Integrated biotechnological and artificial intelligence innovations for plant improvement

Haibo Wu[1], Man Luo[1], Yubin Liu[2,3], Jinghao Yang[2,3] and Yunpeng Cao[2,3*]

[1]School of Health and Nursing, Wuchang University of Technology, Wuhan, China, [2]Guangxi Colleges and Universities Key Laboratory for Cultivation and Utilization of Subtropical Forest Plantation, Guangxi Key Laboratory of Forest Ecology and Conservation, College of Forestry, Guangxi University, Nanning, China, [3]Key Laboratory of National Forestry and Grassland Administration on Cultivation of Fast-Growing Timber in Central South China, State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, College of Forestry, Guangxi University, Nanning, China

1 Introduction

Facing climate change and a growing population, Artificial Intelligence (AI) is leading a revolution in plant breeding, offering unprecedented opportunities to secure our future food supply. The global food system faces dual challenges: a rising population projected to reach 10 billion by 2050 is increasing demand for food, feed, and fiber (Van Dijk et al., 2021), while climate change simultaneously erodes limited arable land and threatens food system stability through extreme weather, soil salinization, and evolving pests and diseases (Hasegawa et al., 2021; Singh et al., 2023). Accelerating plant breeding to develop higher-yielding, better-quality, and more resilient crop varieties is a strategic imperative for global food security and sustainable development. Over the past century, plant breeding has progressed through several technology-driven leaps that have greatly increased agricultural productivity. These advances include early 20th-century hybrid breeding that leveraged heterosis in maize (Crow, 1998), the semi-dwarf varieties of the 1960s "Green Revolution" (Pingali, 2012), and the subsequent genetic engineering of insect-resistant and herbicide-tolerant crops (Lu et al., 2012; Li et al., 2025). However, these once-revolutionary methods are proving too slow for today's complex challenges. Traditional breeding is inefficient for improving multi-gene traits like yield and stress resistance, and even with accelerated techniques like marker-assisted selection, bringing a superior gene from discovery to the field can take over a decade.

To overcome the bottlenecks of speed and accuracy in plant breeding, the integration of modern biotechnology and AI is driving a technological revolution to reshape the entire process in response to rapid environmental changes. Transcending the limitations of natural variation and random mutagenesis, genome editing technologies like CRISPR-Cas provide the surgical precision to rewrite the genome, enabling the rapid "*de novo* domestication" of wild plants and vastly expanding the available genetic pool (Li et al., 2018). Simultaneously, AI, particularly deep learning, is deciphering life's complex code. It can mine massive multi-omics data for "elite alleles," predict protein structures with models like AlphaFold (Abramson et al., 2024), and even design novel functional proteins

(Kortemme, 2024). This is transforming plant science from a discipline of observation and experimentation into a precise science guided by prediction and design.

Overall, by combining AI's predictive design with biotechnology's precise implementation, a "Design-Build-Test-Learn" (DBTL) closed-loop accelerator is formed. In this cycle, AI models analyze vast genetic, phenotypic, and environmental data to propose optimal designs. Biotechnology tools like gene editing then build these designs into organisms. High-throughput phenotyping with drones and sensors rapidly tests the results, feeding new data back to the AI for continuous learning and optimization. This self-improving process promises to shorten breeding cycles from years to months and enable the improvement of previously unattainable complex traits. This article will highlight how AI, in synergy with gene editing and high-throughput phenotyping, is creating a 'DBTL' cycle that accelerates breeding cycles and enables the improvement of complex traits, while also exploring the associated challenges and opportunities.

2 AI-driven precision genomics and allele mining

Advances in sequencing technology have generated unprecedented amounts of multi-omics data, spanning genomics, transcriptomics, and proteomics. For example, researchers have sequenced the genomes of over ten species and accumulated vast transcriptome data from diverse tissues and developmental stages in the Rosaceae family (Cao et al., 2024, Cao et al., 2025; Jiang et al., 2025). However, the real challenge lies in identifying the key genes and genetic variations related to target traits from high-dimensional and complex data, since data itself is not knowledge. AI offers a transformative solution for complex genomic analysis. Recent studies demonstrate that advanced AI models, particularly those employing deep learning, can efficiently integrate large-scale multi-omics data to achieve outcomes that surpass the capabilities of traditional bioinformatics, such as predicting gene function, identifying crucial regulatory elements in non-coding regions, and deciphering intricate gene regulatory networks (Li et al., 2025). For example, Wu et al. developed the AutoGP AI-powered breeding platform by integrating multi-omics data from maize (genomics, transcriptomics, and metabolomics) to guide the selection of superior hybrid varieties and enhance breeding efficiency (Wu et al., 2025).

