Plant Biotechnology Journal

REVIEW ARTICLE OPEN ACCESS

Challenges in Bringing Pangenome Research Into Breeding: A Case Study in Rice

Shuai Nie[1] | Fangping Li[2] | Risheng Li[1] | Jian Wang[1] | Yamei Ma[1] | Chon Kit Kenneth Chan[3] | Junliang Zhao[1] | Haifei Hu[1]

[1]Rice Research Institute, Guangdong Academy of Agricultural Sciences; Guangdong Key Laboratory of Rice Science and Technology; Guangdong Rice Engineering Laboratory; Key Laboratory of Genetics and Breeding of High Quality Rice in Southern China (Co-Construction by Ministry and Province), Ministry of Agriculture and Rural Affairs, Guangzhou, China | [2]Guangdong Provincial Key Laboratory of Plant Molecular Breeding, Guangdong Laboratory for Lingnan Modern Agriculture, State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, South China Agricultural University, Guangzhou, China | [3]Faculty of Digital Science and Technology, Macau Millennium College, Macau, China

Correspondence: Chon Kit Kenneth Chan (kenneth.chan@mmc.edu.mo) | Junliang Zhao (zhao_junliang@gdaas.cn) | Haifei Hu (huhaifei@gdaas.cn)

Received: 16 June 2025 | **Revised:** 27 October 2025 | **Accepted:** 2 February 2026

Keywords: artificial intelligence | molecular breeding | rice pangenome | structural variants (SVs)

ABSTRACT

Crop breeding has entered the pangenomics era, unlocking a far more comprehensive view of genetic diversity than a single reference genome can capture. In rice (*Oryza sativa*), a staple crop critical to global food security, the construction of pangenome resources has uncovered extensive structural variations (SVs), presence/absence variations (PAVs) and novel genes that underpin key agronomic traits. As the rice pangenome matures from a research resource into a practical breeding tool, it promises to accelerate the development of higher-yielding, stress-resilient and disease-resistant varieties. This transition represents a pivotal advance toward sustainable agriculture and enhanced global food security, while also establishing a model for applying pangenomics to other crops. Here, we review how rice pangenome research, encompassing both cultivated and wild species, has advanced trait discovery from yield improvement and disease resistance to stress tolerance and enabled new molecular breeding strategies. Despite these advances, several challenges remain before pangenomic data can be routinely integrated into breeding pipelines. The complexity of graph-based data structures, difficulties in detecting multiallelic variants from population-wide resequencing data and the lack of breeder-friendly genotyping tools are significant barriers. Additionally, while artificial intelligence (AI) and machine learning (ML) approaches show great promise for interpreting complex pangenomic data and accelerating trait discovery by genomic selection, their practical adoption is hindered by the absence of breeder-oriented interfaces, integration challenges with multi-omics data and high computational demands. Overcoming these issues will require interdisciplinary collaboration, robust infrastructure and innovations focused on practical breeding needs across diverse crop species.

1 | Introduction

Crop breeding has entered the genomics era, yet progress has been constrained by reliance on a single reference genome to represent a species' genomic diversity. While this single-reference approach has been foundational, it overlooks the extensive genetic diversity within crops, as evidenced by the increasing number of genome assemblies (Mishra et al. 2024; Schreiber et al. 2024). These advances reveal that single reference genome missed significant number of SVs, PAVs and novel genes that distinguish cultivated varieties and wild relatives (Jayakodi et al. 2024; Zhou et al. 2023). The pangenome, representing the combined genetic diversity of multiple accessions, has emerged to encompass core genes found

---

Shuai Nie, Fangping Li and Risheng Li contributed equally to this study.

https://doi.org/10.1111/pbi.70591

in every individual and dispensable gene missing from one or more accessions (Hu et al. 2025). By capturing more diverse genes and SVs present from both cultivated and wild species, the pangenome provides a more comprehensive blueprint of crop genetics than any single genome alone (Chen et al. 2019; Huang et al. 2021).

Despite its potential to uncover hidden genetic variation, substantial gaps remain between generating pangenomic datasets and applying them in breeding programs. This gap arises from the lack of breeder friendly SV or PAV genotyping platforms, integrated pipelines that connect pangenomic and phenotypic data, and sufficient computational resources and expertise to analyse graph-based pangenome. These constraints raise practical questions for plant geneticists and breeders: How can we harness this richer genomic resource to accelerate molecular breeding, and what technical and practical barriers most hinder the integration of pangenomics data into routine breeding workflows?

As a staple crop for over half the global population and a model organism in cereal genomics, rice serves as an ideal model species to address above challenges. This review uses rice to evaluate the promise of the pangenome for crop improvement, highlighting recent breakthroughs enabled by pangenomics based analyses and examining the technical and practical challenges that must be addressed before pangenome-assisted breeding becomes routine. Although our focus is on rice pangenomics as a model system, the challenges and solutions discussed are broadly applicable to pangenome-enabled breeding in other crops. Whereas prior reviews have focused on pangenome construction and variant discovery (He et al. 2025; Li et al. 2022; Tao et al. 2019), we emphasise the translational steps required to embed pangenomic data into breeder workflows. By bridging pangenomes with breeder-friendly implementation strategies, we outline actionable solutions to overcome the complexity of genomics analyses and the challenges of AI integration, providing a roadmap for pangenome-assisted crop improvement across species.

2 | The Promise of Rice Pangenome for Bridging Genomics and Breeding

Rice pangenome research has advanced markedly in recent years, with more than 10 pangenomics studies published to date (Table 1), reflecting both technological progress and a widening research scope (Figure 1). Early pangenomics studies focused primarily on Asian cultivated rice (*Oryza sativa*) and were limited to only a few representative varieties (Schatz et al. 2014). Subsequent improvements in sequencing technologies and computational methods enabled expansion to wild rice species (*Oryza rufipogon*) (Zhao et al. 2018) and African cultivated rice (*Oryza glaberrima*) (Qin et al. 2021).

Notably, a linear rice pangenome was constructed using 3010 Asian cultivated rice accessions to delineate nine distinct subpopulations, providing unprecedented resolution of population structure and genetic diversity (Wang et al. 2018). This comprehensive genomics resource establishes a robust framework for exploring allelic variation and informing future rice breeding strategies. Another landmark study constructed a graph-based pangenome from 145 wild and cultivated rice genomes. This effort revealed 3.87Gb of novel sequences absent from the reference genome and identified 69 531 pan-genes, including 19.74% wild-specific genes linked to disease resistance and environmental adaptation (Guo, Li, et al. 2025). This study also resolved longstanding debates by affirming a single origin of domestication for Asian rice and tracing the emergence of *indica* and *basmati* subpopulations through gene flow (Guo, Li, et al. 2025). These developments underscore the evolution of pangenome studies from linear reference genomes to dynamic graph-based frameworks, capturing SVs and PAVs across thousands of accessions. Complementing these approaches, gene-centric analyses like the Rice Gene Index (RGI) (Yu et al. 2023) integrate gene annotations from multiple individual genomes to map the pangene set of the species, providing a unified Ortholog Gene Index (OGI) for gene functional studies. Moreover, incorporating diverse *Oryza* species into a super-pangenome also provides a more comprehensive view of the genus's genetic diversity, facilitating the discovery of novel genes and SVs essential for breeding (Fornasiero et al. 2025). For instance, a recent pangenome study spanning the *Oryza* genus identified unique NBS-LRR gene clusters in BB- and GG-genome wild rice species, representing untapped resistance loci absent from modern cultivars (Long et al. 2024). Collectively, these rice pangenomes provide unprecedented catalogue of gene pools and genetic variations, which hold promise for breeding applications.

