Review

Exploring Pan-Genomes: An Overview of Resources and Tools for Unraveling Structure, Function, and Evolution of Crop Genes and Genomes

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 97331 , USA; pankaj.jaiswal@oregonstate.edu
[2] Molecular & Digital Breeing Group, New Cultivar Innovation, The New Zealand Institute for Plant and Food Research Limited, Private Bag 92169, Auckland 1142, New Zealand; cecilia.deng@plantandfood.co.nz
[3] State Key Laboratory of Agricultural Genomics, Key Laboratory of Genomics, Ministry of Agriculture, BGI Research, Shenzhen 518083, China; sunilkumarsahu@genomics.cn
* Correspondence: sushma.naithani@oregonstate.edu
† These authors contributed equally to this work.

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 pan-genomes 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 to mitigate the impact of global climate change. Here, we summarize the tools used for pan-genome assembly and annotations, web-portals hosting plant pan-genomes, etc. Furthermore, we highlight a few discoveries made in crops using the pan-genomic approach and future potential of this emerging field of study.

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 sequencing technology (i.e., short-read 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 pan-genomes 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 among genotypes of the same species include insertion deletions (indels) and translocation of the small or large genomic regions that further cause presence–absence variations, copy number variations, chromosomal rearrangements, and variations in repeat 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 the "dispensable" or "accessory" genome [2,24]. As shown in Figure 1, 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 the same species).
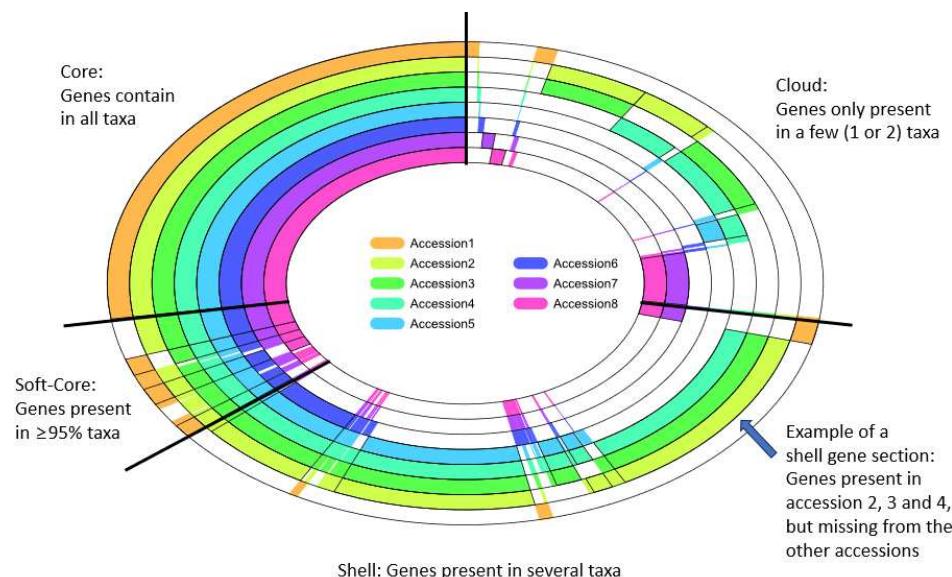
Figure 1. A conceptual depiction of the pan-genes across genomes of eight accessions belonging to the same clade. Each ring represents one accession. The top-left "core genes" represent conserved genes across eight accessions. The white section in a ring indicates the absence of ortholog(s). The "soft cores" represent genes found in ≥95% of accessions. The "cloud genes" are 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 pan-genomes encompass genes found in crop wild relatives and ancestral species [27–31]. We would like to note here that many useful genes lost during crop domestication and extensive plant breeding [32–34] may be found in the "dispensable"/accessory genome of any crop species [15,33,35–39]. Thus, the availability of plant pan-genomes allows researchers and breeders to explore important candidate genes for improving crop yield, nutritional quality, and adaptability to changing climatic conditions and diseases. 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, and yield [40–49]. In addition, the high-quality pan-genomes also make it possible to study previously inaccessible regions of the eukaryotic genomes, including centromeres, long heterochromatic blocks, 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 [1,2,51–54], which focus on pan-genome construction, structure variation detection, challenges associated with polyploid crops, and the application of pan-genomes in crop research. It is important to highlight that none of the published reviews on pan-genomes provide a comprehensive collection of pan-genome tools or resources accessible in the public domain. Thus, our work fills this gap by concentrating on the current landscape of available pan-genome tools and resources tailored to the needs of crop researchers. Here, we review

the current tools used for constructing and visualizing crop pan-genome data, public genomic portals/resources hosting pan-genes, pan-genome data, and pan-genome browsers. Furthermore, we highlight a few studies that have exploited a pan-genomic approach for discovering candidate genes associated with important agronomic traits. We also discuss the potential of pan-genome-driven translational research.

2. Pan-Genome Construction, Visualization, and Data Analysis Tools

The first step in setting up a pan-genome 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 the sequencing of individual genomes. The high-quality reference genomes are of critical importance for building pan-genome assemblies and a complete pan-gene atlas. Therefore, we see overlap and inter-connection between the genomics and pan-genomics. We envision pan-genomics as a natural extension and outgrowth of genomics, not a different field of study. The third step is the assembly and construction of the pan-genome. Previously a few reviews [1,57–59] have been published on several approaches implemented for pan-genome assembly. Here, we briefly describe the basic tenets of three popular methods.

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 (Figure 2A). 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 without access to a reference genome, de novo assembly of individual genotypes is generated, followed by mapping assembled contigs 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 genomes and their comparison by pair-wise sequence alignment is arguably the most powerful and accurate approach to detect sequence variants from base-level resolution to novel genomic elements and rearrangements (Figure 2B). However, generating assemblies of polyploid plant genomes is still challenging, as current methods are limited in detecting and phasing heterozygous structural variants that 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 adding 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 pan-genome assembly (Figure 2C). However, graph construction and traversal can be computationally intensive, especially for large and diverse pan-genomes, and require 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 the identification of shared and unique genomic regions among individuals or strains and aiding in excellent visualization of pan-genomes. A conceptual visualization of the graph-based pan-genome is shown in Figure 3.
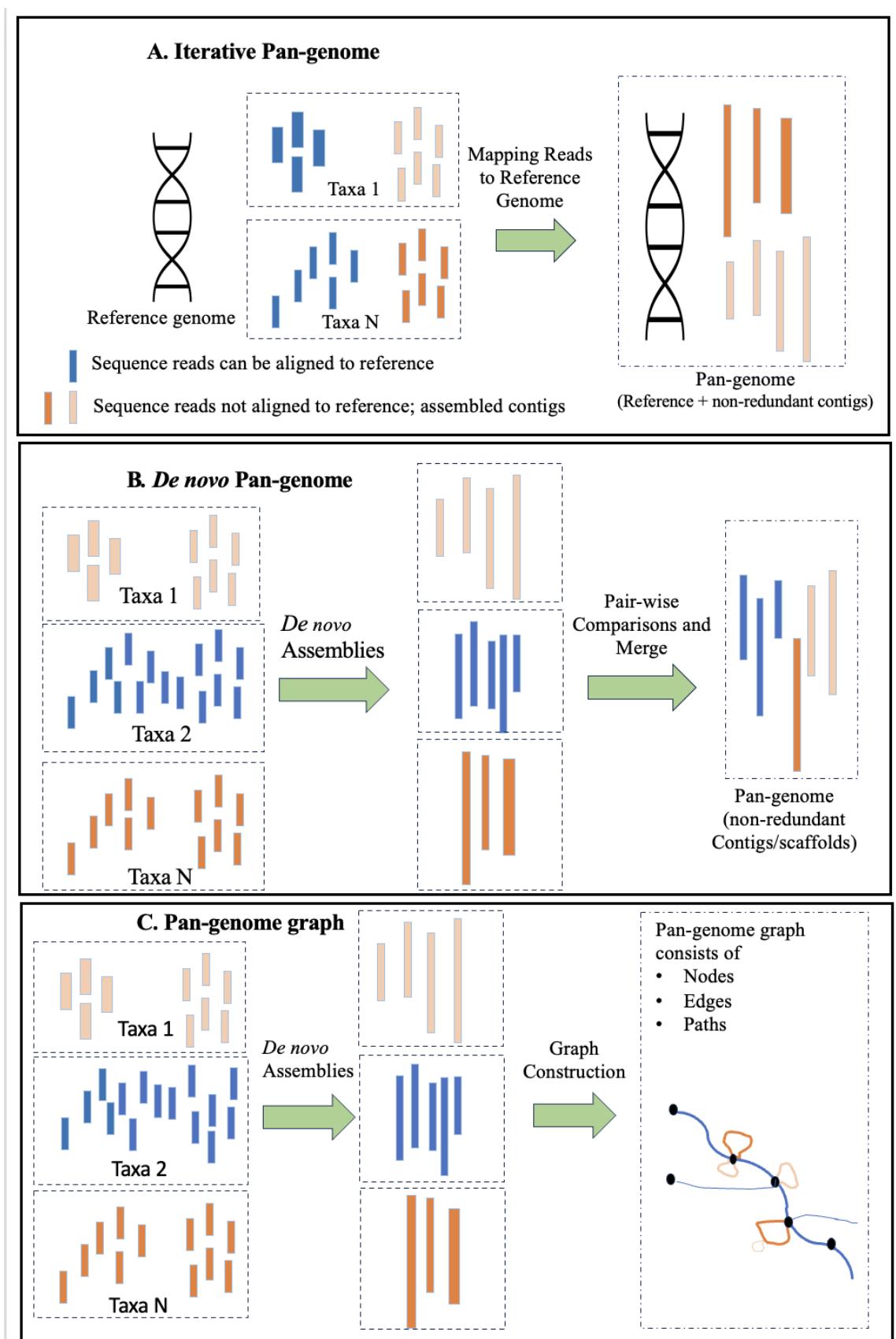
Figure 2. An illustration of three popular approaches currently used for pan-genome construction, including (**A**) reference-based iterative method, (**B**) de novo genome assembly, and (**C**) graph-based pan-genome assembly.
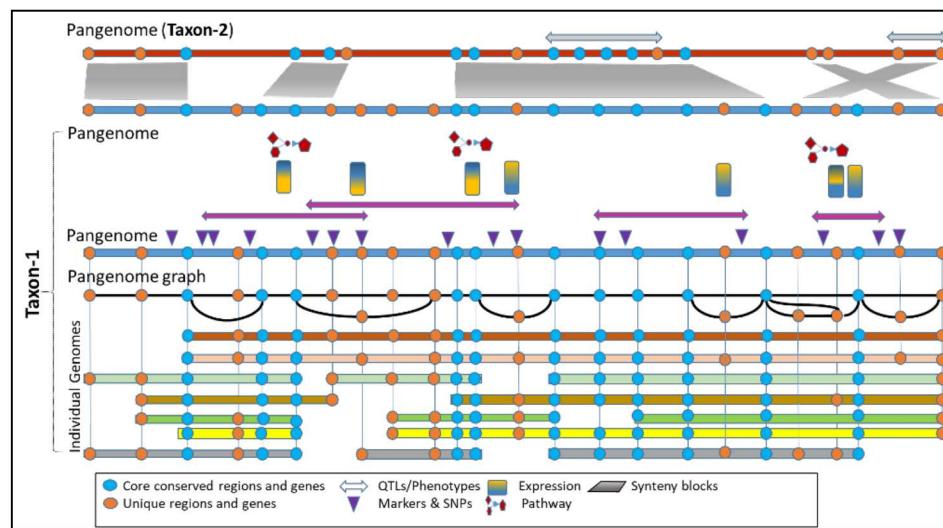
Figure 3. A conceptual view of a pan-genome reference graph carrying chromosomal rearrangements and mapped features. The graph allows views and analysis of whole-genome alignments, pangene sets, gene orthology, expression, pathways, function, and aligned synteny to help accelerate knowledge discovery and hypothesis-driven research.

Recently, Shang et al., 2022 [21] have constructed a '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 *Oryza rufipogon* (the wild ancestor of *O. sativa*), 11 accessions of domesticated African rice *Oryza glaberrima*, and 10 accessions of *Oryza barthii* (the wild ancestor of *O. glaberrima*). They used the de novo long-read assembly and a graph-based approach. The Rice Super Pan-genome Information Resource Database (RiceSuperPIRdb) provides access to a reference-free whole-genome multiple sequence alignment for these 251 rice accessions. This resource hosts fully annotated pan-genome graph visualization using the JBrowse genome browser. It facilitates the integration of structural variations, gene annotations, transposable element annotations, pan-genome graphs, and BLAST tools [21].

A few excellent reviews have previously described the development of computational tools for pan-genome visualizations [57,58,62,63]. We note here that genome sequencing technologies and assembly algorithms are rapidly evolving to achieve high accuracy 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 pan-genome 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. All URLs were checked and confirmed to be valid on 13 September 2023.

Tool Name and URL	Remarks and Citation
Genome assembly	
Hifiasm https://github.com/chhylp123/hifiasm (accessed on 13 September 2023).	Constructs haplotype-resolved assemblies from accurate HiFi Reads [60].
Canu https://github.com/marbl/canu (accessed on 13 September 2023).	Assembles genomes of any size from single molecule sequences and provides graphical fragment assembly that can be integrated with complementary phasing and scaffolding methods [66].

