



PANGENOMICS OF PLANTS

¹Kuluev B.R.*, ²Chemeris D.A., ¹Gerashchenkov G.A., ¹Kuluev A.R., ¹Mikhailova E.V.,
¹Vershinina Z.R., ¹Baymiev An.Kh., ¹Garafutdinov R.R., ¹Baymiev Al.Kh., ¹Chemeris A.V.

¹Institute of Biochemistry and Genetics, Ufa Federal Research Center, Russian Academy of Sciences, Russia, Ufa

²GENVED LLC, Moscow, Russia

*E-mail: kuluev@bk.ru

Resume

The concept of plant pangenomes appeared in 2007, but the preliminary pangenomes of corn and soybeans were created in 2010. First pangenomes of three plant species (*Brassica rapa*, *Glycine soja*, and *Oryza sativa*) were constructed only in 2014. In 2016, several species from *Populus* and *Oryza* were used to construct pangenomes for these genera, which formally made those pangenomes super-pangenomes long before the concept of super-pangenomes was described in 2020, already dealing with a taxon having the rank of genus. In the same year (2020), the first *Malus* pangenome was constructed based on sequenced genomes with phased assembly of haplotypes, and because two more wild apple tree species were involved, that phased pangenome also became a *Malus* super-pangenome. In 2022, hyper-pangenomes were generated for representatives of genera *Musa* and *Ensete* of the Musaceae as well as a *Citrus* hyper-pangenome using data on genomes of several genera from the Rutaceae. To date, more than 150 pangenomes of all these types have been constructed, and there is a clear growth trend in the number of pangenomes being built. At the same time, it can be predicted that the number of conventional pangenomes will grow at a slower rate than that of phased super-pangenomes because the latter are of the greatest interest for breeding to create varieties of agricultural plants that are high-yielding and resistant to adverse environmental factors. The reason for this interest in plant pangenomes is that reference genomes of individual species, owing to mosaic assembly of determined nucleotide sequences, no longer satisfy the needs of breeders because these data are essentially incomplete information about genomic diversity characteristic of a species/genus or a group of closely related genera of the same family in the form of a gene repertoire consisting of different categories of genes: core, softcore, disposable, and private genes. Although the first two categories mostly ensure the main metabolism, the other two are responsible for secondary metabolism and largely determine the diversity of forms, e.g., by allowing a plant to adapt to its changing environmental conditions. It can be said that agricultural science has already entered the pangenomic era. The most correct selection of different cultivars for breeding should now be based on pangenomic data (including super- and hyper-pangenomes) constructed on the basis of a chromosomal assembly of diploid genomes with phased haplotypes. In fact, genomics, even if it retains its former name, should ideologically turn into pangenomics.

Keywords: genome, pangenome, super-pangenome, hyper-pangenome, pangenomics, T2T phased genome, plant

Citation: Kuluev B.R., Chemeris D.A., Gerashchenkov G.A., Kuluev A.R., Mikhailova E.V., Vershinina Z.R., Baymiev An.Kh., Garafutdinov R.R., Baymiev Al.Kh., Chemeris A.V. Pangenomics of plants. *Biomics*. 2025. V.17(1). P. 42 - 64. DOI: 10.31301/2221-6197.bmcs.2025-4

© The Authors

Introduction

The concept of a reference genome has long outlived its usefulness and is being inexorably replaced by the pangenome containing different nucleotide sequences specific to closely related organisms. The reason is the

inevitable loss of genetic information specific not to one given sample but to a whole plant species during the assembly of determined sequences from a reference genome. Moreover, the most serious losses occur due to omission of large structural variations (SVs), rather than

single-nucleotide polymorphisms (SNPs) or small insertion-deletions (indels). In addition, a reference genome can be figuratively compared to a street lamp that illuminates only the nearest space, and with incorrect reflections, while nothing further is visible. Therefore, only pangenomes provide a general idea of entire genomic diversity. Meanwhile, there are no requirements for the number of sequenced samples so that they can be employed to create or not create pangenomes, and there are wide variations of this parameter within studies in this field. Nonetheless, the type of pangenome depends on this parameter, as discussed below.

In some early studies, as a result of sequencing of several or even multiple samples of the same plant species, researchers noted significant variation of nucleotide sequences, but the term "pangenome" was not used. Nevertheless, we decided to mention some of these studies to complete the picture. It should also be noted that in many papers, representatives of not only one species but also species within a genus have been sequenced, but authors have not called the constructed pangenomes super-pangenomes. Moreover, similar work has been carried out both before the appearance of the very concept of the super-pangenome [Khan et al., 2020] and after. A recent review [He et al., 2025] states that among 110 constructed pangenomes, at least 53 are based on genomic data from two or more species, but only nine of them are declared as super-pangenomes. We believe that it is right to pay attention to this pattern and give such pangenomes the logical status of super-pangenomes. Following the terminological logic, we also propose to introduce the term "hyper-pangenome" for pangenomes covering genomes of representatives of closely related genera within the same family, especially because such entities already exist but have different names. This article will also highlight pangenomes and super-pangenomes built on the basis of diploid genomes with phased assembly of haplotypes, which are in fact the most informative data and allow for a better correlation of a genotype with a phenotype.

A mention of plant pangenomes in this article will be given as much as possible in chronological order for different species, whereas for specific species, information about pangenomes after their first construction will be provided up to the present. Moreover, priority will be given to those species for which super-pangenomes have been generated; this is because they are extremely important for breeding work.

Although in this article, we collected information on almost all plant pangenomes, it is likely not exhaustive.

A brief history of (plant) pangenomics

In the mid-1990s, nucleotide sequences of complete genomes of several free-living organisms, first prokaryotic microorganisms, and then eukaryotic yeast,

became known. At the turn of the century, a complete genome of the *Arabidopsis thaliana* weed model plant was sequenced, and two draft human genomes were constructed. A little later, after being inspired by the results on sequencing of complete genomes of organisms at different levels of genetic complexity and by the common origin of all living things, investigators proposed a new term, «пангеном» (in Russian = pangenome), and this one word was the title of the published article [Tets, 2003]. Nonetheless, that term conceptually united too diverse groups of organisms. In addition, an unsuccessful English translation of the title of that article was "Pangenom," and this word went essentially unnoticed. Even earlier, in 2000, the phrase "pan genome" (in French) was once used in an article on cancer genomics when investigators described a special research program aimed at creating a database containing a genome-wide assessment of the changes in nucleotide sequences of a genome and transcriptome observed in tumors, normal cells, and experimental models in the form of a catalog [Sigaux, 2000]. To be fair, it should be said that in 2001, G.D.Erlich proposed the DGH (the distributed genome hypothesis) for bacteria, according to which not all representatives of a species in a population carry the same set of genes and there is no a single strain that carries all genes of the species but rather a "distributed genome" or "supragenome" acts as a reservoir for a panoply of different genes that provide a significant survival benefit at large [Ehrlich et al., 2004; Hammond et al., 2020].

At last, in 2005, an article was published in which the term "pangenome" (used to describe genomes of several isolates of *Streptococcus agalactiae* bacteria that differ from each other) had the modern meaning uniting a group of similar organisms [Tettelin et al., 2005]. It has been shown that one reference genome of a species does not provide all the information. Then, the term "pangenome" spread to other organisms, including plants. Thus, the plant pangenome concept was first proposed in 2007. Analysis of partially sequenced maize genomes has revealed significant differences between inbred lines Mo17 and B73 [Morgante et al., 2007]. Later, information about the diversity of genomes within a species has continued to accumulate, and some of the first plant pangenomes (more correctly termed pre-pangenomes) have been built for corn by means of knowledge about genomes of six inbred lines [Lai et al., 2010] as well as for soybean on the basis of 17 wild and 14 cultivated specimens [Lam et al. 2010]. In 2011, genomes of several *Arabidopsis* specimens were sequenced, showing significant diversity of nucleotide sequences [Cao et al., 2011; Schneeberger et al., 2011], but the term "pangenome" was not used at the time either.

By that time, a new word, "pangenomics," had emerged, which was used in relation to plants only in April 2016 in a review article and was included in the title

[Golicz et al., 2016], when very few plant pangenomes were constructed. On the other hand, by now, most of articles mentioning "pangenomics" are devoted to prokaryotes, some of the articles deal with bioinformatic methods for composing and analyzing pangenomes of various organisms, and only a few so far are related to plants, among which there are many review papers.

With the accumulation of whole-genome sequencing data, including for specific plant species, it became clear over time that it was absolutely impossible to operate with a reference genomic sequence for one species. DNA polymorphism is much greater than assumed at the beginning of sequencing of complete genomes of organisms. Many authors have already emphasized that reference consensus genomes should be replaced by reference pangenomes because the latter reflect full diversity of DNA polymorphism of a species and this concerning not only plants [Bayer et al., 2020; Khan et al., 2020; Eisenstein, 2023]. At the same time, a new level of assembly of pangenomes has already been achieved by way of haplotyped sequences of a particular species with phased assembly of parental chromosomes separately. Besides, many such pangenomes are already known, as outlined in a description below. Moreover, they are of the greatest interest because they better reflect the real polymorphism of nucleotide sequences on the basis of haplotypes, by taking into account their *cis* and *trans* positions, and therefore lead to true amino acid sequences of the proteins they encode; this correctness is important for establishing a relationship between a genotype and phenotype, sometimes called 'haplo-pheno' [Sinha et al., 2020], although this dual term has not yet been applied directly to pangenomics and to haplotype-resolved genomes with phased assembly.

The genetic basis of (plant) pangenomes

DNA polymorphism can take many forms. In addition to the most widespread SNP and a small indel, representatives of the same species of organisms have greater differences. Moreover, larger genome rearrangements in the form of SVs, including copy number variations (CNVs), presence/absence variations (PAVs), primarily gPAVs, which deal with genes, as well as inversions and translocations of genome regions make a much greater contribution to phenotypic manifestations. When this notion became clear, it was necessary to display all these detectable differences between sequenced samples, thereby leading to the concept of a pangenome, which can be considered a collection of all genes and noncoding regions present or absent in genomes of various specimens of a studied group from a certain taxon, usually a species.

Before proceeding to the main material (in this article), which deals with a variety of plant pangenomes and super-pangenomes, it is necessary to briefly address

several concepts about categories of genes — basic and additional especially because their designations also vary. Via an analysis of genomes of individual species forming pangenomes, it becomes possible to determine which genes are included in the main set for this species, and which are additional and are not present in all the studied samples. Thus, pangenomes carry conserved so-called core genes, which are typical for all sequenced samples and perform basic vital functions, including primary metabolism. Another category is composed of some additional genes that are not present in all samples under study. They are mainly responsible for secondary metabolism and largely determine the variety of forms, including allowing the organism carrying them to adapt to changing environmental conditions. In various articles, they are called dispensable or variable. There is also a group of private genes specific to a few studied samples, and this definition gives them some uniqueness. These are the most commonly used designations for different categories of genes, although there are many other designations as mentioned in the text below. Moreover, the proportions of different categories of genes vary widely within different species, as explained in a discussion below.

Taking into account that unified terminology has not yet been developed, categories of genes will be given here according to how they have been designated by authors of the original articles. Moreover, it should be pointed out that in a number of studies, pangenomes have been created on the basis of genomes sequenced by authors, whereas others also used data previously sequenced by other researchers, and some articles describe pangenomes created solely by means of information from publicly available databases.

Currently, construction of pangenomes is based on three approaches: *de novo* assembly and comparison (*de novo*), reference genome-based iterative assembly (iterative), and a graph-based pangenome (map-to-pan). Each of them has its advantages and limitations, which we will not discuss here because they are described in other reviews [Hu et al., 2024; 2025; Kaur et al., 2024]. At the same time, it is necessary to touch on the types of pangenomes, subdivided into "open" and "closed." "Open pangenomes" include those obtained when the sequencing of the next genome of a given representative of a species/genus replenishes the repertoire of genes previously undetected for this taxon. At the same time, there is a decrease in the shares of core and softcore genes and an increase in the shares of disposable and private genes. "Closed pangenomes" include those for which no newly sequenced genomes replenish the pool of genes characteristic of this type and do not change established ratios of the gene categories. The greater the number of sequenced genomes that are analyzed, the closer the constructed pangenome turns out to be to a closed one. At

the same time, the sooner the number of added genes reaches a plateau, the lower is genetic diversity of a given taxon and vice versa. It can be added here that in an article dedicated to the 10th anniversary of the term "genome," its authors stated that to compile a pangenome, it is desirable to compare genomes of at least five specimens [Vernikos et al., 2015].

Information about numbers of genes specific to constructed pangenomes and forming a certain gene pool or repertoire of genes of a given species can be considered the most important for breeding, and therefore, when pangenomes are described below, special attention will be given to them (where such information is provided by authors of the original papers). In some articles, the size of pangenomes and their increase in comparison with reference genomes are given, but from the point of view of breeding, this information is inferior in importance to the number of genes along with their categories; therefore, genome sizes will not be emphasized with a few exceptions.

Plant pangenomes

Pre-pangenomes, pangenomes, and super-pangenomes

As mentioned above, chronological priority will be given to those plant species for which super-pangenomes have been constructed before (or immediately after) the pangenomes. Nevertheless, the beginning was the end of 2010, when almost simultaneously, first attempts to create plant pangenomes were reported, which were corn *Zea mays* [Lai et al., 2010] and soybean *Glycine max* [Lam et al., 2010], which can be regarded as some kind of pre-pangenomes.

Thus, genomes of six elite maize lines were resequenced, and the data were compared with a reference maize genome, thereby showing many differences between them in terms of SNPs, indels, and SVs, including PAVs [Lai et al., 2010]. To eliminate errors, those authors also resequenced the reference genome of corn line B73. In fact, the maize pangenome was not constructed in that work because the researchers focused on various types of polymorphisms of maize genomes. Later, the progress toward the maize pangenome has continued, and articles have been published describing approaches involving pan-transcriptome and genetic mapping of pangenome sequence anchors [Hirsch et al., 2014; Lu et al., 2015]. A comparison of three complete corn genomes of lines Mo17, PH207, and B73 then revealed many differences in PAVs and other SVs, including those possibly related to heterosis [Sun et al., 2018]. Later, a pangenome analysis of four European and two North American maize lines was carried out, revealing groups of core and dispensable genes [Haberer et al., 2020]. After that, sequencing of 26 diverse maize genomes has allowed to annotate 103,033 pangenes, of which 32,052 genes are in the core or near-core portion of the pangenome where 70,981 are dispensable genes [Hufford et al., 2021]. A

meaningful super pan-Zea genome was constructed in 2022 by means of 11 public assemblies of maize genomes and 721 *de novo*-sequenced accessions, including 507 modern maize lines, 31 landraces, and 183 accessions of teosinte [Gui et al., 2022]. The total length of the pan-Zea genome was ~6.71 billion base pairs (Gbp), versus 2.14 Gbp of the reference genome. A total of 58,944 pan-Zea genes were annotated, among them 44.34% are dispensable. The investigators have paid a lot of attention to gPAV patterns for each maize inbred and teosinte individuals. The Maize Genetics and Genomics Database (MaizeGDB) has been available for corn since 1991 <https://www.maizegdb.org>, in which since 2015 there are data on multiple corn genomes as well as pangenome information [Portwood, 2nd et al., 2019; Woodhouse et al., 2021; 2025; Cannon et al., 2024].

