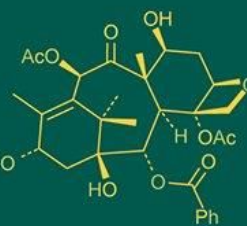
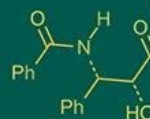


International Journal of Advanced Biochemistry Research



ISSN Print: 2617-4693
ISSN Online: 2617-4707
IJABR 2024; SP-8(7): 568-575
www.biochemjournal.com
Received: 26-06-2024
Accepted: 24-07-2024

Patel Supriya
Acharya N.G. Ranga
Agricultural University,
Tirupati, Andhra Pradesh,
India

Aavula Naveen
Indian Agricultural Research
Institute, New Delhi, India

Thalari Vasanthrao
Professor Jayashankar
Telangana State Agricultural
University, Hyderabad,
Telangana, India

Gampa Mallesh
Indian Agricultural Research
Institute, New Delhi, India

Dharavath Hathiram
Professor Jayashankar
Telangana State Agricultural
University, Hyderabad,
Telangana, India

Hemanth S
Indian Agricultural Research
Institute, New Delhi, India

Danakumara T
Indian Agricultural Research
Institute, New Delhi, India

The role of pan-genomics in understanding genetic diversity and structural variations for developing climate-resilient crops

Patel Supriya, Aavula Naveen, Thalari Vasanthrao, Gampa Mallesh, Dharavath Hathiram, Hemanth S and Danakumara T

DOI: <https://doi.org/10.33545/26174693.2024.v8.i7Sg.2045>

Abstract

Exploring and utilization of genetic variation within the gene pool of modern crop species is a critical step in maintaining and improving crop productivity. Genetic variation, ranging from SNP's (single nucleotide polymorphisms) to large structural variations, can result in variation in the gene content between individuals of the same species. The pan-genome concept was proposed to better capture this variation, as a single reference genome is inadequate to capture the complete genetic diversity of a species. By combining genomic data from multiple accessions, the pan-genome allows for the detection of structural variants, including copy number variations (CNV), presence/absence variation (PAV), inversions and translocations. Advances in sequencing technologies have enabled the assembly of high-quality genome sequences at a low cost, making it possible to construct genus-level pan-genomes across all species. The analysis provides a platform to evaluate the genetic diversity of a species via investigation of its entire genome repertoire and can also be applied to breeding climate resilience into existing crops or to re-domesticate crops, combining environmental adaptation traits with crop productivity. It has been constructed for major crops like rice, maize, Brassica and soybean, this has led to the discovery of valuable genes linked to disease resistance and yield components. High-quality pan-genomes coupled with phenotypic data can identify variant alleles and CRISPR-Cas9 target sites, enhancing editing efficiency. Using a pan-genome graph as a reference aid in comparative analyses for developing climate-resilient, high-performance crops. With diverse species pan-genomes available, we can understand species and higher taxa at the genomic level, offering insights into plant evolution and domestication.

Keywords: Pan-genome, copy number variations, climate-resilience, plant evolution, domestication

Introduction

The genetic variation found in the gene pool of modern crop species is the outcome of a range of changes, including single-nucleotide polymorphisms (SNPs), indels, and large structural variants (SVs). This variation serves as the raw material upon which both human and natural selection can act. To maintain and enhance crop productivity in the face of an increasing world population and climate change, it is essential to understand the processes that cause this variation, assess its extent, and utilize it effectively (Brozynska *et al.*, 2016) [5]. With the availability of reference genomes for most modern crops, genome-wide surveys of SNPs and subsequent marker-trait association studies can now link genetic variation to phenotypic variation. However, our understanding of genetic variation was previously limited due to our reliance on resequencing approaches based on a single reference genome. While some methods may have limited capabilities in identifying structural variations (SVs), mounting evidence suggests that SVs, including copy number variants (CNVs) and presence/absence variants (PAVs), are prevalent in crops and play crucial roles in determining agronomical traits (Springer *et al.*, 2009; Hirsch *et al.*, 2014; Li *et al.*, 2014; Zhao *et al.*, 2014 and Lu *et al.*, 2015) [42, 17, 55].

Genetic variation, particularly structural variation, can result in substantial differences in functional gene complement and gene content between individuals of the same species. To better capture this variation, the concept of the "pan-genome" has been proposed, which involves investigating the entire gene repertoire of a species by sequencing multiple

