R., Ishikawa, M., Araki, S., Soyano, T., Asada, T., and Machida, Y. (2001). The NPK1 mitogen-activated protein kinase kinase kinase is a regulator of cell-plate formation in plant cytokinesis. *Genes Dev.* 15, 352–363. doi: 10.1101/gad.863701

Nowak, M. A., Boerlijst, M. C., Cooke, J., and Smith, J. M. (1997). Evolution of genetic redundancy. *Nature* 388, 167–171. doi: 10.1038/40618

Oh, C.-S., Pedley, K. F., and Martin, G. B. (2010). Tomato 14-3-3 protein 7 positively regulates immunity-associated programmed cell death by enhancing protein abundance and signaling ability of MAPKKKa. *Plant Cell* 22, 260–272. doi: 10.1105/tpc.109.070664

Ortiz-Masia, D., Perez-Amador, M. A., Carbonell, J., and Marcote, M. J. (2007). Diverse stress signals activate the C1 subgroup MAP kinases of *Arabidopsis*. *FEBS Lett.* 581, 1834–1840. doi: 10.1016/j.febslet.2007.03.075

Pedley, K. F., and Martin, G. B. (2005). Role of mitogen-activated protein kinases in plant immunity. *Curr. Opin. Plant Biol.* 8, 541–547. doi: 10.1016/j.pbi.2005.07.006

Posas, F., and Saito, H. (1997). Osmotic activation of the HOG MAPK pathway via Ste11p MAPKKK: Scaffold role of Pbs2p MAPKK. *Science* 276, 1702–1705. doi: 10.1126/science.276.5319.1702

Qi, M., and Elion, E. A. (2005). MAP kinase pathways. *J. Cell Sci.* 118, 3569–3572. doi: 10.1242/jcs.02470

Qiu, J. L., Filii, B. K., Petersen, K., Nielsen, H. B., Botanga, C. J., Thorgrimsen, S., et al. (2008). *Arabidopsis* MAP kinase 4 regulates gene expression through transcription factor release in the nucleus. *Embo J.* 27, 2214–2221. doi: 10.1038/embj.2008.147

Rao, K. P., Richa, T., Kumar, K., Raghuram, B., and Sinha, A. K. (2010). *In silico* analysis reveals 75 members of mitogen-activated protein kinase kinase kinase gene family in rice. *DNA Res.* 17, 139–153. doi: 10.1093/dnares/dsq011

Ren, D., Yang, H., and Zhang, S. (2002). Cell death mediated by MAPK is associated with hydrogen peroxide production in *Arabidopsis*. *J. Biol. Chem.* 277, 559–565. doi: 10.1074/jbc.M109495200

Rodriguez, M., Petersen, M., and Mundy, J. (2010). Mitogen-activated protein kinase signaling in plants. *Annu. Rev. Plant Biol.* 61, 621–649. doi: 10.1146/annurev-arplant-042809-112252

Rohila, J. S., and Yang, Y. (2007). Rice Mitogen-activated protein kinase gene family and its role in biotic and abiotic stress response. *J. Integr. Plant Biol.* 49, 751–759. doi: 10.1111/j.1744-7909.2007.00501.x

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.

Sells, M. A., Knaus, U. G., Bagrodia, S., Ambrose, D. M., Bokoch, G. M., and Chernoff, J. (1997). Human p21-activated kinase (Pak1) regulates actin organization in mammalian cells. *Curr. Biol.* 7, 202–210. doi: 10.1016/S0960-9822(97)70091-5

Song, F., and Goodman, R. M. (2002). OsBIMK1, a rice MAP kinase gene involved in disease resistance responses. *Planta* 215, 997–1005. doi: 10.1007/s00425-002-0794-5

Stulemeijer, I. J. E., Stratmann, J. W., and Joosten, M. H. A. J. (2007). Tomato mitogen-activated protein kinases LeMPK1, LeMPK2, and LeMPK3 are activated during the Cf-4/Avr4-Induced hypersensitive response and have distinct phosphorylation specificities. *Plant Physiol.* 144, 1481–1494. doi: 10.1104/pp.107.101063

Sugden, P. H., and Clerk, A. (1997). Regulation of the ERK subgroup of MAP kinase cascades through G protein-coupled receptors. *Cell Signal.* 9, 337–351. doi: 10.1016/S0898-6568(96)00191-X

Sun, Y., Wang, C., Yang, B., Wu, F., Hao, X., Liang, W., et al. (2014). Identification and functional analysis of mitogen-activated protein kinase kinase kinase (MAPKKK) genes in canola (*Brassica napus* L.). *J. Exp. Bot.* 65, 2171–2188. doi: 10.1093/jxb/eru092

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121

Teige, M., Scheikl, E., Eulgem, T., Doczi, R., Ichimura, K., Shinozaki, K., et al. (2004). The MKK2 pathway mediates cold and salt stress signaling in *Arabidopsis*. *Mol. Cell* 15, 141–152. doi: 10.1016/j.molcel.2004.06.023

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Wang, G., Lovato, A., Liang, Y. H., Wang, M., Chen, F., Tornielli, G. B., et al. (2014a). Validation by isolation and expression analyses of MAPK gene family in grapevine (*Vitis vinifera*). *Aust. J. Grape Wine Res.* 20, 255–262. doi: 10.1111/ajgw.12081

Wang, G., Lovato, A., Polverari, A., Wang, M., Liang, Y. H., Ma, Y. C., et al. (2014b). Genome-wide identification and analysis of mitogen activated protein kinase kinase kinase gene family in grapevine (*Vitis vinifera*). *BMC Plant Biol.* 14:219. doi: 10.1186/s12870-014-0219-1

Wang, J., Pan, C., Wang, Y., Ye, L., Wu, J., Chen, L., et al. (2015). Genome-wide identification of MAPK, MAPKK, and MAPKKK gene families and transcriptional profiling analysis during development and stress response in cucumber. *BMC Genomics* 16:386. doi: 10.1186/s12864-015-1621-2

Wang, L., Liu, Y., Cai, G., Jiang, S., Pan, J., and Li, D. (2014d). Ectopic expression of ZmSIMK1 leads to improved drought tolerance and activation of systematic acquired resistance in transgenic tobacco. *J. Biotechnol.* 20, 18–29. doi: 10.1016/j.biote.2013.11.006

Wang, L., Su, H., Han, L., Wang, C., Sun, Y., and Liu, F. (2014c). Differential expression profiles of poplar MAP kinase kinases in response to abiotic stresses and plant hormones, and overexpression of PtMKK4 improves the drought tolerance of poplar. *Gene* 51, 141–148. doi: 10.1016/j.gene.2014.04.058

Wang, X. J., Zhu, S. Y., Lu, Y. F., Zhao, R., Xin, Q., Wang, X. F., et al. (2010). Two coupled components of the mitogen-activated protein kinase cascade MdMPK1 and MdMKK1 from apple function in ABA signal transduction. *Plant Cell Physiol.* 51, 754–766. doi: 10.1093/pcp/pcq037

Wang, Y., Liu, K., Liao, H., Zhuang, C., Ma, H., and Yan, X. (2008). The plant WNK gene family and regulation of flowering time in *Arabidopsis*. *Plant Biol.* 10, 548–562. doi: 10.1111/j.1438-8677.2008.00072.x

Wilson, C., Voronin, V., Touraev, A., Vicente, O., and Heberle-Bors, E. (1997). A developmentally regulated MAP kinase activated by hydration in tobacco pollen. *Plant Cell* 9, 2093–2100. doi: 10.1105/tpc.9.11.2093

Yang, K.-Y., Liu, Y., and Zhang, S. (2001). Activation of a mitogen-activated protein kinase pathway is involved in disease resistance in tobacco. *Proc. Natl. Acad. Sci. U.S.A.* 98, 741–746. doi: 10.1073/pnas.98.2.741

Yuasa, T., Ichimura, K., Mizoguchi, T., and Shinozaki, K. (2001). Oxidative stress activates ATMPK6, an *Arabidopsis* homologue of MAP kinase. *Plant Cell Physiol.* 42, 1012–1016. doi: 10.1093/pcp/pce123

Zaidi, I., Ebel, C., Touzri, M., Herzog, E., Evrard, J.-L., Schmit, A., et al. (2010). TMK1 is a novel wheat stress responsive MAP kinase phosphatase localized in the nucleus. *Plant Mol. Biol.* 73, 325–338. doi: 10.1007/s11103-010-9617-4

Zhang, S., and Klessig, D. F. (1997). Salicylic acid activates a 48-kD MAP kinase in tobacco. *Plant Cell* 9, 809–824. doi: 10.1105/tpc.9.5.809

Zhang, S., and Klessig, D. F. (2001). MAPK cascades in plant defense signaling. *Trends Plant Sci.* 6, 520–527. doi: 10.1016/S1360-1385(01)02103-3

Conflict of Interest Statement: 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.

Copyright © 2015 Çakır and Kılıçkaya. 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) or licensor 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.

Biodiversity of genes encoding anti-microbial traits within plant associated microbes

Walaa K. Mousa^{1,2} and Manish N. Raizada^{1*}

¹ Department of Plant Agriculture, University of Guelph, Guelph, ON, Canada, ² Department of Pharmacognosy, Faculty of Pharmacy, Mansoura University, Mansoura, Egypt

The plant is an attractive versatile home for diverse associated microbes. A subset of these microbes produces a diversity of anti-microbial natural products including polyketides, non-ribosomal peptides, terpenoids, heterocyclic nitrogenous compounds, volatile compounds, bacteriocins, and lytic enzymes. In recent years, detailed molecular analysis has led to a better understanding of the underlying genetic mechanisms. New genomic and bioinformatic tools have permitted comparisons of orthologous genes between species, leading to predictions of the associated evolutionary mechanisms responsible for diversification at the genetic and corresponding biochemical levels. The purpose of this review is to describe the biodiversity of biosynthetic genes of plant-associated bacteria and fungi that encode selected examples of antimicrobial natural products. For each compound, the target pathogen and biochemical mode of action are described, in order to draw attention to the complexity of these phenomena. We review recent information of the underlying molecular diversity and draw lessons through comparative genomic analysis of the orthologous coding sequences (CDS). We conclude by discussing emerging themes and gaps, discuss the metabolic pathways in the context of the phylogeny and ecology of their microbial hosts, and discuss potential evolutionary mechanisms that led to the diversification of biosynthetic gene clusters.

Keywords: genes, biodiversity, evolution, plant associated microbes, rhizosphere, endophyte, antimicrobial secondary metabolites

OPEN ACCESS

Edited by:

Joanna Marie-France Cross,
Inönü University, Turkey

Reviewed by:

Nancy Keller,
University of Wisconsin, USA
Antoine Danchin,
Amabiotics SAS, France

*Correspondence:

Manish N. Raizada,
Department of Plant Agriculture,
University of Guelph, 50 Stone Road
East, Guelph, ON, N1G 2W1, Canada
raizada@uoguelph.ca

Specialty section:

This article was submitted to Plant Genetics and Genomics, a section of the journal *Frontiers in Plant Science*

Received: 21 January 2015

Accepted: 23 March 2015

Published: 10 April 2015

Citation:

Mousa WK and Raizada MN (2015) Biodiversity of genes encoding anti-microbial traits within plant associated microbes. *Front. Plant Sci.* 6:231.
doi: 10.3389/fpls.2015.00231

Introduction

The plant is an attractive versatile home for diverse microbes that can colonize internal plant tissues (endophytes), live on the surface (epiphytes) or in the soil surrounding the root system (rhizosphere microbiota) (Barea et al., 2005; Johnston-Monje and Raizada, 2011). Plant associated microbes have the potential to be used as biocontrol, the use of living organisms to suppress crop disease (Eilenberg, 2006) through various mechanisms including the production of antibiotics (Compant et al., 2005). Diverse classes of antimicrobial secondary metabolites of microbial origin have been reported (Mousa and Raizada, 2013), including polyketides, non-ribosomal peptides, terpenoids, heterocyclic nitrogenous compounds, volatile compounds, bacteriocins as well as lytic enzymes. Polyketides and non-ribosomal peptides constitute the majority of microbial derived natural products (Cane, 1997). Interestingly, the tremendous structural diversity of antimicrobial secondary metabolites originated via limited metabolic pathways utilizing few primary metabolites as precursors (Keller et al., 2005). Underlying the diversification of antimicrobial metabolites must have

been a corresponding genetic diversification of ancestral genes driven by co-evolutionary pressures (Vining, 1992).

The revolution in genomics, genome mining tools and bioinformatics offers a new opportunity to connect biochemical diversity to the underlying genetic diversity and to analyze the evolutionary events leading to biodiversity (Zotchev et al., 2012; Scheffler et al., 2013; Deane and Mitchell, 2014).

The scope of this review is to describe the biodiversity of biosynthetic coding sequences (CDS) of plant-associated microbes (bacteria and fungi) that encode selected examples of antimicrobial secondary metabolites and lytic enzymes. For each example, the target pathogen(s) and mode of action are described where known, in order to highlight the diversity of biochemical targets. Out of necessity, the review focuses on compounds for which in depth molecular analysis has been conducted. We review data pertaining to the underlying molecular diversity and highlight comparative genomic data of the orthologous genes. The review concludes with a discussion of common themes and gaps in the literature, and discusses the role of evolution in the diversification of biosynthetic gene clusters including horizontal gene transfer (HGT).

Biosynthetic Genes Encode Diverse Chemical Classes of Anti-Microbial Compounds

The diversity of compounds described in this review, the underlying genes, microbes, and pathogenic targets are summarized (Tables 1, 2).

Polyketides

The structures of Polyketides described in this review are shown (Figure 1)

2,4-Diacetylphloroglucinol

2,4-diacetylphloroglucinol (2,4-DAPG) is a well-studied fluorescent polyketide metabolite produced by many strains of fluorescent *Pseudomonas* spp. that contributes to disease-suppressive soils of crops (McSpadden Gardener et al., 2000; Mavrodi et al., 2001). 2,4-DAPG is synthesized by the condensation of three molecules of acetyl coenzyme A and one molecule of malonyl coenzyme A to produce the precursor monoacetylphloroglucinol (MAPG) (Shanahan et al., 1992). In *P. fluorescens* strain Q2-87, four coding sequences (CDS) within the *phl* operon are responsible for biosynthesis of 2,4-DAPG: a single CDS (*phlD*) encoding a type III polyketide synthase is responsible for the production of phloroglucinol from the condensation of three acetyl-CoAs, and then three CDS (*phlACB*) encoding acetyltransferases are sufficient to convert phloroglucinol to 2,4-DAPG via MAPG (Bangera and Thomashow, 1999; Yang and Cao, 2012). It was suggested that the peptides encoded by *phlACB* may exist as a multi-enzyme complex (Bangera and Thomashow, 1999). *phlD* has been the subject of interest, because it has homology to chalcone and stilbene synthases from plants, which suggests horizontal gene transfer (HGT) between plants and their rhizosphere microbial populations (Bangera and Thomashow, 1999). Whereas, *phlACB* coding sequences are highly conserved between eubacteria and

archaeabacteria (Picard et al., 2000), a considerable degree of polymorphism was reported for *phlD* (Mavrodi et al., 2001). *phlA* transcription is negatively regulated by the product of *phlF* (Delany et al., 2000) which also appears to mediate repression by fusaric acid (Delany et al., 2000), a metabolite of pathogenic fungi of plants, that has previously been implicated in repression of biosynthesis of the anti-fungal compound, phenazine (see above) (van Rij et al., 2005). These observations demonstrate the ongoing arms race between plants, their fungal pathogens and associated anti-fungal antagonists, leading to gene diversification.

Mupirocin

The polyketide mupirocin or pseudomonic acid is one of the major antibacterial metabolites produced by *Pseudomonas fluorescens* (Fuller et al., 1971) and is widely used as a clinical antibiotic (Gurney and Thomas, 2011). Mupirocin can inhibit the growth of methicillin resistant *Staphylococci*, *Streptococci*, *Haemophilus influenzae*, and *Neisseria gonorrhoeae* (Sutherland et al., 1985). In terms of the mode of action, mupirocin inhibits isoleucyl-tRNA synthetase, and hence prevents incorporation of isoleucine into newly synthesized proteins, thus terminating protein synthesis (Hughes and Mellows, 1980). Biochemically, mupirocin has a unique chemical structure that contains a C9 saturated fatty acid (9-hydroxynonanoic acid) linked to C17 monic acid A (a heptaketide) by an ester linkage (Whatling et al., 1995). Mupirocin is derived from acetate units incorporated into monic acid A and 9-hydroxynonanoic acid via polyketide synthesis (Whatling et al., 1995). At the molecular level, the mupirocin biosynthetic gene cluster (*mup* operon) in *P. fluorescens* is complex, and includes 6 Type I polyketide synthases that are multi-functional as well as 29 proteins of single function within a 65 kb region, which are incorporated into 6 larger coding sequences (modules *mmpA-F*) (El-Sayed et al., 2003; Gurney and Thomas, 2011). The gene cluster is non-standard as the CDS are not in the same order as the biosynthetic steps (El-Sayed et al., 2003; Gurney and Thomas, 2011). The acyltransferase (AT) domains of the polyketide synthases (PKS) are not present in each genetic module but are instead encoded by a separate CDS (from the *mmpC* module) and this classifies these PKS as *in-trans* AT PKSs (El-Sayed et al., 2003). With respect to gene regulation, two putative regulatory genes, *mupR* and *mupI*, were identified within the cluster that are involved in quorum sensing (QS) dependent regulation (El-Sayed et al., 2001).

An interesting feature of this system in *P. fluorescens* is that self-resistance to mupirocin is also encoded by a CDS (*mupM*) within the biosynthetic gene cluster (El-Sayed et al., 2003). *mupM* encodes a resistant Ile t-RNA synthetase (IleS) due to polymorphisms within the binding site of mupirocin (El-Sayed et al., 2003; Gurney and Thomas, 2011). A second resistant IleS was cloned from *P. fluorescens* NCIMB 10586 outside of the *mup* gene cluster which showed 28% similarity to the *mupM* product (Yanagisawa et al., 1994). Human pathogens that have high level mupirocin-resistance are associated with an additional gene that encode a novel IleS with similarity to eukaryotic counterparts; this resistance gene is associated with transposable elements and is carried on plasmids, facilitating its rapid spread (Eltringham, 1997; Gurney and Thomas, 2011).

TABLE 1 | Summary of anti-microbial compounds belonging to genetic pathways that show evidence of horizontal gene transfer (HGT).

Compound	Chemical class	Underlying genes*	Example of producing microbe	Best studied pathogen target	Significance to plant /human	References
2,4-DAFG	Phenolic polyketide	<i>phID</i> and <i>phACB</i> operon	<i>Pseudomonas fluorescens</i> ^a	<i>Gaeumannomyces graminis</i> var. <i>tritici</i>	Disease suppressive soils of crops	McSpadden Gardener et al., 2000; Mavrodin et al., 2001
Agrocin 84	Nucleotide analog	agn operon	<i>Agrobacterium radiobacter</i> ^b	<i>Agrobacterium</i> sp.	Biocontrol of crown gall	Roberts and Tate, 1977; Wang et al., 1994
Chitinases(family 18)	Lytic enzymes	See text	<i>P. fluorescens</i> ^a <i>Actinoplanes missouriensis</i> <i>Streptomyces</i> sp. <i>Bacillus subtilis</i> <i>Stenotrophomonas maltophilia</i>	<i>Colletotrichum falcatum</i> <i>Plectosporum tabacobium</i>	Suppression of crop diseases such as lupin root rot	Viswanathan and Samiyappan, 2001; Kobayashi et al., 2002; El-Farably, 2003; Guecine et al., 2008
Chitinases(family 19)	Lytic enzymes	See text	<i>Green and purple bacteria</i> Actinobacteria	Various fungal pathogens	Suppression of crop diseases	Prakash et al., 2010
Hevolic acid	Fusidane triterpene	See text	<i>Aspergillus fumigatus</i>	<i>Magnaporthe oryzae</i>	Rice blast disease	Lodeiro et al., 2009; Zhao et al., 2010
Loline	Indole alkaloid	LOL-1 and LOL-2	<i>Neotyphodium uncinatum</i> [*] Epichloë	Insects	Protects its host plants from insects	Blankenship et al., 2001; Schardl, 2010
Mupirocin	Polyketide	<i>mup</i> operon	<i>P. fluorescens</i>	Methicillin resistant <i>Staphylococcus</i> and <i>Streptococci</i> <i>Haemophilus influenzae</i> <i>Neisseria gonorrhoeae</i>	Clinical use for human	Fuller et al., 1971; Sutherland et al., 1985; Whatling et al., 1995; Gurney and Thomas, 2011
Paclitaxel	Diterpene	See text	<i>Penicillium aurantiogriseum</i>	<i>Heterobasidion annosum</i> <i>Phaeolus schweinitzii</i> <i>Perennipora subacida</i>	Anticancer	Zhou et al., 2010; Sollman et al., 2013
Phenazines	Heterocyclic nitrogenous compounds	<i>phzADEFG</i> operon	<i>P. fluorescens</i> ^a <i>P. aeruginosa</i> <i>Burkholderia cepacia</i> <i>Pantoea agglomerans</i> (<i>Erwinia herbicola</i>)	<i>Rhizoctonia solani</i> <i>Gaeumannomyces graminis</i> var. <i>tritici</i> <i>Pythium</i> spp. <i>Fusarium oxyphorum</i>	Combat soil borne pathogens	Ligon et al., 2000; Gunjiddaiah et al., 1986; Anjiah et al., 1998; McDonald et al., 2001; Mavrodin et al., 2013
Pyoluteorin	Phenolic polyketide	<i>pitABCDEG</i> operon	<i>P. fluorescens</i> ^a	<i>P. ultimum</i>	Damping-off disease in cotton	Howell and Stipanovic, 1980

(Continued)

TABLE 1 | Continued

Compound	Chemical class	Underlying genes*	Example of producing microbe	Best studied pathogen target	Significance to plant /human	References
Pyroliptin	Chlorinated phenylpyrrole	<i>pmABC</i> operon	<i>P. fluorescens</i> ^a <i>B. pyrrcinia</i>	<i>Trichophyton</i> sp. <i>Botryosphaeriaceae</i> <i>Rhizoctonia solani</i> <i>G. graminis</i> var. <i>tritic</i>	Treatment of skin mycoses	Arima et al., 1964; Hammer et al., 1993; Chernin et al., 1996; El-Banna and Winkeleman, 1998; Kimerer et al., 1998; Hammer et al., 1999; Tazawa et al., 2000
Zwittermicin A	Polyketide/nonribosomal peptide hybrid	<i>zmaA</i> – <i>zmaV</i> and <i>kabR</i> , <i>kabA</i> – <i>kabD</i>	<i>B. cereus</i> ^b <i>B. thuringiensis</i>	<i>Phytophthora medicaginis</i>	Contro root rot in alfalfa and chickpeas	Raffel et al., 1996; Slo-Suh et al., 1998; Keyany et al., 2009

*Indicates that the genes noted have been studied in this species.

^a*P. fluorescens* (Martínez-García et al., 2015)

^b*Agrobacterium radiobacter* (Zhang et al., 2014)

^c*B. cereus* (Takano et al., 2012).

There is also genetic evidence that the entire *mup* gene cluster in *Pseudomonas* arose by horizontal gene transfer; specifically the genes encoding tRNA_{Val} and tRNA_{Asp} were found upstream of the *mupA* promoter region leading to speculation that the *mup* cluster arose from homologous recombination between chromosomal tRNA genes and possibly a plasmid containing the *mup* cluster (El-Sayed et al., 2003). The inclusion of a resistant IleS (*mupM*) within the *mup* biosynthetic cluster might have facilitated such horizontal gene transfer, as otherwise uptake of the mupirocin gene cluster would have been immediately suicidal.

Difficidin

Difficidin is a polyketide with an interesting geometry that involves four double bonds in the Z configuration (Chen et al., 2006). Difficidin is produced by various *Bacillus* species such as *B. subtilis* and *B. amyloliquefaciens* FZB 42 with broad antibacterial activity against human and crop pathogens (Zimmerman et al., 1987; Chen et al., 2006, 2009). A large gene cluster (*pks3*) encoding difficidin (and oxydifficidin) was characterized in *B. amyloliquefaciens* (Chen et al., 2006). This compound is included in this review, because *pks3* is adjacent to other polyketide synthesis gene clusters, *pks1* and *pks2*, that encode bacillaene and macro-lactin, respectively (Chen et al., 2006; Schneider et al., 2007). All three gene clusters share sequence homology, a similar order of CDS and are located close to another on the chromosome, leading Chen et al. (2006) to hypothesize that they emerged from homologous recombination from a common ancestral gene cluster resulting in gene duplication. This system provides insights into the diversification of polyketides.

Pyoluteorin

Pyoluteorin (PLt) is a phenolic polyketide with bactericidal, herbicidal, and fungicidal properties. PLt can suppress damping-off disease in cotton caused by the fungus, *Pythium ultimum* (Howell and Stipanovic, 1980). Both PLt and phenazine (see below) may act synergistically to suppress such soil-borne fungal diseases in plants, as some studies have suggested that the two biosynthetic pathway interact with one another (Ge et al., 2007; Lu et al., 2009). The biosynthesis of PLt involves condensation of proline with three acetate equivalents through chlorination and oxidation. The carbon skeleton is built up by the action of a single multienzyme complex (Nowak-Thompson et al., 1999). In *Pseudomonas fluorescens* Pf-5, a 24 kb segment contains the PLt biosynthetic operon (*pltABCDEFG*). PLt biosynthesis is catalyzed by type I polyketide synthases (*pltB*, *pltC*), an acyl-CoA dehydrogenase (*pltE*), an acyl-CoA synthetase (*pltF*), a thioesterase (*pltG*), and halogenases (*pltA*, *pltD*, *pltM*) with *pltM* located adjacent to the gene cluster (Nowak-Thompson et al., 1999). A significant delay in the expression of the PLt biosynthetic operon was reported in the cucumber rhizosphere compared to cotton, which correlated to the timing of infection with the fungal root pathogen *Pythium ultimum* (Kraus and Loper, 1995). The authors suggest that such temporal differences may be responsible for differential disease suppression in diverse plant hosts.

The *plt* biosynthetic operon has been shown to be regulated by a LysR family transcriptional activator, encoded by *pltR* (Nowak-Thompson et al., 1999). Interestingly, *pltR* is tightly linked and

TABLE 2 | Summary of anti-microbial compounds belonging to genetic pathways, grouped by different evolutionary levels of diversification.

Compound	Chemical class	Underlying genes ^a	Example of producing microbe	Best studied pathogen target	Significance to plant /human	References
GENOME LEVEL DIVERSIFICATION						
Difficidin	Polyketide	PKS3	<i>B. amylolyticfaciens</i> ^a <i>B. subtilis</i>	Broad spectrum antibacterial	Crop and human pathogens	Zimmerman et al., 1987; Chen et al., 2006, 2009.
Loline	Indole alkaloid	LOL-1 and LOL-2	<i>Neotyphodium uncinatum</i> [*] Epichloë	Insects	Protects its host plants from insects	Blankenship et al., 2001; Schardl, 2010
INTRA GENE CLUSTER DIVERSIFICATION						
Agrocin 84	Nucleotide analog	agn operon	<i>Agrobacterium radiobacter</i> ^b	<i>Agrobacterium</i> sp.	Biocontrol of crown gall	Roberts and Tate, 1977; Wang et al., 1994
Polymyxins A, B, D, E (colistin), and M (mattacidin)	Non-ribosomal lipopeptide	<i>pmxA-E</i> operon	<i>Paenibacillus polymyxa</i> ^c	<i>Klebsiella pneumonia</i> <i>P. aeruginosa</i> <i>Acinetobacter</i> sp.	Limited clinical application due to toxicity	Storm et al., 1977; Gales et al., 2006; Li et al., 2006; Shaheen et al., 2011.
Trichodermin and Harzianum A	Terpenoids	tri cluster	<i>Trichoderma arundinaceum</i> [*] <i>T. brevicompactum</i> [*]	<i>Rhizoctonia solani</i> <i>Alternaria solani</i>	Crop pathogens	Chen et al., 2007; Cardoza et al., 2011; Tijerino et al., 2011
DIVERSIFICATION WITHIN CODING SEQUENCES (CDS)						
Chitinases (family 18)	Lytic enzymes	See text	<i>P. fluorescens</i> ^d <i>Actinoplanes missouriensis</i> <i>Streptomyces</i> sp. <i>Bacillus</i> sp. <i>Stenotrophomonas maltophilia</i>	<i>Colletotrichum falcatum</i> <i>Plectosporium tabacinum</i>	Suppression of crop diseases such as lupin root rot	Viswanathan and Samiyappan, 2001; Kobayashi et al., 2002; El-Tarabily, 2003; Quecine et al., 2008
Ergots	Alkaloid	See text	Epichloë <i>Neotyphodium</i> <i>Aspergillus</i> sp. <i>A. fumigatus</i> <i>Penicillium</i> sp. <i>Claviceps purpurea</i>	Nematodes, insects, and mammalian herbivores	Protect the host plant	de Groot et al., 1998; Gröger and Floss, 1998; Schardl, 2010
Fusariocidins	Non-ribosomal peptides	<i>fusGFEDCBA</i> operon	<i>P. polymyxa</i> ^c	<i>Aspergillus niger</i> <i>Aspergillus oryzae</i> <i>Fusarium oxysporum</i> <i>Penicillium thomii</i> <i>Leptosphaeria maculans</i>	Control black root rot in canola	Kajimura and Kaneda, 1996, 1997; Beatty and Jensen, 2002; Schwarzer et al., 2003
Iturins	Non-ribosomal cyclolipopeptides	<i>ituDABC</i> operon	<i>B. subtilis</i> ^e <i>B. amylolyticfaciens</i> ^a	<i>Rhizoctonia solani</i> <i>Fusarium oxysporum</i> <i>F. graminearum</i>	Targets Gibberella ear rot	Guildner et al., 1988; Constantinescu, 2001; Tsuge et al., 2001; Leclère et al., 2005; Dunlap et al., 2013; Hamdache et al., 2013

(Continued)

TABLE 2 | Continued

Compound	Chemical class	Underlying genes*	Example of producing microbe	Best studied pathogen target	Significance to plant /human	References
COMPLEX DYNAMIC INTERACTION						
Jadomycin	Angucycline polyketide	<i>jad</i> operon	<i>Streptomyces venezuelae</i>	Methicillin-resistant <i>S. aureus</i>	Clinical use	Doull et al., 1994; Jakeman et al., 2009
OTHERS						
Baclysin	Non-ribosomal peptide	<i>ywIB</i> - <i>G</i> operon and <i>ywH</i>	<i>B. subtilis</i> ^a <i>B. amyloliquefaciens</i> <i>B. pumilus</i>	<i>Candida albicans</i>	Clinical use	Kenig and Abraham, 1976; Leoffler et al., 1986; Phister et al., 2004
HCN	Volatile compound	<i>hcnABC</i> operon	<i>P. aeruginosa</i> ^a <i>P. fluorescens</i>	<i>Thielaviopsis basicola</i>	Black root rot of tobacco	Voisard et al., 1989, 1994; Frapolli et al., 2012
Phomone	Sesquiterpene	See text	<i>Xyloara</i> sp.	<i>Cladosporium cladosporioides</i>	Targets fungal wheat pathogen	Silva et al., 2010

*Indicates that the genes noted have been studied in this species.

^aIndicates that the species refers to the recent genome sequence data as indicated below

^a*B. amyloliquefaciens* (Kim et al., 2015).

^b*Agrobacterium radiobacter* (Zhang et al., 2014).

^c*P. polymixa* (Xie et al., 2014).

^d*P. fluorescens* (Martinez-Garcia et al., 2015), ^e*B. subtilis* (Barbe et al., 2009; Beldá et al., 2013).

transcribed divergently to the biosynthetic gene cluster (Nowak-Thompson et al., 1999). In earlier studies involving the biosynthetic operon of phenazine, its LysR transcriptional regulator gene (*phzR*) was also shown to be tightly linked to its corresponding biosynthesis gene cluster (Pierson et al., 1998). As both phenazine and PLT combat soil-borne fungal diseases in plants, we speculate that strong evolutionary pressures in the rhizosphere may have promoted HGT of the biosynthetic operons to new rhizosphere microbial hosts; the activator-cluster gene module would facilitate activation of the biosynthetic CDS following such gene transfer.

Jadomycin

Jadomycin is a member of angucycline antibiotics produced by *Streptomyces* species such as *S. venezuelae*. Jadomycin (*jad*) production is induced under stress conditions such as phage infection or heat shock (Doull et al., 1994; Jakeman et al., 2009). The *jad* biosynthetic gene cluster in *S. venezuelae* is closely related to type II polyketide synthase genes (Han et al., 1994) with a complex biosynthetic gene cluster (Zou et al., 2014). Jadomycin is of interest here because upstream of the *jad* operon are sets of negative regulatory genes including *jadR1R2R3* and *jadW123* (Yang et al., 1995; Zou et al., 2014). *jadW123* encodes enzymes for the biosynthesis of gamma-butyrolactones (GBL), whereas *JadR2* is a pseudoreceptor for GBL which upon its binding activates *JadR1* and *JadR3* that subsequently act as positive and negative transcriptional regulators of the *jad* biosynthetic operon, respectively (Zou et al., 2014). GBLs are becoming well known as regulators of secondary metabolism in gram positive bacteria, analogous to the related acyl homoserine lactone compounds which mediate QS in gram negative bacteria (Nodwell, 2014). QS is a method of communication between bacterial populations that activates genes based on high cell density through the signal molecule N-acyl-homoserine lactone (AHL) (Whitehead et al., 2001). Whereas, QS signaling molecules are thought to be synthesized and sensed by the same species (Nodwell, 2014), the GBL/*jad* system is interesting, because recent data suggests that GBL can signal across different *Streptomyces* species to activate different polyketide biosynthetic pathways (Nodwell, 2014; Zou et al., 2014). Biologically, it has been shown that different *Streptomyces* species, which are soil microbes, can live on the same grain of soil alongside a diversity of bacteria (Keller and Surette, 2006; Vetsigian et al., 2011), suggesting there may have been evolutionary selection for inter-species coordination for antibiotic production (Nodwell, 2014), resulting in enhanced genetic complexity associated with the *jad* locus.

Non-Ribosomal Peptides

The structures of non-ribosomal peptides described in this review are shown (Figure 2).

