

Legume pangenome: Status and scope for crop improvement

Uday Chand Jha¹, Harsh Nayyar², Eric J.B.von Wettberg³, Yogesh Dasharath Naik⁴, Mahender Thudi^{4,5,6}, Kadambot H.M. Siddique⁷ *

¹ Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

² Department of Botany, Panjab University, Chandigarh, India

³ Department and Plant and Soil Science and Gund Institute for the Environment, The University of Vermont

⁴ Dr. Rajendra Prasad Central Agricultural University, Pusa, Bihar, India

⁵ Shandong Academy of Agricultural Sciences, Jinan, Shandong, China

⁶ University of Southern Queensland, Toowoomba, Queensland, Australia

⁷ The UWA Institute of Agriculture, The University of Western Australia, Perth, WA 6001, Australia

* Correspondence: kadambot.siddique@uwa.edu.au

Abstract: In the last decade, legume genomics research has seen a paradigm shift due to advances in genome sequencing technologies, assembly algorithms, and computational genomics that enabled the construction of high-quality reference genome assemblies of major legume crops. These advances have certainly facilitated the identification of novel genetic variants underlying the traits of agronomic importance in many legume crops. Furthermore, these robust sequencing technologies have allowed us to study structural variations across the whole genome in multiple individuals and at the species level using ‘pangenome analysis.’ This review updates the progress of constructing pangenome assemblies for various legume crops and discusses the prospects for these pangenomes and how to harness the information to improve various traits of economic importance through molecular breeding to increase genetic gain in legumes and tackle the increasing global food crisis.

Keywords: pangenome; legumes; structural variation; genome sequence; gene; crop domestication

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Plants* **2022**, *11*, x. <https://doi.org/10.3390/xxxxx>

Academic Editor: Firstname Lastname

Received: date

Accepted: date

Published: date

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Global climate change causing significant crop yield losses, burgeoning human population growth, and increasing evidence of malnutrition and hunger-related issues are serious challenges worldwide [1,2]. Novel scientific interventions are urgently needed to minimize these rising challenges and ensure global food security by developing climate-resilient food crops [3]. Advances in genomics resources, such as molecular markers, have greatly expanded the plant breeding tools for improving various plant traits to meet the global human food demand [3,4]. The paradigm shift and technological advances in DNA sequencing technology, especially next-generation sequencing (NGS), have aided the construction of accurate reference genome assemblies for various legume crops, allowing the identification of underlying target trait candidate gene(s), structural variations, and genetic diversity and insights into domestication and evolutionary history [5–7]. However, a single reference genome assembly is insufficient to capture all the structural variations, nucleotide variations, and gene content information for all individuals related to a particular crop species [8]. Therefore, species-representative genomes or ‘pangenomes’ are needed to capture all of the structural variations (SVs), including presence–absence variations (PAVs), copy number variations (CNVs), and repetitive elements or transposable elements (TEs), across the entire genome of all individuals in a plant species [8–16]. Pangenomes contain ‘core genes’ common to all the individuals of a species and ‘accessory genes’ or ‘variable genes’ present in some but not all individuals [9]. Accessory genes vary more than core genes, enabling plant adaption under diverse environments [17].

Tettelin et al. [9] first proposed the pangenome concept for *Streptococcus agalactiae* bacteria. Subsequently, Morgante et al. [18] reported this concept in plants. Li et al. [17] constructed the first pangenome of soybean to explore the role of various gene(s)/genomic regions contributing to agronomic traits and the domestication process and to track evolutionary history. The pangenomes of various plant species have been constructed, including Arabidopsis [19], rice [20], wheat [21], maize [22], Brachypodium [23,24], sorghum [25], brassica [26,27], sunflower [28], tomato [29], pepper [30], cotton [31], and apple [32], to unravel gene content variation/genome dynamics, domestication, evolution processes and gene(s) lost during selection and breeding activity.

Legume crops play a crucial role in ensuring global food security. Legume reference genomes have been constructed for soybean [33], chickpea [34], common bean [35], pigeon-pea [36], pea [37–38], lupin [39–40], peanut [41], cowpea [42] and mung bean [43]. However, pangenomes are needed to obtain insights into the genome dynamics, gene content variation, genetic basis of agronomic trait variation and evolutionary relationship in various legumes. This review discusses the current status of various pangenomes available in various legume species and how they can be exploited in genomics-assisted trait improvement to increase their genetic gain.

2. Pangenome capturing structural variations not present in single reference genomes

Plant genomes vary between individual genotypes within the same species due to variations in gene content, including CNVs and PAVs, inversions, and the presence of transposable elements (see [44–46]). Several mechanisms cause structural variations in genomes. The replication ability of transposable elements (TEs) allows these TEs to change position across the whole genome and insert to any coding region, functionally disrupting these genes [47–48]. Recently, a pangenome sequence analysis of pea estimated 85.5% deletions and 77.4% duplications in Long Terminal Repeat (LTR) Copia and LTR Gypsy repeat sequences [38]. Thus, these transposable elements cause insertions and deletions in genomes, rendering changes in plant phenotypes [46, 49]. Likewise, unequal recombination or crossover can introduce SVs in genome [50]. The other important mechanism for introducing SVs are segmental genome duplication and whole genome duplication found in case of soybean [33], and peanut [41]. Inversion and transversion and chromosomal translocations are other mechanism accounting genome SVs [6,7,15,51]. Thus, a single reference genome assembly does not capture the entire gene repertoire of a species. In contrast, pangenomes harbor core and dispensable genes encompassing all structural variations in non-coding and coding sequences of all individuals in a particular species (see Fig.1). Pangenome analysis can be performed using three approaches: *de novo* assembly, reference-based assembly and iterative mapping, and graph-based assembly [13, 15, 46, 52].

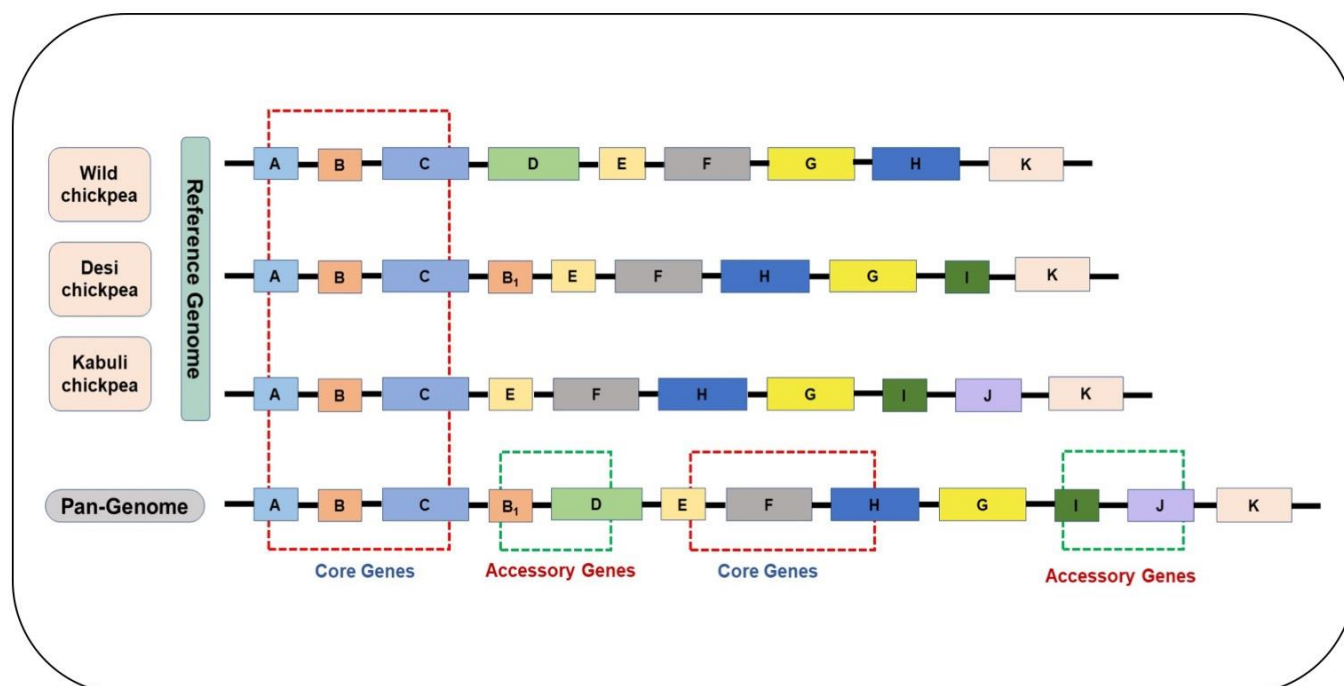


Figure 1. Phylogenetically diverse accession selected for construction of pan-genome that contains core genes and accessory genes: Pan-genome contains all the possible structural variation that may occur during evolution such as B₁: Copy number variation, D: Deletion variation, I and J: Insertional variation and G: Transposable element.

