

Haplotype-based Pangenomics: A Blueprint for Climate Adaptation in Plants

Wanfei Liu (刘万飞) ^{*}, Peng Cui (崔鹏) ^{*}

Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

*Corresponding authors: cuipeng@caas.cn (Cui P), liuwanfei@caas.cn (Liu W).

Handling Editor: Xiangfeng Wang

Escalating impacts of climate change pose serious threats to ecosystems and agriculture, and thus to human societies and their living conditions. Plants — both wild and cultivated — are constantly facing global warming, drought, and precipitation change, rendering plant adaptability a crucial process. Haplotype-based pangenomics is a collection of genomes from each haplotype in multiple strains of a species, offering a cutting-edge approach for identifying traits that foster climate resilience. Although pangenomics focuses on building a pangenome for an agricultural crop, often within cultivars of a plant species, this technology holds significant untapped potential for preserving biodiversity, stabilizing ecosystems, and sequestering carbon.

Pangenomics has already revolutionized crop breeding, with major breakthroughs achieved in rice, maize, wheat, soybean, tomato, and potato [1], just to name a few. These pangenomic studies have been primarily focusing on crop breeding, adaptation, and evolutionary history [2]. In the context of climate-associated variables, developing a pangenome for non-agricultural species, such as forest trees, wetland plants, and grassland species, is also crucial for water regulation, soil stability, carbon storage, and biodiversity conservation. For instance, forest species such as oak (*Quercus*) and poplar (*Populus*) absorb significant amounts of CO₂, acting as natural carbon sinks; wetland species such as *Spartina alterniflora* help prevent floods and store carbon in waterlogged soils; and grassland species in savannas and temperate grasslands have deep-rooted systems that help stabilize soils and store carbon, thereby preventing soil erosion and enhancing soil carbon. Applying pangenomics to these species helps uncover structural variants (SVs) associated with key traits, such as species adaptation and resilience to environmental changes [3,4]. The sexual reproduction genome always originates from a combination of two parental genomes, which may retain a highly heterozygous state. To solve the problem of high heterozygosity, a common strategy is to assemble two parental haplotypes into a chimeric genome. However, this strategy corrodes genomic variations between parents, which may be critical for the heritability of diseases and phenotypes, and allele-specific gene expression [5,6]. This underscores the need for haplotype-based pangenomics and comparative genomics; the latter emphasizes cross-lineage genomic studies for major crops, such as those among the popular Poaceae and Solanaceae species.

Haplotype-based pangenomics presents a promising approach for analyzing environmental adaptability, as it can capture specific combinations of genetic variations that are crucial for resilience to changing conditions. Previous studies have demonstrated the role of haplotypes in plant climate adaptation and fruit quality heterosis [4,6]. In some highly heterozygous plants, such as trees that are closely related to ecosystems and climate, incorporating haplotypes facilitates the identification of key genes related to environmental adaptation.

A recent study led by Hansheng Zhao built a haplotype-based pangenome from 16 representative moso bamboo (*Phyllostachys edulis*) accessions (RMAs), providing valuable insights into genetic diversity, genomic architecture, climate adaptation, and population risks under future climate conditions [7]. This is the first pangenome of moso bamboo at the haplotype level, thus providing extensive genomic resources for future bamboo research.

One of the key findings is the extensive genetic variations between haplotypes rather than between accessions (97.0% *vs.* 3%), which assumed that there has been a significant difference between the two ancestral haplotypes in the common ancestor of moso bamboo populations. Among these RMAs, most inter-haplotype variations (68.5% and 68.8% for short variations and SVs, respectively) are detected across all accessions, while only a few inter-accession variations are present in all accessions (0.2% and 0.9% for short variations and SVs, respectively). Further analysis reveals that the number of single nucleotide polymorphisms (SNPs) among the same haplotypes is significantly lower than that among different haplotypes (637808 SNPs *vs.* 2933471 SNPs on average) by comparison among haplotypes of the three well-phased genomes.

To systematically compare genes and alleles of the 32 haplotype assemblies, the authors categorize their gene sets into core, softcore, dispensable, and private groups by their presence across accessions (called gene frequency), or into double-allele, single-allele, and variable-allele groups based on allele composition. The 12 groups, resulting from the combination of gene frequency and allele composition, possess different characteristics in terms of gene structure, expression pattern, and functional features. The core and double-allele gene sets have greater gene

Received: 2 January 2025; Revised: 3 March 2025; Accepted: 5 March 2025.

