

Review

Not peer-reviewed version

Exploring Pan-Genomes of Crops: Genomic Resources and Tools for Unraveling Gene Evolution, Function, and Adaptation

[Sushma Naithani](#)^{*}, [Cecilia H. Deng](#), [Sunil Kumar Sahu](#), [Pankaj Jaiswal](#)

Posted Date: 23 August 2023

doi: 10.20944/preprints202308.1641.v1

Keywords: pan-genomes; comparative genomics; plant pathways; genomic databases; gravitropism; Gene Ontology



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

Exploring Pan-Genomes of Crops: Genomic Resources and Tools for Unraveling Gene Evolution, Function, and Adaptation

Sushma Naithani ^{1,*}, Cecilia H. Deng ^{2,†}, Sunil Kumar Sahu ^{3,†} and Pankaj Jaiswal ¹

¹ Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, 97333, USA

² The New Zealand Institute for Plant and Food Research Limited, Auckland, New Zealand

³ State Key Laboratory of Agricultural Genomics, BGI Research, Shenzhen 518083, China

* Correspondence: sushma.naithani@oregonstate.edu

Abstract: The availability of multiple sequenced genomes from a single species made it possible to explore intra- and inter-specific genomic comparisons at higher resolution and build clade-specific pangenomes of several crops. The pan-genomes of crops constructed from various cultivars/accessions, landraces, and wild ancestral species represent a compendium of genes and structural variations and allow researchers to search for the novel genes and alleles that were inadvertently lost in domesticated crops during the historical process of crop domestication or in the process of extensive plant breeding. Fortunately, many valuable genes and alleles associated with desirable traits like disease resistance, abiotic stress tolerance, plant architecture, and nutrition qualities exist in landraces, ancestral species, and crop wild relatives. The novel genes from the wild ancestors and landraces can be introduced back to high-yielding varieties of modern crops by implementing classical plant breeding, genomic selection, and transgenic/gene editing approaches. Thus, pan-genomic represents a great leap in plant research and offers new avenues for targeted breeding for mitigating the impact of global climate change. Here we summarize the tools used for pangenome assembly and annotations, web-portals hosting plant pangenomes. Furthermore, we highlight a few discoveries made in crops using the pan-genomic approach and its future potential.

Keywords: pan-genomes; comparative genomics; plant pathways; gene annotation; Gene Ontology; gravitropism

1. Introduction

In recent years, advancements in affordable sequencing platforms and computational resources have helped to generate reference-quality genome assemblies for multiple crop accessions/varieties belonging to a single species. Thus, crop genomics has transitioned to the pan-genomic era. This shift has been made possible by advances in both low-cost sequencing (i.e., short-read data Illumina sequencing, long-read PacBio sequencing with low errors, Nanopore sequencing) and bioinformatics tools employed for genomic data processing and genome assemblies (for excellent reviews, see [1–7]). Whole-genome assembly-based pangenomes have been reported for a few plant species, including rice [8–10], barley [11], wheat [12], maize [13], soybean [14], a wild relative of soybean *Glycine soja* [15], and brassica [16]. In particular, PacBio HiFi reads proved very useful for high-quality polyploid genome assemblies for peanut [17], wheat [18], oilseed [16], strawberry [19] and potato [20].

The intraspecies genome comparisons of crops suggest extensive structural variation across diverse genotypes that affect both the genomic contents and plant function [14,21–23]. The structural variations (SVs) among genotypes of the same species include insertion, deletions (indels), and translocation of the small or large genomic regions that further cause presence-absence variation (PAV), copy number variation (CNV), chromosomal rearrangements, repetitive sequences (i.e., tandem gene duplications, repetitive sequences in non-coding regions of the genome, transposable elements, centromere repeats, etc.). The conserved genes present in all cultivars/genotypes/subspecies/strains

within a species constitute the “core” genome, and the variable genes represent as the “dispensable” or “accessory” genome [2,24]. The accessory genome consists of the “shell genes” (found in most cultivars within a species) and the “cloud genes” (present in only a small fraction of cultivars of a same species) as shown in Figure 1.

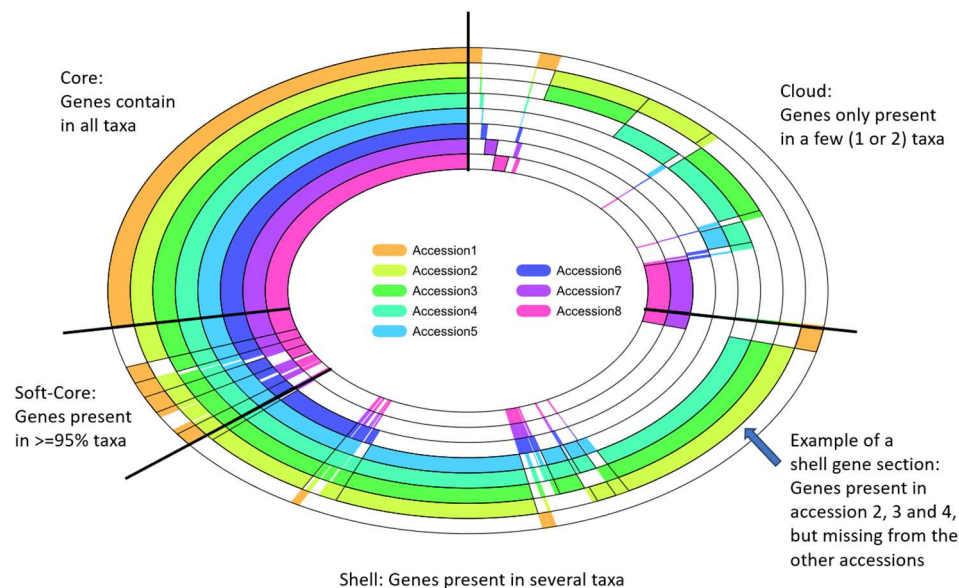


Figure 1. A conceptual depiction of the pan-genes across genomes of eight accessions belonging to the same clade. Each ring represents one accession. The top-left “core genes” represent conserved genes across eight accessions. The white section in a ring indicates the absence of ortholog(s). The “soft cores” represents genes found in $\geq 95\%$ of accessions. The “cloud genes” found only in one or two taxa. The rest between the “cloud genes” and “soft core genes” are “shell genes”.

Thus, a single reference genome represents only a fraction of the species-wide genomic space, and a pan-genome represents species-wide genomic space [2,13,15,25,26]. Often, the pangenomes encompass genes found in crop wild relatives and ancestral species [27–31]. We like to note here that many useful genes which were lost during the crop domestication and extensive plant breeding [32–34] may be found in the “dispensable”/accessory genome of any crop species which are needed for adaptation to a range of biotic and abiotic stress conditions, and improving physiology, plant structure, yield and nutrient quality of crops [15,33,35–39]. Thus, the availability of plant pangenomes allows plant researchers and breeders to explore important candidate genes for improving yield and nutritional quality, and resilience of crops to a range of climatic conditions and pathogens. For instance, a few comparative genomic studies have revealed that gene amplification plays a vital role in disease resistance, abiotic stress tolerance, and other agronomic traits associated with plant development, architecture, yield, etc. [40–49]. The high-quality pangenomes also make it possible to study previously inaccessible regions of the eukaryotic genomes, including centromeres, long heterochromatic blocks, and rDNA regions, etc. that exhibit low recombination, and provide new insights into crop genome evolution [50].

Recently, many excellent reviews have been published on plant pan-genomes (see [1,2,51–54]. Here, we review the current strategies for constructing and visualizing crop pan-genome data, public genomic portals hosting pan-genes, pan-genome data, pan-genome browsers, and associated analytical tools. Furthermore, we highlight a few studies that have exploited a pan-genomic approach to discovering candidate genes associated with important agronomic traits. We also discuss the potential of pan-genome-driven translational research.

2. Pangenome Construction, Visualization, and Data Analysis Tools

The first step in setting up a pangenome infrastructure is the selection of a diverse set of representative genotypes for sequence assembly that capture as many genetic variants as possible with a limited panel of genotypes [14,21,55,56]. The second step is sequencing of individual genomes. The third step is assembly and construction of pan-genome. Several popular approaches are implemented for pan-genome assembly (for detailed reviews, see [1,57–59]. Here we briefly describe the basic tenets of three popular methods (see Figure 2).

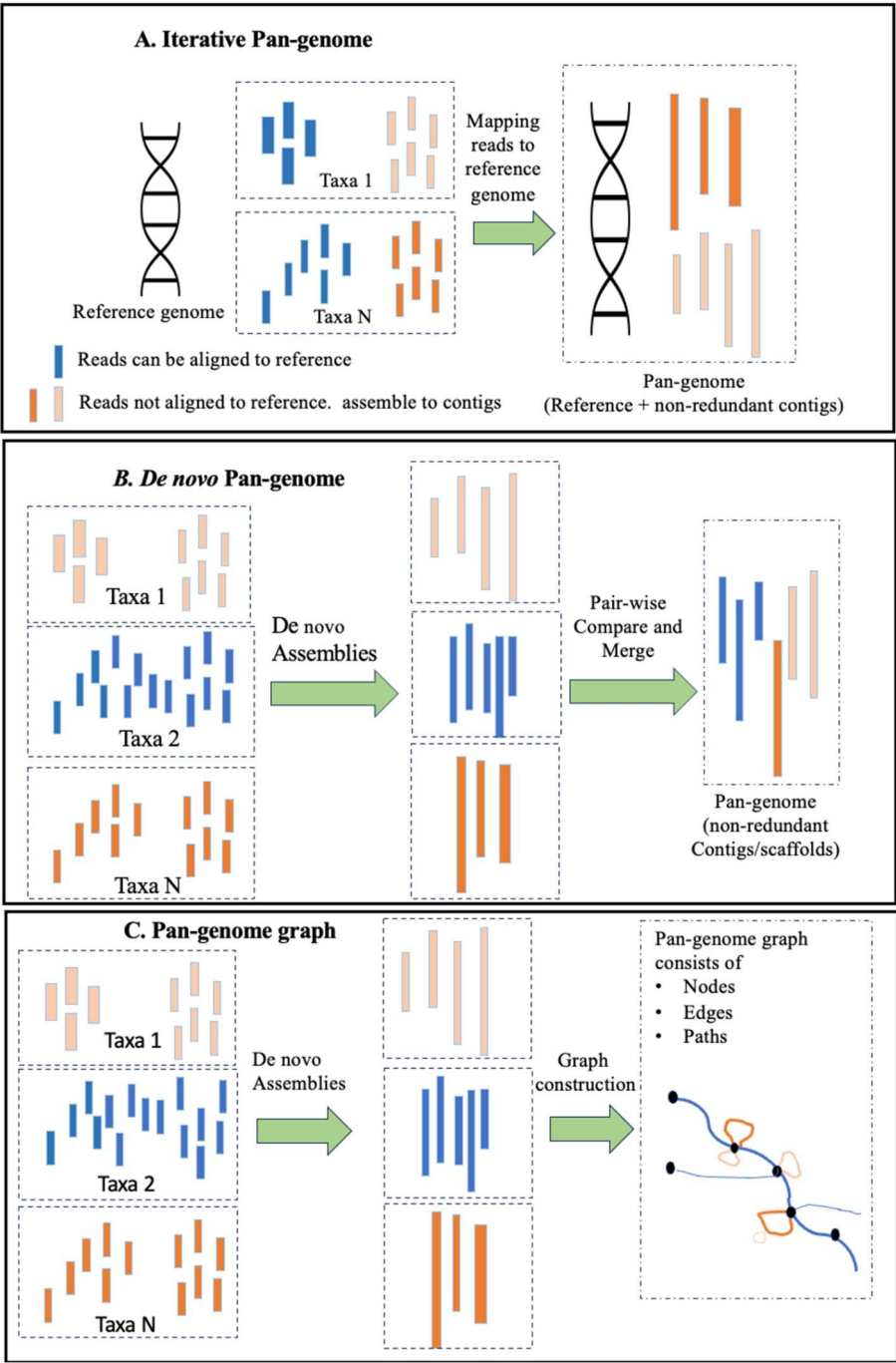


Figure 2. An illustration of three popular approaches currently used for pan-genome construction are the reference-based iterative method, de novo genome assembly, and graph-based pangenome assembly.

The first approach uses a high-quality reference genome for mapping sequence reads generated from all other genotypes. Iterative refinement allows for a progressive improvement of the assembly

with additional data. This strategy can minimize errors by exploiting the information from a high-quality reference genome and limiting the coordinate consolidation issue. However, this method requires the availability of a high-quality reference genome, which may not be available for all species or strains. Secondly, it is sensitive to misalignment errors or inaccuracies in the reference genome, which can potentially propagate errors throughout the assembly. Likewise, bias towards the reference genome may limit the detection of novel or divergent sequences.

In species where a reference is unavailable, *de novo* assembly of individual genotypes is generated, and then assembled contigs are mapped to each other [27]. The *de-novo* genome assemblies have become a method of choice due to the advances in long-read sequencing and the availability of fast algorithms for aligning long-reads to call structural variants [7].

Conceptually, *de novo* assembly of multiple high-quality reference genome sequences and their comparison by pair-wise sequence alignment is arguably the most powerful and accurate approach to detect all types of sequence variants at base-level resolution, including novel genomic elements and rearrangements. However, generating assemblies of polyploid plant genomes is still a challenge, and current methods are limited in detecting and phasing heterozygous structural variants, erroneously produce chimeric contigs joining different haplotypes or ignore alternative haplotypes [60,61]. This approach is time-intensive and requires significant computational and bioinformatic resources, especially for large genomes and complex variations. Repeat resolution can be challenging, leading to fragmented assemblies, as it relies heavily on sequencing depth to overcome repetitive regions and complex variations.

The third graph-based approach allows the addition of any variant to the reference as a node at the genomic location where it is discovered, and then haplotypes are associated with one of the reference genomes used to build the graph. Reads are then realigned to this genome leading to more accurate mapping. This method can accommodate new genomic data through iterative refinement, allowing for continuous improvement of the pangenome assembly. However, graph construction and traversal can be computationally intensive, especially for large and diverse pangenomes, and requires substantial computational resources. Typically, graph complexity increases with the addition of more genomes, potentially impacting scalability, and computational efficiency. Nonetheless, the graph-based pan-genomes can represent complex variations, including structural variants and large-scale rearrangements, facilitating identification of shared and unique genomic regions among individuals or strains, and aiding in excellent visualization of pan-genomes. A conceptual visualization of the pan-genome is shown in Figure 3.

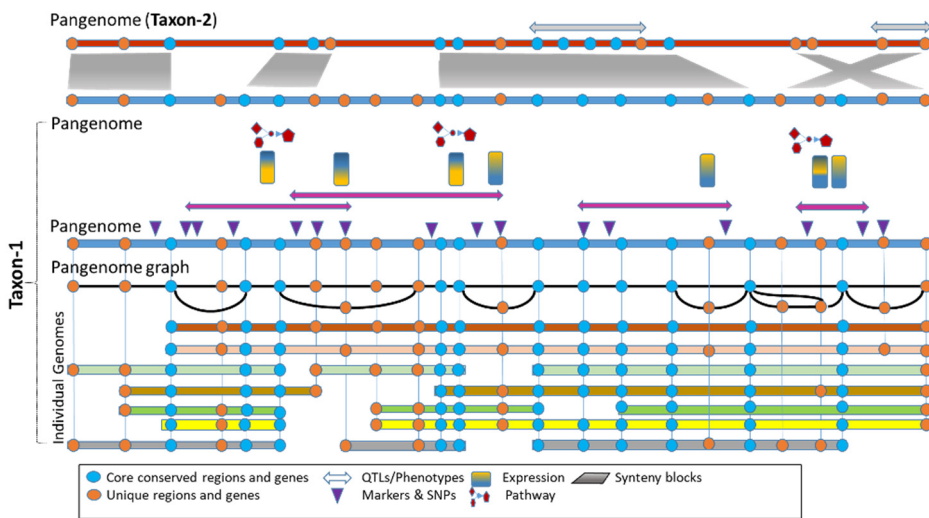


Figure 3. A conceptual view of pan-genome data and its applications in creating a common reference set of genome graph carrying all the possible chromosomal rearrangements and mapped features along with whole-genome alignments, gene orthology, expression, pathways, function, and aligned synteny views to help accelerate the knowledge discovery and hypothesis-driven research.

Recently, Shang et al., 2022 [21] have constructed ‘Super Pan-genome of rice’ containing high-quality assemblies of 251 rice genomes including 202 accessions of domesticated Asian rice *Oryza sativa*, 28 accession of a *Oryza rufipogon* (the wild ancestor of *O. sativa*), 11 accessions of domesticated African rice *Oryza glaberrima*, and ten accessions of *Oryza barthii* (the wild ancestor of *O. glaberrima*) using de novo long-read assembly and a graph-based approach. A reference-free whole-genome multiple sequence alignment for the 251 rice accessions that can be visualized using any assembly as the reference with the Rice Super Pan-genome Information Resource Database (<http://www.ricesuperpir.com>). This fully annotated pan-genome graph, visualized using the JBrowse genome browser, also integrated the SVs, gene annotations, transposable elements annotations, pan-genome graph, and BLAST tools [21].

A few excellent reviews have previously been published on the development of computational tools for pan-genome visualizations [57,58,62,63]. We note here that genome sequencing technologies and assembly algorithms, as well as the pan-genome visualization tools, are rapidly evolving to achieve high accuracy and are being complemented by additional independent mapping approaches, such as optical maps and Hi-C to validate structural variant calls (i.e., inversions and translocations). It is important to acknowledge that the details and outcomes of each method may vary based on the specific pangenome assembly tools, parameters, and characteristics of the genomic data employed in the process [64,65]. Here we compiled a summary of the latest representative tools in Table 1. It is crucial to recognize that the development of pan-genome tools is an active field, and the list could not be exhaustive.

Table 1. A list of popular open-source tools for pan-genome assembly and visualization.