As a result of sequencing of 17 wild-soybean and 14 cultivated-soybean genomes, it has become possible to compare the obtained data with the reference genome, and 4444 large PAVs were found, in which 856 genes were identified, more than 40% of which proved to be responsible for metabolism and other cellular processes [Lam et al., 2010]. In this case, it is more logical to assume that we are dealing with a pre-pangenome. A few years later, based on sequencing of seven genomes, a pangenome was constructed for the wild relative (*G. soja*) of soybean, and it was shown that more than 80% of the genes are core [Li et al., 2014]. After *de novo* sequencing of genomes of 26 wild and cultivated representatives of *G. max* using the reference genome of this species, another soybean pangenome was built [Liu et al., 2020]. A total of 20,623 gene families were identified as core genes, 8163 as softcore genes, 28,679 as dispensable genes, and only 27 genes were found to be private genes. In another study, as a result of sequencing of 204 *G. max* representatives of different geographical origins, a soybean pangenome was constructed, named PanSoy [Torkamaneh et al., 2021]. Those authors gave slightly different designations to the gene categories, and according to their data, 49,431 genes (90.6%) were classified as hardcore and core, 1,402 genes (2.6%) as softcore, 3,402 genes (6.2%) as shell genes, and 297 (0.5%) were classified as cloud genes. In the same year, a Glycine super-pangenome was generated based on 1110 genomes (1042 from *G. max*, 56 from *G. soja*, and two unknown genomes), in which 3765 genes missing from the reference genome were identified [Bayer et al., 2022a]. Other authors have constructed a Glycine super-pangenome based on sequencing of genomes of six perennial species (five diploids: *G. falcata*, *G. stenophita*, *G. cyrtoloba*, *G. syndetika*, and *G. tomentella* and one tetraploid: *G. dolichocarpa*) using 26 genomes of annual species (*G. max* and *G. soja*) [Zhuang et al., 2022]. It was found that 31,936 genes of perennials are core out of a total of 109,827 genes, while the rest were classified as noncore genes; 97,442 genes from a total of 12,9006 genes of annuals were assigned to the same category. Two more reference pangenomes for soybean were recently built based on 12

genomes, 11 of which were sequenced using ONT, and SVs were analyzed [Yano et al., 2025].

One of the first pangenomes was constructed for rice *Oryza sativa* on the basis of its three varieties [Schatz et al., 2014]. In a Venn diagram, one can see that 37,200 genes are core genes, 1,076 are dispensable genes, and 1,464 are private genes. Other authors have created R PAN: a rice pangenome browser for 3,010 genomes of several groups of rice with 28,890 core and candidate core genes and 22,095 dispensable genes [Sun et al., 2017]. Then, a pan-genome analysis of 3010 diverse accessions of Asian cultivated rice allowed researchers to identify more than 10,000 novel protein-coding genes [Wang et al., 2018]. Another pangenome for Asian rice has been constructed on the basis of 12 sequenced genomes [Zhou et al., 2020]. A *japonica* rice pangenome was also generated after sequencing of 239 genomes, and 1131 novel genes were discovered [Liu et al., 2021]. In the same year, a rice pangenome was created based on 33 genomes [Qin et al., 2021]. Long-read sequencing of 111 rice genomes has revealed 19,319 novel protein-coding genes [Zhang et al., 2022]. After a number of other studies, another rice pangenome has been constructed [Wang et al., 2023; Hamilton et al., 2025], including a gene-based pangenome database for rice: Rice Gene Index (RGI) (<https://riceome.hzau.edu.cn/>) [Yu et al., 2023].

Aside from pangenomes for rice, super-pangenomes have also been created in a number of studies, where African rice *O. glaberrima* (three accessions) was involved in the analysis [Monat et al., 2017]. Another super-pangenome of rice has been built based on sequencing results from 66 accessions of *O. sativa* and 74 accessions of *O. rufipogon* [Zhao et al., 2018]. A syntelog-based rice supergenome has also been constructed based on 74 accessions of *O. sativa* and *O. rufipogon* [Wu et al., 2023]. In another work, *O. punctata* was added to the study besides from *O. sativa* and *O. rufipogon* [Zhou et al., 2023]. Nevertheless, there were no mentions of super-pangenomes in these papers, although they were essentially created. In another article [Shang et al., 2023], devoted to the *Oryza* super-pangenome, a “super pangenomic” definition was included the title. Thus, as a result of sequencing of 251 accessions (202 from *O. sativa*, 28 from *O. rufipogon*, 11 from *O. glaberrima*, and 10 from *O. barthii*), investigators generated a graph-based super-pangenome containing 51,359 nonredundant genes among them 21,888 core genes and 29,471 dispensable genes. Those authors also created the Rice Super Pan-genome Information Resource Database — RiceSuperPIRdb (<http://www.ricesuperpir.com/>) based on 215 accessions of four rice species: *O. sativa*, *O. rufipogon*, *O. glaberrima*, and *O. barthii*; another super-pangenome was constructed [Lv et al., 2024]. Recently, with the involvement of 13 wild rice species (*O. glumaepatula*, *O. punctata*, *O. minuta*, *O. malampuzhaensis*, *O. eichingeri*, *O. officinalis*, *O. rhizomatis*, *O. latifolia*, *O. grandiglumis*, *O. alta*, *O. australiensis*, *O. brachyantha*, and *O. meyeriana*) as well as

cultivated species *O. rufipogon*. *O. glaberrima*, *O. sativa* ssp. *japonica*, and *O. sativa* ssp. *indica* were used to create an *Oryza* super-pangenome containing 9,834 core genes (9.66%), 57,822 disposable genes (56.84%), and 34,067 private genes (33.48%) [Long et al., 2024].

Brassica pangenomes, which have already been constructed for a whole group of related species, both diploid and tetraploid, consist of the same subgenomes. For instance, the genus *Brassica* contains three diploid species: *B. rapa* (AA, n = 10), *B. nigra* (BB, n = 8), and *B. oleracea* (CC, n = 9) as well as three amphidiploids: *B. juncea* (AABB, n = 19), *B. carinata* (BBCC, n = 17), and *B. napus* (AACC, n = 19). All of them, with the exception of *B. nigra*, have their own species pangenomes, and there are also super-pangenomes of the genus.

The pangenome of turnip *B. rapa* was the first to be constructed for this species, on the basis of three sequenced genomes [Lin et al., 2014]. The genes present in all genomes (38,186 genes) were named “the common genes,” the genes present in more than one genome were named “the dispensable genes,” and genes specific to only one of the genomes (3,672) were named “the unique genes.” Subsequently, new *B. rapa* pangenomes have been obtained by different authors by means of a larger number of genomes: 16 genomes [Cai et al., 2021], 18 accessions [Wu et al., 2022], and 71 lines [Amas et al., 2023].

Following the turnip pangenome, a formal super-pangenome for cabbage *B. oleracea* has been constructed based on nine sequenced genomes of cultivated forms and the wild species *B. macrocarpa* [Golicz et al., 2016a]. It was shown there that the constructed pangenome contains 61,379 genes among them 49,895 (81.3%) are core genes, 11,484 (18.7%) are variable genes, and 1,322 genes (2.2%) are present only in one line. Not so long ago, a number of pangenomes were built for *B. oleracea* by means of a larger number of samples: six morphotypes [Guo et al., 2024b], seven assembled genomes [Ji et al., 2024], and 27 high-quality genomes [Li et al., 2024b].

A little later, pangenomes for rapeseed *B. napus* began to be created, and the first one was generated based on genomes of two cultivars [Bayer et al., 2017]. After that, 53 sequenced samples were analyzed, due to which 43,327 orthologous genes were identified for rapeseed, of which 28,239 (65.2%) are core genes and 15,088 (34.8%) are variable genes [Hurgobin et al., 2018]. Later, pangenomes for *B. napus* have been constructed based on eight genomes [Song et al., 2020], 1,689 accessions [Song et al., 2021], and 2,902 accessions [Cui et al., 2023]. Moreover, the last two articles describe created Web resources called Bdrip (<http://cbi.hzau.edu.cn/bnapus/>) and BnaOmics (<https://bnaomics.ocri-genomics.net/>), respectively.

In recent years, there was growing interest in involving wild plants as well as other related species in breeding programs, as evidenced by a number of reviews [Khan et al., 2020; Raza et al., 2023]. In this regard, *Brassica*

super-pangenomes are of greater interest. One of these, named by authors as a “cross-species pan-genome,” was built on the basis of 15 genomes, including reference genomes of *B. rapa*, *B. nigra*, *B. oleracea*, *B. napus*, *B. juncea*, and *B. carinata* as well as previously constructed pangenomes of *B. oleracea* and *B. napus* [He et al., 2021]. Considering that several of these species are tetraploids, 22 subgenomes A, B, and C were actually compared. A total of 197,465 gene models were identified, and those authors traced changes of the introduced genes to the pangenome when adding another genome to the analysis, taking into account its subgenome composition. In the same year, a super-pangenome for the genus *Brassica* was generated on the basis of three species, two of which are diploid *B. rapa* (77 accessions) and *B. oleracea* (87 accessions), which served as parent forms for tetraploid *B. napus* (79 accessions) [Bayer et al., 2021]. At the same time, pangenomes were composed for each of these species, and it was reported that *B. rapa*, *B. oleracea*, and *B. napus* contain 59,864, 58,315, and 10,858 genes, respectively, of which 67%, 79%, and 62% are core genes. A total of 711, 360, and 955 genes turned out to be unique for each of these species. For *Brassica*, investigators recently constructed a super-pangenome based on 41 genomes (21 from *B. napus*, 15 from *B. rapa*, and five from *B. oleracea*); it was named by those authors a “multi-species graph pangenome” [MacNish et al., 2025]. In the course of that research, those authors created The Brassica Panache Web portal (http://brassicagenome.net/brassica_panache/). Pangenomes for tetraploid species *B. carinata* [Niu et al., 2024] and *B. juncea* [Zhang et al., 2025a] were constructed recently. The pangenome of the former was generated on the basis of four sequenced genomes and was found to be characterized by 88,307 core genes, 21,262 softcore genes, 16,852 shell genes, and 792 cloud genes. The *B. juncea* pangenome was constructed on the basis of 149 accessions and is characterized by 55,962 core genes, 24,042 softcore genes, 45,337 shell genes, and 309 cloud genes.

As a result of sequencing of genomes of 12 potato clones, noticeable variation was found among them by way of PAVs and other SVs [Hardigan et al., 2016]. In fact, that work can be considered the first attempt to build a potato pangenome because those authors identified groups of core genes and dispensable genes, indicating that the latter play an important role in adaptation to environmental conditions. It was also noted that the core potato gene set contains 30,401 genes (77.4%), with 32% of the genes missing from at least one of the 12 clones. On the other hand, first meaningful potato pangenomes appeared only in 2022 and 2023 [Tang et al., 2022; Hoopes et al., 2022], and the latter two were created via a phased genome assembly. Such an approach is still not widespread due to its complexity, despite the importance of such results, as emphasized in a few articles [Michael, VanBuren, 2020; Jonkheer et al., 2025]. Therefore, similar plant pangenomes of different species, including

other phased pangenomes of potato [Achakkagari et al., 2024; Cheng et al., 2025], will be discussed here separately.

In a paper [Tang et al., 2022], a *Petota* super-pangenome was constructed based on a large number of wild potato species and cultivars from the *Petota* section of the genus *Solanum*: *Solanum tuberosum*, *S. candolleianum*, *S. pinnatisectum*, *S. andreanum*, *S. burkartii*, *S. lignicaule*, *S. buesii*, *S. multiinterruptum*, *S. brevicaule*, *S. jamesii*, *S. piuriae*, *S. morelliforme*, *S. chomatophilum*, *S. paucisectum*, *S. sogarandinum*, *S. vernei*, *S. chacoense*, *S. commersonii*, *S. boliviense*, *S. bulbocastanum*, *S. cajamarquense*, and *S. neorossii* as well as two species of a sister section, *Etuberosum*: *S. etuberosum* and *S. palustre*. Forty-four potato genomes and two *Etuberosum* genomes were *de novo* sequenced, for which 51,401 pangene clusters classified into four categories were identified: core clusters (13,123 genes, 25.5%), softcore clusters (5,743, 11.2%), a shell cluster (28,471, 55.4%), and an accession-specific cluster (4064, 7.9%). The following year, a potato pangenome was compiled on the basis of 15 cultivars grown in Russia, for which a total of 1,050,536 genes were annotated [Karetnikov et al., 2023]. Proportions of the four gene categories—core, softcore, shell, and cloud—for this pangenome were 44.3%, 15.1%, 37.6%, and 3.0%, respectively.

Recently, two articles were published reporting on the creation of potato super-pangenomes [Zhang et al., 2025; Zhu et al., 2025]. In the first work, interspecies pan-genomes (a super-pangenome, actually a hyper-pangenome) were constructed on the basis of 30 genomes, 13 of which belong to the genus *Solanum* (*S. tuberosum*, *S. lycopersicum*, *S. etuberosum*, *S. muricatum*, *S. viarum*, *S. wrightii*, *S. retroflexum*, *S. lyratum*, *S. melongena*, *S. erianthum*, *S. spirale*, *S. laciniatum*, and *S. seaforthianum*), and 17 to non-*Solanum* genera (*Capsicum annuum*, *Iochroma cyaneum*, *Physalis floridana*, *Lycianthes biflora*, *Tubocapsicum anomalum*, *Datura stramonium*, *Lycium barbarum*, *Anisodus luridus*, *Atropa belladonna*, *Przewalskia tangutica*, *Anisodus acutangulus*, *Nicotiana tabacum*, *N. benthamiana*, *N. longiflora*, *Petunia axillaris*, *P. inflata*). The authors [Zhang et al., 2025] classified 43,395 synthetic gene families into three groups: high-retention (present in 26–30 genomes, 51%), medium-retention (present in 6–25 genomes, 17%), and low-retention (present in 2–5 genomes, 31%). Besides, in 30 genomes, the researchers identified 295,763 species-specific genes with 9,858 genes on average (25.36%) per genome. In the second article [Zhu et al. 2025], a *Solanum* super-pangenome was constructed on the basis of *de novo* sequenced genomes of the following species: *Solanum verrucosum*, *S. polyadenium*, *S. tuberosum*, *S. berthaultii*, *S. etuberosum*, *S. ehrenbergii*, *S. violaceimarmoratum*, *S. bulbocastanum*, *S. pinnatisectum*, *S. lignicaule*, *S. cardiophyllum*, *S. brevicaule*, *S. tuberosum*, *S. chomatophilum*, *S. infundibuliforme*, *S. chacoense*, *S. kurtzianum*, *S. brevicaule*, *S. vernei*, *S. microdontum*, *S. commersonii*, *S. raphanifolium*, *S. okadae*, *S. morelliforme*,

S. andreaeanum, *S. jamesii*, *S. candolleianum*, *S. boliviense*, and *S. chacoense*, as well as using a number of other known genomes: there were 50 accessions in total. It was found that the core set consists of 7,485 gene families (536,166 genes), the softcore set of 8,798 gene families, the dispensable set of 24,869 gene families; the private set consisted of 1,254 gene families.