Table 1. *Cont.*

Tool Name and URL	Remarks and Citation
Flye https://github.com/fenderglass/Flye (accessed on 13 September 2023).	Assembles single molecule, long-read sequencing data into genomes using repeat graphs [67].
PAGIT https://www.sanger.ac.uk/tool/pagit (accessed on 13 September 2023).	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 (accessed on 13 September 2023).	Ultra-fast NGS assembler for metagenomes [69].
SPADes https://cab.spbu.ru/software/spades/ (accessed on 13 September 2023).	A set of genome assembly and analysis tools that can use long- and short-read sequence data [70,71].
Pan-genome 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 (accessed on 13 September 2023).	Toolset for eukaryotic pan-genome graph construction, read mapping, variant calling, and graph visualization [58].
Minigraph https://github.com/lh3/minigraph (accessed on 13 September 2023).	Tool for graph construction, mapping, and variant calling [72].
ODGI https://github.com/pangenome/odgi (accessed on 13 September 2023).	Optimized Dynamic Genome/Graph Implementation (ODGI) is a tool suite representing graphs, including structurally complex regions, with minimal memory overhead [73]. It is a pan-genome toolbox with more than 30 tools to transform, analyze, simplify, validate, annotate, and visualize pan-genome graphs.
PGGB https://github.com/pangenome/pggb (accessed on 13 September 2023).	Uses ODGI as the backbone for pan-genome graph construction, normalization, and visualization [74].
MGRgraph https://github.com/LeilyR/Multi-genome-Reference (accessed on 13 September 2023).	An algorithm for building a multi-genome graph.
Cactus https://github.com/ComparativeGenomicsToolkit/cactus (accessed on 13 September 2023).	A reference-free multiple genome alignment program that can use progressive mode to build pan-genome across different species [75,76]
PanTools https://pantools.readthedocs.io/en/latest/user_guide/install.html (accessed on 13 September 2023).	A platform for pan-genome graph construction, read mapping, phylogeny analysis, pan-graph query, and pan-gene annotation [77].
Smoothxg https://github.com/pangenome/smoothxg (accessed on 13 September 2023).	A tool for local reconstruction of variation graphs.
nf-core/pangenome https://github.com/nf-core/pangenome (accessed on 13 September 2023).	Nextflow pipeline for all-vs-all alignment, pan-genome graph construction, normalization, remove redundancy, and visualization (through ODGI) [78].
SeqWish https://github.com/ekg/seqwish (accessed on 13 September 2023).	Builds a variation graph from pairwise alignments [73].
PanPipe https://github.com/USDA-ARS-GBRU/PanPipes (accessed on 13 September 2023).	An end-to-end pan-genome graph construction and genetic analysis pipeline [79].

Table 1. *Cont.*

Tool Name and URL	Remarks and Citation
PanGene https://github.com/lh3/pangene (accessed on 13 September 2023).	Used for ortholog and paralog analysis and for building pan-gene graphs.
Minimap2 https://github.com/lh3/minimap2 (accessed on 13 September 2023).	A fast DNA or long mRNA sequence aligner to a reference genome [80].
NGMLR https://github.com/philres/ngmlr (accessed on 13 September 2023).	This program aligns PacBio long reads to genomes for detecting complex structural variations [5].
MUMmer4 https://mummer.sourceforge.net/ (accessed on 13 September 2023). https://github.com/mummer4/mummer (accessed on 13 September 2023).	A genome-to-genome aligner tool [81].
GraphAligner https://github.com/maickrau/GraphAligner (accessed on 13 September 2023).	A tool for aligning long reads to genome graphs [82].
V-ALIGN https://github.com/tcsatc/V-ALIGN (accessed on 13 September 2023).	V-ALIGN allows gapped sequence alignment directly on the input graph and supports affine and linear gaps [83].
PaSGAL https://github.com/ParBLiSS/PaSGAL (accessed on 13 September 2023).	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 (accessed on 13 September 2023).	A tool for mapping short-read sequence data to the pan-genome graph [84].
DeepVariant https://github.com/google/deepvariant (accessed on 13 September 2023).	A deep learning-based variant caller that uses sequence read alignments in BAM and CRAM format to produce image tensors and convolutional neural networks to identify universal SNP and small-indel variants [85,86].
SpeedSeq https://github.com/hall-lab/speedseq (accessed on 13 September 2023).	A platform for alignment, variant calling, and functional annotation [87].
graphTyper https://github.com/DecodeGenetics/graphtyper (accessed on 13 September 2023).	This graph-based variant caller realigns short-read sequence data to a pan-genome for discovering sequence variants [88].
PanGenie https://github.com/eblerjana/pangenie (accessed on 13 September 2023).	An alignment-free Kmer-based genotyper for structural variation detection on pan-genome graphs. It uses short-read sequencing data to genotype a broad spectrum of genetic variation [89].
VEP https://ensembl.gramene.org/tools.html (accessed on 13 September 2023).	The Variant Effect Prediction (VEP) tool helps in analyzing the consequences of sequence variations on transcript structure and gene function [90].
OrthoFinder https://github.com/davidemms/OrthoFinder (accessed on 13 September 2023).	This method is used for finding orthologs in proteomes [91].
OrthoMCL https://orthomcl.org/orthomcl/app (accessed on 13 September 2023).	A scalable method for constructing orthology groups from eukaryotic proteomes [92].
JustOrthologs https://github.com/ridgelab/JustOrthologs/ (accessed on 13 September 2023).	JustOrthologs is a fast ortholog identification algorithm that uses the conservation of gene structure [93].

Table 1. *Cont.*

Tool Name and URL	Remarks and Citation
PhyloMCL https://sourceforge.net/projects/phylomcl/files/Materials/ (accessed on 13 September 2023).	PhyloMCL provides accurate clustering of hierarchical orthogroups guided by phylogenetic relationships and inference of polyploidy events [94].
OMA https://github.com/DessimozLab/OmaStandalone/tree/v2.4.0 (accessed on 13 September 2023). https://omabrowser.org/oma/home/ (accessed on 13 September 2023).	Orthologous Matrix (OMA) is a method for ortholog identification from genomes [95].
InParanoid-Diamond https://bitbucket.org/sonnhammergroup/inparanoid/src (accessed on 13 September 2023).	The tool is used for the identification of gene-orthologs and gene family clustering [96]. This is used for orthology projection in the Plant Reactome (https://plantreactome.gramene.org) [97].
Panache https://github.com/SouthGreenPlatform/panache (accessed on 13 September 2023).	A web-based tool for viewing linearized pan-genomes [98]. For example, the banana genome hub [99].
MoMI-G https://github.com/MoMI-G/MoMI-G/ (accessed on 13 September 2023).	Genome graph browser for viewing structural variations. Users can filter and visualize annotations and inspect read alignments over the genome graph [100].
panGraphViewer https://github.com/TF-Chan-Lab/panGraphViewer (accessed on 13 September 2023).	panGraphViewer, based on Python3, is used for pan-genome graph visualization and runs on all major operating systems.
Bandage https://rrwick.github.io/Bandage/ (accessed on 13 September 2023).	An interactive tool for visualizing de novo assembled genomes [101].
Bandage-NG https://github.com/asl/BandageNG (accessed on 13 September 2023).	GUI program to interact with assembly graphs based on the Open Graph Drawing Framework (OGDF) and Open Graph Algorithms and Data Structures Framework).
sequenceTubeMaps https://github.com/vgteam/sequenceTubeMap (accessed on 13 September 2023).	Interactive visualization of genomes [102].
GfaViz https://github.com/ggonnella/gfaviz (accessed on 13 September 2023).	Interactive visualization of Graphical Fragment Assembly (GFA) genome graphs [103].
AGB https://github.com/almiheenko/AGB (accessed on 13 September 2023).	Assembly Graph Browser (AGB) is used for constructing and visualizing large assembly graphs and repeat sequence analysis [104].
IGGE https://github.com/immersivegraphgenomeexplorer/IGGE (accessed on 13 September 2023).	An interactive graph genomes browser.
GFAViewer https://lh3.github.io/gfatools/ (accessed on 13 September 2023).	Used for online visualization of GFA files.
SGTK https://github.com/olga24912/SGTK (accessed on 13 September 2023).	The scaffold graph toolkit is used for the construction and interactive visualization of scaffold graphs using sequencing data [105].
Maffer https://github.com/pangenome/maffer (accessed on 13 September 2023).	It converts sorted graphs to multiple alignment format (MAF).
Gfatools https://github.com/lh3/gfatools (accessed on 13 September 2023).	A set of tools to parse, subgraph, and convert GFA or rGFA format to FASTA/BED format.

Table 1. *Cont.*

Tool Name and URL	Remarks and Citation
Pgge https://github.com/pangenome/pgge (accessed on 13 September 2023).	It is a pan-genome graph evaluator
WGT https://github.com/Kuanhao-Chao/Wheeler_ Graph_Toolkit (accessed on 13 September 2023).	This package contains tools and algorithms for recognizing, visualizing, and generating Wheeler graphs.
GBWT https://github.com/jltsiren/gbwt (accessed on 13 September 2023).	A tool used for haplotype matching and storage using the positional Burrows-Wheeler Transform (PBWT) approach [106,107].
Spodgi https://github.com/pangenome/spodgi (accessed on 13 September 2023).	Convert ODGI genome graph file to SPARQL database.
GraphPeakCaller https://github.com/uio-bmi/graph_peak_caller (accessed on 13 September 2023).	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 (accessed on 13 September 2023).	It is a pan-genome analysis pipeline (PSVCP) to construct a pan-genome, call structural variants, and run population genotyping. It was used for rice pan-genome [109].
ppsPCP http://cbi.hzau.edu.cn/ppsPCP/ (accessed on 13 September 2023).	It is designed specifically for constructing fully annotated plant pan-genomes. It scans presence/absence variants [110].

3. A Survey of Crop Pan-Genome Portals and Data Resources

With technical advances and the affordability of the sequencing and assembly of genomes, we are experiencing a deluge of big data in biology. The plant research community now faces a bigger challenge of making 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 analyses and 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 providing 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]). In particular, the pathway databases and species-specific metabolic networks curate data at the species level and, thus, can easily accommodate the knowledge gained from the genome analysis of multiple accessions belonging to the same species. If some of the critical genetic hotspots or genomic loci associated with metabolism or production of specific metabolites production are absent in the reference genome (or not annotated correctly), the pan-genome data can help pathway database biocurators to incorporate data from multiple representative genotypes and build the accurate representation of metabolome present in a species or clade. The availability of pan-genomes would reduce the occurrence of false negatives. However, the availability of plant pan-genome 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-genome portals and data resources for crops. All URLs were checked and confirmed to be valid on 13 September 2023.

Pan-Genome Resource	Remarks
Gramene Link: https://www.gramene.org/pansites (accessed on 13 September 2023). Species: maize, rice, grapevine, and sorghum.	Gramene hosts 128 reference plant genomes [115] and pan-genome sites for maize, rice, grapevine, and sorghum.
SorghumBase Link: https://www.sorghumbase.org (accessed on 13 September 2023). Species: sorghum.	SorghumBase portal hosts a sorghum pan-genome browser comprising 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 (accessed on 13 September 2023). In addition to RPAN, the data and analyzed outputs from 3K RGP are available at the following websites: http://snp-seek.irri.org/ (accessed on 13 September 2023). http://www.rmbreeding.cn/index.php (accessed on 13 September 2023). http://www.ricecloud.org (accessed on 13 September 2023). https://aws.amazon.com/public-data-sets/3000-rice-genome (accessed on 13 September 2023). Species: rice (*O. sativa*) and its wild relatives.	The Rice Pan-genome Browser (RPAN) hosts genomic variation data from 3010 diverse rice accessions [8,9,133,134]. It contains ~370 Mbp IRGSP genome and ~260 Mbp novel sequences comprising 50,995 genes (23,914 core genes). RPAN provides a phylogenetic tree browser to view the phylogeny of rice accessions and a genome browser to view gene annotation and presence-absence variations. Users can access pan-gene views and associated genetic variations.
RiceSuperPIRdb Link: http://www.ricesuperpir.com (accessed on 13 September 2023). Species: 251 genomes representing domesticated rice accessions and wild relatives (202 *O. sativa*, 28 *O. rufipogan*, 11 *O. glaberrima*, and 10 *O. barthii* accessions).	The RiceSuperPIRdb hosts a genome browser for the rice super pan-genome built using reference-free, high-quality whole genome alignment of 251 independent genome assemblies. Genome annotations and node-specific K-mer spectrum pan-genome graphs are available for each assembly. In addition, genetic variation graphs support linking query data and the identification of lineage-specific haplotypes for trait-associated genes [21].
PanOryza Link: https://panoryza.org (accessed on 13 September 2023). Species: magic-16 rice accessions; see https://panoryza.org (accessed on 13 September 2023).	PanOryza provides consistency in the rice gene annotation across all rice varieties and the rice pan-genome browser supported by the JBrowse genome browser.
MaizeGDB Link: https://nam-genomes.org (accessed on 13 September 2023). Species: maize.	MaizeGDB hosts 48 maize genomes, including 26 high-quality PacBio genome assemblies of the Nested Associated Mapping (NAM) population founder lines. It allows users to connect genomes, gene models, expression, methylome, sequence variations, structural variations, transposable elements, etc., across the maize pan-genome supported by the Jbrowse browser [125].
ZEAMAP Link: www.zeamap.com (accessed on 13 September 2023). Species: maize.	The ZEAMAP database incorporates multiple annotated reference genomes, data from transcriptomes, open chromatin regions, chromatin interactions, high-quality genetic variants, phenotypes, metabolomics, genetic maps, population structures, and populational DNA methylation signals from maize inbred lines [135].
GreenPhylDB Link: https://www.greenphyl.org/cgi-bin/index.cgi (accessed on 13 September 2023). Species: 46 plant species and 19 pan-genomes, 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.