Tool name and URL	Remarks and Citation
Genome assembly	
Hifiasm https://github.com/chhylp123/hifiasm	Construct haplotype-resolved assemblies from accurate HiFi Reads [60]
Canu https://github.com/marbl/canu	Assemble genomes of any size from single molecule sequences and provides graphical fragment assembly that can be integrated with complementary phasing and scaffolding methods [66].
Flye https://github.com/fenderglass/Flye	Assemble single molecule, long-read sequencing data into genomes using repeat graphs [67].
PAGIT https://www.sanger.ac.uk/tool/pagit	PAGIT is a package of tools for generating high-quality draft genome sequences by ordering contigs, closing gaps, correcting sequence errors, and transferring annotation. PAGIT is compiled for Linux/UNIX systems and is available as a virtual machine [68].
MEGAHIT https://github.com/voutcn/megahit	Ultra-fast NGS assembler for metagenomes [69].
SPADes https://cab.spbu.ru/software/spades/	A set of genome assembly and analysis tools for long- and short-reads [70,71].
Pangenome graph construction, normalization, identification of structural variants, and visualization	
vgtools (vg construct, 'vg call', 'vg giraffe', 'vg map', or 'vg mpmap') https://github.com/vgteam/vg	Toolset for eukaryotic pangenome graph construction, read mapping, variant calling, and graph visualization [58].
Minigraph https://github.com/lh3/minigraph	Tool for graph construction, mapping, and variant calling [72].
PGGB https://github.com/pangenome/pggb	Pan-genome graph construction, normalization, and visualization, using ODGI as the backbone [73].
MGRgraph https://github.com/LeilyR/Multi-genome-Reference	An Algorithm to Build a Multi-genome.
Cactus https://github.com/ComparativeGenomicsToolkit/cactus	A reference-free multiple genome alignment program that can use progressive mode to build pan-genome across different species [74,75]
PanTools https://pantools.readthedocs.io/en/latest/user_guide/install.html	A platform for pan-genome graph construction, read mapping, phylogeny analysis, pan-graph query, and pan-gene annotation [76].
Smoothxg	Local re-construction of variation graphs

https://github.com/pangenome/smoothxg	
ODGI https://github.com/pangenome/odgi	Optimized Dynamic Genome/Graph Implementation (ODGI) is a tool suite that represents graphs, including structurally complex regions with minimal memory overhead [77]. It is a pangenome toolbox with more than 30 tools to transform, analyze, simplify, validate, annotate, and visualize pangenome graphs.
nf-core/pangenome https://github.com/nf-core/pangenome	Nextflow pipeline for all-vs-all alignment, pangenome graph construction, normalization, remove redundancy, and visualization (through ODGI) [78].
SeqWish https://github.com/ekg/seqwish	Build a variation graph from pairwise alignments [73].
PanPipe https://github.com/USDA-ARS-GBRU/PanPipes	End-to-end pan-genomic graph construction and genetic analysis pipeline [79].
PanGene https://github.com/lh3/pangene	It is used for ortholog and paralog analysis and building pan-gene graphs.
Minimap2 https://github.com/lh3/minimap2	This alignment program can be used for mapping DNA or long mRNA sequences to a reference [80].
NGMLR https://github.com/philres/ngmlr	This program align PacBio long reads to genomes and for detecting complex structural variations [5].
MUMmer4 https://mummer.sourceforge.net/ https://github.com/mummer4/mummer	This is a genome to genome aligner [81].
GraphAligner https://github.com/maickrau/GraphAligner	A tool for aligning long reads to genome graphs [82].
V-ALIGN https://github.com/tcsatc/V-ALIGN	V-ALIGN allows gapped sequence alignment directly on the input graph and supports affine and linear gaps [83].
PaSGAL https://github.com/ParBLISS/PaSGAL	Parallel Sequence to Graph Aligner (PaSGAL) facilitates local sequence alignment of sequences to variation graphs, splicing graphs, etc.
GED-MAP https://github.com/thomas-buechler-ulm/gedmap	Prototype of efficient short read mapping to pangenome graph [84].
DeepVariant https://github.com/google/deepvariant	A universal SNP and small-indel variant caller [85,86].
SpeedSeq https://github.com/hall-lab/speedseq	A platform for alignment, variant calling, and functional annotation [87].
graphTyper https://github.com/DecodeGenetics/graphtyper	This is a graph-based variant caller. It realigns short-read sequence data to a pangenome and is used for discovering and genotyping sequence variants [88].
PanGenie https://github.com/eblerjana/pangenie	Kmer-based genotyper for structural variation detection on pangenome graphs from short-read sequencing data to genotype a broad spectrum of genetic variation [89].
VEP https://ensembl.gemene.org/tools.html	Variant Effect Prediction (VEP) helps in analyzing the consequences of sequence variations [90].
OrthoFinder https://github.com/davidemms/OrthoFinder	This an ortholog inference method [91].
OrthoMCL https://orthomcl.org/orthomcl/app	A method for identification of ortholog groups for eukaryotic genomes [92].
JustOrthologs https://github.com/ridgelab/JustOrthologs/	JustOrthologs is a fast ortholog identification algorithm that uses conservation of gene structure [93].
PhyloMCL https://sourceforge.net/projects/phyloMCL/files/Materials/	PhyloMCL: Accurate clustering of hierarchical orthogroups guided by phylogenetic relationship and inference of polyploidy events [94].
OMA https://github.com/DessimozLab/OmaStandalone/tree/v2.4.0 https://omabrowser.org/oma/home/	Orthologous Matrix (OMA) is a method and database for orthologs identification from public domains and customized genomes [95].
InParanoid-Diamond https://bitbucket.org/sonnhammergroup/inparanoid/src	InParanoid algorithm for orthology analysis and gene family identification [96]. This is used for orthology

	projection of plant gene families in the Plant Reactome (https://plantreactome.gramene.org) [97].
Panache https://github.com/SouthGreenPlatform/panache	This is a web browser-based viewer for linearized pangenomes [98]. It is being used in visualizing pan-genome data in several plant portals, e.g., the banana genome hub [99].
MoMI-G https://github.com/MoMI-G/MoMI-G/	Genome graph browser for SVs visualization. Users can filter and visualize annotations and inspect SVs with read alignments over the genome graph [100].
panGraphViewer https://github.com/TF-Chan-Lab/panGraphViewer	PanGraphViewer was developed using Python3 for pangenome graph visualization and runs on all major operating systems.
Bandage https://rrwick.github.io/Bandage/	This is an interactive tool for visualizing de novo assembled genomes [101].
Bandage-NG https://github.com/asl/BandageNG	GUI program to interact with assembly graphs based on OGDF (The Open Graph Drawing Framework and Open Graph algorithms and Data structures Framework)
sequenceTubeMaps https://github.com/vgteam/sequenceTubeMap	Interactive visualization of genomes [102].
GfaViz https://github.com/ggonnella/gfaviz	Interactive visualization of graphical fragment assembly (GFA) genome graphs [103].
AGB https://github.com/almiheenko/AGB	Assembly Graph Browser (AGB) is an Interactive tool for constructing and visualization of large assembly graphs, repeat sequence analysis [104].
IGGE https://github.com/immersivegraphgenomeexplorer/IGGE	An interactive graph genomes browser.
GFAViewer https://lh3.github.io/gfatools/	Part of the GFA server for online visualization of GFA files
SGTK https://github.com/olga24912/SGTK	Scaffold graph toolkit for construction and interactive visualization of scaffold graphs using sequencing data [105].
Maffer https://github.com/pangenome/maffer	Convert sorted graphs to multiple alignment format (MAF).
Gfatools https://github.com/lh3/gfatools	A set of tools to parse, subgraph, and convert GFA or rGFA format to FASTA/BED format.
Pgge https://github.com/pangenome/pgge	This is a pan-genome graph evaluator
WGT https://github.com/Kuanhao-Chao/Wheeler_Graph_Toolkit	Tools and algorithms for recognizing, visualizing, and generating Wheeler graphs.
GBWT https://github.com/jltsiren/gbwt	A tool used for haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT) approach [106,107].
Spodgi https://github.com/pangenome/spodgi	Convert ODGI genome graph file to SPARQL database.
GraphPeakCaller https://github.com/uio-bmi/graph_peak_caller	A tool for calling transcription factor peaks on graph-based reference genomes using ChIP-seq data [108].
PSVCP https://github.com/wjian8/psvcp_v1.01	A pangenome analysis pipeline (PSVCP) to construct pan-genome, call structural variants, and run population genotyping used for rice pan-genome [109].
ppsPCP http://cbi.hzau.edu.cn/ppsPCP/	It is designed specifically for constructing fully annotated plant pan-genomes. It scans presence/absence variants [110].

3. A Survey of Crop Pan-genomes and Genomic Resources

With technical advances and affordability of the sequencing and assembly of the genomes, we are experiencing a deluge of big data in biology. The plant research community now faces a bigger challenge of making the genomic data findable, accessible, interoperable, and reusable (FAIR) [111,112]. Public databases and genomic resources play a crucial role in making genomic data FAIR and provide tools for visualization of genomic, transcriptomic, proteomic and metabolomic data [113–120]. Furthermore, the secondary knowledgebases synthesize and curate knowledge graphs

providing information for gene-gene interactions, metabolic networks, and pathways, and provide the tools for analyses of user's data in the context of plant genome browser or pathways [90,97,114,115,119,121–130]. Currently, a substantial number of plant genome browsers, amounting to a few hundred, can be accessed through platforms like Plant Ensembl [120], Phytozome [131], and various clade-specific community databases (for a recent review see [132]). However, the availability of plant pangenome portals is limited and experiencing slow growth. To provide an overview of the current state of crop pan-genomic research, resources, and portals, we compiled Table 2.

Table 2. A list of pan-genomic portals and data resources for various crops.

Pan-genome resource	Remarks
Gramene Link: https://www.gramene.org/pansites Species: maize, rice, grapevine, and sorghum	The Gramene hosts 128 reference plant genomes [115] and a pan-genome site for four crops: maize, rice, grapevine, and sorghum.
SorghumBase Sorghum: https://www.sorghumbase.org	SorghumBase portal hosts a sorghum pan-genome browser consisting of five sorghum reference genome assemblies and genetic variant information for natural diversity panels and ethyl methanesulfonate (EMS)-induced mutant populations [118].
RPAN Link: https://cgm.sjtu.edu.cn/3kricedb In addition to RPAN, the data and analyzed outputs from 3K RGP are available at the following websites: http://snp-seek.irri.org/ http://www.rmbreeding.cn/index.php http://www.ricecloud.org https://aws.amazon.com/public-data-sets/3000-rice-genome Species: rice (<i>O. sativa</i>) and its wild relatives.	The Rice Pan-genome Browser (RPAN) hosts the Rice Pan-genome browser and genomic variation data generated from 3,010 diverse rice accessions generated by 3000 Rice Genome Project (3K RGP) [8,9,133,134]. It contains ~370Mbp IRGSP genome and ~260Mbp novel sequences comprising 50,995 genes (23,914 core genes). RPAN provides a phylogenetic tree browser to view the phylogeny of 3K rice accessions and a genome browser to view gene annotation and presence-absence variations (PAVs). Users can access pan-gene views and associated genetic variations.
RiceSuperPIRdb Link: http://www.ricesuperpir.com Species: 251 genomes representing domesticated rice accessions and wild relatives (202 <i>O. sativa</i> , 28 <i>O. rufipogon</i> , 11 <i>O. glaberrima</i> , and 10 <i>O. barthii</i> accessions).	The Rice Super Pan-genome Information Resource Database (RiceSuperPIRdb) provides an interactive web-based browser for the rice super pan-genome. It was built using reference-free high-quality whole genome alignment of 251 independent genome assemblies. Genome annotations (including annotations for transposable elements) and node-specific K-mer spectrum pangenome graphs are available for each assembly. In addition, genetic variation graphs support linking query data. This super pan-genome also facilitates the identification of lineage-specific haplotypes for trait-associated genes [21].
PanOryza Link: https://panoryza.org Species: It hosts the data for Magic-16 rice accessions; see https://panoryza.cgrb.oregonstate.edu/node/5	PanOryza provides consistency in the rice gene annotation across all rice varieties and the rice pan-genome browser supported by JBrowse Genome Browser.
MaizeGDB Link: https://nam-genomes.org Species: maize	MaizeGDB hosts 48 maize genomes, including 26 high-quality PacBio genome assemblies of the Nested Associated Mapping (NAM) population founder lines and their associated datasets. It allows users to connect genomes, gene models, expression, methylome, sequence variation, structural variation, transposable elements, and diversity data across the maize pan-genome framework. That is supported by the JBrowse genome browser [125].
ZEAMAP Link: www.zeamap.com Species: maize	ZEAMAP is a comprehensive database incorporating multiple reference genomes, annotations, comparative genomics, transcriptomes, open chromatin regions, chromatin interactions, high-quality genetic variants, phenotypes, metabolomics, genetic maps, genetic mapping loci, population structures, and populational DNA methylation signals within maize inbred lines" [135]
GreenPhylDB Link: https://www.greenphyl.org/cgi-bin/index.cgi Species: 46 plant species. 19 pangenomes, including rice, maize, banana, grape, and cacao. In addition, it hosts 27 reference genomes.	GreenPhylDB is part of the South Green Bioinformatics platform (https://www.southgreen.fr) [136]. It aids exploration of gene families and homologous relationships among plant genomes.

The Wheat Panache web portal Link: http://www.appliedbioinformatics.com.au/wheat_panache Species: wheat	This wheat pangenome graph visualization is supported by Panache tool. It allows users to explore SVs across the chosen wheat accessions [137].
GrainGenes Link: https://wheat.pw.usda.gov/GG3/pangenome Species: wheat, barley, rye, oat	GrainGenes hosts molecular and phenotypic information for wheat, barley, rye, oat, etc. including several genome assemblies, genome browsers, and a <i>T. aestivum</i> (bread wheat) pangenome [138].
Wheat-Pangenome Link: http://appliedbioinformatics.com.au/cgi-bin/gb2/gbrowse/WheatPan/ Species: bread wheat (<i>Triticum aestivum</i>)	Wheat-Pangenome facilitates comparison of an improved reference for the Chinese Spring wheat genome with 18 wheat cultivars [139].
SGN Links: https://solgenomics.net Subsites: http://solomics.agis.org.cn/tomato https://solgenomics.net/projects/tgg https://solgenomics.net/organism/Solanum_lycopersicum/genome/ https://solgenomics.net/organism/Solanum_melon/gena/genome Species: tomato, potato, pepper, petunia, and eggplant.	The Solanaceae Genomics Network (SGN) database hosts genomic information about the nightshade family. Currently, it has pan-genome data for tomato and eggplant. International Tomato Genome Sequencing Project produced the tomato pan-genome data, that provides a collection of tomato reference genome assemblies from 46 accessions (22 <i>Solanum lycopersicum</i>, 13 <i>Solanum lycopersicum</i> var. <i>cerasiforme</i>; and 11 <i>Solanum pimpinellifolium</i>) that can be viewed at http://solomics.agis.org.cn/tomato/tool/jbrowse_nav and downloaded from http://solomics.agis.org.cn/tomato/ftp/ [140]. For details about the eggplant pan-genome and pan-plastome data see Barchi et al., 2021 [141].
PepperPan Link: http://www.pepperpan.org:8012/ Species: <i>Capsicum annuum</i> (pepper) and its wild relatives	The pepper pan-genome was constructed by mapping the sequences of 383 cultivars to the Zunla-1 genome as the reference [142]. The novel contig sequences (accession number GWHAAAT000000000) is available at http://bigd.big.ac.cn/gwh.
BRIDGEcereal Link: https://bridgrecereal.scinet.usda.gov Species: wheat, maize, barley, sorghum, and rice.	The Blastn Recovered Insertion and Deletion near Gene Explorer (BRIDGEcereal) is a web application for mining pangenomes of cereals. It facilitates the identification of potential indels (insertion or deletions) for genes of interest with publicly accessible pangenomes of five major cereal crops, including wheat, maize, barley, sorghum, and rice [143].
PanSoy Link: https://www.soybase.org/projects/SoyBase.C2021.01.php Species: <i>Glycine soja</i> (wild soybean) and <i>Glycine max</i> (soybean)	From GmHapMap collection 204 phylogenetically and geographically distinct soybean accessions were chosen for construction of soybean pan-genome using de novo genome assembly method [144].
Sunflower Genome Database Link: https://www.sunflowergenome.org Species: <i>Helianthus annuus</i> (sunflower)	Sunflower pan-genome was generated using sequence from 287 cultivated lines, 17 Native American landraces, and 189 wild accessions representing 11 compatible wild species. Raw data used for pan-genome construction is available at NCBI and SNPs data is available at Sunflower Genome Database [145].
COTTONOMICS Link: http://cotton.zju.edu.cn Species: Cotton	It provides genome-wide, gene-scale structural variations detected from 11 assembled allopolyploid cotton genomes and are linked to important agronomic traits [146].
BGH Link: https://banana-genome-hub.southgreen.fr Species: <i>Musa Ensete</i>, and genomics data of 15 Musaceae species	The Banana Genome Hub (BGH), a web-based platform, supports users in exploring genes and gene families, gene expression patterns, associated SNP markers, etc. Users can also view chromosome structures, synteny, presence, absence variation, and genome ancestry mosaics [99].
CPBD Links: http://citrus.hzau.edu.cn/ Species: sweet orange (<i>Citrus sinensis</i>), mandarin (<i>Citrus reticulata</i>), pummelo (<i>Citrus grandis</i>), grapefruit (<i>Citrus paradisi</i>), and lemon (<i>Citrus limon</i>).	The Citrus Pan-genome to Breeding Database (CPBD) was built using 23 genomes of 17 citrus species and has genetic variation data from 167 citrus accessions mapped to two reference genomes [147].
CitGVD Links: http://citgvd.cric.cn/home/index Species: citrus accessions	The Citrus Genome Database (CitGVD) hosts citrus genomic data, genetic variation data, and built-in analysis tools. It contains 1493258964 non-redundant SNPs, INDELs, and 84 phenotypes from 346 citrus individuals. Users can browse/search annotated genetic variations and visualize results graphically in a genome browser or