© The Author(s) 2025. Published by Oxford University Press and Science Press on behalf of the Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

length, cDNA length, CDS number, and CDS size than the private and single-allele gene sets. Moreover, the average gene expression status gradually decreases from core to private gene sets in all allele composition groups, and gradually decreases from double-allele to single-allele gene sets in all gene frequency groups except for the private gene sets. Furthermore, the core and single-allele gene sets, which represent genes present in all accessions but only in one haplotype assembly, are related to environmental adaptations, such as stress tolerance, disease resistance, and DNA damage repair. These adaptation-related genes are also very important and exploitable in other species.

Allele-specific expression gene sets have a high degree of accession specificity (81.8% in 1–2 accessions *vs.* 0.1% in all accessions), exhibit tissue-specific expression patterns (72% variable tissue expression patterns *vs.* 28% consistent tissue expression patterns), and are associated with environmental adaptations, such as stimuli and defense responses and tissue-specific developmental processes.

The genotype–environment association analysis identifies 1050 adaptive variations associated with bioclimatic variables, revealing a greater contribution of climate effects than that of geography (35% *vs.* 13% of adaptive variations). The risk of non-adaptedness analysis and gradient forest modeling indicate that the northern and western regions may have been facing adaptive risks under the high-emission scenario.

However, current genomic offset models — used to predict how species will respond to climate changes — are limited [8]. These models often fail to account for epigenetic factors (changes in gene activity not caused by changes in DNA sequences) and complex interactions between genes and environments. To overcome these barriers, future models must integrate multi-omics approaches (such as population genomics, transcriptomics, epigenetics, metabolomics, and microbiomics) and adjust to changing environmental conditions (both present and future).

In addition, the application of pangenomics to climate adaptation poses significant technical challenges. First, the large genomes of many plants make sequencing expensive and resource-intensive. Next, current pangenome methodologies largely rely on linear genome models, which pose challenges for integrating multiple genomes from the same species and ensuring accessibility of the integrated representation to biologists. Furthermore, many types of genomic resources, such as haplotype information critical for understanding climate adaptation, remain underutilized or are wasted in pangenome analyses due to technical limitations [7]. This emphasizes the need for the development of new analytical techniques and computational methods to more accurately represent and analyze pangenomic data.

Although promising, the application of pangenomics in climate adaptation poses several ethical and socioeconomic challenges. For example, intellectual property concerns exacerbate inequalities by restricting access to climate-resilient technologies in developing nations. Smallholder farmers in sub-Saharan Africa may struggle to afford drought-resistant seeds due to high costs and patent restrictions. The Nagoya Protocol, which aims to ensure fair benefit-sharing with indigenous communities, provides a framework for equitable access. However, more must be done to enhance fair benefit-sharing mechanisms and to ensure equitable technology distribution. Expanding global funding and supporting subsidized research will help bridge the gap between technological innovation and global food security [9].

Artificial Intelligence (AI) and Machine Learning (ML) are transforming genomic analyses by accelerating the discovery of complex traits required for climate adaptation. AI models are increasingly used to identify SVs and gene–environment interactions [8,10]. When integrated with haplotype-based pangenomes, AI and ML enhance the ability to discern intricate genetic patterns and predict phenotypic outcomes under varying environmental conditions [7]. Additionally, advanced techniques, such as HiFi sequencing, combined with SV analysis, improve the detection of large genomic changes, thereby deepening our understanding of adaptive traits. However, high costs remain a barrier, particularly for developing countries.

Modern-day genome editing tools, such as CRISPR-Cas9, have been extensively explored to understand and develop stress-tolerant plant varieties to mitigate the negative effects of climate changes. CRISPR-Cas9 has frequently been employed to understand the stress-responsive genes of heat, drought, salt, and heavy metals, which can be used to enhance climate resilience [11]. Genome editing tools also provide solutions and strategies for plant conservation or restoration, although there are challenges, such as obtaining efficient transformants, producing transgene-free plants, achieving homozygous mutants, and assessing ecological risks [12].

The success of the pangenome depends on global collaboration, open-access data, and continued integration of AI and gene-editing technologies. Projects such as the Earth BioGenome Project and the Consultative Group on International Agricultural Research are already promoting cross-border sharing of genomic data, allowing researchers to apply insights where they are most needed. Reducing costs and expanding data-sharing efforts are essential to ensure that AI-enhanced genomic tools can be effectively leveraged for global climate resilience.