Zwittermicin A

Zwittermicin A is a polyketide/nonribosomal peptide hybrid antibiotic produced by *B. cereus* and *B. thuringiensis* (Raffel et al., 1996) with activity against oomycetes such as *Phytophthora medicaginis* and some other pathogenic fungi (Silo-Suh et al., 1998). Zwittermicin A has a unique structure that includes glycolyl

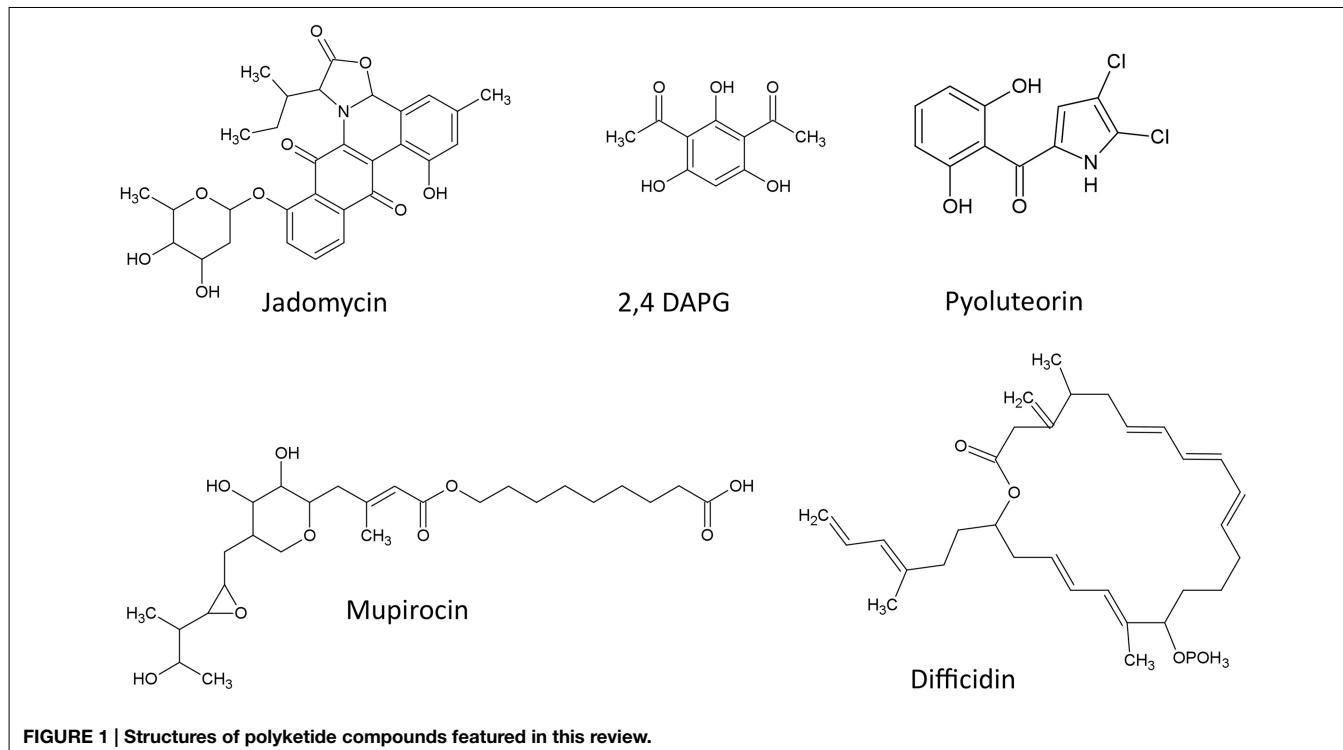


FIGURE 1 | Structures of polyketide compounds featured in this review.

moieties, D amino acid, and ethanolamine in addition to the unusual terminal amide produced from the ureidoalanine (non-proteinogenic amino acid) (Kevany et al., 2009). Zwittermicin A is thought to be biosynthesized as part of a larger metabolite that is processed twice to form zwittermicin A and two other metabolites (Kevany et al., 2009). The complete biosynthetic operon encoding zwittermicin A includes 27 open reading frames (CDS, *zmaA*, and *zmaV*) that extend over 62.5 kb of the *Bacillus cereus* UW85 genome, in addition to five individual genes (*kabR* and *kabA*—*kabD*) (Kevany et al., 2009). In this study, support was gained for the hypothesis that the skeleton of zwittermicin A is catalyzed by a megasynthase enzyme involving multiple nonribosomal peptide synthetases (NRPS) and PKS; the megasynthase has multiple modules containing distinct domains that catalyze the different steps in the pathway (Emmert et al., 2004; Kevany et al., 2009). Evidence suggested that the CDS included 5 NRPS modules (Kevany et al., 2009). It is noteworthy that a similar gene cluster was characterized on a plasmid in *B. cereus* AH1134, suggesting that the pathway can be transferred horizontally (Kevany et al., 2009). Consistent with the mobility of this operon, an orthologous 72-kb region encoding for zwittermicin A in *Bacillus thuringiensis*, was shown to be flanked by putative transposase genes on both edges, suggesting that it may be a mobile element that was gained by *B. cereus* through horizontal gene transfer. Since zwittermicin A has been reported to enhance the activity of protein toxins that attack insects (Broderick et al., 2000), it was hypothesized that transfer of this operon into *B. thuringiensis* permitted the microbe to gain insecticide-promoting factors to combat insects during co-evolution (Luo et al., 2011).

Fusaricidins A–D

Fusaricidins are guanidinylated β -hydroxy fatty acids attached to a cyclic hexapeptide including four D-amino acids (Kajimura and Kaneda, 1997; Schwarzer et al., 2003). These antibiotics are produced by *Paenibacillus polymyxa* strains and exhibit antifungal activity against diverse plant pathogens including, *Aspergillus niger*, *Aspergillus oryzae*, *Fusarium oxysporum*, and *Penicillium thomii* (Kajimura and Kaneda, 1996, 1997) as well as *Lepidosphaeria maculans*, the causal agent of black root rot in canola (Beatty and Jensen, 2002). The amino acid chains of fusaricidins are linked together and modified by a non-ribosomal peptide synthetase (NRPS). The multi-domain NRPS consists of up to 15,000 amino acids and is therefore considered among the longest proteins in nature (Schwarzer et al., 2003). NRPS incorporation is not limited to the 21 standard amino acids translated by the ribosome, and this promiscuity contributes to the great structural diversity and biological activity of non-ribosomal peptides (Li and Jensen, 2008).

In *P. polymyxa* E68, the fusaricidin biosynthetic gene cluster (*fusGFEDCBA*) has been characterized in which the NRPS coding sequence, the largest CDS in the cluster, was observed to encode a six-module peptide (Choi et al., 2008; Li and Jensen, 2008; Li et al., 2013). The biosynthetic cluster includes other CDS responsible for biosynthesis of the lipid moiety but does not contain transporter genes (Li and Jensen, 2008). In *P. polymyxa*, a promoter for the *fus* operon was identified and shown to be bound by a transcriptional repressor (AbrB) which previous studies implicated as a regulator of sporulation; this is of interest since fusaricidin was observed to be synthesized during sporulation,

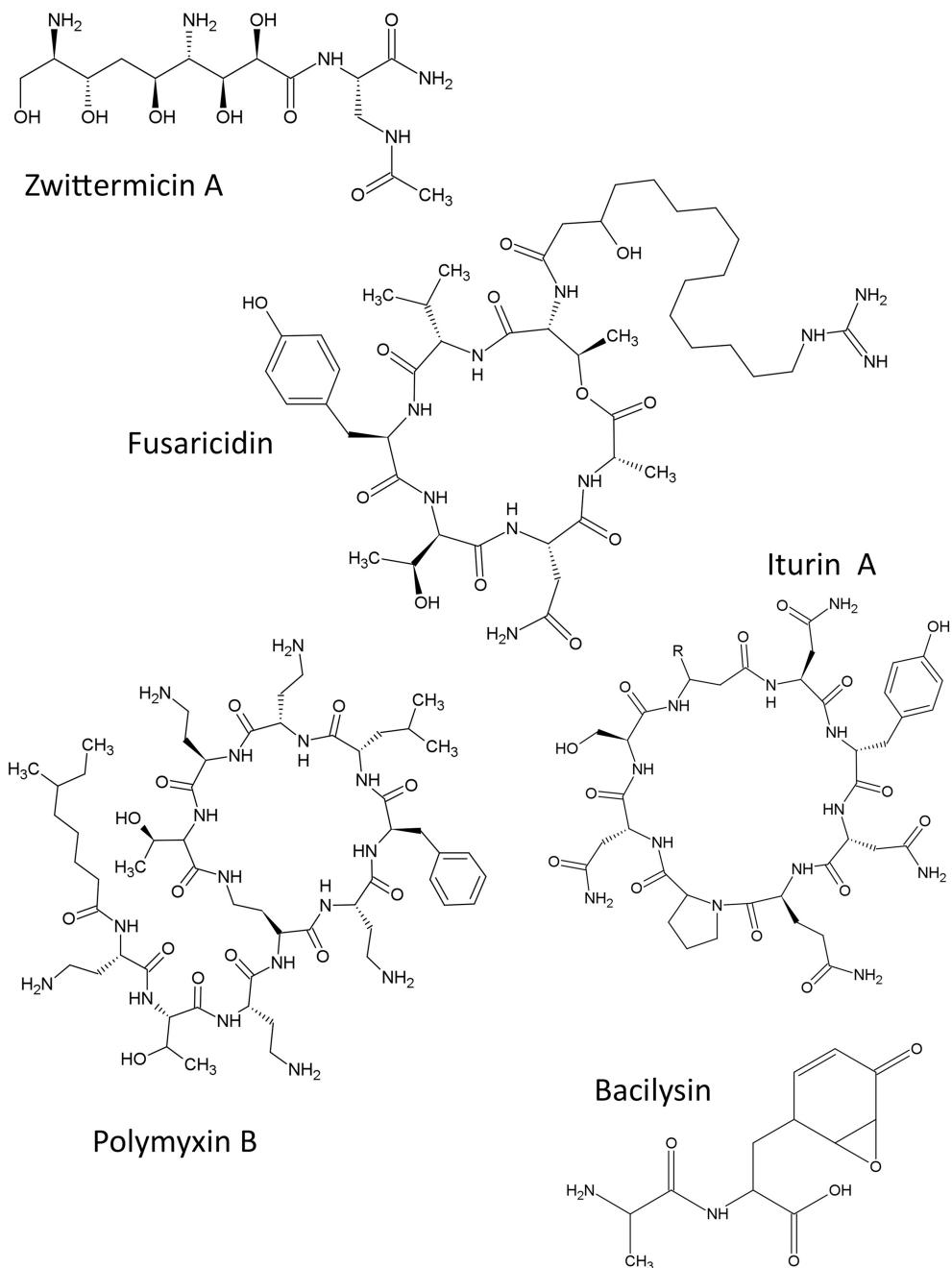


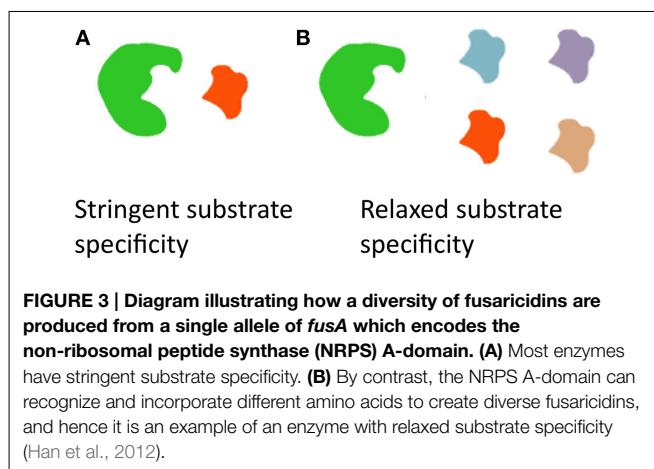
FIGURE 2 | Structures of non-ribosomal peptide compounds featured in this review.

thus coordinating the microbe's secondary metabolism with its life cycle (Li et al., 2013).

Allelic diversity is typically thought to be responsible for producing chemical diversity. However, an interesting feature of the *fus* cluster is that a diversity of fusaricidins, differing in their incorporated amino acids (Tyr, Val, Ile, allo-Ile, Phe), can be produced by a single allele of *fusA*; the underlying mechanism is that the NRPS A-domain, responsible for recognition of amino acids, has relaxed substrate specificity (Figure 3) (Han et al., 2012).

Polymyxins

Polymyxins are a family of non-ribosomal lipopeptide antibiotics composed of ten amino acids, a polycationic heptapeptide ring and a fatty acid derivative at the N terminus (Storm et al., 1977). They are produced by Gram positive bacteria and target Gram negative species, by altering the structure of the cell membrane. The polymyxin family includes polymyxins A, B, D, E (colistin), and M (mattacin) (Shaheen et al., 2011). Polymyxin B exhibits potent antibacterial activity against



Klebsiella pneumoniae, *Pseudomonas aeruginosa*, and *Acinetobacter* spp. (Gales et al., 2006). However, polymyxins exhibit a remarkable degree of neurotoxicity and nephrotoxicity which limit their clinical use (Li et al., 2006).

In *Paenibacillus polymyxa* PKB1 (the same strain that controls plant fungi by producing fusaricidins, see above), a 40.8 kb polymyxin biosynthetic gene cluster was shown to encode five coding sequences, *pmxA-E*. Three CDS (*pmxA, B, E*) encode subunits of NRPS, each responsible for the modular incorporation of amino acids, while two genes (*pmxC, D*) encode a permease belonging to the ABC-type transporter family (Shaheen et al., 2011). In both *P. polymyxa* PKB1 and *P. polymyxa* E681, the arrangement of the NRPS coding sequences in the *pmx* cluster does not match the amino acid sequence in the produced polymyxin, which is unusual for NRPS-encoded peptides (Choi et al., 2009; Shaheen et al., 2011).

With respect to the diversity of polymyxins, polymyxins differ in the amino acid composition of residues 3, 6, and 7, in the D vs. L stereochemistry of the incorporated amino acids, as well as in the lipid moiety (Choi et al., 2009; Shaheen et al., 2011). In *P. polymyxa* SC2 and *P. polymyxa* PKB1, an allelic variant was uncovered within NRPS domain 3 using bioinformatic analysis of the genome which correlated with incorporation of the D rather than L form of 2,4-diaminobutyrate in amino acid position 3, explaining the mechanism for the production of two subtypes of polymyxin B (Shaheen et al., 2011). With respect to the diversity of residues 6 and 7, *P. polymyxa* E681 and *P. polymyxa* PKB1 produce polymyxins that differ in these amino acids, producing polymyxin A and B, respectively (Shaheen et al., 2011). Bioinformatic analysis revealed that the DNA sequences of the *pmx* gene clusters were 92% conserved at the nucleotide level, but differed considerably in the domains corresponding to modules 6 and 7 (Shaheen et al., 2011). These two sets of observations led the authors to suggest that the diversity of polymyxins arises from mixing and matching of alleles of the NRPS modular domains, hence combinatorial chemistry, rather than relaxed substrate specificity as seen in other secondary metabolites such as fusaricidins (see above).

Another interesting feature of the *pmx* gene clusters is that the polymyxin transporters might also transport fusaricidin, since the *fus* biosynthetic cluster lacks any transporter genes (see above), and as both antibiotics are cationic lipopeptides (Shaheen et al., 2011). The authors found support for this hypothesis, as deletion mutations in *pmxC* and *D* genes also reduced the antifungal activity of fusaricidin against *Leptosphaeria maculans* although the two biosynthetic gene clusters are not linked. It is worth noting that there is no evidence yet of genes responsible for lipidation of the peptide residue in the characterized polymyxin clusters, suggesting that this function might be encoded elsewhere in the genome (Shaheen et al., 2011).

Iturins

Iturins are a family of non-ribosomal cyclolipopeptides consisting of seven α -amino acid residues and one β -amino acid, the latter noted as a unique feature compared to other lipopeptide antibiotics (Constantinescu, 2001; Leclère et al., 2005; Hamdache et al., 2013). The iturin family includes compounds such as bacillomycins D, F and L, bacillopeptins, iturins A, C, E and E, and mycosubtilins (Hamdache et al., 2013). Iturins are produced by different strains of *B. subtilis* and *B. amyloliquefaciens*, and exhibit potent antifungal activity against major phytopathogens including *R. solani*, *Fusarium oxysporum*, and *F. graminearum*, the latter responsible for Fusarium head blight in wheat (Gueldner et al., 1988; Constantinescu, 2001; Tsuge et al., 2001; Dunlap et al., 2013). The mechanism of action involves disruption of the target fungal plasma membrane (Thimon et al., 1995). In both *B. subtilis* RB14 and *B. amyloliquefaciens* AS43.3, the iturin A biosynthetic operons were shown to contain four coding sequences (*ituDABC*) coding for: a putative malonyl coenzyme A transacylase, a protein with three functions (fatty acid synthetase, amino acid transferase, and peptide synthetase), and two peptide synthetases, respectively (Tsuge et al., 2001; Dunlap et al., 2013).

Regarding diversification within the chemical family, iturin A from *B. subtilis* RB14 has a similar structure as mycosubtilin that is produced by *B. subtilis* ATCC 6633 but with inverted amino acids at the 6th and 7th positions (Tsuge et al., 2001). By comparative analysis of orthologous CDS between these two strains (*ituC* and *mycC*, respectively), it was suggested that the NRPS amino acid adenylation domain may have been intragenetically swapped during evolution, which would also imply a HGT event (Tsuge et al., 2001). Comparative genome analysis between at least three sequenced *itu* clusters may reveal further information concerning the diversification of the iturin family (Tsuge et al., 2001; Blom et al., 2012; Dunlap et al., 2013).

Bacilysin

Bacilysin is a non-ribosomally produced dipeptide composed of an L-alanine residue at the N terminus and a non-proteinogenic amino acid, L-anticapsin, at the C terminus (Walker and Abraham, 1970; Stein, 2005). Compared to the more elaborate non-ribosomal peptides noted above, bacilysin is noteworthy because it is amongst the simplest peptides in nature, adding to the structural diversity of observed non-ribosomal peptides. Bacilysin is produced by *Bacillus* species such as *B. pumilus*, *B. amyloliquefaciens*, and *B. subtilis* (Leoffler et al., 1986; Phister et al., 2004)

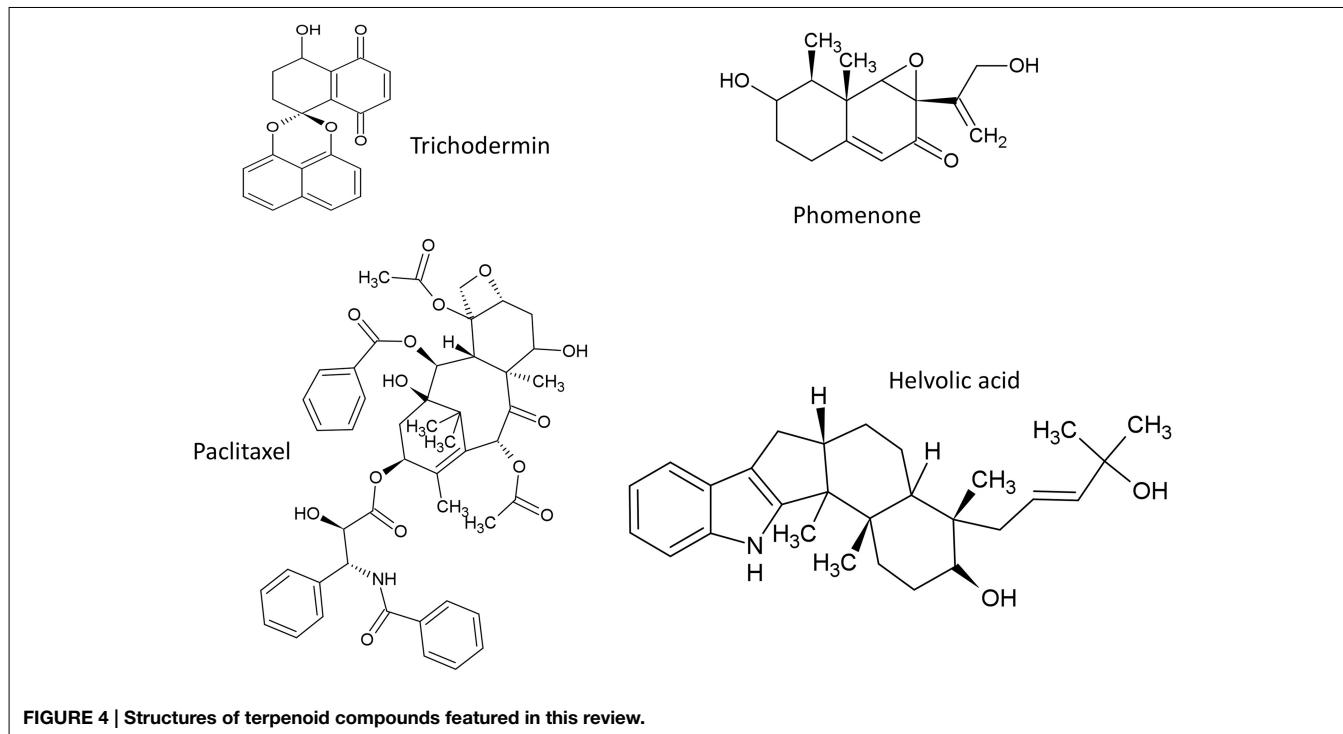


FIGURE 4 | Structures of terpenoid compounds featured in this review.

and shown to have antimicrobial activity against various bacteria and fungi such as *Candida albicans* (Kenig and Abraham, 1976). Mechanistically, bacilysin is a prodrug that is activated by the action of a peptidase enzyme that releases the active moiety, anticapsin (Rajavel et al., 2009). Anticapsin inhibits bacterial peptidoglycan or fungal protein biosynthesis through blockage of glucosamine synthetase, resulting in cell lysis (Kenig et al., 1976). Biosynthesis of bacilysin originates from the prephenate aromatic amino acid pathway (Hilton et al., 1988; Parker and Walsh, 2012).

In *B. subtilis* the biosynthesis of bacilysin is encoded by the operon, *bacABCDE* (*ywfB-G*), in addition to a monocistronic gene (*ywfH*) (Inaoka et al., 2003). *bacABC* is likely responsible for the biosynthesis of anticapsin while *bacDE* (*ywfEF*) encodes a ligase and an efflux transporter protein for self protection, respectively (Steinborn et al., 2005; Rajavel et al., 2009). The bacilysin biosynthetic operon is positively regulated by QS pheromones, in particular PhrC (Yazgan et al., 2001; Köroğlu et al., 2011) and negatively regulated by ScoC, a transition state regulator (Inaoka et al., 2009). The transition state in bacteria is a period of decision making.

Terpenoids

The structures of terpenoids described in this review are summarized (Figure 4).

Trichodermin and Harzianum A

Trichothecene mycotoxins are produced by some fungal genera such as deoxynivalenol (DON) from *Fusarium*, and harzianum and trichodermin from *Trichoderma arundinaceum* and *T. brevicompactum*, respectively (Cardoza et al., 2011). Trichodermin was reported to have antifungal activity against the fungal

pathogens *Rhizoctonia solani* and *Alternaria solani* (Chen et al., 2007) as well as other fungal genera (Tijerino et al., 2011). Trichodermin inhibits protein synthesis in eukaryotes by inhibiting peptidyl transferase that catalyzes translational elongation and/or termination (Wei et al., 1974) and by inhibiting peptide-bond formation at the initiation stage of translation (Carter et al., 1976).

Comparative analysis has been conducted on the CDS responsible for trichothecene biosynthesis in *Fusarium* and *Trichoderma*. In *Fusarium*, trichothecenes are encoded by a gene cluster called the TRI cluster; this cluster also encodes regulatory and transport proteins (Proctor et al., 2009). In *Trichoderma*, an orthologous TRI cluster was discovered in which 7 CDS were conserved with *Fusarium*, but the two clusters showed interesting evolutionary divergence (Cardoza et al., 2011) which may be informative for understanding the genetics underlying other anti-fungal metabolites. In *Fusarium*, the TRI cluster includes *tri5* that encodes trichodiene synthase, the first committed step in trichothecene biosynthesis, which catalyzes the cyclization of farnesyl pyrophosphate to form trichodiene (Hohn and Beremand, 1989). In *Fusarium*, *tri5* is located within the TRI cluster, but surprisingly it is not associated with the orthologous cluster in *Trichoderma*. Three additional CDS responsible for trichothecene biosynthesis in *Fusarium* (*tri7*, *tri8*, *tri13*) are missing from the *Trichoderma* cluster, along with an CDS of unknown function (*tri9*) (Cardoza et al., 2011). Interestingly, two of the apparently conserved biosynthetic CDS (*tri4* and *tri11*, based on sequence homology) were demonstrated to have diverged functionally between *Trichoderma* and *Fusarium* based on heterologous expression analysis: in *Trichoderma*, *tri4* catalyzes three out of four oxygenation reactions carried out

by its corresponding *Fusarium* ortholog; *tri11* catalyzes distinctive hydroxylation reactions in *Fusarium* (C-15) and *Trichoderma* (C-4). Finally, amongst the CDS which are conserved between *Fusarium* and *Trichoderma*, head-to-tail vs. head-to-head rearrangements are observed (e.g., *tri3*, *tri4*) (Cardoza et al., 2011). These results demonstrate multiple evolutionary events (rearrangement, functional diversification, gene loss, gene gain) within one biosynthetic gene cluster (Figure 5).

Phomenone

Phomenone is a sesquiterpene synthesized by various fungi including *Xylaria* sp., an endophytic fungus isolated from *Piper aduncum*, and reported to have antifungal activity against the pathogen *Cladosporium cladosporioides* (Silva et al., 2010). Phomenone is structurally similar to the PR toxin metabolite of *Penicillium roqueforti* which functions by inhibiting RNA polymerase and thus inhibits protein synthesis at the initiation and elongation steps (Moule et al., 1976). A biosynthetic precursor for phomenone A is aristolochene (Proctor and Hohn, 1993). In *P. roqueforti* NRRL 849, a gene required for aristolochene (*aril*) biosynthesis was characterized and shown to encode a sesquiterpene cyclase named aristolochene synthase (AS) (Proctor and Hohn, 1993). Expression of *aril* occurs in stationary phase cultures and is regulated transcriptionally (Proctor and Hohn, 1993).

Paclitaxel (Taxol)

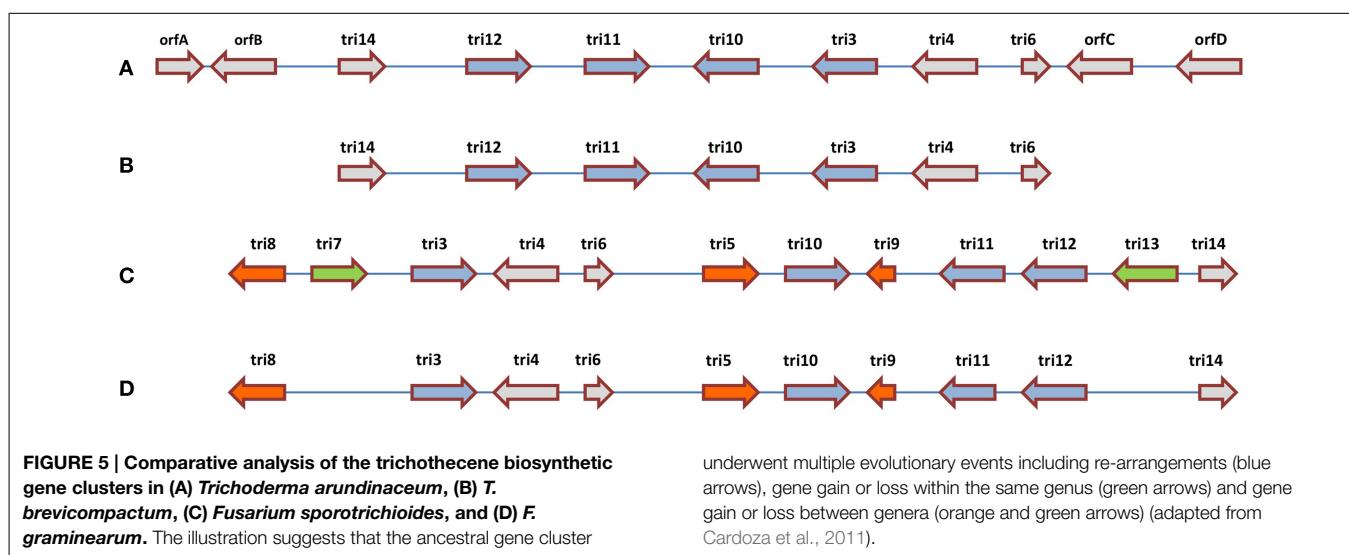
The diterpene paclitaxel (Taxol) is reported to be produced by at least 20 diverse fungal endophyte genera inhabiting various plant species (Zhou et al., 2010). Taxol was reported to be produced by some fungal endophytes that inhabit conifer wood and its ecological function was suggested to be a fungicide against host pathogens (Soliman et al., 2013). Taxol acts by stabilizing microtubules and inhibiting spindle function leading to disruptions in normal cell division (Horwitz, 1994). However, Taxol was originally purified from *Taxus* trees (Wani et al., 1971) and shown to be encoded by plant nuclear genes, apparently redundantly. As

the number of plant genera that produce Taxol is very few, it is interesting to speculate whether its biosynthetic genes may have been transferred horizontally from fungi to plants.

The Taxol biosynthetic pathway in plants requires 19 enzymatic steps. The first committed step in biosynthesis of plant Taxol is cyclization of GGDP to tax- $(4,5),(11,12)$ -diene catalyzed by taxadiene synthase (TS) (Hezari et al., 1995). Thirteen plant Taxol biosynthetic genes from *Taxus* were used in BLASTP searches to identify potential homologs in *Penicillium aurantiogriseum* NRRL 62431 (Yang et al., 2014). Seven putative homologous genes were identified though the homology scores were as low as 19%; these genes were claimed to encode: phenylalanine aminomutase (PAM), geranylgeranyl diphosphate synthase (GGPPS), taxane 5 α -hydroxylase (T5OH), taxane 13 α -hydroxylase (T13OH), taxane 7 β -hydroxylase (T7OH), taxane2 α -hydroxylase (T2OH) and taxane 10 β -hydroxylase (T10OH). Another gene encoding an AT (PAU_P11263) was identified by using BLASTP against the GenBank database. However, no homologs were identified to plant TS; the authors claimed that the fungus might catalyze taxadiene synthesis by a unique enzymatic system (Yang et al., 2014). Position-Specific Iterative BLAST showed one gene from the bacterial genus *Mycobacterium* with potential similarity to plant TS suggesting lateral gene transfer from plants to mycobacteria (Yang et al., 2014).

In a parallel study to isolate fungal Taxol biosynthetic genes, a different approach was taken where PCR primers designed from the plant genes that encode Taxol were used as a primary screen against fungi (Xiong et al., 2013). The study identified putative homologs of fungal TS as well as BAPT (which encodes the critical C-13 phenylpropanoid side-chain CoA acyltransferase) with \sim 40% sequence identities to their plant counterparts. Despite this progress, other reports remain skeptical that fungi actually encode Taxol (Heinig et al., 2013).

Recent studies have demonstrated complex three-way interactions in Taxol biosynthesis between a Taxol-producing fungal endophyte, other endophytes and the host plant. Host endophytic



fungi appear to elicit plant TS transcription or transcript accumulation. Specifically, TS transcript and the corresponding protein were reduced upon treating both young plantlets and old *Taxus* wood with fungicide (Soliman et al., 2013). In a parallel study, co-culture of the Taxol-producing endophyte *Paraconiothyrium SSM001* with two presumptive fungal endophytes of the same yew tree host elicited paclitaxel accumulation from the endophyte, suggesting inter-species interactions between endophytes inhabiting the same host niche (Soliman and Raizada, 2013).

Helvolic Acid

Helvolic acid is a fusidane triterpene produced by *Aspergillus fumigatus* (Lodeiro et al., 2009) and the yeast, *Pichia guilliermondii* Ppf9 (Zhao et al., 2010). Helvolic acid was reported to inhibit the spore germination of *Magnaporthe oryzae*, the causal agent of rice blast disease (Zhao et al., 2010). The biosynthetic genes for helvolic acid are clustered as nine genes coding for protostadienol synthase which catalyzes the precursor (17Z)-protosta-17(20),24-dien-3-ol, along with genes that encode squalene-hopene cyclase, four cytochrome P450 monooxygenases, short chain dehydrogenase, two transferases and 3-ketosteroid 1-dehydrogenase (Lodeiro et al., 2009). The authors reported that the P450 monooxygenases from different fungi shared substantial sequence identity across recent evolution, while the transferases duplicated and diversified into paralogous gene families (Lodeiro et al., 2009). This observation suggests that even within a single gene cluster, there may be different selection pressures on adjacent genes belonging to the same biosynthetic pathway. Interestingly, the helvolic acid biosynthetic gene cluster in *A. fumigatus* is located in the subtelomere chromosome region (Lodeiro et al., 2009) which is associated with high rates of evolutionary recombination and diversification. However, the gene cluster lacks introns which is a trait sometimes associated with subtelomeric regions, but this observation might also be evidence of HGT from bacteria (Lodeiro et al., 2009).

Alkaloids

The structures of alkaloids described in this review are summarized (Figure 6).

Ergot

Ergot alkaloids are produced from the sexual *Epichloe* fungi and their asexual derivatives *Neotyphodium* within the Clavicipitaceae family which inhabit Pooideae grasses (Schardl, 2010). Ergot alkaloids can interact with receptors of the central nervous system and exhibit toxic effect on nematodes, insects, and mammalian herbivores including livestock which graze these grasses (de Groot et al., 1998; Gröger and Floss, 1998). In Europe in the Middle Ages, consumption of ergot-infected grain or grasses caused convulsions, paranoia and hallucinations in livestock and humans, known as St. Anthony's Fire (Dotz, 1980). The diverse ergot alkaloids share a tetracyclic ergoline backbone derived from tryptophan and dimethylallyl diphosphate (Flieger et al., 1997). Gene clusters for ergot alkaloid biosynthesis have been identified in various Ascomycete species belonging to *Aspergillus*, *Penicillium*, and *Claviceps*. Seven genes encode the ergoline

scaffold including dimethylallyltryptophan synthase (DMATS) which catalyzes the first committed step. DMATS is responsible for the prenylation of L-tryptophan with dimethylallylpyrophosphate (DMAPP) to produce 4-dimethylallyltryptophan (4-DMAT) (Heinstein et al., 1971). Ergots have diversified into three classes, caused by diverse substituents attached to the carboxyl group of the tetracyclic ergoline backbone, in particular the presence of an amide group (creating ergoamides), a peptide-like amide moiety (creating ergopeptines) or the absence of these moieties (creating clavine alkaloids) (Wallwey and Li, 2011). These structural modifications are responsible for the differential physiological and pharmacological effects of the ergot family, that include treatment of postpartem hemorrhage, leukemia, and Parkinson's disease. The genetic basis for ergot diversification into these 3 major classes is associated with the presence or absence of nonribosomal peptide synthases (NRPS) which catalyze the biosynthesis of the peptide moieties on the ergoline backbone (Wallwey and Li, 2011). For example, four NRPS genes are present in *Claviceps purpurea* (which encodes ergopeptines) but absent in *Aspergillus fumigatus* (which produces clavine alkaloids). Inactivation of these genes suggests that two of the NRPS genes (*lpsA* and *lpsB*) are also responsible for synthesis of the ergoamides (Haarmann et al., 2008). Interestingly, further diversification of the peptide moiety within *C. purpurea* has been reported to be caused by fine-scale allelic diversification of the NRPS genes (Haarmann et al., 2005). There is additional evidence to suggest that diversification of the ergot alkaloid gene clusters is associated with DNA transposons and retroelements, which were observed in the cluster encoding ergovaline, an ergot alkaloid from *Epichloe festucae* associated with livestock toxicity (Fleetwood et al., 2007). As an interesting note, the genes encoding ergovaline were highly expressed only during biotrophic growth of the fungus within the host grass plant not when the mycelia were cultured *in vitro*, suggesting that the host might have a regulatory role in the expression of the fungal gene cluster (Fleetwood et al., 2007).