Corresponding Author:
Patel Supriya
Acharya N.G. Ranga
Agricultural University,
Tirupati, Andhra Pradesh,
India

individuals of the species (Tettelin *et al.*, 2005) [47]. Recent advances in sequencing technology and analysis tools have made it feasible to sequence multiple accessions of crop species (Golicz *et al.*, 2016a) [14]. A wave of pan-genomic studies on major crops and related species, including *Brachypodium distachyon* (Gordon *et al.*, 2017) [16], *Brassica napus* (Hurgobin *et al.*, 2018) [19], *Brassica oleracea* (Golicz *et al.*, 2016b) [14], *Brassica rapa* (Lin *et al.*, 2014), maize (Hirsch *et al.*, 2014) [17], rice (Schatz *et al.*, 2014; Zhao *et al.*, 2015 and Wang *et al.*, 2018) [39, 54, 44], soybean (Li *et al.*, 2014) and wheat (Montenegro *et al.*, 2017) [34] has revolutionized our understanding of genetic diversity within species, revealing a more comprehensive picture of gene content variation. This method has been particularly impactful in crop research, unveiling previously hidden genetic resources that could be valuable for crop improvement and adaptation to changing environments. The growing body of pan-genomic studies across various crop species has provided insights into the genetic basis of important agronomic traits and the evolutionary history of domesticated plants. This pan-genomics has shed new light on the landscape of crop diversity and improvement. In this review, we highlight the progress made in understanding crop pan-genomics, discuss the biological processes that cause SVs, review important agronomic traits affected by SVs, and present our perspective on the application of pan-genomics in crop improvement.

Pan-genome

Pan-genomes were first introduced by Tettelin to describe gene diversity in *Streptococcus agalactiae*. Pan-genomes can be constructed through the sequencing of individual genomes or survey of genetic variations within a species to describe the extensive repertoire of variation. The use of pangenomes removes the sample bias caused by using single reference genome assemblies, allowing the identification of structural variations (SVs) to assess the diversity within species. This diversity can encompass PAVs and non-genic regions (Tettelin *et al.*, 2005) [47].

Classification of genes in pan-genome

The pan-genome represents the full complement of genes within a species, which can be partitioned into a set of Core genome that are shared by all individuals, a set of dispensable genome/Accessory /Variable genome that are partially shared and Unique genome is individual specific (Tettelin *et al.*, 2005) [47].

Types of pan-genome

1. **Open pan-genome:** Open pan-genome approach refers to the condition when there is no fixed quantity of the genes in a species and whenever a new organism is included in the pan-genome analysis, there is a probability of identification of the novel genes.
2. **Closed pan-genome:** In this approach the gene pool is restricted and the pan-genome analyzed based on the definite number of individuals (or samples, strains) would not be extended in terms of gene numbers with the genome of additional individuals analyzed.

Pan-genome assembly methods

Pangenome can be constructed based on genome sequencing data. Genome sequencing reads for genomes A, B and C are shown in the image, each colour represents a gene in the

genome (Figure 1). The genome sequencing reads may be assembled into a pangenome using de novo, iterative and graph-based assemblies which may influence the positioning of the assembled genes. The * indicates that genome A is used as reference genome in iterative assembly method. Pangenomes are mostly constructed in one of three ways. 1) Denovo assembly 2) Iterative mapping 3) Graph based assembly

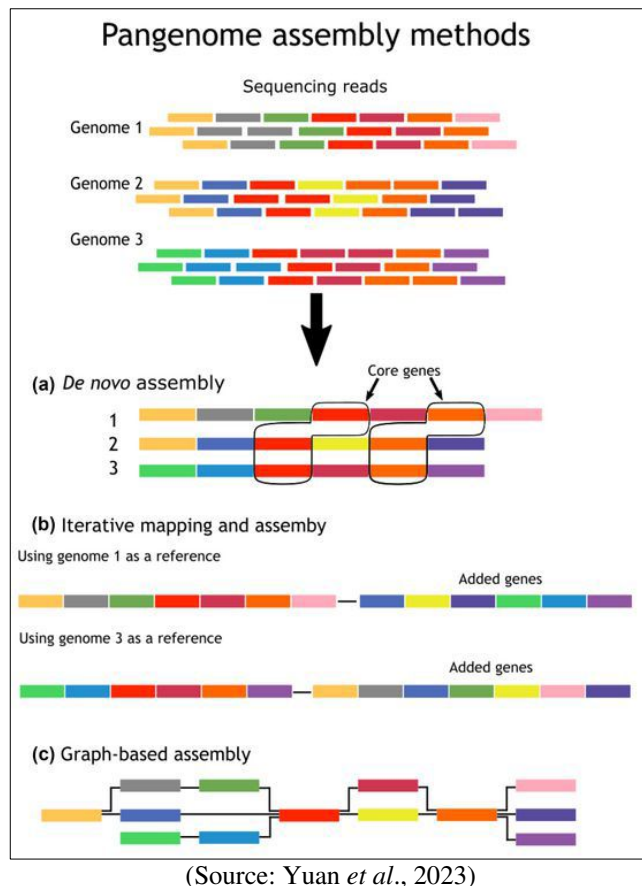


Fig 1: Pan genome construction methods

1. Denovo assembly

The first, de novo sequencing and comparison, involves the sequencing, assembly, and comparison of multiple genomes to identify core and variable genes and genomic regions. This approach reveals the physical position of genes and other genomic elements. However, errors in assembly and annotation may lead to the false calling of variation. Furthermore, this approach is costly and requires high-quality data with high sequencing coverage, limiting the application to relatively few individuals (Lischer & Shimizu, 2017) [30].