Table 2. *Cont.*

Pan-Genome Resource	Remarks
The Wheat Panache Web Portal Link: http://www.appliedbioinformatics.com.au/wheat_panache (accessed on 13 September 2023). Species: wheat.	This wheat pan-genome graph visualization is supported by the Panache tool. It allows users to explore structural variations across the selected wheat accessions [137].
GrainGenes Link: https://wheat.pw.usda.gov/GG3/pangenome (accessed on 13 September 2023). Species: wheat, barley, rye, oat.	GrainGenes hosts molecular and phenotype data for wheat, barley, rye, oat, etc., including several genome assemblies, genome browsers, and a *T. aestivum* (bread wheat) pan-genome [138].
Wheat Pan-genome Link: http://appliedbioinformatics.com.au/cgi-bin/gb2/gbrowse/WheatPan/ (accessed on 13 September 2023). Species: bread wheat (*Triticum aestivum*).	The wheat Pan-genome facilitates comparison of an improved reference for the Chinese Spring wheat genome with 18 wheat cultivars [139].
SGN Links: https://solgenomics.net (accessed on 13 September 2023). Subsites: http://solomics.agis.org.cn/tomato/tool/jbrowse_nav (accessed on 13 September 2023). https://solgenomics.net/projects/tgg (accessed on 13 September 2023). https://solgenomics.net/organism/Solanum_melongena/genome (accessed on 13 September 2023). Species: tomato, potato, petunia, and eggplant.	The Solanaceae Genomics Network (SGN) database hosts pan-genome data for tomato and eggplant. International Tomato Genome Sequencing Project produced the tomato pan-genome data consisting of genome assemblies from 46 accessions (22 *Solanum lycopersicum*, 13 *Solanum lycopersicum* var. cerasiforme; and 11 *Solanum pimpinellifolium*) [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/ (accessed on 13 September 2023). Species: *Capsicum annuum* (pepper) and its wild relatives.	The PepperPan was constructed by mapping the sequences of 383 pepper cultivars to the Zunla-1 genome as the reference [142]. The novel contig sequences (accession number GWHAAAT00000000) are available at http://bigd.big.ac.cn/gwh.
BRIDGEcereal Link: https://bridgecereal.scinet.usda.gov (accessed on 13 September 2023). Species: wheat, maize, barley, sorghum, and rice.	The Blastn Recovered Insertion and Deletion near Gene Explorer (BRIDGEcereal) web application supports mining publicly accessible pan-genomes of five major cereal crops, including wheat, maize, barley, sorghum, and rice [143]. It facilitates the identification of potential indels (insertion or deletions) for genes of interest.
PanSoy Link: https://www.soybase.org/projects/SoyBase.C2021.01.php (accessed on 13 September 2023). Species: *Glycine soja* (wild soybean) and *Glycine max* (soybean).	PanSoy is a soybean pan-genome assembly consisting of the genome sequence data from 204 phylogenetically and geographically distinct soybean accessions (GmHapMap collection). It was built using the de novo genome assembly method [144].
Sunflower Genome Database Link: https://www.sunflowergenome.org (accessed on 13 September 2023). Species: *Helianthus annuus* (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 SNP data is available at the Sunflower Genome Database [145].
COTTONOMICS Link: http://cotton.zju.edu.cn (accessed on 13 September 2023). Species: cotton.	It provides genome-wide, gene-scale structural variations detected from 11 assembled allopolyploid cotton genomes and is linked to important agronomic traits [146].
BGH Link: https://banana-genome-hub.southgreen.fr (accessed on 13 September 2023). Species: *Musa Ensete,* 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].

Table 2. *Cont.*

Pan-Genome Resource	Remarks
CPBD Links: http://citrus.hzau.edu.cn/ (accessed on 13 September 2023). Species: sweet orange (*Citrus sinensis*), mandarin (*Citrus reticulata*), pummelo (*Citrus grandis*), grapefruit (*Citrus paradisi*), and lemon (*Citrus limon*).	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 (accessed on 13 September 2023). Species: citrus accessions.	The Citrus Genome Database (CitGVD) hosts 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 [148].
Apple pan-genome Link: http://bioinfo.bti.cornell.edu/apple_genome (accessed on 13 September 2023). Species: apple (*Malus domestica*) and its wild progenitors *M. sieversii* and *M. sylvestris*.	Apple pan-genome was constructed using phased diploid genome assemblies of *Malus domestica* cv. Gala, *M. sieversii*, and *M. sylvestris*, and 91 sequenced genomes of additional accessions [149].
BnPIR Link: http://cbi.hzau.edu.cn/bnapus (accessed on 13 September 2023). Species: *Brassica oleracea*, *Brassica macrocarpa* (cultivated and wild cabbage), *Brassica napus*.	The Brassica napus pan-genome information resource (BnPIR) hosts eight high-quality *B. napus* reference genomes generated using PacBio sequencing and re-sequencing data from 1688 rapeseed accessions. It provides a pan-gene module, pan-genome Browser, and synteny data. It also hosts multi-omics data and common bioinformatics tools [150].
Cassava pan-genome Link: https://cassavabase.org/ (accessed on 13 September 2023). Species: cassava (*Manihot esculenta*).	Two high-quality, chromosome-scale haploid genome assemblies for African cassava cultivar TME204 (resistant to cassava mosaic diseases caused by African cassava mosaic viruses) were generated using a combination of short-read and long-read sequencing methods (Illumina PE reads, PacBio CLRs, 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/home (accessed on 13 September 2023). 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/D2/RIO2QM (accessed on 13 September 2023). 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]. It has 209935 assembled contig sequences from 176 sorghum accessions. This data represent 35,719 genes (including 34,211 genes from reference).
Barley pan-genome Link: https://bitbucket.org/ipk_dg_public/barley_pangenome/src/master/ (accessed on 13 September 2023). https://galaxy-web.ipk-gatersleben.de/libraries (accessed on 13 September 2023). Species: barley cultivars and a wild relative.	This first-generation barley pan-genome consists of chromosome-scale sequence assemblies for the 20 barley varieties (including landraces, cultivars, and wild barley from global barley diversity collection) and whole-genome shotgun sequencing data from additional 300 barley accessions [11].
Soybean pan-genome The genetic diversity data is available at https://figshare.com/s/689ae685ad2c368f2568 (accessed on 13 September 2023). SNPs and small indels data from the 2,898 accessions are available at (http://bigd.big.ac.cn/gvm/getProjectDetail?project=GVM000063 (accessed on 13 September 2023). Species: *Glycine soja* (wild soybean), *Glycine max* (soybean).	This graph-based pan-genome assembly was generated using de novo genome assemblies of 26 representative soybean accessions [14]. The sequencing data, assembled chromosomes, unplaced scaffolds, and annotations from this project 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.

Table 2. *Cont.*

Pan-Genome Resource	Remarks
Chickpea pan-genome Links: Pan-genome assembly and annotations: https://doi.org/10.6084/m9.figshare.16592819 (accessed on 13 September 2023). The variant calls: https://cegresources.icrisat.org/cicerseq (accessed on 13 September 2023).	The chickpea pan-genome consists of genome sequence data from 3366 chickpea lines (including 3171 cultivated and 195 wild accessions) [153]. Additional data, including Manhattan and QQ-plots for Genome-Wide Association Study (GWAS) analysis, is available at https://doi.org/10.6084/m9.figshare.15015309.
Pigeon pea Link: https://research-repository.uwa.edu.au/en/datasets/pigeon-pea-pangenome-contig-assembly-annotation-snps-pav (accessed on 13 September 2023). Species: pigeon pea (*Cajanus cajan*).	The pigeon pea pan-genome consists of genome sequence data from 89 pigeon pea accessions, including 70 from South Asia, 8 from sub-Saharan Africa, 7 from South East Asia, 2 from Mesoamerica, and 1 from Europe. This pan-genome was generated using the reference genome assembly (C. cajan_V1.0) and iterative mapping and assembly method [154].
Sesame pan-genome Species: sesame (*Sesamum indicum* L.).	The sesame pan-genome was constructed by mapping genome sequence data from two landraces, *S. indicum* cv. Baizhima and Mishuozhima and two cultivars, Yuzhi11 and Swetha, to the *S. indicum* var. Zhongzhi13 reference genome [155].
Cotton Variome Links: Genetic variation is available at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA576032 and https://figshare.com/s/cb3c104782a1dcd90ab0 (accessed on 13 September 2023). Species: *Gossypium hirsutum* and *Gossypium barbadense.*	Cotton Variome provides genetic variation data from 1961 cotton accessions [156].
Melon pan-genome Link: https://figshare.com/articles/dataset/melon_pangenome/17195072 (accessed on 13 September 2023).	Pan-genome of *Cucumismelo* 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 PRJNA657438.	The cucumber pan-genome graph was constructed using genome sequence data from 11 representative accessions from the 115-line core collection. The genome assemblies were generated using long-read and short-read sequence 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 (accessed on 13 September 2023). Species: cultivated and wild strawberry.	This strawberry pan-genome was generated using chromosome-scale reference genome assemblies of five diploid strawberry species (*Fragaria mandschurica*, *Fragaria daltoniana*, *Fragaria pentaphylla*, *F. nilgerrensis*, and *F. viridis*) and genome resequencing data of 128 accessions [159].
Walnut pan-genome Link:https://db.cngb.org/search/project/CNP0001209 (accessed on 13 September 2023). Species: walnut (*Juglans nigra*).	A high-quality reference genome assembly of black walnut (*Juglans nigra*) genotype NWAFU168 was constructed using short-read and long-read sequence data (Illumina, Pacbio, and Hi-C). A Walnut pan-genome was built using this reference genome and mapping sequence data from 74 walnut accessions [56].
SalviaGDB Link: https://salviagdb.org/ (accessed on 13 September 2023). Species: *Salvia hispanica* (Chia), *S. miltiorrhiza* (Danshen), *S. bowleyana* (nan Denshen), *S. splendens* (sage), and *S. rosmarinus* (rosemary).	The high-quality genome assembly and annotations of orphan crop *Salvia hispanica* (Chia) (4 genomes), and one each for the herbs used in culinary and traditional medicine.

It is clear from Table 2 that 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 analysis of the user's data. However, the few platforms and genomic databases that support pan-gene analysis in a phylogenomic context and support the user's query show the potential of pan-genome data for supporting basic research as well as translational applications for crop improvement. Here, we highlight an

example of visualizing the pan-gene data for the *TILLER ANGLE CONTROL 1* (*OsTAC1*) transcription factor coding gene from various accessions of cultivated rice *O. sativa* and other members of *Oryza* genus at Gramene (https://oryza.gramene.org; accessed on July 20, 2023). *OsTAC1* is induced by gravity stimulation and promotes horizontal shoot growth by negatively regulating shoot gravitropism [160]. Thus, it is involved in regulating tiller angle and modulating plant architectural traits of agronomic 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 (see Figure 4A). Indeed, a previously published study has shown that a point mutation in the *OsTAC1* gene at the 3′-splicing site of the 1.5-kb intron ('GGGA') in japonica rice accessions caused a 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 with large tiller angles contain 'AGGA' sequences at the 3′-splicing site of the 1.5-kb intron [161]. The *OsTAC1* gene is on chromosome 9 in the rice genome and shows high conservation across all rice accessions (Figure 4B). However, we also see that the *OsTAC1* gene neighborhood on the left-hand side is not much conserved, and it remains to be explored for candidate genes involved in the functional adaptation of rice accessions.
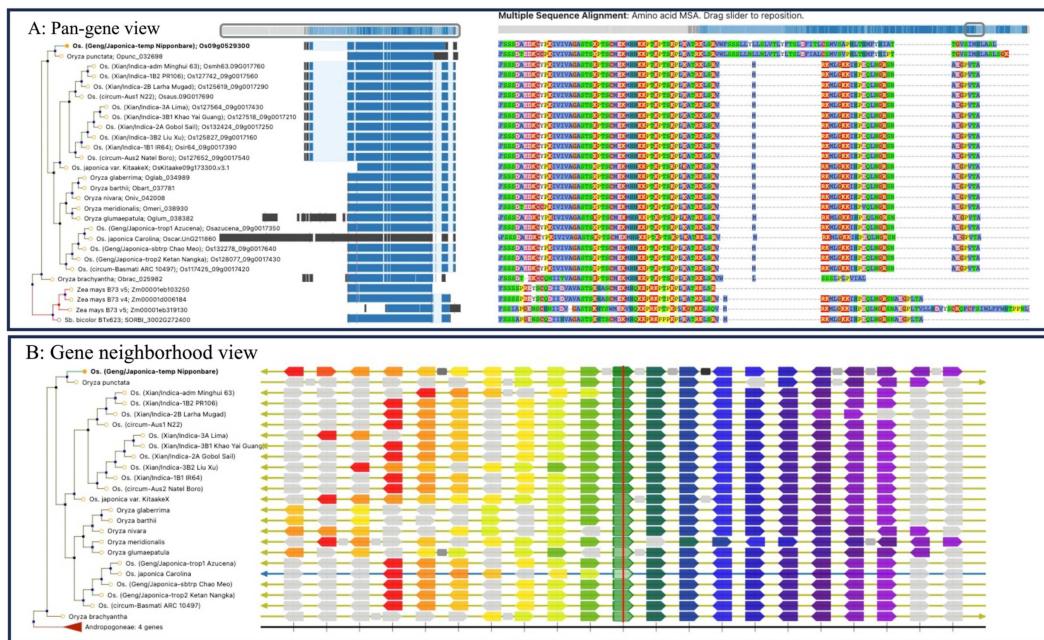
Figure 4. A pan-gene overview for TAC1 transcription factor (reference gene *OsTAC1*; Os09g0529300) and its orthologs from various accessions of cultivated rice *O. sativa,* other members of *Oryza* genus, and two other monocots maize and sorghum at Gramene oryza pansite. Users can explore (**A**) a multiple protein sequence alignment of TAC1 orthologs and (**B**) gene neighborhood conservation.

In addition to the Gramene database, a few more pan-genome portals provide a similar view of pan-genes visualization and analysis tools. For example, the Banana Genome Hub uses the Panache platform to visualize pan-genome data on *Musaceae*. However, the well-established crop genome portals generally support users in exploring genes and gene families, chromosome structures, synteny, structural variations, 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 on maize, rice, and wheat, and discoveries are being implemented for their improvement.