Integrating pangenomic data into breeding strategies represents a transformative shift in rice improvement. Traditional breeding and association studies that rely on a single reference genome often overlook critical genes, especially those absent from the reference, as well as SVs underlying essential traits (Aganezov et al. 2022; Hu, Wang, et al. 2024). In contrast, a pangenomic approach encompasses both core genes and dispensable genes, thereby harnessing a wider spectrum of genetic variation. For example, a rice pangenome analysis of 3010 Asian accessions uncovered ~268 Mb of sequence absent from the Nipponbare reference, revealing 12 465 novel genes and 19 721 dispensable genes (Wang et al. 2018). Later studies employing long-read sequencing of 111 genomes further expanded the known pangenome, identifying a total of 879 Mb of new sequences and around 19 000 new genes when wild rice relatives were included (Zhang et al. 2022). Rice pangenomes thus reveal a vast reservoir of previously hidden alleles associated with disease resistance, stress tolerance, yield and grain quality, offering breeders valuable new targets for breeding selection (Abbai et al. 2019; Chen et al. 2025; Lin et al. 2023; Wei et al. 2024). By moving beyond the constraints of a single reference genome, researchers can now capture 'missing heritability' by exploring SVs and PAVs across diverse germplasm (Yuan et al. 2021; Zhou et al. 2022). Beyond merely cataloguing genes, pangenome references can be integrated with modern molecular breeding strategies such as genome-wide association studies (GWAS) (Jayakodi et al. 2021) and genomic selection (Raza et al. 2025; Yang, He, et al. 2025), allowing PAVs and SVs to serve alongside SNPs as informative markers for linking genotype to phenotype (Jiao et al. 2025). As a result, SV-controlled traits can be pinpointed more effectively, and candidate genes absent from a single reference can be identified within the pangenome. Overall, the rice pangenome provides a holistic view of genetic diversity, enabling more precise

TABLE 1 | Summary of research progress in rice pangenomics studies.

Pangenome composition	Pangenome construction method	Pangenome representations	Number of accessions	Sequencing platform	References
Asian cultivated rice and African cultivated rice	Iterative mapping and assembly	64.10 Mb PAVs and 43 232 pan-genes	12	PacBio	Wang, Yang, et al. (2023)
Asian cultivated rice	Iterative mapping and assembly	268 Mb PAVs and 53 758 pan-genes	3010	Illumina	Wang et al. (2018)
Asian cultivated rice and wild progenitor of Asian cultivated rice	*De novo* assembly and gene annotations comparison	10 872 gene PAVs and 42 580 pan-genes	66	Illumina	Zhao et al. (2018)
Asian cultivated rice and African cultivated rice	Graph pangenome	~24 469 PAVs and 66 636 pan-genes	33	PacBio	Qin et al. (2021)
Asian and African wild and cultivated rice	Graph pangenome	Pangenome of 1.52 Gb and 51 359 pan-genes	251	Nanopore	Shang et al. (2022)
Asian and African wild and cultivated rice	*De novo* assembly and gene annotations comparison	604 Mb PAVS and 60 293 pan-genes	111	PacBio	Zhang et al. (2022)
Asian and African wild and cultivated rice, weedy rice	*De novo* assembly and gene annotations comparison	175 528 pan-gene families	74	PacBio	Wu et al. (2023)
Asian cultivated rice	*De novo* assembly and gene annotations comparison	297 to 786 genome-specific loci	3	Illumina	Schatz et al. (2014)
Asian cultivated rice and wild progenitor of Asian cultivated rice	*De novo* assembly and gene annotations comparison	3.87 Gb of non-reference sequences and 69 531 pan-genes	129	PacBio and Nanopore	Guo, Li, et al. (2025)
African cultivated rice	*De novo* assembly and gene annotations comparison	The gene number ranging from 49 662 to 51 262	3	Illumina	Monat et al. (2017)
Asian cultivated rice	Iterative mapping and assembly	38 998 pan-genes and 71.74 Mb non-reference sequences	60	Illumina	Woldegiorgis et al. (2022)
Wild and cultivated Oryza species	*De novo* assembly and gene annotations comparison	101 723 pan-gene families	13	PacBio	Long et al. (2024)
Asian cultivated rice	*De novo* assembly and gene annotations comparison	119 783 pan-gene families	16	PacBio	Yu et al. (2023)

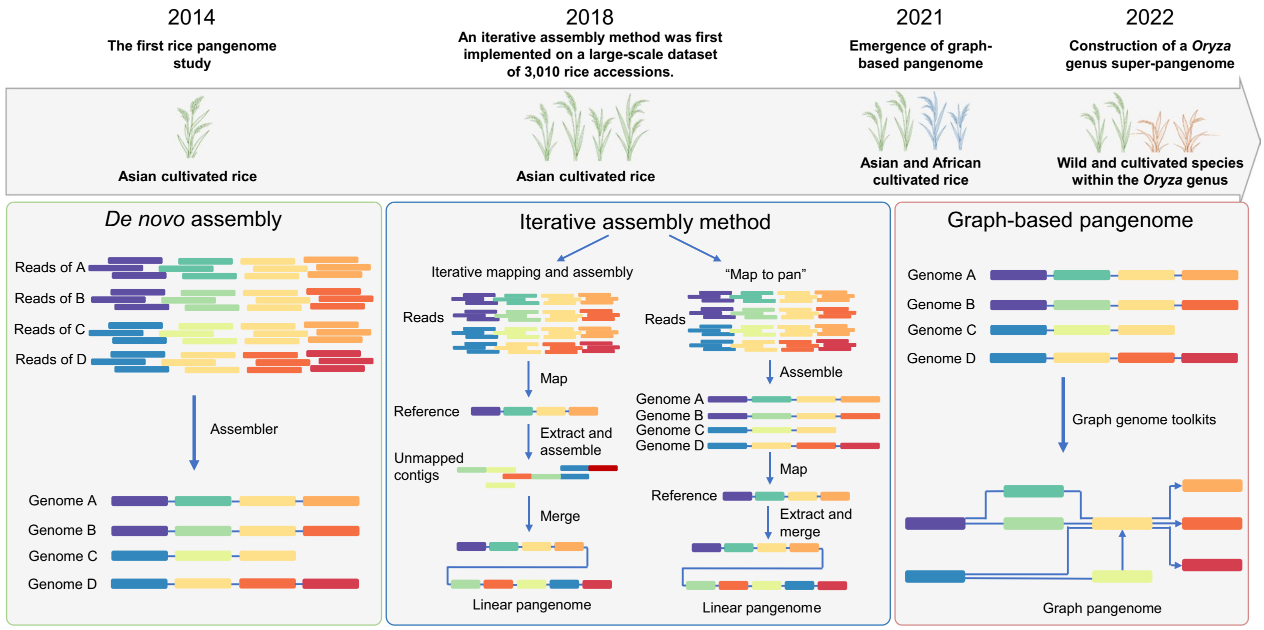

FIGURE 1 | Timeline of major milestones in rice pangenomics research. The first rice pangenome was constructed in 2013 (Schatz et al. 2014). Iterative assembly methods were subsequently applied to a large population of 3010 Asian cultivated rice accessions (Wang et al. 2018). In 2021, the first graph-based rice pangenome was developed (Qin et al. 2021), and by 2022, pangenomics studies had expanded to the genus *Oryza* level (Shang et al. 2022).

and efficient breeding via marker-assisted selection and SV-based genomic selection.

3 | Breakthroughs in Trait Discovery Enabled by Rice Pangenomics

The potential impact of pangenomics on crop improvement is immense, as demonstrated by numerous recent case studies and applications in rice breeding (Huang et al. 2021; Varshney, Roorkiwal, et al. 2021; Wei et al. 2024). Integrated pangenome analyses have unlocked important agronomic traits that previously eluded detection by conventional single-reference approaches (Wei et al. 2024). Several notable examples exemplify how pangenomics has transformed rice research and breeding. A major breakthrough involves the discovery of novel loci for yield and plant architecture. For example, using a 12-genome rice pangenome reference, Wang, Yang, et al. (2023) applied PAV-based GWAS to 413 diverse rice accessions, successfully identifying causal structural variants affecting grain weight and plant height that single reference SNP-based GWAS had failed to detect. Notably, a new quantitative trait locus (QTL) for plant height on chromosome 8 (*qPH8-1*) was discovered exclusively through the pangenomics analysis, highlighting its power to reveal hidden genetic factors influencing key yield components. Building on this discovery, *qPH8-1* was mapped to chromosome 8 (27.9–28.1 Mb) and resolved to *PH8* (*LOC_Os08g44590*) (Chen et al. 2025). Meanwhile, a naturally occurring haplotype, $PH8^{Hap0}$, lowers *PH8* expression and produces moderately shorter plants without reducing tiller number and panicle length. This haplotype is enriched in high-latitude landraces and now predominates in modern *japonica* cultivars, consistent with strong directional selection. In breeding practice, *qPH8-1* acts as height-tuning allele that moderates plant height more gently than *SD1* while maintaining fertility at low temperatures, facilitating adaptation and the geographic expansion of temperate *japonica* rice (Chen et al. 2023).

