

MiniReview

Exploring the Frontier of Genetic Diversity: Recent Advances in Pan-genomics

Chao Shen^{1,2*}

¹Guangdong Provincial Key Laboratory for Green Agricultural Production and Intelligent Equipment, Guangdong Province, China

²College of Biological and Food Engineering, Guangdong University of Petrochemical Technology, Guangdong Province, China

*Correspondence to: Chao Shen, PhD, Lecturer, Guangdong Provincial Key Laboratory for Green Agricultural Production and Intelligent Equipment, College of Biological and Food Engineering, Guangdong University of Petrochemical Technology, Maoming, 525000, Guangdong, China; Email: cshen@gdupt.edu.cn

Received: August 17, 2024 Revised: November 5, 2024 Accepted: December 26, 2024 Published: January 24, 2025

Abstract

Pan-genomics has revolutionized our understanding of genetic diversity within species by providing a comprehensive view of both core and variable genes across multiple genomes. Recent advances in this field, driven by innovations in sequencing technologies, computational methods, and integrative approaches with other -omics fields, have unveiled new layers of complexity in genome architecture, gene regulation, and evolutionary dynamics. This mini-review explores key developments in pan-genomics, including the construction of graph-based pan-genomes, applications in crop improvement, integration with epigenomics, and the emerging role of machine learning. These advances underscore the potential of pan-genomics to drive future discoveries in genomics, with significant implications for agriculture, evolutionary biology, and personalized medicine.

Keywords: Pan-genomics, genetic diversity, epigenetics, machine learning

Citation: Shen C. Exploring the Frontier of Genetic Diversity: Recent Advances in Pan-genomics. *J Mod Agric Biotechnol*, 2025; 4: 1. DOI: 10.53964/jmab.2025001.

1 INTRODUCTION

In the symphony of life, where genes play the notes that compose the music of existence, pan-genomics has emerged as a maestro, orchestrating the vast ensemble of genetic diversity. This area of study not only delves deeply into the genetic diversity within species but also shows great potential in population genetics, evolutionary biology, and crop improvement. This reviews the latest advances in pan-genomics and explores its profound impact on genetic diversity research.

Pan-genomics studies the total genomic sequences of

all individuals within a species, collectively known as the "pan-genome." With decreasing sequencing costs and increasing computational power, methods for constructing pan-genomes have advanced significantly. Recent studies have introduced graph-based pan-genome construction methods, which not only better represent genetic diversity but also more accurately capture structural variations between genes. For instance, a graph-based pan-genome to thoroughly analyze the genetic variation in the model plant *Setaria*, providing new insights for crop improvement^[1]. The new algorithm SVarp discovered many previously unidentified structural variations through the pan-genome

reference sequence, which played an important role in domestication selection on different evolutionary paths (<https://github.com/asylvz/SVarp>).

2 PAN-GENOMICS HAS NOTABLE APPLICATIONS IN CROP IMPROVEMENT

By comparing the pan-genomes of different cultivars and wild species, researchers can identify genes associated with important agronomic traits and optimize them through molecular breeding. For example, Wang et al. constructed a pan-genome of 33 genetically diverse rice accessions, revealing hidden genomic variations that are closely related to key agronomic traits such as disease resistance and yield, providing new molecular markers for rice breeding^[2]. Similar research on tomato has identified genomic variations associated with crucial agricultural traits^[3], providing a roadmap for breeding programs aimed at improving crop resistance and productivity.

2.1 Combining Pan-genomes with Epigenetics Facilitates the Dissection of Complex Traits and Adaptations

The integration of pan-genomics, epigenetics and has been a significant research direction in recent years. Epigenetic regulation plays a crucial role in gene expression and functional diversity, and pan-genomics provides a comprehensive genetic background to understand these regulatory mechanisms. Liu et al. established a regulatory model in which structural variations leading to pan-genome differences significantly affect gene expression patterns, which in turn affects phenotypic variation in *Brassica oleracea*, deepening the understanding of the mechanism by which structural variation regulates trait formation^[4]. In addition, micro-collinearity and macro-collinearity analysis combined with functional genomic data can reveal the conservation and evolutionary patterns of genes and regulatory elements, which advance our in-depth understanding of gene expression regulation and environmental adaptation^[5]. This integrated approach helps us to understand how genome structure and regulatory networks evolve together, thereby affecting the complex traits and adaptability.