Notably, more recently, 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 helped to identify over 400,000 genomic structural variants and provided insights into heat tolerance. This study also identified the RWP-RK gene conferring enhanced heat tolerance [55]. Another group of previously understudied crops is legumes that have 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 peas identified 225 SNPs associated with nine agronomically important traits. These associations will aid pigeon pea germplasm improvement [154]. In another study, Liu et al. (2022) analyzed 217 mung bean accessions and discovered many novel genes associated with agronomic traits, including 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 gene presence-absence variation in a multi-gene locus associated with color-related traits [163]. Mung bean is an excellent plant-based source of protein and is grown in temperate, subtropical, and tropical regions.

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-genome research are being used for improving banana 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 fruits 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 an understanding of the genetic basis of diverse traits in domesticated crops and offer promising prospects for introducing candidate genes (for disease resistance and quality traits) through molecular breeding or precise genome editing into elite cultivars.

In conclusion, the acquisition of pan-genomes holds immense potential for pursuing fundamental questions related to the evolution of crop genomes as well as for breeding high-yielding crops that are resilient to a range of biotic and abiotic stresses. Utilizing

pan-genomes 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 correlate to 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, and gain knowledge. 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 data and their visualization in the context of pan-genome is crucial for supporting marker-assisted selection, genomic selection, and gene editing efforts towards developing crops that can adapt and thrive in a range of climates and environments, securing global food security and sustainability (see Figure 5).
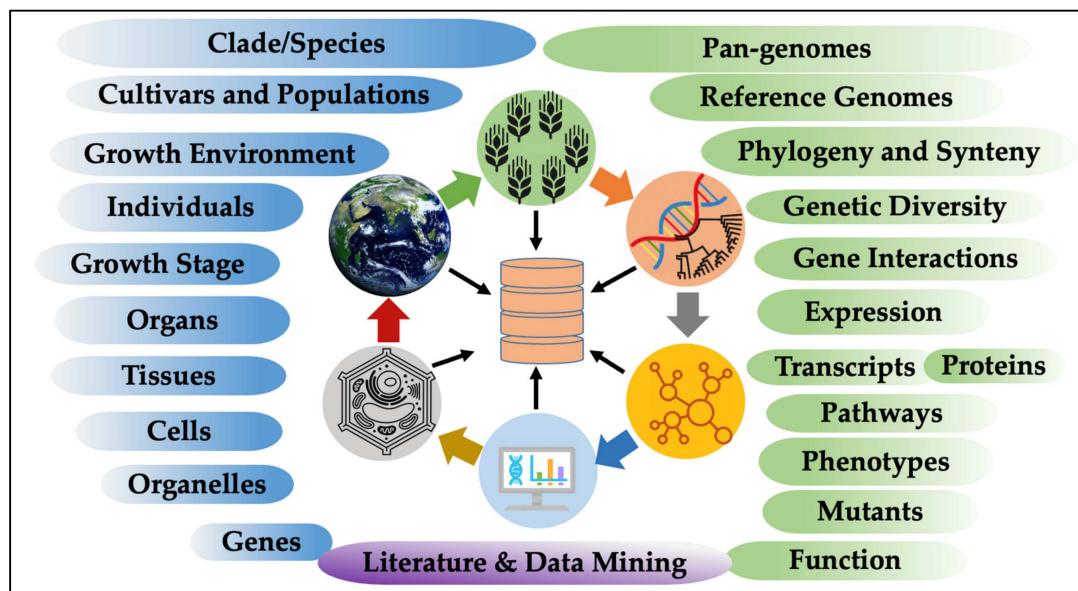
Figure 5. Plant pan-genome browsers can help to integrate heterogeneous omics data to understand gene function, genome evolution and speciation; to establish genotype to phenotype connections; and enable genomic selection, genome editing, and phenotype prediction to support and sustain agriculture production.

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 next-generation sequencing and computational methods required for processing large-scale genomic and transcriptomic data have facilitated rapid and cost-effective whole-genome sequencing [7,174,175] and are now driving pan-genomic research. For the first time in history, researchers are able to explore intra- and inter-species structural variations at the resolution of nucleotide sequence level. We expect a significant expansion in pan-genomes availability across a broader range of plant species. The pan-genomes facilitate the identification of genes conserved across species and genes unique to specific species or a subset of accessions. The availability of species-specific or genus-level pan-genomes 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-pan-genome of the *Citrullus* genus comprising 346 cultivated watermelon accessions and 201 wild accessions suggested that a duplication of the sugar transporter gene *ClTST2* was likely selected during domestication for higher fruit sweetness, and the wild accessions harbor many genes related to disease resistance [178].

In addition, plant pan-genomes can help to advance the fundamental understanding of the plant kingdom at various scales (see Figure 5). 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, pan-genomes can help in understanding 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 pan-genomes, 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 responses to biotic and abiotic stress or stimulants [179,180,187–189]. Integrating pan-genomes 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 drives functional genomics, evolutionary studies, and biodiversity exploration and 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 and 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 making experimental plans [90,97,115,128]. Here, we have reviewed the resources and tools available to the plant research community for pan-genomic research.

We note here that the resources and tools to support researchers in exploring pan-genomic data are limited and at an early developmental stage. We have witnessed a 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 plays an essential role in making sense of the big data and ensuring quality controls at various steps [190]. It could allow researchers to study pan-genomes more thoroughly 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]. However, 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 orphaned crops. We expect that this review will help researchers find appropriate tools and resources relevant to pan-genome construction and analysis. Secondly, it will help to understand and evaluate the strategies employed for pan-genomic studies in crops and encourage them to seek necessary collaborations within the community. We take this opportunity to advocate for increased funding for developing the infrastructure, tools, and biocuration of genomic data. Furthermore, we hope this review can help students and young researchers learn about the status 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.H.D., S.K.S. and P.J. conducted a review of the literature and contributed to writing the manuscript. S.N., C.H.D. and P.J. created all the figures for the manuscript. All authors have read and agreed to the published version of the manuscript.

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

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.* **2018**, *19*, 118–135.
2. Della Coletta, R.; Qiu, Y.; Ou, S.; Hufford, M.B.; Hirsch, C.N. How the Pan-Genome Is Changing Crop Genomics and Improvement. *Genome Biol.* **2021**, *22*, 3. [CrossRef] [PubMed]
3. Ho, S.S.; Urban, A.E.; Mills, R.E. Structural Variation in the Sequencing Era. *Nat. Rev. Genet.* **2020**, *21*, 171–189. [CrossRef]
4. Kyriakidou, M.; Tai, H.; Anglin, N.L.; Ellis, D.; Stromvik, M.V. Current Strategies of Polyploid Plant Genome Sequence Assembly. *Front. Plant Sci.* **2018**, *9*, 1660. [CrossRef] [PubMed]
5. Sedlazeck, F.J.; Rescheneder, P.; Smolka, M.; Fang, H.; Nattestad, M.; von Haeseler, A.; Schatz, M.C. Accurate Detection of Complex Structural Variations Using Single-Molecule Sequencing. *Nat. Methods* **2018**, *15*, 461–468. [CrossRef] [PubMed]

6. Wang, Y.; Yu, J.; Jiang, M.; Lei, W.; Zhang, X.; Tang, H. Sequencing and Assembly of Polyploid Genomes. *Methods Mol. Biol.* **2023**, *2545*, 429–458. [PubMed]

7. Sahu, S.K.; Liu, H. Long-Read Sequencing (Method of the Year 2022): The Way Forward for Plant Omics Research. *Mol. Plant* **2023**, *16*, 791–793. [CrossRef] [PubMed]

8. Zhou, Y.; Chebotarov, D.; Kudrna, D.; Llaca, V.; Lee, S.; Rajasekar, S.; Mohammed, N.; Al-Bader, N.; Sobel-Sorenson, C.; Parakkal, P.; et al. A Platinum Standard Pan-Genome Resource That Represents the Population Structure of Asian Rice. *Sci. Data* **2020**, *7*, 113. [CrossRef] [PubMed]

9. Wang, W.; Mauleon, R.; Hu, Z.; Chebotarov, D.; Tai, S.; Wu, Z.; Li, M.; Zheng, T.; Fuentes, R.R.; Zhang, F.; et al. Genomic Variation in 3010 Diverse Accessions of Asian Cultivated Rice. *Nature* **2018**, *557*, 43–49. [CrossRef] [PubMed]

10. Schatz, M.C.; Maron, L.G.; Stein, J.C.; Hernandez Wences, A.; Gurtowski, J.; Biggers, E.; Lee, H.; Kramer, M.; Antoniou, E.; Ghiban, E.; et al. Whole Genome De Novo Assemblies of Three Divergent Strains of Rice, Oryza Sativa, Document Novel Gene Space of Aus and Indica. *Genome Biol.* **2014**, *15*, 506. [PubMed]

11. Jayakodi, M.; Padmarasu, S.; Haberer, G.; Bonthala, V.S.; Gundlach, H.; Monat, C.; Lux, T.; Kamal, N.; Lang, D.; Himmelbach, A.; et al. The Barley Pan-Genome Reveals the Hidden Legacy of Mutation Breeding. *Nature* **2020**, *588*, 284–289. [CrossRef] [PubMed]

12. Walkowiak, S.; Gao, L.; Monat, C.; Haberer, G.; Kassa, M.T.; Brinton, J.; Ramirez-Gonzalez, R.H.; Kolodziej, M.C.; Delorean, E.; Thambugala, D.; et al. Multiple Wheat Genomes Reveal Global Variation in Modern Breeding. *Nature* **2020**, *588*, 277–283. [CrossRef] [PubMed]

13. Hirsch, C.N.; Foerster, J.M.; Johnson, J.M.; Sekhon, R.S.; Muttoni, G.; Vaillancourt, B.; Penagaricano, F.; Lindquist, E.; Pedraza, M.A.; Barry, K.; et al. Insights into the Maize Pan-Genome and Pan-Transcriptome. *Plant Cell* **2014**, *26*, 121–135. [CrossRef]

14. Liu, Y.; Du, H.; Li, P.; Shen, Y.; Peng, H.; Liu, S.; Zhou, G.A.; Zhang, H.; Liu, Z.; Shi, M.; et al. Pan-Genome of Wild and Cultivated Soybeans. *Cell* **2020**, *182*, 162–176.e13. [CrossRef] [PubMed]

15. Li, Y.H.; Zhou, G.; Ma, J.; Jiang, W.; Jin, L.G.; Zhang, Z.; Guo, Y.; Zhang, J.; Sui, Y.; Zheng, L.; et al. De Novo Assembly of Soybean Wild Relatives for Pan-Genome Analysis of Diversity and Agronomic Traits. *Nat. Biotechnol.* **2014**, *32*, 1045–1052. [CrossRef]

16. Song, J.M.; Guan, Z.; Hu, J.; Guo, C.; Yang, Z.; Wang, S.; Liu, D.; Wang, B.; Lu, S.; Zhou, R.; et al. Eight High-Quality Genomes Reveal Pan-Genome Architecture and Ecotype Differentiation of Brassica Napus. *Nat. Plants* **2020**, *6*, 34–45. [CrossRef] [PubMed]

17. Zhuang, W.; Chen, H.; Yang, M.; Wang, J.; Pandey, M.K.; Zhang, C.; Chang, W.C.; Zhang, L.; Zhang, X.; Tang, R.; et al. The Genome of Cultivated Peanut Provides Insight into Legume Karyotypes, Polyploid Evolution and Crop Domestication. *Nat. Genet.* **2019**, *51*, 865–876. [CrossRef] [PubMed]

18. International Wheat Genome Sequencing, Consortium. Shifting the Limits in Wheat Research and Breeding Using a Fully Annotated Reference Genome. *Science* **2018**, *361*, 6403.