Similarly, pangenomic approaches have significantly enhanced the discovery of disease resistance genes in rice (Long et al. 2024). Resistance loci often occur in clusters and exhibit PAV polymorphism, making them difficult to detect with single-reference methods. By harnessing pangenome-based GWAS, a recent study discovered 74 QTLs for blast resistance, including the novel *qPBR1* (conferring both panicle and leaf blast resistance) and *qPBR12* (co-localised with the known broad-spectrum resistance gene *ptr*) (Wang et al. 2024). Within *qPBR1*, six candidate genes were pinpointed, one of which (*LOC_Os1g14580*) showed a strong association with enhanced blast resistance. This pangenome-driven approach not only revealed new resistance loci but also reaffirmed cloned Resistance(R) genes (such as *Pi9*, *Pi5*, *Pid1* and *Pita*), demonstrating the power of pangenome-based methods to capture both known and novel disease resistance genes.

Moreover, pangenomics approaches have shed light on the genetic basis of heat stress tolerance in rice. A recent study constructed a pangenome from 60 rice cultivars (45 heat-tolerant and 15 susceptible) to identify genes associated with high-temperature tolerance (Woldegiorgis et al. 2022). This analysis revealed 1141 genes exclusive to heat-tolerant varieties and absent from the reference genome; many of these genes were differentially expressed under heat stress. By intersecting these findings with known heat-tolerance QTLs, researchers pinpointed two strong candidate genes from non-reference regions, suggesting promising targets for developing heat-resilient rice cultivars. Furthermore, the development of the Rice Pangenome Genotyping Array (RPGA), which incorporates approximately

TABLE 2 | Current challenges and potential solutions in rice pangenomics studies and their breeding applications.

Category	Challenge	Potential solution
Data volume and variant complexity	Pangenomes encompass tens of thousands of variable genes and millions of SVs, while multi-allelic variants (e.g., tandem repeats) remain underrepresented, complicating downstream analysis.	Integrate transcriptomic, epigenomic and phenotypic data through high-throughput pipelines and employ gene-centric summarisation workflows to extract core and dispensable gene sets for targeted allele discovery.
Computational and bioinformatic complexity	Linear pangenome representations lack positional context, and graph-based tools (vg, GraphTyper2) demand substantial compute and memory, limiting scalability for large, repeat-rich plant genomes.	Develop scalable software tools such as VRPG that combine linear reference coordinate projection, annotation integration and advanced data structures for graph-based pangenome analysis.
Tool adaptation and resource constraints	Many pangenome tools were developed for human datasets, and breeding programs often lack high-performance computing and specialised bioinformatics expertise.	Establish plant-specific benchmarking frameworks and optimise human-derived tools for crop genomes, following best practices from recent methodological reviews.
Genotyping platform limitations	Conventional SNP arrays miss non-reference and SVs, and novel pangenome arrays require integration with existing breeding decision-support systems.	Integrate RPGA outputs with genome navigation tools like RiceNavi to streamline QTL pyramiding and breeding-route optimisation within familiar breeder interfaces.
AI and machine learning gaps	AI/ML shows promise for variant detection and trait prediction but faces usability, data, and trust issues.	Develop accessible, explainable AI/ML tools tailored for breeding; standardise and share high-quality breeding datasets; invest in collaborative training and infrastructure; design user-friendly decision-support platforms; prioritise model transparency and integration with existing breeding workflows.
Translational and organisational hurdles	Introgression of novel alleles via traditional backcrossing is time-consuming and prone to linkage drag, while CRISPR/Cas9 editing faces regulatory and breeder-acceptance barriers.	Encourage partnerships among breeders, bioinformaticians, and policymakers to align pangenome data with regulations and breeding workflows through training and clear communication.

80000 markers (including SNPs and PAVs) from the 3K rice pangenome, has underscored the practical value of pangenomics tools (Daware et al. 2023). GWAS using the RPGA, identified 42 loci associated with grain size and weight, including eight novel loci that were undetected by traditional single-reference approaches. For example, a dispensable gene on chromosome 7, encoding a WD40-repeat protein at the *qLWR7* locus, was identified as a regulator of the grain length/width ratio. Validation with mapping populations confirmed that pangenome-based markers can uncover new yield-related genes, offering targets for marker-assisted selection or gene editing.

Collectively, these studies illustrate that incorporating pangenome data into rice breeding programs enables the capture of 'missing' genetic variations related to yield, plant architecture, stress tolerance and disease resistance. Tools such as pangenome arrays and PAV-GWAS broaden the search for beneficial alleles and provide a direct pathway from genomic discovery to breeding application. Furthermore, from a translational and practical standpoint, the expanding set of pangenome-identified loci is advancing from discovery to validation, laying the groundwork for crop improvement. Future work should prioritise developing breeder-ready markers, confirming the stability of effects across environments and introgressing favourable alleles into elite germplasm.

4 | Key Barriers and Potential Solutions to Translating Pangenomics Insights Into Practical Breeding

However, transitioning these proof-of-concept successes into routine breeding strategies will require further efforts to address remaining technical and logistical challenges. Several challenges continue to impede its integration into routine molecular breeding (Table 2). Below, we highlight the major barriers along with some actionable solutions.

4.1 | Data Structural Complexity Limit Functional Variant Discovery

A primary obstacle is the enormous volume and structural complexity of data generated. For instance, a comprehensive rice pangenome can encompass tens of thousands of variable genes and millions of SVs, making it difficult to extract insights of the genomic information (Naithani et al. 2023). Most studies remain focused on bi-allelic variants, largely due to limitations of short-read sequencing technologies and the lack of robust analytical frameworks capable of handling complex multi-allelic variation. Tandem repeats and other complex SVs are frequently underrepresented because their accurate detection and interpretation require high-quality genome assemblies, graph-based reference structures and multi-omics integration (Guo, Schreiber, et al. 2025).

Prioritising functional alleles requires more than DNA sequence alone, as transcriptomes indicate gene activity, epigenomic marks inform chromatin context and allele usage, and phenotypes link functional variants. By systematically integrating diverse omics data with pangenomes through high-throughput analytical pipelines, researchers can refine hundreds of thousands of polymorphisms to the subset that directly underlies meaningful biological functions or agronomic traits. For instance, integrating developmental stages transcriptomes with rice pangenome variants has been essential for uncovering the regulatory roles of multiallelic variants (He et al. 2024). In another study, genome-wide DNA methylome and transcriptome data were combined in rice hybrids and their parents to identify differentially methylated regions associated with allele-specific expression and phenotypic traits such as tiller number and biomass, demonstrating that epimutations can drive heritable expression changes relevant to agronomic performance (Chodavarapu et al. 2012). Such integrative analysed approaches enable the identification of expression quantitative trait loci (eQTLs) that remain undetectable using bi-allelic markers alone, offering deeper insights into the genetic regulation of complex traits and supporting more precise breeding decisions (Han et al. 2023).

4.2 | Computational Challenges, Emerging Resources and Standardisation in Pangenome Representation

Rice pangenome analyses face substantial computational and bioinformatic complexity. Traditional linear representations typically concatenate novel sequences or represent them as separate contigs, but often lack positional context. Owing to limitations of early assembly and alignment tools, non-reference sequences are appended without anchoring to a unified genomic framework, complicating placement relative to known genes and structural features (Jonkheer et al. 2025). Graph-based pangenome representations mitigate these issues by resolving structural variants and complex rearrangements with greater fidelity (Tao et al. 2020). However, tools for read mapping and variant calling in graph-based frameworks, such as vg (Garrison et al. 2018) and GraphTyper2 (Eggertsson et al. 2019), are still immature. Many pangenome computational tools were originally developed and benchmarked on human genome datasets, and their performance on plant genomes, which are larger, more repetitive and have greater heterozygosity, remains poorly characterised (Du et al. 2024). For example, vg (Garrison et al. 2018) was the first graph-based tool demonstrated to scale to the 3 Gb human genome; however, its read-mapping accuracy and computational requirements have yet to be systematically benchmarked on highly repetitive genomes typical of crop species. Additionally, many breeding programs lack access to the high-performance computing infrastructure and specialised bioinformatics expertise necessary to manage such complex pangenomic data structures (Li et al. 2022).