2. Iterative mapping

This is the second method for pangenome construction. Iterative mapping, and assembly, uses a single reference genome as a base for the pangenome. Whole genome sequence data for multiple individuals is aligned to the reference genome and any non-aligning sequence reads are assembled and added to the reference to build a pangenome. This approach is less expensive than de novo assembly and comparison as it requires much less data, and so permits the assessment of large numbers of individuals with relatively low sequencing coverage. After pangenome construction, gene PAVs can be called by realigning the sequencing data

from each individual back to the final assembled pangenome.

This approach usually only calls PAV within genes and requires further analysis to accurately place the non-reference contigs within a genomic context. However, a combination of de novo assembly for a small number of representative individuals together with iterative assembly using large numbers of low coverage individuals provides both genomic context and PAV at a population scale, permitting in depth diversity analysis (Fernandez, 2022 and Hu *et al.*, 2020) ^[46].

3. Graph based assembly

This method of assembly uses graph-based approaches, including sequence and variation graphs (VGs) and practical haplotype graphs. Pan genome graphs can be constructed from whole genome assemblies or by de novo graph genome assembly. Instead of a single representative sequence, the graph represents genomic variants as multiple paths. Sequence regions shared between individuals are collapsed into a single path and SVs are added to the graph as a node at the genomic location of their discovery. In, doing so, variant information for dispensable regions are stored as unique paths through the graph, displaying genomic diversity and sequence conservation. Graph-based pangenomes can provide gene position information, but are computationally intensive to assemble, and graph quality correlates directly with the quality of the input data. However, with further advancements in DNA sequencing and data processing, particularly the expansion of high quality long read data, they become standard approach for pangenome construction (Chandra *et al.*, 2024; Cheng *et al.*, 2021; Feng *et al.*, 2020; Garrison *et al.*, 2017) ^[6, 7, 11, 13].

Super pan-genome

Super pan-genome is a pan-genome constructed from the genomes of different species within a genus. Useful to transfer genes from the species belonging to distantly related gene pools. Unifying genomic features by the super pan-genome enables functional and evolutionary studies of genes across different species or populations Khan *et al.* (2020) ^[20].

How a pangenome is generated?

A pangenome is generated by sequencing and analyzing the complete set of genes present across multiple individuals of a species, capturing both the core genome, which is shared by all individuals, and the accessory genome, which includes genes that are present in some but not all individuals. This comprehensive approach allows for the identification of genetic diversity and structural variants, such as insertions, deletions, inversions, and duplications, that contribute to phenotypic variation and adaptation. By studying these structural variants within the pangenome, researchers can gain insights into the evolutionary dynamics and adaptive potential of species, which is crucial for breeding programs aimed at improving crop resilience and productivity.

Structural variants

Structural variation was defined as the genomic alterations which includes insertions, deletions, duplications, inversions and translocations covering at least 1kb. Later adjusted to include small variants Della Coletta *et al.* (2021) ^[10].

Utilizing pan-genomics for crop improvement

- 1. Pan-genomics in utilizing crop wild relatives (CWRs):** Pan genomic studies characterize full genomic content of CWRs uncovering genetic diversity including SVs and SNPs, that defines uniqueness of individual CWRs laying the foundation for utilization of this untapped resource.
- 2. Domestication of new and existing species:** Utilization of pan genome information in identifying causal variants underlying domestication traits. Variation for fruit weight QTL fw3.2 caused by tandem duplication of the cytochrome P450 gene elucidated with the help of tomato pangenome Alonge *et al.* (2020) ^[2]. De novo domestication of wild plants to utilize genetic variation from secondary and tertiary gene pool. Utilization of pan genome information in identifying causal variants underlying domestication traits. The tomato pan-genome has revealed that variation at the fruit weight QTL fw3.2 is caused by tandem duplication of the cytochrome P450 gene rather than a SNP in the gene's promoter as proposed earlier. De novo domestication of wild plants to utilize genetic variation from secondary and tertiary gene pool. Zsogon *et al.* (2018) edited six loci [general plant growth habit (SELFPRUNING), fruit shape (OVATE) and size (FASCIATED and FRUIT WEIGHT 2.2), fruit number (MULTIFLORA), and nutritional quality (LYCOPENE BETA CYCLASE)] in wild tomato (*Solanum pimpinellifolium*) and significantly increased its yield, productivity and nutritional value resulting in de novo domestication of tomato.
- 3. Relevance of transposable elements in crop improvement:** Transposable elements are largely responsible for extensive variation in both intergenic and local genic content. TE (Hopschotch) inserted in a regulatory element of *tb1* was shown to act as an enhancer and inhibits side branching in maize Studer *et al.* (2011) ^[43].
- 4. Easing Mapping and GWAS:** A major concern in QTL mapping and GWAS based on SNPs from a single reference genome is reference bias. Maize gene conferring resistance to sugarcane mosaic virus is identified by GWAS using markers based on B73 but not present in PH207 genome assembly. In, pangenome perspective, the contribution of SVs to trait variation is clear. In Brassica napus, GWAS was performed with PAVs identified from eight whole-genome assemblies, and causal associations between SVs and silique length, seed weight, and flowering time were discovered that were not captured by SNP-GWAS. GWAS conducted using rice pan-genome genotyping array (RPGA) includes 80K genome-wide SNPs genotyping data of a rice diversity panel detected total of 42 loci, regulating grain size/weight traits in rice. (Daware *et al.*, 2022) ^[8].
- 5. Employing Genomic prediction in crop improvement:** Genomic selection is an alternative approach for complex traits controlled by QTLs with small effects. Uses SNPs as predictors for important agronomic traits predicted with fairly high accuracy and it is biased with the use of single reference genome. New approaches for identifying markers within a pangenome frame work are needed to improve prediction accuracy and one such approach is practical

haplotype graph approach (PHG).