19. Edger, P.P.; Poorten, T.J.; VanBuren, R.; Hardigan, M.A.; Colle, M.; McKain, M.R.; Smith, R.D.; Teresi, S.J.; Nelson, A.D.L.; Wai, C.M.; et al. Origin and Evolution of the Octoploid Strawberry Genome. *Nat. Genet.* **2019**, *51*, 541–547. [CrossRef] [PubMed]

20. Kyriakidou, M.; Anglin, N.L.; Ellis, D.; Tai, H.H.; Stromvik, M.V. Genome Assembly of Six Polyploid Potato Genomes. *Sci. Data* **2020**, *7*, 88. [CrossRef] [PubMed]

21. Shang, L.; Li, X.; He, H.; Yuan, Q.; Song, Y.; Wei, Z.; Lin, H.; Hu, M.; Zhao, F.; Zhang, C.; et al. A Super Pan-Genomic Landscape of Rice. *Cell Res.* **2022**, *32*, 878–896. [CrossRef] [PubMed]

22. He, Q.; Tang, S.; Zhi, H.; Chen, J.; Zhang, J.; Liang, H.; Alam, O.; Li, H.; Zhang, H.; Xing, L.; et al. A Graph-Based Genome and Pan-Genome Variation of the Model Plant Setaria. *Nat. Genet.* **2023**, *55*, 1232–1242. [CrossRef]

23. Yap, I.V.; Schneider, D.; Kleinberg, J.; Matthews, D.; Cartinhour, S.; McCouch, S.R. A Graph-Theoretic Approach to Comparing and Integrating Genetic, Physical and Sequence-Based Maps. *Genetics* **2003**, *165*, 2235–2247. [CrossRef] [PubMed]

24. Tettelin, H.; Masignani, V.; Cieslewicz, M.J.; Donati, C.; Medini, D.; Ward, N.L.; Angiuoli, S.V.; Crabtree, J.; Jones, A.L.; Durkin, A.S.; et al. Genome Analysis of Multiple Pathogenic Isolates of Streptococcus Agalactiae: Implications for the Microbial Pan-Genome. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13950–13955. [CrossRef]

25. Springer, N.M.; Ying, K.; Fu, Y.; Ji, T.; Yeh, C.T.; Jia, Y.; Wu, W.; Richmond, T.; Kitzman, J.; Rosenbaum, H.; et al. Maize Inbreds Exhibit High Levels of Copy Number Variation (Cnv) and Presence/Absence Variation (Pav) in Genome Content. *PLoS Genet.* **2009**, *5*, e1000734. [CrossRef]

26. Anderson, J.E.; Kantar, M.B.; Kono, T.Y.; Fu, F.; Stec, A.O.; Song, Q.; Cregan, P.B.; Specht, J.E.; Diers, B.W.; Cannon, S.B.; et al. A Roadmap for Functional Structural Variants in the Soybean Genome. *G3* **2014**, *4*, 1307–1318. [CrossRef]

27. Golicz, A.A.; Bayer, P.E.; Barker, G.C.; Edger, P.P.; Kim, H.; Martinez, P.A.; Chan, C.K.; Severn-Ellis, A.; McCombie, W.R.; Parkin, I.A.; et al. The Pangenome of an Agronomically Important Crop Plant Brassica Oleracea. *Nat. Commun.* **2016**, *7*, 13390. [CrossRef]

28. Tao, Y.; Luo, H.; Xu, J.; Cruickshank, A.; Zhao, X.; Teng, F.; Hathorn, A.; Wu, X.; Liu, Y.; Shatte, T.; et al. Extensive Variation within the Pan-Genome of Cultivated and Wild Sorghum. *Nat. Plants* **2021**, *7*, 766–773. [CrossRef]

29. Xu, X.; Liu, X.; Ge, S.; Jensen, J.D.; Hu, F.; Li, X.; Dong, Y.; Gutenkunst, R.N.; Fang, L.; Huang, L.; et al. Resequencing 50 Accessions of Cultivated and Wild Rice Yields Markers for Identifying Agronomically Important Genes. *Nat. Biotechnol.* **2011**, *30*, 105–111. [CrossRef] [PubMed]

30. Lam, H.M.; Xu, X.; Liu, X.; Chen, W.; Yang, G.; Wong, F.L.; Li, M.W.; He, W.; Qin, N.; Wang, B.; et al. Resequencing of 31 Wild and Cultivated Soybean Genomes Identifies Patterns of Genetic Diversity and Selection. *Nat. Genet.* **2010**, *42*, 1053–1059. [CrossRef]

31. Gui, S.; Wei, W.; Jiang, C.; Luo, J.; Chen, L.; Wu, S.; Li, W.; Wang, Y.; Li, S.; Yang, N.; et al. A Pan-Zea Genome Map for Enhancing Maize Improvement. *Genome Biol.* **2022**, *23*, 178. [CrossRef]

32. Allaby, R.G.; Ware, R.L.; Kistler, L. A Re-Evaluation of the Domestication Bottleneck from Archaeogenomic Evidence. *Evol. Appl.* **2019**, *12*, 29–37. [CrossRef]

33. Tirnaz, S.; Zandberg, J.; Thomas, W.J.W.; Marsh, J.; Edwards, D.; Batley, J. Application of Crop Wild Relatives in Modern Breeding: An Overview of Resources, Experimental and Computational Methodologies. *Front. Plant Sci.* **2022**, *13*, 1008904. [CrossRef] [PubMed]

34. Papa, R.; Gepts, P. Asymmetry of Gene Flow and Differential Geographical Structure of Molecular Diversity in Wild and Domesticated Common Bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theor. Appl. Genet.* **2003**, *106*, 239–250. [CrossRef]

35. McNally, K.L.; Childs, K.L.; Bohnert, R.; Davidson, R.M.; Zhao, K.; Ulat, V.J.; Zeller, G.; Clark, R.M.; Hoen, D.R.; Bureau, T.E.; et al. Genomewide Snp Variation Reveals Relationships among Landraces and Modern Varieties of Rice. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12273–12278. [CrossRef]

36. Brozynska, M.; Furtado, A.; Henry, R.J. Genomics of Crop Wild Relatives: Expanding the Gene Pool for Crop Improvement. *Plant Biotechnol. J.* **2016**, *14*, 1070–1085. [CrossRef] [PubMed]

37. Bohra, A.; Kilian, B.; Sivasankar, S.; Caccamo, M.; Mba, C.; McCouch, S.R.; Varshney, R.K. Reap the Crop Wild Relatives for Breeding Future Crops. *Trends Biotechnol.* **2022**, *40*, 412–431. [CrossRef]

38. McCouch, S.R.; Rieseberg, L.H. Harnessing Crop Diversity. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2221410120. [CrossRef]

39. McCouch, S. Toward a Plant Genomics Initiative: Thoughts on the Value of Cross-Species and Cross-Genera Comparisons in the Grasses. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 1983–1985. [CrossRef]

40. Wurschum, T.; Rapp, M.; Miedaner, T.; Longin, C.F.H.; Leiser, W.L. Copy Number Variation of Ppd-B1 Is the Major Determinant of Heading Time in Durum Wheat. *BMC Genet.* **2019**, *20*, 64. [CrossRef]

41. Knox, A.K.; Dhillon, T.; Cheng, H.; Tondelli, A.; Pecchioni, N.; Stockinger, E.J. Cbf Gene Copy Number Variation at Frost Resistance-2 Is Associated with Levels of Freezing Tolerance in Temperate-Climate Cereals. *Theor. Appl. Genet.* **2010**, *121*, 21–35. [CrossRef] [PubMed]

42. Maron, L.G.; Guimaraes, C.T.; Kirst, M.; Albert, P.S.; Birchler, J.A.; Bradbury, P.J.; Buckler, E.S.; Coluccio, A.E.; Danilova, T.V.; Kudrna, D.; et al. Aluminum Tolerance in Maize Is Associated with Higher Mate1 Gene Copy Number. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5241–5246. [CrossRef] [PubMed]

43. Cook, D.E.; Lee, T.G.; Guo, X.; Melito, S.; Wang, K.; Bayless, A.M.; Wang, J.; Hughes, T.J.; Willis, D.K.; Clemente, T.E.; et al. Copy Number Variation of Multiple Genes at Rhg1 Mediates Nematode Resistance in Soybean. *Science* **2012**, *338*, 1206–1209. [CrossRef] [PubMed]

44. Liu, Q.; Xu, J.; Zhu, Y.; Mo, Y.; Yao, X.F.; Wang, R.; Ku, W.; Huang, Z.; Xia, S.; Tong, J.; et al. The Copy Number Variation of Osmtd1 Regulates Rice Plant Architecture. *Front. Plant Sci.* **2020**, *11*, 620282. [CrossRef]

45. Wang, Y.; Xiong, G.; Hu, J.; Jiang, L.; Yu, H.; Xu, J.; Fang, Y.; Zeng, L.; Xu, E.; Xu, J.; et al. Copy Number Variation at the Gl7 Locus Contributes to Grain Size Diversity in Rice. *Nat. Genet.* **2015**, *47*, 944–948. [CrossRef]

46. Bosman, R.N.; Vervalle, J.A.; November, D.L.; Burger, P.; Lashbrooke, J.G. Grapevine Genome Analysis Demonstrates the Role of Gene Copy Number Variation in the Formation of Monoterpenes. *Front. Plant Sci.* **2023**, *14*, 1112214. [CrossRef]

47. Falginella, L.; Castellarin, S.D.; Testolin, R.; Gambetta, G.A.; Morgante, M.; Di Gaspero, G. Expansion and Subfunctionalisation of Flavonoid 3′,5′-Hydroxylases in the Grapevine Lineage. *BMC Genom.* **2010**, *11*, 562. [CrossRef]

48. Nilsen, K.T.; Walkowiak, S.; Xiang, D.; Gao, P.; Quilichini, T.D.; Willick, I.R.; Byrns, B.; N'Diaye, A.; Ens, J.; Wiebe, K.; et al. Copy Number Variation of Tddof Controls Solid-Stemmed Architecture in Wheat. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 28708–28718. [CrossRef]

49. Gao, L.; Gonda, I.; Sun, H.; Ma, Q.; Bao, K.; Tieman, D.M.; Burzynski-Chang, E.A.; Fish, T.L.; Stromberg, K.A.; Sacks, G.L.; et al. The Tomato Pan-Genome Uncovers New Genes and a Rare Allele Regulating Fruit Flavor. *Nat. Genet.* **2019**, *51*, 1044–1051. [CrossRef]

50. Liu, J.; Dawe, R.K. Large Haplotypes Highlight a Complex Age Structure within the Maize Pan-Genome. *Genome Res.* **2023**, *33*, 359–370. [CrossRef] [PubMed]

51. Tao, Y.; Zhao, X.; Mace, E.; Henry, R.; Jordan, D. Exploring and Exploiting Pan-Genomics for Crop Improvement. *Mol. Plant* **2019**, *12*, 156–169. [CrossRef] [PubMed]

52. Bayer, P.E.; Golicz, A.A.; Scheben, A.; Batley, J.; Edwards, D. Plant Pan-Genomes Are the New Reference. *Nat. Plants* **2020**, *6*, 914–920. [CrossRef] [PubMed]

53. Jayakodi, M.; Schreiber, M.; Stein, N.; Mascher, M. Building Pan-Genome Infrastructures for Crop Plants and Their Use in Association Genetics. *DNA Res.* **2021**, *28*, dsaa030. [CrossRef] [PubMed]

54. Li, W.; Liu, J.; Zhang, H.; Liu, Z.; Wang, Y.; Xing, L.; He, Q.; Du, H. Plant Pan-Genomics: Recent Advances, New Challenges, and Roads Ahead. *J. Genet. Genom.* **2022**, *49*, 833–846. [CrossRef] [PubMed]

55. Yan, H.; Sun, M.; Zhang, Z.; Jin, Y.; Zhang, A.; Lin, C.; Wu, B.; He, M.; Xu, B.; Wang, J.; et al. Pangenomic Analysis Identifies Structural Variation Associated with Heat Tolerance in Pearl Millet. *Nat. Genet.* **2023**, *55*, 507–518. [CrossRef] [PubMed]

56. Zhou, H.; Yan, F.; Hao, F.; Ye, H.; Yue, M.; Woeste, K.; Zhao, P.; Zhang, S. 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.* **2023**, *10*, uhad015. [CrossRef]