	tabular outputs. This portal supports comparative genomics between two or among several citrus accessions [148].
Apple pan-genome Link: http://bioinfo.bti.cornell.edu/apple_genome Species: Apple (<i>Malus domestica</i>) and its wild progenitors <i>M. sieversii</i> , and <i>M. sylvestris</i>	Apple pan-genome was constructed using phased diploid genome assemblies of <i>Malus domestica</i> cv. Gala, <i>M. sieversii</i> and <i>M. sylvestris</i> , and 91 sequenced genomes of additional accessions [149].
BnPIR Link: http://cbi.hzau.edu.cn/bnapus Species: <i>Brassica oleracea</i> , <i>Brassica macrocarpa</i> (cultivated and wild cabbage), <i>Brassica napus</i>	The Brassica napus pan-genome information resource (BnPIR) hosts eight high-quality <i>B. napus</i> reference genomes generated using PacBio sequencing and a collection of 1688 rapeseed re-sequencing data. This pangenome resource provides a pan-gene module, pan-genome Browser, and synteny data. It contains multi-omics data and common bioinformatics tools [150].
Cassava Pan-genome Link: https://cassavabase.org/ Species: cassava (<i>Manihot esculenta</i>)	Two high-quality, chromosome-scale haploid genomes assemblies for African cassava cultivar TME204 (resistant to cassava mosaic diseases caused by African cassava mosaic viruses) were generated using both short-read and long-read sequencing methods (Illumina PE reads, PacBio CLR, and HiFi reads [151].
Other public pan-genome data available (not yet included in crop databases or supported by Genome Browser and associated tools)	
Pearl millet Link: http://117.78.45.2:91/download/ Species: Pearl millet	Pearl millet pan-genome was constructed using whole genome assemblies of 11 accessions generated using a combination of PacBio long-read sequences, Bionano optical mapping data, Hi-C data, and Illumina short-read sequence data [55].
Sorghum pan-genome Link: The bulk data is available at http://dataverse.icrisat.org/dataset.xhtml?persistentId=doi:10.21421/D2RIO2OM . Species: sorghum	This pan-genome was assembled using iterative mapping of whole-genome sequence data from 176 sorghum accessions to a sorghum reference assembly v3.0.1 from Phytozome [152]. The sorghum pan-genome consists of 210,805 sequences, including sequences from the sorghum reference genome assembly v3.1.0. In addition, it has 209935 assembled contig sequences from 176 sorghum accessions. The total genes represented in this data are 35719 genes (34211 genes from reference).
Barley pan-genome Link: https://bitbucket.org/ipk_dg_public/barley_pangenome/src/master/ https://galaxy-web.ipk-gatersleben.de/libraries Species: Barley cultivars, landraces, and a wild relative.	This first-generation barley pan-genome is based on chromosome-scale sequence assemblies for the 20 barley varieties, including landraces, cultivars, and a wild barley from global barley diversity, and by mapping whole-genome shotgun sequencing data from additional 300 barley accessions [11].
Soybean pan-genome Links: The genetic diversities from the 29 genomes is available at https://figshare.com/s/689ae685ad2c368f2568 SNPs and small indels) data from the 2,898 accessions is available at (http://bigd.big.ac.cn/gvm/getProjectDetail?project=GVM000063). Species: Glycine soja (wild soybean) and 26 Glycine max (soybean) accession	This graph-based pan-genome assembly was generated using de novo genome assemblies for 26 representative soybeans selected from 2,898 deeply sequenced accessions [14]. The sequencing data, assembled chromosomes, unplaced scaffolds, and annotations are available at the Genome Sequence Archive and Genome Warehouse database in BIG Data Center (https://bigd.big.ac.cn/gsa/index.jsp) under Accession Number PRJCA002030.
Chickpea pan-genome Links: Pan-genome assembly and annotations: https://doi.org/10.6084/m9.figshare.16592819 The variant calls: https://cegresearch.icrisat.org/cicerseq Manhattan and QQ-plots for Genome-Wide Association Study (GWAS) analysis: https://doi.org/10.6084/m9.figshare.15015309	The chickpea pan-genome was constructed using sequenced 3,366 germplasm lines, including 3,171 cultivated and 195 wild accessions, using iterative mapping and assembly method [153].
Pigeon pea Link: https://research-repository.uwa.edu.au/en/datasets/pigeon-pea-pangenome-contig-assembly-annotation-snps-pav Species: pigeon pea (<i>Cajanus cajan</i>)	The pigeon pea pan-genome based on 89 accessions, including 70 from South Asia, 8 from sub-Saharan Africa, 7 from South-East Asia, 2 from Mesoamerica and 1 from Europe) was constructed. The pangenome was generated using the reference genome assembly (C. cajan_V1.0) and iterative mapping and assembly method [154].

Sesame pan-genome Species: sesame (<i>Sesamum indicum</i> L.)	The sesame pan-genome was constructed by mapping genome sequencing data from two landraces <i>S. indicum</i> cv. Baizhima and Mishuozhima and two cultivars, Yuzhi11, and Swetha to the <i>S. indicum</i> var. Zhongzhi13 reference genome [155].
Cotton Variome Links: Genetic variation is available at https://www.ncbi.nlm.nih.gov/bioproject/PRINA576032 and https://figshare.com/s/cb3c104782a1dcd90ab0 Species: <i>Gossypium hirsutum</i> and <i>Gossypium barbadense</i>	This is a comprehensive resource providing genetic variation data from 1961 cottons accessions from <i>G. hirsutum</i> and <i>G. barbadense</i> [156].
Melon pan-genome Link: https://figshare.com/articles/dataset/melon_pangenome/17195072	Pan-genome of <i>Cucumis melo</i> L. consists of genome sequence data from 297 accessions [157]
Cucumber pan-genome Data availability: Genome assemblies of the 11 cucumber accessions have been deposited in NCBI GenBank under the accession number PRINA657438 .	Pan-genome graph constructed for 11 cucumber accessions. These 11 representative accessions from the 115-line core collection were sequenced using the PacBio platform, and the genome assemblies were generated using long-read and short-read data [158].
Strawberry pan-genome The genome assembly and annotation files are available in the Genome Database for Rosaceae. The pan-genome browser or query support is not available. Link: https://www.rosaceae.org/species/fragaria/all Species: cultivated and wild strawberry accessions	This strawberry pan-genome was generated using chromosome-scale reference genomes assemblies of five diploid strawberry species (<i>Fragaria mandschurica</i> , <i>Fragaria daltoniana</i> , <i>Fragaria pentaphylla</i> , <i>F. nilgerrensis</i> , and <i>F. viridis</i>) strawberry species and genome resequencing data of 128 accessions [159].
Walnut pan-genome Link: https://db.cngb.org/search/project/CNP0001209 . Species: walnut (<i>Juglans nigra</i>)	A high-quality genome assembly of black walnut (<i>Juglans nigra</i>) genotype NWAUFU168 using both short-read and long-read platforms (Illumina, Pacbio, and Hi-C) was constructed. A Walnut pan-genome was built using this reference genome and mapping sequence data from 74 walnut accessions [56].

It is clear from the Table 2 that at present crop pan-genome research is at its early stage. The pan-genome browsers are available for a few crops, and thus, most of the data is not supported for user-friendly query, visualization and supporting the analysis of the user’s data. However, the very few platforms and genomic databases that support pan-gene analysis in phylogenomic context and support the user’s query show the potential of pan-genome data for answering both basic research and translational applications of this knowledge for crop improvement. Here, we highlight an example of visualizing the pan-gene data for TILLER ANGLE CONTROL 1 (TAC1) transcription factor from various accessions of cultivated rice *O. sativa*, other members of *Oryza* genus at Gramene oryza pansite (<https://oryza.gramene.org>). *OsTAC1* is induced by gravistimulation and promotes horizontal shoot growth by negatively regulating shoot gravitropism [160]. Thus, it is involved in regulation of tiller angle and modulating plant architectural traits of agronomic tremendous importance. A comparison of TAC1 protein sequences shows a significant variability at the carboxy terminal between the domesticated rice cultivars of japonica and indica accessions (Figure 4A). Indeed, a previously published study has shown that a point mutation in *OsTAC1* gene at the 3’-splicing site of the 1.5-kb intron (‘GGGA’) in japonica rice accessions caused reduction in the expression of this gene leading to a smaller tiller angle. This trait was selected in the japonica rice accessions. In contrast, wild rice accessions and indica rice accessions, which have large tiller angle, contain ‘AGGA’ sequence at 3’-splicing site of the 1.5-kb intron [161]. *OsTAC 1* gene is on chromosome 9 in rice genome and shows high conservation across all rice accession (Figure 4B). However, we also see that genomic neighborhood of the *TAC1* on left-hand side is not much conserved and it remains to be explored if any of these genes are involved in functional adaptation of rice accessions.



Figure 4. A pan-gene overview for TAC1 transcription factor (reference gene OsTAC1; Os09g0529300) and its orthologs from various accessions of cultivated rice *O. sativa*, other members of *Oryza* genus, and two other monocots maize and sorghum at Gramene oryza pansite. Users can explore (A) an overall pan-gene alignment view that can be expanded to a multiple sequence alignment (using amino acid sequences of the encoded proteins) and gene neighborhood conservation.

Beside Gramene database, a few more pan-genome portals provide a similar view of pan-genes visualization. The other portal host different views and analysis tools. For example, the Banana Genome Hub uses Panache platform for visualization of pan-genomic data on *Musaceae*. However, in general the well-established crop genome portals supports users in exploring genes and gene families, chromosome structures, synteny, SVs, gene expression patterns, SNP markers, etc.

4. Plant pan-genomics-driven insights for understanding the basis of agronomic traits

Cereal crops have been the prime subject of agriculture research. Thus, we have matured genomic resources, enriched genome annotations, and genotype and sequence data facilitating comparative genomics and pan-genomics studies (See Table 2). The focus of all categories of genomic research on cereal crops aims to increase the grain yield or plant developmental, physiological, and architectural traits that can support the higher yield. For example, Wang et al. (2022) [109] used rice pan-genome to identify GW5 genes associated with the trait ‘thousand-grain weight’ (TGW), and a novel locus *qPH8-1* involved in the regulation of plant height. In another study by Shang et al., 2022 [21] a super pan-genome of rice was constructed helped to identify genetic variants associated with submergence tolerance, seed shattering, and plant architecture [21]. Many important studies have been published in maize, rice, wheat and are being implemented for their improvement.

Notably, more investments are being made in the genomic and pan-genomic research of minor cereal crops, including sorghum and millets suitable for growing in diverse and marginal lands (see Table 2). These crops have a high degree of in-built tolerance for mitigating the impact of harsh environments and resistance against many pests and pathogens. For example, foxtail millet (*Setaria italica*), is a model plant for studying C4 photosynthesis and developing climate-resilient crops. A pan-genomic study of foxtail millet identified an important genetic variation in the promoter region of *SiGW3* that is associated with yield improvements [22]. Another study identified 13 marker-trait associations using proso millet (*Panicum miliaceum* L.) pan-genome [162]. Similarly, Yan et al. (2023) recently constructed a pearl millet pan-genome that identified over 400,000 genomic structural variants (SVs) and provided insights into heat tolerance. This study identified the RWP-RK gene conferring enhanced heat tolerance [55]. Another group of previously understudied crops is legumes

that has gained from genomic and pan-genomic research and breeding efforts (see Table 2). For example, pigeon pea is an important orphan crop mainly grown by smallholder farmers in the tropics and subtropical regions of the world. It has an in-built tolerance for drought stress and is very productive in marginal land with small inputs. A pan-genome study of pigeon pea identified 225 SNPs associated with nine agronomically important traits. These associations will aid pigeon pea germplasm improvement for breeding legumes [154]. Mung bean is an excellent plant-based source of protein and is grown in temperate, subtropical, and tropical regions. In a study, Liu et al. (2022) analyzed 217 mung bean accessions and discovered many novel genes associated with agronomic traits, for example, an SNP in the candidate genes SWEET10 homolog (jg24043) associated with crude starch content; NRT1/PTR FAMILY 2.13 gene for pod length; a homolog of WUSCHEL-family homeobox gene associated with yield; and a PAV event in a multi-gene locus associated with color-related traits [163].

Pan-genome studies have become increasingly important for understanding the genetic diversity of major crops from tropical regions, including cassava (*Manihot esculenta*) and banana (*Musa* spp.). These crops are a vital component of diverse ecosystems and play essential roles in the livelihoods of local communities. Insights gained from genomic and pan-genomic research are being deemed vital for improving banana for resistance against Fusarium wilt to safeguard the global banana industry [99,164]. Similarly, the pan-genome of cassava (*Manihot esculenta*), a staple crop for millions of people in Africa, is helping to score genomic variations, particularly in genes associated with disease resistance and starch biosynthesis [151]. More studies are being performed on important fruit and vegetables (i.e., banana, apple, tomato, melon, citrus, and grape) and oilseed crops (i.e., Brassica, soybean, sunflower, etc.) listed in Table 2. A few studies have uncovered new genes and rare alleles that regulate secondary metabolites associated with color and flavor [46,49,165]; pathogen resistance [43,47,56,79,145,165–167]; and abiotic stress tolerance [41,42,55]. These findings foster understanding of the genetic basis of diverse traits in the domesticated crops and offer promising prospects for the introduction of candidate genes (for disease resistance breeding and quality traits) through precise genome editing into elite cultivars.

In conclusion, the acquisition of pangenome holds immense potential for pursuing fundamental questions, especially for climate-change resilient breeding and overall crop improvement. Utilizing pangenomes enables researchers to (1) Identify gains and losses of genetic regions and structural variations strongly associated with desirable fitness phenotypes such as abiotic stress and disease tolerance, growth and development, yield, biomass, and performance. (2) Identify gains and losses in protein-coding regions and/or epigenetic features between crops and closely related species. (3) Identify pathways, gene networks, expression profiles, and transcript isoforms that are enriched for given major and minor quantitative trait loci (QTLs) with desired phenotypes and adaptation traits. (4) Project functional and phenotype homologs from a well-studied species onto a new/less-studied species through whole genome comparisons and synteny. (5) Advance plant breeding efforts by mapping/querying/visualizing public or personal project data to build or test hypotheses, discover markers, gain knowledge, and support state-of-the-art breeding applications. These goals and the knowledge required to achieve them represent important pieces of an overall puzzle and therefore require exploration. In summary, harnessing the full potential of pan-genomes for targeted crop breeding is an important goal of the plant research and breeding community. The integration of diverse genomic tools, including pan-genome, marker-assisted selection, genomic selection, and gene editing techniques, will play a pivotal role in developing climate-resilient crops that can adapt to and thrive in the face of climate changes, securing global food security and sustainability (see Figure 5).

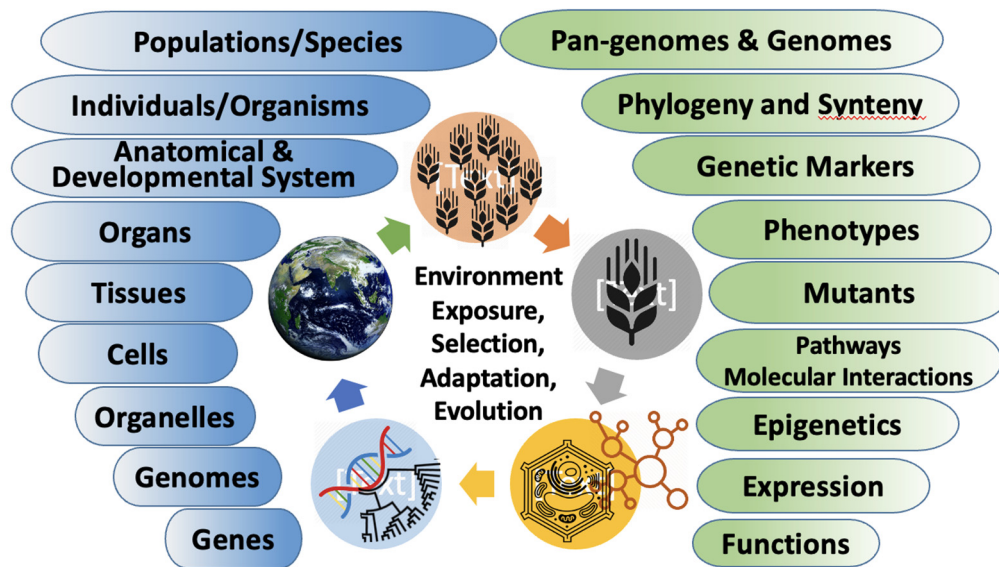


Figure 5. Plant pan-genomes can help to integrate heterogeneous omics data to understand gene function, genome evolution, and speciation and enable genomic selection, genome editing, and phenotype prediction to support and sustain agriculture production.

5. Outlook, opportunities, and innovations in plant pan-genome research

Plant genomes are often very large and complex, making it difficult to produce high-quality genome assemblies with accurate gene annotations. For over last three decades, genotypic variations have been captured using various genetic markers, including RFLP, RAPD, SNPs, SSRs, microsatellite markers, etc. to establish connections between genotype and phenotype to aid plant breeding and cultivar improvement [168–173]. However, the advances in the short-read and long-read-based next-generation sequencing platforms and computational methods required for processing large-scale genomic and transcriptomic data have facilitated the rapid and cost-effective whole genome sequences, including de-novo genome assemblies [7,174,175]. These whole genome sequence comparisons are driving pan-genomic research and, first time in history, allowing researchers to explore intra- and inter-species structural variations at the resolution of nucleotide sequence level. We can expect a significant expansion in the availability of pangenomes across a broader range of plant species. The pangenomes facilitate the identification of genes that are conserved across species and genes that are unique to specific species or a subset of accessions. The availability of species-specific or genus-level pangenomes of crops is crucial for understanding the dynamics of their genome evolution, including the impact of artificial selection and domestication, crop diversification, and adaptation under varied environments [50,174,176,177]. For instance, a super-pangenome of the *Citrullus* genus comprising of 346 cultivated watermelon accessions and 201 wild accessions suggested that a duplication of the sugar transporter gene *CITST2* was likely selected during domestication for higher fruit sweetness; the wild accessions harbor many genes related to disease resistance compared to cultivated watermelon accessions; and identified domestication sweeps in cultivated watermelon [178].