2.2 Pan-genome Accelerates Research Progress on Genome Structural Variation

The pan-genome can reveal differences in genome structure between different individuals, and these differences are often reflected in the organization of the three-dimensional genome. For example, three-dimensional genome analysis of 27 soybean germplasms found that the distribution and variation of A/B chromatin regions are closely related to genomic characteristics. By comparing these regions, it was found that the transition between A/B regions mainly occurred in regions with intermediate genomic characteristics. These variations, in addition, can reshape the evolution of the three-dimensional genome by driving these SVs^[6]. Pan-genome and pan-3D

genome analysis of cotton genus clarified the evolutionary relationship between transposon-driven genome size changes and chromatin higher-order structural remodeling. Further combined with population genetics research, the genomic basis of fiber quality evolution was revealed, providing theoretical support for the use of wild cotton germplasm resources to improve cultivated varieties^[7]. The rearrangement of genes in three-dimensional space may make some genes more easily activated, allowing species to adapt to new environments more quickly^[8]. In addition, this dynamic genome folding ability provides flexibility and adaptability for species to survive and reproduce in different ecological environments^[9]. Some genes may form regulatory networks due to spatial proximity, thereby responding together in a specific environment^[10]. These findings are of great significance for agricultural breeding and crop improvement, helping us develop plant varieties that are more environmentally adaptable and stress-resistant.

2.3 Pan-genomes Drive Evolution and Adaptability

Pan-genomics also plays an essential role in understanding species evolution and adaptation. For example, *Brassica oleracea* has uncovered signatures of domestication selection in genome evolution and gene variations associated with agronomic traits^[11]. Broomcorn millet have revealed genetic changes in these species during adaptation to different environments and domestication processes, providing new perspectives on adaptive evolution^[12]. The construction of the rice pan-genome provides a new reference framework for functional genomics research in rice, enabling the identification of specific genes related to environmental adaptability, which is crucial for improving rice stress resistance^[13]. It also reveals the evolutionary path of rice. The genetic changes and genome reorganizations exhibited by different rice varieties when adapting to different environments provide important information for studying their evolutionary process^[14]. At the same time, pan-genomes can help identify genes related to climate adaptation, thereby predicting the performance of species under future climate conditions and revealing how genetic variation affects phenotypic plasticity and adaptability^[15]. This insight into the genetic architecture of adaptive traits provides a roadmap for breeding programs aimed at improving crop resistance and productivity.

2.4 Machine Learning Drives Progress in Pan-genomic Research

Pan-genome data are large and complex, making them difficult to handle with traditional analysis methods. Machine learning can effectively extract useful features from large-scale Pan-genome data for pattern recognition and functional prediction. For example, graph-based machine learning methods can be used to analyze relationships and interactions between genes^[16] and can be used to identify gene patterns associated with specific

phenotypes^[17]. In addition, the BRIDGEcereal platform accelerates gene discovery and understanding of trait variation through the use of novel unsupervised machine learning algorithms such as CHOICE and CLIPS for large insertion-deletion polymorphisms in major cereal crops, especially in it is of great significance in crop improvement^[18]. The Maize Feature Store (MFS) uses basic and advanced machine learning models to predict core and non-core genes. This approach significantly helps distinguish different gene classes and is crucial for well-targeted crop breeding programs^[19]. Supervised learning, such as support vector machines (SVM) or random forest models, can extract useful information from complex genomic data and perform classification and regression analysis, greatly improving the efficiency of crop genomic data analysis and providing a scientific basis for breeding decisions^[20,21]. Cluster analysis of different crop genomes using unsupervised learning methods revealed multiple new genetic variations that are closely related to crop growth characteristics, providing important clues for subsequent gene function research and data support for precision breeding^[22,23]. The above mentioned researches show that pan-genomics combined with machine learning can greatly improve the prediction and selection capabilities of crop traits and genetic variations.