Looking forward, the convergence of long-read sequencing, AI, and CRISPR is set to revolutionize agriculture and biodiversity conservation. Pangenomics based on long-read sequencing can identify genomic variations related to key traits, AI-based genomic analyses can uncover genomic variations associated with agronomic traits and climate changes, and CRISPR-based genome editing tools can accelerate breeding improvements for environmental adaptation. By fostering global cooperation and making genomic technologies accessible, the scientific community can help create climate-resilient species and protect critical ecosystems.

In conclusion, haplotype-based pangenomes are a promising solution for climate adaptation challenges. By extending genomic tools beyond agriculture to include keystone species that are critical for carbon sequestration and ecosystem stability, climate resilience can be built across multiple fronts. Through global collaboration, integration of AI and precision breeding, and addressing ethical and socioeconomic barriers, we can ensure that genomic innovations benefit all regions. With sustained investments in data-driven research, plants — and the ecosystems they support — can be equipped to thrive in a rapidly changing world.

CRedit author statement

Wanfei Liu: Writing – original draft, Writing – review & editing. **Peng Cui:** Conceptualization, Funding acquisition, Writing – original draft, Writing – review & editing. Both authors have read and approved the final manuscript.

Competing interests

Both authors have declared no competing interests.

Acknowledgments

We are grateful to Jun Yu from the Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation for help with language polishing. This work was supported by the Biological Breeding-National Science and Technology Major Project (Grant No. 2022ZD04017 to Peng Cui), China.

ORCID

0000-0002-6637-6807 (Wanfei Liu)

0000-0003-2989-571X (Peng Cui)

References

- [1] Shi J, Tian Z, Lai J, Huang X. Plant pan-genomics and its applications. *Mol Plant* 2023;16:168–86.
- [2] Li W, Liu J, Zhang H, Liu Z, Wang Y, Xing L, et al. Plant pan-genomics: recent advances, new challenges, and roads ahead. *J Genet Genomics* 2022;49:833–46.
- [3] He Q, Tang S, Zhi H, Chen J, Zhang J, Liang H, et al. A graph-based genome and pan-genome variation of the model plant *Setaria*. *Nat Genet* 2023;55:1232–42.
- [4] Shi T, Zhang X, Hou Y, Jia C, Dan X, Zhang Y, et al. The super-pangenome of *Populus* unveils genomic facets for its adaptation and diversification in widespread forest trees. *Mol Plant* 2024;17:725–46.
- [5] He Y, Chu Y, Guo S, Hu J, Li R, Zheng Y, et al. T2T-YAO: a telomere-to-telomere assembled diploid reference genome for Han Chinese. *Genomics Proteomics Bioinformatics* 2023;21:1085–100.
- [6] Li Q, Qiao X, Li L, Gu C, Yin H, Qi K, et al. Haplotype-resolved T2T genome assemblies and pangenome graph of pear reveal diverse patterns of allele-specific expression and the genomic basis of fruit quality traits. *Plant Commun* 2024;5:101000.
- [7] Hou Y, Gan J, Fan Z, Sun L, Garg V, Wang Y, et al. Haplotype-based pangenomes reveal genetic variations and climate adaptations in moso bamboo populations. *Nat Commun* 2024;15:8085.
- [8] Feng J, Dan X, Cui Y, Gong Y, Peng M, Sang Y, et al. Integrating evolutionary genomics of forest trees to inform future tree breeding amid rapid climate change. *Plant Commun* 2024;5:101044.
- [9] Rosegrant MW, Sulser TB, Wiebe K. Global investment gap in agricultural research and innovation to meet Sustainable Development Goals for hunger and Paris Agreement climate change mitigation. *Front Sustain Food Syst* 2022;6:965767.
- [10] Ahsan MU, Liu Q, Perdomo JE, Fang L, Wang K. A survey of algorithms for the detection of genomic structural variants from long-read sequencing data. *Nat Methods* 2023;20:1143–58.
- [11] Choudry MW, Riaz R, Nawaz P, Ashraf M, Ijaz B, Bakhsh A. CRISPR-Cas9 mediated understanding of plants' abiotic stress-responsive genes to combat changing climatic patterns. *Funct Integr Genomics* 2024;24:132.
- [12] Yin K, Chung MY, Lan B, Du FK, Chung MG. Plant conservation in the age of genome editing: opportunities and challenges. *Genome Biol* 2024;25:279.