In addition, plant pan-genome can help to advance the fundamental understanding of the plant kingdom at various scales. It can aid in improving functional annotations of genes and genomes [179,180], and provide insights into specific roles and interactions of different genetic elements. Furthermore, pangenomes could help to understand the evolution of metabolic diversity across diverse taxonomic clades [97]. This knowledge has implications for improving crop traits and developing more resilient and sustainable agricultural practices. Moreover, by analyzing the genomic diversity represented in pangenomes, scientists can understand the distribution and composition of vegetation across different environments. This information is critical for conservation efforts, ecological studies, and developing strategies to protect and sustainably manage plant resources.

Finally, we see a great application of pan-genome data in improving gene annotations and identifying evolutionary conserved sets of genes associated with important agronomic traits. Likewise, comparative genomic studies can help to identify and annotate clade-specific unique genes that determine metabolite compositions of important fruit and vegetable crops or other categories. These specialized metabolic pathways and associated entities can be easily curated in the pathway databases [97,121,123,124,128,181,182]. The workflows for interspecies gene family comparisons, GO annotations, and standard protocols for gene biocuration are very efficient and established; they can easily accommodate the insights gained from intra-species comparisons. For instance, the availability of whole genome sequences of plants has contributed tremendously to the knowledge of gene duplications, gene family evolution, and functional diversification of homologous genes [179,180,183]. Gene Ontology (GO) and Plant Ontology (PO) annotations have played a central role in accessing the potential gene functions [184–186]. Furthermore, the comprehensive analysis of plant transcriptomes has helped us to link genes with potential biological processes, pathways, and response to biotic and abiotic stress or stimulants [179,180,187–189]. Integrating pangenomes with other omics, such as transcriptomics, epigenomics, proteomics, and phenomics data, will enable a comprehensive understanding of gene regulation and functional mechanisms underlying important agronomic traits. In conclusion, as shown in Figure 5, pan-genomic research is driving functional genomics, evolutionary studies, and biodiversity exploration and consequently, holds great potential for crop improvement, environmental conservation, and sustainable agriculture.

It is important to emphasize that pan-genome construction and whole genome-level comparative analysis require substantial computational infrastructure, expertise in various facets including sequence generation, genome assembly and annotation; and subsequent bioinformatic analyses and data visualization. In general, all these tasks are beyond the capacity of individual research laboratories, and thus, they require extensive infrastructure and bioinformatic support from their institutions and public databases. Public data repositories, databases, genomic resources, and secondary knowledgebases play an essential role in aiding the community of researchers in providing ontologies [184], archiving and annotating genomic data, and supporting analysis and visualization of omics data to support data-driven hypotheses and experimental plans [90,97,115,128]. Here we have reviewed the tools and genomic resources available to the plant research community for conducting pan-genomic research.

We note here that the resources and tools for pan-genome visualization and to support researchers for exploring pan-genomic data are limited and at an early developmental stage. We have witnessed the growing number of publications on crop pan-genomes in the last five years (see Table 2); however, the pan-genomic and genomic diversity data for the majority of the crops are stored and archived with no associated tools/features required for visualization and effective use by other researchers. Thus, advancement and innovations in pan-genomic data visualization, analysis tools, and additional biocuration of genomic data are needed to facilitate meaningful intraspecies and interspecies genomic comparisons. Community biocuration can play an essential role in making sense of the big data and ensuring quality controls at various steps [190]. This will allow researchers to study pan-genomes in more detail and identify genes and genomic regions associated with important traits. Equally important is integrating crop pan-genomes in comprehensive knowledgebases that host analyzed and annotated genomic and pathway data for ensuring the FAIR data policy implementations [111,132].

From a technical aspect, integrating machine learning and artificial intelligence techniques will expedite the analysis of complex pan-genomic datasets, aid in identifying patterns, and accelerate the construction of predictive models and gene biocuration.

Furthermore, the genomic resources are more mature for the model plants and major crop species [132]. But the necessary financial and infrastructure support for minor crops, fruit and vegetable crops, and orphan crops is still insufficient [191–193]. Although the cost of sequencing a single genome has decreased significantly in recent years, sequencing multiple genomes can still be prohibitive, a significant barrier to conducting pan-genome research in poorly studied crops or orphan crops.

We expect that this review will help researchers to understand and evaluate the status and strategies being employed for pan-genomic studies in crops, and encourage them to seek necessary collaborations within the community or help to advocate for increased funding for developing the infrastructure, tools, and biocuration of genomic data that is vital for making needed progress. Furthermore, we hope this review can help students and young researchers learn about the status of crop genomic research and the future potential of pan-genomics.

Author Contributions: S.N. conceptualized the initial study plan and coordinated the collaboration with all the authors. S.N., C.D., S.K.S., and P.J. conducted a review of the literature and contributed to writing the manuscript. S.N., C.D. and P.J. created all the figures for the manuscript. All authors have read and agreed to this version of the manuscript.

Funding: SN acknowledges funding from National Aeronautics and Space Administration (NASA), USA #80NSSC22K0891. PJ acknowledges funding from NASA #80NSSC22K0855 and National Science Foundation, USA #2029854.

Data Availability Statement: We have provided all the data within this manuscript.

Acknowledgments: SN and PJ thank the members of the Gramene project team at Cold Spring Harbor Laboratory, the Ensembl Plants team at European Bioinformatics Institute, and the human Reactome project and the Ontario Institute for Cancer Research for their support and interest in strong cross-platform data integration. SKS acknowledges support from the 10KP project (<https://db.cngb.org/10kp>) and China National GeneBank (CNGB; <https://www.cngb.org/>). CD thanks the bioinformatics teams at The New Zealand Institute for Plant and Food Research Limited (PFR), breeders at Kiwifruit Breeding Centre (KBC), and researchers at China Agricultural University (CAU) for their support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Computational Pan-Genomics, Consortium. "Computational Pan-Genomics: Status, Promises and Challenges." *Brief Bioinform* 19, no. 1 (2018): 118-35.
2. Della Coletta, R., Y. Qiu, S. Ou, M. B. Hufford, and C. N. Hirsch. "How the Pan-Genome Is Changing Crop Genomics and Improvement." *Genome Biology* 22, no. 1 (2021): 3.
3. Ho, S. S., A. E. Urban, and R. E. Mills. "Structural Variation in the Sequencing Era." *Nature Reviews Genetics* 21, no. 3 (2020): 171-89.
4. Kyriakidou, M., H. H. Tai, N. L. Anglin, D. Ellis, and M. V. Stromvik. "Current Strategies of Polyploid Plant Genome Sequence Assembly." *Frontiers of Plant Science* 9 (2018): 1660.
5. Sedlazeck, F. J., P. Rescheneder, M. Smolka, H. Fang, M. Nattestad, A. von Haeseler, and M. C. Schatz. "Accurate Detection of Complex Structural Variations Using Single-Molecule Sequencing." *Nature Methods* 15, no. 6 (2018): 461-68.
6. Wang, Y., J. Yu, M. Jiang, W. Lei, X. Zhang, and H. Tang. "Sequencing and Assembly of Polyploid Genomes." *Methods in Molecular Biology* 2545 (2023): 429-58.
7. Sahu, S. K., and H. Liu. "Long-Read Sequencing (Method of the Year 2022): The Way Forward for Plant Omics Research." *Mol Plant* 16, no. 5 (2023): 791-93.
8. Zhou, Y., D. Chebotarov, D. Kudrna, V. Llaca, S. Lee, S. Rajasekar, N. Mohammed, N. Al-Bader, C. Sobel-Sorenson, P. Parakkal, L. J. Arbelaez, N. Franco, N. Alexandrov, N. R. S. Hamilton, H. Leung, R. Mauleon, M. Lorieux, A. Zuccolo, K. McNally, J. Zhang, and R. A. Wing. "A Platinum Standard Pan-Genome Resource That Represents the Population Structure of Asian Rice." *Sci Data* 7, no. 1 (2020): 113.
9. Wang, W., R. Mauleon, Z. Hu, D. Chebotarov, S. Tai, Z. Wu, M. Li, T. Zheng, R. R. Fuentes, F. Zhang, L. Mansueto, D. Copetti, M. Sanciangco, K. C. Palis, J. Xu, C. Sun, B. Fu, H. Zhang, Y. Gao, X. Zhao, F. Shen, X. Cui, H. Yu, Z. Li, M. Chen, J. Detras, Y. Zhou, X. Zhang, Y. Zhao, D. Kudrna, C. Wang, R. Li, B. Jia, J. Lu, X. He, Z. Dong, J. Xu, Y. Li, M. Wang, J. Shi, J. Li, D. Zhang, S. Lee, W. Hu, A. Poliakov, I. Dubchak, V. J. Ulat, F. N. Borja, J. R. Mendoza, J. Ali, J. Li, Q. Gao, Y. Niu, Z. Yue, M. E. B. Naredo, J. Talag, X. Wang, J. Li, X. Fang, Y. Yin, J. C. Glaszmann, J. Zhang, J. Li, R. S. Hamilton, R. A. Wing, J. Ruan, G. Zhang, C. Wei, N. Alexandrov, K. L. McNally, Z. Li, and H. Leung. "Genomic Variation in 3,010 Diverse Accessions of Asian Cultivated Rice." *Nature* 557, no. 7703 (2018): 43-49.
10. Schatz, M. C., L. G. Maron, J. C. Stein, A. Hernandez Wences, J. Gurtowski, E. Biggers, H. Lee, M. Kramer, E. Antoniou, E. Ghiban, M. H. Wright, J. M. Chia, D. Ware, S. R. McCouch, and W. R. McCombie. "Whole

- Genome De Novo Assemblies of Three Divergent Strains of Rice, *Oryza Sativa*, Document Novel Gene Space of Aus and Indica." *Genome Biology* 15, no. 11 (2014): 506.
11. Jayakodi, M., S. Padmarasu, G. Haberer, V. S. Bonthala, H. Gundlach, C. Monat, T. Lux, N. Kamal, D. Lang, A. Himmelbach, J. Ens, X. Q. Zhang, T. T. Angessa, G. Zhou, C. Tan, C. Hill, P. Wang, M. Schreiber, L. B. Boston, C. Plott, J. Jenkins, Y. Guo, A. Fiebig, H. Budak, D. Xu, J. Zhang, C. Wang, J. Grimwood, J. Schmutz, G. Guo, G. Zhang, K. Mochida, T. Hirayama, K. Sato, K. J. Chalmers, P. Langridge, R. Waugh, C. J. Pozniak, U. Scholz, K. F. X. Mayer, M. Spannagl, C. Li, M. Mascher, and N. Stein. "The Barley Pan-Genome Reveals the Hidden Legacy of Mutation Breeding." *Nature* 588, no. 7837 (2020): 284-89.
 12. Walkowiak, S., L. Gao, C. Monat, G. Haberer, M. T. Kassa, J. Brinton, R. H. Ramirez-Gonzalez, M. C. Kolodziej, E. Delorean, D. Thambugala, V. Klymiuk, B. Byrns, H. Gundlach, V. Bandi, J. N. Siri, K. Nilsen, C. Aquino, A. Himmelbach, D. Copetti, T. Ban, L. Venturini, M. Bevan, B. Clavijo, D. H. Koo, J. Ens, K. Wiebe, A. N'Diaye, A. K. Fritz, C. Gutwin, A. Fiebig, C. Fosker, B. X. Fu, G. G. Accinelli, K. A. Gardner, N. Fradgley, J. Gutierrez-Gonzalez, G. Halstead-Nussloch, M. Hatakeyama, C. S. Koh, J. Deek, A. C. Costamagna, P. Fobert, D. Heavens, H. Kanamori, K. Kawaura, F. Kobayashi, K. Krasileva, T. Kuo, N. McKenzie, K. Murata, Y. Nabeka, T. Paape, S. Padmarasu, L. Percival-Alwyn, S. Kagale, U. Scholz, J. Sese, P. Juliana, R. Singh, R. Shimizu-Inatsugi, D. Swarbreck, J. Cockram, H. Budak, T. Tameshige, T. Tanaka, H. Tsuji, J. Wright, J. Wu, B. Steuernagel, I. Small, S. Cloutier, G. Keeble-Gagnere, G. Muehlbauer, J. Tibbets, S. Nasuda, J. Melonek, P. J. Hucl, A. G. Sharpe, M. Clark, E. Legg, A. Bharti, P. Langridge, A. Hall, C. Uauy, M. Mascher, S. G. Krattinger, H. Handa, K. K. Shimizu, A. Distelfeld, K. Chalmers, B. Keller, K. F. X. Mayer, J. Poland, N. Stein, C. A. McCartney, M. Spannagl, T. Wicker, and C. J. Pozniak. "Multiple Wheat Genomes Reveal Global Variation in Modern Breeding." *Nature* 588, no. 7837 (2020): 277-83.
 13. Hirsch, C. N., J. M. Foerster, J. M. Johnson, R. S. Sekhon, G. Muttoni, B. Vaillancourt, F. Penagaricano, E. Lindquist, M. A. Pedraza, K. Barry, N. de Leon, S. M. Kaeppler, and C. R. Buell. "Insights into the Maize Pan-Genome and Pan-Transcriptome." *Plant Cell* 26, no. 1 (2014): 121-35.
 14. Liu, Y., H. Du, P. Li, Y. Shen, H. Peng, S. Liu, G. A. Zhou, H. Zhang, Z. Liu, M. Shi, X. Huang, Y. Li, M. Zhang, Z. Wang, B. Zhu, B. Han, C. Liang, and Z. Tian. "Pan-Genome of Wild and Cultivated Soybeans." *Cell* 182, no. 1 (2020): 162-76 e13.
 15. Li, Y. H., G. Zhou, J. Ma, W. Jiang, L. G. Jin, Z. Zhang, Y. Guo, J. Zhang, Y. Sui, L. Zheng, S. S. Zhang, Q. Zuo, X. H. Shi, Y. F. Li, W. K. Zhang, Y. Hu, G. Kong, H. L. Hong, B. Tan, J. Song, Z. X. Liu, Y. Wang, H. Ruan, C. K. Yeung, J. Liu, H. Wang, L. J. Zhang, R. X. Guan, K. J. Wang, W. B. Li, S. Y. Chen, R. Z. Chang, Z. Jiang, S. A. Jackson, R. Li, and L. J. Qiu. "De Novo Assembly of Soybean Wild Relatives for Pan-Genome Analysis of Diversity and Agronomic Traits." *Nature Biotechnology* 32, no. 10 (2014): 1045-52.
 16. Song, J. M., Z. Guan, J. Hu, C. Guo, Z. Yang, S. Wang, D. Liu, B. Wang, S. Lu, R. Zhou, W. Z. Xie, Y. Cheng, Y. Zhang, K. Liu, Q. Y. Yang, L. L. Chen, and L. Guo. "Eight High-Quality Genomes Reveal Pan-Genome Architecture and Ecotype Differentiation of Brassica Napus." *Nat Plants* 6, no. 1 (2020): 34-45.
 17. Zhuang, W., H. Chen, M. Yang, J. Wang, M. K. Pandey, C. Zhang, W. C. Chang, L. Zhang, X. Zhang, R. Tang, V. Garg, X. Wang, H. Tang, C. N. Chow, J. Wang, Y. Deng, D. Wang, A. W. Khan, Q. Yang, T. Cai, P. Bajaj, K. Wu, B. Guo, X. Zhang, J. Li, F. Liang, J. Hu, B. Liao, S. Liu, A. Chitkineni, H. Yan, Y. Zheng, S. Shan, Q. Liu, D. Xie, Z. Wang, S. A. Khan, N. Ali, C. Zhao, X. Li, Z. Luo, S. Zhang, R. Zhuang, Z. Peng, S. Wang, G. Mamadou, Y. Zhuang, Z. Zhao, W. Yu, F. Xiong, W. Quan, M. Yuan, Y. Li, H. Zou, H. Xia, L. Zha, J. Fan, J. Yu, W. Xie, J. Yuan, K. Chen, S. Zhao, W. Chu, Y. Chen, P. Sun, F. Meng, T. Zhuo, Y. Zhao, C. Li, G. He, Y. Zhao, C. Wang, P. B. Kavikishor, R. L. Pan, A. H. Paterson, X. Wang, R. Ming, and R. K. Varshney. "The Genome of Cultivated Peanut Provides Insight into Legume Karyotypes, Polyploid Evolution and Crop Domestication." *Nature Genetics* 51, no. 5 (2019): 865-76.
 18. International Wheat Genome Sequencing Consortium. "Shifting the Limits in Wheat Research and Breeding Using a Fully Annotated Reference Genome." *Science* 361, no. 6403 (2018).
 19. Edger, P. P., T. J. Poorten, R. VanBuren, M. A. Hardigan, M. Colle, M. R. McKain, R. D. Smith, S. J. Teresi, A. D. L. Nelson, C. M. Wai, E. I. Alger, K. A. Bird, A. E. Yocca, N. Pumphlin, S. Ou, G. Ben-Zvi, A. Brodt, K. Baruch, T. Swale, L. Shiue, C. B. Acharya, G. S. Cole, J. P. Mower, K. L. Childs, N. Jiang, E. Lyons, M. Freeling, J. R. Puzey, and S. J. Knapp. "Origin and Evolution of the Octoploid Strawberry Genome." *Nature Genetics* 51, no. 3 (2019): 541-47.
 20. Kyriakidou, M., N. L. Anglin, D. Ellis, H. H. Tai, and M. V. Stromvik. "Genome Assembly of Six Polyploid Potato Genomes." *Sci Data* 7, no. 1 (2020): 88.