Loline Alkaloid

Loline is an indole alkaloid produced by *Neotyphodium uncinatum* fungus, the asexual mutualistic derivative of *Epichloe*, which is known to protect its host plants from insects (Blankenship et al., 2001; Schardl, 2010). The loline biosynthetic pathway was suggested to involve proline and homoserine (Spiering et al., 2005). In *N. uncinatum*, two homologous gene clusters encoding loline were identified, named LOL-1 and LOL-2 (Spiering et al., 2005). The cluster LOL-1 involves nine genes-(*lolF-1*, *lolC-1*, *lolD-1*, *lolO-1*, *lolA-1*, *lolU-1*, *lolP-1*, *lolT-1*, *lolE-1*) within a 25-kb chromosomal segment, while the LOL-2 cluster contains the same homologs (except for *lolF*) ordered and oriented the same as in LOL-1. This evidence suggests that the loline clusters may represent a recent segmental duplication event (Spiering et al., 2005).

An interesting ecological situation exists in grasses infected with *Epichloe* fungi (sexual form of *Neotyphodium*): the fungus reduces the ability of these plants to propagate sexually (they choke the inflorescences), which, without compensatory mechanisms, would prevent vertical transmission of the fungus (Zhang

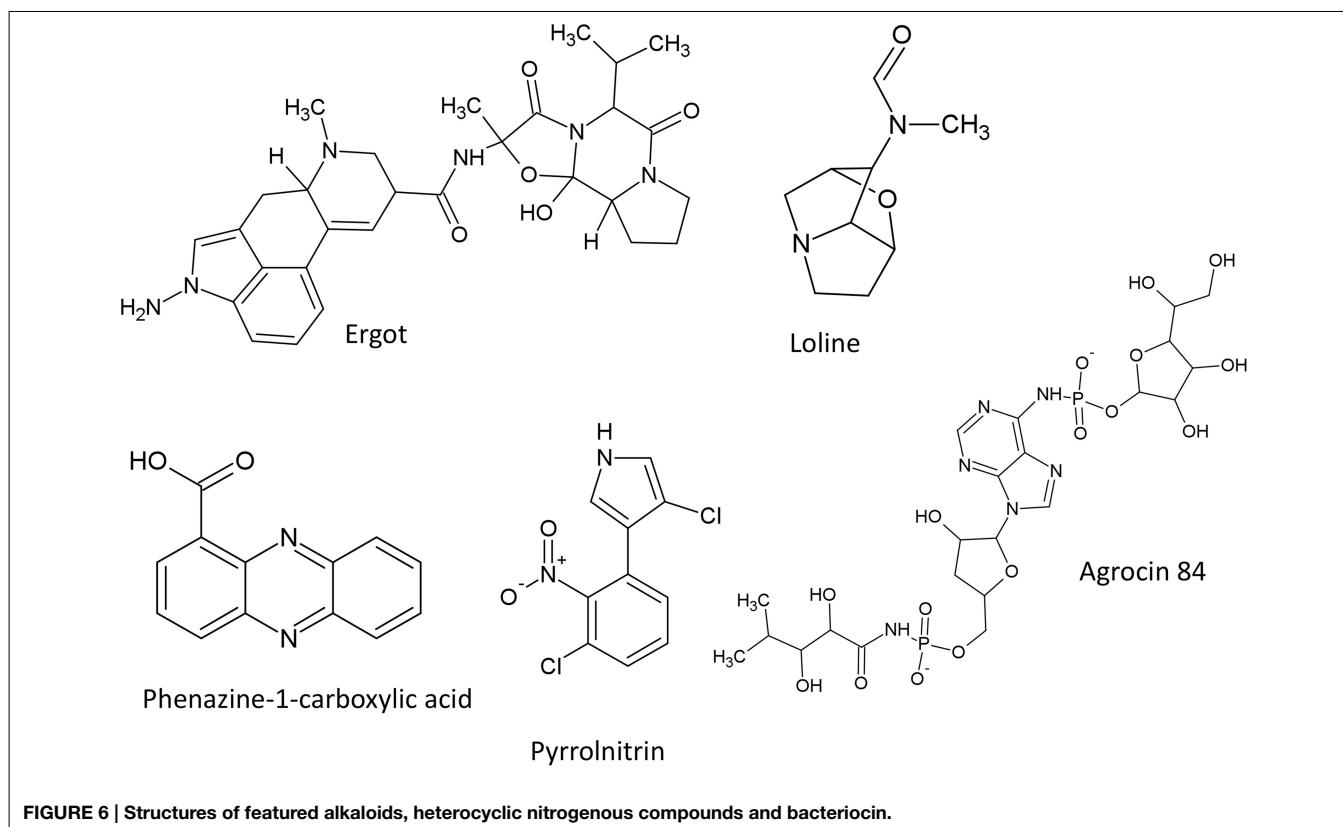


FIGURE 6 | Structures of featured alkaloids, heterocyclic nitrogenous compounds and bacteriocin.

et al., 2010). However, to compensate, the fungal stromata attract fly vectors which transmit the fungal spores to other plants, permitting horizontal transfer of the fungus. Loline accumulates in young tissues of the grasses, providing insect protection to these young hosts; however if loline was also to accumulate in the grass inflorescences, it would kill the fly vector of the fungus. Upon further investigation, this apparent paradox was resolved: in these grass inflorescences, transcription of the loline biosynthesis genes was dramatically downregulated compared to plants with healthy inflorescences (infected with the symbiotic asexual *Neotyphodium*), permitting the fly vectors to survive (Zhang et al., 2010). These observations suggest strong selection pressure to evolve the regulatory elements of these genes.

Heterocyclic Nitrogenous Compounds

The structures of heterocyclic nitrogenous compounds described in this review are summarized (Figure 6).

Phenazines

Phenazines are a group of naturally occurring heterocyclic nitrogenous antibiotics produced exclusively by bacteria and widely reported in fluorescent *Pseudomonas* (Mavrodi et al., 2006, 2013). Phenazines are potent antifungal compounds that can combat soil borne pathogens (Ligon et al., 2000) such as *Rhizoctonia solani*, *Gaeumannomyces graminis* var. *tritici*, *Pythium* spp. (Gurusiddaiah et al., 1986) and *Fusarium oxysporum* (Anjiah et al., 1998). Mechanisms of action include: (1) accumulation of toxic molecules such as hydrogen peroxide and superoxide due

to the redox potential of phenazine (Hassan and Fridovich, 1980; Hassett et al., 1995); and (2) elicitation of induced host resistance (Audenaert et al., 2002). Ecologically, the evidence suggests that the plant rhizosphere promotes phenazine-producing bacteria to combat pathogens (Mazzola et al., 1992; Mavrodi et al., 2013).

Phenazine is derived from the shikimic acid pathway, with amino-2-deoxyisochorismic acid (ADIC) as the branchpoint to phenazine (McDonald et al., 2001). ADIC is then converted to trans-2, 3-dihydro-3-hydroxy anthranilic acid which undergoes dimerization to form phenazine-1-carboxylic acid, the first derivative of the phenazines (McDonald et al., 2001). Phenazine biosynthesis in *Pseudomonas fluorescens* is encoded by a single or duplicated core of five CDS, *phzADEFG*, that encode ketosteroid isomerase, isochorismatase, anthranilate synthase, trans-2,3-dihydro-3-hydroxyanthranilate isomerase, and pyridoxamine oxidase respectively (Mavrodi et al., 2013). In *Pseudomonas*, the core may include other CDS such as *phzB* which was duplicated from *phzA*, and *phzC* which encodes 3-deoxy-D-arabino-heptulose-7-phosphate synthase that is responsible for diverting carbon from the shikimate pathway to phenazine (Pierson and Pierson, 2010).

Comparisons between *Pseudomonas* species and other genera have revealed conservation yet diversity of the core phenazine biosynthetic CDS. For example, the phenazine biosynthesis operon in *Burkholderia cepacia* maintains the five core enzymes observed in *Pseudomonas* as reviewed (Mavrodi et al., 2006). However, there is evidence to suggest that these coding sequences spread to enteric bacteria and *Burkholderia* species via horizontal

gene transfer, because these genes can be observed in plasmids and transposons (Mavrodi et al., 2010). For example, in *Erwinia herbicola*, a biosynthetic cluster of 16 CDS (*ehp*) was isolated from a plasmid, of which 15 coded for D-alanyl griseoluteic acid (AGA) while *ehpR* was observed to encode for resistance to AGA (Giddens et al., 2002). Other differences in the core have also been observed between *Pseudomonas* species and others; for example in both *Burkholderia cepacia* and *Erwinia herbicola*, *phzA* is not duplicated (Mavrodi et al., 2006).

Structural diversity of phenazines in different species is achieved by specific genes that may be located within the cluster or elsewhere in the genome. For example, in *P. chlororaphi*, the *phzH* gene is located downstream of the phenazine operon, where it encodes an aminotransferase responsible for converting phenazine-1-carboxylic acid (PCA) to phenazine-1-carboxamide, the characteristic green pigment of *P. chlororaphi* (Chin-A-Woeng et al., 1998). In *P. aureofaciens*, *phzO* was identified as the gene that encodes an aromatic monooxygenase, responsible for catalyzing the hydroxylation of PCA to form the broad spectrum antifungal compound, 2-OH-PCA (Delaney et al., 2001). In *P. aeruginosa* (PAO1), two diversification genes were discovered: *phzM* was shown to be involved in the synthesis of pyocyanin while *phzS* gene encodes a monooxygenase that catalyzes the production of 1-hydroxy phenazine (Mavrodi et al., 2001).

Pyrrolnitrin

Pyrrolnitrin is a chlorinated phenylpyrrole antibiotic purified initially from *Burkholderia pyrrocinia* (Arima et al., 1964) then subsequently from other species including pseudomonads, *Myxococcus fulvus*, *Enterobacter agglomerans*, and *Serratia* sp (Chernin et al., 1996; Kirner et al., 1998; Hammer et al., 1999). Pyrrolnitrin was initially used for treatment of skin mycoses caused by *Trichophyton* fungus, then was developed as an effective fungicide for crops against *Botrytis cinerea* (Hammer et al., 1993), *Rhizoctonia solani* (El-Banna and Winkelmann, 1998) and *Gaeumannomyces graminis* var. *tritici* (Tazawa et al., 2000). In *P. fluorescens*, the pyrrolnitrin biosynthetic operon consists of four coding sequences (*prnABCD*) coding for tryptophan halogenase (*prnA*), a decarboxylase (*prnB*), monodechloroaminopyrrolnitrin halogenase (*prnC*), and an oxidase (*prnD*) (Hammer et al., 1997; Kirner et al., 1998). Comparative analysis indicates that the pyrrolnitrin biosynthetic operon is differentially conserved between divergent species with 59% similarity among diverse bacterial strains such as *Pseudomonas*, *Myxococcus fulvus*, and *Burkholderia cepacia*, with lower similarity shown for *prnA* in *M. fulvus* (45%) (Hammer et al., 1999). Furthermore, RFLP-based polymorphisms within a 786 bp *prnD* fragment suggested that there may have been lateral gene transfer of the *prn* operon from *Pseudomonas* to *Burkholderia pyrrocinia* (Souza and Raaijmakers, 2003). Consistent with such mobility, transposase-encoding genes surrounding the *prn* biosynthetic operon were observed in *Burkholderia pseudomallei* (Costa et al., 2009).

Volatile Compounds

In this section, only the most well studied volatile compound, hydrogen cyanide, is discussed.

Hydrogen Cyanide (HCN)

Hydrogen cyanide (HCN) is a volatile secondary metabolite produced by *P. aeruginosa*, and diverse rhizosphere fluorescent pseudomonads, where they exhibit biocontrol activity against pathogenic fungi such as *Thielaviopsis basicola*, the fungal causal agent of black root rot of tobacco (Voisard et al., 1989, 1994; Frapolli et al., 2012). Mechanistically, HCN functions by inhibiting important metalloenzymes such as cytochrome *c* oxidase (Blumer and Haas, 2000) and/or by complexing metals in the soil (Brandl et al., 2008). HCN is biosynthesized from glycine (Castric, 1977) in an oxidative reaction catalyzed by HCN synthase, a membrane-bound flavoenzyme (Castric, 1994; Blumer and Haas, 2000). The biosynthesis of HCN occurs in the presence of an electron acceptor such as phenazine methosulfate (Wissing, 1974).

In *P. aeruginosa* PAO1, the HCN synthase biosynthetic operon *hcnABC* was characterized (Pessi and Haas, 2000). *hcnA* was reported to encode a protein similar to formate dehydrogenase while *hcnB* and *hcnC* encode products with similarity to amino acid oxidases (Laville et al., 1998; Svercel et al., 2007). In a phylogenetic analysis of 30 fluorescent pseudomonads, no evidence was found for HGT of the *hcn* gene cluster, but rather that the locus appears to be exclusively inherited vertically (Frapolli et al., 2012). HCN has also been detected in *Chromobacterium violaceum* but the underlying genes have not been reported which might otherwise give new insights into HCN biosynthesis outside of the pseudomonads (Blom et al., 2011).

Bacteriocin

In this section, only the most well studied compound from this class is discussed. The structure of agrocin 84 described in this review is included (Figure 6).

Agrocin 84

Agrocin 84 is a 6-N-phosphoramidate of an adenine nucleotide analog (Roberts and Tate, 1977). This compound is produced by non-pathogenic strains of *Agrobacterium radiobacter* to biocontrol crown gall, a tumorous disease resulting from overproduction of auxin and cytokinin hormones stimulated by the Ti plasmid after it has transferred from *A. radiobacter* and integrated within host plant chromosomal DNA (Wang et al., 1994). Recently, it was shown that agrocin 84 employs a novel mechanism to inhibit leucyl-tRNA synthetases and hence inhibit translation (Chopra et al., 2013), though it was earlier suggested that agrocin 84 acts by inhibiting DNA synthesis (Das et al., 1978).

In *Agrobacterium radiobacter* K84, the biosynthesis and immunity to agrocin 84 is encoded by 17 coding sequences (the *agn* operon) located on a 44-kb conjugal plasmid, *pAgK84*, though the plasmid has 36 CDS in total (Kim et al., 2006). The two most interesting CDS are *agnB2* and *agnA* which encode aminoacyl tRNA synthetase homologs. The agrocin 84 antibiotic is essentially a nucleotide attached to an amino acid-like moiety (methyl pentanamide), and its mode of action was proposed to involve competitive binding to the active site of leucyl-tRNA synthetases (Reader et al., 2005). *agnB2* encodes a leucyl-tRNA synthetase homolog that confers self-immunity to agrocin 84 since it does not bind the antibiotic (Kim et al., 2006). Normally, a

tRNA synthetase acts as a ligase that catalyzes the attachment of an amino acid to a tRNA which includes an anticodon; the catalysis results in a phosphoanhydride bond between the amino acid and ATP as the initial step in the aminoacetylation of tRNA (Ibba and Söll, 2000). Surprisingly, *agnA* encodes a truncated homolog of an asparaginyl-tRNA synthetase which lacks the anticodon-binding domain, but maintains the catalytic domain. Thus, *agnA* appears to be a fascinating example of a gene that evolved from an ancient tRNA synthetase (for arginine), but now is a biosynthetic enzyme for an antibiotic that inhibits a paralogous enzyme (for leucine attachment) (Kim et al., 2006).

The *agn* operon may have an evolutionary history of horizontal gene transfer, as *pAgK84* is inter and intra species transferable: *Rhizobium* that received the *pAgK84* plasmid from *Agrobacterium* as trans-conjugates could synthesize agrocin 84 and received immunity as well (Farrand et al., 1985).

A final fascinating feature of the *agn* system is an apparent second form of ancient evolutionary pressure on the genes responsible for the biosynthesis of the antibiotic. Agrocin 84 is a chemical mimic of agrocinopines, a class of compounds that is a source of plant-derived nitrogen for the pathogens targeted by the antibiotic; the pathogens have their own Ti plasmids that encode for transporters that not only transport agrocinopines but also agrocin 84 (Ellis and Murphy, 1981; Hayman and Farrand, 1988; Kim and Farrand, 1997). Hence the *agn* biosynthetic genes evolved to create a chemical structure that not only mimics the tRNA synthetase substrate of the pathogen target, but also targets its nitrogen uptake machinery.

Enzymes

In this section, only the most well studied anti-fungal enzyme, chitinase, is discussed.

Chitinase

Chitinases are enzymes that break down chitin, one of the fungal cell wall components composed of repeated units of N-acetyl-D-glucos-2-amine, linked by β -1,4 glycosidic bonds (Bhattacharya et al., 2007). Fungi and hence chitin are enriched in soil and thus soil microbes are abundant sources of chitinases (also to target insects) (Hjort et al., 2010). Examples of chitinase-producing microbes include: fluorescent *Pseudomonas* strains isolated from the sugarcane rhizosphere that can target *Colletotrichum falcatum*, the causative agent of red rot disease in this crop (Viswanathan and Samiyappan, 2001); *Actinoplanes*

missouriensis that antagonizes *Plectosporium tabacinum*, the causal agent of lupin root rot in Egypt (El-Tarabily, 2003); and *Stenotrophomonas maltophilia* that suppresses summer patch disease in Kentucky bluegrass (Kobayashi et al., 2002). Chitinases are produced by diverse bacterial genera including *Pseudomonas*, *Streptomyces*, *Bacillus*, and *Burkholderia* (Quecine et al., 2008). Chitinases are divided into two major categories, exochitinases and endochitinases. Of the four reported endochitinase family members (glycoside hydrolase families 18, 19, 23, and 48), primarily families 18 and 19 have been reported in bacteria, with only a single example of a family 23 chitinase (Prakash et al., 2010).

In *Stenotrophomonas maltophilia* 34S1, the chitinase family 18 gene has one CDS that encodes for a protein with seven domains: a catalytic domain, a chitin binding domain, three putative binding domains, a fibronectin type III domain and a polycystic kidney disease domain (Kobayashi et al., 2002). Bacterial chitinase family 18 has been shown to display different types of diversity. First, sequence analysis has shown that the catalytic domain and substrate binding domain, which are separated by a linker, have evolved independently. As the domain sequences do not match the taxonomies of their hosts, it has been suggested that domain swapping has been an important generator of diversity in this family, combined with HGT (Figure 7) (Karlsson and Stenlid, 2008). Unusual examples of chitinase genes are those that contain multiple family 18 catalytic domains within the same peptide that appear to function independently of one another (Howard et al., 2004). Additional examples of family 18 biodiversity include genes that contain non-consensus sequences at the catalytic site, as well as a bacterial subgroup that consists solely of a catalytic domain (Karlsson and Stenlid, 2008).

Unlike family 18 chitinases that are widely distributed among the prokaryotes, family 19 chitinases are restricted to green non-sulfur and purple bacteria, as well as actinobacteria (Prakash et al., 2010). Based on sequence alignments of family 19 chitinases in prokaryotes and eukaryotes, strong evidence has emerged that this gene family in actinobacteria and purple bacteria was derived from flowering plants by HGT (Prakash et al., 2010). Furthermore, HGT from plants to purple bacteria may have occurred as two independent events in the distant past, followed more recently by HGT to actinobacteria (Prakash et al., 2010). The core architecture and catalytic sites of bacterial and plant family 19 chitinases are nearly identical. The sequence analysis further suggests that there was subsequent HGT from purple bacteria

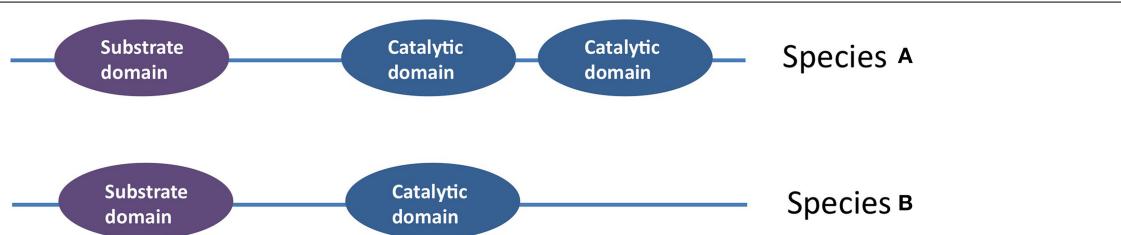


FIGURE 7 | An example of intra-coding sequences diversification within an anti-microbial gene cluster: amongst the Family 18 chitinases is an example of a chitinase in which the catalytic domain has been duplicated (Howard et al., 2004).

and actinobacteria to nematodes and arthropods, respectively (Prakash et al., 2010).

Discussion

The objective of this paper was to review the biodiversity of antimicrobial compounds, their mode(s) of action and underlying biosynthetic genes within plant associated microbes. This review covered diverse biosynthetic gene clusters that encode polyketides, non-ribosomal peptides, terpenoids, alkaloids, heterocyclic nitrogenous compounds, volatile compounds, bacteriocins, and lytic enzymes. The reviewed evidence suggests that these biosynthetic genes have diversified at different orders, each based on distinct evolutionary mechanisms:

Species Level Diversification

An emerging theme from the literature is that horizontal gene transfer (HGT) appears to have played a major role in the evolutionary diversification of plant-associated microbial species through inheritance of anti-microbial traits. There is evidence that HGT may have occurred from: bacteria to bacteria such as those that inhabit the rhizosphere (e.g., phenazine); from bacteria to fungi (e.g., helvolic acid); from bacteria to nematodes and arthropods (e.g., chitinase family 19); possibly from fungi to plants (e.g., Taxol); from plants to bacteria (e.g., phenazine and chitinase family 19); and even from higher eukaryotes to bacteria (e.g., IleS, pseudomonic acid resistance protein) (Figure 8). As noted in the literature, diverse factors might have facilitated these remarkable gene transfer events including: (1) the clustering of genes encoding the secondary metabolite; (2) homologous recombination between chromosomes and transconjugated plasmids (e.g., phenazine, zwittermicin A); (3) the presence of mobile elements (DNA transposons and retroelements) flanking the biosynthetic operons (e.g., zwittermicin A, phenazine and ergovaline); and (4) the presence of genes that

encode self-immunity to the antibiotic within the biosynthetic cluster as otherwise receiving the cluster would have caused immediate suicide (e.g., mupirocin). It is worth noting that some gene clusters show no evidence of HGT (e.g., hydrogen cyanide).

Genome Level Diversification

A second emerging theme from the literature is that a subset of associated plant- associated microbial genomes have diversified with respect to duplications of entire gene clusters responsible for the synthesis of antimicrobial compounds. For example, as noted above, in *Neotyphodium uncinatum*, there are two homologous gene clusters that encode loline, LOL-1 and LOL-2, the likely result of a segmental duplication event within this fungus. Another noted example is from *Bacillus* species in which three tandemly duplicated gene clusters, *pks1*, *pks2*, and *pks3*, encode the polyketides, bacillaene, macrolactin, and difficidin, respectively, the likely result of a homologous recombination event.

Intra Gene Cluster Diversification

A third interesting theme from the literature is that gene clusters encoding anti-microbial compounds have extensively diversified within, to permit biochemical diversification. The biosynthetic genes for these compounds are clustered in fungi or organized into operons in bacteria—in the latter, they are generally located on chromosomes but occasionally on plasmids (e.g., agrocin 84). Diversity within each cluster can include varying combinations of biosynthetic coding sequences (CDS), transporters for the respective compound, regulatory genes and CDS that confer self-immunity (e.g., mupirocin, agrocin 84). The biosynthetic operons vary in how many CDS synthesize the core skeleton (e.g., synthetases) as well as in how many encode decoration enzymes (e.g., hydroxylases, acyltransferases). However, the decoration enzymes may be encoded outside the gene cluster (e.g., phenazine operon). Furthermore, the biosynthetic CDS may be organized into genetic modules (e.g., NRPS) that vary in number. Each gene cluster is also associated with distinct DNA regulatory elements, for example to receive signals such as from quorum sensing. For example, comparative analysis of the trichothecene biosynthetic gene clusters (TRI) in *Fusarium* and *Trichoderma* showed multiple evolutionary diversification events within a single biosynthetic gene cluster family (e.g., head-to-tail vs. head-to-head rearrangements) (Figure 5). In another example, comparative analyses of the polymyxin operon showed mixing and matching of CDS, resulting in diversification of the compounds. Similarly, diversification of the phenazines likely arose through a diversity of biosynthetic decoration enzymes (e.g., hydroxylases). Another intriguing observation is from the helvolic acid biosynthetic gene cluster, in which transferase CDS were shown to have duplicated and diversified into paralogous gene families. As noted above, an interesting feature of this gene cluster is that it located in the subtelomere chromosome region which is associated with high rates of evolutionary recombination.

Diversification Within Coding Sequences (CDS)

A final emerging theme from the literature is that diversification of anti-microbial traits in plant-associated microbes

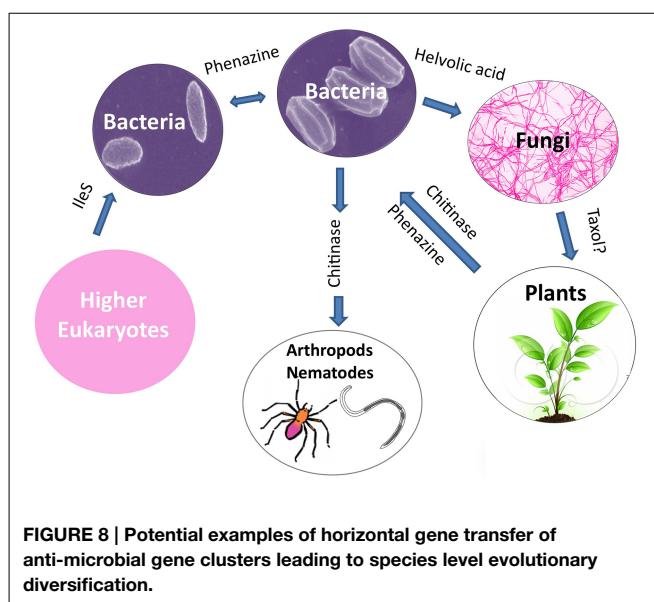


FIGURE 8 | Potential examples of horizontal gene transfer of anti-microbial gene clusters leading to species level evolutionary diversification.

arose from allelic diversification. For example, intragenic swapping of domains was observed within the same genetic module (e.g., iturin A, mycosubtilins). As another example, whereas most chitinase genes possess a single catalytic domain, examples were noted where a single CDS encodes two catalytic domains (**Figure 7**). In general, the literature notes that domains within the same CDS can evolve independently (e.g., catalytic vs. substrate binding domains of chitinase); combined with the existence of linker peptides between domains as sites of homologous recombination, these features can result in novel alleles following domain swapping (e.g., family18 chitinases). Whereas, such allelic diversification plays a major role in the diversification of compound structures, caused for example by DNA mutations within the substrate binding domain, the literature demonstrates examples where biochemical diversity has arisen from relaxed substrate specificity of the biosynthetic enzymes (**Figure 3**). A representative example of the latter is the promiscuous fusaricidin NRPS in which the same recognition domain in different species can recognize and incorporate different amino acids, and furthermore it can recognize amino acids beyond the 21 standard amino acids translated by the ribosome, which results in significant structural diversity.

Dynamic Evolutionary Driven by Selection Pressures

It is interesting to speculate on the evolutionary selection pressures that have led to the diversification at the various biological levels noted above. At the most basic level, diversification was likely driven by a three-way co-evolution between the plant-associated microbe, its target pathogen and the host plant. This co-evolution may have occurred within a specific plant tissue niche or within soil associated with the rhizosphere (e.g., phenazine and PLT to combat soil-borne pathogens). However, there is also evidence for four-way interactions, to also include additional microbes (e.g., Taxol, jadomycin) and insects (e.g., ergovaline). These complex interactions can be bi-directional (e.g., loline). Within the producing organism, there is evidence for selection pressure to coordinate biosynthesis of the anti-microbial compound with the life cycle of the microbe (e.g., fusaricidins). There may also have been selection for genetic efficiency (e.g., potential sharing of transporter genes between polymyxin and fusaricidin). These selection pressures have led to fascinating individual stories, including the evolution of mimicry to facilitate antibiosis (e.g., agrocin).

Ecological and Evolutionary Lessons

When the examples of anti-microbial pathways were grouped by the phylogeny of their host microbes, several trends were observed (**Figure 9**, **Tables 1, 2**). Specifically: (1) some anti-microbial genes are apparently widely distributed among diverse taxonomic classes of bacteria (e.g., chitinases); (2) some metabolic pathways are widely distributed within one taxonomic class such as pyrrolnitrin that shows up in more than half of the presented species of Proteobacteria; (3) other anti-microbial pathways appear to be more restricted (e.g., fusaricidin, polymyxin, jadomycin). These results may correlate

to the evolutionary age of these genetic pathways, or may represent a bias based on how well the pathway has been studied. More widespread genome sequencing and/or the use of orthologous gene probes may help to inform the evolutionary origins of these anti-microbial pathways.

Bacterial Pathway Lessons

The selected examples of anti-microbial pathways from plant-associated bacteria found in the literature and presented in this review are distributed across Proteobacteria, Actinobacteria and Firmicutes (**Figure 9**). This may not be surprising as Proteobacteria and Actinobacteria are among the most widespread bacterial taxa associated with plants, perhaps because of their saprophytic capabilities (Bulgarelli et al., 2012).

Within these phyla, *P. fluorescens* (Proteobacteria) and *B. subtilis* (Firmicutes) were observed to produce a plethora of diverse antimicrobial compounds belonging to diverse chemical classes including polyketides, non-ribosomal peptides, heterocyclic nitrogenous compounds, volatiles and enzymes which reflect the diversity of the metabolic machineries of these species. As *P. fluorescens* and *B. subtilis* are both model systems, these results also support the above note of the bias within this literature.

Bacillus sp. and *Pseudomonas* sp. are ubiquitous microbes that can survive in diverse ecological niches (Compant et al., 2005). Both have elegant survival strategies that involve the production of antibiotics, surfactin, cyanide, biofilms, and induction of host resistance (Espinosa-Urgel, 2004; Dini-Andreote and van Elsas, 2013). These unique adaptations have led to their widespread study and use as biocontrol agents (Santoyo et al., 2012).

Genome analysis of *P. fluorescens* has provided insight into its ecological competency and evolutionary mechanisms. The versatile and rapid adaptability of *P. fluorescens* to diverse environmental clues may be attributed to over 200 characterized signal transduction proteins which enhance its sensing capability (Garbeva and de Boer, 2009; Humair et al., 2010). With respect to co-evolution, the *P. fluorescens* genome is exceptionally rich in repetitive extragenic palindromic (REP) elements, target sites for transposases and recombinases, with 1052 REP elements identified in *P. fluorescens* Pf-5 (compared to 21 in *P. aeruginosa* PAO1 and 365 in *P. syringae* DC 3000) (Tobes and Pareja, 2005, 2006). REPs likely affected genome evolution either by gene gain, loss or rearrangement (Silby et al., 2011). The latest version of the genome sequence and annotation of *P. fluorescens* was recently released (Martínez-García et al., 2015).

B. subtilis is naturally competent genetically, with a cascade of competence-specific DNA-uptake proteins that bind and transport DNA, in addition to a dynamic recombination mechanism which transforms chromosomal or plasmid DNA via different pathways (Chen and Dubnau, 2004; Kidane et al., 2009). Additionally, the *B. subtilis* genome encodes integrative and conjugative element binding (ICEBs1) proteins responsible for excision, integration, transfer of DNA (Lee et al., 2007) that likely have facilitated HGT. Comparative genomic analysis of *B. subtilis* strains revealed 298 accessory segments that potentially originated from mobile elements including plasmids, transposons and phages. This implies extensive HGT events that lead to diversification of the arsenal of anti-microbial pathways within *Bacillus*

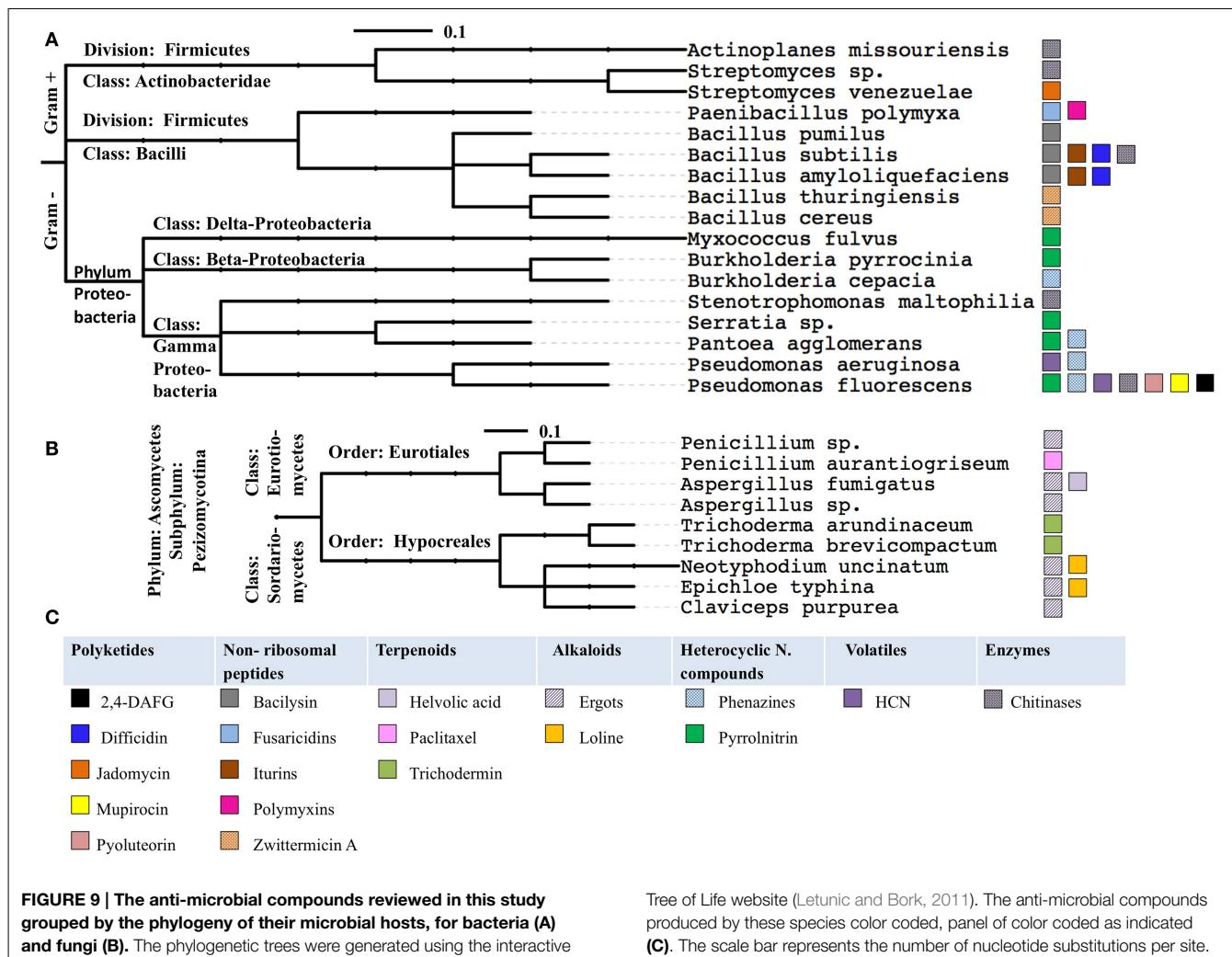


FIGURE 9 | The anti-microbial compounds reviewed in this study grouped by the phylogeny of their microbial hosts, for bacteria (A) and fungi (B). The phylogenetic trees were generated using the interactive

Tree of Life website (Letunic and Bork, 2011). The anti-microbial compounds produced by these species color coded, panel of color coded as indicated (C). The scale bar represents the number of nucleotide substitutions per site.