2.5 The Role of Pan-genomics in Agricultural Management

With the challenges brought by climate change, agricultural management is facing unprecedented pressure. Pan-genomics can help identify crop varieties that are adapted to different climatic conditions, thereby providing farmers with more accurate planting recommendations. For example, specific genomic information can guide farmers to choose crop varieties that perform well under high temperature or low moisture conditions, which will directly affect the growth and yield of crops and improve crop adaptability^[24,25]. In addition, pan-genomic data can be used to develop crop varieties that are resistant to pests and diseases, and reduce the use of pesticides by identifying related genes and introducing them into other crop varieties. This not only helps to protect the ecological environment, but also reduces farmers' production costs and improves economic benefits^[24,26]. Further using pan-genomic information, agricultural managers can formulate fertilization and irrigation plans more scientifically. By understanding the needs and adaptability of crops to different nutrients, farmers can optimize fertilizer use, reduce negative impacts on the environment, and improve crop yields and quality^[25,27].

3 FUTURE DEVELOPMENT

As more species' pan-genome data are accumulated and analysis methods continue to innovate, pan-genomics will play a critical role in broader biological research. Future research directions include: (1) Combining machine learning and big data technologies to enhance the efficiency

and accuracy of pan-genome construction. (2) Using pan-genome data to study how genetic diversity influences species' adaptation to environmental changes. (3) Pan-genome provides powerful data support for precision breeding. By comparing different genomes, breeders can quickly identify genotypes with excellent traits and introduce them into breeding programs. This approach can significantly shorten the breeding cycle and improve breeding efficiency.

As an emerging field in genomics research, pan-genomics is rapidly developing and continually breaking new ground. Applying pan-genomics in crop design breeding and pest and disease control to achieve precise agricultural management. By deeply analyzing genetic diversity, pan-genomics not only provides rich data resources for fundamental biological research but also opens new avenues for applied research, particularly in crop improvement and precision agriculture. We anticipate that future pan-genomics research will bring more exciting discoveries and contribute significantly to sustainable development.

4 CONCLUSION

Pan-genomics research provides strong data support for precision breeding and agricultural management. By identifying genes associated with important traits, farmers and breeders can conduct variety selection and agricultural management more scientifically. With the continuous advancement of technology, the application prospects of pan-genomics in actual agriculture will be broader and can make positive contributions to the sustainable development of global agriculture. By reviewing the latest advances and future prospects in pan-genomics, we gain a clearer understanding of the vast potential and profound impact of this field. It is hoped that this editorial will inspire more researchers to engage in pan-genomics studies and collectively advance the frontier of science.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (NO.32201873), State Key Laboratory of Cotton Bio-breeding and Integrated Utilization Open Fund (CB2024A21), Maoming Science and Technology Project (2021KJZXZJGSPDX006) and the Projects of Talents Recruitment of Guangdong University of Petrochemical Technology (2019rc112).

Conflicts of Interest

The authors declared no conflict of interest.

Data Availability

Data sharing is not applicable to this review as no datasets were generated or analyzed during the current study.

Copyright and Permissions

Copyright © 2025 The Author(s). Published by Innovation

Forever Publishing Group Limited. This open-access article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, sharing, adaptation, distribution, and reproduction in any medium, provided the original work is properly cited.

Author Contribution

The author contributed to the manuscript and approved the final version.