While variation graphs can improve genotyping accuracy, they typically demand far more compute than linear reference methods, limiting scalability for large, complex plant genomes (Du et al. 2024). Therefore, there is a pressing need for scalable, efficient software solutions tailored to graph-based pangenomes. Emerging tools like Interactive Visualisation and Interpretation of Pangenome Graphs (VRPG) offer promising solutions by combining linear reference-based coordinate projection with integrated annotation, alongside advanced data structures specifically optimised for graph-based pangenome analysis (Miao and Yue 2025). By leveraging compact, memory-efficient graphs and parallel rendering, VRPG has demonstrated responsive, interactive visualisation on yeast pangenome graphs built from hundreds of high-quality assemblies. The same graph data structures and rendering strategies can be easily translated to plant pangenomes, enabling smooth, interactive exploration of SV-dense regions and integration with gene and phenotype annotations at scale.

Moreover, even simplified approaches that linearise pangenome data can produce extremely large datasets. For instance, aligning the 3000 Rice Genomes Project data to a single reference genome yielded approximately 17 TB of alignment data (Sun et al. 2017), which posed substantial challenges for data storage and impeded computational analyses. While this project provided invaluable insights into rice genetic diversity, integrating such vast genomic datasets into practical breeding workflows remains a significant challenge. Breeders often struggle with the computational demands of analysing such large datasets and the lack of user-friendly tools tailored to breeding objectives (Xu et al. 2022). Furthermore, the inherent complexity of pangenomics data, including SVs and PAVs, requires advanced analytical pipelines and expertise that may not be readily available in many breeding programs. These challenges underscore the urgent need for developing more accessible, scalable tools and standardised resources that can facilitate the effective utilisation of pangenomics data in crop breeding initiatives (Mishra et al. 2024).

To facilitate the practical use of rice pangenomic information, several community resources and data-sharing initiatives now support discovery and breeding applications. The Rice Pangenome Browser (RPAN) offers visualisation and search across 3000 accessions, including PAV and thousands of genes absent from the Nipponbare, making pangenome content directly explorable (Sun et al. 2017). The SNP-Seek database provides genotypes, phenotypes and metadata from the 3000 Rice Genomes Project, supporting large-scale variant mining and allele tracking in breeding populations (Mansueto et al. 2017). For decision support, the RiceNavi implements genome navigation for QTN pyramiding and breeding-route optimisation, demonstrating how pangenome markers can be operationalised in practice (Wei et al. 2021). Ongoing efforts are underway to standardise genomics and phenotypic data formats. The Graphical Fragment Assembly (GFA/rGFA) specifications encode assembly/variation graphs used by many pangenome toolchains (Gonnella et al. 2019), improving graph portability among pipelines. For phenotype metadata, MIAPPE and associated ontologies are widely adopted

(Ćwiek-Kupczyńska et al. 2016), supporting consistent annotation of traits and variant types across repositories and tools.

4.3 | Advancing Genotyping Platforms to Capture Complex Pangenomic Variation

Another significant challenge lies in developing genotyping solutions that are to accommodate diverse variant types and complex genomic contexts involving repetitive sequences and graph-based representations within the pangenome. Traditional breeding programs predominantly rely on SNP arrays or targeted markers designed from a single reference genome, but these approaches struggle to capture non-reference PAVs and SVs. In rice, the Rice Pangenome Genotyping Array (RPGA) assays over 80 K SNPs and PAV probes selected from the 3 K pangenome, enabling genome-wide association studies to uncover loci invisible to single-reference analyses (Daware et al. 2023). More recently, the Rice Super-Population Variation Map (RSPVM) has catalogued over 54 million SNPs, 11 million indels, and approximately 185 000 PAVs across 10 548 accessions on a graph pangenome framework, dramatically improving the ability to genotype rare and structural variants at population scale (Wang, He, et al. 2023). Together, these advances point toward breeder-ready, pangenome-aware platforms that integrate SNPs, indels, SVs and PAVs within a unified, scalable genotyping workflow.

Breeders of other crops face similar needs. In wheat, early platforms such as the 9K iSelect SNP array, which was validated on nearly 3000 accessions, proved instrumental in mapping multiple rust resistance genes (Hao et al. 2017). More recently, the TaNG Axiom array incorporates over 43 000 haplotype-optimised markers derived from skim sequencing of elite lines and landraces, markedly enhancing copy-number variation detection and GWAS resolution (Burridge et al. 2024). Maize researchers have progressed from the Maize 55K SNP array (Xu et al. 2017), which interrogates 50 000 SNPs, to the Axiom Maize Genotyping Array (Mabire et al. 2019), which assays 600 000 markers and captures hundreds of thousands of polymorphisms, including complex structural variants, across global germplasm collections. Nevertheless, despite these gains, array-based platforms retain inherent limits, including reference bias and a biallelic focus.

Therefore, transitioning to an advanced genotyping paradigm requires cost-effective, high-throughput platforms such as customised arrays or targeted sequencing panels that can accurately capture pangenome-wide multi-allelic and structural variants while remaining compatible with existing breeding databases and decision-support systems such as RiceNavi, which currently accommodate only bi-allelic SNPs and small indels (Wei et al. 2021). By learning from these examples across rice, wheat and maize, future genotyping solutions can be optimised to deliver richer variant information without reference bias, paving the way for more precise, pangenome-informed breeding strategies.

4.4 | Emerging AI Approaches in Pangenome Interpretation: Gaps for Breeding

While AI and ML offer transformative potential for interpreting complex pangenomics data, but integrating these approaches into practical breeding pipelines remains a significant gap (Bayer et al. 2021; Hu, Danilevicz, et al. 2024). Recent advances demonstrate that AI can enhance the detection and genotyping of SVs and PAVs within pangenomes that are normally challenging for traditional bioinformatics methods, especially as data volume and complexity increase (Luo et al. 2024). For example, convolutional neural networks (CNNs) and other deep learning models such as SVcnn (Zheng and Shang 2023), cnnLSV (Ma et al. 2023) and SVision (Lin et al. 2022) have been incorporated into SV detection pipelines, improving sensitivity and accuracy across large population datasets and enabling the construction of high-resolution pan-SV maps (Lin et al. 2022). These advances provide breeders with more precise molecular markers and facilitate the identification of cultivar-specific targets for genome editing. Regarding markers based genomic selection with ML, a recent rice study showed that adding SVs with SNPs improved prediction ability in 87% of comparisons; under across-population validation, deep learning outperformed Bayesian baselines by 24% for leaf senescence and 21% for time-to-flowering trait (Vourlaki et al. 2024). A practical pathway is to genotype on a graph-based pangenome to obtain SNP and SV matrices, assemble multi-environment training data, fit GS baselines (GBLUP/RKHS) alongside deep learning where nonlinear or imaging signals matter, apply explainable ML to prioritise breeder-ready alleles, and validate across environments before converting key markers to KASP assays. Resources such as CropGS-Hub provide ready-to-use workflows (Chen et al. 2024), including phenotypic prediction and user-defined model training, that operationalise genomic selection pipelines and can be readily adapted for rice and other crops. Moreover, the Rice SeedLLM database illustrates how the integration of artificial intelligence with genomics can drive meaningful advances in rice research and breeding (Yang, Kong, et al. 2025). Trained on approximately 1.4 million rice-related publications and linked to the Rice Biological Knowledge Graph, this rice specialised large language model enables integrate multi-omics results that accelerate hypothesis generation and the identification of key functional genes for crop improvement. By effectively coupling AI with species-specific biological knowledge, AI-powered knowledge genomics databases can be applied to other crop species (Li et al. 2025).