For the genus *Populus*, super-pangenomes were constructed in 2016 and in 2019. In the first case, genomes of *P. nigra* (four samples) and *P. deltoids* (two samples) were compared with the reference genome of *Populus trichocarpa* [Pinosio et al., 2016]. In another study [Zhang et al., 2019], 10 poplar species from five sections were already used to compile a super-pangenome: *P. alba*, *P. davidiana*, *P. cathayana*, *P. simonii*, *P. ussuriensis*, *P. maximowiczii*, *P. nigra*, *P. deltoides*, *P. lasiocarpa*, and *P. euphratica*. In these studies, the main focus was on detecting various SVs in the genomes of these species as compared to the reference genome. To construct another super-pangenome of poplar, 19 genomes of 18 species of this genus were employed: *P. pseudoglaucua*, *P. wuana*, *P. szechuanica*, *P. yunnanensis*, *P. koreana*, *P. trichocarpa*, *P. deltoids*, *P. simonii*, *P. lasiocarpa*, *P. davidiana*, *P. rotundifolia*, *P. tremula*, *P. alba*, *P. qionghaensis*, *P. adenopoda*, *P. euphratica*, *P. pruinosa*, and *P. ilicifolia* (and one subspecies, *P. alba pyramidalis*), six of which were sequenced by those authors [Shi et al., 2024]. Nonetheless, considering that three more genomes of sister willow species *Salix purpurea*, *S. suchowensis*, and *S. brachista* were involved in this study, the obtained pangenome can be regarded as a hyper-pangenome. At the same time, the number of genes for poplars ranged from 32,959 to 44,853, and for willow trees, from 30,209 to 36,937; 12,924 gene families were present in all 19 poplar genomes and were designated as core ones. The number of softcore gene families was 6,827, and the number of dispensable genes was 19,668. Meanwhile, most core and softcore genes showed high synteny (86% and 76%) when compared with those of genes from *Salix*, while dispensable genes and private genes showed much lower synteny: 30% and 18%, respectively. Those authors noted that although the sequenced genomes were not phased, it was possible to conduct a haplotype analysis, which indicated that SVs are characterized by fairly high heterozygosity, ranging from 1.59% to 7.6%. Considerable attention was given to hemizygous genes, the proportion of which varied from 0.63% to 1.42%.

A super-pangenome for pepper *Capsicum annuum* (355 cultivars) has been constructed with the involvement of three more species: *C. baccatum* (four samples), *C. chinense* (11 samples), and *C. frutescens* (13 samples) [Ou et al., 2018]; 51,757 genes were subdivided into species-specific genes and core genes, which amounted to 55.7%.

A tomato super-pangenome was built by means of 725 accessions, including 639 accessions of the cultivated species *Solanum lycopersicum* and three wild species *S.*

pimpinellifolium (78 accessions), *S. cheesmaniae* (three accessions), and *S. galapagense* (five accessions) [Gao et al., 2019]. In addition to the reference genome, 4,873 genes were identified. Later, other authors have constructed a graph-based super-pangenome of *S. lycopersicum* and *S. pimpinellifolium* on the basis of 838 sequenced genomes and found 12,507 genes missing from the reference genome among 51,155 genes [Zhou et al., 2022]. Finally, in 2023, a tomato super-pangenome was created based on two cultivars of *S. lycopersicum* and nine wild species *S. lycopersicoides*, *S. habrochaites*, *S. chilense*, *S. peruvianum*, *S. corneliomulleri*, *S. neorickii*, *S. chmielewski*, *S. pimpinellifolium*, and *S. galapagense* [Li et al., 2023]. It was found that core gene families constitute 54%, and dispensable gene families 38%; 7.6% of gene families were categorized as accession-specific genes.

To create a sunflower super-pangenome, genomes of 493 specimens of 11 species (*Helianthus annuus*, *H. anomalus*, *H. argophyllus*, *H. debilis*, *H. divaricatus*, *H. giganteus*, *H. grosseserratus*, *H. neglectus*, *H. paradoxus*, *H. petiolaris*, and *H. praecox*) were sequenced, of which 287 are cultivars of *H. annuus* [Hubner et al. al., 2019]. The sunflower super-pangenome combined 61,205 genes, of which ~73% were assumed to be core genes; 2,464 rare genes (5.4%) were found in <5% of the accessions.

For the first time, a pangenome for *Sesamum indicum* on the basis of five accessions (two landraces and three cultivars) was generated in 2019 [Yu et al., 2019]. It was determined that ~58% of the genes are core, and the remaining 42% are dispensable genes. Five years later, a super-pangenome for *Sesamum* was constructed by means of cultivated *S. indicum* and six wild species *S. alatum*, *S. latifolium*, *S. angolense*, *S. calycinum*, *S. angustifolium*, and *S. radiatum* [Miao et al., 2024]. As expected, the number of core genes decreased and amounted to ~36%; 56% were dispensable genes, and 7.7% private genes.

Four super-pangenomes have been built for watermelon *Citrullus lanatus*, involving all species of this genus: *C. amarus*, *C. colocynthis*, *C. ecirrhosus*, *C. mucosospermus*, *C. naudinianus*, and *C. rehmi* [Guo et al., 2020; Sun et al., 2023; Wu et al., 2023a; Zhang et al., 2024]. In the first of the articles just cited, the *Citrullus* super-pangenome was constructed based on sequenced genomes of 414 accessions (258 cultivars and 87 landraces from *C. lanatus*, 31 from *C. amarus*, 19 from *C. mucosospermus*, 15 from *C. colocynthis*, two from *C. rehmi*, one from *C. ecirrhosus*, and one from *C. naudinianus*) [Guo et al., 2019]. In the next article [Sun et al., 2023], a *Citrullus* super-pangenome was generated on the basis of 400 sequenced samples, and it was reported that in addition to the reference genome of *C. lanatus*, which carries 22,596 protein-coding genes, the constructed pangenome already contains 28,845 genes, which were categorized as 20,050 core and softcore genes, 7,341 shell genes, and 1,454 cloud genes. In an article by other authors, a watermelon super-pangenome was built

on the basis of 201 samples sequenced by them, using other data on genomes of *Citrullus* species for a total of 547 accessions, among which 349 belonged to *C. lanatus* [Wu et al., 2023]. In a subsequent work, T2T genomes of 27 samples of all seven species of the genus *Citrullus* were sequenced [Zhang et al., 2024]. Among them, there were 13 samples of *C. lanatus*, five samples of *C. amarus*, four samples of *C. colocynthis*, two samples of *C. mucosospermus*, and one sample each of *C. rehmi*, *C. ecirrhosus*, and *C. naudinianus*, with genomes ranging from 361 to 413 Mbp. The number of genes for them ranged from 23,369 to 26,969. At the same time, for all these species, genes were classified as core genes (52.72%), dispensable genes (43.81%), and private genes (3.47%).

Several pangenomes and super-pangenomes have been constructed for a number of cotton species of different ploidy levels. For instance, in 2021, based on 1,581 samples of *Gossypium hirsutum* of different geographical origins and 226 accessions of *G. barbadense*, pangenomes of these species were constructed, which carried 32,569 and 8851 genes more than their reference genomes, respectively [Li et al., 2021]. In their next study, those authors by means of 10 representatives of diploid species with different genomes (*G. herbaceum* (A), *G. anomalum* (B), *G. sturtianum* (C), *G. stocksii* (E), *G. longicalyx* (F), and *G. bickii* (G)) in combination with previously investigated species composed a super-pangenome of cotton containing 17,079 core genes in 22 subgenomes (40.3%), 7,832 (27%) softcore genes, 16,803 variable genes, and 680 genome-specific gene families [Wang et al., 2022]. In their next paper [Li et al., 2024], they generated super-pangenomes for diploid and tetraploid species on the basis of 50 *de novo*-sequenced samples. Based on 15 diploid representatives with A2 genomes, it was shown there that the pangenome carries 25,707 core gene families, 4,953 near-core gene families, 17,314 variable gene families, 550 genome-specific families, and 4,734 singletons. Having constructed a pangenome for diploids with the A genome based on 344 accessions, those authors identified 5,479 novel protein-coding genes [He et al., 2024]. Other researchers using sequenced genomes of 11 accessions of several species (*G. hirsutum*, *G. barbadense*, *G. tomentosum*, *G. mustelinum*, and *G. darwinii*) built super-pangenome Pan-SV (based on SVs) for *G. barbadense* [Jin et al., 2023]. Another paper presents the construction of a cotton super-pangenome based on genomes of 17 diploid species and 10 tetraploid species with different genomic formulas; this approach allowed all genes to be classified into 269,549 syntelog groups, including 6,371 core, 11,994 softcore, 84,866 dispensable, and 166,318 private syntelog groups [Song et al., 2024].

By means of 3,366 sequenced genomes and iterative mapping, a Cicer super-pangenome has been constructed for cultivated chickpea *Cicer arietinum* (3,171 accessions) and wild species: *C. bijugum* (40 accessions), *C. cuneatum* (four accessions), *C. echinospermum* (nine accessions), *C. judaicum* (68 accessions), *C. pinnatifidum* (39

accessions), *C. reticulatum* (28 accessions), and *C. yamashitae* (six accessions) [Varshney et al., 2021]. Out of a total of 29,870 genes, 1,582 novel genes were identified there as compared to the reference genome of *C. arietinum*. Subsequently, those authors continued their research on the Cicer super-pangenome by sequencing new genomes of eight wild species, including previously unsequenced *C. chorassanicum* [Khan et al., 2024]. A total of 24,827 gene families were identified, including 14,748 core, 2,958 softcore, 6,212 dispensable, and 909 species-specific gene families.

A super-pangenome has been built for radish *Raphanus sativus* using samples of the wild species *R. raphanistrum* and their hybrid on the basis of 11 sequenced genomes [Zhang et al., 2021]. A total of 449,856 genes were identified in that work and were grouped into 41,952 gene families; 36% of them were identified as core genes, 59% as dispensable, and ~1% as accession-specific genes.

A super-pangenome for strawberry *Fragaria* spp. has been constructed based on seven diploid species of this genus: *F. iinumae*, *F. nilgerrensis*, *F. daltoniana*, *F. mandshurica*, *F. pentaphylla*, *F. viridis*, and *F. vesca* [Qiao et al., 2021]. Their genomes ranged in size from 229 to 305 Mbp and carried 23,665 to 28,131 genes. At the same time, the number of core genes was 10,665; 13,765 genes were specific to a subset of species and 1257 were species-specific orthogroups for each species.

A super-pangenome for sorghum has been generated based on 15 accessions of *Sorghum bicolor* and one accession of *S. propinquum* [Tao et al., 2021]. It was shown in that work that 15,867 (36%) gene families are core, 2,8026 (63.6%) gene families are shell genes, and 186 (0.4%) are cloud genes. In the same year, a *S. bicolor* pangenome was constructed based on 354 accessions, thereby leading to a 24% increase in the pangenome as compared to a reference genome and making it possible to identify 47% of core genes [Ruperao et al., 2021]. Meanwhile, those authors used a different definition of variable/accessory genes: respectively uniquely present or uniquely absent in any one accession. Later, using 10 accessions, other authors have built another *S. bicolor* pangenome, in which proportions of core genes, shell genes, and cloud genes are 36.69%, 50.32%, and 12.99%, respectively [Voelker et al., 2023].

For the family Musaceae on the basis of representatives of two genera *Musa* and *Ensete*, a pangenome of the banana family has been created, which investigators have named an "intergeneric pangenome" [Rijzaani et al., 2022]. Perhaps due to terminological logic that such pangenomes are better named as a hyper-pangenome. The genus *Musa* was represented by 12 specimens in that article, including six representatives of *M. acuminata* (genome A), one representative of *M. balbisiana* (genome B), *M. itinerans* and *M. textilis* (FeI) (genome T) as well as three hybrids with an A-B genome (one hybrid having AB and two having AAB). The genus *Ensete* was represented by *E. ventricosum*

and included three representatives. On average, each studied representative of these two genera was found to have 34,014 genes. The number of main banana hyper-pangenome genes is 18,288 (± 29). The number of observed variable genes was 29,331. If the analysis is limited to genes of the genus *Musa*, then a larger number of core genes ($27,858 \pm 69$) was predicted because they are absent in the genus *Ensete*.

In 2022, the CPBD database was created: *Citrus* Pan-genome to Breeding Database (<http://citrus.hzau.edu.cn/>), which contains information on 23 genomes of 17 citrus species (including the following: *Citrus sinensis*, *C. australasica*, *C. clementina*, *C. hongheensis*, *C. grandis*, *C. reticulata*, *C. liuensis*, *C. mangshanensis*, *C. medica*, and *C. ichangensis*) as well as *Fortunella hindsii*, *Atalantia buxifolia*, *Murraya paniculata*, and *Poncirus trifoliata* Liu et al., 2022a]. Taking into account the presence of genomes belonging to species of other genera of the family Rutaceae, the constructed pangenome can also be regarded as a hyper-pangenome. In 2023, other authors composed their hyper-pangenome of citrus fruits by means of 314 accessions, including 182 genomes sequenced by them, including 12 genomes assembled *de novo* [Huang et al., 2023]. The analysis included the following species: *Citrus sinensis*, *Citrus ryukyuensis*, *Citrus crassifolia*, *Citrus japonica*, *Citrus aurantium*, *Citrus limon*, *Citrus paradisi*, *Zanthoxylum armatum*, *Clausena lansium*, *Glycosmis pentaphylla*, *Bergera koenigii*, *Murraya alata*, *Murraya microphylla*, *Luvunga scandens*, *Aegle marmelos*, *Limonia acidissima*, *Feroniella oblata*, *Hesperethusa crenulata*, *Citropsis daweana*, *Citropsis gabunensis*, and *Citropsis gillettiana* as well as a mandarin hybrid. This hyper-pangenome included 318,791 genes forming 31,235 gene families, of which 29.89% were core genes, 5.89% were softcore genes, 55.37% were dispensable genes, and 8.85% were special genes. Another citrus hyper-pangenome was recently constructed based on 11 species (*Citrus sinensis*, *C. clementina*, *C. grandis*, *C. reticulata*, *C. medica*, *C. ichangensis*, and *C. unshiu* as well as *Fortunella hindsii*, *Atalantia buxifolia*, *Murraya paniculata*, and *Poncirus trifoliata*) [Tahir et al., 2025]. A total of 59,261 pangene clusters were identified, of which 12,770 are core, 2,980 are softcore, 5,406 are shell(R) or new genes, and 4,936 are shell(N) or accessory genes missing in the reference genome.