(Zeigler, 2011). The complete genome sequence and genome annotation of *B. subtilis* is available (Barbe et al., 2009; Belda et al., 2013).

Fungal Pathway Lessons

In contrast to bacteria, all the fungal examples presented in the review belong to a single classification—the Pezizomycotina (filamentous fungi), a subdivision of Ascomycota, the largest phylum of fungi (Blackwell, 2011) including representatives from Eurotiomycetes and Sordariomycetes (Figure 9). Pezizomycotina has an ancient origin in the Cambrian period, ca 530 Mya (Prieto and Wedin, 2013).

Pezizomycotina species are the most ubiquitous fungi with extremely diverse lifestyles, suggesting a corresponding diversity of ecological strategies (Spatafora et al., 2006; Beimforde et al., 2014) reflected in their production of a range of anti-microbial metabolites. There may be at least two reasons for this metabolic diversity, HGT and recombination. First, HGT from bacteria to fungi was previously reported in Ascomycota, of which 65% were observed in Pezizomycotina (Marcel-Toubro and Gabaldón, 2010). Second, secondary metabolism

gene clusters in Pezizomycotina show evidence of recent gene expansion (Arvas et al., 2007). Interestingly, most of these genes are located in the sub-telomere region (Rehmeyer et al., 2006) that is associated with a considerable high rate of recombination and correspondingly rapid evolution compared to other regions in the genome (Freitas-Junior et al., 2000), an example represented in this review by helvolic acid (Table 1).

Gaps and Future Perspectives

Despite the apparent progress in understanding the genetic mechanisms underlying the diversity of anti-microbial compounds produced by plant-associated microbes, significant gaps and opportunities remain. The major challenge is that a vast majority of plant associated microbes are unculturable, a phenomena that, to a far extent, limits our understanding of species diversification and evolution. It is worth noting that considerable progress toward cultivation of unculturable microbes has started to be achieved (Pham and Kim, 2012; Stewart, 2012). The modified cultivation methods attempt to simulate the natural environment, and include community culturing, and the

use of high-throughput microbioreactors and laser microdissection (Pham and Kim, 2012). Another challenge is that the literature appears to be biased for model organisms, with insufficient data from other organisms in the phylogenetic tree for comparative genomics and evolutionary studies. For example, despite our efforts to originally focus this review only on anti-microbial pathways from endophytes, it became clear that the number of associated genes from endophytes has largely been unexplored, compared to free living rhizosphere model species. Indeed, there remains a lack of detailed genetic analysis underlying many anti-microbial compounds across microbes (endophytic and non-endophytic) and a lack of information to connect allelic diversity with compound diversity.

With respect to understanding the biosynthetic pathways of these metabolites, more information is needed as to the extent that diverse anti-microbial pathways coordinate and share biosynthetic enzymes. An important question in metabolic biosynthesis is understanding how chemical substrates are channeled along metabolic pathways from one enzyme to the next; from this review, it appears that some anti-microbial pathways solve this problem by using mega-synthase enzymes (e.g., zwittermycin A), but for other pathways, investigation of enzyme-enzyme interactions will be informative. An interesting future area of study will be to investigate the subcellular location of biosynthetic and storage proteins, especially of self-toxic compounds that may need to be sequestered. To that end, there have been advances in studying compartmentalization and secondary metabolite trafficking machinery (Roze et al., 2011; Lim and Keller, 2014; Kistler and Broz, 2015), which offer strategies to move forward.

A significant challenge in this discipline is the study of anti-microbial compounds in their native ecological context, as most reports are based only on *in vitro* studies. In particular, because the target pathogen affects the host plant, more information is needed as to how the plant and the anti-pathogenic microbe coordinate and regulate one another. For example, in the jadomycin pathway, evidence suggests that the plant sensing of the pathogen stimulates the anti-pathogen pathway in the associated beneficial microbe. The potential complexity of plant-microbe interactions and associated signaling networks are well studied in model systems such as *Rhizobium* (Janczarek et al., 2015). Though *Rhizobium* is a symbiotic microbe of legume plants, these studies suggest that a wealth of information remains to be explored for other plant-associated microbes, in particular endophytes (Kusari et al., 2012).

Also within the ecological context, basic biochemical questions are raised such as whether the anti-microbial pathway is regulated by the target pathogen, for example feedback inhibition once the pathogen has been eliminated. To help understand the genetic regulation of these anti-microbial pathways, analysis of gene expression with respect to the microbial life cycle would be a

useful avenue of investigation, similar to the interesting findings from the fusaricidin pathway. An interesting study concerning aflatoxin, a polyketide mycotoxin, revealed strong evidence for the potential link between the fungal growth stage and polyketide biosynthesis (Zhou et al., 2000). Furthermore, intracellular tracking of aflatoxin biosynthetic enzymes in *Aspergillus parasiticus* showed significant accumulation in the vacuoles of specific cells but its absence in neighboring ones (Hong and Linz, 2008). This surprising result led Roze et al. (2011) to hypothesize the possibility of special and temporal gene expression of the associated biosynthetic pathway, at different developmental resolutions ranging from a single cell to fungal colony.

A related major challenge is that there are many natural products that exist in the literature that were initially isolated as part of screens for new compounds from total extracts, and hence the ecological functions of these compounds, as well as their underlying genes, remain unknown.

As anti-pathogenic metabolites may be self-toxic, the evolution of self-resistance is a particularly fascinating avenue of study, which this review demonstrates has been investigated for a limited number of pathways (e.g., mupirocin). Diverse self-resistance mechanisms have been reported in the microbial literature (Schäferle et al., 2011; Westman et al., 2013; Stegmann et al., 2015), suggesting that each plant-associated microbe with anti-microbial activity may employ unique self-protection strategies.

The recent advances in genome sequencing combined with gene editing tools will facilitate more in-depth analysis of orthologous biosynthetic genes in diverse species. Bioinformatic genome mining of biosynthetic gene clusters, combined with new advances in metabolomics, may also lead to the discovery of a diverse array of novel bio-active natural products. Moreover, merging these techniques with knowledge of microbial co-evolution and ecology (Vizcaino et al., 2014) along with advanced microscopy and imaging techniques will open a new era of discovery to harvest the diversity of natural products to combat evolving pathogens.

Author Contributions

WM wrote the manuscript, and WM and MR edited the manuscript.

Acknowledgments

We thank Travis Goron (University of Guelph) for helpful comments on the manuscript. WM was supported by a generous scholarship from the Government of Egypt. MR was supported by grants from the Grain Farmers of Ontario, OMAFRA, and the CIFSRF program by the International Development Research Centre and DFATD of the Government of Canada.

Fusarium spp. and *Pythium* spp. *Mol. Plant Microbe Interact.* 11, 847–854. doi: 10.1094/MPMI.1998.11.9.847

Arima, K., Imanaka, H., Kousaka, M., Fukuta, A., and Tamura, G. (1964). Pyrrolnitrin, a new antibiotic substance, produced by *Pseudomonas*. *Agric. Biol. Chem.* 28, 575–576. doi: 10.1271/bbb1961.28.575

References

Anjaiah, V., Koedam, N., Nowak-Thompson, B., Loper, J. E., Höfte, M., Tambong, J. T., et al. (1998). Involvement of phenazines and anthranilate in the antagonism of *Pseudomonas aeruginosa* PNA1 and Tn5 derivatives toward

Arvas, M., Kivioja, T., Mitchell, A., Saloheimo, M., Ussery, D., Penttila, M., et al. (2007). Comparison of protein coding gene contents of the fungal phyla Pezizomycotina and Saccharomycotina. *BMC Genomics* 8:325. doi: 10.1186/1471-2164-8-325

Audenaert, K., Pattery, T., Cornelis, P., and Hofte, M. (2002). Induction of systemic resistance to *Botrytis cinerea* in tomato by *Pseudomonas aeruginosa* 7NSK2: role of salicylic acid, pyochelin, and pyocyanin. *Mol. Plant Microbe Interact.* 15, 1147–1156. doi: 10.1094/MPMI.2002.15.11.1147

Bangera, M. G., and Thomashow, L. S. (1999). Identification and characterization of a gene cluster for synthesis of the polyketide antibiotic 2,4-diacetylphloroglucinol from *Pseudomonas fluorescens* Q2-87. *J. Bacteriol.* 181, 3155–3163.

Barbe, V., Cruveiller, S., Kunst, F., Lenoble, P., Meurice, G., Sekowska, A., et al. (2009). From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology* 155, 1758–1775. doi: 10.1099/mic.0.027839-0

Barea, J. M., Pozo, M. J., Azcon, R., and Azcon-Aguilar, C. (2005). Microbial cooperation in the rhizosphere. *J. Exp. Bot.* 56, 1761–1778. doi: 10.1093/jxb/eri197

Beatty, P. H., and Jensen, S. E. (2002). *Paenibacillus polymyxa* produces fusaricidin-type antifungal antibiotics active against *Leptosphaeria maculans*, the causative agent of blackleg disease of canola. *Can. J. Microbiol.* 48, 159–169. doi: 10.1139/w02-002

Beimforde, C., Feldberg, K., Nylander, S., Rikkinen, J., Tuovila, H., Dörfelt, H., et al. (2014). Estimating the phanerozoic history of the Ascomycota lineages: combining fossil and molecular data. *Mol. Phylogenet. Evol.* 78, 386–398. doi: 10.1016/j.ympev.2014.04.024

Belda, E., Sekowska, A., Le Feuvre, F., Morgat, A., Mornico, D., Ouzounis, C., et al. (2013). An updated metabolic view of the *Bacillus subtilis* 168 genome. *Microbiology* 159, 757–770. doi: 10.1099/mic.0.064691-0

Bhattacharya, D., Nagpure, A., and Gupta, R. K. (2007). Bacterial chitinases: properties and potential. *Crit. Rev. Biotechnol.* 27, 21–28. doi: 10.1080/07388550601168223

Blackwell, M. (2011). The fungi: 1, 2, 3 ... 5.1 million species? *Am. J. Bot.* 98, 426–438. doi: 10.3732/ajb.1000298

Blankenship, J. D., Spiering, M. J., Wilkinson, H. H., Fannin, F. F., Bush, L. P., and Schardl, C. L. (2001). Production of loline alkaloids by the grass endophyte, *Neotyphodium uncinatum*, in defined media. *Phytochemistry* 58, 395–401. doi: 10.1016/S0031-9422(01)00272-2

Blom, D., Fabbri, C., Eberl, L., and Weisskopf, L. (2011). Volatile-mediated killing of *Arabidopsis thaliana* by bacteria is mainly due to hydrogen cyanide. *Appl. Environ. Microbiol.* 77, 1000–1008. doi: 10.1128/AEM.01968-10

Blom, J., Rueckert, C., Niu, B., Wang, Q., and Borriss, R. (2012). The complete genome of *Bacillus amyloliquefaciens* subsp. *plantarum* CAU B946 contains a gene cluster for nonribosomal synthesis of iturin A. *J. Bacteriol.* 194, 1845–1846. doi: 10.1128/JB.06762-11

Blumer, C., and Haas, D. (2000). Mechanism, regulation, and ecological role of bacterial cyanide biosynthesis. *Arch. Microbiol.* 173, 170–177. doi: 10.1007/s002039900127

Brandl, H., Lehmann, S., Faramarzi, M. A., and Martinelli, D. (2008). Biomobilization of silver, gold, and platinum from solid waste materials by HCN-forming microorganisms. *Hydrometallurgy* 94, 14–17. doi: 10.1016/j.hydromet.2008.05.016

Broderick, N. A., Goodman, R. M., Raffa, K. F., and Handelsman, J. (2000). Synergy between zwittemycin A and *Bacillus thuringiensis* subsp. kurstaki against gypsy moth (Lepidoptera: Lymantriidae). *Environ. Entomol.* 29, 101–107. doi: 10.1603/0046-225X-29.1.101

Bulgarelli, D., Rott, M., Schlaeppi, K., van Themaat, E., Ahmadinejad, N., Assenza, F., et al. (2012). Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature* 488, 91–95. doi: 10.1038/nature11336

Cane, D. E. (1997). A special thematic issue on polyketide and nonribosomal polypeptide biosynthesis. *Chem. Rev.* 97, 2463–2706. doi: 10.1021/cr970097g

Cardoza, R. E., Malmierca, M. G., Hermosa, M. R., Alexander, N. J., McCormick, S. P., Proctor, R. H., et al. (2011). Identification of loci and functional characterization of trichothecene biosynthesis genes in filamentous fungi of the genus *Trichoderma*. *Appl. Environ. Microbiol.* 77, 4867–4877. doi: 10.1128/AEM.00595-11

Carter, C. J., Cannon, M., and Smith, K. E. (1976). Inhibition of protein synthesis in *Reticulocyte* lysates by trichodermin. *Biochem. J.* 154, 171–178.

Castric, P. (1994). Influence of oxygen on the *Pseudomonas aeruginosa* hydrogen cyanide synthase. *Curr. Microbiol.* 29, 19–21. doi: 10.1007/BF01570186

Castric, P. A. (1977). Glycine metabolism by *Pseudomonas aeruginosa*: hydrogen cyanide biosynthesis. *J. Bacteriol.* 130, 826–831.

Chen, I., and Dubnau, D. (2004). DNA uptake during bacterial transformation. *Nat. Rev. Microbiol.* 2, 241–249. doi: 10.1038/nrmicro844

Chen, L., Chen, J., Zheng, X., Zhang, J., and Yu, X. (2007). Identification and anti-fungal activity of the metabolite of endophytic fungi isolated from *Llex cornuta*. *Chin. J. Pestic. Sci.* 9, 143–150.

Chen, X.-H., Scholz, R., Borriss, M., Junge, H., Mögel, G., Kunz, S., et al. (2009). Difficidin and bacilysin produced by plant-associated *Bacillus amyloliquefaciens* are efficient in controlling fire blight disease. *J. Biotechnol.* 140, 38–44. doi: 10.1016/j.jbiotec.2008.10.015

Chen, X.-H., Vater, J., Piel, J., Franke, P., Scholz, R., Schneider, K., et al. (2006). Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. *J. Bacteriol.* 188, 4024–4036. doi: 10.1128/JB.00052-06

Chernin, L., Brandis, A., Ismailov, Z., and Chet, I. (1996). Pyrrolnitrin production by an *Enterobacter agglomerans* strain with a broad spectrum of antagonistic activity towards fungal and bacterial phytopathogens. *Curr. Microbiol.* 32, 208–212. doi: 10.1007/s002849900037

Chin-A-Woeng, T. F., Bloemberg, G. V., van der Bij, A. J., van der Drift, K. M. G. M., Schripsema, J., Kroon, B., et al. (1998). Biocontrol by phenazine-1-carboxamide-producing *Pseudomonas chlororaphis* PCL1391 of tomato root rot caused by *Fusarium oxysporum* f. sp. *Radicis-lycopersici*. *Mol. Plant Microbe Interact.* 11, 1069–1077. doi: 10.1094/MPMI.1998.11.11.1069

Choi, S.-K., Park, S.-Y., Kim, R., Lee, C.-H., Kim, J. F., and Park, S. H. (2008). Identification and functional analysis of the fusaricidin biosynthetic gene of *Paenibacillus polymyxa* E681. *Biochem. Biophys. Res. Commun.* 365, 89–95. doi: 10.1016/j.bbrc.2007.10.147

Choi, S.-K., Park, S.-Y., Kim, R., Kim, S.-B., Lee, C.-H., and Kim, J. F. (2009). Identification of a polymyxin synthetase gene cluster of *Paenibacillus polymyxa* and heterologous expression of the gene in *Bacillus subtilis*. *J. Bacteriol.* 191, 3350–3358. doi: 10.1128/JB.01728-08

Chopra, S., Palencia, A., Virus, C., Tripathy, A., Temple, B. R., Velazquez-Campoy, A., et al. (2013). Plant tumour biocontrol agent employs a tRNA-dependent mechanism to inhibit leucyl-tRNA synthetase. *Nat. Commun.* 4, 1417. doi: 10.1038/ncomms2421

Compart, S., Duffy, B., Nowak, J., Clément, C., and Barka, E. A. (2005). Use of plant growth-promoting bacteria for biocontrol of plant diseases: principles, mechanisms of action, and future prospects. *Appl. Environ. Microbiol.* 71, 4951–4959. doi: 10.1128/AEM.71.9.4951-4959.2005

Constantinescu, F. (2001). Extraction and identification of antifungal metabolites produced by some *B. subtilis* strains. *Analele Institutului de Cercetari* 31, 17–23.

Costa, R., van Aarle, I. M., Mendes, R., and Elsaes, J. D. (2009). Genomics of pyrrolnitrin biosynthetic loci: evidence for conservation and whole-operon mobility within Gram-negative bacteria. *Environ. Microbiol.* 11, 159–175. doi: 10.1111/j.1462-2920.2008.01750.x

Das, P. K., Basu, M., and Chatterjee, G. C. (1978). Studies on the mode of action of agrocin 84. *J. Antibiot.* 31, 490–492. doi: 10.7164/antibiotics.31.490

Deane, C. D., and Mitchell, D. A. (2014). Lessons learned from the transformation of natural product discovery to a genome-driven endeavor. *J. Ind. Microbiol. Biotechnol.* 41, 315–331. doi: 10.1007/s10295-013-1361-8

de Groot, A. N., van Dongen, P. W., Vree, T. B., Hekster, Y. A., and van Roosmalen, J. (1998). Ergot Alkaloids. *Drugs* 56, 523–535. doi: 10.2165/00003495-199856040-00002

Delaney, S. M., Mavrodi, D. V., Bonsall, R. F., and Thomashow, L. S. (2001). phzO, a gene for biosynthesis of 2-hydroxylate phenazine compounds in *Pseudomonas aeruginosa*. *J. Bacteriol.* 183, 5376–5384. doi: 10.1128/JB.183.1.318-327.2001

Delaney, I., Sheehan, M. M., Fenton, A., Bardin, S., Aarons, S., and O'Gara, F. (2000). Regulation of production of the antifungal metabolite 2,4-diacetylphloroglucinol in *Pseudomonas fluorescens* F113: genetic analysis of phfA as a transcriptional repressor. *Microbiology* 146, 537–546.

Dini-Andreote, F., and van Elsaes, J. D. (2013). Back to the basics: the need for eco-physiological insights to enhance our understanding of microbial behaviour in the rhizosphere. *Plant Soil* 373, 1–15. doi: 10.1007/s11104-013-1687-z

Dotz, W. (1980). St. Anthony's fire. *Am. J. Dermatopathol.* 2, 249–254. doi: 10.1097/00000372-198000230-00013

Doull, J. L., Singh, A. K., Hoare, M., and Ayer, S. W. (1994). Conditions for the production of jadomycin B by *Streptomyces venezuelae* ISP5230: effects of heat shock, ethanol treatment and phage infection. *J. Ind. Microbiol.* 13, 120–125. doi: 10.1007/BF01584109

Dunlap, C. A., Bowman, M. J., and Schisler, D. A. (2013). Genomic analysis and secondary metabolite production in *Bacillus amyloliquefaciens* AS 43.3: a biocontrol antagonist of Fusarium head blight. *Biol. Control* 64, 166–175. doi: 10.1016/j.bioccontrol.2012.11.002

Eilenberg, J. (2006). “Concepts and visions of biological control,” in *An Ecological and Societal Approach to Biological Control*, eds J. Eilenberg and H. M. T. Hokkanen (Dordrecht: Springer), 1–11.

El-Banna, N., and Winkelmann, G. (1998). Pyrrolnitrin from *Burkholderia cepacia*: antibiotic activity against fungi and novel activities against streptomycetes. *J. Appl. Microbiol.* 85, 69–78. doi: 10.1046/j.1365-2672.1998.00473.x

Ellis, J., and Murphy, P. (1981). Four new opines from crown gall tumours – their detection and properties. *Mol. Gen. Genet.* 181, 36–43. doi: 10.1007/BF00339002

El-Sayed, A. K., Hothersall, J., Cooper, S. M., Stephens, E., Simpson, T. J., and Thomas, C. M. (2003). Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem. Biol.* 10, 419–430. doi: 10.1016/S1074-5521(03)00091-7

El-Sayed, A. K., Hothersall, J., and Thomas, C. M. (2001). Quorum-sensing-dependent regulation of biosynthesis of the polyketide antibiotic mupirocin in *Pseudomonas fluorescens* NCIMB 10586. *Microbiology* 147, 2127–2139.

El-Tarably, K. A. (2003). An endophytic chitinase-producing isolate with potential for biological control of root rot of lupin caused by *Plectosporium tabacinum*. *Aust. J. Bot.* 51, 257–266. doi: 10.1071/BT02107

Eltringham, I. (1997). Mupirocin resistance and methicillin-resistant *Staphylococcus aureus* (MRSA). *J. Hosp. Inf.* 35, 1–8. doi: 10.1016/S0195-6701(97)90162-6

Emmert, E. A., Klimowicz, A. K., Thomas, M. G., and Handelsman, J. (2004). Genetics of zwittermicin A production by *Bacillus cereus*. *Appl. Environ. Microbiol.* 70, 104–113. doi: 10.1128/AEM.70.1.104-113.2004

Espinosa-Urgel, M. (2004). Plant-associated *Pseudomonas* populations: molecular biology, DNA dynamics, and gene transfer. *Plasmid* 52, 139–150. doi: 10.1016/j.plasmid.2004.06.004

Farrand, S. K., Slota, J. E., Shim, J. S., and Kerr, A. (1985). Tn5 insertions in the agrocin 84 plasmid: the conjugal nature of pAgK84 and the locations of determinants for transfer and agrocin 84 production. *Plasmid* 13, 106–117. doi: 10.1016/0147-619X(85)90063-0

Fleetwood, D. J., Scott, B., Lane, G. A., Tanaka, A., and Johnson, R. D. (2007). A complex ergovaline gene cluster in *Epichloë* endophytes of grasses. *Appl. Environmen. Microbiol.* 73, 2571–2579. doi: 10.1128/AEM.00257-07

Flieger, M., Wurst, M., and Shelby, R. (1997). Ergot alkaloids-sources, structures and analytical methods. *Folia Microbiol.* 42, 3–30. doi: 10.1007/BF02898641

Frapolli, M., Pothier, J. F., Défago, G., and Moënne, Y. (2012). Evolutionary history of synthesis pathway genes for phloroglucinol and cyanide antimicrobials in plant-associated fluorescent pseudomonads. *Mol. Phylogen. Evol.* 63, 877–890. doi: 10.1016/j.ympev.2012.02.030

Freitas-Junior, L. H., Bottius, E., Pirrit, L. A., Deitsch, K. W., Scheidig, C., Guinet, F., et al. (2000). Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* 407, 1018–1022. doi: 10.1038/35039531

Fuller, A., Mellows, G., Woolford, M., Banks, G., Barrow, K., and Chain, E. (1971). Pseudomonic acid: an antibiotic produced by *Pseudomonas fluorescens*. *Nature* 234, 416–417. doi: 10.1038/234416a0

Gales, A., Jones, R., and Sader, H. (2006). Global assessment of the antimicrobial activity of polymyxin B against 54 731 clinical isolates of Gram negative bacilli: report from the SENTRY antimicrobial surveillance programme (2001–2004). *Clin. Microbiol. Infect.* 12, 315–321. doi: 10.1111/j.1469-0951.2005.01351.x

Garbeva, P., and de Boer, W. (2009). Inter-specific interactions between carbon-limited soil bacteria affect behavior and gene expression. *Microb. Ecol.* 58, 36–46. doi: 10.1007/s00248-009-9502-3

Ge, Y.-H., Pei, D.-L., Zhao, Y.-H., Li, W.-W., Wang, S.-F., and Xu, Y.-Q. (2007). Correlation between antifungal agent phenazine-1-carboxylic acid and pyoluteorin biosynthesis in *Pseudomonas* sp. m18. *Curr. Microbiol.* 54, 277–281. doi: 10.1007/s00284-006-0317-x

Giddens, S. R., Feng, Y., and Mahanty, H. K. (2002). Characterization of a novel phenazine antibiotic gene cluster in *Erwinia herbicola* Eh1087. *Mol. Microbiol.* 45, 769–783. doi: 10.1046/j.1365-2958.2002.03048.x

Gröger, D., and Floss, H. G. (1998). “Biochemistry of ergot alkaloids – achievements and challenges,” in *The Alkaloids: Chemistry and Biology*, ed G. A. Cordell (New York, NY: Academic Press), 171–218.

Gueldner, R. C., Reilly, C. C., Pusey, P. L., Costello, C. E., Arrendale, R. F., Cox, R. H., et al. (1988). Isolation and identification of iturins as antifungal peptides in biological control of peach brown rot with *Bacillus subtilis*. *J. Agric. Food Chem.* 36, 366–370. doi: 10.1021/jf00080a031

Gurney, R., and Thomas, C. M. (2011). Mupirocin: biosynthesis, special features and applications of an antibiotic from a Gram-negative bacterium. *Appl. Microbiol. Biotechnol.* 90, 11–21. doi: 10.1007/s00253-011-3128-3

Gurusiddaiah, S., Weller, D., Sarkar, A., and Cook, R. (1986). Characterization of an antibiotic produced by a strain of *Pseudomonas fluorescens* inhibitory to *Gaeumannomyces graminis* var. *tritici* and *Pythium* spp. *Antimicrob. Agent Chemother.* 29, 488–495. doi: 10.1128/AAC.29.3.488

Haarmann, T., Lorenz, N., and Tudynski, P. (2008). Use of a nonhomologous end joining deficient strain (Deltaku70) of the ergot fungus *Claviceps purpurea* for identification of a nonribosomal peptide synthetase gene involved in ergotamine biosynthesis. *Fungal Genet. Biol.* 45, 35–44. doi: 10.1016/j.fgb.2007.04.008

Haarmann, T., Machado, C., Lübbe, Y., Correia, T., Schardl, C. L., Panaccione, D. G., et al. (2005). The ergot alkaloid gene cluster in *Claviceps purpurea*: extension of the cluster sequence and intra species evolution. *Phytochemistry* 66, 1312–1320. doi: 10.1016/j.phytochem.2005.04.011

Hamdache, A., Azarken, R., Lamarti, A., Aleu, J., and Collado, I. (2013). Comparative genome analysis of *Bacillus* spp. and its relationship with bioactive nonribosomal peptide production. *Phytochem. Rev.* 12, 685–716. doi: 10.1007/s11101-013-9278-4

Hammer, P. E., Burd, W., Hill, D. S., Ligon, J. M., and van Pee, K.-H. (1999). Conservation of the pyrrolnitrin biosynthetic gene cluster among six pyrrolnitrin-producing strains. *FEMS Microbiol. Lett.* 180, 39–44. doi: 10.1111/j.1574-6968.1999.tb08775.x

Hammer, P. E., Hill, D. S., Lam, S. T., van Pee, K. H., and Ligon, J. M. (1997). Four genes from *Pseudomonas fluorescens* that encode the biosynthesis of pyrrolnitrin. *Appl. Environ. Microbiol.* 63, 2147–2154.

Hammer, P., Evensen, K., and Janisiewicz, W. (1993). Postharvest control of *Botrytis cinerea* on cut rose flowers with pyrrolnitrin. *Plant Dis.* 77, 283–286. doi: 10.1094/PD-77-0283

Han, J. W., Kim, E. Y., Lee, J. M., Kim, Y. S., Bang, E., and Kim, B. S. (2012). Site-directed modification of the adenylation domain of the fusaricidin non-ribosomal peptide synthetase for enhanced production of fusaricidin analogs. *Biotechnol. Lett.* 34, 1327–1334. doi: 10.1007/s10529-012-0913-8

Han, L., Yang, K., Ramalingam, E., Mosher, R. H., and Vining, L. C. (1994). Cloning and characterization of polyketide synthase genes for jadomycin B biosynthesis in *Streptomyces venezuelae* ISP5230. *Microbiology* 140, 3379–3389. doi: 10.1099/13500872-140-12-3379

Hassan, H. M., and Fridovich, I. (1980). Mechanism of the antibiotic action pyocyanine. *J. Bacteriol.* 141, 156–163.

Hassett, D. J., Schweizer, H. P., and Ohman, D. E. (1995). *Pseudomonas aeruginosa* sodA and sodB mutants defective in manganese-and iron-cofactored superoxide dismutase activity demonstrate the importance of the iron-cofactored form in aerobic metabolism. *J. Bacteriol.* 177, 6330–6337.

Hayman, G., and Farrand, S. (1988). Characterization and mapping of the agrocinopine-agrocin 84 locus on the nopaline Ti plasmid pTiC58. *J. Bacteriol.* 170, 1759–1767.

Heinig, U., Scholz, S., and Jennewein, S. (2013). Getting to the bottom of taxol biosynthesis by fungi. *Fungal Divers.* 60, 161–170. doi: 10.1007/s13225-013-0228-7

Heinstein, P. F., Lee, S. I., and Floss, H. G. (1971). Isolation of dimethylallylpyrophosphate: tryptophan dimethylallyl transferase from the ergot fungus (*Claviceps spec.*). *Biochem. Biophys. Res. Commun.* 44, 1244–1251. doi: 10.1016/S0006-291X(71)80219-X

Hezari, M., Lewis, N. G., and Croteau, R. (1995). Purification and characterization of tax-4(5),11(12)-diene synthase from pacific yew (*Taxus brevifolia*) that catalyzes the first committed step of taxol biosynthesis. *Arch. Biochem. Biophys.* 322, 437–444. doi: 10.1006/abbi.1995.1486

Hilton, M. D., Alaeddinoglu, N. G., and Demain, A. L. (1988). Synthesis of bacilysin by *Bacillus subtilis* branches from prephenate of the aromatic amino acid pathway. *J. Bacteriol.* 170, 482–484.

Hjort, K., Bergström, M., Adesina, M. F., Jansson, J. K., Smalla, K., and Sjöling, S. (2010). Chitinase genes revealed and compared in bacterial isolates, DNA extracts and a metagenomic library from a phytopathogen suppressive soil. *FEMS Microbiol. Ecol.* 71, 197–207. doi: 10.1111/j.1574-6941.2009.00801.x

Hohn, T. M., and Beremand, P. D. (1989). Isolation and nucleotide sequence of a sesquiterpene cyclase gene from the trichothecene-producing fungus *Fusarium sporotrichoides*. *Gene* 79, 131–138. doi: 10.1016/0378-1119(89)90098-X

Hong, S. Y., and Linz, J. E. (2008). Functional expression and subcellular localization of the aflatoxin pathway enzyme Ver-1 fused to enhanced green fluorescent protein. *Appl. Environ. Microbiol.* 74, 6385–6396. doi: 10.1128/AEM.01185-08

Horwitz, S. B. (1994). Taxol (paclitaxel): mechanisms of action. *Ann. Oncol.* 5(Suppl. 6), S3–S6.

Howard, M. B., Ekborg, N. A., Taylor, L. E., Weiner, R. M., and Hutcheson, S. W. (2004). Chitinase B of “microbulbifer degradans” 2-40 contains two catalytic domains with different chitinolytic activities. *J. Bacteriol.* 186, 1297–1303. doi: 10.1128/JB.186.5.1297-1303.2004

Howell, C. R., and Stipanovic, R. D. (1980). Suppression of *Pythium ultimum*-induced damping-off of cotton seedlings by *Pseudomonas fluorescens* and its antibiotic, pyoluteorin. *Phytopathology* 70, 712–715. doi: 10.1094/Phyto-70-712

Hughes, J., and Mellows, G. (1980). Interaction of pseudomonic acid A with *Escherichia coli* B isoleucyl-tRNA synthetase. *Biochem. J.* 191, 209–219.

Humair, B., Wackwitz, B., and Haas, D. (2010). GacA-controlled activation of promoters for small RNA genes in *Pseudomonas fluorescens*. *Appl. Environ. Microbiol.* 76, 1497–1506. doi: 10.1128/AEM.02014-09

Ibba, M., and Söll, D. (2000). Aminoacyl-tRNA synthesis. *Ann. Rev. Biochem.* 69, 617–650. doi: 10.1146/annurev.biochem.69.1.617

Inaoka, T., Takahashi, K., Ohnishi-Kameyama, M., Yoshida, M., and Ochi, K. (2003). Guanine nucleotides guanosine 5'-diphosphate 3'-diphosphate and GTP co-operatively regulate the production of an antibiotic bacilysin in *Bacillus subtilis*. *J. Biol. Chem.* 278, 2169–2176. doi: 10.1074/jbc.M208722200

Inaoka, T., Wang, G., and Ochi, K. (2009). ScoC regulates bacilysin production at the transcription level in *Bacillus subtilis*. *J. Bacteriol.* 191, 7367–7371. doi: 10.1128/JB.01081-09

Jakeman, D. L., Bandi, S., Graham, C. L., Reid, T. R., Wentzell, J. R., and Douglas, S. E. (2009). Antimicrobial activities of jadomycin B and structurally related analogues. *Antimicrob. Agent Chemother.* 53, 1245–1247. doi: 10.1128/AAC.00801-08

Janczarek, M., Rachwał, K., Marzec, A., Grzą, J., and Palusińska-Szysz, M. (2015). Signal molecules and cell-surface components involved in early stages of the legume–rhizobium interactions. *Appl. Soil Ecol.* 85, 94–113. doi: 10.1016/j.apsoil.2014.08.010

Johnston-Monje, D., and Raizada, M. N. (2011). Conservation and diversity of seed associated endophytes in across boundaries of evolution, ethnography and ecology. *PLoS ONE* 6:e20396. doi: 10.1371/journal.pone.0020396

Kajimura, Y., and Kaneda, M. (1996). Fusaricidin A, a new depsipeptide antibiotic produced by *Bacillus polymyxa* KT-8. Taxonomy, fermentation, isolation, structure elucidation and biological activity. *J. Antibiot.* 49, 129–135. doi: 10.7164/antibiotics.49.129

Kajimura, Y., and Kaneda, M. (1997). Fusaricidins B, C and D, new depsipeptide antibiotics produced by *Bacillus polymyxa* KT-8: isolation, structure elucidation and biological activity. *J. Antibiot.* 50, 220–228. doi: 10.7164/antibiotics.50.220

Karlsson, M., and Stenlid, J. (2008). Evolution of family 18 glycoside hydrolases: diversity, domain structures and phylogenetic relationships. *J. Mol. Microbiol. Biotechnol.* 16, 208–223. doi: 10.1159/000151220

Keller, L., and Surette, M. G. (2006). Communication in bacteria: an ecological and evolutionary perspective. *Nat. Rev. Microbiol.* 4, 249–258. doi: 10.1038/nrmicro1383

Keller, N. P., Turner, G., and Bennett, J. W. (2005). Fungal secondary metabolism: from biochemistry to genomics. *Nat. Rev. Microbiol.* 3, 937–947. doi: 10.1038/nrmicro1286

Kenig, M., and Abraham, E. (1976). Antimicrobial activities and antagonists of bacilysin and anticapsin. *J. Gen. Microbiol.* 94, 37–45. doi: 10.1099/00221287-94-1-37

Kenig, M., vandamme, E. T., and Abraham, E. (1976). The mode of action of bacilysin and anticapsin and biochemical properties of bacilysin-resistant mutants. *J. Gen. Microbiol.* 94, 46–54. doi: 10.1099/00221287-94-1-46

Kevany, B. M., Rasko, D. A., and Thomas, M. G. (2009). Characterization of the complete zwittermicin A biosynthesis gene cluster from *Bacillus cereus*. *Appl. Environ. Microbiol.* 75, 1144–1155. doi: 10.1128/AEM.02518-08

Kidane, D., Carrasco, B., Manfredi, C., Rothmaier, K., Ayora, S., Tadesse, S., et al. (2009). Evidence for different pathways during horizontal gene transfer in competent *Bacillus subtilis* cells. *PLoS Genet.* 5:e1000630. doi: 10.1371/journal.pgen.1000630

Kim, B.-Y., Lee, S.-Y., Ahn, J.-H., Song, J., Kim, W.-G., and Weon, H.-Y. (2015). Complete genome sequence of *Bacillus amyloliquefaciens* subsp. *plantarum* CC178, a phyllosphere bacterium antagonistic to plant pathogenic fungi. *Genome Announc.* 3, e01368–e01314. doi: 10.1128/genomeA.01368-14

Kim, H., and Farrand, S. K. (1997). Characterization of the acc operon from the nopaline-type Ti plasmid pTiC58, which encodes utilization of agrocinopines A and B and susceptibility to agrocin 84. *J. Bacteriol.* 179, 7559–7572.