A pangenome-informed, AI-enabled breeding workflow is now showing promise in rice by combining diverse germplasm panels with sequencing to capture a fuller spectrum of genetic variation through pangenome-based genotyping arrays, thereby improving the explanation of trait heritability (Figure 2). Leveraging heritability-guided phenotype stratification, integrating pangenome loci with AI modelling can support an iterative, increasingly intelligent decision system that delivers varietal profiles, breeding schemes and precision-editing guidelines for prescriptive breeding. Despite these technical breakthroughs, the application of AI in pangenome-assisted breeding still faces several key obstacles. One major issue is usability and accessibility for breeders: most AI-driven tools for pangenome analysis are designed for geneticists and bioinformaticians, not breeders. Breeding programs struggle to adopt pangenome-based methods because intuitive, breeder-friendly interfaces and visualisation platforms are lacking to translate complex graph structures into actionable trait and variant information (Zhang et al. 2024). At the same time, breeding datasets such as field trial records,

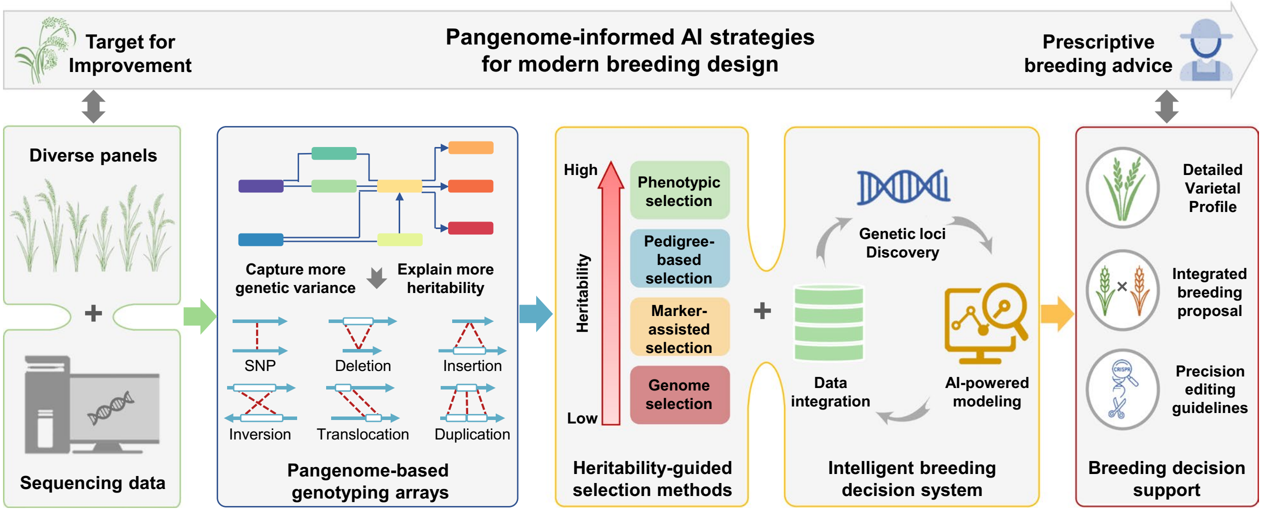

FIGURE 2 | Pangenome-informed AI strategies for enhancing modern rice breeding. This flowchart depicts a pipeline for improving rice traits using pangenome data and AI. Diverse germplasm panels and sequencing data capture extensive genetic variations (e.g., SNPs, deletions, inversions, translocations and duplications) via pangenome-based genotyping arrays, enabling better heritability explanation. Guided by heritability levels for customised phenotype classification, the integration of comprehensive genetic loci from the pangenome with AI's data analysis and modelling capabilities will form an iteratively high-intelligent breeding decision system. This system outputs detailed varietal profiles, integrated breeding proposals and precision editing guidelines to provide prescriptive breeding advice.

remote sensing images and pedigree data exist in separate silos and cannot be integrated smoothly with genomic variant data (Shakoor et al. 2019). Creating unified portals and decision support modules that bring these diverse streams together will be essential to applying pangenomic insights in everyday breeding decisions. There remains a disconnect between the genetic insights generated by AI-powered pangenome analysis and the phenotypic data needed for selection decisions. Effective breeding requires the integration of genotypic, phenotypic and environmental data, yet current ML models often lack access to structured, high-quality breeding datasets with standardised metadata (Varshney, Bohra, et al. 2021). Data scarcity and imbalance further impede the development of robust ML models. Many crops—especially orphan or under-resourced species—suffer from limited genomic and phenotypic data, impeding the development of accurate predictive models for breeding (Hu et al. 2025). In addition, model interpretability and trust remain to be concerns. Breeders need transparent and interpretable models to make confident selection decisions. However, many state-of-the-art AI models, particularly deep learning architectures, function as 'black boxes,' making it difficult to understand how predictions are made or to validate results in a breeding context (Lisboa et al. 2023).

Furthermore, resource and infrastructure limitations present another barrier. The computational demands of AI-driven pangenome analysis can be prohibitive for many breeding programs, particularly in low- and middle-income regions (Talabi et al. 2022). These issues are not merely extensions of general pangenome adoption barriers; they are specific challenges arising from the application of AI/ML as an enabling technology in breeding. For example, while AI/ML could theoretically streamline variant discovery and trait prediction, the current lack of standardised, high-quality datasets and explainable models limits their practical utility in breeding programs (Murmu et al. 2024). Moreover, the resource requirements for deploying advanced AI models often exceed the capacity of many breeding operations, particularly in resource-limited settings. To bridge this gap, future efforts should prioritise the development of AI tools that are tailored for breeding applications, including user-friendly interfaces, standardised data integration frameworks and explainable models (van Dijk et al. 2021). In practical terms, containerised workflows deployed on community platforms such as Galaxy can reduce local operational burdens while providing GUI access and shared compute. Collaborative initiatives to share and standardise breeding datasets, as well as investments in local capacity and infrastructure, will be essential for democratising access to AI and ensuring their impact on crop improvement (Ghamkhar et al. 2025).

4.5 | Translational and Organisational Challenges in Applying Pangenomics Discoveries

Even after novel trait-associated genes are identified, significant translational and organisational challenges remain in bringing these discoveries to fields. Introgression of valuable alleles from wild rice or untapped landraces into elite cultivars often requires multiple generations and can be complicated by issues such as linkage drag or reduced fertility (Wang et al. 2017). While conventional methods like marker-assisted backcrossing are widely used to move traits from wild into cultivated backgrounds, genome editing techniques (particularly CRISPR/Cas9) offer a promising alternative (Dong et al. 2020).

Rice pangenome studies have uncovered beneficial genes, particularly resistance loci lost during domestication but retained in wild rice (Shang et al. 2022). Leveraging these findings through a 'super-pangenome guided' genome editing strategy can rapidly reintroduce desirable traits such as drought tolerance or disease resistance into modern varieties, potentially circumventing the lengthy process of backcrossing (Shang et al. 2022). Crucially, super-pangenomes add knowledge beyond species-level panels by integrating wild and cultivated species across the genus, exposing novel sequences, dispensable genes and cross-species haplotypes that illuminate adaptation and domestication networks and expand the catalogue of edit-ready targets (Khan et al. 2020, 2024). However, the implementation of both traditional and genome editing approaches faces hurdles, including regulatory constraints on genetically modified crops and the need for breeder and public acceptance of new technologies. These issues are common across crop improvement programs.

Breeders typically rely on established selection criteria and operational pipelines, so introducing complex structural variation information or novel PAV markers identified through pangenomics demands a significant shift in breeding practice. This transition requires extensive communication and knowledge exchange between genomic researchers and breeding practitioners (Hu, Scheben, et al. 2024). Furthermore, the technical complexities of interpreting and applying large-scale pangenomic datasets often exceed breeders' traditional expertise. This highlights the critical importance of collaborative training programs, user-friendly bioinformatics tools and decision-support platforms tailored to practical breeding scenarios (Varshney et al. 2016).

5 | Conclusion, Outlook and Implications for Rice and Beyond

The translation of the rice pangenome into routine molecular breeding applications is progressing steadily, albeit with several challenges still to be overcome (Figure 3). Recent proofs-of-concept studies have demonstrated the power of pangenomics approaches to identify trait-associated genes that conventional methods failed to detect (Bayer et al. 2020; Tao et al. 2021). The development of innovative tools, such as pangenome-based genotyping arrays and the expansion of comprehensive *Oryza* genus-wide pangenomes further reinforce the transformative potential of this approach for both functional genomics and practical breeding. Collectively, these advancements indicate that the rice pangenome is emerging as a powerful tool for crop improvement.