AI identifies subtle epistatic genetic combinations overlooked by traditional breeding, driving the shift from marker-assisted selection to predictive breeding. AI empowers us to simulate and predict the phenotypic performance of new plant varieties in specific environments within a computer before they are even created. This offers a significant advantage in plant breeding, drastically shortening breeding cycles and improving success rates (Li et al., 2025). However, a clear weakness lies in the fact that the performance of AI models heavily relies on large-scale, high-quality, and precisely annotated training datasets. For many orphan crops or traits that are not well-studied, the scarcity of effective data is a major bottleneck limiting the application of AI (Li et al., 2025).

3 Synergies of gene editing and AI-powered design

The advent of CRISPR-Cas technology has enabled precise, "surgical scalpel" level modifications to plant genomes (Kim et al., 2025; Li et al., 2025; Ruffolo et al., 2025). Now, AI is equipping this "scalpel" with a smart navigation system. From designing efficient guide RNAs (gRNAs) with low off-target rates to predicting potential genome-wide off-target effects of gene editing, AI tools are significantly improving the efficiency and safety of gene editing (Ruffolo et al., 2025). For example, by constructing a curated dataset of CRISPR operons, comprising Cas proteins, CRISPR arrays, trans-activating CRISPR RNA (tracrRNA), and Protospacer Adjacent Motifs (PAMs), Ruffolo et al. (2025) engineered novel proteins with diversity far exceeding natural variation and predicted structural effectiveness. In addition, AI has also begun to venture into the field of *de novo* design. By learning from vast amounts of protein structure and function data, AI models can design proteins with entirely new functions that do not exist in nature (Kortemme, 2024). For example, GENERA, a *de novo* design algorithm by Lamanna et al. (2023), integrates deep learning with a genetic algorithm to rapidly generate focused compound libraries that score higher than known ACE-2 binders (Lamanna et al., 2023). In plant improvement, this means we can design Rubisco enzymes with higher photosynthetic efficiency, metabolic enzymes capable of degrading novel herbicides, or disease-resistant proteins with broad-spectrum resistance to specific pathogens. The synergy of AI and gene editing is transforming plant breeding from modification to creation, enabling the programming of novel biological functions to develop breakthrough crop traits beyond the limitations of natural variation (Li et al., 2025). However, challenges remain: efficient and universally applicable gene editing delivery systems are still technically challenging in many plants. The scarcity of high-quality training data, especially when the data types are diverse, the standardization level is low, and biases exist, limits the predictive power of AI models. Furthermore, the lack of globally harmonized regulations surrounding gene-edited crops creates uncertainty for their practical application due to complex regulatory landscapes (Li et al., 2025).

4 High-throughput phenotyping and predictive phenomics

The phenotype, as the ultimate manifestation of the interaction between genes and the environment, serves as the direct basis for selection by breeders (Li et al., 2025). However, the labor-intensive and subjective nature of traditional phenotypic assessment methods constitutes a major bottleneck in modern breeding, impeding the efficient and accurate selection of desirable traits. High-throughput phenotyping platforms, composed of drones, ground robots, multispectral cameras, and various sensors, capture plant dynamic data throughout the growing season with unprecedented scale and precision. AI, particularly computer vision, is crucial for unlocking the value of this massive phenotyping data. AI algorithms can automatically and accurately measure thousands of phenotypic traits, such as plant height, leaf area index, disease spot area, fruit

count, and color. More importantly, by integrating continuous phenotyping data, environmental data, and genomic data, AI can construct dynamic genotype-phenotype (G2P) prediction models (Danilevicz et al., 2022; Sharma and Goel, 2025). The integration of high-throughput phenotyping and AI analysis is paving the way to bridge the "last mile" between genotype and phenotype (Sheikh et al., 2024). Breeders can leverage these models to predict the field performance of plants with specific genotypes across different years and regions, enabling more informed breeding decisions. The advantage lies in transforming phenotype identification from an "art" into a precise "science" (Sheikh et al., 2024). However, the challenges should not be underestimated: the initial investment in establishing high-throughput phenotyping platforms is substantial. Furthermore, integrating data from diverse sensors and spatiotemporal scales, while eliminating environmental noise, presents a significant computational challenge.