57. Golicz, A.A.; Batley, J.; Edwards, D. Towards Plant Pangenomics. *Plant Biotechnol. J.* **2016**, *14*, 1099–1105. [CrossRef] [PubMed]
58. Garrison, E.; Siren, J.; Novak, A.M.; Hickey, G.; Eizenga, J.M.; Dawson, E.T.; Jones, W.; Garg, S.; Markello, C.; Lin, M.F.; et al. Variation Graph Toolkit Improves Read Mapping by Representing Genetic Variation in the Reference. *Nat. Biotechnol.* **2018**, *36*, 875–879. [CrossRef] [PubMed]
59. Rakocevic, G.; Semenyuk, V.; Lee, W.P.; Spencer, J.; Browning, J.; Johnson, I.J.; Arsenijevic, V.; Nadj, J.; Ghose, K.; Suciu, M.C.; et al. Fast and Accurate Genomic Analyses Using Genome Graphs. *Nat. Genet.* **2019**, *51*, 354–362. [CrossRef]
60. Cheng, H.; Concepcion, G.T.; Feng, X.; Zhang, H.; Li, H. Haplotype-Resolved De Novo Assembly Using Phased Assembly Graphs with Hifiasm. *Nat. Methods* **2021**, *18*, 170–175. [CrossRef]
61. Padgitt-Cobb, L.K.; Kingan, S.B.; Wells, J.; Elser, J.; Kronmiller, B.; Moore, D.; Concepcion, G.; Peluso, P.; Rank, D.; Jaiswal, P.; et al. A Draft Phased Assembly of the Diploid Cascade Hop (*Humulus lupulus*) Genome. *Plant Genome* **2021**, *14*, e20072. [CrossRef]
62. Eizenga, J.M.; Novak, A.M.; Sibbesen, J.A.; Heumos, S.; Ghaffaari, A.; Hickey, G.; Chang, X.; Seaman, J.D.; Rounthwaite, R.; Ebler, J.; et al. Pangenome Graphs. *Annu. Rev. Genom. Hum. Genet* **2020**, *21*, 139–162. [CrossRef]
63. Hickey, G.; Heller, D.; Monlong, J.; Sibbesen, J.A.; Siren, J.; Eizenga, J.; Dawson, E.T.; Garrison, E.; Novak, A.M.; Paten, B. Genotyping Structural Variants in Pangenome Graphs Using the Vg Toolkit. *Genome Biol.* **2020**, *21*, 35. [CrossRef]
64. Vernikos, G.S. A Review of Pangenome Tools and Recent Studies. In *The Pangenome: Diversity, Dynamics and Evolution of Genomes*; Tettelin, H., Medini, D., Eds.; OAPEN: Cham, Switzerland, 2020; pp. 89–112. [CrossRef]
65. Glick, L.; Mayrose, I. The Effect of Methodological Considerations on the Construction of Gene-Based Plant Pan-Genomes. *Genome Biol. Evol.* **2023**, *15*, evad121. [CrossRef]
66. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and Accurate Long-Read Assembly Via Adaptive K-Mer Weighting and Repeat Separation. *Genome Res.* **2017**, *27*, 722–736. [CrossRef] [PubMed]
67. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of Long, Error-Prone Reads Using Repeat Graphs. *Nat. Biotechnol.* **2019**, *37*, 540–546. [CrossRef] [PubMed]
68. Swain, M.T.; Tsai, I.J.; Assefa, S.A.; Newbold, C.; Berriman, M.; Otto, T.D. A Post-Assembly Genome-Improvement Toolkit (Pagit) to Obtain Annotated Genomes from Contigs. *Nat. Protoc.* **2012**, *7*, 1260–1284. [CrossRef] [PubMed]
69. Li, D.; Liu, C.M.; Luo, R.; Sadakane, K.; Lam, T.W. Megahit: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly Via Succinct De Bruijn Graph. *Bioinformatics* **2015**, *31*, 1674–1676. [CrossRef] [PubMed]
70. Tolstoganov, I.; Bankevich, A.; Chen, Z.; Pevzner, P.A. Cloudspades: Assembly of Synthetic Long Reads Using De Bruijn Graphs. *Bioinformatics* **2019**, *35*, i61–i70. [CrossRef] [PubMed]
71. Meleshko, D.; Mohimani, H.; Tracanna, V.; Hajirasouliha, I.; Medema, M.H.; Korobeynikov, A.; Pevzner, P.A. Biosyntheticspades: Reconstructing Biosynthetic Gene Clusters from Assembly Graphs. *Genome Res.* **2019**, *29*, 1352–1362. [CrossRef]
72. Li, H.; Feng, X.; Chu, C. The Design and Construction of Reference Pangenome Graphs with Minigraph. *Genome Biol.* **2020**, *21*, 265. [CrossRef]
73. Guarracino, A.; Heumos, S.; Nahnsen, S.; Prins, P.; Garrison, E. Odgi: Understanding Pangenome Graphs. *Bioinformatics* **2022**, *38*, 3319–3326. [CrossRef]
74. Guarracino, A.; Heumos, S.; Nahnsen, S.; Prins, P.; Garrison, E. Building Pangenome Graphs. *bioRxiv* **2023**, 535718. [CrossRef]
75. Hickey, G.; Monlong, J.; Ebler, J.; Novak, A.M.; Eizenga, J.M.; Gao, Y.; Human Pangenome Reference, C.; Marschall, T.; Li, H.; Paten, B. Pangenome Graph Construction from Genome Alignments with Minigraph-Cactus. *Nat. Biotechnol.* **2023**, 1277. [CrossRef]
76. Armstrong, J.; Hickey, G.; Diekhans, M.; Fiddes, I.T.; Novak, A.M.; Deran, A.; Fang, Q.; Xie, D.; Feng, S.; Stiller, J.; et al. Progressive Cactus Is a Multiple-Genome Aligner for the Thousand-Genome Era. *Nature* **2020**, *587*, 246–251. [CrossRef] [PubMed]
77. Jonkheer, E.M.; van Workum, D.M.; Sheikhizadeh Anari, S.; Brankovics, B.; de Haan, J.R.; Berke, L.; van der Lee, T.A.J.; de Ridder, D.; Smit, S. Pantools V3: Functional Annotation, Classification and Phylogenomics. *Bioinformatics* **2022**, *38*, 4403–4405. [CrossRef] [PubMed]
78. Ewels, P.A.; Peltzer, A.; Fillinger, S.; Patel, H.; Alneberg, J.; Wilm, A.; Garcia, M.U.; Di Tommaso, P.; Nahnsen, S. The Nf-Core Framework for Community-Curated Bioinformatics Pipelines. *Nat. Biotechnol.* **2020**, *38*, 276–278. [CrossRef] [PubMed]
79. Vaughn, J.N.; Branham, S.E.; Abernathy, B.; Hulse-Kemp, A.M.; Rivers, A.R.; Levi, A.; Wechter, W.P. Graph-Based Pangenomics Maximizes Genotyping Density and Reveals Structural Impacts on Fungal Resistance in Melon. *Nat. Commun.* **2022**, *13*, 7897. [CrossRef]
80. Li, H. Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [CrossRef]
81. Marcais, G.; Delcher, A.L.; Phillippy, A.M.; Coston, R.; Salzberg, S.L.; Zimin, A. Mummer4: A Fast and Versatile Genome Alignment System. *PLoS Comput. Biol.* **2018**, *14*, e1005944. [CrossRef] [PubMed]
82. Rautiainen, M.; Marschall, T. Graphaligner: Rapid and Versatile Sequence-to-Graph Alignment. *Genome Biol.* **2020**, *21*, 253. [CrossRef] [PubMed]
83. Kavya, V.N.S.; Tayal, K.; Srinivasan, R.; Sivadasan, N. Sequence Alignment on Directed Graphs. *J. Comput. Biol.* **2019**, *26*, 53–67. [CrossRef]
84. Buchler, T.; Olbrich, J.; Ohlebusch, E. Efficient Short Read Mapping to a Pangenome That Is Represented by a Graph of Ed Strings. *Bioinformatics* **2023**, *39*, btad320. [CrossRef] [PubMed]

85. Poplin, R.; Chang, P.C.; Alexander, D.; Schwartz, S.; Colthurst, T.; Ku, A.; Newburger, D.; Dijamco, J.; Nguyen, N.; Afshar, P.T.; et al. A Universal Snp and Small-Indel Variant Caller Using Deep Neural Networks. *Nat. Biotechnol.* **2018**, *36*, 983–987. [CrossRef] [PubMed]

86. Yun, T.; Li, H.; Chang, P.C.; Lin, M.F.; Carroll, A.; McLean, C.Y. Accurate, Scalable Cohort Variant Calls Using Deepvariant and Glnexus. *Bioinformatics* **2021**, *36*, 5582–5589. [CrossRef]

87. Chiang, C.; Layer, R.M.; Faust, G.G.; Lindberg, M.R.; Rose, D.B.; Garrison, E.P.; Marth, G.T.; Quinlan, A.R.; Hall, I.M. Speedseq: Ultra-Fast Personal Genome Analysis and Interpretation. *Nat. Methods* **2015**, *12*, 966–968. [CrossRef]

88. Eggertsson, H.P.; Jonsson, H.; Kristmundsdottir, S.; Hjartarson, E.; Kehr, B.; Masson, G.; Zink, F.; Hjorleifsson, K.E.; Jonasdottir, A.; Jonasdottir, A.; et al. Graphtyper Enables Population-Scale Genotyping Using Pangenome Graphs. *Nat. Genet.* **2017**, *49*, 1654–1660. [CrossRef] [PubMed]

89. Ebler, J.; Ebert, P.; Clarke, W.E.; Rausch, T.; Audano, P.A.; Houwaart, T.; Mao, Y.; Korbel, J.O.; Eichler, E.E.; Zody, M.C.; et al. Pangenome-Based Genome Inference Allows Efficient and Accurate Genotyping across a Wide Spectrum of Variant Classes. *Nat. Genet.* **2022**, *54*, 518–525. [CrossRef]

90. Naithani, S.; Geniza, M.; Jaiswal, P. Variant Effect Prediction Analysis Using Resources Available at Gramene Database. *Methods Mol. Biol.* **2017**, *1533*, 279–297.

91. Emms, D.M.; Kelly, S. Orthofinder: Phylogenetic Orthology Inference for Comparative Genomics. *Genome Biol.* **2019**, *20*, 238. [CrossRef]

92. Li, L.; Stoeckert, C.J., Jr.; Roos, D.S. Orthomcl: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* **2003**, *13*, 2178–2189. [CrossRef] [PubMed]

93. Miller, J.B.; Pickett, B.D.; Ridge, P.G. Justorthologs: A Fast, Accurate and User-Friendly Ortholog Identification Algorithm. *Bioinformatics* **2019**, *35*, 546–552. [CrossRef] [PubMed]

94. Zhou, S.; Chen, Y.; Guo, C.; Qi, J. Phylomcl: Accurate Clustering of Hierarchical Orthogroups Guided by Phylogenetic Relationship and Inference of Polyploidy Events. *Methods Ecol. Evol.* **2020**, *11*, 943–954. [CrossRef]

95. Altenhoff, A.M.; Train, C.M.; Gilbert, K.J.; Mediratta, I.; Mendes de Farias, T.; Moi, D.; Nevers, Y.; Radoykova, H.S.; Rossier, V.; Warwick Vesztrocy, A.; et al. Oma Orthology in 2021: Website Overhaul, Conserved Isoforms, Ancestral Gene Order and More. *Nucleic Acids Res.* **2021**, *49*, D373–D379. [CrossRef]

96. Persson, E.; Sonnhammer, E.L.L. Inparanoid-Diamond: Faster Orthology Analysis with the Inparanoid Algorithm. *Bioinformatics* **2022**, *38*, 2918–2919. [CrossRef]

97. Naithani, S.; Gupta, P.; Preece, J.; D'Eustachio, P.; Elser, J.L.; Garg, P.; Dikeman, D.A.; Kiff, J.; Cook, J.; Olson, A.; et al. Plant Reactome: A Knowledgebase and Resource for Comparative Pathway Analysis. *Nucleic Acids Res.* **2020**, *48*, D1093–D1103. [CrossRef]

98. Durant, E.; Sabot, F.; Conte, M.; Rouard, M. Panache: A Web Browser-Based Viewer for Linearized Pangenomes. *Bioinformatics* **2021**, *37*, 4556–4558. [CrossRef] [PubMed]

99. Droc, G.; Martin, G.; Guignon, V.; Summo, M.; Sempere, G.; Durant, E.; Soriano, A.; Baurens, F.C.; Cenci, A.; Breton, C.; et al. The Banana Genome Hub: A Community Database for Genomics in the Musaceae. *Hortic. Res* **2022**, *9*, uhac221. [CrossRef] [PubMed]

100. Yokoyama, T.T.; Sakamoto, Y.; Seki, M.; Suzuki, Y.; Kasahara, M. Momi-G: Modular Multi-Scale Integrated Genome Graph Browser. *BMC Bioinform.* **2019**, *20*, 548. [CrossRef] [PubMed]

101. Wick, R.R.; Schultz, M.B.; Zobel, J.; Holt, K.E. Bandage: Interactive Visualization of De Novo Genome Assemblies. *Bioinformatics* **2015**, *31*, 3350–3352. [CrossRef]

102. Beyer, W.; Novak, A.M.; Hickey, G.; Chan, J.; Tan, V.; Paten, B.; Zerbino, D.R. Sequence Tube Maps: Making Graph Genomes Intuitive to Commuters. *Bioinformatics* **2019**, *35*, 5318–5320. [CrossRef] [PubMed]

103. Gonnella, G.; Niehus, N.; Kurtz, S. Gfaviz: Flexible and Interactive Visualization of Gfa Sequence Graphs. *Bioinformatics* **2019**, *35*, 2853–2855. [CrossRef]

104. Mikheenko, A.; Kolmogorov, M. Assembly Graph Browser: Interactive Visualization of Assembly Graphs. *Bioinformatics* **2019**, *35*, 3476–3478. [CrossRef]

105. Kunyavskaya, O.; Prjibelski, A.D. Sgtk: A Toolkit for Visualization and Assessment of Scaffold Graphs. *Bioinformatics* **2019**, *35*, 2303–2305. [CrossRef] [PubMed]

106. Durbin, R. Efficient Haplotype Matching and Storage Using the Positional Burrows-Wheeler Transform (Pbwt). *Bioinformatics* **2014**, *30*, 1266–1272. [CrossRef]

107. Novak, A.M.; Garrison, E.; Paten, B. A Graph Extension of the Positional Burrows-Wheeler Transform and Its Applications. *Algorithms Mol. Biol.* **2017**, *12*, 18. [CrossRef] [PubMed]

108. Grytten, I.; Rand, K.D.; Nederbragt, A.J.; Storvik, G.O.; Glad, I.K.; Sandve, G.K. Graph Peak Caller: Calling Chip-Seq Peaks on Graph-Based Reference Genomes. *PLoS Comput. Biol.* **2019**, *15*, e1006731. [CrossRef]

109. Wang, J.; Yang, W.; Zhang, S.; Hu, H.; Yuan, Y.; Dong, J.; Chen, L.; Ma, Y.; Yang, T.; Zhou, L.; et al. A Pangenome Analysis Pipeline Provides Insights into Functional Gene Identification in Rice. *Genome Biol.* **2023**, *24*, 19. [CrossRef] [PubMed]

110. Tahir Ul Qamar, M.; Zhu, X.; Xing, F.; Chen, L.L. Ppspcp: A Plant Presence/Absence Variants Scanner and Pan-Genome Construction Pipeline. *Bioinformatics* **2019**, *35*, 4156–4158. [CrossRef] [PubMed]

111. Harper, L.; Campbell, J.; Cannon, E.K.S.; Jung, S.; Poelchau, M.; Walls, R.; Andorf, C.; Arnaud, E.; Berardini, T.; Birkett, C.; et al. Agbiodata Consortium Recommendations for Sustainable Genomics and Genetics Databases for Agriculture. *Database* **2018**, *2018*, bay088. [CrossRef]

112. Adam-Blondon, A.F.; Alaux, M.; Pommier, C.; Cantu, D.; Cheng, Z.M.; Cramer, G.R.; Davies, C.; Delrot, S.; Deluc, L.; Di Gaspero, G.; et al. Towards an Open Grapevine Information System. *Hortic. Res* **2016**, *3*, 16056. [CrossRef]

113. Bolser, D.; Staines, D.M.; Pritchard, E.; Kersey, P. Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data. *Methods Mol. Biol.* **2016**, *1374*, 115–140.

114. Gupta, P.; Naithani, S.; Preece, J.; Kim, S.; Cheng, T.; D'Eustachio, P.; Elser, J.; Bolton, E.E.; Jaiswal, P. Plant Reactome and Pubchem: The Plant Pathway and (Bio)Chemical Entity Knowledgebases. *Methods Mol. Biol.* **2022**, *2443*, 511–525.