6. **Using pangenomes to dissect agronomic traits:** Dispensable genome is enriched with genes involved in environmental responses. Pangenomes are being increasingly used in the detection of sequences associated with agronomically relevant traits, such as yield and stress resistance. This lead to the transition from so-called genomic-assisted to pan genomic-assisted breeding strategies like:
 - a. **Resistance to biotic stress/Disease resistance:** Deletions in the Pi21 gene result in quantitative and durable resistance against blast diseases in rice. Fukuoka *et al.* (2009) [12]. In rapeseed 70% of the predicted resistance genes found in variable proportion of pangenome (53 accessions), which is absent from 'Darmor- bzh' reference Hurgobin *et al.* (2018) [19]. The CNV at the Rhg1 locus is resistant to cystic nematodes in soybean. The sunflower (*Helianthus annuus* L.) pangenome 1.5% of genes were derived from introgression with wild relatives, including two candidate genes associated with resistance to downy mildew: a syntaxin (SYP132) [bacterial resistance] and PR-1 protein secretion, and a GDSL-motif lipase involved in plant resistance to fungal infection. Among the dispensable sequences in the *Brachypodium distachyon* (L.) Beauv. pangenome assembly Brdisv1Bd1 11011965m was identified as a dispensable gene absent from the reference line (Bd21) and induced during infection of wheat stem rust in Bd1-1, a line resistant to *Puccinia graminis* f. sp. tritici
 - b. **Vernalization and flowering time:** The *Brachypodium distachyon* reference (Bd21) has the shortest flowering time recorded in the species, which will be useful for identifying PAVs involved in flower variation. This dispensable gene, Brdisv1ABR21022861m, is present in delayed-and extremely delayed-flowering species. This gene encodes an NF-Y subunit transcription factor, which is a class of transcription factors known to regulate flowering in Arabidopsis, wheat and rice. In Brassica napus, pangenome-based comparative analysis and PAV GWAS identified variations causing several agronomically relevant traits including silique length, seed weight and flowering time (Song *et al.*, 2020) [40].
 - c. **Fruit, grain, yield and seed quality:** The Assembled pepper (*Capsicum* spp.) pangenome (383 cultivars) was used as a reference for PAV GWAS analysis to detect deletions in genes involved in carotenoid and capsaicinoid biosynthetic pathways (Ou *et al.*, 2018). A 2.5-kb deletion in the predicted Pungent gene 1 (Pun1, pan02g021380) was uncovered in 50 cultivars with known low capsaicin content. In tomato, the shorter, non-reference allele in the Tom LoxC promoter, which was abundant in wild relatives, was neglected/negatively selected in domestication, which is responsible for the increase in apocarotenoid levels in the varieties it is identified. In rice, a 1212-bp deletion 5 kb downstream of GW5 causes variation in grain width and weight. (Liu *et al.* 2017) [31]
 - d. **Abiotic stress tolerance and resistance to abiotic stress:** A sorghum pangenome was used to identify drought-responsive genes in transcriptomic data generated from both resistant and susceptible genotypes (Abdel-Ghany *et al.*, 2020; Ruperao *et al.*, 2021; Varoquaux *et al.*, 2019) [49]. A total of 79 genes absent

from the reference genome were confirmed to be differentially expressed during drought stress, including two resistance-specific genes, Sobic.005G069800 and Sobic.006G127800, which were co-mapped with plant height and leaf pigment traits. Analysis of structural variations in a soybean pangenome identified a gene encoding an Fe²⁺/Zn²⁺-regulated transporter associated with iron deficiency chlorosis (Liu *et al.*, 2020) [32]. The PAVs in Sub1A gene encoding ERF like genes-submergence tolerance&2 ERF genes SNORKEL 1&SNORKEL 2 - deep water response in rice (Xu *et al.* 2006).

- e. **Plant architecture:** In wheat, a extra copy of Rht-D1b resulting from a duplication of a >1Mb region causes >70%reduction in plant height (Li *et al.* 2022). In rice, a deletion in qPE9-1, resulted in loss of function of this gene, results in erect panicles.