21. Shang, L., X. Li, H. He, Q. Yuan, Y. Song, Z. Wei, H. Lin, M. Hu, F. Zhao, C. Zhang, Y. Li, H. Gao, T. Wang, X. Liu, H. Zhang, Y. Zhang, S. Cao, X. Yu, B. Zhang, Y. Zhang, Y. Tan, M. Qin, C. Ai, Y. Yang, B. Zhang, Z. Hu, H. Wang, Y. Lv, Y. Wang, J. Ma, Q. Wang, H. Lu, Z. Wu, S. Liu, Z. Sun, H. Zhang, L. Guo, Z. Li, Y. Zhou, J. Li, Z. Zhu, G. Xiong, J. Ruan, and Q. Qian. "A Super Pan-Genomic Landscape of Rice." *Cell Research* 32, no. 10 (2022): 878-96.
22. He, Q., S. Tang, H. Zhi, J. Chen, J. Zhang, H. Liang, O. Alam, H. Li, H. Zhang, L. Xing, X. Li, W. Zhang, H. Wang, J. Shi, H. Du, H. Wu, L. Wang, P. Yang, L. Xing, H. Yan, Z. Song, J. Liu, H. Wang, X. Tian, Z. Qiao, G. Feng, R. Guo, W. Zhu, Y. Ren, H. Hao, M. Li, A. Zhang, E. Guo, F. Yan, Q. Li, Y. Liu, B. Tian, X. Zhao, R. Jia, B. Feng, J. Zhang, J. Wei, J. Lai, G. Jia, M. Purugganan, and X. Diao. "A Graph-Based Genome and Pan-Genome Variation of the Model Plant *Setaria*." *Nature Genetics* 55, no. 7 (2023): 1232-42.
23. Yap, I. V., D. Schneider, J. Kleinberg, D. Matthews, S. Cartinhour, and S. R. McCouch. "A Graph-Theoretic Approach to Comparing and Integrating Genetic, Physical and Sequence-Based Maps." *Genetics* 165, no. 4 (2003): 2235-47.
24. Tettelin, H., V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini, N. L. Ward, S. V. Angiuoli, J. Crabtree, A. L. Jones, A. S. Durkin, R. T. Deboy, T. M. Davidsen, M. Mora, M. Scarselli, I. Margarit y Ros, J. D. Peterson, C. R. Hauser, J. P. Sundaram, W. C. Nelson, R. Madupu, L. M. Brinkac, R. J. Dodson, M. J. Rosovitz, S. A. Sullivan, S. C. Daugherty, D. H. Haft, J. Selengut, M. L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K. J. O'Connor, S. Smith, T. R. Utterback, O. White, C. E. Rubens, G. Grandi, L. C. Madoff, D. L. Kasper, J. L. Telford, M. R. Wessels, R. Rappuoli, and C. M. Fraser. "Genome Analysis of Multiple Pathogenic Isolates of *Streptococcus Agalactiae*: Implications for the Microbial "Pan-Genome"." *Proceedings of the National Academy of Sciences of the United States of America* 102, no. 39 (2005): 13950-5.
25. Springer, N. M., K. Ying, Y. Fu, T. Ji, C. T. Yeh, Y. Jia, W. Wu, T. Richmond, J. Kitzman, H. Rosenbaum, A. L. Iniguez, W. B. Barbazuk, J. A. Jeddloh, D. Nettleton, and P. S. Schnable. "Maize Inbreds Exhibit High Levels of Copy Number Variation (Cnv) and Presence/Absence Variation (Pav) in Genome Content." *PLoS Genetics* 5, no. 11 (2009): e1000734.
26. Anderson, J. E., M. B. Kantar, T. Y. Kono, F. Fu, A. O. Stec, Q. Song, P. B. Cregan, J. E. Specht, B. W. Diers, S. B. Cannon, L. K. McHale, and R. M. Stupar. "A Roadmap for Functional Structural Variants in the Soybean Genome." *G3 (Bethesda)* 4, no. 7 (2014): 1307-18.
27. Golicz, A. A., P. E. Bayer, G. C. Barker, P. P. Edger, H. Kim, P. A. Martinez, C. K. Chan, A. Severn-Ellis, W. R. McCombie, I. A. Parkin, A. H. Paterson, J. C. Pires, A. G. Sharpe, H. Tang, G. R. Teakle, C. D. Town, J. Batley, and D. Edwards. "The Pangenome of an Agronomically Important Crop Plant *Brassica Oleracea*." *Nat Commun* 7 (2016): 13390.
28. Tao, Y., H. Luo, J. Xu, A. Cruickshank, X. Zhao, F. Teng, A. Hathorn, X. Wu, Y. Liu, T. Shatte, D. Jordan, H. Jing, and E. Mace. "Extensive Variation within the Pan-Genome of Cultivated and Wild Sorghum." *Nat Plants* 7, no. 6 (2021): 766-73.
29. Xu, X., X. Liu, S. Ge, J. D. Jensen, F. Hu, X. Li, Y. Dong, R. N. Gutenkunst, L. Fang, L. Huang, J. Li, W. He, G. Zhang, X. Zheng, F. Zhang, Y. Li, C. Yu, K. Kristiansen, X. Zhang, J. Wang, M. Wright, S. McCouch, R. Nielsen, J. Wang, and W. Wang. "Resequencing 50 Accessions of Cultivated and Wild Rice Yields Markers for Identifying Agronomically Important Genes." *Nature Biotechnology* 30, no. 1 (2011): 105-11.
30. Lam, H. M., X. Xu, X. Liu, W. Chen, G. Yang, F. L. Wong, M. W. Li, W. He, N. Qin, B. Wang, J. Li, M. Jian, J. Wang, G. Shao, J. Wang, S. S. Sun, and G. Zhang. "Resequencing of 31 Wild and Cultivated Soybean Genomes Identifies Patterns of Genetic Diversity and Selection." *Nature Genetics* 42, no. 12 (2010): 1053-9.
31. Gui, S., W. Wei, C. Jiang, J. Luo, L. Chen, S. Wu, W. Li, Y. Wang, S. Li, N. Yang, Q. Li, A. R. Fernie, and J. Yan. "A Pan-Zea Genome Map for Enhancing Maize Improvement." *Genome Biology* 23, no. 1 (2022): 178.
32. Allaby, R. G., R. L. Ware, and L. Kistler. "A Re-Evaluation of the Domestication Bottleneck from Archaeogenomic Evidence." *Evol Appl* 12, no. 1 (2019): 29-37.
33. Tirnaz, S., J. Zandberg, W. J. W. Thomas, J. Marsh, D. Edwards, and J. Batley. "Application of Crop Wild Relatives in Modern Breeding: An Overview of Resources, Experimental and Computational Methodologies." *Frontiers of Plant Science* 13 (2022): 1008904.
34. Papa, R., and P. Gepts. "Asymmetry of Gene Flow and Differential Geographical Structure of Molecular Diversity in Wild and Domesticated Common Bean (*Phaseolus Vulgaris* L.) from Mesoamerica." *Theoretical and Applied Genetics* 106, no. 2 (2003): 239-50.

35. McNally, K. L., K. L. Childs, R. Bohnert, R. M. Davidson, K. Zhao, V. J. Ulat, G. Zeller, R. M. Clark, D. R. Hoen, T. E. Bureau, R. Stokowski, D. G. Ballinger, K. A. Frazer, D. R. Cox, B. Padhukasahasram, C. D. Bustamante, D. Weigel, D. J. Mackill, R. M. Bruskiewich, G. Ratsch, C. R. Buell, H. Leung, and J. E. Leach. "Genomewide Snp Variation Reveals Relationships among Landraces and Modern Varieties of Rice." *Proceedings of the National Academy of Sciences of the United States of America* 106, no. 30 (2009): 12273-8.
36. Brozynska, M., A. Furtado, and R. J. Henry. "Genomics of Crop Wild Relatives: Expanding the Gene Pool for Crop Improvement." *Plant Biotechnology Journal* 14, no. 4 (2016): 1070-85.
37. Bohra, A., B. Kilian, S. Sivasankar, M. Caccamo, C. Mba, S. R. McCouch, and R. K. Varshney. "Reap the Crop Wild Relatives for Breeding Future Crops." *Trends in Biotechnology* 40, no. 4 (2022): 412-31.
38. McCouch, S. R., and L. H. Rieseberg. "Harnessing Crop Diversity." *Proceedings of the National Academy of Sciences of the United States of America* 120, no. 14 (2023): e2221410120.
39. McCouch, S. "Toward a Plant Genomics Initiative: Thoughts on the Value of Cross-Species and Cross-Genera Comparisons in the Grasses." *Proceedings of the National Academy of Sciences of the United States of America* 95, no. 5 (1998): 1983-5.
40. Wurschum, T., M. Rapp, T. Miedaner, C. F. H. Longin, and W. L. Leiser. "Copy Number Variation of Ppd-B1 Is the Major Determinant of Heading Time in Durum Wheat." *BMC Genetics* 20, no. 1 (2019): 64.
41. Knox, A. K., T. Dhillon, H. Cheng, A. Tondelli, N. Pecchioni, and E. J. Stockinger. "Cbf Gene Copy Number Variation at Frost Resistance-2 Is Associated with Levels of Freezing Tolerance in Temperate-Climate Cereals." *Theoretical and Applied Genetics* 121, no. 1 (2010): 21-35.
42. Maron, L. G., C. T. Guimaraes, M. Kirst, P. S. Albert, J. A. Birchler, P. J. Bradbury, E. S. Buckler, A. E. Coluccio, T. V. Danilova, D. Kudrna, J. V. Magalhaes, M. A. Pineros, M. C. Schatz, R. A. Wing, and L. V. Kochian. "Aluminum Tolerance in Maize Is Associated with Higher *Mate1* Gene Copy Number." *Proceedings of the National Academy of Sciences of the United States of America* 110, no. 13 (2013): 5241-6.
43. Cook, D. E., T. G. Lee, X. Guo, S. Melito, K. Wang, A. M. Bayless, J. Wang, T. J. Hughes, D. K. Willis, T. E. Clemente, B. W. Diers, J. Jiang, M. E. Hudson, and A. F. Bent. "Copy Number Variation of Multiple Genes at *Rhg1* Mediates Nematode Resistance in Soybean." *Science* 338, no. 6111 (2012): 1206-9.
44. Liu, Q., J. Xu, Y. Zhu, Y. Mo, X. F. Yao, R. Wang, W. Ku, Z. Huang, S. Xia, J. Tong, C. Huang, Y. Su, W. Lin, K. Peng, C. M. Liu, and L. Xiao. "The Copy Number Variation of *Osmtd1* Regulates Rice Plant Architecture." *Frontiers of Plant Science* 11 (2020): 620282.
45. Wang, Y., G. Xiong, J. Hu, L. Jiang, H. Yu, J. Xu, Y. Fang, L. Zeng, E. Xu, J. Xu, W. Ye, X. Meng, R. Liu, H. Chen, Y. Jing, Y. Wang, X. Zhu, J. Li, and Q. Qian. "Copy Number Variation at the *Gl7* Locus Contributes to Grain Size Diversity in Rice." *Nature Genetics* 47, no. 8 (2015): 944-8.
46. Bosman, R. N., J. A. Vervalle, D. L. November, P. Burger, and J. G. Lashbrooke. "Grapevine Genome Analysis Demonstrates the Role of Gene Copy Number Variation in the Formation of Monoterpenes." *Frontiers of Plant Science* 14 (2023): 1112214.
47. Falginella, L., S. D. Castellarin, R. Testolin, G. A. Gambetta, M. Morgante, and G. Di Gaspero. "Expansion and Subfunctionalisation of Flavonoid 3',5'-Hydroxylases in the Grapevine Lineage." *BMC Genomics* 11 (2010): 562.
48. Nilsen, K. T., S. Walkowiak, D. Xiang, P. Gao, T. D. Quilichini, I. R. Willick, B. Byrns, A. N'Diaye, J. Ens, K. Wiebe, Y. Ruan, R. D. Cuthbert, M. Craze, E. J. Wallington, J. Simmonds, C. Uauy, R. Datla, and C. J. Pozniak. "Copy Number Variation of *Tdd8* Controls Solid-Stemmed Architecture in Wheat." *Proceedings of the National Academy of Sciences of the United States of America* 117, no. 46 (2020): 28708-18.
49. Gao, L., I. Gonda, H. Sun, Q. Ma, K. Bao, D. M. Tieman, E. A. Burzynski-Chang, T. L. Fish, K. A. Stromberg, G. L. Sacks, T. W. Thannhauser, M. R. Foolad, M. J. Diez, J. Blanca, J. Canizares, Y. Xu, E. van der Knaap, S. Huang, H. J. Klee, J. J. Giovannoni, and Z. Fei. "The Tomato Pan-Genome Uncovers New Genes and a Rare Allele Regulating Fruit Flavor." *Nature Genetics* 51, no. 6 (2019): 1044-51.
50. Liu, J., and R. K. Dawe. "Large Haplotypes Highlight a Complex Age Structure within the Maize Pan-Genome." *Genome Research* 33, no. 3 (2023): 359-70.
51. Tao, Y., X. Zhao, E. Mace, R. Henry, and D. Jordan. "Exploring and Exploiting Pan-Genomics for Crop Improvement." *Mol Plant* 12, no. 2 (2019): 156-69.
52. Bayer, P. E., A. A. Golicz, A. Scheben, J. Batley, and D. Edwards. "Plant Pan-Genomes Are the New Reference." *Nat Plants* 6, no. 8 (2020): 914-20.
53. Jayakodi, M., M. Schreiber, N. Stein, and M. Mascher. "Building Pan-Genome Infrastructures for Crop Plants and Their Use in Association Genetics." *DNA Research* 28, no. 1 (2021).

54. Li, W., J. Liu, H. Zhang, Z. Liu, Y. Wang, L. Xing, Q. He, and H. Du. "Plant Pan-Genomics: Recent Advances, New Challenges, and Roads Ahead." *Journal of Genetics and Genomics* 49, no. 9 (2022): 833-46.
55. Yan, H., M. Sun, Z. Zhang, Y. Jin, A. Zhang, C. Lin, B. Wu, M. He, B. Xu, J. Wang, P. Qin, J. P. Mendieta, G. Nie, J. Wang, C. S. Jones, G. Feng, R. K. Srivastava, X. Zhang, A. Bombarely, D. Luo, L. Jin, Y. Peng, X. Wang, Y. Ji, S. Tian, and L. Huang. "Pangenomic Analysis Identifies Structural Variation Associated with Heat Tolerance in Pearl Millet." *Nature Genetics* 55, no. 3 (2023): 507-18.
56. Zhou, H., F. Yan, F. Hao, H. Ye, M. Yue, K. Woeste, P. Zhao, and S. Zhang. "Pan-Genome and Transcriptome Analyses Provide Insights into Genomic Variation and Differential Gene Expression Profiles Related to Disease Resistance and Fatty Acid Biosynthesis in Eastern Black Walnut (*Juglans Nigra*)." *Hortic Res* 10, no. 3 (2023): uhad015.
57. Golicz, A. A., J. Batley, and D. Edwards. "Towards Plant Pangenomics." *Plant Biotechnology Journal* 14, no. 4 (2016): 1099-105.
58. Garrison, E., J. Siren, A. M. Novak, G. Hickey, J. M. Eizenga, E. T. Dawson, W. Jones, S. Garg, C. Markello, M. F. Lin, B. Paten, and R. Durbin. "Variation Graph Toolkit Improves Read Mapping by Representing Genetic Variation in the Reference." *Nature Biotechnology* 36, no. 9 (2018): 875-79.
59. Rakocevic, G., V. Semenyuk, W. P. Lee, J. Spencer, J. Browning, I. J. Johnson, V. Arsenijevic, J. Nadj, K. Ghose, M. C. Suci, S. G. Ji, G. Demir, L. Li, B. C. Toptas, A. Dolgoborodov, B. Pollex, I. Spulber, I. Glotova, P. Komar, A. L. Stachyra, Y. Li, M. Popovic, M. Kallberg, A. Jain, and D. Kural. "Fast and Accurate Genomic Analyses Using Genome Graphs." *Nature Genetics* 51, no. 2 (2019): 354-62.
60. Cheng, H., G. T. Concepcion, X. Feng, H. Zhang, and H. Li. "Haplotype-Resolved De Novo Assembly Using Phased Assembly Graphs with Hifiasm." *Nature Methods* 18, no. 2 (2021): 170-75.
61. Padgitt-Cobb, L. K., S. B. Kingan, J. Wells, J. Elser, B. Kronmiller, D. Moore, G. Concepcion, P. Peluso, D. Rank, P. Jaiswal, J. Henning, and D. A. Hendrix. "A Draft Phased Assembly of the Diploid Cascade Hop (*Humulus Lupulus*) Genome." *Plant Genome* 14, no. 1 (2021): e20072.
62. Eizenga, J. M., A. M. Novak, J. A. Sibbesen, S. Heumos, A. Ghaffaari, G. Hickey, X. Chang, J. D. Seaman, R. Rounthwaite, J. Ebler, M. Rautiainen, S. Garg, B. Paten, T. Marschall, J. Siren, and E. Garrison. "Pangenome Graphs." *Annu Rev Genomics Hum Genet* 21 (2020): 139-62.
63. Hickey, G., D. Heller, J. Monlong, J. A. Sibbesen, J. Siren, J. Eizenga, E. T. Dawson, E. Garrison, A. M. Novak, and B. Paten. "Genotyping Structural Variants in Pangenome Graphs Using the Vg Toolkit." *Genome Biology* 21, no. 1 (2020): 35.
64. Vernikos, G. S. "A Review of Pangenome Tools and Recent Studies." In *The Pangenome: Diversity, Dynamics and Evolution of Genomes*, edited by H. Tettelin and D. Medini, 89-112. Cham (CH), 2020.
65. Glick, L., and I. Mayrose. "The Effect of Methodological Considerations on the Construction of Gene-Based Plant Pan-Genomes." *Genome Biol Evol* 15, no. 7 (2023).
66. Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, and A. M. Phillippy. "Canu: Scalable and Accurate Long-Read Assembly Via Adaptive K-Mer Weighting and Repeat Separation." *Genome Research* 27, no. 5 (2017): 722-36.
67. Kolmogorov, M., J. Yuan, Y. Lin, and P. A. Pevzner. "Assembly of Long, Error-Prone Reads Using Repeat Graphs." *Nature Biotechnology* 37, no. 5 (2019): 540-46.
68. Swain, M. T., I. J. Tsai, S. A. Assefa, C. Newbold, M. Berriman, and T. D. Otto. "A Post-Assembly Genome-Improvement Toolkit (Pagit) to Obtain Annotated Genomes from Contigs." *Nat Protoc* 7, no. 7 (2012): 1260-84.
69. Li, D., C. M. Liu, R. Luo, K. Sadakane, and T. W. Lam. "Megahit: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly Via Succinct De Bruijn Graph." *Bioinformatics* 31, no. 10 (2015): 1674-6.
70. Tolstoganov, I., A. Bankevich, Z. Chen, and P. A. Pevzner. "Cloudspades: Assembly of Synthetic Long Reads Using De Bruijn Graphs." *Bioinformatics* 35, no. 14 (2019): i61-i70.
71. Meleshko, D., H. Mohimani, V. Tracanna, I. Hajirasouliha, M. H. Medema, A. Korobeynikov, and P. A. Pevzner. "Biosyntheticspades: Reconstructing Biosynthetic Gene Clusters from Assembly Graphs." *Genome Research* 29, no. 8 (2019): 1352-62.
72. Li, H., X. Feng, and C. Chu. "The Design and Construction of Reference Pangenome Graphs with Minigraph." *Genome Biology* 21, no. 1 (2020): 265.
73. Garrison, E., A. Guarracino, S. Heumos, F. Villani, Z. Bao, L. Tattini, J. Hagmann, S. Vorbrugg, S. Marco-Sola, C. Kubica, D. G. Ashbrook, K. Thorell, R. L. Rusholme-Pilcher, G. Liti, E. Rudbeck, S. Nahnsen, Z.