Despite this progress, significant obstacles remain before pangenome-informed breeding becomes widespread. Many breeding programs are only beginning to incorporate massive genomics datasets into their decision-making processes. Continued refinement of analytical methods, reductions in costs and capacity building through training and infrastructure investment are essential to transition these advances from research settings into routine breeding practice (Tuggle et al. 2024). A critical emerging challenge lies in the integration of AI and ML approaches into pangenome interpretation and breeding pipelines. AI/ML techniques have shown remarkable potential to enhance the detection and genotyping of complex SVs and PAVs, tasks that are difficult for traditional methods, but several practical gaps currently limit their adoption in plant breeding (Aziz and Masmoudi 2025). These include the lack of breeder-friendly interfaces and decision-support tools, insufficient integration of genotypic, phenotypic and environmental data, scarcity of large, well-annotated datasets for model training, issues of model interpretability and high computational demands that many breeding programs cannot meet. Addressing these AI/ML-specific challenges through the development of accessible, explainable and integrated tools, alongside collaborative

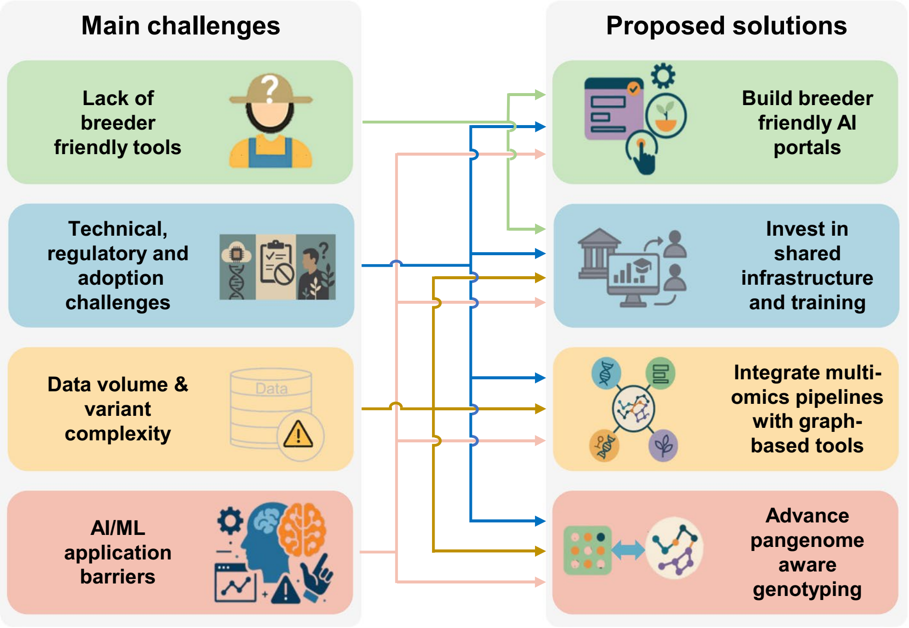

FIGURE 3 | Main challenges and proposed solutions in pangenome-guided rice breeding. This diagram outlines key barriers to integrating pangenomics into practical rice breeding and suggests targeted solutions. The integration of pangenomics into routine breeding remains constrained by several key challenges: the overwhelming volume and complexity of variant data, immature graph-based analytical tools, the absence of breeder-friendly interfaces and the limited adoption of explainable AI models. Addressing these limitations requires integrating automated multi-omics workflows with scalable graph-based tools and pangenome-aware genotyping, developing intuitive AI-driven decision-support portals compatible with breeding databases and investing in shared infrastructure, standardised training and interdisciplinary collaborations.

data-sharing initiatives and infrastructure investment, will be essential to fully harness AI/ML's transformative potential in crop breeding.

Looking ahead, rice breeders are expected to routinely use pangenome panels to select parental lines with complementary novel gene content. They could leverage pangenome-based GWAS to identify critical genomic regions for introgression from exotic donors and consult expansive pangenome databases to guide precise gene editing strategies. Ultimately, the rice pangenome has evolved from an academic concept into a practical asset that promises to enhance molecular breeding (Huang et al. 2021). This transformation offers valuable lessons and best practices with broad applicability, indicating that pangenome-informed strategies could benefit diverse crops beyond rice. While the integration of pangenomic data into breeding pipelines is already underway, a few more years of collaborative research, infrastructure investment and breeder engagement are needed to fully harness this emerging resource. The breakthroughs achieved to date offer a tantalising glimpse of a future where higher-yielding, stress-resilient and disease-resistant rice varieties are developed by exploiting the full spectrum of genetic diversity. In summary, although the journey is ongoing, the fusion of pangenomic data, advanced genotyping platforms and AI/ML-driven analytical tools with modern breeding strategies is well underway, bringing global agriculture closer to realising this transformative promise and providing a framework for other crops to follow. To fully realise these opportunities, stakeholders should focus on breeder-friendly pangenome tools, robust data infrastructure and cross-crop collaborations.

Author Contributions

H.H. and C.K.K.C. wrote the manuscript. S.N., F.L., R.L., J.W., Y.M., C.K.K.C., J.Z. and H.H. contributed to editing the manuscript.

Acknowledgements

We thank Dr. Runxuan Zhang of the James Hutton Institute for providing valuable feedback during manuscript editing.

Funding

This research was supported by National Key R&D Program of China (2024YFD1200801), the National Natural Science Foundation of China (32400512), Guangdong Pearl River Talents Program (2024QN11N336), The GuangDong Basic and Applied Basic Research Foundation (2024A1515011981, 2025A1515012969), Guangdong Provincial Association for Science and Technology Youth Talent Cultivation Program (SKXRC2025531), The 'YouGu' Plan and 'Outstanding youth Researcher' of Rice Research Institute of Guangdong Academy of Agricultural Sciences (2023YG04, 2024YG01, 2024YG07 and 2025YG03), the Introduction of Young Key Talents of Guangdong Academy of Agricultural Sciences (R2023YJ-QC001), Guangzhou Science and Technology Plan Project (2024B03J320).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

Abbai, R., V. K. Singh, V. V. Nachimuthu, et al. 2019. "Haplotype Analysis of Key Genes Governing Grain Yield and Quality Traits Across 3K RG Panel Reveals Scope for the Development of Tailor-Made Rice With Enhanced Genetic Gains." *Plant Biotechnology Journal* 17: 1612–1622.

Aganezov, S., S. M. Yan, D. C. Soto, et al. 2022. "A Complete Reference Genome Improves Analysis of Human Genetic Variation." *Science* 376: eabl3533.

Aziz, M. A., and K. Masmoudi. 2025. "Molecular Breakthroughs in Modern Plant Breeding Techniques." *Horticultural Plant Journal* 11: 15–41.

Bayer, P. E., A. A. Golicz, A. Scheben, J. Batley, and D. Edwards. 2020. "Plant Pan-Genomes Are the New Reference." *Nature Plants* 6: 914–920.

Bayer, P. E., J. Petereit, M. F. Danilevicz, R. Anderson, J. Batley, and D. Edwards. 2021. "The Application of Pangenomics and Machine Learning in Genomic Selection in Plants." *Plant Genome* 14: e20112.

Burridge, A. J., M. Winfield, A. Przewieslik-Allen, et al. 2024. "Development of a Next Generation SNP Genotyping Array for Wheat." *Plant Biotechnology Journal* 22: 2235–2247.

Chen, C., X. Zhang, J. Chen, et al. 2025. "Natural Variation of PH8 Allele Improves Architecture and Cold Tolerance in Rice." *Rice* 18: 1–10.

Chen, E., X. Huang, Z. Tian, R. A. Wing, and B. Han. 2019. "The Genomics of Oryza Species Provides Insights Into Rice Domestication and Heterosis." *Annual Review of Plant Biology* 70: 639–665.

Chen, J., C. Tan, M. Zhu, et al. 2024. "CropGS-Hub: A Comprehensive Database of Genotype and Phenotype Resources for Genomic Prediction in Major Crops." *Nucleic Acids Research* 52: D1519–D1529.

Chen, Z., Q. Bu, G. Liu, et al. 2023. "Genomic Decoding of Breeding History to Guide Breeding-By-Design in Rice." *National Science Review* 10: nwad029.

Chodavarapu, R. K., S. Feng, B. Ding, et al. 2012. "Transcriptome and Methylome Interactions in Rice Hybrids." *Proceedings of the National Academy of Sciences* 109: 12040–12045.

Ćwiek-Kupczyńska, H., T. Altmann, D. Arend, et al. 2016. "Measures for Interoperability of Phenotypic Data: Minimum Information Requirements and Formatting." *Plant Methods* 12: 44.

Daware, A., A. Malik, R. Srivastava, et al. 2023. "Rice Pangenome Genotyping Array: An Efficient Genotyping Solution for Pangenome-Based Accelerated Genetic Improvement in Rice." *Plant Journal* 113: 26–46.

Dong, O. X., S. Yu, R. Jain, et al. 2020. "Marker-Free Carotenoid-Enriched Rice Generated Through Targeted Gene Insertion Using CRISPR-Cas9." *Nature Communications* 11: 1178.