Complexity of polyploid genomes

High-quality genome assembly of polyploid species has been difficult to achieve because of the inclusion of multiple closely related sub genomes and the associated challenges in discriminating homoeologous loci and creating non-mosaic sub-genome scaffolds. Many researchers have resorted to sequencing diploid progenitors or closely related species. However, closely related diploids fail to capture lineage-specific SNPs, SVs, and other forms of variation that have accumulated post-polyploidization. The degree of structural variation within polyploid species is greater than that within diploid species, and SVs may be particularly useful markers for genomic approaches to polyploid crop improvement. Allopolyploidy (the result of interspecific or intergeneric hybridization and chromosome doubling) and auto polyploidy (the result of whole-genome duplication) are particularly common in plant species (Kumar, 2014; Pozo & Ramirez-Parra, 2015; Qiu *et al.*, 2020; Spoelhof *et al.*, 2017; Trojak-Goluch *et al.*, 2021) [22, 9, 37, 41, 48]. All angiosperms have undergone at least two rounds of polyploidy in their evolutionary history. Many have returned to a diploid state, bearing remnants of this evolutionary history in their genome. As a natural mechanism, polyploidization can increase allelic diversity, expand gene complements, generate novel phenotypic variations, and aid in adaptation to new environments. Taking advantage of this, plant breeders have also generated artificial polyploids, resulting in increased grain yield, fruit size, and seedless fruits. Although polyploid crops are vitally important for sustaining human life, genomic studies of these species have traditionally been very challenging for a number of reasons. Beyond these difficulties in genome assembly, genomic approaches to polyploid crop improvement face further complications: (1) dissection of the genetic architecture of complex traits can be confounded when variants are not mapped to the correct sub genome, which is a technical limitation, and (2) biologically, the more extensive epistatic interactions in polyploids and regulatory feedback between sub genomes can complicate the accurate prediction of phenotypes based on genotype. Advances in sequencing technologies and assembly algorithms have addressed the technical challenges in crop genomic research on polyploids. Long-read sequencing with low error rates (e.g., PacBio HiFi reads) has produced high-quality polyploids.

Editing regulatory elements for crop improvement

CREs are attractive genome editing targets. In tomatoes, inspired by the natural variation observed between wild and domesticated relatives, researchers used CRISPR-Cas9 mutagenesis to generate novel alleles of the promoter of *SICLV3*, a gene affecting fruit size, and engineered a continuum of phenotypic variation. In maize, it was possible to engineer a spectrum of variation for yield-related traits by targeting homologues of *CLV3* (Zhu *et al.*, 2021) [57].

Current developments in pan-genomics

1. RPAN - Rice Pan-Genome Browser developed by Institute of Crop Sciences, Chinese Academy of Agricultural Sciences for 3010 rice accessions.
2. Panache - a tool for the visualization and exploration of linear representations of gene-based and sequence-based pan genomes.
3. Pan X is a software package for comprehensive analysis, interactive visualization and dynamic exploration of bacterial pan-genomes.

Studies on pan-genomics

Bayer *et al.* (2019) [3] performed comparative analysis of RGA candidates in the pangenome of *Brassica oleracea*. In total, 59 RGA candidates were linked to Sclerotinia, clubroot and Fusarium wilt resistance QTL, and these findings have implications for crop breeding in *B. oleracea*, which may also be applicable to other crop species. Normally sequenced reference cultivars do not contain all *B. oleracea* genes because of gene presence/absence variation between individuals. Qin *et al.* (2021) [36] had built a gene-based pan-genome of 32 rice accessions of *Oryza sativa* and 1 accession of *Oryza glaberrima* as inference with Nipponbare genes as the base set and obtained a pan-genome of cultivated rice comprising 66,636 protein-coding genes. Distribution analysis showed that 20,374 genes were categorized as “core genes” 46,262 genes were categorized as “variable genes” which included 14,609 accession-private genes. Varshney *et al.* (2021) [50] constructed the chickpea pan genome (592.58 Mb) using the iterative mapping and assembly approach. They constructed this pan-genome to determine species divergence, migration of the cultigen (*C. arietinum*), rare allele burden, and fitness loss in chickpea. The predicted performance for 100 seed weight increased by up to 23% and 12% with the use of OCS- and haplotype based genomic approaches. Wang *et al.* (2022) [51] had developed the graph pangenomes to overcome the limits of current pangenome identification which will not provide information about the position information of the novel sequences. PAV-based GWAS analysis identified causal structural variations in rice grain weight and plant height, whereas SNP-based GWAS only identified approximate genomic locations. They also identified a new locus (qPH8-1) that was found to be associated with plant height on chromosome 8, which could not be detected using SNP-based GWAS. The pan genomic studies can provide plant geneticists and breeders a comprehensive genome resource. New tools are required to support variation graph assembly, pan-genome construction, and visualization. There is a need to develop an integrated pangenome browser capable of representing SNPs and SVs in multireference coordinate systems for genome analysis. Expanding the pangenome beyond species will increase the use of wild gene sequence diversity in crop improvement.

Future prospects

Pan genomic studies can provide plant geneticists and breeders with a comprehensive genome resource. New tools are required to support variation graph assembly, pan-genome construction, and visualization. An integrated pangenome browser should be developed that is capable of representing SNPs and SVs in a multi-reference coordinate system for genome analysis. Expanding the pangenome beyond species will increase the use of wild gene sequence diversity for crop improvement.

Conclusions

A dispensable genome (structural variants) is not essential for survival. Pangenomes can be used to map dispensable genomes, highlighting the varieties of host genes that provide agronomic benefits. The application of pangenome-wide SNPs, as well as the inclusion of PAV in association analysis, increases the power of genomic associations and the identification of causal variants for agronomic traits. Variants identified from pan-genomes can support genome editing approaches. The adoption of the pangenome graph as a reference facilitates comparative analyses to develop the next generation of climate-resilient high-performance crops. Once pangenomes are available for a large number of diverse species, we can understand how species and higher taxa are defined at the genomic level, providing insights into plant evolution and domestication.