3. *Glycine max* L.

In soybean, Li et al. [17] performed pangenome analysis by constructing a *de novo* genome assembly of seven accessions of *Glycine soja* collected from different geographic regions to shed light on the genetic basis of untapped genetic diversity and evolution-related novel insights. Comparing the genome sequences obtained from this pangenome analysis and a reference genome sequence of *G. max* W82 revealed a plethora of loss or gain of CNVs in multiple genes ranging from biotic stress tolerance (including R genes) to various transcription factors contributing to local environment acclimation against invading pathogens [17]. In addition, comparing the *G. max* and *G. soja* genomes provided novel insights into the *FLOWERING LOCUS T* gene controlling flowering time, *PHYA* regulating light receptor, and *LEAFY* genes via various ‘indels’ and nonsynonymous SNPs in the given loci in the *G. soja* genome [17]. In *Arabidopsis*, 1,332 genes related to seed protein and oil contents were annotated in the *G. max* W82 genome based on homology to genes related to lipid acyl pathways [53]. A comparative genomics analysis revealed that the *G. max* and *G. soja* genomes differ for several indels and large-effect single nucleotide variants (SNPs) and CNVs for these genes attributing to oil seed content [17]. A *de novo* assembly of 26 soybean accessions (including 14 cultivated, 3 wild, and 9 landraces) from 2,898 deeply sequenced soybean genomes and pangenome analysis provided novel insights into the thousands of structural variations related to various candidate genes of agronomic interest (including iron deficiency chlorosis, seed luster, and flowering) and involved in domestication and evolution [54](see Fig.2). The assembled sequences contained 20,623 gene families designated as core genes, with 8,163 gene families in 25–26 accessions designated soft-core genes, and 28,679 families in 2–24 accessions designated dispensable genes [54] [see Table 1]. A comparative analysis of this pangenome with the

previously assembled reference genomes for Zhonghuang [63,64], W05 [65], and Wm82 [33] indicated the presence of myriad PAVs and CNVs in various accessions of current soybean accessions, demanding a thorough pangenome analysis of diverse and geographically distinct soybean accessions [54]. Torkamaneh et al. [16] developed the soybean pangenome known as ‘PanSoy’ by constructing a *de novo* assembly of 204 geographically diverse soybean accessions following the ‘map to pan’ approach using the EUPAN pipeline [66]. This pangenome contained 108 Mbp novel genome sequences—with 1,659 novel genes not present in the reference Wm82 genome—revealing 49,431 hard-core genes, 1,401 soft-core genes, 3,402 shell genes, and 297 ‘cloud genes’ [16]. The PanSoy resource provided novel insights into intraspecific genetic variation in *Glycine max* and offered a platform for genomic diversity and evolutionary and domestication studies in soybean.

Subsequently, a pangenome measuring 1,213 Mbp was constructed from >1,000 accessions collected from the USDA Soybean Germplasm collection using the chromosome level ‘Lee’ soybean genome assembly as the reference sequence [55]. The results identified 3,765 genes missing in the ‘Lee’ genome sequence assembly, indicating a loss of genetic diversity during domestication and breeding activity [55]. This pangenome also contained an additional 198.34 Mbp sequence assembly compared to the ‘Lee’ reference genome. The PAV analysis revealed 86.8% core genes and 13.2% dispensable genes [55]. The authors also uncovered 110 genomic regions containing 1,266 protein coding genes in a ‘domestication-selective sweep’ and 86 genomic regions containing 1,434 genes in a ‘breeding-selective sweep’ [55]. Based on this pangenome analysis, the authors concluded that wild soybean had higher nucleotide diversity than landraces, old cultivars, and modern cultivars due to the loss of diversity during domestication and breeding efforts.

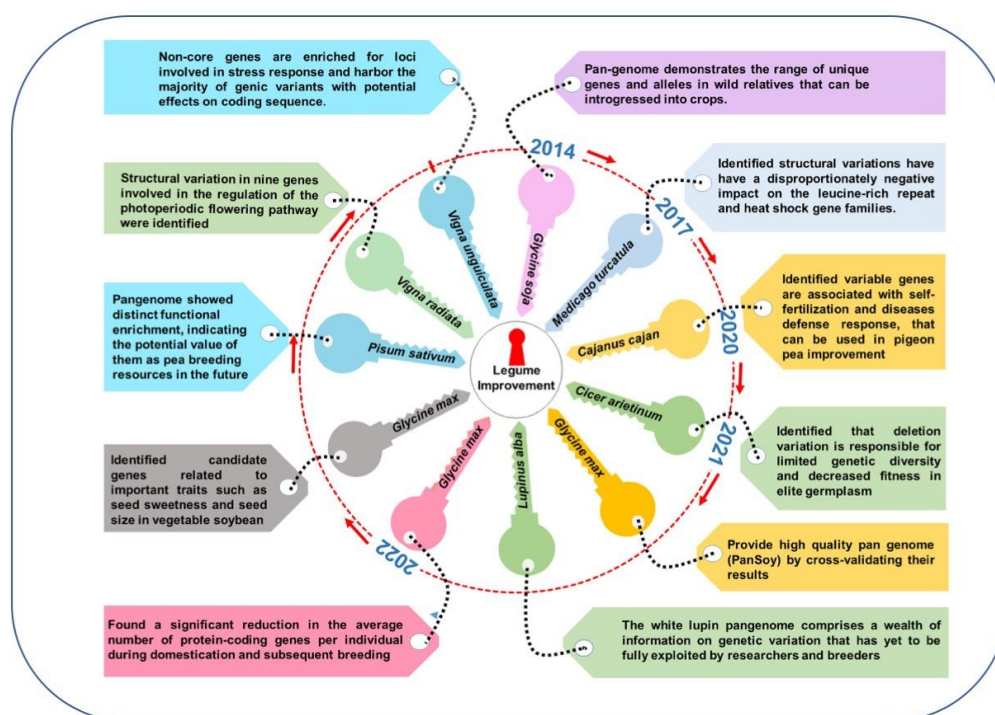


Figure 2. Legume pangenomes and key learnings for crop improvement.

Table 1. List of various pan-genome constructed in various legume crops.

Crop	Size of pangenome	Number of core and dispensable genes	Accessions used	Reference
<i>Glycine soja</i> (wild soybean)	889.33–1,118.34Mb	48.6% of the gene families	7	[17]

		were core genomic units and 51.4% of gene families (30,364) were noncore		
<i>Glycine max</i> (soybean)	108 Mb	49 431 (90.6%) hardcore genes , 7.2% variable genes	204	[16]
<i>Glycine max</i> (soybean)	992.3 Mb to 1059.8 Mb	20,623 core genes, and 28679 dispensable genes	26	[54]
<i>Glycine max</i> (soybean)	1,213 Mbp	86.8% of genes are core and 13.2% are dispensable	1110	[55]
<i>Vigna radiata</i>	762.92 Mb	83.1% were core genes and 16.9% were variable	217	[56]
<i>Vigna unguiculata</i>	449.91 Mb	21,330 core and 23,531 non-core genes	6	[57]
<i>Cajanus cajan</i>	622 Mbp	48 067 (86.6%) core genes and 7445 (13.4%) were noncore	89	[58]
<i>Cicer arietinum</i>	592.58 Mbp	-	3366	[59]
<i>Medicago turcatula</i>	388 Mbp to 428 Mbp	250Mbp core sequence and 180 Mbp is dispensable	15	[60]
<i>Lupinus alba</i>	462.7Mbp	sequence 32 068 core genes, 14 822 non-core genes	39	[61]
<i>Lens culinaris</i>	-	15,910 core genes , 24,226 accessory genes	8	[62]
<i>Pisum sativum</i>		15,470 core genes, softcore genes, 6,170, 41,028 shell genes	116	[38]
		50,108 cloud genes		

4. Chickpea pangenome

139

Chickpea is a major pulse crop, ranking third among pulses for global production and serving as a plant source of protein and many essential micronutrients [59]. In the last decade, unprecedented advancements in bioinformatics and NGS technologies allowed for the assembly of a chickpea genome sequence [34,67] which has served as a tool for chickpea improvement. To further build chickpea genomic resources, and to study speciation, phylogenetic analysis, genomic diversity of cultivated species and its wild progenitor and migration of cultivated chickpea species, Varshney et al. [59] constructed the first chickpea pangenome(see Fig.2) (measuring 592.58 Mb with 29,870 genes) by sequencing 3,171 cultivated and 195 wild chickpea accessions using an iterative mapping and assembly approach. From which, 124,833 SNPs were selected to form linkage disequilibrium blocks; these SNPs and the phenotyping data for 100 seed weight (100SW) and yield per plant (YPP) from 2,980 genotypes were used for genomic prediction to improve these traits [59]. Likewise, 3.94 million SNPs and the phenotyping data for 16 phenological and yield-related traits in 2,980 genotypes revealed associations between 205 SNPs and 11 traits, with 152 SNPs associated with 79 unique genes governing seed size and development [59].