For pea *Pisum sativum*, a super-pangenome with 112,776 pangenes has been constructed based on genomes of 116 cultivars and two samples of *P. fulvum* and *P. abyssinicum*, which were subdivided into four categories: core genes (35%), softcore genes (15%), shell genes (44%), and cloud genes (5%) [Yang et al., 2022].

Investigators have also built a super-pangenome of water caltrop *Trapa* based on diploid (AA) and tetraploid (AABB) forms of *T. natans* as well as the diploid species *T. incisa* (AA) [Zhang et al., 2023]. Sizes of the diploid genomes there ranged from 464 to 480 Mbp, and the tetraploid had a genome of 1,057 Mbp with 68,946 annotated

genes versus 32,457–34,940 for diploids. It was determined that the *Trapa* super-pangenome, taking into account subgenomes, is characterized by a 48% core gene family, a 29% dispensable gene family, and a 23% private gene family.

Based on 13 accessions of apple, including nine cultivars of *Malus domestica* and four wild species *M. sylvestris*, *M. sieversii*, *M. orientalis*, and *M. asiatica*, a gene-based super-pangenome has been constructed [Wang et al., 2023a]. The number of genes in all genomes proved to be 590,746, among which 287,868 genes were revealed as core genes, 123,759 as softcore genes, 159,297 as dispensable genes, and 19,822 genes as specific clusters. Two more pangenomes of apple trees have been compiled by other authors, but because they are based on a phased assembly of genomes, they will be discussed in an appropriate section below.

Based on the T2T genome of *Rosa gigantea* as a reference and four other genomes of species from this genus *R. persica*, *R. chinensis*, *R. rugosa*, and *R. wichuraiana* — a super-pangenome has been constructed consisting of 15,703 core gene families, 13,981 dispensable gene families, and 1,646 species-specific gene families [Zhou et al., 2024].

A super-pangenome was recently built for *Hevea*, which is a tree important for humanity [Fang et al., 2024]. For this purpose, *de novo* sequencing and assembly of genomes were performed on three cultivated and two wild-growing *Hevea brasiliensis* specimens as well as three other species of this genus, *H. nitida*, *H. pauciflora*, and *H. benthamiana*, totaling 94 accessions. Sizes of their assembled genomes ranged from 1.49 to 1.58 Gbp, and the number of genes ranged from 42,386 to 46,095. The investigators identified four groups of genes: core, softcore, dispensable, and private, the proportions of which were 58.58%, 11.54%, 28.95%, and 0.93%, respectively.

To create a lettuce super-pangenome, researchers have used an “iterative map and build” approach with pangenomes of four species (*Lactuca sativa*, *L. saligna*, *L. serriola*, and *L. virosa*) with sizes of 2,185 to 3,459 Mbp based on 474 sequenced accessions [van Workum et al., 2024]. In the meantime, the number of transcripts increased by 1,856–3,576 compared to the reference genomes.

Based on 11 sequenced genomes of ash trees *Fraxinus excelsior*, *F. angustifolia*, *F. sogdiana*, *F. mandshurica*, *F. hupehensis*, *F. chinensis*, *F. baroniana*, *F. ornus*, *F. pennsylvanica*, *F. velutina*, and *F. americana*, a super-pangenome with 54,035 pangene clusters has been constructed using data from 28 previously sequenced genomes of 25 species, with 12,156 core gene clusters, 40,664 dispensable gene clusters, and 1,215 private genes [Liu et al., 2025]. At the same time, genomes of these species contained 48,389 to 53,800 protein-coding genes. Gene sets were categorized as core (53.9%), softcore (16.94%), dispensable (28.06%), and private genes (1.1%).

Recently, 11 species of the genus *Chenopodium* of different ploidy levels, including *C. pallidicaule* (2x, A), *C. watsonii* (2x, A), *C. ficifolium* (2x, B), *C. acuminatum* (2x, D), *C. pamiricum* (2x, E), *C. vulvaria* (H), *Chenopodium quinoa* (4x, AB), *C. berlandieri* ssp. *nuttalliae* (4x, AB), *C. sosnowskyi* (4x, AG), *C. strictum* (4x, CD), *C. opulifolium* (6x, BCF), and *C. formosanum* (6x, BCD), were employed to construct a Quinoa super-pangenome carrying 33,457 annotated pangenes [Jaggi et al., 2025]. It was shown there that core genes account for 49%, shell genes for 23%, softcore and cloud genes for 13% each, and cloud-private genes for 2%.

Haplotype-resolved pangenomes

As mentioned above, the construction of pangenomes based on data from genomes with phased assembly is of the greatest interest because this approach allows scientists to see a different level of genomic diversity; two nuclear genomes of one organism in fact constitute a “mini-pangenome.” Many such pangenomes have already been built using phased assembly data, and some plant species have aroused interest among researchers. Nonetheless, here we will follow the chronology, starting with the species for which such information appeared earlier.

The first one is a (super)pangenome constructed for apple tree *Malus domestica* (43 accessions) along with two wild species: *M. sieversii* (37 accessions) and *M. sylvestris* (11 accessions) [Sun et al., 2020]. It was demonstrated in that work that closed pangenomes with a proportion of core genes varying from 81.3% to 87.3% were successfully generated for all species. Another phased pangenome for apple tree *M. domestica* has been constructed elsewhere using four cultivars and three wild accessions from *M. sieversii*, *M. sylvestris*, and *M. baccata* [Su et al., 2024]. It was shown there that ~34% of the genes are core, 63% are dispensable genes, and 1.15% are private genes.

As a result of sequencing of four cultivars of pecan tree *Carya illinoensis*, a pangenome has been constructed, and a diploid genome has been assembled for one cultivar [Lovell et al., 2021]. The number of annotated genes in that paper ranged from 31,042 to 33,280, 21,196 of which were present in all samples. At the same time, it was noted that there are 3,889 blocks of five or more genes missing in other samples. Those authors once again drew attention to the notion that the single-reference-genome paradigm is not sufficient for functional genomics.

A pangenome has been constructed for tetraploid potato *Solanum tuberosum* as a result of sequencing and phased assembly of genomes of six cultivars. In that work, 713,568 genes were annotated in these taxa, 562,550 of which were identified as core genes because they were present in all 12 haplotypes (genomes); 142,225 genes

were identified as shell genes, and 8,793 as cloud genes [Hoopes et al., 2022]. Investigators have also created a potato super-pangenome based on 296 accessions of 60 species and hybrids of the section *Petota*, including cultivars of *S. tuberosum* with different ploidy levels (2×, 3×, 4×, or 5×) and 33 publicly available genome assemblies [Bozan et al., 2023]. The pangenome reached a plateau there after a reading of ~80 genomes. The potato *Petota* pangenome consists of 24 haplotype-resolved haplotypes (12 chromosomes per haploid) with 23,055 core genes, 83,524 shell genes, and 25,776 cloud genes.

As a result of sequencing and phased assembly of the genome of the diploid potato species *S. okadae* with the involvement of a number of other species (*S. brevicaulis*, *S. boliviense*, *S. bukasovii*, *S. chacoense*, *S. chiquidenum*, *S. commersonii*, *S. etuberosum*, *S. gracilifrons*, *S. megistacrolobum*, *S. lignicaule*, *S. megistacrolobum*, *S. paucissectum*, *S. piurae*, *S. tarijense*, *S. gandarillasii*, *S. commersonii*, *S. pinnatisectum*, *S. sparsipilum*, *S. tacnaense*, and *S. tarapatanum*) and hybrids (*S. commersonii* × *S. andigena* and *S. tarnii* × *S. tuberosum*), a super-pangenome has been constructed for the section *Petota* of the genus *Solanum* [Achakkagari et al., 2024]. Another super-pangenome for diploid potatoes was recently generated based on a phased assembly of 60 haplotypes (genomes) including 20 genomes from 10 wild accessions, 38 genomes from 19 domesticated diploid accessions, and two haplotypes from two inbred lines [Cheng et al., 2025]. Those authors noticed elevated heterozygosity in cultured diploids compared to wild specimens: 14.5% versus 9.5%.

Based on the sequenced phased genome of *Rhododendron* × *pulchrum* for the genus *Rhododendron* involving a number of species (*R. griersonianum*, *R. henanense*, *R. ovatum*, *R. ripense*, *R. simsii*, and *R. williamsianum*), a super-pangenome has been constructed, in which groups of genes have been identified: core, softcore, dispensable, and private; the latter category in the above-mentioned species has turned out to constitute 452, 1,382, 1,138, 1,048, 529, 539, and 356 genes, respectively [Shen et al., 2023]. It is also worth noting an increase in the size of the *Rhododendron* super-pangenome, which reached 766 Mbp by the sixth round in that work (with the addition of the next, sixth genome) as compared to initial 509 Mbp and a similar increase in the number of genes: from 35,610 to 53,179. Another super-pangenome for *Rhododendron* was recently constructed based on 15 genomes of 13 species (*R. liliiflorum*, *R. decorum*, *R. platypodum*, *R. concinnum*, *R. delavayi*, *R. griersonianum*, *R. henanense*, *R. irroratum*, *R. kiyosumense*, *R. ripense*, *R. vialii*, *R. nivale*, and *R. williamsianum*) [Wang et al., 2025]. Of the 45,731 genes, 5,734 proved to be core genes, 37,027 dispensable, and 2,970 private genes.

For grapevine *Vitis* spp., several super-pangenomes have been constructed, based on genomes with haplotype-resolved assemblies. For example, in 2023, a pangenome was constructed for three muscadine cultivars of *Vitis rotundifolia* with genome sizes of 393 to 413 Mbp including also one accession of *V. vinifera* [Huff et al., 2023]. For this pangenome, 34,970 synteny-constrained orthogroups were predicted, among them 17,457 are a core gene set; 17,513 formed accessory outgroups, 4,919 orthogroups were “single-individual” groups. Not so long ago, investigators constructed a super-pangenome based on *V. vinifera* and *V. davidii* [Luo et al., 2024]. Another super-pangenome has been created based on nine specimens of *V. vinifera* (18 haplotype genomes), and it has been estimated that the core genome is 48%, whereas dispensable and private genomes constitute 36% and 16% of the genome, respectively [Cochetel et al., 2023]. Another super-pangenome, Grapepan v.1.0, has been built based on 18 haplotype-resolved assembled genomes of eight cultivars of *V. vinifera* and one wild diploid species *V. rotundifolia* [Liu et al., 2024]. Recently, another grape super-pangenome was constructed by means of 72 haplotype-resolved T2T genomes (144 haplotype genomes) including 25 wild and 47 cultivated grapevines; among them, 60 were released for the first time [Guo et al., 2025]; 12.3% genes are core, 14.34% are softcore genes, 71.28% are dispensable genes, and 1.63% are private genes. Those authors concluded that such a super-pangenome would contribute to the improvement of breeding and provide new knowledge on the biology of grapes.

A super-pangenome has been created for spinach on the basis of eight sequenced genomes of cultivars of *Spinacia oleracea*, one genome of the wild species *S. turkestanica*, and two genomes of another wild species: *S. tetrandra* [She et al., 2024]. In two *S. tetrandra* samples, the number of genes was higher: 28,755 and 31,377, while in the remaining samples, it ranged from 25,540 to 26,513. For the super-pangenome, it was determined that 10,623 (37%) are core gene families, 62% (17,831 gene families) are dispensable, and only 308 (1%) are accession-specific gene families.

Based on 13 genomes of the section *Armeniaca*, a graph-based apricot super-pangenome has been generated, including the *Prunus zhengheensis* genome sequenced in haplotype-resolved assembly [Tan et al., 2024]. Genomes of such wild species as *P. mandshurica*, *P. sibirica*, *P. mume*, and *P. hongpingensis* were also utilized there. In total, the number of genes was 30,702. Core and softcore gene families constituted 44.1%, dispensable genes 53.2%, and private gene families 2.7%.

Pangenomes were constructed for two jujube species in 2024. Thus, for Indian jujube *Ziziphus mauritiana*, eight monoploid genomes were built for two accessions of tetraploid cultivated and wild plants with

haplotype-resolved T2T genomes [Guo et al., 2024]. The number of annotated genes varied from 25,205 to 25,728. The resulting pangenome consists of 650 Mbp, which is 1.55 times larger than the average size of eight monoploid genomes. In another work by those authors, a pangenome was composed for the species *Ziziphus jujuba* after *de novo* assembly of genomes for one wild and three cultivated jujube accessions and also using four previously sequenced genomes of these species [Guo et al., 2024a]. A reference genome contained 32,567 gene families, out of which 35% were present as core genes, 63% were categorized as dispensable genes, and 446 gene families (1.37%) were termed accession-specific genes.

A graph-based pangenome for kiwifruit *Actinidia chinensis* has been constructed based on 14 cultivars with chromosomal-scale haplotype-resolved genome assemblies [Wang et al., 2024]. In that paper, 49,770 gene families with 46.6% core genes and dispensable genes were identified, constituting 53.4%; among them, 12.6% were softcore genes and 0.1% cloud genes. A super-pangenome was recently built for kiwifruit [Yu et al., 2025]. To this end, 15 specimens (five males and nine females) of eight species of the genus *Actinidia* (*A. arguta*, *A. polygama*, *A. chinensis*, *A. eriantha*, *A. hemsleyana*, *A. latifolia*, *A. rufa*, and *A. zhejiangensis*) were analyzed by sequencing, with genomes ranging in size from 608 to 652 Mbp and the number of genes ranging from 40,311 to 46,308. A total of 61,465 families of orthologous genes were identified, of which 14,492 were core genes, 5,347 softcore genes, 21,326 dispensable genes, and 20,300 were cloud genes (present only in one assembly).

Based on 11 pear genomes, among which there were cultivars of *Pyrus communis*, *P. pyrifolia*, *P. bretschneideri*, *P. sinkiangensis*, and *P. betulifolia*, hybrids *P. ussuriensis* × *communis*, and four haplotype genomes of two hybrid varieties, a *Pyrus* super-pangenome has been constructed [Li et al., 2024a]. Resequencing data from 139 accessions were mapped on to the pangenome there. Core pangenome size was ~13% of the total pangenome size, confirming extensive evolution within this genus.

A pangenome of moso bamboo *Phyllostachys edulis* has been generated based on genomes of 16 accessions, whose 32 genomes have been assembled in a haplotype-resolved format and ranged in size from 1880 to 2000 Mbp [Hou et al., 2024]. At the same time, numbers of genes in different genomes showed some variation: the smallest difference for one sample between its two genomes amounted to 438 genes, and the largest difference for one sample amounted to 3178 genes. Because a phased assembly was performed for moso bamboo, it was possible to evaluate the presence of core genes from genomes of a single sample. Thus, it turned out that they overwhelmingly have two alleles, whereas for genes with one allele, the frequency was only 0.3%. Those authors paid considerable attention to genes of all

categories in the hemizygous state. In total, 1,738,962 genes were classified into categories (core, softcore, dispensable, and private) in the following proportions: 53.90%, 16.94%, 28.06%, and 1.10%, respectively.