Du, Z.-Z., J.-B. He, and W.-B. Jiao. 2024. "A Comprehensive Benchmark of Graph-Based Genetic Variant Genotyping Algorithms on Plant Genomes for Creating an Accurate Ensemble Pipeline." *Genome Biology* 25: 91.

Eggertsson, H. P., S. Kristmundsdottir, D. Beyter, et al. 2019. "GraphTyper2 Enables Population-Scale Genotyping of Structural Variation Using Pangenome Graphs." *Nature Communications* 10: 5402.

Fornasiero, A., T. Feng, N. Al-Bader, et al. 2025. "Oryza Genome Evolution Through a Tetraploid Lens." *Nature Genetics* 57: 1287–1297.

Garrison, E., J. Sirén, A. M. Novak, et al. 2018. "Variation Graph Toolkit Improves Read Mapping by Representing Genetic Variation in the Reference." *Nature Biotechnology* 36: 875–879.

Ghamkhar, K., F. R. Hay, M. Engbers, H. Dempewolf, and U. Schurr. 2025. "Realizing the Potential of Plant Genetic Resources: The Use of Phenomics for Genebanks." *Plants, People, Planet* 7: 23–32.

Gonnella, G., N. Niehus, and S. Kurtz. 2019. "GfaViz: Flexible and Interactive Visualization of GFA Sequence Graphs." *Bioinformatics* 35: 2853–2855.

Guo, D., Y. Li, H. Lu, et al. 2025. "A Pangenome Reference of Wild and Cultivated Rice." *Nature* 642: 1–10.

Guo, W., M. Schreiber, V. B. Marosi, et al. 2025. "A Barley Pan-Transcriptome Reveals Layers of Genotype-Dependent Transcriptional Complexity." *Nature Genetics* 57: 1–10.

Han, S. K., M. T. McNulty, C. J. Benway, et al. 2023. "Mapping Genomic Regulation of Kidney Disease and Traits Through High-Resolution and Interpretable eQTLs." *Nature Communications* 14: 2229.

Hao, C., Y. Wang, S. Chao, et al. 2017. "The iSelect 9 K SNP Analysis Revealed Polyploidization Induced Revolutionary Changes and Intense Human Selection Causing Strong Haplotype Blocks in Wheat." *Scientific Reports* 7: 41247.

He, H., Y. Leng, X. Cao, et al. 2024. "The Pan-Tandem Repeat Map Highlights Multiallelic Variants Underlying Gene Expression and Agronomic Traits in Rice." *Nature Communications* 15: 7291.

He, W., X. Li, Q. Qian, and L. Shang. 2025. "The Developments and Prospects of Plant Super-Pangenomes: Demands, Approaches, and Applications." *Plant Communications* 6: 101230.

Hu, H., M. F. Danilevicz, C. Li, and D. Edwards. 2024. "Pangenomics and Machine Learning in Improvement of Crop Plants." In *Plant Molecular Breeding in Genomics Era: Concepts and Tools*, 321–347. Springer.

Hu, H., A. Scheben, J. Wang, et al. 2024. "Unravelling Inversions: Technological Advances, Challenges, and Potential Impact on Crop Breeding." *Plant Biotechnology Journal* 22: 544–554.

Hu, H., J. Wang, S. Nie, J. Zhao, J. Batley, and D. Edwards. 2024. "Plant Pangenomics, Current Practice and Future Direction." *Agriculture Communications* 2: 100039.

Hu, H., J. Zhao, W. J. W. Thomas, J. Batley, and D. Edwards. 2025. "The Role of Pangenomics in Orphan Crop Improvement." *Nature Communications* 16: 118.

Huang, C., Z. Chen, and C. Liang. 2021. "Oryza Pan-Genomics: A New Foundation for Future Rice Research and Improvement." *Crop Journal* 9: 622–632.

Jayakodi, M., Q. Lu, H. Pidon, et al. 2024. "Structural Variation in the Pangenome of Wild and Domesticated Barley." *Nature* 636: 654–662.

Jayakodi, M., M. Schreiber, N. Stein, and M. Mascher. 2021. "Building Pan-Genome Infrastructures for Crop Plants and Their Use in Association Genetics." *DNA Research* 28: dsaa030.

Jiao, C., X. Xie, C. Hao, et al. 2025. "Pan-Genome Bridges Wheat Structural Variations With Habitat and Breeding." *Nature* 637: 384–393.

Jonkheer, E. M., D. de Ridr, T. A. van Der Lee, J. R. de Haan, L. Berke, and S. Smit. 2025. "Exploring Intra- and Intergenomic Variation in Haplotype-Resolved Pangenomes." *Plant Biotechnology Journal* 23: 874–886.

Khan, A. W., V. Garg, M. Roorkiwal, A. A. Golicz, D. Edwards, and R. K. Varshney. 2020. "Super-Pangenome by Integrating the Wild Side of a Species for Accelerated Crop Improvement." *Trends in Plant Science* 25: 148–158.

Khan, A. W., V. Garg, S. Sun, et al. 2024. "Cicer Super-Pangenome Provides Insights Into Species Evolution and Agronomic Trait Loci for Crop Improvement in Chickpea." *Nature Genetics* 56: 1225–1234.

Li, F., J. Chen, W. Luo, et al. 2025. "DeepPGDB: A Novel Paradigm for AI-Guided Interactive Plant Genomic Database." *Plant Communications* 6: 10.1016/j.xplc.2025.101494.

Li, W., J. Liu, H. Zhang, et al. 2022. "Plant Pan-Genomics: Recent Advances, New Challenges, and Roads Ahead." *Journal of Genetics and Genomics* 49: 833–846.

Lin, J., S. Wang, P. A. Audano, et al. 2022. "SVision: A Deep Learning Approach to Resolve Complex Structural Variants." *Nature Methods* 19: 1230–1233.

Lin, Y., Y. Zhu, Y. Cui, et al. 2023. "Identification of Natural Allelic Variation in TTL1 Controlling Thermotolerance and Grain Size by a Rice Super Pan-Genome." *Journal of Integrative Plant Biology* 65: 2541–2551.

Lisboa, P. J., S. Saralajew, A. Vellido, R. Fernández-Domenech, and T. Villmann. 2023. "The Coming of Age of Interpretable and Explainable Machine Learning Models." *Neurocomputing* 535: 25–39.

Long, W., Q. He, Y. Wang, et al. 2024. "Genome Evolution and Diversity of Wild and Cultivated Rice Species." *Nature Communications* 15: 9994.

Luo, C., Y. H. Liu, and X. M. Zhou. 2024. "VolcanoSV Enables Accurate and Robust Structural Variant Calling in Diploid Genomes From Single-Molecule Long Read Sequencing." *Nature Communications* 15: 6956.

Ma, H., C. Zhong, D. Chen, H. He, and F. Yang. 2023. "cnnLSV: Detecting Structural Variants by Encoding Long-Read Alignment Information and Convolutional Neural Network." *BMC Bioinformatics* 24: 119.

Mabire, C., J. Duarte, A. Darracq, et al. 2019. "High Throughput Genotyping of Structural Variations in a Complex Plant Genome Using an Original Affymetrix Axiom Array." *BMC Genomics* 20: 1–25.

Mansueto, L., R. R. Fuentes, F. N. Borja, et al. 2017. "Rice SNP-Seek Database Update: New SNPs, Indels, and Queries." *Nucleic Acids Research* 45: D1075–D1081.

Miao, Z., and J.-X. Yue. 2025. "Interactive Visualization and Interpretation of Pangenome Graphs by Linear Reference–Based Coordinate Projection and Annotation Integration." *Genome Research* 35: 296–310.

Mishra, S., A. K. Srivastava, A. W. Khan, L.-S. P. Tran, and H. T. Nguyen. 2024. "The Era of Panomics-Driven Gene Discovery in Plants." *Trends in Plant Science* 29: 995–1005.

Monat, C., B. Pera, M.-N. Ndjiondjop, et al. 2017. "De Novo Assemblies of Three *Oryza glaberrima* Accessions Provide First Insights About Pan-Genome of African Rices." *Genome Biology and Evolution* 9: 1–6.

Murmu, S., D. Sinha, H. Chaurasia, et al. 2024. "A Review of Artificial Intelligence-Assisted Omics Techniques in Plant Defense: Current Trends and Future Directions." *Frontiers in Plant Science* 15: 1292054.