References

1. Abdel-Ghany SE, Ullah F, Ben-Hur A, Reddy AS. Transcriptome analysis of drought-resistant and drought-sensitive Sorghum bicolor genotypes in response to PEG-induced drought stress. International Journal of Molecular Sciences. 2020;21(3):772. <https://doi.org/10.3390/ijms21030772>
2. Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, *et al.* Major impacts of widespread structural variation on gene expression and crop improvement in tomato. Cell. 2020;182(1):145-161. <https://doi.org/10.1016/j.cell.2020.05.027>
3. Bayer PE, Golicz AA, Tirnaz S, Chan CKK, Edwards D, Batley J. Variation in abundance of predicted resistance genes in the Brassica oleracea pangenome. Plant Biotechnology Journal. 2019;17(4):789-800. <https://doi.org/10.1111/pbi.13015>
4. Brockhurst MA, Harrison E, Hall JP, Richards T, McNally A, MacLean C. The ecology and evolution of pangenomes. Current Biology. 2019, 29(20) <https://doi.org/10.1016/j.cub.2019.08.012>
5. Brozynska M, Furtado A, Henry RJ. Genomics of crop wild relatives: expanding the gene pool for crop improvement. Plant Biotechnology Journal. 2016;14(4):1070-1085. <https://doi.org/10.1111/pbi.12449>
6. Chandra G, Jain C, Gibney D. Haplotype-aware sequence alignment to pangenome graphs. Genome Research; c2024. <https://doi.org/10.1101/gr.279143.124>
7. Cheng H, Concepcion GT, Feng X, Li H, Zhang H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nature Methods. 2021;18(2):170-175. <https://doi.org/10.1038/s41592-020-01056-5>
8. Daware A, Srivastava R, Das D, Malik N, Tyagi AK, Parida SK. GWAShub: A Web-based resource to