115. Tello-Ruiz, M.K.; Naithani, S.; Gupta, P.; Olson, A.; Wei, S.; Preece, J.; Jiao, Y.; Wang, B.; Chougule, K.; Garg, P.; et al. Gramene 2021: Harnessing the Power of Comparative Genomics and Pathways for Plant Research. *Nucleic. Acids Res.* **2021**, *49*, D1452–D1463. [CrossRef] [PubMed]

116. Pasha, A.; Subramaniam, S.; Cleary, A.; Chen, X.; Berardini, T.; Farmer, A.; Town, C.; Provart, N. Araport Lives: An Updated Framework for Arabidopsis Bioinformatics. *Plant Cell* **2020**, *32*, 2683–2686. [CrossRef]

117. Shamimuzzaman, M.; Gardiner, J.M.; Walsh, A.T.; Triant, D.A.; Le Tourneau, J.J.; Tayal, A.; Unni, D.R.; Nguyen, H.N.; Portwood, J.L., 2nd; Cannon, E.K.S.; et al. Maizemine: A Data Mining Warehouse for the Maize Genetics and Genomics Database. *Front. Plant Sci.* **2020**, *11*, 592730. [CrossRef] [PubMed]

118. Gladman, N.; Olson, A.; Wei, S.; Chougule, K.; Lu, Z.; Tello-Ruiz, M.; Meijs, I.; Van Buren, P.; Jiao, Y.; Wang, B.; et al. Sorghumbase: A Web-Based Portal for Sorghum Genetic Information and Community Advancement. *Planta* **2022**, *255*, 35. [CrossRef]

119. Arkin, A.P.; Cottingham, R.W.; Henry, C.S.; Harris, N.L.; Stevens, R.L.; Maslov, S.; Dehal, P.; Ware, D.; Perez, F.; Canon, S.; et al. Kbase: The United States Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.* **2018**, *36*, 566–569. [CrossRef] [PubMed]

120. Yates, A.D.; Allen, J.; Amode, R.M.; Azov, A.G.; Barba, M.; Becerra, A.; Bhai, J.; Campbell, L.I.; Carbajo Martinez, M.; Chakiachvili, M.; et al. Ensembl Genomes 2022: An Expanding Genome Resource for Non-Vertebrates. *Nucleic. Acids Res.* **2022**, *50*, D996–D1003. [CrossRef]

121. Naithani, S.; Preece, J.; D'Eustachio, P.; Gupta, P.; Amarasinghe, V.; Dharmawardhana, P.D.; Wu, G.; Fabregat, A.; Elser, J.L.; Weiser, J.; et al. Plant Reactome: A Resource for Plant Pathways and Comparative Analysis. *Nucleic. Acids Res.* **2017**, *45*, D1029–D1039. [CrossRef] [PubMed]

122. Tello-Ruiz, M.K.; Naithani, S.; Stein, J.C.; Gupta, P.; Campbell, M.; Olson, A.; Wei, S.; Preece, J.; Geniza, M.J.; Jiao, Y.; et al. Gramene 2018: Unifying Comparative Genomics and Pathway Resources for Plant Research. *Nucleic Acids Res.* **2018**, *46*, D1181–D1189. [CrossRef] [PubMed]

123. Naithani, S.; Raja, R.; Waddell, E.N.; Elser, J.; Gouthu, S.; Deluc, L.G.; Jaiswal, P. Vitiscyc: A Metabolic Pathway Knowledgebase for Grapevine (*Vitis vinifera*). *Front. Plant Sci.* **2014**, *5*, 644. [CrossRef]

124. Naithani, S.; Partipilo, C.M.; Raja, R.; Elser, J.L.; Jaiswal, P. Fragariacyc: A Metabolic Pathway Database for Woodland Strawberry Fragaria Vesca. *Front. Plant Sci.* **2016**, *7*, 242. [CrossRef]

125. Woodhouse, M.R.; Cannon, E.K.; Portwood, J.L., 2nd; Harper, L.C.; Gardiner, J.M.; Schaeffer, M.L.; Andorf, C.M. A Pan-Genomic Approach to Genome Databases Using Maize as a Model System. *BMC Plant Biol.* **2021**, *21*, 385. [CrossRef] [PubMed]

126. Kanehisa, M.; Furumichi, M.; Sato, Y.; Kawashima, M.; Ishiguro-Watanabe, M. Kegg for Taxonomy-Based Analysis of Pathways and Genomes. *Nucleic Acids Res.* **2023**, *51*, D587–D592. [CrossRef] [PubMed]

127. Paley, S.; Karp, P.D. The Biocyc Metabolic Network Explorer. *BMC Bioinform.* **2021**, *22*, 208. [CrossRef]

128. Naithani, S.; Jaiswal, P. Pathway Analysis and Omics Data Visualization Using Pathway Genome Databases: Fragariacyc, a Case Study. *Methods Mol. Biol.* **2017**, *1533*, 241–256. [PubMed]

129. Hawkins, C.; Ginzburg, D.; Zhao, K.; Dwyer, W.; Xue, B.; Xu, A.; Rice, S.; Cole, B.; Paley, S.; Karp, P.; et al. Plant Metabolic Network 15: A Resource of Genome-Wide Metabolism Databases for 126 Plants and Algae. *J. Integr. Plant Biol.* **2021**, *63*, 1888–1905. [CrossRef]

130. Foerster, H.; Bombarely, A.; Battey, J.N.D.; Sierro, N.; Ivanov, N.V.; Mueller, L.A. Solcyc: A Database Hub at the Sol Genomics Network (Sgn) for the Manual Curation of Metabolic Networks in Solanum and Nicotiana Specific Databases. *Database* **2018**, *2018*, bay035. [CrossRef] [PubMed]

131. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; et al. Phytozome: A Comparative Platform for Green Plant Genomics. *Nucleic Acids Res.* **2012**, *40*, D1178–D1186. [CrossRef] [PubMed]

132. Deng, C.H.; Naithani, S.; Kumari, S.; Cobo-Simon, I.; Quezada-Rodriguez, E.H.; Skrabisova, M.; Gladman, N.; Correll, M.J.; Sikiru, A.B.; Afuwape, O.O.; et al. Agricultural Sciences in the Big Data Era: Genotype and Phenotype Data Standardization, Utilization and Integration. *Preprints* **2023**, 2023061013. [CrossRef]

133. Sun, C.; Hu, Z.; Zheng, T.; Lu, K.; Zhao, Y.; Wang, W.; Shi, J.; Wang, C.; Lu, J.; Zhang, D.; et al. Rpan: Rice Pan-Genome Browser for Approximately 3000 Rice Genomes. *Nucleic Acids Res.* **2017**, *45*, 597–605. [CrossRef] [PubMed]

134. Zhao, Q.; Feng, Q.; Lu, H.; Li, Y.; Wang, A.; Tian, Q.; Zhan, Q.; Lu, Y.; Zhang, L.; Huang, T.; et al. Pan-Genome Analysis Highlights the Extent of Genomic Variation in Cultivated and Wild Rice. *Nat. Genet.* **2018**, *50*, 278–284. [CrossRef]

135. Gui, S.; Yang, L.; Li, J.; Luo, J.; Xu, X.; Yuan, J.; Chen, L.; Li, W.; Yang, X.; Wu, S.; et al. Zeamap, a Comprehensive Database Adapted to the Maize Multi-Omics Era. *iScience* **2020**, *23*, 101241. [CrossRef] [PubMed]

136. Valentin, G.; Abdel, T.; Gaetan, D.; Jean-Francois, D.; Matthieu, C.; Mathieu, R. Greenphyldb V5: A Comparative Pangenomic Database for Plant Genomes. *Nucleic Acids Res.* **2021**, *49*, D1464–D1471.

137. Bayer, P.E.; Petereit, J.; Durant, E.; Monat, C.; Rouard, M.; Hu, H.; Chapman, B.; Li, C.; Cheng, S.; Batley, J.; et al. Wheat Panache: A Pangenome Graph Database Representing Presence-Absence Variation across Sixteen Bread Wheat Genomes. *Plant Genome* **2022**, *15*, e20221. [CrossRef]

138. Blake, V.C.; Woodhouse, M.R.; Lazo, G.R.; Odell, S.G.; Wight, C.P.; Tinker, N.A.; Wang, Y.; Gu, Y.Q.; Birkett, C.L.; Jannink, J.L.; et al. Graingenes: Centralized Small Grain Resources and Digital Platform for Geneticists and Breeders. *Database* **2019**, *2019*, baz065. [CrossRef]

139. Montenegro, J.D.; Golicz, A.A.; Bayer, P.E.; Hurgobin, B.; Lee, H.; Chan, C.K.; Visendi, P.; Lai, K.; Dolezel, J.; Batley, J.; et al. The Pangenome of Hexaploid Bread Wheat. *Plant J.* **2017**, *90*, 1007–1013. [CrossRef]

140. Li, N.; He, Q.; Wang, J.; Wang, B.; Zhao, J.; Huang, S.; Yang, T.; Tang, Y.; Yang, S.; Aisimutuola, P.; et al. Super-Pangenome Analyses Highlight Genomic Diversity and Structural Variation across Wild and Cultivated Tomato Species. *Nat. Genet.* **2023**, *55*, 852–860. [CrossRef] [PubMed]

141. Barchi, L.; Rabanus-Wallace, M.T.; Prohens, J.; Toppino, L.; Padmarasu, S.; Portis, E.; Rotino, G.L.; Stein, N.; Lanteri, S.; Giuliano, G. Improved Genome Assembly and Pan-Genome Provide Key Insights into Eggplant Domestication and Breeding. *Plant J.* **2021**, *107*, 579–596. [CrossRef] [PubMed]

142. Ou, L.; Li, D.; Lv, J.; Chen, W.; Zhang, Z.; Li, X.; Yang, B.; Zhou, S.; Yang, S.; Li, W.; et al. Pan-Genome of Cultivated Pepper (Capsicum) and Its Use in Gene Presence-Absence Variation Analyses. *New Phytol.* **2018**, *220*, 360–363. [CrossRef]

143. Zhang, B.; Huang, H.; Tibbs-Cortes, L.E.; Vanous, A.; Zhang, Z.; Sanguinet, K.; Garland-Campbell, K.A.; Yu, J.; Li, X. Streamline Unsupervised Machine Learning to Survey and Graph Indel-Based Haplotypes from Pan-Genomes. *Mol. Plant* **2023**, *16*, 975–978. [CrossRef] [PubMed]

144. Torkamaneh, D.; Lemay, M.A.; Belzile, F. The Pan-Genome of the Cultivated Soybean (Pansoy) Reveals an Extraordinarily Conserved Gene Content. *Plant Biotechnol. J.* **2021**, *19*, 1852–1862. [CrossRef]

145. Hubner, S.; Bercovich, N.; Todesco, M.; Mandel, J.R.; Odenheimer, J.; Ziegler, E.; Lee, J.S.; Baute, G.J.; Owens, G.L.; Grassa, C.J.; et al. Sunflower Pan-Genome Analysis Shows That Hybridization Altered Gene Content and Disease Resistance. *Nat. Plants* **2019**, *5*, 54–62. [CrossRef] [PubMed]

146. Jin, S.; Han, Z.; Hu, Y.; Si, Z.; Dai, F.; He, L.; Cheng, Y.; Li, Y.; Zhao, T.; Fang, L.; et al. Structural Variation (Sv)-Based Pan-Genome and Gwas Reveal the Impacts of Svs on the Speciation and Diversification of Allotetraploid Cottons. *Mol. Plant* **2023**, *16*, 678–693. [CrossRef] [PubMed]

147. Liu, H.; Wang, X.; Liu, S.; Huang, Y.; Guo, Y.X.; Xie, W.Z.; Liu, H.; Tahir Ul Qamar, M.; Xu, Q.; Chen, L.L. Citrus Pan-Genome to Breeding Database (Cpbd): A Comprehensive Genome Database for Citrus Breeding. *Mol. Plant* **2022**, *15*, 1503–1505. [CrossRef] [PubMed]

148. Li, Q.; Qi, J.; Qin, X.; Dou, W.; Lei, T.; Hu, A.; Jia, R.; Jiang, G.; Zou, X.; Long, Q.; et al. Citgvd: A Comprehensive Database of Citrus Genomic Variations. *Hortic. Res* **2020**, *7*, 12. [CrossRef] [PubMed]

149. Sun, X.; Jiao, C.; Schwaninger, H.; Chao, C.T.; Ma, Y.; Duan, N.; Khan, A.; Ban, S.; Xu, K.; Cheng, L.; et al. Phased Diploid Genome Assemblies and Pan-Genomes Provide Insights into the Genetic History of Apple Domestication. *Nat. Genet.* **2020**, *52*, 1423–1432. [CrossRef] [PubMed]

150. Song, J.M.; Liu, D.X.; Xie, W.Z.; Yang, Z.; Guo, L.; Liu, K.; Yang, Q.Y.; Chen, L.L. Bnpir: Brassica Napus Pan-Genome Information Resource for 1689 Accessions. *Plant Biotechnol. J.* **2021**, *19*, 412–414. [CrossRef] [PubMed]

151. Qi, W.; Lim, Y.W.; Patrignani, A.; Schlapfer, P.; Bratus-Neuenschwander, A.; Gruter, S.; Chanez, C.; Rodde, N.; Prat, E.; Vautrin, S.; et al. The Haplotype-Resolved Chromosome Pairs of a Heterozygous Diploid African Cassava Cultivar Reveal Novel Pan-Genome and Allele-Specific Transcriptome Features. *Gigascience* **2022**, *11*, giac028. [CrossRef]

152. Ruperao, P.; Thirunavukkarasu, N.; Gandham, P.; Selvanayagam, S.; Govindaraj, M.; Nebie, B.; Manyasa, E.; Gupta, R.; Das, R.R.; Odeny, D.A.; et al. Sorghum Pan-Genome Explores the Functional Utility for Genomic-Assisted Breeding to Accelerate the Genetic Gain. *Front. Plant Sci.* **2021**, *12*, 666342. [CrossRef]