5. Cowpea pangenome

156

Cowpea is an important warm season, climate-resilient, multipurpose legume crop in sub-Saharan Africa [68]. Significant progress has been achieved in developing genomics resources in cowpea. Lonardi et al. [42] developed the genome assembly of cowpea genotype IT97K-499-35. Liang et al. [57] constructed the first cowpea pangenome by producing *de novo* assemblies of six cowpea accessions with a mean size 449.91 Mb, identifying 21,330 core and 23,531 non-core genes, with the large number of frame shift variations in the non-core genes attributing large diversity in domesticated cowpea (see Table 1). Comparative and genome synteny analyses of the IT97K-499-35 reference genome and the genome assemblies of the six accessions indicated the presence of numerous structural rearrangements, including PAVs, inversions, indels, and CNVs [57]. The authors confirmed that the existing PAVs are responsible for black seed coat color in cowpea (see Table 2).

Table 2. Applications of pan-genome in various legume crops.

168

Crop	Major finding from the pangenome analysis	Reference
<i>Glycine soja</i> (wild soybean)	2.3–3.9 Mbp of <i>G. soja</i> -specific PAV related to defence response, cell growth and photosynthesis. Pangenome analysis informed variation for protein, oil, flowering time and organ size traits in the wild and cultivated soybean species	[17]
<i>Glycine max</i> (soybean)	PanSoy sheds novel insights into the intraspecific variation in <i>G. max</i>	[16]
<i>Glycine max</i> (soybean)	The function of core genes were related to growth, immune system, reproductive, cellular, and cellular component organization or biogenesis. The function of non-core genes were related to abiotic and biotic response genes, such as different NBS gene families. Structural variation related to domestication traits and agronomic trait viz., iron deficiency chlorosis	[54]
<i>Glycine max</i> (soybean)	Domestication selection sweeps on chromosome Gm20, breeding-related selective sweep region on Gm20, 110 genomic regions with signatures of domestication-selective sweeps harbouring 1,266 protein-coding genes, 86 genomic regions with signatures of breeding-selective sweeps harbouring 1,434 protein-coding genes	[55]
<i>Vigna radiata</i>	A total of nine presence/absence variations controlling flowering regulation, bruchid resistant genes Pang34265, Pang44622, Pang57772, Pang58608, and Pang64254	[56]
<i>Vigna unguiculata</i>	PAVs contributing to black seed coat colour in cowpea	[57]
<i>Cajanus cajan</i>	PAVs attributing phenotypic diversity including various phenotypic traits viz., seed weight, days to 50% flowering and plant height	[58]
<i>Cicer arietinum</i>	643 gene-gain and 247 gene-loss CNVs in <i>C. reticulatum</i> accessions, insertions (139,483), deletions (47,882), inversions (61,171), intra-chromosomal translocations (417) and inter-chromosomal translocations (2,410) in cultivated and 287,854 insertions, 67,351 deletions, 58,070 inversions, 446 intra-chromosomal translocations and 2,066 inter-chromosomal translocations among <i>C. reticulatum</i> accessions	[59]
<i>Medicago truncatula</i>	500,000–1,500,000 short InDels (<50 bp), 27,000–110,000 large InDels, 49,000–169,000 copy number variants (CNVs), and 2,700–12,700 translocations, NBS-LRRs showed high SNP diversity,	[60]
<i>Medicago sativa</i>	circRNA, lncRNA responsive to salinity and drought	[59]
<i>Lupinus alba</i>	Alkaloid related genes/QTLs/genomic region, 1195 PAVs	[61]
<i>Lens culinaris</i>	Comparative analysis and evolutionary analysis, transcriptome of <i>L. culinaris</i> contained genes 58,375, <i>L. nigricans</i> transcriptome contained the minimum number 46,742 genes	[62]
<i>Pisum sativum</i>	Two seed dormancy genes <i>Psat02G0081200</i> and <i>Psat02G0507900</i> were elucidated, pan genes of <i>P. abyssinicum</i> were unique in response to chemical and stimuli, pan genes of <i>P. fulvum</i> showed their role in development, cytoskeleton and tropism	[38]

6. Pigeonpea pangenome

Pigeon pea is an important grain legume, mostly grown in India and Africa, with an essential role in ensuring protein nutrition security across the semi-arid tropics [36]. The pigeon pea reference genome sequence [36] has been used to improve various traits in pigeon pea; however, the pangenome assembly of pigeon pea provides further opportunities to harness untapped genetic variation not present in the reference genome sequence. Zhao et al. [58] built the pigeon pea pangenome (measuring 622 Mbp) using iterative mapping and an assembly approach [13], sequencing 89 pigeon pea accessions collected across the globe and the reference pigeon pea with >9.5× coverage. The pangenome contained 55,512 more genes than the previously assembled reference genome (53,612), with 48,067 core genes and 7,445 accessory genes [58]. A functional analysis of the 'variable genes' indicated their possible role in proteolysis, pollen tube reception and signal transduction process [58]. Of 909 identified R genes, 836 were core genes and 73 were variable genes.

169

170

171

172

173

174

175

176

177

178

179

180

181

To confirm the role of PAVs attributing phenotypic diversity, the authors detected several additional SNPs for various phenotypic traits (seed weight, days to 50% flowering and plant height) compared to Varshney et al. [70].

7. Mung bean pangenome

Mung bean is an important grain legume crop with high nutritional benefits, mostly grown in south and southeast Asian countries [71]. Draft genome assemblies are available for mung bean genotypes VC1973A [43] and JL7 [56]. Liu et al. [56] assembled the first mung bean pangenome (measuring 762.92 Mb), with 43,462 annotated genes, by deep-sequencing JL7 and 217 mung bean accessions, including cultivars and landraces. The predicted genes comprised 33,258 hard-core genes, 2,872 soft-core genes, 7,154 shell genes, and 178 cloud genes [56]. Of the nine PAVs controlling flowering regulation, three genes (*jg13350*, *jg13746*, and *Pang80812*) were present in Chinese landraces and breeding lines [56], while six genes (*jg1521*, *jg5273*, *jg5274*, *jg5281*, *jg5284*, and *Pang68295*) were absent but present in non-Chinese lines [56].

8. White lupin pangenome

White lupin (*Lupinus albus* L.) is a protein-rich grain legume crop domesticated in the Mediterranean region [72,73]. However, the presence of quinolizidine alkaloids in lupin seed causes a bitter taste and toxicity; thus lowering seed alkaloids is a prime objective of lupin breeding programs. Hufnagel et al. [40] and Xu et al. [74] constructed the white lupin genome assembly. Hufnagel et al. [61] established the first white lupin pangenome by sequencing a diverse set of 39 white lupin accessions using a ‘map-to-pan’ approach [66]. The pangenome comprised 32,068 (78.5%) core genes, 6,046 soft-core genes, and 8,776 shell genes (Hufnagel et al. 2021)[61] (see Table 1). Several selection sweeps related to breeding for low-alkaloid accessions were identified on chromosomes 18, 15, 1, and 3, overlapping the low-alkaloid content QTLs previously reported by Ksiazkiewicz et al. [75] and Lin et al. (2009)[76].