Kim, J.-G., Park, B. K., Kim, S.-U., Choi, D., Nahm, B. H., Moon, J. S., et al. (2006). Bases of biocontrol: sequence predicts synthesis and mode of action of agrocin 84, the trojan horse antibiotic that controls crown gall. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8846–8851. doi: 10.1073/pnas.0602965103

Kirner, S., Hammer, P. E., Hill, D. S., Altmann, A., Fischer, I., Weislo, L. J., et al. (1998). Functions encoded by pyrrolnitrin biosynthetic genes from *Pseudomonas fluorescens*. *J. Bacteriol.* 180, 1939–1943.

Kistler, H., and Broz, K. (2015). Cellular compartmentalization of secondary metabolism. *Front. Microbiol.* 6:68. doi: 10.3389/fmicb.2015.00068

Kobayashi, D. Y., Reedy, R. M., Bick, J., and Oudemans, P. V. (2002). Characterization of a chitinase gene from *Stenotrophomonas maltophilia* strain 34S1 and its involvement in biological control. *Appl. Environ. Microbiol.* 68, 1047–1054. doi: 10.1128/AEM.68.3.1047-1054.2002

Köroğlu, T. E., Öğülür, İ., Mutlu, S., Yazgan-Karataş, A., and Özçengiz, G. (2011). Global regulatory systems operating in bacilysin biosynthesis in *Bacillus subtilis*. *J. Mol. Microbiol. Biotechnol.* 20, 144–155. doi: 10.1159/000328639

Kraus, J., and Loper, J. E. (1995). Characterization of a genomic region required for production of the antibiotic pyoluteorin by the biological control agent *Pseudomonas fluorescens* Pf-5. *Appl. Environ. Microbiol.* 61, 849–854.

Kusari, S., Hertweck, C., and Spiteller, M. (2012). Chemical ecology of endophytic fungi: origins of secondary metabolites. *Chem. Biol.* 19, 792–798. doi: 10.1016/j.chembiol.2012.06.004

Laville, J., Blumer, C., Von Schroetter, C., Gaia, V., Défago, G., Keel, C., et al. (1998). Characterization of the hcnABC gene cluster encoding hydrogen cyanide synthase and anaerobic regulation by ANR in the strictly aerobic biocontrol agent *Pseudomonas fluorescens* CHA0. *J. Bacteriol.* 180, 3187–3196.

Leclère, V., Béchet, M., Adam, A., Guez, J., Waitelet, B., Ongena, M., et al. (2005). Mycosubtilin overproduction by *Bacillus subtilis* BBG100 enhances the organism's antagonistic and biocontrol activities. *Appl. Environ. Microbiol.* 71, 4577–4584. doi: 10.1128/AEM.71.8.4577-4584.2005

Lee, C. A., Auchtung, J. M., Monson, R. E., and Grossman, A. D. (2007). Identification and characterization of int (integrase), xis (excisionase) and chromosomal attachment sites of the integrative and conjugative element ICEBs1 of *Bacillus subtilis*. *Mol. Microbiol.* 66, 1356–1369. doi: 10.1111/j.1365-2958.2007.06000.x

Leöffler, W., Tschen, J., Venittanakom, N., Kugler, M., Knorpp, E., Hsieh, T. F., et al. (1986). Antifungal effects of bacilysin and fengycin from *Bacillus subtilis* F-29-3: a comparison with activities of other *Bacillus* antibiotics. *J. Phytopathol.* 115, 204–213. doi: 10.1111/j.1439-0434.1986.tb00878.x

Letunic, I., and Bork, P. (2011). Interactive Tree of Life v2: online annotation and display of phylogenetic trees made easy. *Nucl. Acids Res.* 39, W475–W478. doi: 10.1093/nar/gkr201

Ligon, J. M., Hill, D. S., Hammer, P. E., Torkewitz, N. R., Hofmann, D., Kempf, H. J., et al. (2000). Natural products with antifungal activity from *Pseudomonas* biocontrol bacteria. *Pest Manag. Sci.* 56, 688–695. doi: 10.1002/1526-4998(200008)56:8

Li, J., and Jensen, S. E. (2008). Nonribosomal biosynthesis of fusaricidins by *Paenibacillus polymyxa* pkb1 involves direct activation of a d-amino acid. *Chem. Biol.* 15, 118–127. doi: 10.1016/j.chembiol.2007.12.014

Li, J., Nation, R. L., Turnidge, J. D., Milne, R. W., Coulthard, K., Rayner, C. R., et al. (2006). Colistin: the re-emerging antibiotic for multidrug-resistant Gram-negative bacterial infections. *Lancet Infect. Dis.* 6, 589–601. doi: 10.1016/S1473-3099(06)70580-1

Li, S., Zhang, R., Wang, Y., Zhang, N., Shao, J., Qiu, M., et al. (2013). Promoter analysis and transcription regulation of fus gene cluster responsible for fusaricidin synthesis of *Paenibacillus polymyxa* SQR-21.

Appl. Microbiol. Biotechnol. 97, 9479–9489. doi: 10.1007/s00253-013-5157-6

Lim, F. Y., and Keller, N. P. (2014). Spatial and temporal control of fungal natural product synthesis. *Nat. Prod. Rep.* 31, 1277–1286. doi: 10.1039/C4NP00083H

Lodeiro, S., Xiong, Q., Wilson, W. K., Ivanova, Y., Smith, M. L., May, G. S., et al. (2009). Protostadienol biosynthesis and metabolism in the pathogenic fungus *Aspergillus fumigatus*. *Org. Lett.* 11, 1241–1244. doi: 10.1021/ol802696a

Lu, J., Huang, X., Li, K., Li, S., Zhang, M., Wang, Y., et al. (2009). LysR family transcriptional regulator PqsR as repressor of pyoluteorin biosynthesis and activator of phenazine-1-carboxylic acid biosynthesis in *Pseudomonas* sp. M18. *J. Biotechnol.* 143, 1–9. doi: 10.1016/j.biote.2009.06.008

Luo, Y., Ruan, L.-F., Zhao, C.-M., Wang, C.-X., Peng, D.-H., and Sun, M. (2011). Validation of the intact zwettermicin A biosynthetic gene cluster and discovery of a complementary resistance mechanism in *Bacillus thuringiensis*. *Antimicrob. Agent Chemother.* 55, 4161–4169. doi: 10.1128/AAC.00111-11

Marçet-Houben, M., and Gabaldón, T. (2010). Acquisition of prokaryotic genes by fungal genomes. *Trends Genet.* 26, 5–8. doi: 10.1016/j.tig.2009.11.007

Martínez-García, P. M., Ruano-Rosa, D., Schilirò, E., Prieto, P., Ramos, C., Rodríguez-Palenzuela, P., et al. (2015). Complete genome sequence of *Pseudomonas fluorescens* strain PICF7, an indigenous root endophyte from olive (*Olea europaea* L.) and effective biocontrol agent against *Verticillium dahliae*. *Stand. Genomic Sci.* 10:10. doi: 10.1186/1944-3277-10-10

Mavrodi, D. V., Blankenfeldt, W., and Thomashow, L. S. (2006). Phenazine compounds in fluorescent *Pseudomonas* spp. biosynthesis and regulation. *Annu. Rev. Phytopathol.* 44, 417–445. doi: 10.1146/annurev.phyto.44.013106.145710

Mavrodi, D. V., Parejko, J. A., Mavrodi, O. V., Kwak, Y., Weller, D. M., Blankenfeldt, W., et al. (2013). Recent insights into the diversity, frequency and ecological roles of phenazines in fluorescent *Pseudomonas* spp. *Environ. Microbiol.* 15, 675–686. doi: 10.1111/j.1462-2920.2012.02846.x

Mavrodi, D. V., Peever, T. L., Mavrodi, O. V., Parejko, J. A., Raaijmakers, J. M., Lemanceau, P., et al. (2010). Diversity and evolution of the phenazine biosynthesis pathway. *Appl. Environ. Microbiol.* 76, 866–879. doi: 10.1128/AEM.00209-09

Mavrodi, O. V., McSpadden, G. B. B., Mavrodi, D. V., Bonsall, R. F., Weller, D. M., and Thomashow, L. S. (2001). Genetic diversity of phlD from 2,4-diacetylphloroglucinol-producing fluorescent *Pseudomonas* spp. *Phytopathology* 91, 35–43. doi: 10.1094/PHYTO.2001.91.1.35

Mazzola, M., Cook, R. J., Thomashow, L. S., Weller, D., and Pierson, L. (1992). Contribution of phenazine antibiotic biosynthesis to the ecological competence of fluorescent pseudomonads in soil habitats. *Appl. Environ. Microbiol.* 58, 2616–2624.

McDonald, M., Mavrodi, D. V., Thomashow, L. S., and Floss, H. G. (2001). Phenazine biosynthesis in *Pseudomonas fluorescens*: branchpoint from the primary shikimate biosynthetic pathway and role of phenazine-1,6-dicarboxylic acid. *J. Am. Chem. Soc.* 123, 9459–9460. doi: 10.1021/ja011243+

McSpadden Gardener, B. B., Schroeder, K. L., Kalloge, S. E., Raaijmakers, J. M., Thomashow, L. S., and Weller, D. M. (2000). Genotypic and phenotypic diversity of phlD containing *Pseudomonas* strains isolated from the rhizosphere of wheat. *Appl. Environ. Microbiol.* 66, 1939–1946. doi: 10.1128/AEM.66.5.1939-1946.2000

Moule, Y., Jemmali, M., and Rousseau, N. (1976). Mechanism of the inhibition of transcription by pr toxin, a mycotoxin from *Penicillium roqueforti*. *Chem. Biol. Interact.* 14, 207–216. doi: 10.1016/0009-2779(76)90101-0

Mousa, W. K., and Raizada, M. N. (2013). The diversity of anti-microbial secondary metabolites produced by fungal endophytes: an interdisciplinary perspective. *Front. Microbiol.* 4:65. doi: 10.3389/fmicb.2013.00065

Nodwell, J. R. (2014). Are you talking to me? a possible role for γ butyrolactones in interspecies signalling. *Mol. Microbiol.* 94, 483–485. doi: 10.1111/mmi.12787

Nowak-Thompson, B., Chaney, N., Wing, J. S., Gould, S. J., and Loper, J. E. (1999). Characterization of the pyoluteorin biosynthetic gene cluster of *Pseudomonas fluorescens* Pf-5. *J. Bacteriol.* 181, 2166–2174.

Parker, J. B., and Walsh, C. T. (2012). Olefin isomerization regiochemistries during tandem action of bacA and bacB on prephenate in bacilysin biosynthesis. *Biochemistry* 51, 3241–3251. doi: 10.1021/bi300254u

Pessi, G., and Haas, D. (2000). Transcriptional control of the hydrogen cyanide biosynthetic genes hcnABC by the anaerobic regulator ANR and the quorum-sensing regulators LasR and RhlR in *Pseudomonas aeruginosa*. *J. Bacteriol.* 182, 6940–6949. doi: 10.1128/JB.182.24.6940-6949.2000

Pham, V. H., and Kim, J. (2012). Cultivation of unculturable soil bacteria. *Trends Biotechnol.* 30, 475–484. doi: 10.1016/j.tibtech.2012.05.007

Phister, T. G., O'Sullivan, D. J., and McKay, L. L. (2004). Identification of bacilysin, chlorotetaine, and iturin a produced by *Bacillus* sp. strain cs93 isolated from pozol, a Mexican fermented maize dough. *Appl. Environ. Microbiol.* 70, 631–634. doi: 10.1128/AEM.70.1.631-634.2004

Picard, C., Di Cello, F., Ventura, M., Fani, R., and Guckert, A. (2000). Frequency and biodiversity of 2, 4-diacetylphloroglucinol-producing bacteria isolated from the maize rhizosphere at different stages of plant growth. *Appl. Environ. Microbiol.* 66, 948–955. doi: 10.1128/AEM.66.3.948-955.2000

Pierson, I. L. S., and Pierson, E. A. (2010). Metabolism and function of phenazines in bacteria: impacts on the behavior of bacteria in the environment and biotechnological processes. *Appl. Microbiol. Biotechnol.* 86, 1659–1670. doi: 10.1007/s00253-010-2509-3

Pierson, I. L., Wood, D., Pierson, E., and Chancey, S. (1998). N-acyl-homoserine lactone-mediated gene regulation in biological control by fluorescent pseudomonads: current knowledge and future work. *Eur. J. Plant Pathol.* 104, 1–9. doi: 10.1023/A:1008674417007

Prakash, N. U., Jayanthi, M., Sabarinathan, R., Kangueane, P., Mathew, L., and Sekar, K. (2010). Evolution, homology conservation, and identification of unique sequence signatures in GH19 family chitinases. *J. Mol. Evol.* 70, 466–478. doi: 10.1007/s00239-010-9345-z

Prieto, M., and Wedin, M. (2013). Dating the diversification of the major lineages of Ascomycota (Fungi). *PLoS ONE* 8:e65576. doi: 10.1371/journal.pone.0065576

Proctor, R. H., and Hohn, T. M. (1993). Aristolochene synthase. isolation, characterization, and bacterial expression of a sesquiterpenoid biosynthetic gene (Ari1) from *Penicillium roqueforti*. *J. Biol. Chem.* 268, 4543–4548.

Proctor, R. H., McCormick, S. P., Alexander, N. J., and Desjardins, A. E. (2009). Evidence that a secondary metabolic biosynthetic gene cluster has grown by gene relocation during evolution of the filamentous fungus *Fusarium*. *Mol. Microbiol.* 74, 1128–1142. doi: 10.1111/j.1365-2958.2009.06927.x

Quecine, M. C., Araujo, W. L., Marcon, J., Gai, C. S., Azevedo, J. L., and Pizzirani-Kleiner, A. A. (2008). Chitinolytic activity of endophytic *Streptomyces* and potential for biocontrol. *Lett. Appl. Microbiol.* 47, 486–491. doi: 10.1111/j.1472-765X.2008.02428.x

Raffel, S. J., Stabb, E. V., Milner, J. L., and Handelsman, J. (1996). Genotypic and phenotypic analysis of zwettermicin A-producing strains of *Bacillus cereus*. *Microbiology* 142, 3425–3436. doi: 10.1099/13500872-142-12-3425

Rajavel, M., Mitra, A., and Gopal, B. (2009). Role of *Bacillus subtilis* BacB in the synthesis of bacilysin. *J. Biol. Chem.* 284, 31882–31892. doi: 10.1074/jbc.M109.014522

Reader, J. S., Ordoukhian, P. T., Kim, J.-G., de Crécy-Lagard, V., Hwang, I., Farand, S., et al. (2005). Major biocontrol of plant tumors targets tRNA synthetase. *Science* 309, 1533–1533. doi: 10.1126/science.1116841

Rehmeyer, C., Li, W., Kusaba, M., Kim, Y.-S., Brown, D., Staben, C., et al. (2006). Organization of chromosome ends in the rice blast fungus, *Magnaporthe oryzae*. *Nucl. Acids Res.* 34, 4685–4701. doi: 10.1093/nar/gkl588

Roberts, W., and Tate, M. (1977). Agrocin 84 is a 6-N-phosphoramidate of an adenine nucleotide analogue. *Nature* 265, 379–381. doi: 10.1038/265379a0

Roze, L. V., Chanda, A., and Linz, J. E. (2011). Compartmentalization and molecular traffic in secondary metabolism: a new understanding of established cellular processes. *Fungal Genet. Biol.* 48, 35–48. doi: 10.1016/j.fgb.2010.05.006

Santoyo, G., Orozco-Mosqueda, M. D. C., and Govindappa, M. (2012). Mechanisms of biocontrol and plant growth-promoting activity in soil bacterial species of *Bacillus* and *Pseudomonas*: a review. *Bio. Sci. Technol.* 22, 855–872. doi: 10.1080/09583157.2012.694413

Schäberle, T. F., Vollmer, W., Frasch, H.-J., Hüttel, S., Kulik, A., Röttgen, M., et al. (2011). Self-resistance and cell wall composition in the glycopeptide producer *Amycolatopsis balhimycina*. *Antimicrob. Agent. Chemother.* 55, 4283–4289. doi: 10.1128/AAC.01372-10

Schardl, C. L. (2010). The epichloae, symbionts of the grass subfamily poöideae1. *Ann. Mo. Bot. Gard.* 97, 646–665. doi: 10.3417/2009144

Scheffler, R., Colmer, S., Tynan, H., Demain, A., and Gullo, V. (2013). Antimicrobials, drug discovery, and genome mining. *Appl. Microbiol. Biotechnol.* 97, 969–978. doi: 10.1007/s00253-012-4609-8

Schneider, K., Chen, X.-H., Vater, J., Franke, P., Nicholson, G., Borriss, R., et al. (2007). Macrolactin is the polyketide biosynthesis product of the pks2 cluster of *Bacillus amyloliquefaciens* FZB42. *J. Nat. Prod.* 70, 1417–1423. doi: 10.1021/np070070k

Schwarzer, D., Finking, R., and Marahiel, M. A. (2003). Nonribosomal peptides: from genes to products. *Nat. Prod. Rep.* 20, 275–287. doi: 10.1039/b111145k

Shaheen, M., Li, J., Ross, A. C., Vedera, J. C., and Jensen, S. E. (2011). *Paenibacillus polymyxa* PKB1 produces variants of polymyxin B-type antibiotics. *Chem. Biol.* 18, 1640–1648. doi: 10.1016/j.chembiol.2011.09.017

Shanahan, P., Borro, A., O'Gara, F., and Glennon, J. D. (1992). Isolation, trace enrichment and liquid chromatographic analysis of diacetylphloroglucinol in culture and soil samples using UV and amperometric detection. *J. Chromatogr.* 606, 171–177. doi: 10.1016/0021-9673(92)87022-Z

Silby, M. W., Winstanley, C., Godfrey, S. A., Levy, S. B., and Jackson, R. W. (2011). *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol. Rev.* 35, 652–680. doi: 10.1111/j.1574-6976.2011.00269.x

Silo-Suh, L. A., Stabb, E. V., Raffel, S. J., and Handelsman, J. (1998). Target range of zwittermicin A, an aminopolysaccharide antibiotic from *Bacillus cereus*. *Curr. Microbiol.* 37, 6–11. doi: 10.1007/s002849900328

Silva, G. H., de Oliveira, C. M., Teles, H. L., Pauletti, P. M., Castro-Gamboa, I., Silva, D. H. S., et al. (2010). Sesquiterpenes from *Xylaria* sp., an endophytic fungus associated with *Piper aduncum* (Piperaceae). *Phytochem. Lett.* 3, 164–167. doi: 10.1016/j.phytol.2010.07.001

Soliman, S., and Raizada, M. N. (2013). Interactions between co-habitating fungi elicit synthesis of Taxol from an endophytic fungus in host Taxus plants. *Front. Microbiol.* 4:3. doi: 10.3389/fmicb.2013.00003

Soliman, S., Trobacher, C., Tsao, R., Greenwood, J., and Raizada, M. N. (2013). A fungal endophyte induces transcription of genes encoding a redundant fungicide pathway in its host plant. *BMC Plant Biol.* 13:93. doi: 10.1186/1471-2229-13-93

Souza, J. T., and Raaijmakers, J. M. (2003). Polymorphisms within the prnD and pltC genes from pyrrolnitrin and pyoluteorin-producing *Pseudomonas* and *Burkholderia* spp. *FEMS Microbiol. Ecol.* 43, 21–34. doi: 10.1111/j.1574-6941.2003.tb01042.x

Spatafora, J. W., Sung, G.-H., Johnson, D., Hesse, C., O'Rourke, B., Serdani, M., et al. (2006). A five-gene phylogeny of Pezizomycotina. *Mycologia* 98, 1018–1028. doi: 10.3852/mycologia.98.6.1018

Spiering, M. J., Moon, C. D., Wilkinson, H. H., and Schardl, C. L. (2005). Gene clusters for insecticidal lolaine alkaloids in the grass-endophytic fungus *Neotyphodium uncinatum*. *Genetics* 169, 1403–1414. doi: 10.1534/genetics.104.035972

Stegmann, E., Frasch, H.-J., Kilian, R., and Pozzi, R. (2015). Self-resistance mechanisms of actinomycetes producing lipid II-targeting antibiotics. *Int. J. Med. Microbiol.* 305, 190–195. doi: 10.1016/j.ijmm.2014.12.015

Stein, T. (2005). *Bacillus subtilis* antibiotics: structures, synthesis and specific functions. *Mol. Microbiol.* 56, 845–857. doi: 10.1111/j.1365-2958.2005.04587.x

Steinborn, G., Hajirezaei, M.-R., and Hofemeister, J. (2005). bac genes for recombinant bacilysin and anticapsin production in *Bacillus* host strains. *Arch. Microbiol.* 183, 71–79. doi: 10.1007/s00203-004-0743-8

Stewart, E. J. (2012). Growing unculturable bacteria. *J. Bacteriol.* 194, 4151–4160. doi: 10.1128/JB.00345-12

Storm, D. R., Rosenthal, K. S., and Swanson, P. E. (1977). Polymyxin and related peptide antibiotics. *Ann. Rev. Biochem.* 46, 723–763. doi: 10.1146/annurev.bi.46.070177.003451

Sutherland, R., Boon, R., Griffin, K., Masters, P., Slocum, B., and White, A. (1985). Antibacterial activity of mupirocin (pseudomonic acid), a new antibiotic for topical use. *Antimicrob. Agents Chemother.* 27, 495–498. doi: 10.1128/AAC.27.4.495

Svercel, M., Duffy, B., and Defago, G. (2007). PCR amplification of hydrogen cyanide biosynthetic locus hcnaB in *Pseudomonas* spp. *J. Microbiol. Methods* 70, 209–213. doi: 10.1016/j.mimet.2007.03.018

Takeno, A., Okamoto, A., Tori, K., Oshima, K., Hirakawa, H., Toh, H., et al. (2012). Complete genome sequence of *Bacillus cereus* NC7401, which produces high levels of the emetic toxin cereulide. *J. Bacteriol.* 194, 4767–4768. doi: 10.1128/JB.01015-12

Tazawa, J., Watanabe, K., Yoshida, H., Sato, M., and Homma, Y. (2000). Simple method of detection of the strains of fluorescent *Pseudomonas* spp. producing antibiotics, pyrrolnitrin and phloroglucinol. *Soil Microorg.* 54, 61–67.

Thimon, L., Peypoux, F., Wallach, J., and Michel, G. (1995). Effect of the lipopeptide antibiotic, iturin A, on morphology and membrane ultrastructure of yeast cells. *FEMS Microbiol. Lett.* 128, 101–106. doi: 10.1111/j.1574-6968.1995.tb07507.x

Tijerino, A., Elena, C. R., Moraga, J., Malmierca, M. G., Vicente, F., Aleu, J., et al. (2011). Overexpression of the trichodiene synthase gene tri5 increases trichodermin production and antimicrobial activity in *Trichoderma brevicompactum*. *Fungal Gene. Biol.* 48, 285–296. doi: 10.1016/j.fgb.2010.11.012

Tobes, R., and Pareja, E. (2005). Repetitive extragenic palindromic sequences in the *Pseudomonas syringae* pv. tomato DC3000 genome: extragenic signals for genome reannotation. *Res. Microbiol.* 156, 424–433. doi: 10.1016/j.resmic.2004.10.014

Tobes, R., and Pareja, E. (2006). Bacterial repetitive extragenic palindromic sequences are DNA targets for insertion sequence elements. *BMC Genomics* 7:62. doi: 10.1186/1471-2164-7-62

Tsuge, K., Akiyama, T., and Shoda, M. (2001). Cloning, sequencing, and characterization of the iturin A operon. *J. Bacteriol.* 183, 6265–6273. doi: 10.1128/JB.183.21.6265-6273.2001

van Rij, E. T., Girard, G., Lugtenberg, B. J., and Bloemberg, G. V. (2005). Influence of fusaric acid on phenazine-1-carboxamide synthesis and gene expression of *Pseudomonas chlororaphis* strain PCL1391. *Microbiology* 151, 2805–2814. doi: 10.1099/mic.0.28063-0

Vetsigian, K., Jajoo, R., and Kishony, R. (2011). Structure and evolution of *Streptomyces* interaction networks in soil and *in silico*. *PLoS Biol.* 9:e1001184. doi: 10.1371/journal.pbio.1001184

Vining, L. C. (1992). Secondary metabolism, inventive evolution and biochemical diversity - a review. *Gene* 115, 135–140. doi: 10.1016/0378-1119(92)90551-Y

Viswanathan, R., and Samiyappan, R. (2001). Antifungal activity of chitinases produced by some fluorescent pseudomonads against *Colletotrichum falcatum* Went causing red rot disease in sugarcane. *Microbiol. Res.* 155, 309–314. doi: 10.1016/S0944-5013(01)80009-4

Vizcaino, M. I., Guo, X., and Crawford, J. M. (2014). Merging chemical ecology with bacterial genome mining for secondary metabolite discovery. *J. Ind. Microbiol. Biotechnol.* 41, 285–299. doi: 10.1007/s10295-013-1356-5

Voisard, C., Bull, C. T., Keel, C., Laville, J., Maurhofer, M., Schnider, U., et al. (1994). "Biocontrol of root diseases by *Pseudomonas fluorescens* CHA0: current concepts and experimental approaches," in *Molecular Ecology of Rhizosphere Microorganisms*, eds F. O'Gara, D. Dowling, and B. Boesten (Weinheim: VCH Publishers), 67–89.

Voisard, C., Keel, C., Haas, D., and D'Efago, G. (1989). Cyanide production by *Pseudomonas fluorescens* helps suppress black root rot of tobacco under gnotobiotic conditions. *EMBO J.* 8, 351–358.

Walker, J., and Abraham, E. (1970). The structure of bacilysin and other products of *Bacillus subtilis*. *Biochem. J.* 118, 563–570.

Wallwey, C., and Li, S. (2011). Ergot alkaloids: structure diversity, biosynthetic gene clusters and functional proof of biosynthetic genes. *Nat. Prod. Rep.* 28, 496–510. doi: 10.1039/C0NP00060D

Wang, C.-L., Farrand, S. K., and Hwang, I. (1994). Organization and expression of the genes on pAgK84 that encode production of agrocin 84. *Mol. Plant Microbe Interact.* 7, 472–481. doi: 10.1094/MPMI-7-0472

Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P., and McPhail, A. T. (1971). Plant antitumor agents. VI. isolation and structure of Taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J. Am. Chem. Soc.* 93, 2325–2327. doi: 10.1021/ja00738a045

Wei, C., Hansen, B. S., Vaughan, M. H., and McLaughlin, C. S. (1974). Mechanism of action of the mycotoxin trichodermin, a 12,13-epoxytrichothecene. *Proc. Natl. Acad. Sci. U.S.A.* 71, 713–717. doi: 10.1073/pnas.71.3.713

Westman, E. L., Yan, M., Waglechner, N., Koteva, K., and Wright, G. D. (2013). Self resistance to the atypical cationic antimicrobial peptide edeine of *Brevibacillus brevis* Vm4 by the N-acetyltransferase EdeQ. *Chem. Biol.* 20, 983–990. doi: 10.1016/j.chembiol.2013.06.010

Whatling, C. A., Hodgson, J. E., Burnham, M. K. R., Clarke, N. J., Franklin, F. C. H., and Thomas, C. M. (1995). Identification of a 60 kb region of the chromosome of *Pseudomonas fluorescens* NCIB 10586 required for the biosynthesis of pseudomonic acid. *Microbiology* 141, 973–982. doi: 10.1099/13500872-141-4-973

Whitehead, N. A., Barnard, A. M., Slater, H., Simpson, N. J., and Salmond, G. P. (2001). Quorum-sensing in Gram-negative bacteria. *FEMS Microbiol. Rev.* 25, 365–404. doi: 10.1111/j.1574-6976.2001.tb00583.x

Wissing, F. (1974). Cyanide formation from oxidation of glycine by a *Pseudomonas* species. *J. Bacteriol.* 117, 1289–1294.

Xie, N.-Z., Li, J.-X., Song, L.-F., Hou, J.-F., Guo, L., Du, Q.-S., et al. (2014). Genome sequence of type strain *Paenibacillus polymyxa* DSM 365, a highly efficient producer of optically active (R, R)-2, 3-butanediol. *J. Biotechnol.* 195, 72–73. doi: 10.1016/j.jbiotec.2014.07.441

Xiong, Z.-Q., Yang, Y.-Y., Zhao, N., and Wang, Y. (2013). Diversity of endophytic fungi and screening of fungal paclitaxel producer from Anglo-jap yew, *Taxus x media*. *BMC Microbiol.* 13:71. doi: 10.1186/1471-2180-13-71

Yanagisawa, T., Lee, J. T., Wu, H. C., and Kawakami, M. (1994). Relationship of protein-structure of isoleucyl-transfer-RNA synthetase with pseudomonic acid resistance of *Escherichia coli* - proposed mode of action of pseudomonic acid as an inhibitor of isoleucyl-transfer-RNA synthetase. *J. Biol. Chem.* 269, 24304–24309.

Yang, F., and Cao, Y. (2012). Biosynthesis of phloroglucinol compounds in microorganisms - review. *Appl. Microbiol. Biotechnol.* 93, 487–495. doi: 10.1007/s00253-011-3712-6

Yang, K., Han, L., and Vining, L. C. (1995). Regulation of jadomycin B production in *Streptomyces venezuelae* ISP5230: involvement of a repressor gene, jadR2. *J. Bacteriol.* 177, 6111–6117.

Yang, Y., Zhao, H., Barrero, R. A., Zhang, B., Sun, G., Wilson, I. W., et al. (2014). Genome sequencing and analysis of the paclitaxel-producing endophytic fungus *Penicillium aurantiogriseum* NRRL 62431. *BMC Genomics* 15:69. doi: 10.1186/1471-2164-15-69

Yazgan, A., Özçengiz, G., and Marahiel, M. A. (2001). Tn10 insertional mutations of *Bacillus subtilis* that block the biosynthesis of bacilysin. *Biochim. Biophys. Acta* 1518, 87–94. doi: 10.1016/S0167-4781(01)00182-8

Zeigler, D. R. (2011). The genome sequence of *Bacillus subtilis* subsp. *spizizenii* W23: insights into speciation within the *B. subtilis* complex and into the history of *B. subtilis* genetics. *Microbiology* 157, 2033–2041. doi: 10.1099/mic.0.048520-0

Zhang, D.-X., Nagabhyru, P., Blankenship, J. D., and Schardl, C. L. (2010). Are loline alkaloid levels regulated in grass endophytes by gene expression or substrate availability? *Plant Signal. Behav.* 5, 1419–1422. doi: 10.4161/psb.5.11.13395

Zhang, L., Li, X., Zhang, F., and Wang, G. (2014). Genomic analysis of *Agrobacterium radiobacter* DSM 30147T and emended description of *A. radiobacter* (Beijerinck and van Delden 1902) Conn 1942 (Approved Lists 1980) emend. Sawada et al. 1993. *Stand. Genomic Sci.* 9, 574. doi: 10.4056/sigs.4688352

Zhao, J., Mou, Y., Shan, T., Li, Y., Zhou, L., Wang, M., et al. (2010). Antimicrobial metabolites from the endophytic fungus *Pichia guilliermondii* isolated from *Paris polyphylla* var. *yunnanensis*. *Molecules* 15, 7961–7970. doi: 10.3390/molecules15117961

Zhou, R., Rasooly, R., and Linz, J. E. (2000). Isolation and analysis of fluP, a gene associated with hyphal growth and sporulation in *Aspergillus parasiticus*. *Mol. Gen. Genet.* 264, 514–520. doi: 10.1007/s004380000335

Zhou, X., Zhu, H., Liu, L., Lin, J., and Tang, K. (2010). A review: recent advances and future prospects of Taxol-producing endophytic fungi. *Appl. Microbiol. Biotechnol.* 86, 1707–1717. doi: 10.1007/s00253-010-2546-y

Zimmerman, S. B., Schwartz, C. D., Monaghan, R. L., Pelak, B. A., Weissberger, B., Gilfillan, E. C., et al. (1987). Difidin and oxydifidin: novel broad spectrum antibacterial antibiotics produced by *Bacillus subtilis*. I. production, taxonomy and antibacterial activity. *J. Antibiot.* 40, 1677–1681. doi: 10.7164/antibiotics.40.1677

Zotchev, S. B., Sekurova, O. N., and Katz, L. (2012). Genome-based bioprospecting of microbes for new therapeutics. *Curr. Opin. Biotechnol.* 23, 941–947. doi: 10.1016/j.copbio.2012.04.002

Zou, Z., Du, D., Zhang, Y., Zhang, J., Niu, G., and Tan, H. (2014). A γ -butyrolactone-sensing activator/repressor, JadR3, controls a regulatory mini-network for jadomycin biosynthesis. *Mol. Microbiol.* 94, 490–505. doi: 10.1111/mmi.12752

Conflict of Interest Statement: 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.

Copyright © 2015 Mousa and Raizada. 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) or licensor 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.