Therefore, approximately two dozen constructed pangenomes and super-pangenomes of different plant species are already known, with phased assembly of all or only some of them. They are of the greatest interest for breeding, as noted in many articles [Chandra et al., 2024; Sarashetti et al., 2024; Ruperao et al., 2025]. As for the phased plant genomes, ~250 have already been assembled for one and a half hundred species, and we have addressed this topic thoroughly [Baymiev et al., 2025], although this is less than 5% of all plant genomes sequenced to date; this state of affairs is explained by the complexity of the procedure.

Other pangenomes

There are much more pangenomes of individual species, and therefore we will touch on only some of them in more detail, while we will only mention the rest.

In 2017, a pangenome of Chinese spring wheat *Triticum aestivum* was constructed based on a reference genome and sequenced genomes of 18 varieties available at the time, thus creating the WheatPan database (<http://appliedbioinformatics.com.au/cgi-bin/gb2/gbrowse/WheatPan/>) [Montenegro et al., 2017]; furthermore, it now contains information about pangenomes of cabbage, rapeseed *B. napus*, and turnip *B. rapa*, described above. It was found that on average, each wheat variety contains an average of 128,656 genes, of which 89,795 (64.3%) are present in all samples, while 49,952 genes represent a variable part of the genome, and on average, 49 genes are unique for each variety. Later, other authors, via a comparison with a refined genome of common wheat, have constructed a new wheat pangenome based on sequenced genomes of 16 varieties and built a database of wheat with a scope (https://www.appliedbioinformatics.com.au/wheat_panache/) [Bayer et al., 2022]. It was shown there that 109,071 genes (68%) are present in all varieties, while 51,460 genes (32%) are present in each species.

Pangenomes have been constructed for barley *Hordeum vulgare* by means of 1,140 genotypes using machine learning [Gao et al., 2020] and 20 accessions [Jayakodi et al., 2020]. The latter authors later have compiled a new pangenome based on long-read sequence assemblies of 76 domesticated and wild accessions and short-read sequence data from 1,315 accessions [Jayakodi et al., 2024].

On the basis of 23 accessions of eggplant *Solanum melongena*, a pangenome has been created containing 35,732 protein-coding genes; among them, 816 additional genes are absent in a reference genome. The pangenome genes were classified in that paper into the following categories: 31,424 core genes, 922 softcore

genes, 1,556 shell genes, and 1,246 cloud genes [Barchi et al., 2021].

Cucumis melo pangenomes have been generated from three accessions [Vaughn et al., 2022], 297 accessions [Sun et al., 2022], 27 accessions [Oren et al., 2022], and nine accessions [Lyu et al., 2023].

A pangenome has been constructed for *Arabidopsis* by means of 32 *de novo*-assembled genomes from different ecotypes in Europe, Asia, Africa, and North America [Kang et al., 2023]. For them, 27,239 to 28,735 protein-coding genes were predicted in that paper; 21,545 gene clusters (68.8%) were present in all 32 ecotypes and were identified as core genes, whereas 3,743 (12.0%) gene clusters were identified as softcore, 3,929 gene clusters (12.6%) as dispensable, and 2,101 (6.7%) as private.

Sequencing of complete genomes of three cultivars of durian *Duriozibethinus* with a predicted number of genes of 45,705, 44,814, and 47,980, respectively, has revealed their substantial genetic variation [Nawae et al., 2023]. A comparison with a previously sequenced reference genome of another variety of this species showed there that it features the absence of 3,074, 6,326, and 4,994 genes, respectively, which allowed to create a draft pangenome of these species.

Pangenomes have also been compiled for a number of species from different families: grass *Brachypodium distachyon* (54 lines) [Gordon et al., 2017], model legume *Medicago truncatula* (15 accessions) [Zhou et al., 2017], pigeon pea *Cajanus cajan* (89 accessions) [Zhao et al., 2020], white lupin *Lupinus albus* (39 accessions) [Hufnagel et al., 2021], Chinese chestnut *Castanea mollissima* (three cultivars) [Hu et al., 2022], mung bean *Vigna radiata* (217 accessions) [Liu et al., 2022], asparagus bean *Vigna unguiculata* (4 accessions) [Liang et al., 2022], narrow-leaved lupin *Lupinus angustifolius* (55 lines) [Garg et al., 2022], cucumber *Cucumis sativus* (12 cultivars) [Li et al., 2022], foxtail millet *Setaria italica* (110 representatives) [He et al., 2023], pearl millet *Pennisetum glaucum* (10 accessions) [Yan et al., 2023], broomcorn millet *Panicum miliaceum* (24 cultivars and eight wild accessions) [Chen et al., 2023a], and tea plant *Camelia sinensis* (22 elite cultivars) [Chen et al., 2023] and on the basis of 10 genomes of cultivars and one wild tea plant [Tariq et al., 2024], common bean *Phaseolus vulgaris* (306 domesticated forms and 33 wild accessions) [Cortinovis et al., 2024].

The past, present, and future of plant pangenomes

Since the construction of the first three true plant pangenomes [Li et al., 2014; Lin et al., 2014; Schatz et al., 2014] and the first review article on plant pangenomics [Golicz et al., 2016], the number of constructed plant pangenomes has reached approximately one and a half hundred over the past decade, as readers can see in Table 1, which lists pangenomes of different types in chronological order with scientific plant names for ranking. Dozens of reviews have also been published, some of which are listed below

Table

Chronology of the constructing different types of plant pangenomes

Year	2010	2011	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
Species Genus Family														
<i>Glycine max</i>	prPG								PG	PG				PG
<i>Zea mays</i>	prPG		prPG	prPG			prPG		PG	PG	PG			
<i>Arabidopsis thaliana</i>		2×prPG									PG			
<i>Brassica rapa</i>			PG							PG	PG	PG		
<i>Glycine soja</i>			PG											
<i>Oryza sativa</i>			PG	PG	PG	PG	PG	PG	PG	2×PG	PG	2×PG		PG
<i>Brassica oleraceae</i>				PG									3×PG	
<i>Oryza</i>				SPG	SPG		SPG					3×SPG	SPG	
<i>Populus</i> Salicaceae				SPG	SPG			SPG						HPG
<i>Solanum tuberosum</i>			PG								PG	PG	PPG	PPG
<i>Brachypodium distachion</i>						PG								
<i>Brassica napus</i>					PG	PG	PG		PG	PG		PG		
<i>Medicago truncatula</i>					PG									
<i>Triticum aestivum</i>					PG						PG			
<i>Capsicum annuum</i>							SPG							
<i>Helianthus annuus</i>								SPG						
<i>Sesamum indicum</i>								PG						
<i>Solanum lycopersicum</i>								SPG			SPG	SPG		
<i>Cajanus cajan</i>									PG					
<i>Citrus</i>									SPG					
<i>Hordeum vulgare</i>									SPG			2×SPG	SPG	
<i>Malus domestica</i>									2×PG				PG	
<i>Malus sieversii</i>									PPG				SPG	PPG
<i>Malus sylvestris</i>									PG					
<i>Brassica</i>									PG					
<i>Carya illinoensis</i>										2×SPG				SPG
<i>Cicer</i>										PPG				
<i>Fragaria</i>										SPG			SPG	
<i>Glycine</i>										SPG				
<i>Gossypium</i>										SPG	SPG			
<i>Lupinus albus</i>										SPG			3×SPG	
<i>Raphanus sativum</i>										PG				
<i>Solanum melongena</i>										PG				
<i>Sorghum</i>										SPG				
<i>Sorghum bicolor</i>										PG			PG	
<i>Castanea mollissima</i>											PG			
<i>Citrus</i> Rutaceae											HPG	HPG		HPG

	2010	2011	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
<i>Cucumis melo</i>											3×PG		PG	
<i>Cucumis sativus</i>											PG			
<i>Lupinus angustifolius</i>											PG			
<i>Musa, Ensete</i> Musaceae											HPG			
<i>Pisum sativum</i>											SPG			
<i>Solanum</i> (potato)											PPG	PPG	PPG	2×SPG PPG
<i>Vigna radiata</i>											PG			
<i>Vigna unguiculata</i>											PG			
<i>Camelia sinensis</i>											PG	PG		
<i>Panicum miliaceum</i>											PG			
<i>Pennisetum glaucum</i>											PG			
<i>Rhododendron</i>											PPG			
<i>Setaria italica</i>											PG			
<i>Trapa natans</i>											SPG			
<i>Vitis</i>											2×PPG	PPG	SPG	
<i>Actinidia chinensis</i>											PG			
<i>Brassica carinata</i>											PG			
<i>Durio zibethinus</i>											PG			
<i>Hevea</i>											SPG			
<i>Lactuca saligna</i>											PG			
<i>Lactuca sativa</i>											PG			
<i>Lactuca serriola</i>											PG			
<i>Lactuca virosa</i>											PG			
<i>Lactuca</i>											SPG			
<i>Malus</i>											PPG			
<i>Phaseolus vulgaris</i>											PG			
<i>Phyllostachys edulis</i>											PPG			
<i>Prunus</i>											PPG			
<i>Pyrus</i>											PPG			
<i>Sesamum</i>											SPG			
<i>Spinacia</i>											SPG			
<i>Ziziphus jujuba</i>											PG			
<i>Ziziphus mauritiana</i>											PG			
<i>Actinidia</i>											SPG			
<i>Brassica juncea</i>											PG			
<i>Chenopodium</i>											SPG			
<i>Fraxinus</i>											SPG			
Years	2010	2011	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
Total	2×prPG	2×prPG	pPG	3×PG	pPG	3×PG	2×SPG	3×PG	9×PG	10×PG	13×PG	12×PG	15×PG	3×PG
						4×PG	2×PG	PG	SPG	7×SPG	4×SPG	9×SPG	10×SPG	7×SPG
							2×SPG	3×SPG	PPG	PPG	2×HPG	HPG	HPG	HPG
											2×PPG	5×PPG	8×PPG	2×PPG
	2	2	4	1	5	4	5	4	11	18	21	27	34	13

prPG – pre-pangenome (7); PG – pangenome (75); SPG – superpangenome (45); HPG – hyperpangenome (5); PPG – phased pangenome (19); Σ - 151

To date, ~150 pangenomes have been compiled for representatives of 29 families. The most strongly represented families are Fabaceae, Brassicaceae, and Poaceae. Among their representatives are soybean, other legume species, rice, and several *Brassica* species. Since 2016, super-pangenomes have begun to appear, which had a different name at the time. In 2020, the first pangenome

was assembled based on a haplotyped assembly of *Malus domestica*, and two wild apple tree species were involved; formally that phased pangenome was also a super-pangenome. Starting from 2022, researchers have begun to compose cross-genus pangenomes or otherwise hyper-pangenomes.

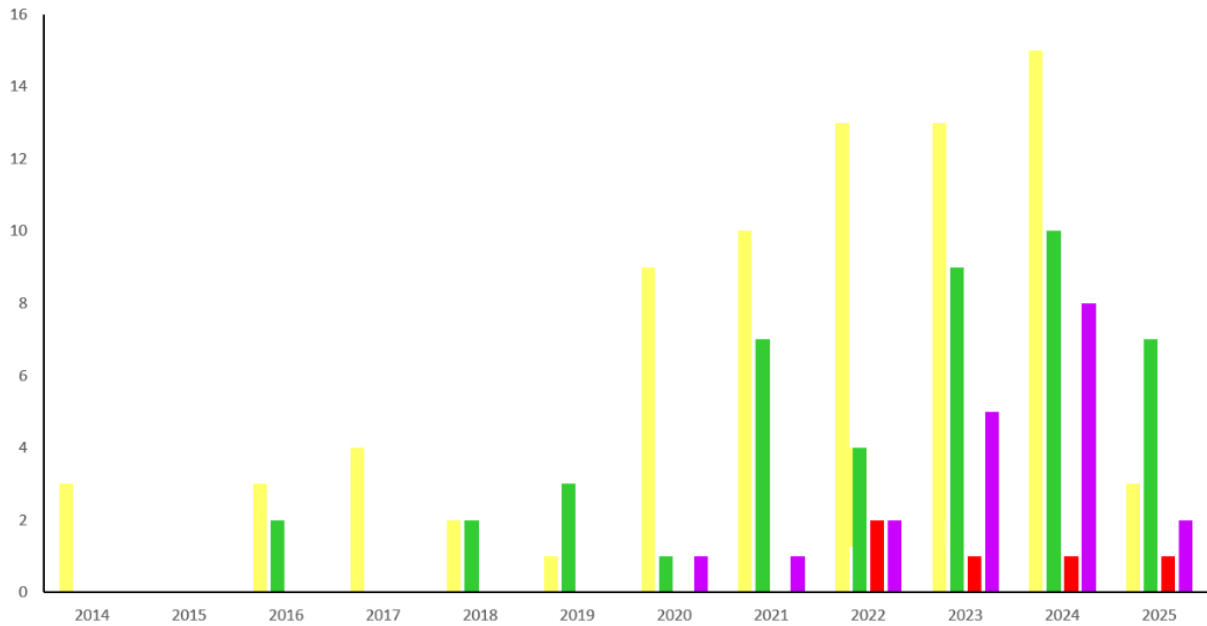


Fig. 1. The number of different types of plant pangenomes constructed from 2014 to 2025. **PG**: pangenome (75); **SPG**: super-pangenome (45); **HPG**: hyper-pangenome (5); **PPG**: phased pangenome (19)

Figure 1 shows that starting from 2020, the number of plant pangenomes of various types has been continuously increasing. Extrapolation of the data for the first 3 months of 2025 suggests that for the whole year 2025, there will be a greater number of composed plant pangenomes than in the previous year.

Although, most likely, the number of common pangenomes will grow to a lesser extent because other types of pangenomes in particular super-pangenomes and phased pangenomes are of greater interest. Accordingly, it has been stated that haplotype phasing is a new frontier in the assembly of plant pangenomes [Michael, VanBuren, 2020], which are required for breeding work of better quality. Another review article provides information on almost three dozen pangenomes of agricultural plants as a timeline, with a mention (in some cases) of phased assembly by haplotypes [Li et al., 2022].

Many reviews have come out about plant pangenomes in general, including many that directly link pangenomics to crop improvement and yield increases, and contain an analysis of relevant resources and specialized software tools for compiling and analyzing

pangenomes [Bayer et al., 2020; Della Coletta et al., 2021; Lei et al., 2021; Hameed et al., 2022; Petereit et al., 2022; Tay Fernandez et al., 2022; Naithani et al., 2023; Shi et al., 2023]. In these reviews, one can even find statements that agricultural science has entered the pangenomic era, and it is hard to disagree with this notion.

It is noteworthy that in 2005, a review article on modern breeding was titled “Genomics-assisted breeding for crop improvement” [Varshney et al., 2005], and a similar article in 2024 is titled “Haplotype-based breeding: A new insight in crop improvement” [Sivabharathi et al., 2024]. This observation indicates that views on breeding have changed markedly, and conventional genomics with consensus genomes is clearly no longer sufficient, but the latter article does not mention pangenomes at all (the same in the former). On the other hand, in 2020, an article titled “5Gs for crop genetic improvement” was published [Varshney et al., 2020]. These 5Gs are genome assembly, germplasm characterization, gene function identification, genomic breeding, and gene editing. In that article [Varshney et al., 2020] researchers emphasize the need to take haplotypes into account because conventional

genomic information as a mosaic assembly does not provide the necessary information. Unfortunately, pangenomes are not mentioned either. The most correct selection of plants should now be based on pangenomic data (including super- and hyper-pangenomes) compiled on the basis of chromosomal assembly of diploid genomes with phased haplotypes. Such pangenomes already exist and were given special attention in our article. Such pangenomes are the future.