9. Medicago truncatula pangenome

Medicago truncatula is an important model legume crop used for investigating symbiotic relationships with rhizobia and mycorrhizae and root development [77,78]. Zhou et al. [60] developed the first *M. truncatula* pangenome (measuring 388–428 Mbp, with an additional 63 Mbp novel sequence) by *de novo* assembly of 15 *M. truncatula* accessions, adding 16% genome space to the reference genome sequence [79] of this model legume (see Table 1). This pangenome sheds light on causative structural variants, including transposable elements in different gene families such as nucleotide-binding site leucine-rich repeat rapid evolution.

11. Lentil pangenome

Lentil is an important source of dietary protein, carbohydrates, and iron for humans [80]. Due to the large size of the genome (4Gb) and the presence of high number repetitive elements, the completion of a lentil genome assembly has taken some time to complete. However, an incomplete draft genome sequence of *Lens culinaris* cv. CDC Redberry v2.0 [81, 82] has facilitated new directions in lentil genomics research. Thus, in the absence of a complete reference genome sequence, pan-transcriptomes [83] could provide a platform to establish the lentil pangenome. Guerra-Garcia et al. (2022)[84] have used an exome capture array built on the draft lentil genome and transcriptomic data [85] to examine copy number variation across cultivated lentil germplasm. This work has uncovered a previously hidden amount of variation across lentils[84]. Furthermore, Gutierrez-Gonzalez et al.[62] assembled and annotated the transcriptomes of eight lentil accessions, including cultivated and wild species (*L. orientalis*, *L. tomentosus*, *L. ervoides*, *L. lamottei*, *L. nigricans*, and two *L. odemensis*), identifying 15,910 core genes and 24,226 accessory genes. A

comparative analysis of these assembled transcriptomes with the incomplete CDC Redberry reference genome assembly indicated that transcripts of *L. culinaris* had the greatest similarity with CDC Redberry transcripts, while transcripts of *L. lamottei* and *L. nigricans* had the least [62]. Further long-read-based sequencing efforts could assist in decoding the complete genome and pangenome assembly of lentil.

12. Pea pangenome

Pea is the fourth most globally important grain legume, enriched with protein, starch, and minerals [86,38] and considered a ‘founder crop’ member; it is also one of the oldest domesticated crops [87]. Despite having an assembled pea reference genome [37], genetic diversity in cultivated and wild pea and the domestication history of pea remain elusive. Yang et al. [38] constructed a second pea *de novo* genome assembly of ‘Zwo’ and a pangenome assembly of 116 accessions containing cultivated and wild pea genotypes, identifying 112,117 pangenes, including 15,470 core genes, 6,170 soft-core genes, 41,028 shell genes, and 50,108 cloud genes. The study confirmed that *Pisum abyssinicum* is separate from *P. fulvum* and *P. sativum*. The pangenes from *Pisum abyssinicum* were unique in stimulus and chemical response, while those of *P. fulvum* were involved in growth, tropism, and development; thus, the novel genes harboring in these species could be used to improve disease resistance and yield in modern pea cultivars [38].

13. Pangenome-derived PAVs for GWAS analysis to explore novel genes

Pangenome analysis allows the discovery of a plethora of novel SVs, including gene PAVs that may be missing in the reference genome assembly [27,16,56]. Thus, these pangenome-derived PAVs can be used in a GWAS to uncover novel agronomic trait gene allele(s). Liu et al. [56] uncovered five novel genes missing in the mung bean reference genome assembly that contributed to bruchid resistance (*Pang34265*, *Pang44622*, *Pang57772*, *Pang58608*, and *Pang64254*). Zhao et al. [58] identified 225 significant SNPs associated with various phenological and yield-related traits (e.g., days to 50% flowering, days to maturity, pods/plant, yield/plant) from SNPs and PAVs derived from a pigeon pea pangenome analysis, which overlapped those previously identified by Varshney et al. [70]. However, several new SNPs were associated with these studied traits [58]. Likewise, Varshney et al. [59] identified 205 significantly associated SNPs for agronomic traits in a GWAS study on 16 traits in a panel of 2,980 chickpea accessions using 3.94 million SNPs, of which 152 SNPs were in 79 unique genes related to seed size and seed development [59]. GWAS analysis using SVs derived from 2,898 soybean accessions elucidated the genetic basis of soybean seed luster, uncovering the presence of 10 kb PAVs in the hydrophobic protein encoding gene causing seed luster [54].

14. Pangenome and scope of crop domestication for new species

The availability of reference genomes and pangenome assemblies in various legume crops facilitates identifying causal variants, such as SVs, related to domestication traits (see Fig.3). Genome editing tools allow us to edit these traits from wild species to make them agronomically desirable [45]. For instance, pangenome analysis indicated that *G. soja*, a crop wild relative of soybean, contained more NBS-R genes (25) than the cultivated counterpart *G. max* (14), providing increased advantages against various biotic stress [17]. Similarly, *GmFT2c* a candidate gene regulating flowering time, was present as a complete gene structure in *G. soja*; however, *GmaxW82* had *GmFT2c* gene with missing up to 50% N-terminal peptide [17]. Pangenome analysis revealed that the change in seed coat color from black in wild soybean cultivars to yellow in cultivated cultivars related to SV inversion in chalcone synthase gene (*CSH*) and *CSH* gene duplication during soybean domestication [54]. Likewise, the authors confirmed that wild soybean contains more SVs than landraces and cultivated soybean from 26 *de novo* genome assemblies and data for 2,872 soybean lines. Bayer et al. [55] also reported gene losses during domestication in a

pangenome analysis of >1,000 soybean lines, with wild soybean containing more genes (48,785) than landraces (48,371) and modern cultivars (48,165). In another pangenome analysis, Varshney et al. [59] indicated that present-day chickpea cultivars contain several deleterious alleles due to domestication, affecting genetic fitness. Intensive breeding activities and domestication have reduced genetic diversity [55,59]. North American soybean breeding lines possess an estimated 85% of the genes derived from only 19 landraces of soybean [88]. In another study, 79% of rare alleles harbored by diverse landraces were lost during breeding in soybean [89]. A pangenome analysis indicated that wild pea genomes (e.g., *P. fulvum* and *P. abyssinicum*) contained higher SVs than the cultivated species (*P. sativum*), confirming that wild genomes are rich in novel genes of agronomic traits [38]. Thus, genetic variants differ between cultivated and wild types, and the genome editing could target deleterious alleles for domestication [45].

Particularly useful targets of breeding for *de novo* domestication or for enhanced utilization of wild relatives of existing crops come from our growing understanding of the genetic basis of domestication in some of the most extensively studied legumes. The domestication syndrome has long been a useful concept for this, with a few essential traits repeated, selected by farmers to convert wild plants into crops that could be effectively planted, harvested, and consumed (e.g. [90–94]). Cultivated legumes share similarities in indehiscent pods [95, 96], loss of seed dormancy [97], increased seed size (e.g., [98,99] , and reduced defensive chemistry (e.g., [100,101]. These have long been suspected to have a common genetic basis (e.g.,[102]. Furthermore, post-domestication traits involved in crop spread from centers of origin and diversification into distinct agroecologies may also have shared genetic bases (e.g.,[103]. Phylogenetically informed super-pangenomes are the ideal tool to examine questions of repeated selection on particular loci, pathways, or gene families across legumes (e.g.,[15]. In legume clades with multiple domestications, such as the genera *Vigna*, *Phaseolus*, *Lupinus*, and the closely related cool season legumes in *Pisum*, *Lens*, and *Vicia*, genus or even tribe legume pan-genomes would be immense tools for understanding shared and contrasting patterns of selection. These tools may also assist in harnessing germplasm for crop improvement. In forage legumes, such as *Medicago sativa* and *Trifolium*, these tools could have a similar impact [78]. And in legumes that are potential targets of neodomestication, such as legumes making industrially useful compounds like *Derris*, *Tephrosia*, and *Indigofera*, these tools could be equally powerful.