Target or barrier? The cell wall of early- and later-diverging plants vs cadmium toxicity: differences in the response mechanisms

Luigi Parrotta¹, Gea Guerriero², Kjell Sergeant², Giampiero Cai^{1*} and Jean-Francois Hausman^{2*}

¹ Dipartimento Scienze della Vita, Università di Siena, Siena, Italy

² Environmental Research and Innovation, Luxembourg Institute of Science and Technology, Esch-sur-Alzette, Luxembourg

Edited by:

Joanna M.-F. Cross, Inönü University, Turkey

Reviewed by:

Zoë A. Popper, National University of Ireland, Ireland

Danuta M. Antosiewicz, University of Warsaw, Poland

Metin Turan, Yeditepe University, Turkey

***Correspondence:**

Giampiero Cai, Dipartimento Scienze della Vita, Università di Siena, via Mattioli 4, I-53100 Siena, Italy

e-mail: giampiero.cai@unisi.it;

Jean-Francois Hausman, Environmental Research and Innovation, Luxembourg Institute of Science and Technology, 5, Avenue des Hauts-Fourneaux, L4362 Esch-sur-Alzette, Luxembourg

e-mail: jean-francois.hausman@list.lu

Increasing industrialization and urbanization result in emission of pollutants in the environment including toxic heavy metals, as cadmium and lead. Among the different heavy metals contaminating the environment, cadmium raises great concern, as it is ecotoxic and as such can heavily impact ecosystems. The cell wall is the first structure of plant cells to come in contact with heavy metals. Its composition, characterized by proteins, polysaccharides and in some instances lignin and other phenolic compounds, confers the ability to bind non-covalently and/or covalently heavy metals via functional groups. A strong body of evidence in the literature has shown the role of the cell wall in heavy metal response: it sequesters heavy metals, but at the same time its synthesis and composition can be severely affected. The present review analyzes the dual property of plant cell walls, i.e., barrier and target of heavy metals, by taking Cd toxicity as example. Following a summary of the known physiological and biochemical responses of plants to Cd, the review compares the wall-related mechanisms in early- and later-diverging land plants, by considering the diversity in cell wall composition. By doing so, common as well as unique response mechanisms to metal/cadmium toxicity are identified among plant phyla and discussed. After discussing the role of hyperaccumulators' cell walls as a particular case, the review concludes by considering important aspects for plant engineering.

Keywords: cadmium, plant cell wall, heavy metal stress, heavy metal biosorption, cell wall polysaccharides, lignin

INTRODUCTION

Plants, differently from animals, are sessile organisms and therefore cannot escape from potentially life-threatening conditions. This, in part, explains the great metabolic plasticity of plant cells, which have evolved mechanisms enabling them to adapt to and cope with environmental challenges. Plants growing on contaminated soils typically display either tolerance or avoidance. Plants tolerate heavy metals by sequestering them in specific plant organelles to keep them segregated from vital cellular components, or by the synthesis of enzymes involved in detoxification. Alternatively, the uptake and translocation of the heavy metal is decreased. However, there is a wide diversity in the details of the response, and species- or even clone-specific molecular responses have been recorded. Cadmium (Cd) is known to be phytotoxic and to affect plant physiological processes, from roots to shoots. Adverse effects caused by Cd are of considerable importance for all plants but, from an agricultural point of view, Cd is a toxic environmental pollutant that affects crop productivity (Du et al., 2012). While numerous reports have analyzed the cellular and enzymatic mechanisms involved in the response to Cd stress in plants, only a handful of studies have focused on the role of the plant cell wall. The cell wall is the outermost structure of plant

cells and therefore the first in contact with the environment. It has important chemical characteristics, which make it a very good biosorbent of heavy metals. The presence of different functional groups (e.g., carboxyl, hydroxyl) deriving from the different wall polysaccharides favors for instance ion-exchange mechanisms with wall counter-ions. The cell wall is in its turn affected by heavy metals, since both its biosynthesis and composition can be altered. This leads to modifications in its physico-chemical properties which increase the binding capacity and at the same time lower the entry of heavy metals in the protoplast (Krzesłowska, 2011).

IMPACT OF Cd POLLUTION ON PLANT CELLS AND ORGANS

Cd can have many consequences on plant physiology. It can affect both shoot growth and leaf biomass in *Zea mays* by impairing chlorophyll synthesis and by promoting the expression of defense proteins (Lagriffoul et al., 1998). In garlic (*Allium sativum*), Cd reduces root growth in a concentration-dependent manner (Liu et al., 2003).

It is also capable of damaging processes related to plant reproduction, with logical impacts on plant dispersal and biodiversity. For example, Cd can affect pollen germination and pollen tube growth by altering the polarization mechanism and

by inducing abnormalities within the cell, such as blocking cytoplasmic streaming and alteration of the cytoplasmic organization (Sawidis, 2008). Moreover, in a recent study, the influence of exposure to a variety of heavy metals on the germination of 23 flax cultivar seeds was assessed. Using root elongation as the main parameter, large cultivar-dependent differences were found for some heavy metals (notably Cd, Ni and Co), while for other metals a more homogenous influence is described (Soudek et al., 2010).

The impact of Cd on crops can also be exerted by negatively influencing the translocation of nutrients, especially of minerals; in Cd-exposed tomato K, Fe, Mn, and Zn are reported to be poorly translocated in roots, while P and Mn uptake is drastically reduced in fruits (Moral et al., 1994). In pea, Cd has effects on both roots and leaves, and a significant inhibition of growth is combined with a reduction of transpiration and photosynthesis rate, as well as a general deterioration of the nutrient status (Sandhalio et al., 2001). These are just a few examples describing the broad range of Cd-caused impacts on plants. For more information on the overall effects caused by Cd, readers can refer to specific reviews (Benavides et al., 2005; Qadir et al., 2014).

Cd is essentially absorbed by roots, but it localizes to all plant organs and tissues. Leaves are usually the preferential accumulation site of Cd. In leaves, Cd accumulates in different regions, which correspond to less metabolically active areas and to forthcoming necrotic regions. This suggests that an active metabolism might be required to detoxify Cd in plants and that the absence of detoxification mechanisms precedes necrotic events (Cosio et al., 2005). In the leaves of sensitive plants, Cd exerts its toxic activity by inducing chlorosis, thereby inhibiting growth. The effects caused by Cd range from morphological changes to reduction of the photosynthesis rate to decreased transpiration up to cell apoptosis (Souza et al., 2011).

Cd also affects the structure and function of roots. In *Allium cepa*, Cd treatment induces abnormalities, such as extensive vacuolization, condensation of the cytoplasm, damages to mitochondrial cristae, plasmolysis, and condensation of chromatin. Although dense granules were detected between the cell wall and the plasma membrane, no specific sites of Cd-accumulation were found in the cell walls of *Allium* roots (Liu and Kottke, 2004).

As plants are exposed to different environmental stresses, the effects of Cd contamination cannot be isolated from injuries caused by other environmental stresses and it might sometimes be difficult to predict the combinatory effect (Mittler, 2006). The simultaneous presence of different stresses might multiply the negative effects; however, as already observed in rice seedlings, one type of stress can also partially suppress the effects of another. In rice, for example, heat stress can ameliorate the negative effects of Cd pollution through the activation of protective systems based on anti-oxidant activities (Shah et al., 2013). However, the study of combined stresses is much less advanced than that of single stresses, although some studies have been performed combining Cd-exposure with other treatments (Sergeant et al., 2014). Given that polluted soils are generally not contaminated with Cd alone, the exposure of plants to mixtures of metals (Printz et al., 2013a,b), or the use of real-life polluted soils in experiments (Evlard et al., 2014a), could give new insights into the metabolic

adjustments of plants when exposed to high concentrations of trace nutrients.

ROLE OF PHYTOCHELATINS AND METALLOTHIONEINS IN THE RESPONSE OF PLANTS TO Cd STRESS

Although Cd is a generally toxic contaminant of the ecosystem, plants have evolved diverse mechanisms to respond to Cd contamination of their growth substrate. As for many compounds with detrimental effects on cellular metabolism, the toxicity of Cd is often alleviated through its sequestration to specific cellular compartments, such as vacuoles, or within specialized cells, such as trichomes (Harada et al., 2010). Adaptation of plants to Cd-contaminated soils might also rely on the symbiotic cooperation with other organisms. In *Medicago sativa*, colonization of roots by arbuscular mycorrhizae increases the tolerance of plants to Cd and the decreased Cd toxicity observed is somewhat proportional to the extent of colonization. It is likely that fungi possess a battery of enzymatic activities/mechanisms capable of chemically modifying Cd, thereby making it less toxic (Wang et al., 2012).

The cell wall can act effectively as a biosorbent of Cd, alleviating the toxic effects of this heavy metal. It is clear that the cell wall-based protection mechanism is only one of the processes used by plants to cope with the damage induced by Cd. Before exploring more specifically the molecular mechanisms in which the cell wall is involved, it is necessary to introduce the other protective mechanisms with which plants are equipped. These are based on specialized oligopeptides (namely phytochelatins and metallothioneins), on biochemical responses and on the intracellular sequestration of Cd.

Plants can respond by producing specific chelating agents, such as phytochelatins, that complex with Cd, thereby reducing its toxic potential (Carrier et al., 2003). The production of phytochelatins coincides with the activation of sulfur metabolism. This increases the synthesis of cysteine and of reduced glutathione (GSH), an antioxidant precursor of phytochelatins (Zhang and Shu, 2006). Increased levels of sulfur can also affect the distribution of Cd in plants by remobilizing it from the cell wall to the cytosolic compartment. As sulfur is important in the biosynthesis of sulfhydryl proteins, it is suggested that a higher content of sulfur generates larger amounts of proteins capable of sequestering Cd in the cytosol, thereby reducing the accumulation of the metal in the cell wall (Zhang et al., 2014c). Therefore, the availability of sulfur might control the synthesis rate of these sulfur-rich proteins (Loeffler et al., 1989; Uegsegger et al., 1990). Although this information is inferred from indirect observations (i.e., by determining that, in the presence of higher levels of sulfur, Cd is preferentially located in the vacuole), this also suggests that the balance between cell wall-bound and cytosol-associated Cd depends on the availability of proteins capable of sequestering Cd.

The use of phytochelatins for the detoxification of Cd-stressed plants is considered an important mechanism, but its actual role is still unclear and seems to be organ-dependent. Roots of the hyperaccumulator *Sedum alfredii* do not make efficient use of the phytochelatin-based mechanism for protection against Cd, while this process is conversely used in the shoots of the same plant (Zhang et al., 2010b). The use of different tolerance mechanisms in shoots and roots is also found in other hyperaccumulators

such as *Typha angustifolia*, one based on GSH-related antioxidant systems (in leaves) and the other based on GSH-related chelation system (in roots; Xu et al., 2011).

An important class of cysteine-rich proteins binding heavy metals in plants and animals are metallothioneins (MTs). Their role in heavy metal stress in plants is known and their potential use in biotechnology is supported by recent studies, which have analyzed the effects of MTs overexpression in *Arabidopsis thaliana* and tobacco (Gu et al., 2014; Zhou et al., 2014). The transformed plants showed increased tolerance to Cd stress, which was accompanied by a lower accumulation of H₂O₂ in *A. thaliana* and by a higher ROS scavenging activity in tobacco.

Data in the literature indicate that a universal mechanism of Cd tolerance is not present and that different, often closely-related plants, respond differently when exposed to Cd. For instance, Cd-tolerant cultivars of black oat (*Avena strigosa*) accumulate Cd in the leaves, mainly in the cell wall. Those plants use a phytochelatin-based response when exposed to Cd, but also induce the up-regulation of ascorbate peroxidase and superoxide dismutase, indicating that an anti-oxidant response is triggered upon Cd treatment. By contrast, the phytochelatin-based response appears unimportant in Cd-sensitive plants, although their total content increased upon heavy metal exposure (Uraguchi et al., 2009).

OTHER FACTORS AFFECTING THE RESPONSE/ADAPTATION/DEFENSE OF PLANTS TO Cd

The number and diversity of resistance mechanisms used by plants might be larger than expected. For example, it was reported that treatment of Cd-stressed plants with salicylic acid alleviates a number of effects caused by Cd. It is likely that the positive effects of salicylic acid are not directly correlated with the removal of Cd, but with the activation of response and defense genes, which in turn may code for enzymes involved in the maintenance of homeostasis of plant cells (Moussa and El-Gamal, 2010). Activation of pathogenesis-related defense proteins of the glucanase family is likely to be a common trait of many plants in response to Cd. This process was observed in different plants, for instance maize and soybean treated with Cd and other metals, suggesting that this defense mechanism is shared by relatively distant species (Pirselová et al., 2011).

The response of plants to Cd also requires the activation of enzymes capable of refolding proteins, as one of the most dramatic effects of intracellular Cd is protein denaturation. Therefore, some of the enzymes activated during response and adaptation to Cd belong to the chaperone family and chaperone-like proteins are likely active during the response to Cd pollution. As Cd can denature proteins, heat-shock proteins (HSPs) are involved in the refolding of denatured proteins and indeed levels of HSPs have been reported to increase after Cd treatment (Sergio et al., 2007). Increased levels of HSP70 were also detected in the bryophyte *Conocephalum conicum* exposed to Cd and Pb, suggesting that the need to refold proteins is a prerequisite to maintain cellular activity during heavy metal stress (Basile et al., 2013).

In pea roots, Cd and Cu treatment were reported to affect the enzymatic activity of proteins involved in the oxidation and peroxidation of cell wall components, such as guaiacol peroxidase, ascorbate peroxidase, coniferyl alcohol peroxidase, NADH oxidase and indole-3-acetic acid (IAA) oxidase (Chaoui et al., 2004).

The evidence that Cd might increase the activity of cell wall-bound coniferyl alcohol peroxidase is of particular interest because this enzymatic activity metabolizes coniferyl alcohol, a monolignol used as substrate presumably involved in the lignification process (Quiroga et al., 2001). This suggests that Cd pollution might have severe repercussions on the development of secondary cell walls, by reducing their rigidity and robustness. An increase in guaiacol peroxidase activity was also reported in lichens exposed to higher concentrations of Cd (Sanità Di Toppi et al., 2005). As lichens are fungi-algae symbionts generally resistant and tolerant to heavy metal pollution, the use of plasma membrane-associated peroxidases might be a general mechanism to counteract the negative effects caused by Cd.

Changes in cell wall-associated peroxidase activity were also observed in roots of *Brassica juncea* stressed by Cd treatment. Here, Cd increases the activity of ionically cell wall-bound proteins, thereby increasing peroxidase activity too (Verma et al., 2008). The upregulated peroxidase activity is supposedly involved in the adaptation mechanism to Cd stress.

Although specific studies have not been conducted to determine the impact of Cd exposure on the cell wall proteome and on the diversity of responses in closely related species, numerous cell wall localized proteins were identified in general proteome studies on the impact of heavy metal pollution, namely Cd excess, on plants. These include general stress responsive proteins. For instance, in a study on two flax cultivars with different tolerance levels to Cd exposure, the cell wall-localized chitinase increased in both cultivars, but the protein accumulated at significantly higher levels in the more tolerant cultivar (Hradilová et al., 2010).

Proteomics studies in poplar found that exposure to Cd resulted in an increased accumulation of β -1,3-glucanase and chitinase (Kieffer et al., 2008; Durand et al., 2010). Moreover Kieffer et al. (2008) identified the significantly higher accumulation of different isoforms of cell wall localized peroxidase, linking the cell wall and anti-oxidative enzymes found in this compartment with Cd-stress. An intra-specific comparison of different *Populus nigra* ecotypes found differences in the mechanism of tolerance, indicating the variability and highlighting the difficulty to generate a physiological model, even at the species level. In Cd-stressed poplar plants, the activity of superoxide dismutase decreases, while the enzymatic activity of both peroxidase and catalase increased significantly in roots; this suggests that these plants trigger a response based on enhanced oxidase activity. Although the cytology of root cells is significantly altered, these plants exhibit a considerable capacity as hyperaccumulators and they have been consequently proposed as phytoremediators (Ge et al., 2012).

In addition to the various mechanisms outlined above, extracellular compartmentalization is another strategy used by many plants to limit the injuries caused by Cd. This mechanism involves the binding and accumulation of Cd in the cell wall, which serves as a reservoir for Cd. It is capable of concentrating and accumulating Cd, thus preventing it from penetrating inside the cell where it can cause damages. In the legume white lupin, the ability of the cell wall and vacuole-deposited phytochelatins

to sequester Cd is relatively comparable in stems, but the cell wall is more effective in both roots and leaves (Vázquez et al., 2006). This suggests that these organs have specialized cell walls capable of sequestering Cd, while stems have mostly developed the phytochelatin-based mechanism for protection against Cd. Predominant accumulation of Cd in the cell wall of roots was also observed in *Spartina alterniflora* (Pan et al., 2012) and *Kandelia obovata* (Weng et al., 2012) treated with different concentrations of Cd. Enhanced accumulation of Cd in the cell wall of leaf cells was conversely reported in *Alternanthera philoxeroides* (Xu et al., 2012a).

Before discussing how Cd affects plant cell wall and how it can be a barrier against Cd, we need to introduce its key features.

THE PLANT CELL WALL: A LIVING STRUCTURE

Plant cells are enveloped by a layer composed of structural proteins and polysaccharides (which can be impregnated by the aromatic polymer lignin), interwoven to form an intricate mesh, the cell wall. This structure has a complex three-dimensional organization and is considered an example of a natural biocomposite. The plant cell wall constitutes a material with exceptional mechanical performance and is taken as a model for the creation of materials with superior properties.

Cell walls protect the living protoplasts against external insults but, at the same time, they are very plastic since modifications in their composition and structure are responsible for enabling cell growth and development. Two types of cell walls can be distinguished: primary and secondary cell walls. The first is typically found in actively growing plant cells and is characterized by the presence of cellulose, pectin and hemicellulose. It is quite thin and flexible, thus enabling cell expansion (Guerriero et al., 2014a). Secondary walls are specialized structures synthesized when plant cells have ceased to elongate. They are responsible for mechanical strengthening (namely via the deposition of lignin) and accompany the phase of secondary growth, which determines the increase in stem girth (Guerriero et al., 2014a,b). The progressive stem thickening ensures resistance to bending. The main components of secondary walls are cellulose, xylan and lignin (Cosgrove and Jarvis, 2012).

A third type of wall (more specifically a wall layer) can be distinguished, which is found in specialized cell types and/or in response to mechanical stimuli: the gelatinous wall layer (G-layer). This type of wall is characterized by the presence of a thick layer of crystalline cellulose and is typically found in tension wood. Cells with a G-layer are also found in the stem of fiber crops like hemp, flax and nettle and are associated with the phloem. These cells are very long and known as bast fibers (Mellerowicz and Gorshkova, 2012; Guerriero et al., 2013).

Besides encasing and shielding plant protoplasts, the cell wall can be affected by nutritional stress and undergoes variations in composition. For example grapevine callus subjected to deficiency in S, N, and P showed walls with a decreased content in cellulose, increased lignin amount, alterations in the methylesterification of pectin (Fernandes et al., 2013). These results show how plant cell walls are remodeled in response to mineral deficits and how their composition is modified according to the stress condition applied. The plant cell wall indeed takes active part in transmitting

exogenous signals to the interior of the cells: evidences show the existence of a cell wall integrity (CWI) maintenance mechanism which is triggered upon exogenous stresses (Hamann and Denness, 2011; Wolf et al., 2012; Engelsdorf and Hamann, 2014). The knowledge concerning CWI maintenance is far from being complete; however, given the similarity existing with that of yeast, both chemical and physical signals are supposed to contribute to it (Hamann and Denness, 2011). An example of a physical signal is represented by the weakening of the cell wall (for instance as a consequence to treatments with drugs like dichlobenil or isoxaben) up to a point that it can no longer counteract the internal turgor pressure with consequent cell swelling and bursting. A weakened wall causes stretching of the plasma membrane and activation of mechanosensitive channels which trigger an increase in the cytoplasmic concentration of calcium (Humphrey et al., 2007; Hamann and Denness, 2011).

Chemical signals are generated upon degradation of the cell wall by pathogens: for example, pectin-derived oligogalacturonides (OGAs) with a specific length, degree of acetylation and methylesterification, are capable of inducing rapid defense responses (Hamann and Denness, 2011; Vallarino and Osorio, 2012; Ferrari et al., 2013). The presence of signals, which allow the wall to “sense” the status of the cell, implies the existence of receptors that can intercept those signals. Several receptors have been described in the literature and among them wall associated kinases (WAKs) have been well studied. Those receptors are located in the plasma membrane, but have a domain protruding into the cell wall and a cytoplasmic kinase domain (Anderson et al., 2001; Humphrey et al., 2007). Their structure is therefore ideal to create a continuum cell wall-plasma membrane-cytoplasm and to transfer the signal from the outside to the inside of the cell. The role of these receptors is particularly interesting if one considers the relationship existing between heavy metal stress and water balance within plant cells. Gating of aquaporins was recorded within 10 min of heavy metal application to onion epidermal cells, which results in alterations in the turgor pressure (Przedpelska-Wasowicz and Wierzbicka, 2011). Since the WAKs membrane receptors are activated upon changes in turgor pressure (a phenomenon observed in pollen grains of *Picea wilsonii* after Cd stress; Wang et al., 2014), their functional study could provide important insights into the sensing mechanism and reveal downstream players involved in the signaling cascade.

THE CELL WALL AS A BARRIER TO Cd

The composition of the cell wall shows many differences among phyla and it has played a vital role for their survival throughout evolution (Sarkar et al., 2009 and references therein). In the attempt to understand how different taxonomic groups respond to Cd (or heavy metals in general), it is valuable to compare the chemical composition and structure of cell walls from different phyla. Information might shed light either on alternative resistance mechanisms, or on how to improve the strategy used to tolerate Cd. The rationale at the basis of this assumption is that the variety of responses to Cd pollution, as found among different phyla, might be mirrored by molecular and cytological “tricks” that plants have evolved at the cell wall level. In discussing this issue, it is necessary to consider that evidence for the cell wall

acting as a barrier against Cd (and other heavy metals) is not immediately clear from every study. The availability of molecular and structural data is low, which leads to fragmented pictures strongly affecting the development of a unifying model. Moreover, it is not possible to rule out that a unifying model does not exist and therefore it is necessary to present individual situations. The ability of either a plant cell or a plant *in toto* to tolerate Cd can be implemented at different levels and with different mechanisms. In the attempt to outline the different resistance mechanisms in which cell walls are involved, it is possible to rank them based on (a) the morphology of the barrier organ (for example the root), (b) the transport capacity of the contaminant, (c) the absorption of Cd into the cell wall by polysaccharides or proteins, (d) the impediment of Cd to penetrate through the plasma membrane.

The first part of the section aims at surveying the wall-related strategies developed by early- and later-diverging embryophytes. Charophytes, the closest algal relatives of embryophytes, are also briefly treated, given their importance in the study of land plant cell wall origin (Sørensen et al., 2011; Mikkelsen et al., 2014). The second part of the section discusses the mechanisms of Cd absorption/translocation in different plant organs and tissues, while considering the cell wall organization/composition.

PLANT CELL WALL ACROSS DIFFERENT PHYLA: COMPOSITION AND MECHANISMS OF RESISTANCE TO Cd

Charophyte cell walls possess high amounts of mannose-containing hemicellulose, glucuronic acid, mannuronic acid, 3-O-methyl rhamnose (Popper and Fry, 2003) and possess neither lignin nor cutin. More importantly, they are capable of calcifying and calcite is known to bind and sequester heavy metals (Gomes and Asaeda, 2013). The macrophyte alga *Chara australis* can accumulate Cd and be used for remediation of contaminated soils (Clabeaux et al., 2013). Studies carried out on another species of macrophyte, *Chara fragilis*, showed that binding of uranyl species to the cell walls was mainly due to the presence of calcite (Daković et al., 2008). The biomineralization potential of Charophytes is therefore very promising for phytoremediation (i.e., the use of algae for the removal of pollutants) and further studies should be performed to link the cell wall composition of these algae to their calcifying potential.

Bryophytes, a group of early-diverging non-vascular plants, possess walls with mannose-containing hemicellulose, uronic acids and 3-O-methyl rhamnose, similarly to Charophytes (Sarkar et al., 2009). They lack lignin, although they contain lignans and other lignin-like polymers (Sarkar et al., 2009 and references therein). Bryophytes are known bioindicators of heavy metal pollution and their walls show high biosorption capacity because of the numerous ion-exchange sites (e.g., from uronic acids). The wall of the moss *Pohlia drummondii* was shown to form, together with the plasma membrane, an efficient shield against elevated zinc doses (Lang and Wernitznig, 2011); therefore, the living protoplast is protected from harmful effects. The moss *Scorpiurum circinatum* treated with different heavy metals, including Cd, immobilized the toxic ions in the cell walls, which is therefore used as the main detoxification site (Basile et al., 2012).

The Pteridophytes, another group of early-diverging plants, are very interesting organisms, as their cell walls show special-

ization to support the differentiation of a vascular tissue: galactomannan, glucomannan are abundant in secondary walls of leptosporangiates and lignin, together with xylan, has also been observed (Sarkar et al., 2009 and references therein). In the fern *Lygodium japonicum* it was shown that the wall pectins bind copper via homogalacturonans (Konno et al., 2005). In *Salvinia auriculata* Cd was shown to induce severe deformations at the cell wall level. Moreover an opaque layer could be observed along the middle lamella (Wolff et al., 2012). This finding might be explained by the binding capacity of pectins, the chief component of middle lamellas, a feature which is shared by seed plants. Among Pteridophytes, Equisetales certainly deserve attention, as their walls show the occurrence of mixed-linkage glucans and they are known for their biosilicification activity. Interestingly, a relationship was proposed between these hemicelluloses and Si mineralization (Fry et al., 2008). It would be interesting to investigate whether the sequestration of heavy metals by Equisetales takes place via mixed-linkage glucans.

These data indicate that the cell wall is a target for the accumulation of Cd in early-diverging plants, such as Bryophytes and Pteridophytes. These organisms use molecular mechanisms that are compatible with the chemical composition of their cell wall. However, as outlined above, both the structure and composition of the cell wall changed during evolution and it is therefore expected that chemically-different cell walls might respond differently to Cd pollution. The in-depth study of early-diverging plant cell wall biosynthesis and composition can reveal fine molecular details and inspire further avenues for improving the cell wall capacity to accumulate toxic heavy metals (see section “Future outlook”).

In later-diverging plants (such as spermatophytes), the increasing number of tissues and organs, coupled with the more sophisticated architecture, also generated cell walls with an increasing variability at the chemical and physical level. Consequently, it is likely that the diversity in Cd tolerance extends to the organ, tissue and even to the cellular level (see next paragraph).

In seed plants, and more specifically in flowering plants, most of the information was obtained from a limited number of species and, sometimes, on crop plants.

The evidence that the cell wall can exert a protective role was supported by studying Cd toxicity in crop plants. Pea plants were shown to be more sensitive than maize plants to Cd exposure; compared to pea plants, maize plants exhibited a higher percentage of cell wall-associated Cd, while pea showed higher content of Cd in the cytoplasm (Lozano-Rodríguez et al., 1997). This finding suggests that the cell wall acts as a barrier, preventing Cd from entering the cytoplasm where it is extremely toxic for plant cells. It is consequential that different cell wall structures and textures might bind Cd differently and thus provide different levels of protection. Comparable results were found in strawberry, where the treatment with different concentrations of Cd highlighted that the cell wall of leaves and roots is the primary reservoir for Cd, with root cell walls showing higher binding capacity than leaf cell walls (Zhang et al., 2010a). The fact that roots display a higher accumulating capacity than leaves can be an index of their protective activity preventing the entry and accumulation of Cd in the cytoplasm. If the concept of the cell wall as a barrier is true,

the differential accumulating capacity of plant cell walls has to be found in their chemical and physical diversity. This diversity can cover not only the polysaccharide component of the cell wall, but also the protein fraction or modifications induced on one, the other or both components.

PLANT CELL WALL COMPOSITION IN DIFFERENT ORGANS AND TISSUES: LINK WITH Cd ABSORPTION AND TRANSLOCATION

Since Cd is primarily adsorbed at the root level, it follows that the anatomy and molecular structure of the cell wall in root cells is a critical parameter. Findings in different plant species indicate that roots with higher content of suberin and lignin may be more impermeable to Cd and therefore more resistant to its absorption and translocation (Lux, 2010). Increased accumulation of suberin was observed in the roots of the monocotyledonous medicinal plant *Merwilla plumbea* exposed to Cd, a process interpreted as a protective response against penetration of Cd in the cells (Lux et al., 2011). The presence of root mucilage is another factor playing an important role in Cd adsorption, as shown in sunflower (Yang and Pan, 2013). The differential structure of roots might explain why Cd translocation is different between plants and variable even within the same taxonomic group.

Screening of different clones of *Salix* grown on identical real-life soil revealed remarkable differences in the translocation of Cd and other metals (Evlard et al., 2014b). Similar differences were found in solanaceous plants (Yamaguchi et al., 2011), different poplar species (He et al., 2013) and between ecotypes of the hyperaccumulators *Thlaspi caerulescens* and *T. praecox* (Xing et al., 2008).

In plants restraining Cd in roots, Cd is sequestered in the endoderm and the cortical regions of roots. When this filter fails, Cd can penetrate the xylem vessels and translocate to the shoot. Therefore, the specific composition of cell walls in the endoderm and cortex is critical (Akhter et al., 2014). Ultimately, the cell wall of roots is considered as an effective barrier against penetration of Cd and a relationship can be established between higher accumulation of Cd in the roots vs lower accumulation in the leaves (Sun et al., 2013). This suggests that Cd-tolerance goes through the capacity of roots to limit the diffusion of Cd towards the entire plant.

The protective activity exerted by the cell wall may not only depend on the polysaccharide fraction, but also on other cell wall molecules that can bind Cd and other heavy metals. As earlier suggested by Sanità Di Toppi and Gabbrielli (1999), the response of later-diverging plants to Cd may involve different types of molecular mechanisms, from the phytochelatin-based sequestration, to cell wall immobilization and plasma membrane exclusion, to the use of stress proteins. In this context, the cell wall (by acting as a molecular barrier) may be part of both adaptation mechanisms to permanent pollution and response mechanisms to acute stress. However, additional factors might be required to achieve full functionality and protection against Cd pollution. In fact, in addition to the binding of Cd to polysaccharides and proteins, tolerance to Cd might also involve the sequestration of Cd complexes in the cell wall. In crop plants such as maize and wheat, addition of P concomitantly to Cd treatment increased the tolerance of plants to Cd (Zhimin et al., 1999). P may

associate with Cd thereby generating insoluble complexes that remain within the cell wall. Combination of binding to P and polysaccharides is also possible.

Therefore, the addition of P to contaminated soils seems to modify their physiochemical status, making Cd less available for plants, thereby improving the tolerance of plants to Cd pollution. In addition, Cd accumulates more consistently in the cell wall of plants growing on P-supplemented soils, which therefore hinder its entry in the cytoplasm (Siebers et al., 2013).

If the differential capacity of plants to resist Cd is linked to the chemical variability of their cell walls, it is clear that such variability can be consistently increased by modifying the protein component of the cell wall. Specific cell wall proteins could provide an additional barrier by binding Cd and thus preventing it from penetrating into the cytoplasm. In barley treated with Ni and/or Cd, the apoplastic fraction of leaves showed increased amount of apoplastic proteins that accumulate concomitantly with the metal treatment (Blinda et al., 1997). Therefore, the response of plants to contaminating metals may also involve the specific *de novo* synthesis of proteins that potentially make plants more tolerant. Proteins not directly involved in Cd binding have also been reported to increase Cd tolerance in plants. As previously discussed, an example is represented by cysteine-rich proteins as identified in *Digitaria ciliaris* and *Oryza sativa*. The genes coding for these proteins were found to be upregulated following Cd treatment and transgenic plants overexpressing cysteine-rich proteins are more tolerant to Cd stress by preventing entry into the cytoplasm (Kuramata et al., 2009). This finding also highlights the possibility of manipulating plants in order to increase their mechanisms of tolerance to pollutant agents.

Accumulation of Cd in the cell wall (or more generally in the apoplast) is also dependent on the efficient transport of Cd through the plasma membrane so that cytoplasmic Cd can be transported outside and incorporated in the cell wall. While doing so, the content of cytoplasmic Cd decreases while Cd accumulates in the cell wall. Transport of Cd through the plasma membrane is likely dependent on membrane transporters. Their involvement is also suggested by indirect evidence showing that, while Cd is promptly adsorbed in the cell wall, its influx in the symplast is linear with Cd concentration, which suggests the activity of low-affinity transport systems (Redjala et al., 2009). Recently the transporter IRT1 (IRON-REGULATED TRANSPORTER 1) was shown to be not only involved in the uptake of iron, but also of toxic metals, among which Cd (Barberon et al., 2014). This root transporter is recycled via a mechanism involving a phosphatidylinositol-3-phosphate-binding protein, FYVE1, which controls its delivery to the outer plasma membrane domain of root epidermal cells.

These observations stress the importance of studying the cell wall structure of both tolerant and resistant plants in order to identify those molecules that provide the highest protection against Cd. One example of an important molecule is lignin. The protective activity of lignin is suggested by pretreatment of wheat plants with wheat germ agglutinin, which generally exerts an amelioration of Cd effects. This pretreatment likely involves a different balance between hormone levels but an important phenotypic effect is an accelerated lignification of cell walls (Bezrukova et al.,

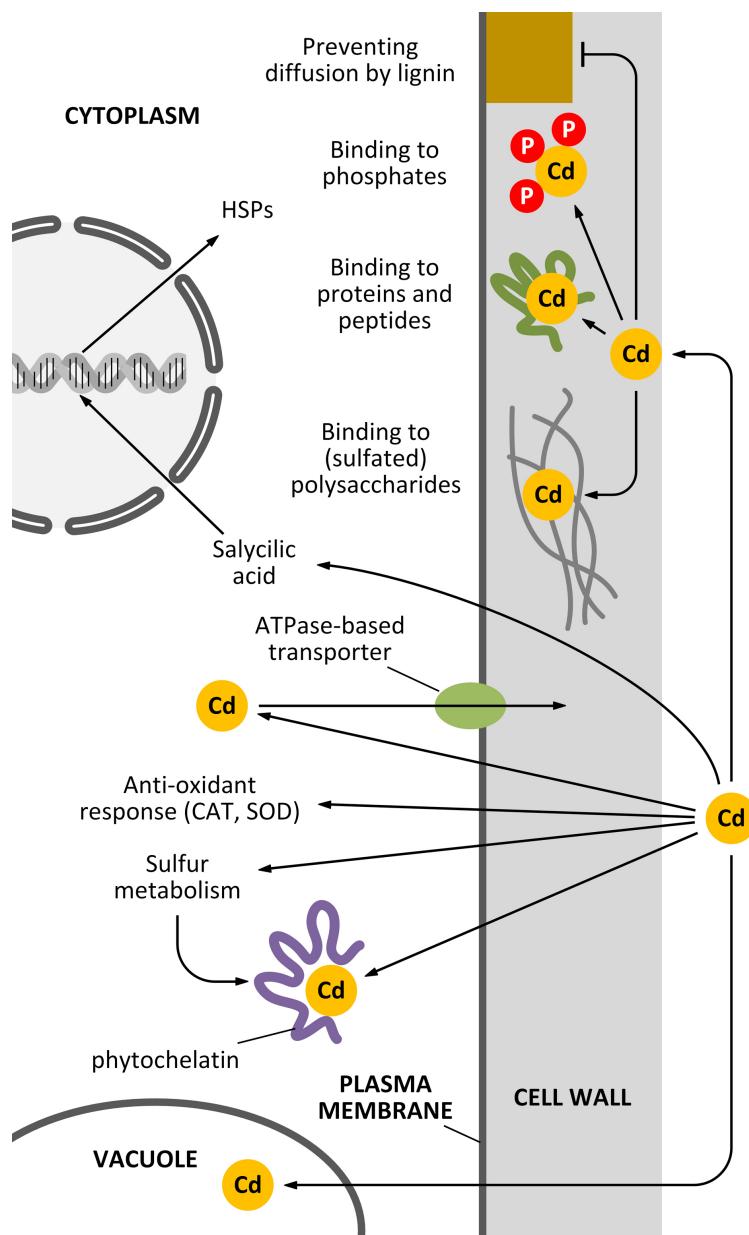


FIGURE 1 | How the cell wall can act as a barrier against Cd pollution. Plant cells can counteract the toxicity of Cd in several ways and the cell wall actively participates to most of these mechanisms. Apart from mechanisms of sequestration based on the activity of vacuoles and phytochelatins, additional decontaminating activities can be exerted by the cell wall. For example, specific polysaccharides of the cell wall can bind Cd

through sulfated moieties (although this situation is restricted to very few cases, e.g., seaweeds and vascular plants adapted to inhabit saline environments). Cd might also be sequestered in the cell wall by the activity of either cell wall-associated proteins or cell wall phosphates. An additional barrier might be represented by secondary modifications of the cell wall, such as lignin deposition.