5 Discussion

The convergence of biotechnology and artificial intelligence is more than just a linear combination of technologies; it's creating a revolutionary "Design-Build-Test-Learn" closed-loop accelerator (Alemu et al., 2024). Data from each breeding cycle, from AI-driven gene discovery to high-throughput phenotyping, continuously refines the AI model. By continuously learning, AI improves its predictive accuracy to drive more efficient, forward-thinking designs. This shift from passive adaptation to active creation holds potential far surpassing any previous agricultural revolution (Li et al., 2025). Li et al. (2025) envision an "AI-assisted crop design" model serving as a "smart co-pilot" for future breeders by integrating vast multi-modal data to understand and predict the genetic basis of complex traits. More importantly, it can optimize the search across millions of potential genetic combinations based on breeding goals to output a comprehensive improvement strategy, detailing everything from specific gene editing targets to blueprints for novel protein designs (Wei et al., 2021; Li et al., 2025; Wu and Xie, 2025). This capability enables the systematic use of minor-effect alleles and breaks linkage drag, overcoming challenges that are difficult for traditional breeding methods to address. However, realizing this vision requires overcoming key bottlenecks, the most critical being data. The performance of AI models depends on the scale, quality, and diversity of training data, yet high-quality, standardized multi-omics and phenomics datasets are scarce, especially for "orphan crops" vital to regional food security (Wu and Xie, 2025; Yetgin, 2025). A global plant science data sharing consortium is therefore essential, demanding not only technological breakthroughs but also open collaboration between research institutions, governments, and businesses (Gibbs et al., 2019). Without this data foundation, even the most advanced algorithms are ineffective. Secondly, we must navigate complex ethical, regulatory, and social challenges. On this front, as Li et al. (2025) noted, regulatory policies on gene-edited crops are becoming more scientific and rational in many countries, creating favorable conditions for new technologies. In addition, cutting-edge technologies like *de novo* designed proteins require rigorous biosafety and ecological impact assessments,

necessitating urgent regulatory frameworks (Wu et al., 2023; Kortemme, 2024; Li et al., 2025). To build public trust, the scientific community must transparently communicate the science, potential benefits, such as reduced pesticide use and enhanced nutrition, and risk management measures to the public and policymakers. To bridge the widening digital divide, international cooperation, open-source tools, and knowledge sharing are essential to ensure these advanced technologies benefit everyone (Li et al., 2025). AI is a powerful tool to augment human intelligence, not replace breeders. For the foreseeable future, a breeder's expertise, intuition, and creative vision will remain the soul of the improvement process (Moor et al., 2023). Overall, AI provides optimal mathematical solutions, while experienced breeders assess their real-world feasibility and suitability for local practices and markets.

Author contributions

HW: Writing – original draft. ML: Writing – original draft. YL: Writing – original draft. JY: Writing – original draft. YC: Writing – review & editing, Writing – original draft.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

The author YC declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

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

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500. doi: 10.1038/s41586-024-07487-w

Alemu, A., Åstrand, J., Montesinos-Lopez, O. A., Y Sanchez, J. I., Fernandez-Gonzalez, J., Tadesse, W., et al. (2024). Genomic selection in plant breeding: Key factors shaping two decades of progress. *Mol. Plant* 17, 552–578. doi: 10.1016/j.molp.2024.03.007

Cao, Y., Feng, X., Ding, B., Huo, H., Abdullah, M., Hong, J., et al. (2025). Gap-free genome assemblies of two Pyrus bretschneideri cultivars and GWAS analyses identify a CCCH zinc finger protein as a key regulator of stone cell formation in pear fruit. *Plant Commun.* 6, 101238. doi: 10.1016/j.xplc.2024.101238

Cao, Y., Hong, J., Zhao, Y., Li, X., Feng, X., Wang, H., et al. (2024). *De novo* gene integration into regulatory networks via interaction with conserved genes in peach. *Horticulture Res.* 11, uhae252. doi: 10.1093/hr/uhae252

Crow, J. F. (1998). 90 years ago: the beginning of hybrid maize. *Genetics* 148, 923–928. doi: 10.1093/genetics/148.3.923

Danilevicz, M. F., Gill, M., Anderson, R., Batley, J., Bennamoun, M., Bayer, P. E., et al. (2022). Plant genotype to phenotype prediction using machine learning. *Front. Genet.* 13, 822173. doi: 10.3389/fgene.2022.822173

Gibbs, J. A., Pound, M. P., French, A. P., Wells, D. M., Murchie, E. H., and Pridmore, T. P. (2019). Active vision and surface reconstruction for 3D plant shoot modelling. *IEEE/ACM Trans. Comput. Biol. Bioinf.* 17, 1907–1917. doi: 10.1109/TCBB.2019.2896908

Hasegawa, T., Sakurai, G., Fujimori, S., Takahashi, K., Hijioka, Y., and Masui, T. (2021). Extreme climate events increase risk of global food insecurity and adaptation needs. *Nat. Food* 2, 587–595. doi: 10.1038/s43016-021-00335-4