153. Varshney, R.K.; Roorkiwal, M.; Sun, S.; Bajaj, P.; Chitikineni, A.; Thudi, M.; Singh, N.P.; Du, X.; Upadhyaya, H.D.; Khan, A.W.; et al. A Chickpea Genetic Variation Map Based on the Sequencing of 3,366 Genomes. *Nature* **2021**, *599*, 622–627. [CrossRef]

154. Zhao, J.; Bayer, P.E.; Ruperao, P.; Saxena, R.K.; Khan, A.W.; Golicz, A.A.; Nguyen, H.T.; Batley, J.; Edwards, D.; Varshney, R.K. Trait Associations in the Pangenome of Pigeon Pea (*Cajanus cajan*). *Plant Biotechnol. J.* **2020**, *18*, 1946–1954. [CrossRef] [PubMed]

155. Yu, J.; Golicz, A.A.; Lu, K.; Dossa, K.; Zhang, Y.; Chen, J.; Wang, L.; You, J.; Fan, D.; Edwards, D.; et al. Insight into the Evolution and Functional Characteristics of the Pan-Genome Assembly from Sesame Landraces and Modern Cultivars. *Plant Biotechnol. J.* **2019**, *17*, 881–892. [CrossRef]

156. Li, J.; Yuan, D.; Wang, P.; Wang, Q.; Sun, M.; Liu, Z.; Si, H.; Xu, Z.; Ma, Y.; Zhang, B.; et al. Cotton Pan-Genome Retrieves the Lost Sequences and Genes During Domestication and Selection. *Genome Biol.* **2021**, *22*, 119. [CrossRef] [PubMed]

157. Sun, Y.; Wang, J.; Li, Y.; Jiang, B.; Wang, X.; Xu, W.H.; Wang, Y.Q.; Zhang, P.T.; Zhang, Y.J.; Kong, X.D. Pan-Genome Analysis Reveals the Abundant Gene Presence/Absence Variations among Different Varieties of Melon and Their Influence on Traits. *Front. Plant Sci.* **2022**, *13*, 835496. [CrossRef]

158. Li, H.; Wang, S.; Chai, S.; Yang, Z.; Zhang, Q.; Xin, H.; Xu, Y.; Lin, S.; Chen, X.; Yao, Z.; et al. Graph-Based Pan-Genome Reveals Structural and Sequence Variations Related to Agronomic Traits and Domestication in Cucumber. *Nat. Commun.* **2022**, *13*, 682. [CrossRef] [PubMed]

159. Qiao, Q.; Edger, P.P.; Xue, L.; Qiong, L.; Lu, J.; Zhang, Y.; Cao, Q.; Yocca, A.E.; Platts, A.E.; Knapp, S.J.; et al. Evolutionary History and Pan-Genome Dynamics of Strawberry (*Fragaria spp*.). *Proc. Natl. Acad. Sci. USA* **2021**, *118*, 5. [CrossRef] [PubMed]

160. Wang, H.; Tu, R.; Ruan, Z.; Chen, C.; Peng, Z.; Zhou, X.; Sun, L.; Hong, Y.; Chen, D.; Liu, Q.; et al. Photoperiod and Gravistimulation-Associated Tiller Angle Control 1 Modulates Dynamic Changes in Rice Plant Architecture. *Theor. Appl. Genet.* **2023**, *136*, 160. [CrossRef]

161. Yu, B.; Lin, Z.; Li, H.; Li, X.; Li, J.; Wang, Y.; Zhang, X.; Zhu, Z.; Zhai, W.; Wang, X.; et al. Tac1, a Major Quantitative Trait Locus Controlling Tiller Angle in Rice. *Plant J.* **2007**, *52*, 891–898. [CrossRef]

162. Boukail, S.; Macharia, M.; Miculan, M.; Masoni, A.; Calamai, A.; Palchetti, E.; Dell'Acqua, M. Genome Wide Association Study of Agronomic and Seed Traits in a World Collection of Proso Millet (*Panicum miliaceum* L.). *BMC Plant Biol.* **2021**, *21*, 330. [CrossRef]

163. Liu, C.; Wang, Y.; Peng, J.; Fan, B.; Xu, D.; Wu, J.; Cao, Z.; Gao, Y.; Wang, X.; Li, S.; et al. High-Quality Genome Assembly and Pan-Genome Studies Facilitate Genetic Discovery in Mung Bean and Its Improvement. *Plant Commun.* **2022**, *3*, 100352. [CrossRef]

164. D'Hont, A.; Denoeud, F.; Aury, J.M.; Baurens, F.-C.; Carreel, F.; Garsmeur, O.; Noel, B.; Bocs, S.; Droc, G.; Rouard, M.; et al. The Banana (*Musa acuminata*) Genome and the Evolution of Monocotyledonous Plants. *Nature* **2012**, *488*, 213–217. [CrossRef] [PubMed]

165. Fernie, A.R.; Aharoni, A. Pan-Genomic Illumination of Tomato Identifies Novel Gene-Trait Interactions. *Trends Plant Sci.* **2019**, *24*, 882–884. [CrossRef]

166. Huff, M.; Hulse-Kemp, A.M.; Scheffler, B.E.; Youngblood, R.C.; Simpson, S.A.; Babiker, E.; Staton, M. Long-Read, Chromosome-Scale Assembly of Vitis Rotundifolia Cv. Carlos and Its Unique Resistance to Xylella Fastidiosa Subsp. Fastidiosa. *BMC Genom.* **2023**, *24*, 409. [CrossRef]

167. Oren, E.; Dafna, A.; Tzuri, G.; Halperin, I.; Isaacson, T.; Elkabetz, M.; Meir, A.; Saar, U.; Ohali, S.; La, T.; et al. Pan-Genome and Multi-Parental Framework for High-Resolution Trait Dissection in Melon (*Cucumis melo*). *Plant J.* **2022**, *112*, 1525–1542. [CrossRef] [PubMed]

168. Hasan, N.; Choudhary, S.; Naaz, N.; Sharma, N.; Laskar, R.A. Recent Advancements in Molecular Marker-Assisted Selection and Applications in Plant Breeding Programmes. *J. Genet. Eng. Biotechnol.* **2021**, *19*, 128. [CrossRef]

169. Garrido-Cardenas, J.A.; Mesa-Valle, C.; Manzano-Agugliaro, F. Trends in Plant Research Using Molecular Markers. *Planta* **2018**, *247*, 543–557. [CrossRef]

170. Moncada, P.; McCouch, S. Simple Sequence Repeat Diversity in Diploid and Tetraploid Coffea Species. *Genome* **2004**, *47*, 501–509. [CrossRef]

171. McCouch, S.R.; Chen, X.; Panaud, O.; Temnykh, S.; Xu, Y.; Cho, Y.G.; Huang, N.; Ishii, T.; Blair, M. Microsatellite Marker Development, Mapping and Applications in Rice Genetics and Breeding. *Plant Mol. Biol.* **1997**, *35*, 89–99. [CrossRef]

172. Tanksley, S.D.; McCouch, S.R. Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. *Science* **1997**, *277*, 1063–1066. [CrossRef]

173. Morales, K.Y.; Singh, N.; Perez, F.A.; Ignacio, J.C.; Thapa, R.; Arbelaez, J.D.; Tabien, R.E.; Famoso, A.; Wang, D.R.; Septiningsih, E.M.; et al. An Improved 7k Snp Array, the C7air, Provides a Wealth of Validated Snp Markers for Rice Breeding and Genetics Studies. *PLoS ONE* **2020**, *15*, e0232479. [CrossRef]

174. Miller, J.R.; Zhou, P.; Mudge, J.; Gurtowski, J.; Lee, H.; Ramaraj, T.; Walenz, B.P.; Liu, J.; Stupar, R.M.; Denny, R.; et al. Hybrid Assembly with Long and Short Reads Improves Discovery of Gene Family Expansions. *BMC Genom.* **2017**, *18*, 541. [CrossRef]

175. Cheng, C.; Fei, Z.; Xiao, P. Methods to Improve the Accuracy of Next-Generation Sequencing. *Front. Bioeng. Biotechnol.* **2023**, *11*, 982111. [CrossRef] [PubMed]

176. Myburg, A.A.; Grattapaglia, D.; Tuskan, G.A.; Hellsten, U.; Hayes, R.D.; Grimwood, J.; Jenkins, J.; Lindquist, E.; Tice, H.; Bauer, D.; et al. The Genome of Eucalyptus Grandis. *Nature* **2014**, *510*, 356–362. [CrossRef]

177. Shulaev, V.; Sargent, D.J.; Crowhurst, R.N.; Mockler, T.C.; Folkerts, O.; Delcher, A.L.; Jaiswal, P.; Mockaitis, K.; Liston, A.; Mane, S.P.; et al. The Genome of Woodland Strawberry (*Fragaria vesca*). *Nat. Genet.* **2011**, *43*, 109–116. [CrossRef] [PubMed]

178. Wu, S.; Sun, H.; Gao, L.; Branham, S.; McGregor, C.; Renner, S.S.; Xu, Y.; Kousik, C.; Wechter, W.P.; Levi, A.; et al. A Citrullus Genus Super-Pangenome Reveals Extensive Variations in Wild and Cultivated Watermelons and Sheds Light on Watermelon Evolution and Domestication. *Plant Biotechnol. J.* **2023**, *6*, 544282. [CrossRef]

179. Naithani, S.; Dikeman, D.A.; Garg, P.; Al-Bader, N.; Jaiswal, P. Beyond Gene Ontology (Go): Using Biocuration Approach to Improve the Gene Nomenclature and Functional Annotation of Rice S-Domain Kinase Subfamily. *PeerJ* **2021**, *9*, e11052. [CrossRef]

180. Naithani, S.; Komath, S.S.; Nonomura, A.; Govindjee, G. Plant Lectins and Their Many Roles: Carbohydrate-Binding and Beyond. *J. Plant Physiol.* **2021**, *266*, 153531. [CrossRef]

181. Monaco, M.K.; Sen, T.Z.; Dharmawardhana, P.D.; Ren, L.; Schaeffer, M.; Naithani, S.; Amarasinghe, V.; Thomason, J.; Harper, L.; Gardiner, J.; et al. Maize Metabolic Network Construction and Transcriptome Analysis. *Plant Genome* **2013**, *6*, 1–12. [CrossRef]

182. Jaiswal, P.; Usadel, B. Plant Pathway Databases. *Methods Mol. Biol.* **2016**, *1374*, 71–87.

183. Naithani, S.; Nonogaki, H.; Jaiswal, P. Exploring Crossroads between Seed Development and Stress-Response. In *Mechanism of Plant Hormone Signaling under Stress*; Pandey, G.K., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2017; pp. 415–454. [CrossRef]

184. Gene Ontology, C.; Aleksander, S.A.; Balhoff, J.; Carbon, S.; Cherry, J.M.; Drabkin, H.J.; Ebert, D.; Feuermann, M.; Gaudet, P.; Harris, N.L.; et al. The Gene Ontology Knowledgebase in 2023. *Genetics* **2023**, *224*, iyad031. [CrossRef]

185. Cooper, L.; Jaiswal, P. The Plant Ontology: A Tool for Plant Genomics. *Methods Mol. Biol.* **2016**, *1374*, 89–114.

186. Walls, R.L.; Cooper, L.; Elser, J.; Gandolfo, M.A.; Mungall, C.J.; Smith, B.; Stevenson, D.W.; Jaiswal, P. The Plant Ontology Facilitates Comparisons of Plant Development Stages across Species. *Front. Plant Sci.* **2019**, *10*, 631. [CrossRef]

187. Naithani, S.; Mohanty, B.; Elser, J.; D'Eustachio, P.; Jaiswal, P. Biocuration of a Transcription Factors Network Involved in Submergence Tolerance During Seed Germination and Coleoptile Elongation in Rice (*Oryza sativa*). *Plants* **2023**, *12*, 1. [CrossRef] [PubMed]

188. Naithani, S.; Dharmawardhana, P.; Nasrallah, J.B. SCR. In *Handbook of Biologically Active Peptides*; Kastin, A.J., Ed.; Elsevier Science: Amsterdam, Netherlands, 2013; pp. 58–66. ISBN 978-0-12-385095-9.

189. Bolger, M.; Schwacke, R.; Usadel, B. Mapman Visualization of Rna-Seq Data Using Mercator4 Functional Annotations. *Methods Mol. Biol.* **2021**, *2354*, 195–212. [PubMed]

190. Naithani, S.; Gupta, P.; Preece, J.; Garg, P.; Fraser, V.; Padgitt-Cobb, L.K.; Martin, M.; Vining, K.; Jaiswal, P. Involving Community in Genes and Pathway Curation. *Database* **2019**, *2019*, bay146. [CrossRef]

191. Gupta, P.; Geniza, M.; Naithani, S.; Phillips, J.L.; Haq, E.; Jaiswal, P. Chia (*Salvia hispanica*) Gene Expression Atlas Elucidates Dynamic Spatio-Temporal Changes Associated with Plant Growth and Development. *Front. Plant Sci.* **2021**, *12*, 667678. [CrossRef]

192. Hendre, P.S.; Muthemba, S.; Kariba, R.; Muchugi, A.; Fu, Y.; Chang, Y.; Song, B.; Liu, H.; Liu, M.; Liao, X.; et al. African Orphan Crops Consortium (Aocc): Status of Developing Genomic Resources for African Orphan Crops. *Planta* **2019**, *250*, 989–1003. [CrossRef]

193. Chang, Y.; Liu, H.; Liu, M.; Liao, X.; Sahu, S.K.; Fu, Y.; Song, B.; Cheng, S.; Kariba, R.; Muthemba, S.; et al. The Draft Genomes of Five Agriculturally Important African Orphan Crops. *Gigascience* **2019**, *8*, giy152. [CrossRef]