2011). Such secondary modifications might render cell walls less permeable to Cd thus limiting its entry in the cell. Lignin can also bind Cd thereby inhibiting its diffusion into the cells. Higher abundance of lignin in secondary cell walls might thus make plants more tolerant to the effects of Cd treatment.

In *Pisum sativum* leaves, lignin was shown to be deposited following Cd stress and this process was accompanied by an oxidative burst in the xylem vessel cell walls (Rodríguez-Serrano

et al., 2009). Moreover, Cd was shown to favor the process of xylogenesis, by inducing the activity of enzymes via H_2O_2 accumulation (Schützendübel et al., 2001). The use of sophisticated labeling techniques involving quantum dot technologies can highlight the interaction between Cd and specific cell wall components (Djikanovic et al., 2012). The multitude of activities of the cell wall as a barrier against Cd pollution is schematically summarized in Figure 1.

THE CELL WALL AS TARGET: EFFECTS OF Cd ON CELL WALL SYNTHESIS AND COMPOSITION

In the previous paragraphs, the cell wall has been considered as a potential reservoir for Cd and thus as an efficient barrier to its penetration in the plant. In some cases, Cd is not efficiently neutralized and it might become an injuring factor by affecting the structure of cell walls. The toxic activity of Cd can be exerted at different levels. Cd can alter the physical structure of the cell wall by interacting with negatively charged molecules (i.e., carboxyl groups and sulfates). In this way, Cd can substantially modify the cell wall resistance to turgor pressure thereby making the cells weaker. In addition to interacting with the polysaccharide component, Cd can interfere with the enzymatic activities present in the cell walls by inhibiting the enzymatic reactions that strengthen the cell wall structure. Furthermore, Cd can affect directly or indirectly the synthesis of cell wall components, or interfere with the transport of the latter to the final destination.

Presence of Cd also affects the deposition pattern of pectins. In the hypocotyl of *Linum usitatissimum*, the presence of an excess of Cd changes the ratio between low and high methyl-esterified pectins, causing the former to accumulate consistently in epidermal cells and determining the collapse of the sub-epidermal cell layer (Douchiche et al., 2007). This unbalance affects the primary walls and may cause swelling of hypocotyl tissues. It is not clear how Cd determines such effects, but it is likely that it alters the expression level of pectin-metabolizing enzymes such as pectin methyl-esterase (PME). Overexpression of this enzyme can change the ratio between low and high methyl-esterified pectins by altering their level of esterification. This leads to a relatively higher abundance of acidic pectins and extremely impairs the shaping of plant cells, which is based on a precise balance in degree of esterification. In addition to changing the degree of esterification of pectins, Cd can also affect more generally the chemistry of these polysaccharides for example by altering the expression of peroxidases, which in their turn alter the chemical structure of homogalacturonans (Paynel et al., 2009). The effects caused by Cd on the development of cell walls and, more specifically, on pectins might be more subtle and not immediately appreciable. This is the case for flax fibers where Cd treatment affects the deposition of secondary cell walls in terms of adhesion of cellulose microfibrils. On the contrary, the effects on both low and high methyl-esterified pectins are not immediately explicable (Douchiche et al., 2011).

That pectins are important in binding Cd is demonstrated by experiments with plants grown in P deficiency. Lower levels of P alleviate the negative effects caused by Cd because P reduces the content of pectins and PME in the cell wall. The reduced level of this enzyme enhances the ratio between low and high methyl-esterified pectins further, making a lower number of active sites available for Cd binding, thereby triggering a heavy metal-exclusion mechanism (Zhu et al., 2012).

Further studies on the role of pectins in response to Cd exposure were carried out in rice. The addition of NO donors concurrently with Cd treatment makes plants more resistant to Cd injuries. The resistance is mainly based on the overproduction

of pectins and on the simultaneous reduction of cellulose levels (Xiong et al., 2009). It is likely that pectins are Cd adsorbent via carboxylic groups, thus preventing the metal from entering the cytoplasm. It would be fascinating to analyze if NO donors alter the expression of genes coding for enzymes involved in pectin synthesis. The possible role of NO in alleviating the symptoms caused by Cd treatment has been already reviewed (Xiong et al., 2010).

The evidence that toxic elements such as heavy metals could bind to pectin in the cell walls has been already observed and described (Malovikova and Kohn, 1982). There is a vast literature on the chemical interaction between heavy metals and pectins (Krzeslowska, 2011) although the exact mechanism of damage caused by Cd on pectins is not completely clear. It is likely that Cd can substitute calcium ions thereby modifying the rigidity of the pectin skeleton. This fact has been suggested in the alga *Ulva lactuca* where the rhamnose subunits are cross-linked by calcium and the replacement of calcium by Cd can decrease the rigidity of the cell wall (Webster and Gadd, 1996). It is noteworthy that calcium can alleviate the injuries caused by Cd, likely either because calcium can compete for the binding to cell wall polysaccharides, or Cd might use calcium channels to penetrate inside cells (Suzuki, 2005). At the cellular level, Cd can induce remarkable aberrations on the cell wall structure resulting in changes in cell shape (Sawidis and Reiss, 1995; Sawidis, 2008). These observations are consistent with the hypothesis that Cd interferes with the process of cell wall structuring probably by altering the exact arrangement of pectins. In some experimental cases, exposure to soils contaminated with Cd induces a general increase of the levels of pectin and a reduction of methyl-esterified pectins (Astier et al., 2014). Nevertheless, understanding exactly the alterations induced by Cd on the structure of the pectin skeleton is not simple and the use of antibodies directed against different chemical forms of pectin cannot help to define exactly how Cd damages the pectin cell wall (Douchiche et al., 2011).

Direct effects of Cd on the synthesis of other polysaccharides (such as cellulose and callose) are not known. However, it was reported that Cd triggers the accumulation of specific Cd-induced glycine-rich proteins (cdiGRPs). These are cell wall-associated proteins that might positively regulate the synthesis of callose. The cdiGRPs likely work in cooperation with an additional cell wall protein, the so-called GrIP (cdiGRP-interacting protein; Ueki and Citovsky, 2005). The model proposed by the authors suggests that Cd increases the post-translational accumulation of cdiGRP, which is also dependent on the expression of GrIP. Once accumulated in the cell wall, cdiGRP enhances the production and/or accumulation of callose, which consequently represents a cell wall adaptation to Cd pollution. Accumulation of callose in response to metal stress probably does not involve altered mechanisms of callose removal. In maize and soybean, treatment with metals (including Cd) induces the accumulation of callose in the cell wall. Results indicate that the accumulation is not related to decreased enzymatic activity of glucanases and reinforce the hypothesis that accumulation of callose is primarily dependent on a higher rate of synthesis and/or deposition (Piršelová et al., 2012). Like callose, there are few reports on the relationship

between Cd and cellulose and the relationship between Cd pollution and cellulose synthesis is far from being completely understood.

The effects of Cd on cell wall synthesis might also be indirect. Cd is known to affect the level of apoplastic sucrose probably by altering the activity of cell wall invertase. Consequently, sucrose accumulates in the cell wall and decreases in the cytoplasm (Podazza et al., 2006). Lower levels of cytoplasmic sucrose can have detrimental effects on cell wall synthesis by making substrates for the biosynthesis of cell wall precursors less available. Synthesis of both cellulose and callose requires the addition of UDP-glucose, which can be produced by two different metabolic pathways. The first requires the enzyme UDP-glucose pyrophosphorylase (UGPase) that catalyzes the reversible production of UDP-glucose and pyrophosphate from glucose-1-phosphate and UTP; the second mechanism involves the activity of sucrose synthase, which breaks down sucrose (in the presence of UTP) into UDP-glucose and fructose (Kleczkowski et al., 2010). Lower levels of cytoplasmic sucrose may reduce the activity of sucrose synthase and consequently may lower the production rate of UDP-glucose. As sucrose synthase is associated with both callose synthase and cellulose synthase (Brill et al., 2011), it follows that the Cd-induced reduction of cytoplasmic sucrose might affect the synthesis of cellulose and callose.

Another pathway by which Cd can affect the synthesis and deposition of cell wall is indirect and relates to the effects of Cd on cytoplasmic streaming. It is known that delivery of cell wall synthesizing enzymes requires an intact cytoskeleton and an intact membrane system: cell wall synthesizing enzymes traffic along the cytoskeleton, which guides the enzymes towards the final insertion sites in the plasma membrane (Crowell et al., 2010). The precise balance between insertion and removal of cell wall synthesizing enzymes is critical to correctly deposit the cell wall and to assemble a proper cell wall texture. Cd affects the intracytoplasmic movement of organelles and vesicles in the epidermal cells of the model plant *A. cepa* (Wierzbicka et al., 2007). Cytoplasmic streaming disorders are likely to impact negatively on the cell wall assembly. We do not know how Cd affects organelle streaming, if through an unspecific pathway (for example, by altering the ionic homeostasis of cells), or a more specific mechanism that directly inhibits the proteins involved in organelle transport. Nevertheless, it is clear that Cd has a considerable influence on the structure of actin filaments and consequently on cytoplasmic streaming. In *A. thaliana* root hairs, Cd alters the balance between influx and efflux of Ca^{2+} , which consequently abolishes the Ca^{2+} gradient. Since Ca^{2+} regulates the dynamics of actin filaments in tip-growing cells, the Cd-induced imbalance of Ca^{2+} has negative effects on actin filaments that consecutively impair the motility of organelles and vesicles, as reviewed by Wan and Zhang (2012). As observed by the authors, the cell wall structure is also affected because of the altered cytoplasmic streaming (Fan et al., 2011). The relationship between Cd and Ca^{2+} is likely more complex than expected in view of the finding that a member of the plant cadmium-resistance (PCR) protein family is a Ca^{2+} efflux transporter, which is involved in the finer regulation of intracellular Ca^{2+} concentration (Song et al., 2011).

The global effects of Cd on cell wall synthesis are summarized in Figure 2.

Cd ACCUMULATION IN PLANTS: THE CASE OF HYPERACCUMULATORS AND THE EFFECT OF NUTRIENTS

In the previous sections, data concerning how Cd can be a harmful element for plants and how plants can respond to damage caused by Cd were presented and discussed. Emphasis was put on how evolutionarily-distant plants can respond in different ways to Cd treatment; this is an indication that several different mechanisms have evolved in response to Cd. In this section, the focus will be put on those plants showing heavy metal hyperaccumulating capacity, since they represent an extreme case. Moreover, the study of their cell walls can lead to important observations, which can be transferred to biotechnological approaches. In this section considerations on the role of macro- and micro-nutrients in the increased Cd accumulation/tolerance will also be made, since this aspect is linked with the potential of enhancing the plant capacity to accumulate Cd.

Hyperaccumulator plants are characterized by the extraordinary capacity of accumulating Cd and of translocating it towards the aerial parts of plants (namely leaves) and of detoxifying the element locally (Rascio and Navari-Izzo, 2011). At the level of leaves, Cd can be accumulated in specific sites, such as the vacuoles or the cell wall. It is therefore possible to assume that in hyperaccumulators the ability of the cell wall to absorb Cd is so high that Cd fails to actively penetrate inside the cytoplasm and remains confined within the cell wall. Even if the molecular mechanisms working in hyperaccumulator plants are the same as those present in Cd-resistant plants, the main difference is that the former can concentrate Cd in the cell wall at extreme levels without apparent damage. There are many examples of hyperaccumulators and, as the study is continuing, new species are added to this list. The root hairs of *T. caerulescens*, a metal hyperaccumulator, can accumulate consistent levels of Cd in their cell walls thereby preventing Cd from penetrating in the cytoplasm (Nedelkoska and Doran, 2000). Massive accumulation of Cd in hyperaccumulator plants is also based on complexes with phosphates within the cell wall. However, not all cell types accumulate Cd at the same level (Küpper et al., 2000). Accumulation of Cd and Zn may have particular relevance in hyperaccumulator plants, such as *Sedum plumbizincicola*, in which heavy metals are chiefly accumulated in shoots and relatively less in roots (Cao et al., 2014). *Dittrichia viscosa* is another model-accumulator plant that accumulates large amounts of Cd in the cell wall (Fernández et al., 2014).

The ability of certain species to accumulate Cd is linked not only to the higher ROS scavenging activity, but also to Fe homeostasis: studies carried out on the Cd-accumulator *Solanum nigrum* and the low Cd-accumulating *Solanum torvum* showed that Cd treatment caused a lower Fe accumulation in *S. torvum* compared to *S. nigrum* (Xu et al., 2012b). This difference is related to the increased expression of Fe transporters, namely IRT1, and IRT2, in the roots of the Cd-accumulator *S. nigrum* (Xu et al., 2012b).

The most important application of hyperaccumulator plants is linked to the possibility of decontaminating polluted sites

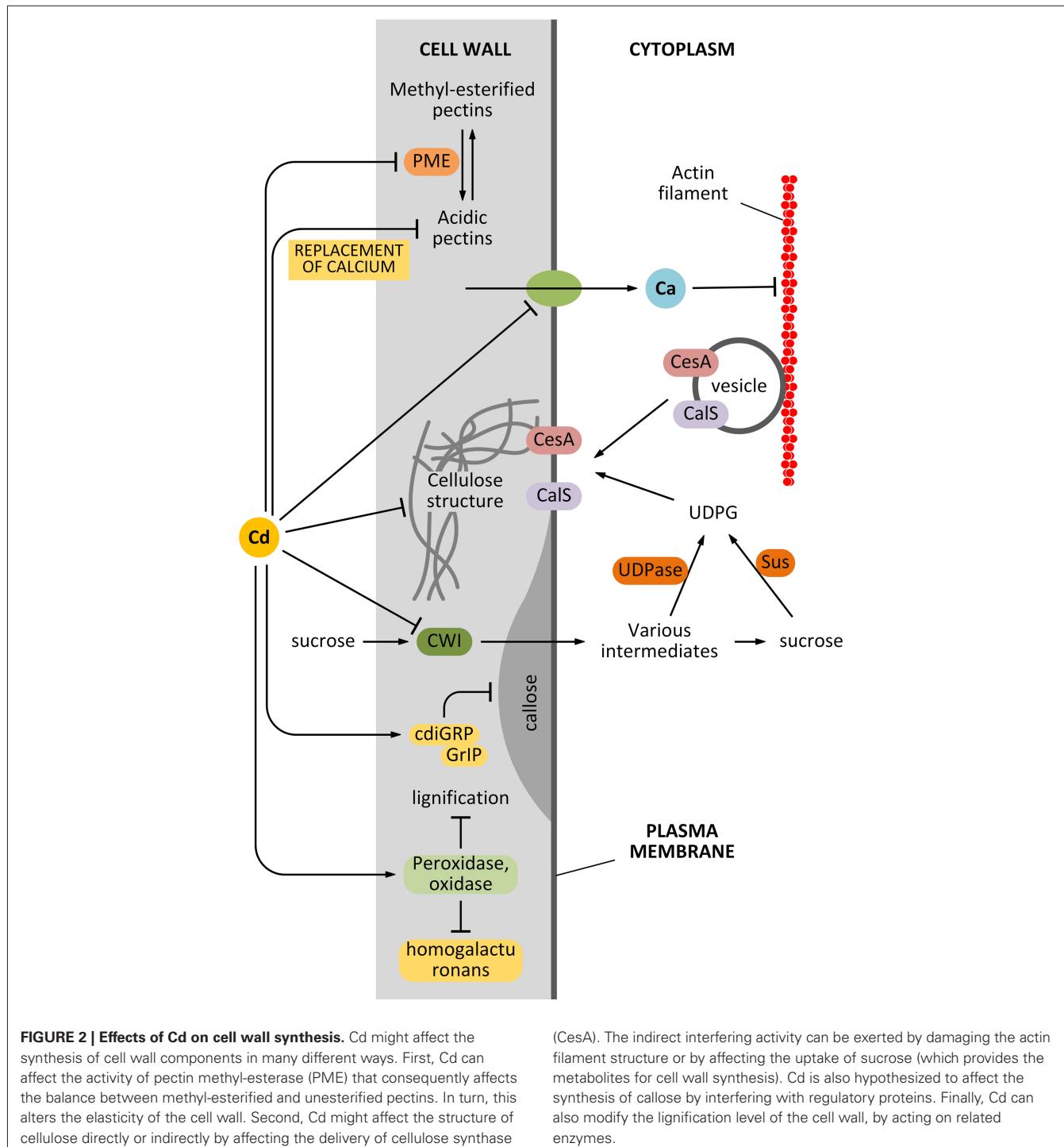


FIGURE 2 | Effects of Cd on cell wall synthesis. Cd might affect the synthesis of cell wall components in many different ways. First, Cd can affect the activity of pectin methyl-esterase (PME) that consequently affects the balance between methyl-esterified and unesterified pectins. In turn, this alters the elasticity of the cell wall. Second, Cd might affect the structure of cellulose directly or indirectly by affecting the delivery of cellulose synthase

(CesA). The indirect interfering activity can be exerted by damaging the actin filament structure or by affecting the uptake of sucrose (which provides the metabolites for cell wall synthesis). Cd is also hypothesized to affect the synthesis of callose by interfering with regulatory proteins. Finally, Cd can also modify the lignification level of the cell wall, by acting on related enzymes.

(see also “Future outlook”). Consequently, researchers are looking for hyperaccumulators that can be used to detoxify contaminated soils in an economically feasible and reasonable way. This has been already proposed for two species of *Iris* (Han et al., 2007). Those plants, capable of restoring optimal conditions in contaminated soils, have evolved molecular mechanisms to tolerate high concentrations of contaminants. The acquisition of tolerance to contaminants involves the development of mechanisms for

the accumulation of pollutants in specific cell sites, such as the vacuole and cell wall of root cells where Cd might be complexed with insoluble phosphates that strongly limit the translocation of Cd to aboveground tissues and organs (Zhang et al., 2014b). This suggests that the hyperaccumulating capacity of cell walls can be induced or improved by adding specific chemical groups. This can involve both the polysaccharide and the protein component of the cell wall. In this respect, studies on transgenic tobacco plants

overexpressing a xyloglucan endotransglucosylase/hydrolase gene support this statement. In those plants, xyloglucans accumulated at lower levels, which prevents plants from binding Cd. This suggests that xyloglucans could be relevant binding sites for Cd and are involved in Cd tolerance; conversely, plants with lower levels of xyloglucans accumulate lower amounts of this heavy metal (Han et al., 2014).

Another significant example is represented by the addition of Si. Si considerably reduced the net influx of Cd by likely producing Si-cell wall complexes that efficiently adsorb Cd thus preventing its entry in the cytoplasm (Liu et al., 2013). In such a way, the cell wall can be potentiated, thus becoming a more effective barrier against Cd. The ameliorating activity of Si is dependent on its concentration, as well as on the concentration of Cd, and Si has a major effect on Cd influx by modifying the cell wall texture. However, Si does not affect intracellular processes like photosynthesis (Lukacová et al., 2013). In mangrove plants, Si is known to alleviate the effects of Cd pollution by restricting Cd to the cell wall of roots and reducing the concentration of Cd in the symplast. Therefore Si enhances the capacity of cell walls to restrain Cd, thereby limiting its diffusion (Zhang et al., 2014a).

The possibility of improving the hyperaccumulator capacity of plants may not be simple, as the number of genes involved may be huge. A preliminary assay in the hyperaccumulator *S. alfredii* revealed that plants under Cd stress respond by modifying the expression level of more than 100 genes responsible for different functions (including cell wall modification; Gao et al., 2013). Comparable analyses by using suppression subtractive hybridization screens showed that the activity of several genes is important in making plants of *Salix caprea* tolerant to Cd pollution. Some of the investigated genes belong to the family of cell wall modifying enzymes (Konlechner et al., 2013). This finding opens the way to the identification of potential genes (and gene products) capable of raising the tolerance capacity of plants against Cd.

The study of hyperaccumulators' cell walls will lead to the identification of those functional chemical groups capable of absorbing high amounts of Cd thereby preventing its entry in the cytoplasm. In addition to understanding the chemistry of this process, it is also necessary to identify the cellular mechanisms allowing the addition of such chemical groups in the cell wall (e.g., specific enzymes that modify the chemical structure of cell wall components). It follows that the development of hyperaccumulator plants requires a high level of knowledge. Ideally, the most convenient result would be to combine the hyperaccumulating capacity of non-crop plants (in which this process has been widely studied) with the economically advantageous crop plants. By transferring the genetic information at the base of Cd resistance into crop plants, it will be possible to design improved plants capable at the same time to detoxify the soil and to maintain the soil productivity rate (Wu et al., 2010).

PLANT BIOTECHNOLOGY AND Cd TOXICITY: WHAT SHOULD BE CONSIDERED?

The examples listed previously show that plants are equipped with different mechanisms to cope with the toxicity induced by heavy metals (Cd, specifically). Studying the variety of responses will allow us to identify specific mechanisms used to tolerate large

doses of pollutant. Those mechanisms (based either on chelating proteins, on general biochemical responses, or on sequestering in the vacuole/cell wall) may not necessarily be present in plants of economic interest and may concern either plant products or plant waste. For example, an interesting applicative aspect concerning the detoxifying ability of cell walls is the use of waste vegetable components rich in pectins as biosorbent of heavy metals (Wan Ngah and Hanafiah, 2008). The knowledge of these biosorbent mechanisms can be used to develop plants capable of sequestering Cd (hyperaccumulators; Douchiche et al., 2010). Pectins can also find applications outside the plant context. Fruits rich in pectin can absorb significant amounts of heavy metals (Schiewer and Patil, 2008). Such technologies can be extended to pectin gels extracted from commercial plants, such as sugar beet. Such studies, in addition to demonstrating that pectin-enriched matrices can actually be applied for Cd removal, confirm that Cd binds pectin matrices by removing calcium ions (Mata et al., 2009, 2010). Other food wastes have been proposed in this regard, for example citrus peels (Schiewer and Iqbal, 2010).

Several studies have shown altered Cd accumulation in plants engineered to express genes involved in heavy metal uptake, translocation or chelation. For example, as previously mentioned, tobacco plants expressing the Zn and Cd translocator *HMA4* from *A. thaliana*, which controls translocation to the shoot (Wong and Cobbett, 2009), show decreased Cd accumulation, but also deep transcriptional remodeling leading to increased lignification (Siemianowski et al., 2014). Restricting the examples to the cell wall, which is the topic of this review, the upregulation of a peroxidase, O-methyltransferase and hydroxycinnamoyl reductase leads to an increased deposition of lignin in a specific layer between the root epidermis and the first cortical layer, which consequently blocks Cd apoplastic movement towards the stele (Siemianowski et al., 2014). These transgenic tobacco plants exposed to Cd, however, show an enhanced Fe and Zn deficiency status linked to the overexpression of two metal uptake genes, *ZIP1* and *IRT1* (Siemianowski et al., 2014). These results show how challenging biotechnological prospects for improving heavy metal tolerance in plants are, as plants overexpressing a specific gene might show unfavorable/unexpected features linked to the response of the host to the transgene.

The insertion of new genes whose products change the chemical structure of cell wall might be a way to engineer plants that are more Cd-resistant. An alternative strategy would be to introduce specific mutations in genes already present. Mutation in one gene coding for a specific cellulose synthase subunit significantly alters the content of cellulose and decreases the cell wall thickness in rice plants. Although there were no clear phenotypical differences in comparison to control plants, mutant plants showed a significant decrease in the translocation rate of Cd along xylem vessels, which appeared as abnormal in shape. The new morphological structure of xylem, although not impairing the total shape of plants, determines a consistent reduction in the transport of Cd to leaves, thereby making rice plants more tolerant to Cd pollution (Song et al., 2013). Together with other studies, this finding highlighted that the translocation of Cd from roots to shoots is likely more critical than the mere uptake of Cd at the root level (Xin et al., 2013). Therefore, the accumulation of Cd

in the leaves seems more closely related to the translocation rate than to Cd uptake and reinforces the necessity of studying how Cd is translocated through the xylem and the factors that can be modified/regulated to delay such process.

The presence of the so-called “metal cross-homeostasis” has been identified as a primary factor influencing the phenotype of plants with a modified expression of genes involved in metal homeostasis (Antosiewicz et al., 2014). This cross-homeostasis is primarily due to the cross-talk existing between the homeostatic mechanisms of different metals: for example the concentration of different metals regulates the same players of the transcriptional wiring, the same transporters and chelators (Antosiewicz et al., 2014).

Approaches in biotechnology represent the most effective way to engineer heavy metal accumulation in plants. In particular, three elements are relevant to devise effective engineering approaches for heavy metal uptake/accumulation in plants: (1) the choice of specific promoters, for instance tissue-specific promoters (Antosiewicz et al., 2014), which restrict eventual modifications to a specific tissue or organ; (2) integrative studies aimed at deciphering the response of plants to a specific metal at a transcriptional, protein and metabolic level; (3) the choice of the host to express a specific transgene. This last feature is very relevant, as fast-growing plants with high biomass yield should be preferred. As an example, fiber crops like hemp (*Cannabis sativa* L.) represent very good candidates for engineering approaches. Hemp has a deep root system, a fast growth rate and is known to require less water than other crops (like cotton): all these features make it extremely interesting for biotechnology. Moreover the hemp genome has been sequenced (van Bakel et al., 2011), a feature which can greatly favor the design of constructs, e.g., for synthetic biology. Last but not least, the tissues of hemp stems show a great diversity of chemical composition: the core is lignified, while the cortex is rich in cellulosic bast fibers. This heterogeneity in cell wall composition offers a variety of functional groups potentially useful for heavy metal sequestration in hemp stem tissues.

FUTURE OUTLOOK: WHAT COMES NEXT?

To conclude this review, a list of points that need further study is hereafter proposed to achieve a more comprehensive knowledge of the mechanisms involved in Cd toxicity in plants and to devise new strategies for the biotechnological improvement of plant tolerance to toxic heavy metals.

- (1) Studying how early- and later-diverging plants progressively adapted and responded to Cd will reveal a particular mechanism or even a combination of molecular mechanisms, which could be used to improve the response of crop plants grown on contaminated soils. In this context the coupling of -omics sciences (transcriptomics, proteomics, metabolomics) and advanced imaging techniques (namely mass spectrometry imaging, like Secondary Ion Mass Spectrometry, SIMS) can generate a wealth of information covering different levels of biological complexity.
- (2) Considering the chemical variety of cell walls and the potential diversity in polysaccharide and protein composition,

there might be several possibilities for the cell wall to interfere with the toxic activity of Cd. Therefore, the study of cell walls (in terms of assembly and composition) in hyperaccumulators can provide important information on how to improve the survival or the productivity of plants growing on Cd-contaminated soils. In this context, it is important to identify the natural targets of Cd at the cell wall level and to figure out which of these targets can be modified to prevent the negative effects of Cd. One approach to develop this point is the functional study of those targets via generation of mutants/overexpressors. The purpose of this approach is to generate plants that either cannot absorb Cd or can sequester Cd in the cell wall thereby reducing or avoiding detrimental effects in terms of growth or productivity.

- (3) Studies on the effects induced at the cell wall-level by the application of beneficial mineral nutrients deserve further attention. In particular the role of Si and P in alleviating the toxic effects of Cd should be analyzed via an integrative biology approach to identify the targets of this interaction.

ACKNOWLEDGMENTS

GG, KS, and JFH thank partial financial support obtained through the National Research Fund, FNR Project CANCAN C13/SR/57742025774202 and project CADWALL INTER/FWO/12/14. The authors also thank Lucien Hoffmann for critical reading. The conceiving of this Review is linked to activities conducted within the COST FP1105 and COST FP1106 actions.

REFERENCES

Akhter, M. F., Omelon, C. R., Gordon, R. A., Moser, D., and Macfie, S. M. (2014). Localization and chemical speciation of cadmium in the roots of barley and lettuce. *Environ. Exp. Bot.* 100, 10–19. doi: 10.1016/j.envexpbot.2013.12.005

Anderson, C. M., Wagner, T. A., Perret, M., He, Z. H., He, D., and Kohorn, B. D. (2001). WAKs: cell wall-associated kinases linking the cytoplasm to the extracellular matrix. *Plant Mol. Biol.* 47, 197–206. doi: 10.1023/A:1010691701578

Antosiewicz, D. M., Barabasz, A., and Siemianowski, O. (2014). Phenotypic and molecular consequences of overexpression of metal-homeostasis genes. *Front. Plant Sci.* 5:80. doi: 10.3389/fpls.2014.00080

Astier, C., Gloaguen, V., and Faugeron, C. (2014). Phytoremediation of cadmium-contaminated soils by young douglas fir trees: effects of cadmium exposure on cell wall composition. *Int. J. Phytorem.* 16, 790–803. doi: 10.1080/15226514.2013.856849

Barberon, M., Dubeaux, G., Kolb, C., Isono, E., Zelazny, E., and Vert, G. (2014). Polarization of IRON-REGULATED TRANSPORTER 1 (IRT1) to the plant-soil interface plays crucial role in metal homeostasis. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8293–8298. doi: 10.1073/pnas.1402262111

Basile, A., Sorbo, S., Conte, B., Cardi, M., and Esposito, S. (2013). Ultrastructural changes and Heat Shock Proteins 70 induced by atmospheric pollution are similar to the effects observed under in vitro heavy metals stress in *Conocephalum conicum* (Marchantiiales–Bryophyta). *Environ. Pollut.* 182, 209–216. doi: 10.1016/j.envpol.2013.07.014

Basile, A., Sorbo, S., Pisani, T., Paoli, L., Munzi, S., and Loppi, S. (2012). Bioaccumulation and ultrastructural effects of Cd, Cu, Pb, and Zn in the moss *Scorpiurum circinatum* (Brid.) Fleisch. Loeske. *Environ. Pollut.* 166, 208–211. doi: 10.1016/j.envpol.2012.03.018

Benavides, M. P., Gallego, S. M., and Tomaro, M. L. (2005). Cadmium toxicity in plants. *Braz. J. Plant Physiol.* 17, 21–34. doi: 10.1590/S1677-04202005000100003

Bezrukova, M. V., Fatkhutdinova, R. A., Lubyanova, A. R., Murzabaev, A. R., Fedyaev, V. V., and Shakirova, F. M. (2011). Lectin involvement in the development of wheat tolerance to cadmium toxicity. *Russ. J. Plant Physiol.* 58, 1048–1054. doi: 10.1134/S1021443711060021

Blinda, A., Koch, B., Ramanjulu, S., and Dietz, K. J. (1997). De novo synthesis and accumulation of apoplastic proteins in leaves of heavy metal-

exposed barley seedlings. *Plant Cell Environ.* 20, 969–981. doi: 10.1111/j.1365-3040.1997.tb00674.x

Brill, E., van Thournout, M., White, R. G., Llewellyn, D., Campbell, P. M., Engelen, S., et al. (2011). A novel isoform of sucrose synthase is targeted to the cell wall during secondary cell wall synthesis in cotton fiber. *Plant Physiol.* 157, 40–54. doi: 10.1104/pp.111.178574

Cao, D., Zhang, H., Wang, Y., and Zheng, L. (2014). Accumulation and distribution characteristics of zinc and cadmium in the hyperaccumulator plant *Sedum plumbizincicola*. *Bull. Environ. Contam. Toxicol.* 93, 171–176. doi: 10.1007/s00128-014-1284-8

Carrier, P., Baryla, A., and Havaux, M. (2003). Cadmium distribution and microlocalization in oilseed rape (*Brassica napus*) after long-term growth on cadmium-contaminated soil. *Planta* 216, 939–950. doi: 10.1007/s00425-002-0947-6

Chaoui, A., Jarrar, B., and El Ferjani, E. (2004). Effects of cadmium and copper on peroxidase, NADH oxidase and IAA oxidase activities in cell wall, soluble and microsomal membrane fractions of pea roots. *J. Plant Physiol.* 161, 1225–1234. doi: 10.1016/j.jplph.2004.02.002

Clabeaux, B. L., Navarro, D. A., Aga, D. S., and Bisson, M. A. (2013). Combined effects of cadmium and zinc on growth, tolerance, and metal accumulation in *Chara australis* and enhanced phytoextraction using EDTA. *Ecotoxicol. Environ. Saf.* 98, 236–243. doi: 10.1016/j.ecoenv.2013.08.014

Cosgrove, D. J., and Jarvis, M. C. (2012). Comparative structure and biomechanics of plant primary and secondary cell walls. *Front. Plant Sci.* 3:204. doi: 10.3389/fpls.2012.00204

Cosio, C., De Santis, L., Frey, B., Diallo, S., and Keller, C. (2005). Distribution of cadmium in leaves of *Thlaspi caerulescens*. *J. Exp. Bot.* 56, 765–775. doi: 10.1093/jxb/eri062

Crowell, E. F., Gonreau, M., Stierhof, Y. D., Höfte, H., and Vernhettes, S. (2010). Regulated trafficking of cellulose synthases. *Curr. Opin. Plant Biol.* 13, 700–705. doi: 10.1016/j.pbi.2010.07.005

Daković, M., Kovacević, M., Andjus, P. R., and Bacić, G. (2008). On the mechanism of uranium binding to cell wall of *Chara fragilis*. *Eur. Biophys. J.* 37, 1111–1117. doi: 10.1007/s00249-008-0282-3

Djikanovic, D., Kalauzi, A., Jeremic, M., Xu, J., Micic, M., Whyte, J. D., et al. (2012). Interaction of the CdSe quantum dots with plant cell walls. *Colloids Surf. B Biointerfaces* 91, 41–47. doi: 10.1016/j.colsurfb.2011.10.032

Douchiche, O., Driouich, A., and Morvan, C. (2011). Impact of cadmium on early stages of flax fibre differentiation: ultrastructural aspects and pectic features of cell walls. *Plant Physiol. Biochem.* 49, 592–599. doi: 10.1016/j.jplph.2011.03.008

Douchiche, O., Rihouey, C., Schaumann, A., Driouich, A., and Morvan, C. (2007). Cadmium-induced alterations of the structural features of pectins in flax hypocotyl. *Planta* 225, 1301–1312. doi: 10.1007/s00425-006-0425-7

Douchiche, O., Soret-Morvan, O., Chaibi, W., Morvan, C., and Payne, F. (2010). Characteristics of cadmium tolerance in 'Hermes' flax seedlings: contribution of cell walls. *Chemosphere* 81, 1430–1436. doi: 10.1016/j.chemosphere.2010.09.011

Du, Y. P., Li, H. J., Yin, K. L., and Zhai, H. (2012). Cadmium accumulation, subcellular distribution, and chemical forms in *Vitis vinifera* cv. *Chardonnay* grapevine. *Chi. J. Appl. Ecol.* 23, 1607–1612.