- Yang, M. N. Moses, F. L. Nobrega, Y. Wu, H. Chen, J. de Ligt, P. H. Sudmant, N. Soranzo, V. Colonna, R. W. Williams, and P. Prins. "Building Pangenome Graphs." *bioRxiv* (2023).
74. Hickey, G., J. Monlong, J. Ebler, A. M. Novak, J. M. Eizenga, Y. Gao, Consortium Human Pangenome Reference, T. Marschall, H. Li, and B. Paten. "Pangenome Graph Construction from Genome Alignments with Minigraph-Cactus." *Nature Biotechnology* (2023).
 75. Armstrong, J., G. Hickey, M. Diekhans, I. T. Fiddes, A. M. Novak, A. Deran, Q. Fang, D. Xie, S. Feng, J. Stiller, D. Genereux, J. Johnson, V. D. Marinescu, J. Alföldi, R. S. Harris, K. Lindblad-Toh, D. Haussler, E. Karlsson, E. D. Jarvis, G. Zhang, and B. Paten. "Progressive Cactus Is a Multiple-Genome Aligner for the Thousand-Genome Era." *Nature* 587, no. 7833 (2020): 246-51.
 76. Jonkheer, E. M., D. M. van Workum, S. Sheikhezadeh Anari, B. Brankovics, J. R. de Haan, L. Berke, T. A. J. van der Lee, D. de Ridder, and S. Smit. "Pantools V3: Functional Annotation, Classification and Phylogenomics." *Bioinformatics* 38, no. 18 (2022): 4403-05.
 77. Guarracino, A., S. Heumos, S. Nahnsen, P. Prins, and E. Garrison. "Odgi: Understanding Pangenome Graphs." *Bioinformatics* 38, no. 13 (2022): 3319-26.
 78. Ewels, P. A., A. Peltzer, S. Fillinger, H. Patel, J. Alneberg, A. Wilm, M. U. Garcia, P. Di Tommaso, and S. Nahnsen. "The Nf-Core Framework for Community-Curated Bioinformatics Pipelines." *Nature Biotechnology* 38, no. 3 (2020): 276-78.
 79. Vaughn, J. N., S. E. Branham, B. Abernathy, A. M. Hulse-Kemp, A. R. Rivers, A. Levi, and W. P. Wechter. "Graph-Based Pangenomics Maximizes Genotyping Density and Reveals Structural Impacts on Fungal Resistance in Melon." *Nat Commun* 13, no. 1 (2022): 7897.
 80. Li, H. "Minimap2: Pairwise Alignment for Nucleotide Sequences." *Bioinformatics* 34, no. 18 (2018): 3094-100.
 81. Marcais, G., A. L. Delcher, A. M. Phillippy, R. Coston, S. L. Salzberg, and A. Zimin. "Mummer4: A Fast and Versatile Genome Alignment System." *PLoS Comput Biol* 14, no. 1 (2018): e1005944.
 82. Rautiainen, M., and T. Marschall. "Graphaligner: Rapid and Versatile Sequence-to-Graph Alignment." *Genome Biology* 21, no. 1 (2020): 253.
 83. Kavya, V. N. S., K. Tayal, R. Srinivasan, and N. Sivadasan. "Sequence Alignment on Directed Graphs." *Journal of Computational Biology* 26, no. 1 (2019): 53-67.
 84. Buchler, T., J. Olbrich, and E. Ohlebusch. "Efficient Short Read Mapping to a Pangenome That Is Represented by a Graph of Ed Strings." *Bioinformatics* 39, no. 5 (2023).
 85. Poplin, R., P. C. Chang, D. Alexander, S. Schwartz, T. Colthurst, A. Ku, D. Newburger, J. Dijamco, N. Nguyen, P. T. Afshar, S. S. Gross, L. Dorfman, C. Y. McLean, and M. A. DePristo. "A Universal Snp and Small-Indel Variant Caller Using Deep Neural Networks." *Nature Biotechnology* 36, no. 10 (2018): 983-87.
 86. Yun, T., H. Li, P. C. Chang, M. F. Lin, A. Carroll, and C. Y. McLean. "Accurate, Scalable Cohort Variant Calls Using Deepvariant and Glnexus." *Bioinformatics* 36, no. 24 (2021): 5582-89.
 87. Chiang, C., R. M. Layer, G. G. Faust, M. R. Lindberg, D. B. Rose, E. P. Garrison, G. T. Marth, A. R. Quinlan, and I. M. Hall. "Speedseq: Ultra-Fast Personal Genome Analysis and Interpretation." *Nature Methods* 12, no. 10 (2015): 966-8.
 88. Eggertsson, H. P., H. Jonsson, S. Kristmundsdottir, E. Hjartarson, B. Kehr, G. Masson, F. Zink, K. E. Hjorleifsson, A. Jonasdottir, A. Jonasdottir, I. Jonsdottir, D. F. Gudbjartsson, P. Melsted, K. Stefansson, and B. V. Halldorsson. "GraphTyper Enables Population-Scale Genotyping Using Pangenome Graphs." *Nature Genetics* 49, no. 11 (2017): 1654-60.
 89. Ebler, J., P. Ebert, W. E. Clarke, T. Rausch, P. A. Audano, T. Houwaart, Y. Mao, J. O. Korbel, E. E. Eichler, M. C. Zody, A. T. Dilthey, and T. Marschall. "Pangenome-Based Genome Inference Allows Efficient and Accurate Genotyping across a Wide Spectrum of Variant Classes." *Nature Genetics* 54, no. 4 (2022): 518-25.
 90. Naithani, S., M. Geniza, and P. Jaiswal. "Variant Effect Prediction Analysis Using Resources Available at Gramene Database." *Methods in Molecular Biology* 1533 (2017): 279-97.
 91. Emms, D. M., and S. Kelly. "Orthofinder: Phylogenetic Orthology Inference for Comparative Genomics." *Genome Biology* 20, no. 1 (2019): 238.
 92. Li, L., C. J. Stoeckert, Jr., and D. S. Roos. "Orthomcl: Identification of Ortholog Groups for Eukaryotic Genomes." *Genome Research* 13, no. 9 (2003): 2178-89.
 93. Miller, J. B., B. D. Pickett, and P. G. Ridge. "Justorthologs: A Fast, Accurate and User-Friendly Ortholog Identification Algorithm." *Bioinformatics* 35, no. 4 (2019): 546-52.
 94. Zhou, S., Y. Chen, C. Guo, and J. Qi. "Phylomcl: Accurate Clustering of Hierarchical Orthogroups Guided by Phylogenetic Relationship and Inference of Polyploidy Events." *Methods Ecol. Evol.* 11 (2020): 943-54.

95. Altenhoff, A. M., C. M. Train, K. J. Gilbert, I. Mediratta, T. Mendes de Farias, D. Moi, Y. Nevers, H. S. Radoykova, V. Rossier, A. Warwick Vesztrocy, N. M. Glover, and C. Dessimoz. "Oma Orthology in 2021: Website Overhaul, Conserved Isoforms, Ancestral Gene Order and More." *Nucleic Acids Res* 49, no. D1 (2021): D373-D79.
96. Persson, E., and E. L. L. Sonnhammer. "Inparanoid-Diamond: Faster Orthology Analysis with the Inparanoid Algorithm." *Bioinformatics* 38, no. 10 (2022): 2918-19.
97. Naithani, S., P. Gupta, J. Preece, P. D'Eustachio, J. L. Elser, P. Garg, D. A. Dikeman, J. Kiff, J. Cook, A. Olson, S. Wei, M. K. Tello-Ruiz, A. F. Mundo, A. Munoz-Pomer, S. Mohammed, T. Cheng, E. Bolton, I. Papatheodorou, L. Stein, D. Ware, and P. Jaiswal. "Plant Reactome: A Knowledgebase and Resource for Comparative Pathway Analysis." *Nucleic Acids Res* 48, no. D1 (2020): D1093-D103.
98. Durant, E., F. Sabot, M. Conte, and M. Rouard. "Panache: A Web Browser-Based Viewer for Linearized Pangenomes." *Bioinformatics* 37, no. 23 (2021): 4556-58.
99. Droc, G., G. Martin, V. Guignon, M. Summo, G. Sempere, E. Durant, A. Soriano, F. C. Baurens, A. Cenci, C. Breton, T. Shah, J. M. Aury, X. J. Ge, P. H. Harrison, N. Yahiaoui, A. D'Hont, and M. Rouard. "The Banana Genome Hub: A Community Database for Genomics in the Musaceae." *Hortic Res* 9 (2022): uhac221.
100. Yokoyama, T. T., Y. Sakamoto, M. Seki, Y. Suzuki, and M. Kasahara. "Momi-G: Modular Multi-Scale Integrated Genome Graph Browser." *BMC Bioinformatics* 20, no. 1 (2019): 548.
101. Wick, R. R., M. B. Schultz, J. Zobel, and K. E. Holt. "Bandage: Interactive Visualization of De Novo Genome Assemblies." *Bioinformatics* 31, no. 20 (2015): 3350-2.
102. Beyer, W., A. M. Novak, G. Hickey, J. Chan, V. Tan, B. Paten, and D. R. Zerbino. "Sequence Tube Maps: Making Graph Genomes Intuitive to Commuters." *Bioinformatics* 35, no. 24 (2019): 5318-20.
103. Gonnella, G., N. Niehus, and S. Kurtz. "Gfavis: Flexible and Interactive Visualization of Gfa Sequence Graphs." *Bioinformatics* 35, no. 16 (2019): 2853-55.
104. Mikheenko, A., and M. Kolmogorov. "Assembly Graph Browser: Interactive Visualization of Assembly Graphs." *Bioinformatics* 35, no. 18 (2019): 3476-78.
105. Kunyavskaya, O., and A. D. Pribelski. "Sgtk: A Toolkit for Visualization and Assessment of Scaffold Graphs." *Bioinformatics* 35, no. 13 (2019): 2303-05.
106. Durbin, R. "Efficient Haplotype Matching and Storage Using the Positional Burrows-Wheeler Transform (Pbwt)." *Bioinformatics* 30, no. 9 (2014): 1266-72.
107. Novak, A. M., E. Garrison, and B. Paten. "A Graph Extension of the Positional Burrows-Wheeler Transform and Its Applications." *Algorithms Mol Biol* 12 (2017): 18.
108. Grytten, I., K. D. Rand, A. J. Nederbragt, G. O. Storvik, I. K. Glad, and G. K. Sandve. "Graph Peak Caller: Calling Chip-Seq Peaks on Graph-Based Reference Genomes." *PLoS Comput Biol* 15, no. 2 (2019): e1006731.
109. Wang, J., W. Yang, S. Zhang, H. Hu, Y. Yuan, J. Dong, L. Chen, Y. Ma, T. Yang, L. Zhou, J. Chen, B. Liu, C. Li, D. Edwards, and J. Zhao. "A Pangenome Analysis Pipeline Provides Insights into Functional Gene Identification in Rice." *Genome Biology* 24, no. 1 (2023): 19.
110. Tahir Ul Qamar, M., X. Zhu, F. Xing, and L. L. Chen. "Ppspcp: A Plant Presence/Absence Variants Scanner and Pan-Genome Construction Pipeline." *Bioinformatics* 35, no. 20 (2019): 4156-58.
111. Harper, L., J. Campbell, E. K.S. Cannon, S. Jung, M. Poelchau, R. Walls, C. Andorf, E. Arnaud, T. Berardini, C. Birkett, S. Cannon, J. Carson, B. Condon, L. Cooper, N. Dunn, C. G. Elsik, A. Farmer, S. Ficklin, D. Grant, E. Grau, N. Herndon, Z.-L. Hu, J. Humann, P. Jaiswal, C. Jonquet, M.-A. Laporte, P. Larmande, G. Lazo, F. McCarthy, N. Menda, C. Mungall, M. Munoz-Torres, S. Naithani, R. Nelson, D. Nesdill, C. Park, J. Reecy, L. Reiser, L.-A. Sanderson, T. Z. Sen, M. Staton, S. Subramaniam, M. K. Tello-Ruiz, V. Unda, D. Unni, L. Wang, D. Ware, J. Wegrzyn, J. Williams, M. Woodhouse, J. Yu, and D. Main. "Agbiodata Consortium Recommendations for Sustainable Genomics and Genetics Databases for Agriculture." *Database (Oxford)* (accepted) (2018).
112. Adam-Blondon, A. F., M. Alaux, C. Pommier, D. Cantu, Z. M. Cheng, G. R. Cramer, C. Davies, S. Delrot, L. Deluc, G. Di Gaspero, J. Grimplet, A. Fennell, J. P. Londo, P. Kersey, F. Mattivi, S. Naithani, P. Neveu, M. Nikolski, M. Pezzotti, B. I. Reisch, R. Topfer, M. A. Vivier, D. Ware, and H. Quesneville. "Towards an Open Grapevine Information System." *Hortic Res* 3 (2016): 16056.
113. Bolser, D., D. M. Staines, E. Pritchard, and P. Kersey. "Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data." *Methods in Molecular Biology* 1374 (2016): 115-40.

114. Gupta, P., S. Naithani, J. Preece, S. Kim, T. Cheng, P. D'Eustachio, J. Elser, E. E. Bolton, and P. Jaiswal. "Plant Reactome and Pubchem: The Plant Pathway and (Bio)Chemical Entity Knowledgebases." *Methods in Molecular Biology* 2443 (2022): 511-25.
115. Tello-Ruiz, M. K., S. Naithani, P. Gupta, A. Olson, S. Wei, J. Preece, Y. Jiao, B. Wang, K. Chougule, P. Garg, J. Elser, S. Kumari, V. Kumar, B. Contreras-Moreira, G. Naamati, N. George, J. Cook, D. Bolser, P. D'Eustachio, L.D. Stein, A. Gupta, W. Xu, J. Regala, I. Papatheodorou, P. J. Kersey, C. Taylor, P. Jaiswal, and D. Ware. "Gramene 2021: Harnessing the Power of Comparative Genomics and Pathways for Plant Research." *Nucleic Acids Res* 49, no. D1 (2021): D1452-63.
116. Pasha, A., S. Subramaniam, A. Cleary, X. Chen, T. Berardini, A. Farmer, C. Town, and N. Provart. "Araport Lives: An Updated Framework for Arabidopsis Bioinformatics." *Plant Cell* 32, no. 9 (2020): 2683-86.
117. Shamimuzzaman, M., J. M. Gardiner, A. T. Walsh, D. A. Triant, J. J. Le Tourneau, A. Tayal, D. R. Unni, H. N. Nguyen, J. L. Portwood, 2nd, E. K. S. Cannon, C. M. Andorf, and C. G. Elsik. "Maizemine: A Data Mining Warehouse for the Maize Genetics and Genomics Database." *Frontiers of Plant Science* 11 (2020): 592730.
118. Gladman, N., A. Olson, S. Wei, K. Chougule, Z. Lu, M. Tello-Ruiz, I. Meijs, P. Van Buren, Y. Jiao, B. Wang, V. Kumar, S. Kumari, L. Zhang, J. Burke, J. Chen, G. Burow, C. Hayes, Y. Emendack, Z. Xin, and D. Ware. "Sorghumbase: A Web-Based Portal for Sorghum Genetic Information and Community Advancement." *Planta* 255, no. 2 (2022): 35.
119. Arkin, A. P., R. W. Cottingham, C. S. Henry, N. L. Harris, R. L. Stevens, S. Maslov, P. Dehal, D. Ware, F. Perez, S. Canon, M. W. Sneddon, M. L. Henderson, W. J. Riehl, D. Murphy-Olson, S. Y. Chan, R. T. Kamimura, S. Kumari, M. M. Drake, T. S. Brettin, E. M. Glass, D. Chivian, D. Gunter, D. J. Weston, B. H. Allen, J. Baumohl, A. A. Best, B. Bowen, S. E. Brenner, C. C. Bun, J. M. Chandonia, J. M. Chia, R. Colasanti, N. Conrad, J. J. Davis, B. H. Davison, M. DeJongh, S. Devoid, E. Dietrich, I. Dubchak, J. N. Edirisinghe, G. Fang, J. P. Faria, P. M. Frybarger, W. Gerlach, M. Gerstein, A. Greiner, J. Gurtowski, H. L. Haun, F. He, R. Jain, M. P. Joachimiak, K. P. Keegan, S. Kondo, V. Kumar, M. L. Land, F. Meyer, M. Mills, P. S. Novichkov, T. Oh, G. J. Olsen, R. Olson, B. Parrello, S. Pasternak, E. Pearson, S. S. Poon, G. A. Price, S. Ramakrishnan, P. Ranjan, P. C. Ronald, M. C. Schatz, S. M. D. Seaver, M. Shukla, R. A. Sutormin, M. H. Syed, J. Thomason, N. L. Tintle, D. Wang, F. Xia, H. Yoo, S. Yoo, and D. Yu. "Kbase: The United States Department of Energy Systems Biology Knowledgebase." *Nature Biotechnology* 36, no. 7 (2018): 566-69.
120. Yates, A. D., J. Allen, R. M. Amode, A. G. Azov, M. Barba, A. Becerra, J. Bhai, L. I. Campbell, M. Carbajo Martinez, M. Chakiachvili, K. Chougule, M. Christensen, B. Contreras-Moreira, A. Cuzick, L. Da Rin Fioretto, P. Davis, N. H. De Silva, S. Diamantakis, S. Dyer, J. Elser, C. V. Filippi, A. Gall, D. Grigoriadis, C. Guijarro-Clarke, P. Gupta, K. E. Hammond-Kosack, K. L. Howe, P. Jaiswal, V. Kaikala, V. Kumar, S. Kumari, N. Langridge, T. Le, M. Luypaert, G. L. Maslen, T. Maurel, B. Moore, M. Muffato, A. Mushtaq, G. Naamati, S. Naithani, A. Olson, A. Parker, M. Paulini, H. Pedro, E. Perry, J. Preece, M. Quinton-Tulloch, F. Rodgers, M. Rosello, M. Ruffier, J. Seager, V. Sitnik, M. Szpak, J. Tate, M. K. Tello-Ruiz, S. J. Trevanion, M. Urban, D. Ware, S. Wei, G. Williams, A. Winterbottom, M. Zarowiecki, R. D. Finn, and P. Flicek. "Ensembl Genomes 2022: An Expanding Genome Resource for Non-Vertebrates." *Nucleic Acids Res* 50, no. D1 (2022): D996-D1003.
121. Naithani, S., J. Preece, P. D'Eustachio, P. Gupta, V. Amarasinghe, P. D. Dharmawardhana, G. Wu, A. Fabregat, J. L. Elser, J. Weiser, M. Keays, A. M. Fuentes, R. Petryszak, L. D. Stein, D. Ware, and P. Jaiswal. "Plant Reactome: A Resource for Plant Pathways and Comparative Analysis." *Nucleic Acids Res* 45, no. D1 (2017): D1029-D39.
122. Tello-Ruiz, M. K., S. Naithani, J. C. Stein, P. Gupta, M. Campbell, A. Olson, S. Wei, J. Preece, M. J. Geniza, Y. Jiao, Y. K. Lee, B. Wang, J. Mulvaney, K. Chougule, J. Elser, N. Al-Bader, S. Kumari, J. Thomason, V. Kumar, D. M. Bolser, G. Naamati, E. Tapanari, N. Fonseca, L. Huerta, H. Iqbal, M. Keays, A. Munoz-Pomer Fuentes, A. Tang, A. Fabregat, P. D'Eustachio, J. Weiser, L. D. Stein, R. Petryszak, I. Papatheodorou, P. J. Kersey, P. Lockhart, C. Taylor, P. Jaiswal, and D. Ware. "Gramene 2018: Unifying Comparative Genomics and Pathway Resources for Plant Research." *Nucleic Acids Res* 46, no. D1 (2018): D1181-D89.
123. Naithani, S., R. Raja, E. N. Waddell, J. Elser, S. Gouthu, L. G. Deluc, and P. Jaiswal. "Vitisyc: A Metabolic Pathway Knowledgebase for Grapevine (*Vitis Vinifera*)." *Frontiers of Plant Science* 5 (2014): 644.
124. Naithani, S., C. M. Partipilo, R. Raja, J. L. Elser, and P. Jaiswal. "Fragariacyc: A Metabolic Pathway Database for Woodland Strawberry *Fragaria Vesca*." *Frontiers in Plant Science* 7 (2016).