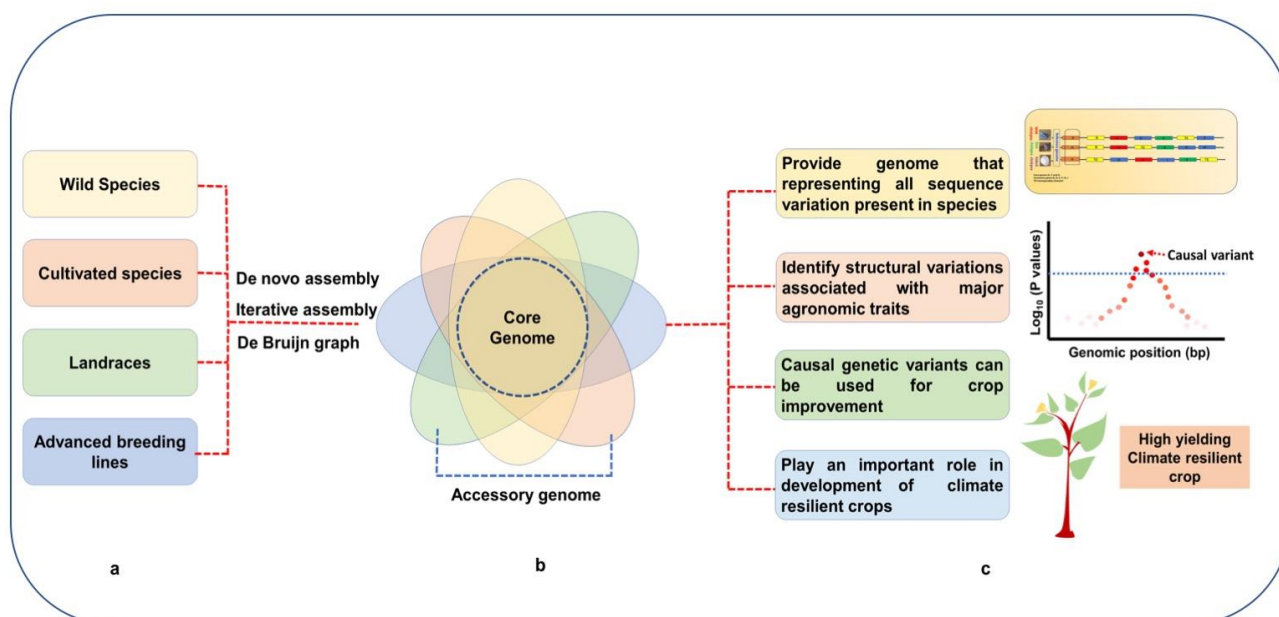


Figure 3. Summarizing pan-genome and their application in crop improvement, a; Different re-courses and various assembly approaches for pangenome construction, **b;** Represent pangenome with shared (core genome) and non-shared genome (accessory genome) between the accessions, **c;** Mention the application of pan-genome in crop improvement.

15. Prospects and conclusion

The increasing human population, decreasing land for agriculture, and drastic changes in global climate ~~change~~ pose serious challenges for crop production to ensure food security [3,4], requiring novel breeding and genomics approaches. Pangenome resources of major legume species could be instrumental in shedding light on crop domestication and evolution and interrogating the genetic basis of agronomic traits to improve genetic grain (see Fig. 3). Further, these pangenome resources could be used for developing novel single nucleotide polymorphism, performing GWAS, mapping QTL, mining superior haplotypes of various traits, and genomic prediction in crop breeding programs [26,15,52,104]. Combining pangenome information with CRISPR-Cas9 based genome editing technology could open up new avenues for *de novo* domestication of wild species [45]. Leveraging pangenome, transcriptome, epigenome, metabolome, and phenome data with machine learning approaches could greatly benefit genomics-assisted breeding [104]. Further, these resources could help improve underutilized legumes with no developed reference genome or pangenome assembly. Further advances in long-read sequencing and improved genome assembly construction and genome annotation will help develop high-quality pangenomes with few gaps and assist in predicting PAVs in targeted trait genes with more precision [105,106]. With decreasing sequencing costs and advances in computational biology, including various software and analysis pipelines, pangenome studies could encompass all species in a genus to construct a ‘super pangenome’ [8,15] providing a comprehensive repertoire of genomic information at the genus and even family level.

Author Contributions: Conceptualization, U.C.J. and K.H.M.S.; writing—original draft preparation, U.C.J., H.N., MT. and E.J.B.v.W.; overall manuscript editing, K.H.M.S. and M.T. Figures drawn, Y.D.N. All authors have read and agreed to the published version of the manuscript.

Funding: APC was provided by The University of Western Australia. E.J.B.v.W is supported for pangenomic work by USDA NIFA 2021-07634, USDA Hatch 2210, USDA SARE LNE21-428R and NSF BII 2120153

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the article.

Acknowledgments: U.C.J. acknowledge support from ICAR, New Delhi, India.

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

References

- Schmidhuber, J.; Tubiello, F.N. Global food security under climate change. *Proc. Natl. Acad. Sci.* **2007**, *104*, 19703–8.
- Leisner, C.P. Climate change impacts on food security-focus on perennial cropping systems and nutritional value. *Plant Sci.* **2020**, *293*, 110412.
- Lenaerts, B.; Collard, B.C.; Demont, M. Improving global food security through accelerated plant breeding. *Plant Sci.* **2019**, *287*, 110207.
- Rivers, J.; Warthmann, N.; Pogson, B.J.; Borevitz, J.O. Genomic breeding for food, environment and livelihoods. *Food Secur.* **2015**, *7*, 375–82.
- Varshney, R.K.; Nayak, S.N.; May, G.D.; Jackson, S.A. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trend Biotechnol.* **2009**, *27*, 522–530.
- Saxena, R.K.; Edwards, D.; Varshney, R.K. Structural variations in plant genomes. *Briefs. Func. Genom.* **2014**, *13*, 296–307.
- Yuan, Y.; Bayer, P.E.; Batley, J.; Edwards, D. Current status of structural variation studies in plants. *Plant Biotechnol. J.* **2021**, *19*, 2153–63.
- Bayer, P.E.; Golicz, A.A.; Scheben, A.; Batley, J.; Edwards, D. Plant pan-genomes are the new reference. *Nat Plants* **2020**, *6*, 914–20.
- Tettelin, H.; Maignani, V.; Cieslewicz, M.J.; Donati, C.; Medini, D.; Ward, N.L.; Angiuoli, S.V.; Crabtree, J.; Jones, A.L.; Durkin, A.S.; DeBoy, R.T. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc. Natl. Acad. Sci.* **2005**, *102*, 13950–5.
- Tao, Y.; Zhao, X.; Mace, E.; Henry, R.; Jordan, D. Exploring and exploiting pan-genomics for crop improvement. *Mol. Plant* **2019**, *12*, 156–169.
- Tao, Y.; Jordan, D.R.; Mace, E.S. A graph-based pan-genome guides biological discovery. *Mol. Plant* **2020**, *13*, 1247–9.
- Danilevich, M.F.; Fernandez, C.G.T.; Marsh, J.I.; Bayer, P.E.; Edwards, D. Plant pangenomics: approaches, applications and advancements. *Curr. Opin. Plant Biol.* **2020**, *54*, 18e25.
- Golicz, A.A.; Bayer, P.E.; Barker, G.C.; Edger, P.P.; Kim, H.; Martinez, P.A.; Chan, C.K.; Severn-Ellis, A.; McCombie, W.R.; Parkin, I.A.; Paterson, A.H. The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat. Commun.* **2016**, *7*, 1–8.
- Golicz, A. A.; Bayer, P. E.; Bhalla, P. L.; Batley, J.; Edwards, D. (2020).Pangenomics comes of age: From bacteria to plant and animal applications. *Trends Genet.* **2020**, *36*, 132–145.
- Khan, A.W.; Garg, V.; Roorkiwal, M.; Golicz, A.A.; Edwards, D.; Varshney, R.K. Super-pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends Plant Sci.* **2020**, *25*, 148–58.
- Torkamaneh, D.; Lemay, M.A.; Belzile, F. The pan-genome of the cultivated soybean (PanSoy) reveals an extraordinarily conserved gene content. *Plant Biotechnol. J.* **2021**, *19*, 1852–62.