Durand, T. C., Sergeant, K., Planchon, S., Carpin, S., Label, P., Morabito, D., et al. (2010). Acute metal stress in *Populus tremula* × *P. alba* (717-1B4 genotype): leaf and cambial proteome changes induced by cadmium²⁺. *Proteomics* 10, 349–368. doi: 10.1002/pmic.200900484

Engelsdorf, T., and Hamann, T. (2014). An update on receptor-like kinase involvement in the maintenance of plant cell wall integrity. *Ann. Bot.* 114, 1339–1347. doi: 10.1093/aob/mcu043

Evard, A., Sergeant, K., Printz, B., Guignard, C., Renaut, J., Campanella, B., et al. (2014a). A multiple-level study of metal tolerance in *Salix fragilis* and *Salix aurita* clones. *J. Proteom.* 101, 113–129. doi: 10.1104/pp.14.00207

Evard, A., Sergeant, K., Ferrandis, S., Printz, B., Renaut, J., Guignard, C., et al. (2014b). Physiological and proteomic responses of different willow clones (*Salix fragilis* × *alba*) exposed to dredged sediment contaminated by heavy metals. *Int. J. Phytoremed.* 16, 1148–1169. doi: 10.1080/15226514.2013.821448

Fan, J. L., Wei, X. Z., Wan, L. C., Zhang, L. Y., Zhao, X. Q., Liu, W. Z., et al. (2011). Disarrangement of actin filaments and Ca²⁺ gradient by CdCl₂ alters cell wall construction in *Arabidopsis thaliana* root hairs by inhibiting vesicular trafficking. *J. Plant Physiol.* 168, 1157–1167. doi: 10.1016/j.jplph.2011.01.031

Fernandes, J. C., García-Angulo, P., Goulao, L. F., Acebes, J. L., and Amâncio, S. (2013). Mineral stress affects the cell wall composition of grapevine (*Vitis vinifera* L.) callus. *Plant Sci.* 205–206, 111–120. doi: 10.1016/j.plantsci.2013.01.013

Fernández, R., Fernández-Fuego, D., Bertrand, A., and González, A. (2014). Strategies for Cd accumulation in *Dittrichia viscosa* (L.) Greuter: role of the cell wall, non-protein thiols and organic acids. *Plant Physiol. Biochem.* 78, 63–70. doi: 10.1016/j.plaphy.2014.02.021

Ferrari, S., Savatin, D. V., Sicilia, F., Gramegna, G., Cervone, F., and Lorenzo, G. D. (2013). Oligogalacturonides: plant damage-associated molecular patterns and regulators of growth and development. *Front. Plant Sci.* 4:49. doi: 10.3389/fpls.2013.00049

Fry, S. C., Nesselrode, B. H. W. A., Miller, J. G., and Mewburn, B. R. (2008). Mixed linkage (1–3,1–4)- β -D-glucan is a major hemicellulose of *Equisetum* (horsetail) cell walls. *New Phytol.* 179, 104–115. doi: 10.1111/j.1469-8137.2008.02435.x

Gao, J., Sun, L., Yang, X., and Liu, J. X. (2013). Transcriptomic analysis of cadmium stress response in the heavy metal hyperaccumulator *Sedum alfredii* Hance. *PLoS ONE* 8:e64643. doi: 10.1371/journal.pone.0064643

Ge, W., Jiao, Y. Q., Sun, B. L., Qin, R., Jiang, W. S., and Liu, D. H. (2012). Cadmium-mediated oxidative stress and ultrastructural changes in root cells of poplar cultivars. *S. Afr. J. Bot.* 83, 98–108. doi: 10.1016/j.sajb.2012.07.026

Gomes, P. I., and Asaeda, T. (2013). Phytoremediation of heavy metals by calcifying macro-algae (*Nettella pseudoflabellata*): implications of redox insensitive end products. *Chemosphere* 92, 1328–1334. doi: 10.1016/j.chemosphere.2013.05.043

Gu, C. S., Liu, L. Q., Zhao, Y. H., Deng, Y. M., Zhu, X. D., and Huang, S. Z. (2014). Overexpression of *Iris lactea* var. *chinensis* metallothionein lIMT2a enhances cadmium tolerance in *Arabidopsis thaliana*. *Ecotoxicol. Environ. Saf.* 105, 22–28. doi: 10.1016/j.ecoenv.2014.04.002

Guerrero, G., Sergeant, K., and Hausman, J. F. (2013). Integrated -omics: a powerful approach to understanding the heterogeneous lignification of fibre crops. *Int. J. Mol. Sci.* 14, 10958–10978. doi: 10.3390/ijms140610958

Guerrero, G., Hausman, J. F., and Cai, G. (2014a). No stress! Relax! Mechanisms governing growth and shape in plant cells. *Int. J. Mol. Sci.* 15, 5094–5114. doi: 10.3390/ijms15035094

Guerrero, G., Sergeant, K., and Hausman, J. F. (2014b). Wood biosynthesis and typologies: a molecular rhapsody. *Tree Physiol.* 34, 839–855. doi: 10.1093/treephys/tpu031

Hamann, T., and Denness, L. (2011). Cell wall integrity maintenance in plants: lessons to be learned from yeast? *Plant Sign. Behav.* 6, 1706–1709. doi: 10.4161/psb.6.11.17782

Han, Y. L., Yuan, H. Y., Huang, S. Z., Guo, Z., Xia, B., and Gu, J. (2007). Cadmium tolerance and accumulation by two species of *Iris*. *Ecotoxicology* 16, 557–563. doi: 10.1007/s10646-007-0162-0

Han, Y., Sa, G., Sun, J., Shen, Z., Zhao, R., Ding, M., et al. (2014). Overexpression of *Populus euphratica* xyloglucan endotransglucosylase/hydrolase gene confers enhanced cadmium tolerance by the restriction of root cadmium uptake in transgenic tobacco. *Environ. Exp. Bot.* 100, 74–83. doi: 10.1016/j.envexpbot.2013.12.021

Harada, E., Kim, J. A., Meyer, A. J., Hell, R., Clemens, S., and Choi, Y. E. (2010). Expression profiling of tobacco leaf trichomes identifies genes for biotic and abiotic stresses. *Plant Cell Physiol.* 51, 1627–1637. doi: 10.1093/pcp/pcq118

He, J. L., Ma, C. F., Ma, Y. L., Li, H., Kang, J. Q., Liu, T. X., et al. (2013). Cadmium tolerance in six poplar species. *Environ. Sci. Poll. Res.* 20, 163–174. doi: 10.1007/s11356-012-1008-8

Hradilova, J., Rehulkova, P., Rehulkova, H., Vrbova, M., Griga, M., and Brzobohatý, B. (2010). Comparative analysis of proteomic changes in contrasting flax cultivars upon cadmium exposure. *Electrophoresis* 31, 421–431. doi: 10.1002/elps.200900477

Humphrey, T. V., Bonetta, D. T., and Goring, D. R. (2007). Sentinels at the wall: cell wall receptors and sensors. *New Phytol.* 176, 7–21. doi: 10.1111/j.1469-8137.2007.02192.x

Kieffer, P., Dommes, J., Hoffmann, L., Hausman, J. F., and Renaut, J. (2008). Quantitative changes in protein expression of cadmium-exposed poplar plants. *Proteomics* 8, 2514–2530. doi: 10.1002/pmic.200701110

Kleczkowski, L. A., Kunz, S., and Wilczynska, M. (2010). Mechanisms of UDP-glucose synthesis in plants. *Crit. Rev. Plant Sci.* 29, 191–203. doi: 10.1080/07352689.2010.483578

Konlechner, C., Türktaş, M., Langer, I., Vaculík, M., Wenzel, W. W., Puschenreiter, M., et al. (2013). Expression of zinc and cadmium responsive genes in leaves of willow (*Salix caprea* L.) genotypes with different accumulation characteristics. *Environ. Pollut.* 178, 121–127. doi: 10.1016/j.envpol.2013.02.033

Konno, H., Nakato, T., Nakashima, S., and Katoh, K. (2005). *Lygodium japonicum* fern accumulates copper in the cell wall pectin. *J. Exp. Bot.* 56, 1923–1931. doi: 10.1093/jxb/eri187

Krzesłowska, M. (2011). The cell wall in plant cell response to trace metals: polysaccharide remodeling and its role in defense strategy. *Acta Physiol. Plant.* 33, 35–51. doi: 10.1007/s11738-010-0581-z

Küpper, H., Lombi, E., Zhao, F. J., and McGrath, S. P. (2000). Cellular compartmentation of cadmium and zinc in relation to other elements in the hyperaccumulator *Arabidopsis thaliana*. *Planta* 212, 75–84. doi: 10.1007/s004250000366

Kuramata, M., Masuya, S., Takahashi, Y., Kitagawa, E., Inoue, C., Ishikawa, S., et al. (2009). Novel cysteine-rich peptides from *Digitaria ciliaris* and *Oryza sativa* enhance tolerance to cadmium by limiting its cellular accumulation. *Plant Cell Physiol.* 50, 106–117. doi: 10.1093/pcp/pcn175

Lagriffoul, A., Mocquot, B., Mench, M., and Vangronsveld, J. (1998). Cadmium toxicity effects on growth, mineral and chlorophyll contents, and activities of stress related enzymes in young maize plants (*Zea mays* L.). *Plant Soil* 200, 241–250. doi: 10.1023/A:1004346905592

Lang, I., and Wernitznig, S. (2011). Sequestration at the cell wall and plasma membrane facilitates zinc tolerance in the moss *Pohlia drummondii*. *Environ. Exp. Bot.* 74, 186–193. doi: 10.1016/j.envexpbot.2011.05.018

Liu, D., Jiang, W., and Gao, X. (2003). Effects of cadmium on root growth, cell division and nucleoli in root tip cells of garlic. *Biol. Plantar.* 47, 79–83. doi: 10.1023/A:1027384932338

Liu, D., and Kottke, I. (2004). Subcellular localization of cadmium in the root cells of *Allium cepa* by electron energy loss spectroscopy and cytochemistry. *J. Biosci.* 29, 329–335. doi: 10.1007/BF02702615

Liu, J., Ma, J., He, C., Li, X., Zhang, W., Xu, F., et al. (2013). Inhibition of cadmium ion uptake in rice (*Oryza sativa*) cells by a wall-bound form of silicon. *New Phytol.* 200, 691–699. doi: 10.1111/nph.12494

Loeffler, S., Hochberger, A., Grilli, E., Gekeler, W., Winnacker, E. L., and Zenk, M. H. (1989). Termination of the phytochelatin synthase reactions through sequestration of heavy metals by reaction products. *FEBS Lett.* 258, 42–46. doi: 10.1016/0014-5793(89)81611-4

Lozano-Rodríguez, E., Hernández, L. E., Bonay, P., and Carpena-Ruiz, R. O. (1997). Distribution of cadmium in shoot and root tissues of maize and pea plants: physiological disturbances. *J. Exp. Bot.* 48, 123–128. doi: 10.1093/jxb/48.1.123

Lukáčová, Z., Švubová, R., Kohanová, J., and Lux, A. (2013). Silicon mitigates the Cd toxicity in maize in relation to cadmium translocation, cell distribution, antioxidant enzymes stimulation and enhanced endodermal apoplasmic barrier development. *Plant Growth Regul.* 70, 89–103. doi: 10.1007/s10725-012-9781-4

Lux, A. (2010). Does diversity in root structure affect the diversity in cadmium uptake by plants? Opinion paper. *Agrochimica* 54, 342–352.

Lux, A., Vaculík, M., Martinka, M., Liskova, D., Kulkarni, M. G., Stirk, W. A., et al. (2011). Cadmium induces hypodermal periderm formation in the roots of the monocotyledonous medicinal plant *Merwilla plumbea*. *Ann. Bot.* 107, 285–292. doi: 10.1093/aob/mcq240

Malovikova, A. and Kohn, R. (1982). Binding of cadmium cations to pectins. *Collec. Czech. Chem. Commun.* 47, 702–708.

Mata, Y. N., Blázquez, M. L., Ballester, A., González, F., and Muñoz, J. A. (2009). Sugar-beet pulp pectin gels as biosorbent for heavy metals: preparation and determination of biosorption and desorption characteristics. *Chem. Eng. J.* 150, 289–301. doi: 10.1016/j.cej.2009.01.001

Mata, Y. N., Blázquez, M. L., Ballester, A., González, F., and Muñoz, J. A. (2010). Studies on sorption, desorption, regeneration and reuse of sugar-beet pulp pectin gels for heavy metal removal. *J. Haz. Mat.* 178, 243–248. doi: 10.1016/j.jhazmat.2010.01.069

Mellerowicz, E. J., and Gorshkova, T. A. (2012). Tensional stress generation in gelatinous fibres: a review and possible mechanism based on cell-wall structure and composition. *J. Exp. Bot.* 63, 551–565. doi: 10.1093/jxb/err339

Mikkelsen, M. D., Harholt, J., Ulvskov, P., Johansen, I. E., Fangel, J. U., Dobkin, M. S., et al. (2014). Evidence for land plant cell wall biosynthetic mechanisms in charophyte green algae. *Ann. Bot.* 114, 1217–1236. doi: 10.1093/aob/mcu171

Mittler, R. (2006). Abiotic stress, the field environment and stress combination. *Trends Plant Sci.* 11, 15–19. doi: 10.1016/j.tplants.2005.11.002

Moral, R., Gomez, I., Pedreno, J. N., and Mataix, J. (1994). Effects of cadmium on nutrient distribution, yield, and growth of tomato grown in soilless culture. *J. Plant Nutr.* 17, 953–962. doi: 10.1080/01904169409364780

Moussa, H. R., and El-Gamal, S. M. (2010). Role of salicylic acid in regulation of cadmium toxicity in wheat (*Triticum aestivum* L.). *J. Plant Nutr.* 33, 1460–1471. doi: 10.1080/01904167.2010.489984

Nedelkoska, T. V., and Doran, P. M. (2000). Hyperaccumulation of cadmium by hairy roots of *Thlaspi caerulescens*. *Biotech. Bioeng.* 67, 607–615. doi: 10.1002/(SICI)1097-0290(20000305)67:5<607::AID-BIT11>3.0.CO;2-3

Pan, X., Liu, F. C., Chai, M. W., Liu, L. M., Morita, S., and Shi, F. C. (2012). Accumulation, translocation, and subcellular distribution of cadmium in *Spartina alterniflora*. *Chi. J. Ecol.* 31, 526–531.

Paynel, F., Schaumann, A., Arkoun, M., Douchiche, O., and Morvan, C. (2009). Temporal regulation of cell-wall pectin methylesterase and peroxidase isoforms in cadmium-treated flax hypocotyl. *Ann. Bot.* 104, 1363–1372. doi: 10.1093/aob/mcp254

Piršelová, B., Kuna, R., Libantová, J., Moravčíková, J., and Matušková, I. (2011). Biochemical and physiological comparison of heavy metal-triggered defense responses in the monocot maize and dicot soybean roots. *Mol. Biol. Rep.* 38, 3437–3446. doi: 10.1007/s11033-010-0453-z

Piršelová, B., Mistríková, V., Libantová, J., Moravčíková, J., and Matušková, I. (2012). Study on metal-triggered callose deposition in roots of maize and soybean. *Biologia* 67, 698–705. doi: 10.2478/s11756-012-0051-8

Podazza, C., Rosa, M., González, J. A., Hilal, M., and Prado, F. E. (2006). Cadmium induces changes in sucrose partitioning, invertase activities, and membrane functionality in roots of rangpur lime (*Citrus limonia* L. Osbeck). *Plant Biol.* 8, 706–714. doi: 10.1055/s-2006-924171

Popper, Z. A., and Fry, S. C. (2003). Primary cell wall composition of bryophytes and charophytes. *Ann. Bot.* 91, 1–12. doi: 10.1093/aob/mcg013

Printz, B., Sergeant, K., Lutts, S., Guignard, C., Renaut, J., and Hausman, J. F. (2013a). From tolerance to acute metabolic deregulation: contribution of proteomics to dig into the molecular response of alder species under a polymetallic exposure. *J. Proteome Res.* 12, 5160–5179. doi: 10.1021/pr400590d

Printz, B., Sergeant, K., Guignard, C., Renaut, J., and Hausman, J. F. (2013b). Physiological and proteome study of sunflowers exposed to a polymetallic constraint. *Proteomics* 13, 1993–2015. doi: 10.1002/pmic.201200400

Przedpelska-Wasowicz, E. M., and Wierzbicka, M. (2011). Gating of aquaporins by heavy metals in *Allium cepa* L. epidermal cells. *Protoplasma* 248, 663–671. doi: 10.1007/s00709-010-0222-9

Qadir, S., Jamshieed, S., Rasool, S., Ashraf, M., Akram, N. A., and Ahmad, P. (2014). "Modulation of plant growth and metabolism in cadmium-enriched environments," in *Reviews of Environmental Contamination and Toxicology*, ed. D. M. Whitacre (Cham: Springer International Publishing), 51–88. doi: 10.1007/978-3-319-03777-6_4

Quiroga, M., de Forchetti, S. M., Taleisnik, E., and Tigier, H. A. (2001). Tomato root peroxidase isoenzymes: kinetic studies of the coniferyl alcohol peroxidase activity, immunological properties and role in response to salt stress. *J. Plant Physiol.* 158, 1007–1013. doi: 10.1078/0176-1617-00304

Rascio, N., and Navari-Izzo, F. (2011). Heavy metal hyperaccumulating plants: how and why do they do it? And what makes them so interesting? *Plant Sci.* 180, 169–181. doi: 10.1016/j.plantsci.2010.08.016

Redjala, T., Sterckeman, T., and Morel, J. L. (2009). Cadmium uptake by roots: contribution of apoplast and of high- and low-affinity membrane transport systems. *Environ. Exp. Bot.* 67, 235–242. doi: 10.1016/j.envexpbot.2009.05.012

Rodríguez-Serrano, M., Romero-Puertas, M. C., Pazmiño, D. M., Testillano, P. S., Risueño, M. C., del Río, L. A., et al. (2009). Cellular response of pea plants to cadmium toxicity: crosstalk between reactive oxygen species, nitric oxide and calcium. *Plant Physiol.* 150, 229–243. doi: 10.1104/pp.108.131524

Sandalio, L. M., Dalurzo, H. C., Gomez, M., Romero-Puertas, M. C., and Del Río, L. A. (2001). Cadmium-induced changes in the growth and oxidative metabolism of pea plants. *J. Exp. Bot.* 52, 2115–2126. doi: 10.1093/jexbot/52.364.2115

Sanità Di Toppi, L., and Gabbielli, R. (1999). Response to cadmium in higher plants. *Environ. Exp. Bot.* 41, 105–130. doi: 10.1016/S0098-8472(98)00058-6

Sanità Di Toppi, L., Marabottini, R., Vattuone, Z., Musetti, R., Favali, M. A., Sorgonà, A., et al. (2005). Cell wall immobilisation and antioxidant status of *Xanthoria parietina* thalli exposed to cadmium. *Funct. Plant Biol.* 32, 611–618. doi: 10.1071/FP04237

Sarkar, P., Bosneaga, E., and Auer, M. (2009). Plant cell walls throughout evolution: towards a molecular understanding of their design principles. *J. Exp. Bot.* 60, 3615–3635. doi: 10.1093/jxb/erp245

Sawidis, T. (2008). Effect of cadmium on pollen germination and tube growth in *Lilium longiflorum* and *Nicotiana tabacum*. *Protoplasma* 233, 95–106. doi: 10.1007/s00709-008-0306-y

Sawidis, T., and Reiss, H. D. (1995). Effects of heavy metals on pollen tube growth and ultrastructure. *Protoplasma* 185, 113–122. doi: 10.1007/BF01272851

Schiewer, S., and Patil, S. B. (2008). Pectin-rich fruit wastes as biosorbents for heavy metal removal: equilibrium and kinetics. *Bioresour. Technol.* 99, 1896–1903. doi: 10.1016/j.biortech.2007.03.060

Schiewer, S., and Iqbal, M. (2010). The role of pectin in Cd binding by orange peel biosorbents: a comparison of peels, depectinated peels and pectic acid. *J. Hazard. Mater.* 177, 899–907. doi: 10.1016/j.jhazmat.2010.01.001

Schützendübel, A., Schwarz, P., Teichmann, T., Gross, K., Langenfeld-Heyser, R., Godbold, D. L., et al. (2001). Cadmium-induced changes in antioxidative systems, hydrogen peroxide content, and differentiation in Scots pine roots. *Plant Physiol.* 127, 887–898. doi: 10.1104/pp.010318

Sergeant, K., Kieffer, P., Dommes, J., Hausmann, J. F., and Renaut, J. (2014). Proteomic changes in leaves of poplar exposed to both cadmium and low-temperature. *Environ. Exp. Bot.* 106, 112–123. doi: 10.1016/j.envexpbot.2014.01.007

Sergio, E., Cobianchi, R. C., Sorbo, S., Conte, B., and Basile, A. (2007). Ultrastructural alterations and HSP 70 induction in *Elodea canadensis* Michx. exposed to heavy metals. *Caryologia* 60, 115–120. doi: 10.1080/00087114.2007.10589557

Shah, K., Singh, P., and Nahakpam, S. (2013). Effect of cadmium uptake and heat stress on root ultrastructure, membrane damage and antioxidative response in rice seedlings. *J. Plant Biochem. Biotech.* 22, 103–112. doi: 10.1007/s13562-012-0116-3

Siebers, N., Siangliw, M., and Tongcumpou, C. (2013). Cadmium uptake and subcellular distribution in rice plants as affected by phosphorus: soil and hydroponic experiments. *J. Soil Sci. Plant Nutr.* 13, 833–844. doi: 10.4067/S0718-95162013005000066

Siemianowski, O., Barabasz, A., Kendziora, M., Ruszczynska, A., Bulska, E., Williams, L. E., et al. (2014). HMA4 expression in tobacco reduces Cd accumulation due to the induction of the apoplastic barrier. *J. Exp. Bot.* 65, 1125–1139. doi: 10.1093/jxb/ert471

Song, W. Y., Choi, K. S., De Alexis, A., Martinoia, E., and Lee, Y. (2011). *Brassica juncea* plant cadmium resistance 1 protein (BjPCR1) facilitates the radial transport of calcium in the root. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19808–19813. doi: 10.1073/pnas.1104905108

Song, X. Q., Liu, L. F., Jiang, Y. J., Zhang, B. C., Gao, Y. P., Liu, X. L., et al. (2013). Disruption of secondary wall cellulose biosynthesis alters cadmium translocation and tolerance in rice plants. *Mol. Plant* 6, 768–780. doi: 10.1093/mp/sst025

Sørensen, I., Pettolino, F. A., Bacic, A., Ralph, J., Lu, F., O'Neill, M. A., et al. (2011). The charophyan green algae provide insights into the early origins of plant cell walls. *Plant J.* 68, 201–211. doi: 10.1111/j.1365-313X.2011.04686.x

Soudkova, P., Katrusakova, A., Sedlackova, L., Petrovova, S., Koc, V., Marsik, P., et al. (2010). Effect of heavy metals on inhibition of root elongation in 23 cultivars of flax (*Linum usitatissimum* L.). *Arch. Environ. Contam. Toxicol.* 59, 194–203. doi: 10.1007/s00244-010-9480-y

Souza, V. L., De Almeida, A. A. F., Lima, S. G. C., De, M. C., Da, C. S., Mangabeira, P. A. O., et al. (2011). Morphophysiological responses and programmed cell death induced by cadmium in *Genipa americana* L. (Rubiaceae). *Biometals* 24, 59–71. doi: 10.1007/s10534-010-9374-5

Sun, J., Cui, J., Luo, C., Gao, L., Chen, Y., and Shen, Z. (2013). Contribution of cell walls, nonprotein thiols, and organic acids to cadmium resistance in two cabbage varieties. *Arch. Environ. Contam. Toxicol.* 64, 243–252. doi: 10.1007/s00244-012-9824-x

Suzuki, N. (2005). Alleviation by calcium of cadmium-induced root growth inhibition in *Arabidopsis* seedlings. *Plant Biotech.* 22, 19–25. doi: 10.5511/plantbiotechnology.22.19

Uegsgeger, A., Schmutz, D., and Brunold, C. (1990). Regulation of glutathione synthesis by cadmium in *Pisum sativum* L. *Plant Physiol.* 93, 1579–1584. doi: 10.1104/pp.93.4.1579

Ueki, S., and Citovsky, V. (2005). Identification of an interactor of cadmium ion-induced glycine-rich protein involved in regulation of callose levels in plant vasculature. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12089–12094. doi: 10.1073/pnas.0505927102

Uraguchi, S., Kiyono, M., Sakamoto, T., Watanabe, I., and Kuno, K. (2009). Contributions of apoplastic cadmium accumulation, antioxidative enzymes and induction of phytochelatins in cadmium tolerance of the cadmium-accumulating cultivar of black oat (*Avena strigosa* Schreb.). *Planta* 230, 267–276. doi: 10.1007/s00425-009-0939-x

Vallarino, J. G., and Osorio, S. (2012). Signaling role of oligogalacturonides derived during cell wall degradation. *Plant Sign. Behav.* 7, 1447–1449. doi: 10.4161/psb.21779

van Bakel, H., Stout, J. M., Cote, A. G., Tallon, C. M., Sharpe, A. G., Hughes, T. R., et al. (2011). The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol.* 12, R102. doi: 10.1186/gb-2011-12-10-r102

Vázquez, S., Goldsborough, P., and Carpenea, R. O. (2006). Assessing the relative contributions of phytochelatins and the cell wall to cadmium resistance in white lupin. *Physiol. Plant.* 128, 487–495. doi: 10.1111/j.1399-3054.2006.00764.x

Verma, K., Shekhawat, G. S., Sharma, A., Mehta, S. K., and Sharma, V. (2008). Cadmium induced oxidative stress and changes in soluble and ionically bound cell wall peroxidase activities in roots of seedling and 3-4 leaf stage plants of *Brassica juncea* (L.) czern. *Plant Cell Rep.* 27, 1261–1269. doi: 10.1007/s00299-008-0552-7

Wan, L., and Zhang, H. (2012). Cadmium toxicity effects on cytoskeleton, vesicular trafficking and cell wall construction. *Plant Sign. Behav.* 7, 345–348. doi: 10.4161/psb.18992

Wan Ngah, W. S., and Hanafiah, M. A. K. M. (2008). Removal of heavy metal ions from wastewater by chemically modified plant wastes as adsorbents: a review. *Bioresour. Technol.* 99, 3935–3948. doi: 10.1016/j.biortech.2007.06.011

Wang, X., Gao, Y., Feng, Y., Li, X., Wei, Q., and Sheng, X. (2014). Cadmium stress disrupts the endomembrane organelles and endocytosis during *Picea wilsonii* pollen germination and tube growth. *PLoS ONE* 9:e94721. doi: 10.1371/journal.pone.0094721

Wang, Y., Huang, J., and Gao, Y. (2012). Arbuscular mycorrhizal colonization alters subcellular distribution and chemical forms of cadmium in *Medicago sativa* L. and resists cadmium toxicity. *PLoS ONE* 7:e48669. doi: 10.1371/journal.pone.0048669

Webster, E. A., and Gadd, G. M. (1996). Cadmium replaces calcium in the cell wall of *Ulva lactuca*. *Biometals* 9, 241–244. doi: 10.1007/BF00817922

Weng, B., Xie, X., Weiss, D. J., Liu, J., Lu, H., and Yan, C. (2012). *Kandelia obovata* (S., L.) Yong tolerance mechanisms to cadmium: subcellular distribution, chemical forms and thiol pools. *Mar. Pollut. Bull.* 64, 2453–2460. doi: 10.1016/j.marpolbul.2012.07.047

Wierzbicka, M. H., Przedpelska, E., Ruzik, R., Ouerdane, L., Polec-Pawlak, K., Jarosz, M., et al. (2007). Comparison of the toxicity and distribution of cadmium and lead in plant cells. *Protoplasma* 231, 99–111. doi: 10.1007/s00709-006-0227-6

Wolf, S., Hematy, K., and Hofte, H. (2012). Growth control and cell wall signaling in plants. *Ann. Rev. Plant Biol.* 63, 381–407. doi: 10.1146/annurev-aplant-042811-105449

Wolff, G., Pereira, G. C., Castro, E. M., Louzada, J., and Coelho, F. F. (2012). The use of *Salvinia auriculata* as a bioindicator in aquatic ecosystems: biomass and structure dependent on the cadmium concentration. *Braz. J. Biol.* 72, 71–77. doi: 10.1590/S1519-69842012000100009

Wong, C. K. E., and Cobbett, C. S. (2009). HMA P-type ATPases are the major mechanism for root-to-shoot Cd translocation in *Arabidopsis thaliana*. *New Phytol.* 181, 71–78. doi: 10.1111/j.1469-8137.2008.02638.x

Wu, G., Kang, H., Zhang, X., Shao, H., Chu, L., and Ruan, C. (2010). A critical review on the bio-removal of hazardous heavy metals from contaminated soils: issues, progress, eco-environmental concerns and opportunities. *J. Hazard. Mater.* 174, 1–8. doi: 10.1016/j.jhazmat.2009.09.113

Xin, J., Huang, B., Yang, J., Yang, Z., Yuan, J., and Mu, Y. (2013). Role of roots in cadmium accumulation of two water spinach cultivars: reciprocal grafting and histochemical experiments. *Plant Soil* 366, 425–432. doi: 10.1007/s11104-012-1439-5

Xing, J. P., Jiang, R. F., Ueno, D., Ma, J. F., Schat, H., McGrath, S. P., et al. (2008). Variation in root-to-shoot translocation of cadmium and zinc among different

accessions of the hyperaccumulators *Thlaspi caerulescens* and *Thlaspi praecox*. *New Phytol.* 178, 315–325. doi: 10.1111/j.1469-8137.2008.02376.x

Xiong, J., An, L., Lu, H., and Zhu, C. (2009). Exogenous nitric oxide enhances cadmium tolerance of rice by increasing pectin and hemicellulose contents in root cell wall. *Planta* 230, 755–765. doi: 10.1007/s00425-009-0984-5

Xiong, J., Fu, G., Tao, L., and Zhu, C. (2010). Roles of nitric oxide in alleviating heavy metal toxicity in plants. *Arch. Biochem. Biophys.* 497, 13–20. doi: 10.1016/j.abb.2010.02.014

Xu, J., Jia, R., Shi, G. X., Tian, X. L., Yang, H. Y., Xu, X. Y., et al. (2012a). Subcellular distribution and phytotoxicity of cadmium in *Alternanthera philoxeroides* leaves. *Chi. J. Appl. Ecol.* 23, 1070–1076.

Xu, J., Sun, J., Du, L., and Liu, X. (2012b). Comparative transcriptome analysis of cadmium responses in *Solanum nigrum* and *Solanum torvum*. *New Phytol.* 196, 110–124. doi: 10.1111/j.1469-8137.2012.04235.x

Xu, W., Shi, W., Yan, F., Zhang, B., and Liang, J. (2011). Mechanisms of cadmium detoxification in cattail (*Typha angustifolia* L.). *Aquat. Bot.* 94, 37–43. doi: 10.1016/j.aquabot.2010.11.002

Yamaguchi, N., Mori, S., Baba, K., Kaburagi-Yada, S., Arao, T., Kitajima, N., et al. (2011). Cadmium distribution in the root tissues of solanaceous plants with contrasting root-to-shoot Cd translocation efficiencies. *Environ. Exp. Bot.* 71, 198–206. doi: 10.1016/j.envexpbot.2010.12.002

Yang, J., and Pan, X. (2013). Root exudates from sunflower (*Helianthus annuus* L.) show a strong adsorption ability toward Cd(II). *J. Plant Interact.* 8, 263–270. doi: 10.1080/17429145.2012.737030

Zhang, J., and Shu, W. S. (2006). Mechanisms of heavy metal cadmium tolerance in plants. *J. Plant Physiol. Mol. Biol.* 32, 1–8.

Zhang, J., Wu, H., Wang, X., and Huang, W. (2010a). Subcellular distribution and chemical forms of cadmium in the cells of strawberry (*Fragaria × ananassa* Duch.). *J. Hortic. Sci. Biotechnol.* 85, 563–569.

Zhang, Z. C., Chen, B. X., and Qiu, B. S. (2010b). Phytochelatin synthesis plays a similar role in shoots of the cadmium hyperaccumulator *Sedum alfredii* as in non-resistant plants. *Plant Cell Environ.* 33, 1248–1255. doi: 10.1111/j.1365-3040.2010.02144.x

Zhang, Q., Yan, C., Liu, J., Lu, H., Duan, H., Du, J., et al. (2014a). Silicon alleviation of cadmium toxicity in mangrove (*Avicennia marina*) in relation to cadmium compartmentation. *J. Plant Growth Regul.* 33, 233–242. doi: 10.1007/s00344-013-9366-0

Zhang, S. J., Li, T. X., Huang, H. G., Zhang, X. Z., Yu, H. Y., Zheng, Z. C., et al. (2014b). Phytoremediation of cadmium using plant species of *Athyrium wardii* (Hook.). *Int. J. Environ. Sci. Tech.* 11, 757–764. doi: 10.1007/s13762-013-0384-z

Zhang, W., Lin, K., Zhou, J., Zhang, W., Liu, L., and Zhang, Q. (2014c). Cadmium accumulation, sub-cellular distribution and chemical forms in rice seedling in the presence of sulfur. *Environ. Toxicol. Pharmacol.* 37, 348–353. doi: 10.1016/j.etap.2013.12.006

Zhimin, Y., Shaojian, Z., and Aitang, H. U. (1999). Subcellular accumulation of cadmium in corn and wheat plants at different levels of phosphorus. *Pedosphere* 9, 169–176.

Zhou, B., Yao, W., Wang, S., Wang, X., and Jiang, T. (2014). The metallothionein gene, TaMT3, from *Tamarix androssowii* confers Cd²⁺ tolerance in tobacco. *Int. J. Mol. Sci.* 15, 10398–10409. doi: 10.3390/ijms150610398

Zhu, X. F., Lei, G. J., Jiang, T., Liu, Y., Li, G. X., and Zheng, S. J. (2012). Cell wall polysaccharides are involved in P-deficiency-induced Cd exclusion in *Arabidopsis thaliana*. *Planta* 236, 989–997. doi: 10.1007/s00425-012-1652-8

Conflict of Interest Statement: 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.

Received: 27 October 2014; accepted: 19 February 2015; published online: 13 March 2015.

Citation: Parrotta L, Guerrero G, Sergeant K, Cai G and Hausman J-F (2015) Target or barrier? The cell wall of early- and later-diverging plants vs cadmium toxicity: differences in the response mechanisms. *Front. Plant Sci.* 6:133. doi: 10.3389/fpls.2015.00133

This article was submitted to Plant Genetics and Genomics, a section of the journal *Frontiers in Plant Science*.

Copyright © 2015 Parrotta, Guerrero, Sergeant, Cai and Hausman. 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) or licensor 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.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read, for greatest visibility



COLLABORATIVE PEER-REVIEW

Designed to be rigorous – yet also collaborative, fair and constructive



FAST PUBLICATION

Average 85 days from submission to publication (across all journals)



COPYRIGHT TO AUTHORS

No limit to article distribution and re-use



TRANSPARENT

Editors and reviewers acknowledged by name on published articles



SUPPORT

By our Swiss-based editorial team



IMPACT METRICS

Advanced metrics track your article's impact



GLOBAL SPREAD

5'100'000+ monthly article views and downloads



LOOP RESEARCH NETWORK

Our network increases readership for your article

Frontiers

EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on