125. Woodhouse, M. R., E. K. Cannon, J. L. Portwood, 2nd, L. C. Harper, J. M. Gardiner, M. L. Schaeffer, and C. M. Andorf. "A Pan-Genomic Approach to Genome Databases Using Maize as a Model System." *BMC Plant Biology* 21, no. 1 (2021): 385.
126. Kanehisa, M., M. Furumichi, Y. Sato, M. Kawashima, and M. Ishiguro-Watanabe. "Kegg for Taxonomy-Based Analysis of Pathways and Genomes." *Nucleic Acids Res* 51, no. D1 (2023): D587-D92.
127. Paley, S., and P. D. Karp. "The Biocyc Metabolic Network Explorer." *BMC Bioinformatics* 22, no. 1 (2021): 208.
128. Naithani, S., and P. Jaiswal. "Pathway Analysis and Omics Data Visualization Using Pathway Genome Databases: Fragariacyc, a Case Study." *Methods in Molecular Biology* 1533 (2017): 241-56.
129. Hawkins, C., D. Ginzburg, K. Zhao, W. Dwyer, B. Xue, A. Xu, S. Rice, B. Cole, S. Paley, P. Karp, and S. Y. Rhee. "Plant Metabolic Network 15: A Resource of Genome-Wide Metabolism Databases for 126 Plants and Algae." *Journal of Integrative Plant Biology* 63, no. 11 (2021): 1888-905.
130. Foerster, H., A. Bombarely, J. N. D. Battey, N. Sierro, N. V. Ivanov, and L. A. Mueller. "Solcyc: A Database Hub at the Sol Genomics Network (Sgn) for the Manual Curation of Metabolic Networks in Solanum and Nicotiana Specific Databases." *Database (Oxford)* 2018 (2018).
131. Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, and D. S. Rokhsar. "Phytozome: A Comparative Platform for Green Plant Genomics." *Nucleic Acids Res* 40, no. Database issue (2012): D1178-86.
132. Deng, C.H., S. Naithani, S. Kumari, I. Cobo-Simon, E.H. Quezada-Rodriguez, M. Skrabisova, N. Gladman, M.J. Correll, A.B. Sikiru, O.O. Afuwape, A. Marrano, I. Rebollo, W. Zhang, and S. Jung. "Agricultural Sciences in the Big Data Era: Genotype and Phenotype Data Standardization, Utilization and Integration." *Preprints - American Chemical Society, Division of Petroleum Chemistry* (2023).
133. Sun, C., Z. Hu, T. Zheng, K. Lu, Y. Zhao, W. Wang, J. Shi, C. Wang, J. Lu, D. Zhang, Z. Li, and C. Wei. "Rpan: Rice Pan-Genome Browser for Approximately 3000 Rice Genomes." *Nucleic Acids Res* 45, no. 2 (2017): 597-605.
134. Zhao, Q., Q. Feng, H. Lu, Y. Li, A. Wang, Q. Tian, Q. Zhan, Y. Lu, L. Zhang, T. Huang, Y. Wang, D. Fan, Y. Zhao, Z. Wang, C. Zhou, J. Chen, C. Zhu, W. Li, Q. Weng, Q. Xu, Z. X. Wang, X. Wei, B. Han, and X. Huang. "Pan-Genome Analysis Highlights the Extent of Genomic Variation in Cultivated and Wild Rice." *Nature Genetics* 50, no. 2 (2018): 278-84.
135. Gui, S., L. Yang, J. Li, J. Luo, X. Xu, J. Yuan, L. Chen, W. Li, X. Yang, S. Wu, S. Li, Y. Wang, Y. Zhu, Q. Gao, N. Yang, and J. Yan. "Zeamap, a Comprehensive Database Adapted to the Maize Multi-Omics Era." *iScience* 23, no. 6 (2020): 101241.
136. Valentin, G., T. Abdel, D. Gaetan, D. Jean-Francois, C. Matthieu, and R. Mathieu. "Greenphyldb V5: A Comparative Pangenomic Database for Plant Genomes." *Nucleic Acids Res* 49, no. D1 (2021): D1464-D71.
137. Bayer, P. E., J. Petereit, E. Durant, C. Monat, M. Rouard, H. Hu, B. Chapman, C. Li, S. Cheng, J. Batley, and D. Edwards. "Wheat Panache: A Pangenome Graph Database Representing Presence-Absence Variation across Sixteen Bread Wheat Genomes." *Plant Genome* 15, no. 3 (2022): e20221.
138. Blake, V. C., M. R. Woodhouse, G. R. Lazo, S. G. Odell, C. P. Wight, N. A. Tinker, Y. Wang, Y. Q. Gu, C. L. Birkett, J. L. Jannink, D. E. Matthews, D. L. Hane, S. L. Michel, E. Yao, and T. Z. Sen. "Graingenes: Centralized Small Grain Resources and Digital Platform for Geneticists and Breeders." *Database (Oxford)* 2019 (2019).
139. Montenegro, J. D., A. A. Golicz, P. E. Bayer, B. Hurgobin, H. Lee, C. K. Chan, P. Visendi, K. Lai, J. Dolezel, J. Batley, and D. Edwards. "The Pangenome of Hexaploid Bread Wheat." *Plant Journal* 90, no. 5 (2017): 1007-13.
140. Li, N., Q. He, J. Wang, B. Wang, J. Zhao, S. Huang, T. Yang, Y. Tang, S. Yang, P. Aisimutuola, R. Xu, J. Hu, C. Jia, K. Ma, Z. Li, F. Jiang, J. Gao, H. Lan, Y. Zhou, X. Zhang, S. Huang, Z. Fei, H. Wang, H. Li, and Q. Yu. "Super-Pangenome Analyses Highlight Genomic Diversity and Structural Variation across Wild and Cultivated Tomato Species." *Nature Genetics* 55, no. 5 (2023): 852-60.
141. Barchi, L., M. T. Rabanus-Wallace, J. Prohens, L. Toppino, S. Padmarasu, E. Portis, G. L. Rotino, N. Stein, S. Lanteri, and G. Giuliano. "Improved Genome Assembly and Pan-Genome Provide Key Insights into Eggplant Domestication and Breeding." *Plant Journal* 107, no. 2 (2021): 579-96.
142. Ou, L., D. Li, J. Lv, W. Chen, Z. Zhang, X. Li, B. Yang, S. Zhou, S. Yang, W. Li, H. Gao, Q. Zeng, H. Yu, B. Ouyang, F. Li, F. Liu, J. Zheng, Y. Liu, J. Wang, B. Wang, X. Dai, Y. Ma, and X. Zou. "Pan-Genome of

- Cultivated Pepper (*Capsicum*) and Its Use in Gene Presence-Absence Variation Analyses." *New Phytologist* 220, no. 2 (2018): 360-63.
143. Zhang, B., H. Huang, L. E. Tibbs-Cortes, A. Vanous, Z. Zhang, K. Sanguinet, K. A. Garland-Campbell, J. Yu, and X. Li. "Streamline Unsupervised Machine Learning to Survey and Graph Indel-Based Haplotypes from Pan-Genomes." *Mol Plant* 16, no. 6 (2023): 975-78.
 144. Torkamaneh, D., M. A. Lemay, and F. Belzile. "The Pan-Genome of the Cultivated Soybean (Pansoy) Reveals an Extraordinarily Conserved Gene Content." *Plant Biotechnology Journal* 19, no. 9 (2021): 1852-62.
 145. Hubner, S., N. Bercovich, M. Todesco, J. R. Mandel, J. Odenheimer, E. Ziegler, J. S. Lee, G. J. Baute, G. L. Owens, C. J. Grassa, D. P. Ebert, K. L. Ostevik, B. T. Moyers, S. Yakimowski, R. R. Masalia, L. Gao, I. Calic, J. E. Bowers, N. C. Kane, D. Z. H. Swanevelter, T. Kubach, S. Munos, N. B. Langlade, J. M. Burke, and L. H. Rieseberg. "Sunflower Pan-Genome Analysis Shows That Hybridization Altered Gene Content and Disease Resistance." *Nat Plants* 5, no. 1 (2019): 54-62.
 146. Jin, S., Z. Han, Y. Hu, Z. Si, F. Dai, L. He, Y. Cheng, Y. Li, T. Zhao, L. Fang, and T. Zhang. "Structural Variation (Sv)-Based Pan-Genome and Gwas Reveal the Impacts of Svs on the Speciation and Diversification of Allotetraploid Cottons." *Mol Plant* 16, no. 4 (2023): 678-93.
 147. Liu, H., X. Wang, S. Liu, Y. Huang, Y. X. Guo, W. Z. Xie, H. Liu, M. Tahir Ul Qamar, Q. Xu, and L. L. Chen. "Citrus Pan-Genome to Breeding Database (Cpbd): A Comprehensive Genome Database for Citrus Breeding." *Mol Plant* 15, no. 10 (2022): 1503-05.
 148. Li, Q., J. Qi, X. Qin, W. Dou, T. Lei, A. Hu, R. Jia, G. Jiang, X. Zou, Q. Long, L. Xu, A. Peng, L. Yao, S. Chen, and Y. He. "Citgvd: A Comprehensive Database of Citrus Genomic Variations." *Hortic Res* 7 (2020): 12.
 149. Sun, X., C. Jiao, H. Schwaninger, C. T. Chao, Y. Ma, N. Duan, A. Khan, S. Ban, K. Xu, L. Cheng, G. Y. Zhong, and Z. Fei. "Phased Diploid Genome Assemblies and Pan-Genomes Provide Insights into the Genetic History of Apple Domestication." *Nature Genetics* 52, no. 12 (2020): 1423-32.
 150. Song, J. M., D. X. Liu, W. Z. Xie, Z. Yang, L. Guo, K. Liu, Q. Y. Yang, and L. L. Chen. "Bnpir: Brassica Napus Pan-Genome Information Resource for 1689 Accessions." *Plant Biotechnology Journal* 19, no. 3 (2021): 412-14.
 151. Qi, W., Y. W. Lim, A. Patrignani, P. Schlapfer, A. Bratus-Neuenschwander, S. Gruter, C. Chanez, N. Rodde, E. Prat, S. Vautrin, M. A. Fustier, D. Pratas, R. Schlapbach, and W. Gruissem. "The Haplotype-Resolved Chromosome Pairs of a Heterozygous Diploid African Cassava Cultivar Reveal Novel Pan-Genome and Allele-Specific Transcriptome Features." *Gigascience* 11 (2022).
 152. Ruperao, P., N. Thirunavukkarasu, P. Gandham, S. Selvanayagam, M. Govindaraj, B. Nebie, E. Manyasa, R. Gupta, R. R. Das, D. A. Odeny, H. Gandhi, D. Edwards, S. P. Deshpande, and A. Rathore. "Sorghum Pan-Genome Explores the Functional Utility for Genomic-Assisted Breeding to Accelerate the Genetic Gain." *Frontiers of Plant Science* 12 (2021): 666342.
 153. Varshney, R. K., M. Roorkiwal, S. Sun, P. Bajaj, A. Chitkineni, M. Thudi, N. P. Singh, X. Du, H. D. Upadhyaya, A. W. Khan, Y. Wang, V. Garg, G. Fan, W. A. Cowling, J. Crossa, L. Gentzbittel, K. P. Voss-Fels, V. K. Valluri, P. Sinha, V. K. Singh, C. Ben, A. Rathore, R. Punna, M. K. Singh, B. Tar'an, C. Bharadwaj, M. Yasin, M. S. Pithia, S. Singh, K. R. Soren, H. Kudapa, D. Jarquin, P. Cubry, L. T. Hickey, G. P. Dixit, A. C. Thuillet, A. Hamwieh, S. Kumar, A. A. Deokar, S. K. Chaturvedi, A. Francis, R. Howard, D. Chattopadhyay, D. Edwards, E. Lyons, Y. Vigouroux, B. J. Hayes, E. von Wettberg, S. K. Datta, H. Yang, H. T. Nguyen, J. Wang, K. H. M. Siddique, T. Mohapatra, J. L. Bennetzen, X. Xu, and X. Liu. "A Chickpea Genetic Variation Map Based on the Sequencing of 3,366 Genomes." *Nature* 599, no. 7886 (2021): 622-27.
 154. Zhao, J., P. E. Bayer, P. Ruperao, R. K. Saxena, A. W. Khan, A. A. Golicz, H. T. Nguyen, J. Batley, D. Edwards, and R. K. Varshney. "Trait Associations in the Pangenome of Pigeon Pea (*Cajanus cajan*)." *Plant Biotechnology Journal* 18, no. 9 (2020): 1946-54.
 155. Yu, J., A. A. Golicz, K. Lu, K. Dossa, Y. Zhang, J. Chen, L. Wang, J. You, D. Fan, D. Edwards, and X. Zhang. "Insight into the Evolution and Functional Characteristics of the Pan-Genome Assembly from Sesame Landraces and Modern Cultivars." *Plant Biotechnology Journal* 17, no. 5 (2019): 881-92.
 156. Li, J., D. Yuan, P. Wang, Q. Wang, M. Sun, Z. Liu, H. Si, Z. Xu, Y. Ma, B. Zhang, L. Pei, L. Tu, L. Zhu, L. L. Chen, K. Lindsey, X. Zhang, S. Jin, and M. Wang. "Cotton Pan-Genome Retrieves the Lost Sequences and Genes During Domestication and Selection." *Genome Biology* 22, no. 1 (2021): 119.
 157. Sun, Y., J. Wang, Y. Li, B. Jiang, X. Wang, W. H. Xu, Y. Q. Wang, P. T. Zhang, Y. J. Zhang, and X. D. Kong. "Pan-Genome Analysis Reveals the Abundant Gene Presence/Absence Variations among Different Varieties of Melon and Their Influence on Traits." *Frontiers of Plant Science* 13 (2022): 835496.