Conclusion

Obsolescent reference genomes with mosaic assembly are inevitably replaced by pangenomes (all types) because they contain much more information about a gene repertoire characteristic of a species/genus or a group of closely related genera of the same family. In fact, genomics, even if it retains its original name, should ideologically turn into pangenomics. At the same time, as mentioned above, pangenomes should also preferably be based on knowledge about diploid genomes with phased assembly of haplotypes. These entities can formally be considered mini-pangenomes because both nuclear genomes are of the same species/plant varieties (we will disregard dihaploids for now), certainly show differences, and therefore already provide certain information about genomic diversity. They are of the greatest interest because they better reflect actual polymorphism of nucleotide sequences by haplotypes, taking into account their *cis*-, *trans*-positions, and consequently amino acid sequences of the proteins encoded by them; this feature is extremely important for predicting their functioning and hence for a better understanding of the haplo-pheno relationship.

There is no doubt that the sequencing of pangenomes, super-pangenomes, and hyper-pangenomes will continue, and there will even be acceleration of this process (e.g., through the development of new high-performance sequencing technologies). Pangenomics will develop further, including in terms of data storage and presentation, because knowledge about the diversity of genetic material is important for many tasks, such as the creation of new high-yielding and pest-resistant varieties of plants.

References

- Achakkagari SR, Bozan I, Camargo-Tavares JC et al. The phased *Solanum okadae* genome and Petota pangenome analysis of 23 other potato wild relatives and hybrids. *Sci Data*. 2024. V.11(1). 454. doi: 10.1038/s41597-024-03300-5
- Amas J.C., Bayer P.E., Hong Tan W. et al. Comparative pangenome analyses provide insights into the evolution of *Brassica rapa* resistance gene analogues (RGAs). *Plant Biotechnology Journal*. 2023. V. 21(10). P. 2100–2112. doi: 10.1111/pbi.14116
- Barchi L., Rabanus-Wallace M.T., Prohens J. et al. Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding. *Plant Journal*. 2021. V. 107(2). P. 579–596. doi: 10.1111/tpj.15313
- Bayer PE, Golicz AA, Scheben A et al. Plant pan-genomes are the new reference. *Nat Plants*. 2020. V.6(8). P.914-920. doi: 10.1038/s41477-020-0733-0
- Bayer P.E., Hurgobin B., Golicz A.A. et al. Assembly and comparison of two closely related *Brassica napus* genomes. *Plant Biotechnology Journal*. 2017. V. 15(12). P. 1602–1610. doi: 10.1111/pbi.12742
- Bayer P.E., Scheben A., Golicz A.A. et al. Modelling of gene loss propensity in the pangenomes of three *Brassica* species suggests different mechanisms between polyploids and diploids. *Plant Biotechnology Journal*. 2021. V. 19(12). P. 2488–2500. doi: 10.1111/pbi.13674
- Bayer P.E., Petereit J., Durant É. et al. Wheat Panache: A pangenome graph database representing presence-absence variation across sixteen bread wheat genomes. *Plant Genome*. 2022. V. 15(3). e20221. doi: 10.1002/tpg2.20221
- Bayer P.E., Valliyodan B., Hu H. et al. Sequencing the USDA core soybean collection reveals gene loss during domestication and breeding. *Plant Genome*. 2022a. V. 15(1). e20109. doi: 10.1002/tpg2.20109
- Baymiev Al.Kh., Chemeris D.A., Sakhabutdinova A.R. et al. In higher plants as an example, one can see that the era of sequencing of their diploid genomes is coming. *Biomics*. 2025. V.17(1). P. 17 – 41. DOI: 10.31301/2221-6197.bmcs.2025-3
- Bozan I., Achakkagari S.R., Anglin N.L et al. Pangenome analyses reveal impact of transposable elements and ploidy on the evolution of potato species. *Proc Natl Acad Sci USA*. 2023. V. 120(31). e2211117120. doi: 10.1073/pnas.2211117120
- Cai X., Chang L., Zhang T. et al. Impacts of allopolyploidization and structural variation on intraspecific diversification in *Brassica rapa*. *Genome Biology*. 2021. V. 22(1). 166. doi: 10.1186/s13059-021-02383-2
- Cannon E.K., Portwood J.L., Hayford R.K. et al. Enhanced pan-genomic resources at the maize genetics and genomics database. *Genetics*. 2024. V. 227(1). iyae036. doi: 10.1093/genetics/iyae036
- Cao J., Schneeberger K., Ossowski S. et al. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genetics*. 2011. V. 43(10). P. 956–963. doi: 10.1038/ng.911
- Chandra G, Gibney D, Jain C. Haplotype-aware sequence alignment to pangenome graphs. *Genome Res*. 2024. V.34(9). P.1265-1275. doi: 10.1101/gr.279143.124

15. Chen J., Liu Y., Liu M. et al. Pangenome analysis reveals genomic variations associated with domestication traits in broomcorn millet. *Nature Genetics*. 2023. V. 55(12). P. 2243–2254. doi: 10.1038/s41588-023-01571-z
16. Chen S., Wang P., Kong W. et al. Gene mining and genomics-assisted breeding empowered by the pangenome of tea plant *Camellia sinensis*. *Nature Plants*. 2023a. V. 9(12). P. 1986–1999. doi: 10.1038/s41477-023-01565-z
17. Cheng L., Wang N., Bao Z. et al. Leveraging a phased pangenome for haplotype design of hybrid potato. *Nature*. 2025. V. 22. doi: 10.1038/s41586-024-08476-9
18. Cochetel N, Minio A, Guarracino Anet al. A super-pangenome of the North American wild grape species. *Genome Biol*. 2023. V.24(1). 290. doi: 10.1186/s13059-023-03133-2
19. Cortinovis G., Vincenzi L., Anderson R. et al. Adaptive gene loss in the common bean pan-genome during range expansion and domestication. *Nature Communications*. 2024. V. 15(1). 6698. doi: 10.1038/s41467-024-51032-2.
20. Cui X., Hu M., Yao S. et al. BnaOmics: A comprehensive platform combining pan-genome and multi-omics data from *Brassica napus*. *Plant Communications*. 2023. V. 4(5). 100609. doi: 10.1016/j.xplc.2023.100609
21. Della Coletta R, Qiu Y, Ou S et al. How the pan-genome is changing crop genomics and improvement. *Genome Biol*. 2021. V.22(1). 3. doi: 10.1186/s13059-020-02224-8
22. Ehrlich GD, Hu FZ, Post JC. Role for Biofilms in Infectious Disease. In: Ghannoum M, O’Toole GA, editors. *Microbial Biofilms*. Washington, DC: ASM Press; 2004. P. 332–358.
23. Eisenstein M. Every base everywhere all at once: pangenomics comes of age. *Nature*. 2023. V.616(7957). P.618-620. doi: 10.1038/d41586-023-01300-w
24. Fang Y., Xiao X., Lin J. et al. Pan-genome and phylogenomic analyses highlight *Hevea* species delineation and rubber trait evolution. *Nature Communications*. 2024. V. 15(1). 7232. doi: 10.1038/s41467-024-51031-3
25. Gao L., Gonda I., Sun H. et al. The tomato pangenome uncovers new genes and a rare allele regulating fruit flavor. *Nature Genetics*. 2019. V. 51(6). P. 1044–1051. doi: 10.1038/s41588-019-0410-2
26. Gao S., Wu J., Stiller J. et al. Identifying barley pan-genome sequence anchors using genetic mapping and machine learning. *Theoretical and Applied Genetics*. 2020. V. 133(9). P. 2535–2544. doi: 10.1007/s00122-020-03615-y
27. Garg G., Kamphuis L.G., Bayer P.E. et al. A pan-genome and chromosome-length reference genome of narrow-leafed lupin (*Lupinus angustifolius*) reveals genomic diversity and insights into key industry and biological traits. *Plant Journal*. 2022. V. 111(5). P. 1252–1266. doi: 10.1111/tpj.15885
28. Golicz AA, Batley J, Edwards D. Towards plant pangenomics. *Plant Biotechnol J*. 2016. V.14(4). P.1099-1105. doi: 10.1111/pbi.12499
29. Golicz A.A., Bayer P.E., Barker G.C. et al. The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nature Communications*. 2016a. V. 7. 13390. doi: 10.1038/ncomms13390
30. Gordon S.P., Contreras-Moreira B., Woods D.P. et al. Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nature Communications*. 2017. V. 8(1). 2184. doi: 10.1038/s41467-017-02292-8
31. Gui S., Wei W., Jiang C. et al. A pan-*Zea* genome map for enhancing maize improvement. *Genome Biology*. 2022. V. 23(1). 178. doi: 10.1186/s13059-022-02742-7
32. Guo L, Wang X, Ayhan DH et al. Super pangenome of *Vitis* empowers identification of downy mildew resistance genes for grapevine improvement. *Nat Genet*. 2025. V.57(3). P.741-753. doi: 10.1038/s41588-025-02111-7
33. Guo M., Bi G., Wang H. et al. Genomes of autotetraploid wild and cultivated *Ziziphus mauritiana* reveal polyploid evolution and crop domestication. *Plant Physiology*. 2024. V. 196(4). P. 2701–2720. doi: 10.1093/plphys/kiae512
34. Guo M., Lian Q., Mei Y. et al. Analyzes of pan-genome and resequencing atlas unveil the genetic basis of jujube domestication. *Nature Communications*. 2024a. V. 15. 9320. doi: 10.1038/s41467-024-53718-z
35. Guo N., Wang S., Wang T. et al. A graph-based pan-genome of *Brassica oleracea* provides new insights into its domestication and morphotype diversification. *Plant Communications*. 2024b. V. 5(2). 100791. doi: 10.1016/j.xplc.2023.100791
36. Guo S., Zhao S., Sun H. et al. Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. *Nature Genetics*. 2019. V.51(11). P. 1616–1623. doi: 10.1038/s41588-019-0518-4
37. Haberer G., Kamal N., Bauer E. et al. European maize genomes highlight intraspecies variation in repeat and gene content. *Nature Genetics*. 2020. V.52(9). P. 950–957. doi: 10.1038/s41588-020-0671-9
38. Hameed A, Poznanski P, Nadolska-Orczyk A, Orczyk W. Graph Pangenomes Track Genetic Variants for Crop Improvement. *Int J Mol Sci*. 2022. V.23(21). 13420. doi: 10.3390/ijms232113420
39. Hamilton J.P., Li C., Buell C.R. The rice genome annotation project: an updated database for mining the rice genome. *Nucleic Acids Reserch*. 2025. V.53(D1). P. D1614–D1622. doi: 10.1093/nar/gkae1061
40. Hammond JA, Gordon EA, Socarras KM et al. Beyond the pan-genome: current perspectives on the

- functional and practical outcomes of the distributed genome hypothesis. *Biochem Soc Trans.* 2020. V.48(6). P.2437-2455. doi: 10.1042/BST20190713
41. Hardigan M.A., Crisovan E., Hamilton J.P. et al. Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated *Solanum tuberosum*. *Plant Cell.* 2016. V. 28(2). P. 388–405. doi: 10.1105/tpc.15.00538
42. He Q., Tang S., Zhi H. et al. A graph-based genome and pan-genome variation of the model plant *Setaria*. *Nature Genetics.* 2023. V. 55(7). P. 1232–1242. doi: 10.1038/s41588-023-01423-w
43. He W, Li X, Qian Q, Shang L. The developments and prospects of plant super-pangenomes: Demands, approaches, and applications. *Plant Commun.* 2025. V.6(2). 101230. doi: 10.1016/j.xplc.2024.101230
44. He X., Qi Z., Liu Z. et al. Pangenome analysis reveals transposon-driven genome evolution in cotton. *BMC Biology.* 2024. V. 22(1). 92. doi: 10.1186/s12915-024-01893-2
45. He Z., Ji R., Havlickova L. et al. Genome structural evolution in *Brassica* crops. *Nature Plants.* 2021. V. 7(6). P. 757–765. doi: 10.1038/s41477-021-00928-8
46. Hirsch C.N., Foerster J.M., Johnson J.M. et al. Insights into the maize pan-genome and pan-transcriptome. *Plant Cell.* 2014. V. 26(1). P. 121–135. doi: 10.1105/tpc.113.1199821
47. Hoopes G., Meng X., Hamilton J.P. et al. Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. *Molecular Plant.* 2022. V. 15(3). P. 520–536. doi: 10.1016/j.molp.2022.01.003
48. Hou Y, Gan J, Fan Z et al. Haplotype-based pangenomes reveal genetic variations and climate adaptations in moso bamboo populations. *Nat Commun.* 2024. V.15(1). 8085. doi: 10.1038/s41467-024-52376-5
49. Hu G., Cheng L., Cheng Y. et al. Pan-genome analysis of three main Chinese chestnut varieties. *Frontiers in Plant Science.* 2022. V. 13. 916550. doi: 10.3389/fpls.2022.916550
50. Hu H, Li R, Zhao J et al. Technological Development and Advances for Constructing and Analyzing Plant Pangenomes. *Genome Biol Evol.* 2024. V.16(4). evae081. doi: 10.1093/gbe/evae081
51. Hu H, Zhao J, Thomas WJW et al. The role of pangenomics in orphan crop improvement. *Nat Commun.* 2025. V.16(1). 118. doi: 10.1038/s41467-024-55260-4
52. Huang Y., He J., Xu Y. et al. Pangenome analysis provides insight into the evolution of the orange subfamily and a key gene for citric acid accumulation in citrus fruits. *Nature Genetics.* 2023. V. 55(11). P. 1964–1975. doi: 10.1038/s41588-023-01516-6
53. Hübner S., Bercovich N., Todesco M. et al. Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nature Plants.* 2019. V. 5(1). P. 54–62. doi: 10.1038/s41477-018-0329-0
54. Huff M, Hulse-Kemp AM, Scheffler BE et al. Long-read, chromosome-scale assembly of *Vitis rotundifolia* cv. Carlos and its unique resistance to *Xylella fastidiosa* subsp. *fastidiosa*. *BMC Genomics.* 2023. V.24(1). 409. doi: 10.1186/s12864-023-09514-y
55. Hufnagel B., Soriano A., Taylor J. et al. Pangenome of white lupin provides insights into the diversity of the species. *Plant Biotechnology Journal.* 2021. V. 19(12). P. 2532–2543. doi: 10.1111/pbi.13678
56. Hufford M.B., Seetharam A.S., Woodhouse M.R. et al. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science.* 2021. V. 373(6555). P. 655–662. doi: 10.1126/science.abg5289
57. Hurgobin B., Golicz A.A., Bayer P.E. et al. Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnology Journal.* 2018. V. 16(7). P. 1265–1274. doi: 10.1111/pbi.12867
58. Jaggi K.E., Krak K., Štorchová H. et al. A pangenome reveals LTR repeat dynamics as a major driver of genome evolution in *Chenopodium*. *Plant Genome.* 2025. V. 18(1). e70010. doi: 10.1002/tpg2.70010
59. Jayakodi M., Lu Q., Pidon H. et al. Structural variation in the pangenome of wild and domesticated barley. *Nature.* 2024. V. 636(8043). P. 654–662. doi: 10.1038/s41586-024-08187-1
60. Jayakodi M., Padmarasu S., Haberer G. et al. The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature.* 2020. V. 588(7837). P. 284–289. doi: 10.1038/s41586-020-2947-8
61. Ji G., Long Y., Cai G. et al. A new chromosome-scale genome of wild *Brassica oleracea* provides insights into the domestication of *Brassica* crops. *Journal of Experimental Botany.* 2024. V.75(10). P. 2882–2899. doi: 10.1093/jxb/erae079
62. Jin S., Han Z., Hu Y. et al. Structural variation (SV)-based pan-genome and GWAS reveal the impacts of SVs on the speciation and diversification of allotetraploid cottons. *Molecular Plant.* 2023. V. 16(4). P. 678–693. doi: 10.1016/j.molp.2023.02.004
63. Jonkheer EM, de Ridder D, van der Lee TAJ et al. Exploring intra- and intergenomic variation in haplotype-resolved pangenomes. *Plant Biotechnol J.* 2025. V.23(3). P.874-886. doi: 10.1111/pbi.14545
64. Kang M., Wu H., Liu H. et al. The pan-genome and local adaptation of *Arabidopsis thaliana*. *Nature Communications.* 2023. V. 14. 6259. doi: 10.1038/s41467-023-42029-4