Jiang, L., Li, X., Lyu, K., Wang, H., Li, Z., Qi, W., et al. (2025). Rosaceae phylogenomic studies provide insights into the evolution of new genes. *Hortic. Plant J.* 11, 389–405. doi: 10.1016/j.hpj.2024.02.002

Kim, M.-G., Go, M.-J., Kang, S.-H., Jeong, S.-H., and Lim, K. (2025). Revolutionizing CRISPR technology with artificial intelligence. *Exp. Mol. Med.* 1-13. doi: 10.1038/s12276-025-01462-9

Kortemme, T. (2024). *De novo* protein design—From new structures to programmable functions. *Cell* 187, 526–544. doi: 10.1016/j.cell.2023.12.028

Lamanna, G., Delre, P., Marcou, G., Saviano, M., Varnek, A., Horvath, D., et al. (2023). GENERA: a combined genetic/deep-learning algorithm for multiobjective target-oriented *de novo* design. *J. Chem. Inf. Modeling* 63, 5107–5119. doi: 10.1021/acs.jcim.3c00963

Li, G., An, L., Yang, W., Yang, L., Wei, T., Shi, J., et al. (2025). Integrated biotechnological and AI innovations for crop improvement. *Nature* 643, 925–937. doi: 10.1038/s41586-025-09122-8

Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., et al. (2018). Domestication of wild tomato is accelerated by genome editing. *Nat. Biotechnol.* 36, 1160–1163. doi: 10.1038/nbt.4273

Lu, Y., Wu, K., Jiang, Y., Guo, Y., and Desneux, N. (2012). Widespread adoption of Bt cotton and insecticide decrease promotes biocontrol services. *Nature* 487, 362–365. doi: 10.1038/nature11153

Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., et al. (2023). Foundation models for generalist medical artificial intelligence. *Nature* 616, 259–265. doi: 10.1038/s41586-023-05881-4

Pingali, P. L. (2012). Green revolution: impacts, limits, and the path ahead. *Proc. Natl. Acad. Sci.* 109, 12302–12308. doi: 10.1073/pnas.0912953109

Ruffolo, J. A., Nayfach, S., Gallagher, J., Bhatnagar, A., Beazer, J., Hussain, R., et al. (2025). Design of highly functional genome editors by modelling CRISPR–Cas sequences. *Nature* 645, 518–525. doi: 10.1038/s41586-025-09298-z

Sharma, J., and Goel, P. (2025). "The use of AI for phenotype-genotype mapping," in *Artificial Intelligence (AI) in Cell and Genetic Engineering.* (Springer: Methods in Molecular Biology), 369–410.

Sheikh, M., Iqra, F., Ambreen, H., Pravin, K. A., Ikra, M., and Chung, Y. S. (2024). Integrating artificial intelligence and high-throughput phenotyping for crop improvement. *J. Integr. Agric.* 23, 1787–1802. doi: 10.1016/j.jia.2023.10.019

Singh, B. K., Delgado-Baquerizo, M., Egidi, E., Guirado, E., Leach, J. E., Liu, H., et al. (2023). Climate change impacts on plant pathogens, food security and paths forward. *Nat. Rev. Microbiol.* 21, 640–656. doi: 10.1038/s41579-023-00900-7

Van Dijk, M., Morley, T., Rau, M. L., and Saghai, Y. (2021). A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050. *Nat. Food* 2, 494–501. doi: 10.1038/s43016-021-00322-9

Wei, X., Qiu, J., Yong, K., Fan, J., Zhang, Q., Hua, H., et al. (2021). A quantitative genomics map of rice provides genetic insights and guides breeding. *Nat. Genet.* 53, 243–253. doi: 10.1038/s41588-020-00769-9

Wu, H., Han, R., Zhao, L., Liu, M., Chen, H., Li, W., et al. (2025). AutoGP: an intelligent breeding platform for enhancing maize genomic selection. *Plant Commun.* 6. doi: 10.1016/j.xplc.2025.101240

Wu, K., Bai, H., Chang, Y.-T., Redler, R., Mcnally, K. E., Sheffler, W., et al. (2023). *De novo* design of modular peptide-binding proteins by superhelical matching. *Nature* 616, 581–589. doi: 10.1038/s41586-023-05909-9

Wu, Y., and Xie, L. (2025). AI-driven multi-omics integration for multi-scale predictive modeling of genotype-environment-phenotype relationships. *Comput. Struct. Biotechnol. J.* 27, 265–277. doi: 10.1016/j.csbj.2024.12.030

Yetgin, A. (2025). Revolutionizing multi-omics analysis with artificial intelligence and data processing. *Quantitative Biol.* 13, e70002. doi: 10.1002/qub2.70002