- democratize genome-wide association studies in crop plants. *bioRxiv*; c2022.
<https://doi.org/10.1101/2022.02.17.480880>
9. Del Pozo JC, Ramirez-Parra E. Whole genome duplications in plants: an overview from Arabidopsis. *Journal of Experimental Botany*. 2015;66(22):6991-7003. <https://doi.org/10.1093/jxb/erv432>
 10. Della Coletta R, Qiu Y, Ou S, Hufford MB, Hirsch CN. How the pan-genome is changing crop genomics and improvement. *Genome Biology*. 2021;22(1):1-19. <https://doi.org/10.1186/s13059-021-02427-5>
 11. Fukuoka S, Yamamoto SI, Mizobuchi R, Yamanouchi U, Ono K, Kitazawa N, *et al.* Multiple functional polymorphisms in a single disease resistance gene in rice enhance durable resistance to blast. *Scientific Reports*. 2014;4(1):4550. <https://doi.org/10.1038/srep04550>
 12. Golicz AA, Bayer PE, Bhalla PL, Batley J, Edwards D. Pan genomics comes of age: from bacteria to plant and animal applications. *Trends in Genetics*. 2020;36(4):268-279. <https://doi.org/10.1016/j.tig.2019.12.006>
 13. Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, Shu S, *et al.* Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure. *Nature Communications*. 2017;8(1):2184. <https://doi.org/10.1038/s41467-017-02292-8>
 14. Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, *et al.* Insights into the maize pan-genome and pan-transcriptome. *The Plant Cell*. 2014;26(1):121-35. <https://doi.org/10.1105/tpc.113.119982>
 15. Hu H, Scheben A, Fernandez CT, Edwards D, Bayer PE, Golicz AA, *et al.* Legume pangenome construction using an iterative mapping and assembly approach. *Methods in Molecular Biology* (Clifton, N.J.). 2020;2107:35-47. https://doi.org/10.1007/978-1-0716-0236-2_3
 16. Hurgobin B, Golicz AA, Bayer PE, Chan CKK, Tirnaz S, Dolatabadian A, *et al.* Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid Brassica napus. *Plant Biotechnology Journal*. 2018;16(7):1265-1274. <https://doi.org/10.1111/pbi.12867>
 17. Khan AW, Garg V, Roorkiwal M, Golicz AA, Edwards D, Varshney RK. Super-pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends in Plant Science*. 2020;25(2):148-158. <https://doi.org/10.1016/j.tplants.2019.10.012>
 18. Khan M, Khan H, Khan S, Nawaz M. Epidemiological and clinical characteristics of coronavirus disease (COVID-19) cases at a screening clinic during the early outbreak period: A single-centre study. *Journal of Medical Microbiology*. 2020;69(8):1114-1123. <https://doi.org/10.1099/jmm.0.001232>
 19. Li A, Hao C, Wang Z, Geng S, Jia M, Wang F, *et al.* Wheat breeding history reveals synergistic selection of pleiotropic genomic sites for plant architecture and grain yield. *Molecular Plant*. 2022;15(3):504-519. <https://doi.org/10.1016/j.molp.2021.12.021>
 20. Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, *et al.* Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature Genetics*. 2014;46(6):567-572. <https://doi.org/10.1038/ng.2987>
 21. Li YH, Zhou G, Ma J, Jiang W, Jin LG, Zhang Z, *et al.* De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology*. 2014;32(10):1045-52. <https://doi.org/10.1038/nbt.2979>
 22. Lin K, Zhang N, Severing EI, Nijveen H, Cheng F, Visser RG, *et al.* Beyond genomic variation-comparison and functional annotation of three Brassica rapa genomes: a turnip, a rapid cycling and a Chinese cabbage. *BMC Genomics*. 2014;15(1):250. <https://doi.org/10.1186/1471-2164-15-250>
 23. Kumar S. Molecular tools for exploring polyploid genomes in plants. Apple Academic; c2014. p. 103-128. <https://doi.org/10.1201/b16568-10>
 24. Li A, Hao C, Wang Z, Geng S, Jia M, Wang F, Han X, *et al.* Wheat breeding history reveals synergistic selection of pleiotropic genomic sites for plant architecture and grain yield. *Molecular Plant*. 2022;15(3):504-519. <https://doi.org/10.1016/j.molp.2021.12.021>
 25. Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, *et al.* Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature Genetics*. 2014;46(6):567-572. <https://doi.org/10.1038/ng.2987>
 26. Li R, Yang H, Zhu H, Kristiansen K, Shi Z, Wang J, *et al.* De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research*. 2009;20:265-272. <https://doi.org/10.1101/gr.097261.109>
 27. Li Y, Bolund L, Hu Y, Wang J. State of the art de novo assembly of human genomes from massively parallel sequencing data. *Human Genomics*. 2010;4:271. <https://doi.org/10.1186/1479-7364-4-4-271>
 28. Li Y, Ma H, Li R, Huang S, Wu B, Wu H, *et al.* Structural variation in two human genomes mapped at single-nucleotide resolution by whole genome de novo assembly. *Nature Biotechnology*. 2011;29:723-730. <https://doi.org/10.1038/nbt.1904>
 29. Li YH, Zhou G, Ma J, Jiang W, Jin LG, Zhang Z, *et al.* De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology*. 2014;32(10):1045-1052. <https://doi.org/10.1038/nbt.2979>
 30. Lin K, Zhang N, Severing EI, Nijveen H, Cheng F, Visser RG, *et al.* Beyond genomic variation-comparison and functional annotation of three Brassica rapa genomes: a turnip, a rapid cycling and a Chinese cabbage. *BMC Genomics*. 2014;15(1):1-17. <https://doi.org/10.1186/1471-2164-15-250>
 31. Lischer HEL, Shimizu KK. Reference-guided de novo assembly approach improves genome reconstruction for related species. *BMC Bioinformatics*; c2017 .p. 18. <https://doi.org/10.1186/s12859-017-1911-6>
 32. Liu J, Chen J, Zheng X, Wu F, Lin Q, Heng Y, *et al.* GW5 acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. *Nature Plants*. 2017;3(5):1-7. <https://doi.org/10.1038/nplants.2017.43>
 33. Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, *et al.* Pan-genome of wild and cultivated soybeans. *Cell*. 2020;182(1):162-176. <https://doi.org/10.1016/j.cell.2020.05.023>
 34. Lu F, Romay MC, Glaubitz JC, Bradbury PJ, Elshire RJ, Wang T, *et al.* High-resolution genetic mapping of