Naithani, S., C. H. Deng, S. K. Sahu, and P. Jaiswal. 2023. "Exploring Pan-Genomes: An Overview of Resources and Tools for Unraveling Structure, Function, and Evolution of Crop Genes and Genomes." *Biomolecules* 13: 1403.

Qin, P., H. Lu, H. Du, et al. 2021. "Pan-Genome Analysis of 33 Genetically Diverse Rice Accessions Reveals Hidden Genomic Variations." *Cell* 184: 3542–3558.

Raza, A., Q. U. Zaman, S. Shabala, et al. 2025. "Genomics-Assisted Breeding for Designing Salinity-Smart Future Crops." *Plant Biotechnology Journal* 23: 3119–3151.

Schatz, M. C., L. G. Maron, J. C. Stein, et al. 2014. "Whole Genome de Novo Assemblies of Three Divergent Strains of Rice, *Oryza sativa*, Document Novel Gene Space of Aus and Indica." *Genome Biology* 15: 1–16.

Schreiber, M., M. Jayakodi, N. Stein, and M. Mascher. 2024. "Plant Pangenomes for Crop Improvement, Biodiversity and Evolution." *Nature Reviews Genetics* 25: 563–577.

Shakoor, N., D. Northrup, S. Murray, and T. C. Mockler. 2019. "Big Data Driven Agriculture: Big Data Analytics in Plant Breeding, Genomics, and the Use of Remote Sensing Technologies to Advance Crop Productivity." *Plant Phenome Journal* 2: 1–8.

Shang, L., X. Li, H. He, et al. 2022. "A Super Pan-Genomic Landscape of Rice." *Cell Research* 32: 878–896.

Sun, C., Z. Hu, T. Zheng, et al. 2017. "RPAN: Rice Pan-Genome Browser for ~3000 Rice Genomes." *Nucleic Acids Research* 45: 597–605.

Talabi, A. O., P. Vikram, S. Thushar, et al. 2022. "Orphan Crops: A Best Fit for Dietary Enrichment and Diversification in Highly Deteriorated Marginal Environments." *Frontiers in Plant Science* 13: 839704.

Tao, Y., D. R. Jordan, and E. S. Mace. 2020. "A Graph-Based Pan-Genome Guides Biological Discovery." *Molecular Plant* 13: 1247–1249.

Tao, Y., H. Luo, J. Xu, et al. 2021. "Extensive Variation Within the Pan-Genome of Cultivated and Wild Sorghum." *Nature Plants* 7: 766–773.

Tao, Y., X. Zhao, E. Mace, R. Henry, and D. Jordan. 2019. "Exploring and Exploiting Pan-Genomics for Crop Improvement." *Molecular Plant* 12: 156–169.

Tuggle, C. K., J. L. Clarke, B. M. Murdoch, et al. 2024. "Current Challenges and Future of Agricultural Genomes to Phenomes in the USA." *Genome Biology* 25: 8.

van Dijk, A. D. J., G. Kootstra, W. Kruijer, and D. de Ridr. 2021. "Machine Learning in Plant Science and Plant Breeding." *iScience* 24: 101890.

Varshney, R. K., A. Bohra, J. Yu, A. Graner, Q. Zhang, and M. E. Sorrells. 2021. "Designing Future Crops: Genomics-Assisted Breeding Comes of Age." *Trends in Plant Science* 26: 631–649.

Varshney, R. K., M. Roorkiwal, S. Sun, et al. 2021. "A Chickpea Genetic Variation Map Based on the Sequencing of 3,366 Genomes." *Nature* 599: 622–627.

Varshney, R. K., V. K. Singh, J. M. Hickey, et al. 2016. "Analytical and Decision Support Tools for Genomics-Assisted Breeding." *Trends in Plant Science* 21: 354–363.

Vourlaki, I.-T., S. E. Ramos-Onsins, M. Pérez-Enciso, and R. Castanera. 2024. "Evaluation of Deep Learning for Predicting Rice Traits Using Structural and Single-Nucleotide Genomic Variants." *Plant Methods* 20: 121.

Wang, C., S. Hu, C. Gardner, and T. Lübberstedt. 2017. "Emerging Avenues for Utilization of Exotic Germplasm." *Trends in Plant Science* 22: 624–637.

Wang, J., H. Hu, X. Jiang, et al. 2024. "Pangenome-Wide Association Study and Transcriptome Analysis Reveal a Novel QTL and Candidate Genes Controlling Both Panicle and Leaf Blast Resistance in Rice." *Rice (N Y)* 17: 27.

Wang, J., W. Yang, S. Zhang, et al. 2023. "A Pangenome Analysis Pipeline Provides Insights Into Functional Gene Identification in Rice." *Genome Biology* 24: 19.

Wang, T., W. He, X. Li, et al. 2023. "A Rice Variation Map Derived From 10 548 Rice Accessions Reveals the Importance of Rare Variants." *Nucleic Acids Research* 51: 10924–10933.

Wang, W., R. Mauleon, Z. Hu, et al. 2018. "Genomic Variation in 3,010 Diverse Accessions of Asian Cultivated Rice." *Nature* 557: 43–49.

Wei, X., M. Chen, Q. Zhang, et al. 2024. "Genomic Investigation of 18,421 Lines Reveals the Genetic Architecture of Rice." *Science* 385: eadm8762.

Wei, X., J. Qiu, K. Yong, et al. 2021. "A Quantitative Genomics Map of Rice Provides Genetic Insights and Guides Breeding." *Nature Genetics* 53: 243–253.

Woldegiorgis, S. T., T. Wu, L. Gao, et al. 2022. "Identification of Heat-Tolerant Genes in Non-Reference Sequences in Rice by Integrating Pan-Genome, Transcriptomics, and QTLs." *Genes* 13: 1353.

Wu, D., L. Xie, Y. Sun, et al. 2023. "A Syntelog-Based Pan-Genome Provides Insights Into Rice Domestication and De-Domestication." *Genome Biology* 24: 179.

Xu, C., Y. Ren, Y. Jian, et al. 2017. "Development of a Maize 55 K SNP Array With Improved Genome Coverage for Molecular Breeding." *Molecular Breeding* 37: 1–12.

Xu, Y., X. Zhang, H. Li, et al. 2022. "Smart Breeding Driven by Big Data, Artificial Intelligence, and Integrated Genomic-Enviromic Prediction." *Molecular Plant* 15: 1664–1695.

Yang, F., H. Kong, J. Ying, et al. 2025. "SeedLLM· Rice: A Large Language Model Integrated With Rice Biological Knowledge Graph." *Molecular Plant* 18: 1118–1129.

Yang, L., W. He, Y. Zhu, et al. 2025. "GWAS Meta-Analysis Using a Graph-Based Pan-Genome Enhanced Gene Mining Efficiency for Agronomic Traits in Rice." *Nature Communications* 16: 3171.

Yu, Z., Y. Chen, Y. Zhou, et al. 2023. "Rice Gene Index: A Comprehensive Pan-Genome Database for Comparative and Functional Genomics of Asian Rice." *Molecular Plant* 16: 798–801.

Yuan, Y., P. E. Bayer, J. Batley, and D. Edwards. 2021. "Current Status of Structural Variation Studies in Plants." *Plant Biotechnology Journal* 19: 2153–2163.

Zhang, F., H. Xue, X. Dong, et al. 2022. "Long-Read Sequencing of 111 Rice Genomes Reveals Significantly Larger Pan-Genomes." *Genome Research* 32: 853–863.

Zhang, Y., G. Huang, Y. Zhao, et al. 2024. "Revolutionizing Crop Breeding: Next-Generation Artificial Intelligence and Big Data-Driven Intelligent Design." *Engineering* 44: 245–255.

Zhao, Q., Q. Feng, H. Lu, et al. 2018. "Pan-Genome Analysis Highlights the Extent of Genomic Variation in Cultivated and Wild Rice." *Nature Genetics* 50: 278–284.

Zheng, Y., and X. Shang. 2023. "SVcnn: An Accurate Deep Learning-Based Method for Detecting Structural Variation Based on Long-Read Data." *BMC Bioinformatics* 24: 213.

Zhou, Y., Z. Yu, D. Chebotarov, et al. 2023. "Pan-Genome Inversion Index Reveals Evolutionary Insights Into the Subpopulation Structure of Asian Rice." *Nature Communications* 14: 1567.

Zhou, Y., Z. Zhang, Z. Bao, et al. 2022. "Graph Pangenome Captures Missing Heritability and Empowers Tomato Breeding." *Nature* 606: 527–534.