References

- [1] He Q, Tang S, Zhi H et al. A graph-based genome and pan-genome variation of the model plant. *Setaria Nat Genet*, 2023; 55:1232-1242.[DOI]
- [2] Qin P, Wang H, Chen WL et al. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell*, 2021; 184, 3542-3558.[DOI]
- [3] Ning L, Qiang H, Juan W et al. Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat Genet*, 2023; 55, 852-860.[DOI]
- [4] Li X, Wang Y, Cai CC et al. Large-scale gene expression alterations introduced by structural variation drive morphotype diversification in Brassica oleracea. *Nat Genet*, 2024; 56, 517-529.[DOI]
- [5] Yu HW, Li YL, Han WT et al. Pan-evolutionary and regulatory genome architecture delineated by an integrated macro- and microsynteny approach. *Nat Protoc*, 2024; 19, 1623-1678.[DOI]
- [6] Ni L, Liu Y, Ma X et al. Pan-3D genome analysis reveals structural and functional differentiation of soybean genomes. *Genome Biol*, 2023; 24, 12.[DOI]
- [7] Wang MJ, Li JY, Qi ZY et al. Genomic innovation and regulatory rewiring during evolution of the cotton genus *Gossypium*. *Nat Genet*, 2022; 54, 1959-1971.[DOI]
- [8] Marques DA, Jones FC, Di Palma F et al. Genomic changes underlying repeated niche shifts in an adaptive radiation. *Evol*, 2022; 76, 1301-1319.[DOI]
- [9] Li A, Zhao M, Zhang Z et al. Genome architecture and selective signals compensatorily shape plastic response to a new environment. *Innovation*, 2023; 4, 100464.[DOI]
- [10] Mohan AV, Escuer P, Cornet C et al. A three-dimensional genomics view for speciation research. *Trends Genet*, 2024; 40, 638-641.[DOI]
- [11] Guo N, Wang SY, Wang TY et al. A graph-based pan-genome of Brassica oleracea provides new insights into its domestication and morphotype diversification. *Plant Commun*, 2024; 5, 100791.[DOI]
- [12] Chen J, Liu Y, Liu M et al. Pangenome analysis reveals genomic variations associated with domestication traits in broomcorn millet. *Nat Genet*, 2023; 55, 2243-2254.[DOI]
- [13] Liu Y, Tian Z. Super graph-based pan-genome: Bringing rice functional genomic study into a new dawn. *Mol Plant*, 2022; 15, 1409-1411.[DOI]
- [14] Huang C, Chen Z, Liang C. Oryza pan-genomics: A new foundation for future rice research and improvement. *Crop J*, 2021; 9, 622-632.[DOI]
- [15] Bernatchez L, Ferchaud AL, Berger CS et al. Genomics for monitoring and understanding species responses to global climate change. *Nat Rev Genet*, 2024; 25, 165-183.[DOI]
- [16] Kahanda I, Mudge J, Manuweera B et al. Graph-Based Machine Learning Approaches for Pangenomics. Springer International Publishing: Cham, Germany, 2024; 117-132.[DOI]
- [17] Do DT, Yang MR, Vo TNS et al. Unitig-centered pan-genome machine learning approach for predicting antibiotic resistance and discovering novel resistance genes in bacterial strains. *Comput Struct Biotechnol*, 2024; 23, 1864-1876.[DOI]
- [18] Zhang B, Huang H, Tibbs-Cortes LE et al. Streamline unsupervised machine learning to survey and graph indel-based haplotypes from pan-genomes. *Mol Plant*, 2023; 16, 975-978.[DOI]
- [19] Shatabdi S, Woodhouse MR, Portwood JL et al. Maize Feature Store: A centralized resource to manage and analyze curated maize multi-omics features for machine learning applications. *Database-oxford*, 2023.[DOI]
- [20] Tong H, Nikoloski Z. Machine learning approaches for crop improvement: Leveraging phenotypic and genotypic big data. *J Plant Physiol*, 2021; 257, 153354.[DOI]
- [21] Farooq MA, Gao S, Hassan MA et al. Artificial intelligence in plant breeding. *Trends Genet*, 2024; 40, 891-908.[DOI]
- [22] Pan S, Qin B, Bi L et al. An Unsupervised Learning Method for the Detection of Genetically Modified Crops Based on Terahertz Spectral Data Analysis. *Secur Commun Netw*, 2021; 5516253.[DOI]
- [23] Wang X, Zeng H, Lin L et al. Deep learning-empowered crop breeding: intelligent, efficient and promising. *Front Plant Sci*, 2023; 14, 1260089.[DOI]
- [24] Tay Fernandez, C G, Nestor BJ et al. Expanding Gene-Editing Potential in Crop Improvement with Pangenomes. *Int J Mol Sci*, 2022; 23, 2276.[DOI]
- [25] Tay Fernandez, C G, Nestor BJ et al. Pangenomes as a Resource to Accelerate Breeding of Under-Utilised Crop Species. *Int J Mol Sci*, 2022; 23, 2671.[DOI]
- [26] Chapman MA, He Y, Zhou M. Beyond a reference genome: pangenomes and population genomics of underutilized and orphan crops for future food and nutrition security. *New Phytol*, 2022; 234, 1583-1597.[DOI]
- [27] Khan AW, Garg V, Roorkiwal M et al. Super-Pangenome by Integrating the Wild Side of a Species for Accelerated Crop Improvement. *Trends Plant Sci*, 2020; 25, 148-158.[DOI]