17. Li, Y.H.; Zhou, G.; Ma, J.; Jiang, W.; Jin, L.G.; Zhang, Z.; Guo, Y.; Zhang, J.; Sui, Y.; Zheng, L.; Zhang, S.S. et al. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* **2014**, *32*, 1045–52.
18. Morgante, M.; De Paoli, E.; Radovic, S. Transposable elements and the plant pan-genomes. *Curr. Opin. Plant Biol.* **2007**, *10*, 149–55.
19. Jiao, W.B.; Schneeberger, K. Chromosome-level assemblies of multiple Arabidopsis genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nat. Commun.* **2020**, *11*, 1–0.
20. Schatz, M.C.; Maron, L.G.; Stein, J.C.; Wences, A.H.; Gurtowski, J.; Biggers, E.; Lee, H.; Kramer, M.; Antoniou, E.; Ghiban, E.; Wright, M.H. Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of aus and indica. *Genome Biol.* **2014**, *15*, 1–6.
21. Montenegro, J.D.; Golicz, A.A.; Bayer, P.E.; Hurgobin, B.; Lee, H.; Chan, C.K.; Visendi, P.; Lai, K.; Doležel, J.; Batley, J.; Edwards, D. The pangenome of hexaploid bread wheat. *Plant J.* **2017**, *90*, 1007–13.
22. Franco, J.A.V.; Gage, J.L.; Bradbury, P.J.; Johnson, L.C.; Miller, Z.R.; Buckler, E.S.; Roday, M.C.A. Maize practical haplotype graph leverages diverse NAM assemblies. *bioRxiv* **2020**.
23. Gordon, S.P.; Contreras-Moreira, B.; Woods, D.P.; Des Marais, D.L.; Burgess, D.; Shu, S.; Stritt, C.; Roulin, A.C.; Schackwitz, W.; Tyler, L.; Martin, J. Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure. *Nat. Commun.* **2017**, *8*, 1–3.
24. Gordon, S.P.; Contreras-Moreira, B.; Levy, J.J.; Djamei, A.; Czedik-Eysenberg, A.; Tartaglio, V.S.; Session, A.; Martin, J.; Cartwright, A.; Katz, A.; Singan, V.R. (2020) Gradual polyploid genome evolution revealed by pan-genomic analysis of Brachypodium hybridum and its diploid progenitors. *Nat. Commun.* **2020**, *11*, 1–6.
25. Ruperao, P.; Thirunavukkarasu, N.; Gandham, P.; Selvanayagam, S.; Govindaraj, M.; Nebie, B.; Maniyasa, E.; Gupta, R.; Das, R.R.; Odeny, D.A.; Gandhi, H. Sorghum pan-genome explores the functional utility for genomic-assisted breeding to accelerate the genetic gain. *Front. Plant Sci.* **2021**, *12*, 963.
26. Hurgobin, B.; Edwards, D. SNP discovery using a pangenome: has the single reference approach become obsolete? *Biology*, **2017**, *6*, 21.
27. Bayer, P.E.; Scheben, A.; Golicz, A.A.; Yuan, Y.; Faure, S.; Lee, H.; Chawla, H.S.; Anderson, R.; Bancroft, I.; Raman, H. et al. Modelling of gene loss propensity in the pangenomes of three *Brassica* species suggests different mechanisms between polyploids and diploids. *Plant Biotechnol. J.* **2021**, *19*, 2488–2500.
28. Hübner, S.; Bercovich, N.; Todesco, M.; Mandel, J.R.; Odenheimer, J.; Ziegler, E.; Lee, J.S.; Baute, G.J.; Owens, G.L.; Grassa, C.J.; Ebert, D.P., 2019. Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nature Plants*, **2019**, *5*, 54–62.
29. Gao, L.; Gonda, I.; Sun, H.; Ma, Q.; Bao, K.; Tieman, D.M.; Burzynski-Chang, E.A.; Fish, T.L.; Stromberg, K.A.; Sacks, G.L.; Thannhauser, T.W. The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* **2019**, *51*, 1044–51.
30. Ou, L.; Li, D.; Lv, J.; Chen, W.; Zhang, Z.; Li, X.; Yang, B.; Zhou, S.; Yang, S.; Li, W.; Gao, H. Pan-genome of cultivated pepper (*Capsicum*) and its use in gene presence–absence variation analyses. *New Phytol.* **2018**, *220*, 360–3.
31. Li, J.; Yuan, D.; Wang, P.; Wang, Q.; Sun, M.; Liu, Z.; Si, H.; Xu, Z.; Ma, Y.; Zhang, B.; Pei, L. Cotton pan-genome retrieves the lost sequences and genes during domestication and selection. *Genome Biol.* **2021**, *22*, 1–26.
32. Sun, X.; Jiao, C.; Schwaninger, H.; Chao, C.T.; Ma, Y.; Duan, N.; Khan, A.; Ban, S.; Xu, K.; Cheng, L.; Zhong, G.Y. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat. Genet.* **2020**, *52*, 1423–32.

33. Schmutz, J.; Cannon, S.B.; Schlueter, J.; Ma, J.; Mitros, T.; Nelson, W.; Hyten, D.L.; Song, Q.; Thelen, J.J.; Cheng, J.; Xu, D. et al. Genome sequence of the palaeopolyploid soybean. *Nature*, **2010**, *463*, 178–83.
34. Varshney, R.K.; Song, C.; Saxena, R.K.; Azam, S.; Yu, S.; Sharpe, A.G.; Cannon, S.; Baek, J.; et al. (2013a) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* **2013**, *31*, 240–246.
35. Schmutz, J.; McClean, P.E.; Mamidi, S.; Wu, G.A.; Cannon, S.B.; Grimwood, J.; Jenkins, J.; Shu, S.; Song, Q.; Chavarro, C.; Torres-Torres, M. et al. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* **2013**, *46*, 707–13.
36. Varshney, R.K.; Chen, W.; Li, Y.; Bharti, A.K.; Saxena, R.K.; Schlueter, J.A.; Donoghue, M.T.; et al. (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.* **2012**, *30*, 83.
37. Kreplak, J.; Madoui, M.A.; Cápál, P.; Novák, P.; Labadie, K.; Aubert, G.; Bayer, P.E.; Gali, K.K.; Syme, R.A.; Main, D.; Klein, A. et al. A reference genome for pea provides insight into legume genome evolution. *Nat. Genet.* **2019**, *51*, 1411–22.
38. Yang, T.; Liu, Rong.; Luo, Y.; Hu, S.; Wang, D. Improved pea reference genome and pan-genome highlight genomic features and evolutionary characteristics. *Nat. Genet.* **2022** doi.org/10.1038/s41588-022-01172-2
39. Hane, J.K.; Ming, Y.; Kamphuis, L.G.; Nelson, M.N.; Garg, G.; Atkins, C.A.; Bayer, P.E.; Bravo, A.; Bringans, S.; Cannon, S.; Edwards, D. A comprehensive draft genome sequence for lupin (*Lupinus angustifolius*), an emerging health food: insights into plant–microbe interactions and legume evolution. *Plant Biotechnol. J.* **2017**, *15*, 318–30.
40. Hufnagel, B.; Marques, A.; Soriano, A.; Marquès, L.; Divol, F.; Doumas, P.; Sallet, E.; Mancinotti, D.; Carrere, S.; Marande, W.; Arribat, S. (2020) High-quality genome sequence of white lupin provides insight into soil exploration and seed quality. *Nat. Commun.* **2020**, *11*, 1–2.
41. Bertoli, D.J.; Jenkins, J.; Clevenger, J.; Dudchenko, O.; Gao, D.; Seijo, G.; Leal-Bertioli, S.; Ren, L.; Farmer, A.D.; Pandey, M.K.; Samoluk, S.S. et al. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nature Genet* **2019**, *51*, 877–84
42. Lonardi, S.; Muñoz-Amatriaín, M.; Liang, Q.; Shu, S.; Wanamaker, S.I.; Lo, S.; Tanskanen, J.; Schulman, A.H.; Zhu, T.; Luo, M.C.; Alhakami, H.; Ounit, R.; Hasan, A.M.; Verdier, J.; Roberts, P.A.; Santos, J.R.P.; Ndeve, A.; Doležel, J.; Vrána, J.; Hokin, S.A.; Farmer, A.D.; Cannon, S.B.; Close, T.J. The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *Plant J.* **2019**, *98*, 767–782.
43. Kang, Y.J.; Kim, S.K.; Kim, M.Y.; Lestari, P.; Kim, K.H.; Ha, B.K.; Jun, T.H.; Hwang, W.J.; Lee, T.; Lee, J.; Shim, S. (2014) Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat. Commun.* **2014**, *5*, 1–9.
44. Gaut, B.S.; Seymour, D.K.; Liu, Q.; Zhou, Y. Demography and its effects on genomic variation in crop domestication. *Nat. Plants*, **2018**, *4*, 512–520.
45. Della Coletta, R.; Qiu, Y.; Ou, S.; Hufford, M.B.; Hirsch, C.N. How the pan-genome is changing crop genomics and improvement. *Genome Biol.* **2021**, *22*, 1–9.
46. Li, W.; Liu, J.; Zhang, H.; Liu, Z.; Wang, Y.; Xing, L.; He, Q.; Du, H. Plant pan-genomics: recent advances, new challenges, and roads ahead. *J. Genet Genom* **2022** Jun 21.
47. Jiang, N.; Bao, Z.; Zhang, X.; Eddy, S.R.; Wessler, S.R. Pack-MULE transposable elements mediate gene evolution in plants. *Nature* **2004**, *431*, 569–73.
48. Fedoroff, N.V. Transposable elements, epigenetics, and genome evolution. *Science*. **2012**, *338*, 758–67.
49. Wei, L.; Cao, X. The effect of transposable elements on phenotypic variation: insights from plants to humans. *Sci. China Life Sci.* **2016**, *59*, 24–37.