65. Karetnikov D.I., Vasiliev G.V., Toshchakov S.V. et al. Analysis of genome structure and its variations in potato cultivars grown in Russia. *International Journal of Molecular Sciences*. 2023. V. 24(6). 5713. doi: 10.3390/ijms24065713
66. Khan A.W., Garg V., Roorkiwal M., et al. Super-Pangenome by Integrating the Wild Side of a Species for Accelerated Crop Improvement. *Trends Plant Sci*. 2020. V.25(2). P.148-158. doi: 10.1016/j.tplants.2019.10.012
67. Khan A.W., Garg V., Sun S. et al. *Cicer* super-pangenome provides insights into species evolution and agronomic trait loci for crop improvement in chickpea. *Nature Genetics*. 2024. V. 56(6). P. 1225–1234. doi: 10.1038/s41588-024-01760-4
68. Kaur H, Shannon LM, Samac DA. A stepwise guide for pangenome development in crop plants: an alfalfa (*Medicago sativa*) case study. *BMC Genomics*. 2024. V.25(1). 1022. doi: 10.1186/s12864-024-10931-w
69. Lai J., Li R., Xu X. et al. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nature Genetics*. 2010. V. 42(11). P. 1027–1030. doi: 10.1038/ng.684
70. Lam H.M., Xu X., Liu X. et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature Genetics*. 2010. V. 42(12). P. 1053–1059. doi: 10.1038/ng.715
71. Lei L, Goltsman E, Goodstein D et al. Plant Pan-Genomics Comes of Age. *Annu Rev Plant Biol*. 2021. V.72. P.411-435. doi: 10.1146/annurev-arplant-080720-105454
72. Li J., Liu Z., You C. et al. Convergence and divergence of diploid and tetraploid cotton genomes. *Nature Genetics*. 2024. V. 56(11). P. 2562–2573. doi: 10.1038/s41588-024-01964-8
73. Li J., Yuan D., Wang P. et al. Cotton pangenome retrieves the lost sequences and genes during domestication and selection. *Genome Biology*. 2021. V. 22(1). 119. doi: 10.1186/s13059-021-02351-w
74. Li H., Wang S., Chai S. et al. Graph-based pangenome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. *Nature Communications*. 2022. V. 13(1). 682. doi: 10.1038/s41467-022-28362-0
75. Li N., He Q., Wang J. et al. Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nature Genetics*. 2023. V. 55(5). P. 852–860. doi: 10.1038/s41588-023-01340-y
76. Li Q, Qiao X, Li L et al. Haplotype-resolved T2T genome assemblies and pangenome graph of pear reveal diverse patterns of allele-specific expression and the genomic basis of fruit quality traits. *Plant Commun*. 2024a. V.5(10). 101000. doi: 10.1016/j.xplc.2024.101000
77. Li Y.H., Zhou G., Ma J. et al. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology*. 2014. V. 32(10). P. 1045–1052. doi: 10.1038/nbt.2979
78. Li X., Wang Y., Cai C. et al. Large-scale gene expression alterations introduced by structural variation drive morphotype diversification in *Brassica oleracea*. *Nature Genetics*. 2024b. V. 56(3). P. 517–529. doi: 10.1038/s41588-024-01655-4
79. Liang Q., Muñoz-Amatriain M., Shu S. et al. A view of the pan-genome of domesticated Cowpea (*Vigna unguiculata* [L.] Walp.). *Plant Genome*. 2024. V. 17(1). e20319. doi: 10.1002/tpg2.20319
80. Lin K., Zhang N., Severing E.I. et al. Beyond genomic variation--comparison and functional annotation of three *Brassica rapa* genomes: a turnip, a rapid cycling and a Chinese cabbage. *BMC Genomics*. 2014. V. 15(1). 250. doi: 10.1186/1471-2164-15-250
81. Liu C., Peng P., Li W. et al. Deciphering variation of 239 elite japonica rice genomes for whole genome sequences-enabled breeding. *Genomics*. 2021. V. 113(5). P. 3083–3091. doi: 10.1016/j.ygeno.2021.07.002
82. Liu C., Wang Y., Peng J. et al. High-quality genome assembly and pan-genome studies facilitate genetic discovery in mung bean and its improvement. *Plant Communications*. 2022. V. 3(6). 100352. doi: 10.1016/j.xplc.2022.100352
83. Liu H., Wang X., Liu S. et al. *Citrus* Pan-Genome to Breeding Database (CPBD): A comprehensive genome database for citrus breeding. *Molecular Plant*. 2022a. V. 15(10). P. 1503–1505. doi: 10.1016/j.molp.2022.08.006
84. Liu J.N., Yan L., Chai Z. et al. Pan-genome analyses of 11 *Fraxinus* species provide insights into salt adaptation in ash trees. *Plant Communications*. 2025. V. 6(1). 101137. doi: 10.1016/j.xplc.2024.101137
85. Liu Y., Du H., Li P. et al. Pan-genome of wild and cultivated soybeans. *Cell*. 2020. V. 182(1). P. 162–176.e13. doi: 10.1016/j.cell.2020.05.023
86. Liu Z., Wang N., Su Y. et al. Grapevine pangenome facilitates trait genetics and genomic breeding. *Nature Genetics*. 2024. V. 56(12). P. 2804–2814. doi: 10.1038/s41588-024-01967-5
87. Luo Y, Liu Z, Jin Z et al. Phased T2T genome assemblies facilitate the mining of disease-resistance genes in *Vitis davidii*. *Hortic Res*. 2024. V.12(2). uhae306. doi: 10.1093/hr/uhae306
88. Long W., He Q., Wang Y. et al. Genome evolution and diversity of wild and cultivated rice species. *Nature Communications*. 2024. V. 15(1). 9994. doi: 10.1038/s41467-024-54427-3
89. Lovell JT, Bentley NB, Bhattarai G et al. Four chromosome scale genomes and a pan-genome annotation to accelerate pecan tree breeding. *Nat*

- Commun.* 2021. V.12(1). 4125. doi: 10.1038/s41467-021-24328-w
90. Lu F., Romay M.C., Glaubitz J.C. et al. High-resolution genetic mapping of maize pan-genome sequence anchors. *Nature Communications*. 2015. V.6. 6914. doi: 10.1038/ncomms7914
91. Lv Y, Liu C, Li X et al. A centromere map based on super pan-genome highlights the structure and function of rice centromeres. *J Integr Plant Biol.* 2024. V.66(2). P.196-207. doi: 10.1111/jipb.13607
92. Lyu X., Xia Y., Wang C. et al. Pan-genome analysis sheds light on structural variation-based dissection of agronomic traits in melon crops. *Plant Physiology.* 2023. V. 193(2). P. 1330–1348. doi: 10.1093/plphys/kiad405
93. MacNish T.R., Al-Mamun H.A., Bayer P.E. et al. *Brassica* Panache: A multi-species graph pangenome representing presence absence variation across forty-one *Brassica* genomes. *Plant Genome.* 2025. V. 18(1). e20535. doi: 10.1002/tpg2.20535
94. Miao H., Wang L., Qu L. et al. Genomic evolution and insights into agronomic trait innovations of *Sesamum* species. *Plant Communications.* 2024. V. 8. 5(1). 100729. doi: 10.1016/j.xplc.2023.100729
95. Michael TP, VanBuren R. Building near-complete plant genomes. *Curr Opin Plant Biol.* 2020. V.54. P.26-33. doi: 10.1016/j.pbi.2019.12.009
96. Monat C., Pera B., Ndjiondjop M.N. et al. *De novo* assemblies of three *Oryza glaberrima* accessions provide first insights about pan-genome of african rices. *Genome Biology Evolution.* 2017. V. 9(1). P. 1–6. doi: 10.1093/gbe/evw253
97. Montenegro J.D., Golicz A.A., Bayer P.E. et al. The pangenome of hexaploid bread wheat. *Plant Journal.* 2017. V. 90(5). P. 1007–1013. doi: 10.1111/tpj.13515
98. Morgante M., De Paoli E., Radovic S. Transposable elements and the plant pan-genomes. *Current Opinion in Plant Biology.* 2007. V. 10(2). P. 149–155. doi: 10.1016/j.pbi.2007.02.001
99. Naithani S, Deng CH, Sahu SK, Jaiswal P. Exploring Pan-Genomes: An Overview of Resources and Tools for Unraveling Structure, Function, and Evolution of Crop Genes and Genomes. *Biomolecules.* 2023. V.13(9). 1403. doi: 10.3390/biom13091403
100. Nawae W., Naktang C., Charoensri S. et al. Resequencing of durian genomes reveals large genetic variations among different cultivars. *Frontiers in Plant Science.* 2023. V. 14. 1137077. doi: 10.3389/fpls.2023
101. Niu Y., Liu Q., He Z. et al. A *Brassica carinata* pan-genome platform for *Brassica* crop improvement. *Plant Communications.* 2024. V. 5(1). 100725. doi: 10.1016/j.xplc.2023.100725
102. Oren E., Dafna A., Tzuri G. et al. Pan-genome and multi-parental framework for high-resolution trait dissection in melon (*Cucumis melo*). *Plant Journal.* 2022. V. 112(6). P. 1525–1542. doi: 10.1111/tpj.16021
103. Ou L., Li D., Lv J. et al. Pan-genome of cultivated pepper (*Capsicum*) and its use in gene presence-absence variation analyses. *New Phytologist.* 2018. V. 220(2). 360–363. doi: 10.1111/nph.15413
104. Petereit J, Bayer PE, Thomas WJW et al. Pangenomics and Crop Genome Adaptation in a Changing Climate. *Plants (Basel).* 2022. V.11(15). 1949. doi: 10.3390/plants11151949
105. Pinosio S., Giacomello S., Faivre-Rampant P. et al. Characterization of the poplar pan-genome by genome-wide identification of structural variation. *Molecular Biology Evolution.* 2016. V. 33(10). P. 2706–2719. doi: 10.1093/molbev/msw161
106. Portwood J.L. 2nd, Woodhouse M.R., Cannon E.K. et al. MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Reserch.* 2019. V. 47(D1). D1146–D1154. doi: 10.1093/nar/gky1046
107. Qiao Q., Edger P.P., Xue L. et al. Evolutionary history and pan-genome dynamics of strawberry (*Fragaria* spp.). *Proc Natl Acad Sci USA.* 2021. V. 118(45). e2105431118. doi: 10.1073/pnas.2105431118
108. Qin P., Lu H., Du H. et al. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell.* 2021. V. 184(13). P. 3542–3558.e16. doi: 10.1016/j.cell.2021.04.046
109. Raza A, Bohra A, Garg V, Varshney RK. Back to wild relatives for future breeding through super-pangenome. *Mol Plant.* 2023. V.16(9). P.1363-1365. doi: 10.1016/j.molp.2023.08.005
110. Rijzaani H., Bayer P.E., Rouard M. et al. The pangenome of banana highlights differences between genera and genomes. *Plant Genome.* 2022. 15(1):e20100. doi: 10.1002/tpg2.20100
111. Ruperao P, Rangan P, Shah T et al. Developing pangenomes for large and complex plant genomes and their representation formats. *J Adv Res.* 2025. V.S2090-1232(25)00071-2. doi: 10.1016/j.jare.2025.01.052
112. Ruperao P., Thirunavukkarasu N., Gandham P. et al. *Sorghum* pan-genome explores the functional utility for genomic-assisted breeding to accelerate the genetic gain. *Frontiers in Plant Science.* 2021. V. 12. 666342. doi: 10.3389/fpls.2021.666342
113. Sarashetti P., Lipovac J., Tomas F. et al. Evaluating data requirements for high-quality haplotype-resolved genomes for creating robust pangenome references. *Genome Biology.* 2024. V. 25(1). 312. doi: 10.1186/s13059-024-03452-y
114. Schatz M.C., Maron L.G., Stein J.C. et al. Whole genome *de novo* assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of aus and indica. *Genome Biology.* 2014. V. 15(11). 506. doi: 10.1186/PREACCEPT-2784872521277375