158. Li, H., S. Wang, S. Chai, Z. Yang, Q. Zhang, H. Xin, Y. Xu, S. Lin, X. Chen, Z. Yao, Q. Yang, Z. Fei, S. Huang, and Z. Zhang. "Graph-Based Pan-Genome Reveals Structural and Sequence Variations Related to Agronomic Traits and Domestication in Cucumber." *Nat Commun* 13, no. 1 (2022): 682.
159. Qiao, Q., P. P. Edger, L. Xue, L. Qiong, J. Lu, Y. Zhang, Q. Cao, A. E. Yocca, A. E. Platts, S. J. Knapp, M. Van Montagu, Y. Van de Peer, J. Lei, and T. Zhang. "Evolutionary History and Pan-Genome Dynamics of Strawberry (*Fragaria* Spp.)." *Proceedings of the National Academy of Sciences of the United States of America* 118, no. 45 (2021).
160. Wang, H., R. Tu, Z. Ruan, C. Chen, Z. Peng, X. Zhou, L. Sun, Y. Hong, D. Chen, Q. Liu, W. Wu, X. Zhan, X. Shen, Z. Zhou, L. Cao, Y. Zhang, and S. Cheng. "Photoperiod and Gravistimulation-Associated Tiller Angle Control 1 Modulates Dynamic Changes in Rice Plant Architecture." *Theoretical and Applied Genetics* 136, no. 7 (2023): 160.
161. Yu, B., Z. Lin, H. Li, X. Li, J. Li, Y. Wang, X. Zhang, Z. Zhu, W. Zhai, X. Wang, D. Xie, and C. Sun. "Tad1, a Major Quantitative Trait Locus Controlling Tiller Angle in Rice." *Plant Journal* 52, no. 5 (2007): 891-8.
162. Boukail, S., M. Macharia, M. Miculan, A. Masoni, A. Calamai, E. Palchetti, and M. Dell'Acqua. "Genome Wide Association Study of Agronomic and Seed Traits in a World Collection of Proso Millet (*Panicum Miliaceum* L.)." *BMC Plant Biology* 21, no. 1 (2021): 330.
163. Liu, C., Y. Wang, J. Peng, B. Fan, D. Xu, J. Wu, Z. Cao, Y. Gao, X. Wang, S. Li, Q. Su, Z. Zhang, S. Wang, X. Wu, Q. Shang, H. Shi, Y. Shen, B. Wang, and J. Tian. "High-Quality Genome Assembly and Pan-Genome Studies Facilitate Genetic Discovery in Mung Bean and Its Improvement." *Plant Commun* 3, no. 6 (2022): 100352.
164. D'Hont, A., F. Denoeud, J. M. Aury, F. C. Baurens, F. Carreel, O. Garsmeur, B. Noel, S. Bocs, G. Droc, M. Rouard, C. Da Silva, K. Jabbari, C. Cardi, J. Poulain, M. Souquet, K. Labadie, C. Jourda, J. Lengelle, M. Rodier-Goud, A. Alberti, M. Bernard, M. Correa, S. Ayyampalayam, M. R. McKain, J. Leebens-Mack, D. Burgess, M. Freeling, A. Mbeguie D. Mbeguie, M. Chabannes, T. Wicker, O. Panaud, J. Barbosa, E. Hribova, P. Heslop-Harrison, R. Habas, R. Rivallan, P. Francois, C. Poirion, A. Kilian, D. Burthia, C. Jenny, F. Bakry, S. Brown, V. Guignon, G. Kema, M. Dita, C. Waalwijk, S. Joseph, A. Dievert, O. Jaillon, J. Leclercq, X. Argout, E. Lyons, A. Almeida, M. Jeridi, J. Dolezel, N. Roux, A. M. Risterucci, J. Weissenbach, M. Ruiz, J. C. Glaszmann, F. Quetier, N. Yahiaoui, and P. Wincker. "The Banana (*Musa Acuminata*) Genome and the Evolution of Monocotyledonous Plants." *Nature* 488, no. 7410 (2012): 213-7.
165. Fernie, A. R., and A. Aharoni. "Pan-Genomic Illumination of Tomato Identifies Novel Gene-Trait Interactions." *Trends in Plant Science* 24, no. 10 (2019): 882-84.
166. Huff, M., A. M. Hulse-Kemp, B. E. Scheffler, R. C. Youngblood, S. A. Simpson, E. Babiker, and M. Staton. "Long-Read, Chromosome-Scale Assembly of *Vitis Rotundifolia* Cv. Carlos and Its Unique Resistance to *Xylella Fastidiosa* Subsp. *Fastidiosa*." *BMC Genomics* 24, no. 1 (2023): 409.
167. Oren, E., A. Dafna, G. Tzuri, I. Halperin, T. Isaacson, M. Elkabetz, A. Meir, U. Saar, S. Ohali, T. La, C. Romay, Y. Tadmor, A. A. Schaffer, E. S. Buckler, R. Cohen, J. Burger, and A. Gur. "Pan-Genome and Multi-Parental Framework for High-Resolution Trait Dissection in Melon (*Cucumis Melo*)." *Plant Journal* 112, no. 6 (2022): 1525-42.
168. Hasan, N., S. Choudhary, N. Naaz, N. Sharma, and R. A. Laskar. "Recent Advancements in Molecular Marker-Assisted Selection and Applications in Plant Breeding Programmes." *J Genet Eng Biotechnol* 19, no. 1 (2021): 128.
169. Garrido-Cardenas, J. A., C. Mesa-Valle, and F. Manzano-Agugliaro. "Trends in Plant Research Using Molecular Markers." *Planta* 247, no. 3 (2018): 543-57.
170. Moncada, P., and S. McCouch. "Simple Sequence Repeat Diversity in Diploid and Tetraploid *Coffea* Species." *Genome* 47, no. 3 (2004): 501-9.
171. McCouch, S. R., X. Chen, O. Panaud, S. Temnykh, Y. Xu, Y. G. Cho, N. Huang, T. Ishii, and M. Blair. "Microsatellite Marker Development, Mapping and Applications in Rice Genetics and Breeding." *Plant Molecular Biology* 35, no. 1-2 (1997): 89-99.
172. Tanksley, S. D., and S. R. McCouch. "Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild." *Science* 277, no. 5329 (1997): 1063-6.
173. Morales, K. Y., N. Singh, F. A. Perez, J. C. Ignacio, R. Thapa, J. D. Arbelaez, R. E. Tabien, A. Famoso, D. R. Wang, E. M. Septiningsih, Y. Shi, T. Kretzschmar, S. R. McCouch, and M. J. Thomson. "An Improved 7k Snp Array, the C7air, Provides a Wealth of Validated Snp Markers for Rice Breeding and Genetics Studies." *PLoS ONE* 15, no. 5 (2020): e0232479.

174. Miller, J. R., P. Zhou, J. Mudge, J. Gurtowski, H. Lee, T. Ramaraj, B. P. Walenz, J. Liu, R. M. Stupar, R. Denny, L. Song, N. Singh, L. G. Maron, S. R. McCouch, W. R. McCombie, M. C. Schatz, P. Tiffin, N. D. Young, and K. A. T. Silverstein. "Hybrid Assembly with Long and Short Reads Improves Discovery of Gene Family Expansions." *BMC Genomics* 18, no. 1 (2017): 541.
175. Cheng, C., Z. Fei, and P. Xiao. "Methods to Improve the Accuracy of Next-Generation Sequencing." *Front Bioeng Biotechnol* 11 (2023): 982111.
176. Myburg, A. A., D. Grattapaglia, G. A. Tuskan, U. Hellsten, R. D. Hayes, J. Grimwood, J. Jenkins, E. Lindquist, H. Tice, D. Bauer, D. M. Goodstein, I. Dubchak, A. Poliakov, E. Mizrahi, A. R. Kullán, S. G. Hussey, D. Pinard, K. van der Merwe, P. Singh, I. van Jaarsveld, O. B. Silva-Junior, R. C. Togawa, M. R. Pappas, D. A. Faria, C. P. Sansaloni, C. D. Petroli, X. Yang, P. Ranjan, T. J. Tschaplinski, C. Y. Ye, T. Li, L. Sterck, K. Vanneste, F. Murat, M. Soler, H. S. Clemente, N. Saidi, H. Cassan-Wang, C. Dunand, C. A. Hefer, E. Bornberg-Bauer, A. R. Kersting, K. Vining, V. Amarasinghe, M. Ranik, S. Naithani, J. Elser, A. E. Boyd, A. Liston, J. W. Spatafora, P. Dharmawardhana, R. Raja, C. Sullivan, E. Romanel, M. Alves-Ferreira, C. Kulheim, W. Foley, V. Carocha, J. Paiva, D. Kudrna, S. H. Brommonschenkel, G. Pasquali, M. Byrne, P. Rigault, J. Tibbits, A. Spokevicius, R. C. Jones, D. A. Steane, R. E. Vaillancourt, B. M. Potts, F. Joubert, K. Barry, G. J. Pappas, S. H. Strauss, P. Jaiswal, J. Grima-Pettenati, J. Salse, Y. Van de Peer, D. S. Rokhsar, and J. Schmutz. "The Genome of Eucalyptus Grandis." *Nature* 510, no. 7505 (2014): 356-62.
177. Shulaev, V., D. J. Sargent, R. N. Crowhurst, T. C. Mockler, O. Folkerts, A. L. Delcher, P. Jaiswal, K. Mockaitis, A. Liston, S. P. Mane, P. Burns, T. M. Davis, J. P. Slovin, N. Bassil, R. P. Hellens, C. Evans, T. Harkins, C. Kodira, B. Desany, O. R. Crasta, R. V. Jensen, A. C. Allan, T. P. Michael, J. C. Setubal, J. M. Celton, D. J. Rees, K. P. Williams, S. H. Holt, J. J. Ruiz Rojas, M. Chatterjee, B. Liu, H. Silva, L. Meisel, A. Adato, S. A. Filichkin, M. Troggio, R. Viola, T. L. Ashman, H. Wang, P. Dharmawardhana, J. Elser, R. Raja, H. D. Priest, D. W. Bryant, Jr., S. E. Fox, S. A. Givan, L. J. Wilhelm, S. Naithani, A. Christoffels, D. Y. Salama, J. Carter, E. Lopez Girona, A. Zdepski, W. Wang, R. A. Kerstetter, W. Schwab, S. S. Korban, J. Davik, A. Monfort, B. Denoyes-Rothan, P. Arus, R. Mittler, B. Flinn, A. Aharoni, J. L. Bennetzen, S. L. Salzberg, A. W. Dickerman, R. Velasco, M. Borodovsky, R. E. Veilleux, and K. M. Folta. "The Genome of Woodland Strawberry (*Fragaria Vesca*)." *Nature Genetics* 43, no. 2 (2011): 109-16.
178. Wu, S., H. Sun, L. Gao, S. Branham, C. McGregor, S. S. Renner, Y. Xu, C. Kousik, W. P. Wechter, A. Levi, and Z. Fei. "A Citrullus Genus Super-Pangenome Reveals Extensive Variations in Wild and Cultivated Watermelons and Sheds Light on Watermelon Evolution and Domestication." *Plant Biotechnology Journal* (2023).
179. Naithani, S., D.A. Dikeman, P. Garg, N. Al-Bader, and P. Jaiswal. "Beyond Gene Ontology (Go): Using Biocuration Approach to Improve the Gene Nomenclature and Functional Annotation of Rice S-Domain Kinase Subfamily." *PeerJ* 9 (2021): e11052.
180. Naithani, S., S. S. Komath, A. Nonomura, and G. Govindjee. "Plant Lectins and Their Many Roles: Carbohydrate-Binding and Beyond." *Journal of Plant Physiology* 266 (2021): 153531.
181. Monaco, M. K., T. Z. Sen, P. D. Dharmawardhana, L. Ren, M. Schaeffer, S. Naithani, V. Amarasinghe, J. Thomason, L. Harper, J. Gardiner, E. K. S. Cannon, C. J. Lawrence, D. Ware, and P. Jaiswal. "Maize Metabolic Network Construction and Transcriptome Analysis." *Plant Genome* 6, no. 1 (2013): 1-12.
182. Jaiswal, P., and B. Usadel. "Plant Pathway Databases." *Methods in Molecular Biology* 1374 (2016): 71-87.
183. Naithani, S., H. Nonogaki, and P. Jaiswal. "Exploring Crossroads between Seed Development and Stress-Response." In *Mechanism of Plant Hormone Signaling under Stress*, edited by Girdhar K. Pandey, 415-54. New Jersey, USA: John Wiley & Sons, Inc. , 2017.
184. Gene Ontology, Consortium, S. A. Aleksander, J. Balhoff, S. Carbon, J. M. Cherry, H. J. Drabkin, D. Ebert, M. Feuermann, P. Gaudet, N. L. Harris, D. P. Hill, R. Lee, H. Mi, S. Moxon, C. J. Mungall, A. Muruganugan, T. Mushayahama, P. W. Sternberg, P. D. Thomas, K. Van Auken, J. Ramsey, D. A. Siegle, R. L. Chisholm, P. Fey, M. C. Aspromonte, M. V. Nugnes, F. Quaglia, S. Tosatto, M. Giglio, S. Nadendla, G. Antonazzo, H. Attrill, G. Dos Santos, S. Marygold, V. Strelets, C. J. Tabone, J. Thurmond, P. Zhou, S. H. Ahmed, P. Asanithong, D. Luna Buitrago, M. N. Erdol, M. C. Gage, M. Ali Kadhun, K. Y. C. Li, M. Long, A. Michalak, A. Pesala, A. Pritazhara, S. C. C. Saverimuttu, R. Su, K. E. Thurlow, R. C. Lovering, C. Logie, S. Oliferenko, J. Blake, K. Christie, L. Corbani, M. E. Dolan, H. J. Drabkin, D. P. Hill, L. Ni, D. Sitnikov, C. Smith, A. Cuzick, J. Seager, L. Cooper, J. Elser, P. Jaiswal, P. Gupta, P. Jaiswal, S. Naithani, M. Lera-Ramirez, K. Rutherford, V. Wood, J. L. De Pons, M. R. Dwinell, G. T. Hayman, M. L. Kaldunski, A. E. Kwitek, S. J. F. Laulederkind, M. A. Tutaj, M. Vedi, S. J. Wang, P. D'Eustachio, L. Aimò, K. Axelsen, A. Bridge, N. Hyka-Nouspikel, A.

- Morgat, S. A. Aleksander, J. M. Cherry, S. R. Engel, K. Karra, S. R. Miyasato, R. S. Nash, M. S. Skrzypek, S. Weng, E. D. Wong, E. Bakker, T. Z. Berardini, L. Reiser, A. Auchincloss, K. Axelsen, G. Argoud-Puy, M. C. Blatter, E. Boutet, L. Breuza, A. Bridge, C. Casals-Casas, E. Coudert, A. Estreicher, M. Livia Famiglietti, M. Feuermann, A. Gos, N. Gruaz-Gumowski, C. Hulo, N. Hyka-Nouspikel, F. Jungo, P. Le Mercier, D. Lieberherr, P. Masson, A. Morgat, I. Pedruzzi, L. Pourcel, S. Poux, C. Rivoire, S. Sundaram, A. Bateman, E. Bowler-Barnett, A. Jee H. Bye, P. Denny, A. Ignatchenko, R. Ishtiaq, A. Lock, Y. Lussi, M. Magrane, M. J. Martin, S. Orchard, P. Raposo, E. Speretta, N. Tyagi, K. Warner, R. Zaru, A. D. Diehl, R. Lee, J. Chan, S. Diamantakis, D. Raciti, M. Zarowiecki, M. Fisher, C. James-Zorn, V. Ponferrada, A. Zorn, S. Ramachandran, L. Ruzicka, and M. Westerfield. "The Gene Ontology Knowledgebase in 2023." *Genetics* 224, no. 1 (2023).
185. Cooper, L., and P. Jaiswal. "The Plant Ontology: A Tool for Plant Genomics." *Methods in Molecular Biology* 1374 (2016): 89-114.
 186. Walls, R. L., L. Cooper, J. Elser, M. A. Gandolfo, C. J. Mungall, B. Smith, D. W. Stevenson, and P. Jaiswal. "The Plant Ontology Facilitates Comparisons of Plant Development Stages across Species." *Frontiers of Plant Science* 10 (2019): 631.
 187. Naithani, S., B. Mohanty, J. Elser, P. D'Eustachio, and P. Jaiswal. "Biocuration of a Transcription Factors Network Involved in Submergence Tolerance During Seed Germination and Coleoptile Elongation in Rice (*Oryza Sativa*)." *Plants (Basel)* 12, no. 11 (2023).
 188. Naithani, S., P. Dharmawardhana, and J. B. Nasrallah. "Scr." In *Handbook of Biologically Active Peptides*, edited by Abba J. Kastin, 58-66: Academic Press, 2013.
 189. Bolger, M., R. Schwacke, and B. Usadel. "Mapman Visualization of Rna-Seq Data Using Mercator4 Functional Annotations." *Methods in Molecular Biology* 2354 (2021): 195-212.
 190. Naithani, S., P. Gupta, J. Preece, P. Garg, V. Fraser, L. K. Padgitt-Cobb, M. Martin, K. Vining, and P. Jaiswal. "Involving Community in Genes and Pathway Curation." *Database (Oxford)* 2019 (2019).
 191. Gupta, P., M. Geniza, S. Naithani, J. L. Phillips, E. Haq, and P. Jaiswal. "Chia (*Salvia Hispanica*) Gene Expression Atlas Elucidates Dynamic Spatio-Temporal Changes Associated with Plant Growth and Development." *Frontiers of Plant Science* 12 (2021): 667678.
 192. Hendre, P. S., S. Muthemba, R. Kariba, A. Muchugi, Y. Fu, Y. Chang, B. Song, H. Liu, M. Liu, X. Liao, S. K. Sahu, S. Wang, L. Li, H. Lu, S. Peng, S. Cheng, X. Xu, H. Yang, J. Wang, X. Liu, A. Simons, H. Y. Shapiro, R. H. Mumm, A. Van Deynze, and R. Jamnadass. "African Orphan Crops Consortium (Aocc): Status of Developing Genomic Resources for African Orphan Crops." *Planta* 250, no. 3 (2019): 989-1003.
 193. Chang, Y., H. Liu, M. Liu, X. Liao, S. K. Sahu, Y. Fu, B. Song, S. Cheng, R. Kariba, S. Muthemba, P. S. Hendre, S. Mayes, W. K. Ho, A. E. J. Yssel, P. Kendabie, S. Wang, L. Li, A. Muchugi, R. Jamnadass, H. Lu, S. Peng, A. Van Deynze, A. Simons, H. Yana-Shapiro, Y. Van de Peer, X. Xu, H. Yang, J. Wang, and X. Liu. "The Draft Genomes of Five Agriculturally Important African Orphan Crops." *Gigascience* 8, no. 3 (2019).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.