50. Yandea-Nelson, M.D.; Xia, Y.; Li, J.; Neuffer, M.G.; Schnable, P.S. Unequal sister chromatid and homolog recombination at a tandem duplication of the A1 locus in maize. *Genetics*. **2006**, *173*, 2211–26. 485–487
51. Lei, L.; Goltsman, E.; Goodstein, D.; Wu, G.A.; Rokhsar, D.S.; Vogel, J.P. Plant pan-genomics comes of age. *Annu. Rev. Plant Biol.* **2021**, *72*, 411–35. 488–489
52. Jayakodi, M.; Schreiber, M.; Stein, N.; Mascher, M. Building pan-genome infrastructures for crop plants and their use in association genetics. *DNA Res* **2021**, *28*, dsaa030 490–491
53. Li-Beisson, Y.; Shorrosh, B.; Beisson, F.; Andersson, M.X.; Arondel, V.; Bates, P.D.; Baud, S.; Bird, D.; DeBono, A.; Durrett, T.P.; Franke, R.B. Acyl-lipid metabolism. *The Arabidopsis book/American Society of Plant Biologists*, **2013**, *11*. 492–494
54. Liu, Y.; Du, H.; Li, P.; Shen, Y.; Peng, H.; Liu, S.; Zhou, G.A.; Zhang, H.; Liu, Z.; Shi, M.; Huang, X. Pan-genome of wild and cultivated soybeans. *Cell* **2020**, *182*, 162–76 495–496
55. Bayer, P.E.; Valliyodan, B.; Hu, H.; Marsh, J.I.; Yuan, Y.; Vuong, T.D.; Patil, G.; Song, Q.; Batley, J.; Varshney, R.K.; Lam, H.M. Sequencing the USDA core soybean collection reveals gene loss during domestication and breeding. *Plant Genome* **2022**, *15*, e20109. 497–499
56. Liu, C.; Wang, Y.; Peng, J.; Fan, B.; Xu, D.; Wu, J.; Cao, Z.; Gao, Y.; Wang, X.; Li, S.; Su, Q. High-quality genome assembly and pan-genome studies facilitate genetic discovery in mung bean and its improvement. *Plant Commun.* **2022**, *26*, 100352 500–502
57. Liang, Q.; Munoz-Amatriain, M.; Shu, S.; Lo, S.; Wu, X.; Carlson, J.W.; Davidson, P.; Goodstein, D.M.; Phillips, J.; Janis, N.M.; Lee, E.J. A view of the pan-genome of domesticated cowpea (*Vigna unguiculata* [L.] Walp.). *bioRxiv*. 2022 Jan 1 503–505
58. Zhao, J.; Bayer, P.E.; Ruperao, P.; Saxena, R.K.; Khan, A.W.; Golicz, A.A.; Nguyen, H.T.; Batley, J.; Edwards, D.; Varshney, R.K. (2020) Trait associations in the pangenome of pigeon pea (*Cajanus cajan*). *Plant Biotechnol. J.* **2020**, *18*, 1946–54. 506–508
59. Varshney, R.K.; Roorkiwal, M.; Sun, S.; Bajaj, P.; Chitikineni, A.; Thudi, M.; Singh, N.P.; Du, X.; et al. A chickpea genetic variation map based on the sequencing of 3,366 genomes. *Nature* **2021**, *599*, 622–7. 509–511
60. Zhou, P.; Silverstein, K.A.; Ramaraj, T.; Guhlin, J.; Denny, R.; Liu, J.; Farmer, A.D.; Steele, K.P.; Stupar, R.M.; Miller, J.R.; Tiffin, P. (2017) Exploring structural variation and gene family architecture with De Novo assemblies of 15 Medicago genomes. *BMC Genom.* **2007**, *18*, 1–4. 512–514
61. Hufnagel, B.; Soriano, A.; Taylor, J.; Divol, F.; Kroc, M.; Sanders, H.; Yeheyis, L.; Nelson, M.; Péret B. Pangenome of white lupin provides insights into the diversity of the species. *Plant Biotechnol. J.* **2021**, *19*, 2532–43. 515–517
62. Gutierrez-Gonzalez, J.J.; García, P.; Polanco, C.; González, A.I.; Vaquero, F.; Vences, F.J.; Pérez de la Vega, M.; Sáenz de Miera, L.E. Multi-Species Transcriptome Assemblies of Cultivated and Wild Lentils (*Lens sp.*) Provide a First Glimpse at the Lentil Pangenome. *Agronomy* **2022**, *12*, 1619. 518–520
63. Shen, Y.; Liu, J.; Geng, H.; Zhang, J.; Liu, Y.; Zhang, H.; Xing, S.; Du, J.; Ma, S.; Tian, Z. (2018). *De novo* assembly of a Chinese soybean genome. *Sci. China Life Sci.* **2018**, *61*, 871–884. 521–522
64. Shen, Y.; Du, H.; Liu, Y.; Ni, L.; Wang, Z.; Liang, C.; Tian, Z. Update soybean Zhonghuang 13 genome to a golden reference. *Sci. China Life Sci.* **2019**, *62*, 1257–1260. 523–524
65. Xie, M.; Chung, C.Y.; Li, M.W.; Wong, F.L.; Wang, X.; Liu, A.; Wang, Z.; Leung, A.K.; Wong, T.H.; Tong, S.W.; et al. A reference-grade wild soybean genome. *Nat. Commun.* **2019**, *10*, 1216. 525–526
66. Hu, Z.; Sun, C.; Lu, K.C.; Chu, X.; Zhao, Y.; Lu, J.; Shi, J.; Wei, C. (2017). EUPAN enables pan-genome studies of a large number of eukaryotic genomes. *Bioinformatics* **2017**, *33*, 2408–2409. 527–528
67. Jain, M.; Misra, G.; Patel, R.K.; Priya, P.; Jhanwar, S.; Khan, A.W.; Shah, N.; Singh, V.K.; Garg, R.; Jeena, G.; Yadav, M.; Kant, C.; Sharma, P.; Yadav, G.; Bhatia, S.; Tyagi, A.K.; Chattopadhyay, D. A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *Plant J.* **2013**, *74*, 715–729. 529–531
68. Timko, M.P.; Singh, B.B. (2008) Cowpea, a multifunctional legume. In P. H. Moore & R. Ming (Eds.), *Genomics of tropical crop plants* (pp. 227–258). New York, NY: Springer. 532–533