115. Schneeberger K., Ossowski S., Ott F. et al. Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. *Proc Natl Acad Sci USA*. 2011. V. 108(25). P.10249–10254. doi: 10.1073/pnas.1107739108
116. She H, Liu Z, Xu Z et al. Pan-genome analysis of 13 *Spinacia* accessions reveals structural variations associated with sex chromosome evolution and domestication traits in spinach. *Plant Biotechnol J*. 2024. V.22(11). P.3102–3117. doi: 10.1111/pbi.14433
117. Shi J, Tian Z, Lai J, Huang X. Plant pangenomics and its applications. *Mol Plant*. 2023. V.16(1). P.168–186. doi: 10.1016/j.molp.2022.12.009
118. Shi T., Zhang X., Hou Y. et al. The super-pangenome of *Populus* unveils genomic facets for its adaptation and diversification in widespread forest trees. *Molecular Plant*. 2024. V. 17(5). P. 725–746. doi: 10.1016/j.molp.2024.03.009
119. Shang L., Li X., He H. et al. A super pangenomic landscape of rice. *Cell Res*. 2022. V. 32(10). P. 878–896. doi: 10.1038/s41422-022-00685-z
120. Shen J-S., Lan L., Kan S-L. et al. Haplotype-resolved genome for *Rhododendron × pulchrum* and the expression analysis of heat shock genes. *Journal of Systematics and Evolution*. 2024. V. 62(3). P. 489–504. doi: 10.1111/jse.13007
121. Sigaux F. Genome du cancer ou de la construction des cartes d'identité moléculaire des tumeurs. *Bulletin de l'Académie Nationale de Médecine*. 2000. V.184(7). P.1441–1449, including discussion 1448–1449. [Cancer genome or the development of molecular portraits of tumors] (In French)
122. Sinha P., Singh V.K., Saxena R.K. et al. Superior haplotypes for haplotype-based breeding for drought tolerance in pigeonpea (*Cajanus cajan* L.). *Plant Biotechnology Journal*. 2020. V. 18(12). P. 2482–2490. doi: 10.1111/pbi.13422
123. Sivabharathi R.C., Rajagopalan V.R., Suresh R. et al. Haplotype-based breeding: A new insight in crop improvement. *Plant Science*. 2024. V. 346. P. 112129. doi: 10.1016/j.plantsci.2024.112129
124. Song J.M., Guan Z., Hu J. et al. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nature Plants*. 2020. V. 6(1). P. 34–45. doi: 10.1038/s41477-019-0577-7
125. Song J.M., Liu D.X., Xie W.Z. et al. BnPIR: *Brassica napus* pan-genome information resource for 1689 accessions. *Plant Biotechnology Journal*. 2021. V. 19(3). P. 412–414. doi: 10.1111/pbi.13491
126. Song Y., Han S., Wang M. et al. Pangenome identification and analysis of terpene synthase gene family members in *Gossypium*. *International Journal of Molecular Sciences*. 2024. V. 25(17). P. 9677. doi: 10.3390/ijms25179677
127. Su Y, Yang X, Wang Y et al. Phased telomere-to-telomere reference genome and pangenome reveal an expansion of resistance genes during apple domestication. *Plant Physiol*. 2024. V.195(4). P.2799–2814. doi: 10.1093/plphys/kiac258
128. Sun C., Hu Z., Zheng T. et al. RPN: rice pangenome browser for ~3000 rice genomes. *Nucleic Acids Res*. 2017. V. 45(2). P. 597–605. doi: 10.1093/nar/gkw958
129. Sun S., Zhou Y., Chen J. et al. Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nature Genetics*. 2018. V. 50(9). P. 1289–1295. doi: 10.1038/s41588-018-0182-0
130. Sun X, Jiao C, Schwaninger H et al. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat Genet*. 2020. V.52(12). P.1423–1432. doi: 10.1038/s41588-020-00723-9
131. Sun Y., Wang J., Li Y. et al. Pan-genome analysis reveals the abundant gene presence/absence variations among different varieties of melon and their influence on traits. *Frontiers in Plant Science*. 2022. V. 13. P. 835496. doi: 10.3389/fpls.2022.835496
132. Sun Y., Kou D.R., Li Y. et al. Pan-genome of *Citrullus* genus highlights the extent of presence/absence variation during domestication and selection. *BMC Genomics*. 2023. V. 24(1). P. 332. doi: 10.1186/s12864-023-09443-w
133. Tahir U., Qamar M., Fatima K. et al. Comparative genomics profiling of *Citrus* species reveals the diversity and disease responsiveness of the GLP pangenome family. *BMC Plant Biology*. 2025. V. 25(1). P. 388. doi: 10.1186/s12870-025-06397-x
134. Tan W, Zhou P, Huang X et al. Haplotype-resolved genome of *Prunus zhengheensis* provides insight into its evolution and low temperature adaptation in apricot. *Hortic Res*. 2024. V.11(4). P. uhae103. doi: 10.1093/hr/uhae103
135. Tang D., Jia Y., Zhang J. et al. Genome evolution and diversity of wild and cultivated potatoes. *Nature*. 2022. V. 606. P. 535–541. doi: 10.1038/s41586-022-04822-x
136. Tao Y., Luo H., Xu J. et al. Extensive variation within the pan-genome of cultivated and wild sorghum. *Nature Plants*. 2021. V. 7(6). P. 766–773. doi: 10.1038/s41477-021-00925-x
137. Tariq A., Meng M., Jiang X. et al. In-depth exploration of the genomic diversity in tea varieties based on a newly constructed pangenome of *Camellia sinensis*. *Plant Journal*. 2024. V. 119(4). P. 2096–2115. doi: 10.1111/tbj.16874
138. Tay Fernandez CG, Nestor BJ, Danilevicz MF et al. Expanding Gene-Editing Potential in Crop Improvement with Pangenomes. *Int J Mol Sci*. 2022. V.23(4). P. 2276. doi: 10.3390/ijms23042276

139. Tets V.V. Pangenom. *Tsitologiya*. 2003. V.45(5). P.526-531. (In Russian)
140. Tettelin H., Masignani V., Cieslewicz M.J. et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci USA*. 2005. V.102(39). P.13950-13955. doi: 10.1073/pnas.0506758102
141. Torkamaneh D., Lemay M.A., Belzile F. The pan-genome of the cultivated soybean (PanSoy) reveals an extraordinarily conserved gene content. *Plant Biotechnology Journal*. 2021. V. 19(9). P. 1852–1862. doi: 10.1111/pbi.13600
142. van Workum D.M., Mehrem S.L., Snoek B.L. et al. *Lactuca* super-pangenome reduces bias towards reference genes in lettuce research. *BMC Plant Biology*. 2024. V. 24(1). 1019. doi: 10.1186/s12870-024-05712-2
143. Varshney R.K., Graner A., Sorrells M.E. Genomics-assisted breeding for crop improvement. *Trends in Plant Science*. 2005. V. 10(12). P. 621–630. doi: 10.1016/j.tplants.2005.10.004
144. Varshney R.K., Roorkiwal M., Sun S. et al. A chickpea genetic variation map based on the sequencing of 3,366 genomes. *Nature*. 2021. V. 599(7886). P. 622–627. doi: 10.1038/s41586-021-04066-1.
145. Varshney R.K., Sinha P., Singh V.K. et al. 5Gs for crop genetic improvement. *Current Opinion in Plant Biology*. 2020. V. 56. P. 190–196. doi: 10.1016/j.pbi.2019.12.004
146. Vaughn J.N., Branham S.E., Abernathy B. et al. Graph-based pangenomics maximizes genotyping density and reveals structural impacts on fungal resistance in melon. *Nature Communications*. 2022. V. 13(1). 7897. doi: 10.1038/s41467-022-35621-7
147. Vernikos G, Medini D, Riley DR, Tettelin H. Ten years of pan-genome analyses. *Curr Opin Microbiol*. 2015. V.23. P.148-154. doi: 10.1016/j.mib.2014.11.016
148. Voelker W.G., Krishnan K., Chougule K. et al. Ten new high-quality genome assemblies for diverse bioenergy sorghum genotypes. *Frontiers in Plant Science*. 2023. V. 13. 1040909. doi: 10.3389/fpls.2022.1040909
149. Wang M., Li J., Qi Z. et al. Genomic innovation and regulatory rewiring during evolution of the cotton genus *Gossypium*. *Nature Genetics*. 2022. V. 54(12). P. 1959–1971. doi: 10.1038/s41588-022-01237-2
150. Wang J., Yang W., Zhang S. et al. A pangenome analysis pipeline provides insights into functional gene identification in rice. *Genome Biology*. 2023. 24(1). 19. doi: 10.1186/s13059-023-02861-9
151. Wang T., Duan S., Xu C. et al. Pan-genome analysis of 13 *Malus* accessions reveals structural and sequence variations associated with fruit traits. *Nature Communications*. 2023a. V. 14(1). 7377. doi: 10.1038/s41467-023-43270-7
152. Wang W., Mauleon R., Hu Z. et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*. 2018. 557(7703). P. 43–49. doi: 10.1038/s41586-018-0063-9
153. Wang X, Zhou P, Hu X et al. T2T genome, pan-genome analysis, and heat stress response genes in *Rhododendron* species. *Imeta*. 2025. V.4(2). e70010. doi: 10.1002/imt2.70010
154. Wang Y., Li P., Zhu Y. et al. Graph-based pangenome of *Actinidia chinensis* reveals structural variations mediating fruit degreening. *Advanced Science (Weinh)*. 2024. V. 11(28). e2400322. doi: 10.1002/advs.202400322
155. Woodhouse M.R., Cannon E.K., Portwood J.L. 2nd et al. A pan-genomic approach to genome databases using maize as a model system. *BMC Plant Biology*. 2021. V. 21(1). 385. doi: 10.1186/s12870-021-03173-5
156. Woodhouse M.R., Cannon E.K., Portwood J.L. 2nd et al. Tools and resources at the maize genetics and genomics database (MaizeGDB). *Cold Spring Harbor Protocols*. 2025. V. 2025(1). pdb.over108430. doi: 10.1101/pdb.over108430
157. Wu D., Xie L., Sun Y. et al. A syntelog-based pan-genome provides insights into rice domestication and de-domestication. *Genome Biology*. 2023. V. 24(1). 179. doi: 10.1186/s13059-023-03017-5
158. Wu J., Xu X.D., Liu L. et al. A Chromosome level genome assembly of a winter turnip rape (*Brassica rapa* L.) to explore the genetic basis of cold tolerance. *Frontiers in Plant Science*. 2022. V. 13. 936958. doi: 10.3389/fpls.2022.936958
159. Wu S., Sun H., Gao L. et al. A *Citrullus* genus super-pangenome reveals extensive variations in wild and cultivated watermelons and sheds light on watermelon evolution and domestication. *Plant Biotechnology Journal*. 2023a. V. 21(10). P. 1926–1928. doi: 10.1111/pbi.14120
160. Yan H., Sun M., Zhang Z. et al. Pangenomic analysis identifies structural variation associated with heat tolerance in pearl millet. *Nature Genetics*. 2023. V. 55(3). P. 507–518. doi: 10.1038/s41588-023-01302-4
161. Yang T., Liu R., Luo Y. et al. Improved pea reference genome and pan-genome highlight genomic features and evolutionary characteristics. *Nature Genetics*. 2022. V. 54(10). P. 1553–1563. doi: 10.1038/s41588-022-01172-2
162. Yano R., Li F., Hiraga S. et al. The genomic landscape of gene-level structural variations in Japanese and global soybean *Glycine max* cultivars. *Nature Genetics*. 2025. V. 57. P. 973–985. doi: 10.1038/s41588-025-02113-5
163. Yu J., Golicz A.A., Lu K. et al. Insight into the evolution and functional characteristics of the pan-genome assembly from sesame landraces and modern

- cultivars. *Plant Biotechnology Journal*. 2019. V. 17(5). P. 881–892. doi: 10.1111/pbi.13022
164. Yu X., Qu M., Wu P. et al. Super pan-genome reveals extensive genomic variations associated with phenotypic divergence in *Actinidia*. *Molecular Horticulture*. 2025. V. 5(1). 4. doi: 10.1186/s43897-024-00123-1
165. Yu Z., Chen Y., Zhou Y. et al. Rice gene index: A comprehensive pan-genome database for comparative and functional genomics of Asian rice. *Molecular Plant*. 2023. V. 16(5). P. 798–801. doi: 10.1016/j.molp.2023.03.012
166. Zhang B., Zhu W., Diao S. et al. The poplar pangenome provides insights into the evolutionary history of the genus. *Communications Biology*. 2019. V. 2. 215. doi: 10.1038/s42003-019-0474-7
167. Zhang F., Xue H., Dong X. et al. Long-read sequencing of 111 rice genomes reveals significantly larger pan-genomes. *Genome Reserch*. 2022. V. 32(5). P. 853–863. doi: 10.1101/gr.276015.121
168. Zhang L., Liu Y., Huang Y. et al. Solanaceae pan-genomes reveal extensive fractionation and functional innovation of duplicated genes. *Plant Communications*. 2025. V. 6(3). 101231. doi: 10.1016/j.xplc.2024.101231
169. Zhang R., Dai C., Gong R. et al. Gapless genome assembly and pan-genome of *Brassica juncea* provide insights into seed quality improvement and environmental adaptation. *Plant Communications*. 2025a. 101298. doi: 10.1016/j.xplc.2025.101298
170. Zhang X., Liu T., Wang J. et al. Pan-genome of *Raphanus* highlights genetic variation and introgression among domesticated, wild, and weedy radishes. *Molecular Plant*. 2021. V. 14(12). P. 2032–2055. doi: 10.1016/j.molp.2021.08.005
171. Zhang X., Chen Y., Wang L. et al. Pangenome of water caltrop reveals structural variations and asymmetric subgenome divergence after allopolyploidization. *Horticulture Research*. 2023. V. 10(11). uhad203. doi: 10.1093/hr/uhad203
172. Zhang Y., Zhao M., Tan J. et al. Telomere-to-telomere *Citrullus* super-pangenome provides direction for watermelon breeding. *Nature Genetics*. 2024. V. 56(8). P. 1750–1761. doi: 10.1038/s41588-024-01823-6
173. Zhao J., Bayer P.E., Ruperao P. et al. Trait associations in the pangenome of pigeon pea (*Cajanus cajan*). *Plant Biotechnology Journal*. 2020. V. 18(9). P. 1946–1954. doi: 10.1111/pbi.13354
174. Zhao Q., Feng Q., Lu H. et al. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nature Genetics*. 2018. V. 50. P. 278–284. doi: 10.1038/s41588-018-0041-z
175. Zhou L., Wu S., Chen Y. et al. Multi-omics analyzes of *Rosa gigantea* illuminate tea scent biosynthesis and release mechanisms. *Nat Commun*. 2024. V.15(1). 8469. doi: 10.1038/s41467-024-52782-9
176. Zhou P., Silverstein K.A., Ramaraj T. et al. Exploring structural variation and gene family architecture with *de novo* assemblies of 15 *Medicago* genomes. *BMC Genomics*. 2017. V. 18(1). 261. doi: 10.1186/s12864-017-3654-1
177. Zhou Y., Chebotarov D., Kudrna D. et al. A platinum standard pan-genome resource that represents the population structure of Asian rice. *Scientific Data*. 2020. V. 7(1). 113. doi: 10.1038/s41597-020-0438-2
178. Zhou Y., Yu Z., Chebotarov D. et al. Pan-genome inversion index reveals evolutionary insights into the subpopulation structure of Asian rice. *Nature Communications*. 2023. V. 14(1). 1567. doi: 10.1038/s41467-023-37004-y
179. Zhou Y., Zhang Z., Bao Z. et al. Graph pangenome captures missing heritability and empowers tomato breeding. *Nature*. 2022. V. 606(7914). P. 527–534. doi: 10.1038/s41586-022-04808-9
180. Zhu X., Yang R., Liang Q. et al. Graph-based pangenome provides insights into structural variations and genetic basis of metabolic traits in potato. *Molecular Plant*. 2025. S1674-2052(25)00038-3. doi: 10.1016/j.molp.2025.01.017
181. Zhuang Y., Wang X., Li X. et al. Phylogenomics of the genus *Glycine* sheds light on polyploid evolution and life-strategy transition. *Nature Plants*. 2022. V. 8(3). P. 233–244. doi: 10.1038/s41477-022-01102-4