- maize pan-genome sequence anchors. *Nature Communications*. 2015;6(1):1-8. <https://doi.org/10.1038/ncomms7914>.
35. Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Chan CKK, *et al.* The pangenome of hexaploid bread wheat. *The Plant Journal*. 2017;90(5):1007-1013. <https://doi.org/10.1111/tpj.13515>.
 36. Ou L, Li D, Lv J, Chen W, Zhang Z, Li X, *et al.* Pan-genome of cultivated pepper (*Capsicum*) and its use in gene presence-absence variation analyses. *New Phytologist*. 2018;220(2):360-363. <https://doi.org/10.1111/nph.15413>.
 37. Qin P, Lu H, Du H, Wang H. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell*. 2021;184(13):3542-3558. <https://doi.org/10.1016/j.cell.2021.05.016>.
 38. Qiu T, Liu B, Liu Z. The effects of hybridization and genome doubling in plant evolution via allopolyploidy. *Molecular Biology Reports*. 2020;47(7):5549-58. <https://doi.org/10.1007/s11033-020-05597-y>.
 39. Ruperao P, Thirunavukkarasu N, Gandham P, Selvanayagam S, Govindaraj M, Nebie B, *et al.* Sorghum pan-genome explores the functional utility for genomic-assisted breeding to accelerate the genetic gain. *Frontiers in Plant Science*. 2021;12:666342. <https://doi.org/10.3389/fpls.2021.666342>.
 40. Schatz MC, Maron LG, Stein JC, Wences AH, Gurtowski J, Biggers E, *et al.* Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of aus and indica. *Genome Biology*. 2014;15(11):1-16. <https://doi.org/10.1186/s13059-014-0506-z>.
 41. Song JM, Guan Z, Hu J, Guo C, Yang Z, Wang S, *et al.* Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nature Plants*. 2020;6(1):34-45. <https://doi.org/10.1038/s41477-019-0585-3>.
 42. Spoelhof JP, Soltis DE, Soltis PS. Pure polyploidy: Closing the gaps in autopolyploid research. *Journal of Systematics and Evolution*. 2017;55(4):340-52. <https://doi.org/10.1111/jse.12253>.
 43. Springer NM, Ying K, Fu Y, Ji T, Yeh CT, Jia Y, *et al.* Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genetics*. 2009, 5(11). <https://doi.org/10.1371/journal.pgen.1000734>.
 44. Studer A, Zhao Q, Ross-Ibarra J, Doebley J. Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nature Genetics*. 2011;43(11):1160-1063. <https://doi.org/10.1038/ng.940>.
 45. Tao Y, Zhao X, Mace E, Henry R, Jordan D. Exploring and exploiting pan-genomics for crop improvement. *Molecular Plant*. 2019;12(2):156-169. <https://doi.org/10.1016/j.molp.2018.12.004>.
 46. Tay Fernandez C. Making a Pangenome Using the Iterative Mapping Approach. *Methods in Molecular Biology*. 2022;2443:259-271. https://doi.org/10.1007/978-1-0716-1979-8_14.
 47. Tay Fernandez CG, Nestor BJ, Danilevicz MF, Marsh JJ, Petereit J, Bayer PE, *et al.* Expanding Gene-Editing Potential in Crop Improvement with Pangenomes. *International Journal of Molecular Sciences*. 2022;23(4):2276. <https://doi.org/10.3390/ijms23042276>.
 48. Tettelin H, Massignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, *et al.* Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences*. 2005;102(39):13950-13955. <https://doi.org/10.1073/pnas.0506758102>.
 49. Trojak-Goluch A, Kawka-Lipińska M, Wielgus K, Praczyk M. Polyploidy in Industrial Crops: Applications and Perspectives in Plant Breeding. *Agronomy*. 2021;11(12):2370. <https://doi.org/10.3390/agronomy11122370>.
 50. VanBuren R, Wai CM, Colle M, Wang J, Sullivan S, Kwon M, *et al.* A near complete, chromosome-scale assembly of the black raspberry (*Rubus occidentalis*) genome. *Gigascience*. 2018;7(8):1-7. <https://doi.org/10.1093/gigascience/giy094>.
 51. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*. 2017;27(5):737-746. <https://doi.org/10.1101/gr.214270.116>.
 52. Walkowiak S, Gao L, Monat C, Haberer G, Kassa MT, Brinton J, *et al.* Multiple wheat genomes reveal global variation in modern breeding. *Nature*. 2020;588(7837):277-283. <https://doi.org/10.1038/s41586-020-2961-x>.
 53. Wendel JF, Jackson SA, Meyers BC, Wing RA. Evolution of plant genome architecture. *Genome Biology*. 2016;17(1):1-14. <https://doi.org/10.1186/s13059-016-0908-1>.
 54. Xiao S, Jiang L, Chen Z, Jiang L, Ma Y, Li M, *et al.* The genetic mechanism of heterosis utilization in *Brassica napus* breeding revealed by heterotic groups. *Theoretical and Applied Genetics*. 2021;134(12):3939-3951. <https://doi.org/10.1007/s00122-021-03914-3>.
 55. Younis A, Ramzan F, Hameed M, Lim KB. RNA interference (RNAi) induced gene silencing: A promising approach of hi-tech plant breeding. *International Journal of Biological Sciences*. 2014;10(10):1150-1158. <https://doi.org/10.7150/ijbs.10452>.
 56. Yu Y, Ouyang Y, Yao W, Wang L, Xu C, Xing Y, *et al.* Transcriptome divergence of cultivated and wild rice controlled by cis-regulatory elements with a transposable element origin. *Genome Research*. 2017;27(10):1775-1786. <https://doi.org/10.1101/gr.216820.116>.
 57. Zhang Y, Zhang F, Xie Q, Han S, Gao J, Huang X, *et al.* Long-read sequencing and pan-genome analysis reveal genomic diversity and structural variations in *Salvia* species. *Plant Biotechnology Journal*. 2022;20(6):1094-1109. <https://doi.org/10.1111/pbi.13788>.
 58. Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, *et al.* Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nature Genetics*. 2018;50(2):278-284. <https://doi.org/10.1038/s41588-018-0041-z>.
 59. Zhao Q, Li Y, Zhou H, Ren F, Wang J, He R, *et al.* Pan-genome identification of long intergenic non-coding RNA in *Oryza sativa*. *Molecular Plant*. 2020;13(2):351-353. <https://doi.org/10.1016/j.molp.2019.11.006>.
 60. Zhou Q, Zhong Z, Duan L, Wang Q, Zhang C, Zhu Z,

et al. A quadratic regression-based iterative clustering strategy for high-resolution quantification of *Salvia miltiorrhiza* metabolites by liquid chromatography/tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*. 2020, 34(8). <https://doi.org/10.1002/rcm.8672>.