69. Medina, C.A.; Samac, D.A.; Yu, L.X. Pan-transcriptome identifying master genes and regulation network in response to drought and salt stresses in Alfalfa (*Medicago sativa* L.). *Sci. Reps* **2021**, *11*, 1–16. 534–536
70. Varshney, R.; Saxena, R.; Upadhyaya, H.; Khan, A.; Yu, Y.; Kim, C.; Rathore, A. et al. Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nat. Genet.* **2017**, *49*, 1082–1088. 537–539
71. Jha, U.C.; Nayyar, H.; Parida, S.K.; Bakır, M.; von Wettberg, E.J.; Siddique, K.H.M. Progress of Genomics-Driven Approaches for Sustaining Underutilized Legume Crops in the Post-Genomic Era. *Front. Genet.* **2022**, *13*, 831656. 540–542
72. Wolko, B.; Clements, J.C.; Naganowska, B.; Nelson, M.; Huaan, Y. *Lupinus* (Kole, C., ed.). *Wild Crop Relatives: Genomic and Breeding Resources*. Berlin, Heidelberg: Springer Berlin Heidelberg. **2011**, pp. 153–206. 543–545
73. Taylor, J.L.; De Angelis, G.; Nelson, M.N. How have narrow-leaved lupin genomic resources enhanced our understanding of lupin domestication?. In *The Lupin Genome*, (Singh, K. B., Kamphuis, L. G. & Nelson, M. N. eds.), **2020**, pp. 95–108. Cham: Springer. 546–548
74. Xu, W.; Zhang, Q.; Yuan, W.; Xu, F.; Muhammad Aslam, M.; Miao, R.; Li, Y.; Wang, Q.; Li, X.; Zhang, X.; Zhang, K. The genome evolution and low-phosphorus adaptation in white lupin. *Nat. Commun.* **2020**, *11*, 1–13. 549–551
75. Ksiazkiewicz, M.; Nazzicari, N.; Yang, H.; Nelson, M.N.; Renshaw, D.; Rychel, S.; Ferrari, B. et al. A high-density consensus linkage map of white lupin highlights synteny with narrow-leaved lupin and provides markers tagging key agronomic traits. *Sci. Reps.* **2017**, *7*, 15335. 552–554
76. Lin, R.; Renshaw, D.; Luckett, D.; Clements, J.; Yan, G.; Adhikari, K.; Buirchell, B. et al. Development of a sequence-specific PCR marker linked to the gene “pauper” conferring low-alkaloids in white lupin (*Lupinus albus* L.) for marker assisted selection. *Mol. Breed.* **2009**, *23*, 153–161. 555–557
77. Young, N.D.; Udvardi, M. Translating *Medicago truncatula* genomics to crop legumes. *Curr. Opin. Plant Biol.* **2009**, *12*, 193–201. 558–559
78. von Wettberg E.J.; Porter, S.S.; Moriuchi, K.S.; Mukherjee JR. 2020 *Medicago* as a model forage and ecological model. Chapter 2 In *The Model Legume Medicago truncatula*, (ed. de Bruijn, FJ). Wiley Publishers, NJ USA. 560–562
79. Young, N.D.; Debelle, F.; Oldroyd, G.E.; Geurts, R.; Cannon, S.B.; Udvardi, M.K.; Benedito, V.A.; Mayer, K.F.; Gouzy, J.; Schoof, H.; Van de Peer, Y. 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature*, **2011**, *480*, 520–524. 563–565
80. de la Vega M.P.; Torres, A.M.; Cubero, J.I.; Kole C, editors. Genetics, genomics and breeding of cool season grain legumes. Eds. CRC Press: Boca Raton, FL, USA, 2011; pp. 98–150. 566–567
81. Ramsay, L.; Koh, C.; Konkin, D.; Cook, D.; Penmetsa, V.; Dongying, G.; Coyne, C.; Humann, J.; Kaur, S.; Dolezel, J.; Bett, K.E. *Lens culinaris* CDC Redberry genome assembly v2. 0.. Available online: <https://knowpulse.usask.ca/genome-assembly/Lcu.2RBY> (accessed on 7 April 2021) 568–570
82. Ramsay, L.; Koh, C.S.; Kagale, S.; Gao, D.; Kaur, S.; Haile, T.; Gela, T.S.; Chen, L.A.; Cao, Z.; Konkin, D.J.; Toegelová, H. Genomic rearrangements have consequences for introgression breeding as revealed by genome assemblies of wild and cultivated lentil species. *bioRxiv*. 2021 Jan 1. 571–573
83. Hirsch, C.N.; Foerster, J.M.; Johnson, J.M.; Sekhon, R.S.; Muttoni, G.; Vaillancourt, B. et al. Insights into the maize pan-genome and pan-transcriptome. *Plant Cell* **2014**, *26*, 121–35. 574–575
84. Guerra-Garcia, A.; Haile, T.; Ogutcen, E.; Bett, K.E.; von Wettberg, E.J. An evolutionary look into the history of lentil reveals unexpected diversity. *Evolutionary Applications*, **2022**, *15*, 1313–1325. 576–577
85. Ogutcen, E.; Ramsay, L.; von Wettberg, E.B.; Bett, K.E., 2018. Capturing variation in *Lens* (Fabaceae): Development and utility of an exome capture array for lentil. *Appl. Plant Sci.* **2018**, *6*, e01165. 578–579
86. Dahl, W.J.; Foster, L.M.; Tyler, R.T. Review of the health benefits of peas (*Pisum sativum* L.). *Br J Nutr* **2012**, *108*, S3–S10. 580–581

87. Zohary, D.; Hopf, M. Domestication of pulses in the Old World: legumes were companions of wheat and barley when agriculture began in the Near East. *Science* **1973**, *182*, 887–894 .
88. Gizlice, Z.; Carter, T. E. Jr.; Burton, J. W. Genetic diversity patterns in North American public soybean cultivars based on coefficient of parentage. *Crop Sci.* **1996**, *36*, 753–765.
89. Hyten, D. L.; Song, Q.; Zhu, Y.; Choi, I. Y.; Nelson, R. L.; Costa, J. M.; Specht, J. E.; Shoemaker, R. C.; Cregan, P. B. Impacts of genetic bottlenecks on soybean genome diversity. *Proc. Natl. Acad. Sci.* **2006**, *103*, 16666–16671.
90. Hammer, K. The domestication syndrome (in German). *Kulturpflanze*, **1984**, *32*, 11–34.
91. Gepts, P. Crop domestication as a long-term selection experiment. *Plant Breed. Rev.* **2010**, *24*, 1–44.
92. Milla, R.; Osborne, C.P.; Turcotte, M.M.; Violle, C. Plant domestication through an ecological lens. *Trends Ecol. Evol.* **2015**, *30*, 463–469.
93. Smýkal, P.; Nelson, M.N.; Berger, J.D.; Von Wettberg, E.J. The impact of genetic changes during crop domestication. *Agronomy*, **2018**, *8*, 119.
94. Hufford, M.B.; Berny Mier Y Teran, J.C.; Gepts, P. Crop Biodiversity: An Unfinished Magnum Opus of Nature. *Annu. Rev. Plant Biol.* **2019**, *70*, 727–751.
95. Ladizinsky, G. Seed dispersal in relation to the domestication of Middle East legumes. *Economic Botany*, **1979**, *33*, 284–289.
96. Ogutcen, E.; Pandey, A.; Khan, M.K.; Marques, E.; Penmetsa, R.V.; Kahraman, A.; Von Wettberg, E.J. Pod shattering: a homologous series of variation underlying domestication and an avenue for crop improvement. *Agronomy*, **2018**, *8*, 137.
97. Balarynová, J.; Klčová, B.; Sekaninová, J.; Kobrlová, L.; Cechová, M.Z.; Krejčí, P.; Leonova, T.; Gorbach, D.; Ihling, C.; Smržová, L.; Trněný, O. The loss of polyphenol oxidase function is associated with hilum pigmentation and has been selected during pea domestication. *New Phytol.* **2022**, *235*, 1807–1821.
98. Kluyver, T.A.; Charles, M.; Jones, G.; Rees, M.; Osborne, C.P. Did greater burial depth increase the seed size of domesticated legumes?. *J. Expt. Bot.* **2013**, *64*, 4101–4108.
99. Berger, J.D.; Shrestha, D.; Ludwig, C. Reproductive strategies in Mediterranean legumes: trade-offs between phenology, seed size and vigor within and between wild and domesticated *Lupinus* species collected along aridity gradients. *Front. Plant Sci.* **2017**, *8*, 548.
100. Salinas, A.D.; Bonet, A.; Gepts, P., 1988. The Wild Relative of *Phaseolus Vulgaris* in Middle America. In Genetic resources of *Phaseolus* beans (pp. 163–184). Springer, Dordrecht.
101. Fernandez, A.R.; Sáez, A.; Quintero, C.; Gleiser, G.; Aizen, M.A. Intentional and unintentional selection during plant domestication: herbivore damage, plant defensive traits and nutritional quality of fruit and seed crops. *New Phytol.* **2021**, *231*, 1586–1598.
102. Weeden, N.F. Genetic changes accompanying the domestication of *Pisum sativum*: is there a common genetic basis to the ‘domestication syndrome’ for legumes?. *Ann. Bot.* **2007**, *100*, 1017–1025.
103. Abbo, S.; van-Oss, R.P.; Gopher, A.; Saranga, Y.; Ofner, I.; Peleg, Z. Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends Plant Sci.* **2014**, *19*, 351–360.
104. Tay Fernandez, C.G.; Nestor, B.J.; Danilevicz, M.F.; Gill, M.; Petereit, J.; Bayer, P.E.; Finnegan, P.M.; Batley, J.; Edwards, D. Pangenomes as a Resource to Accelerate Breeding of Under-Utilised Crop Species. *Int. J. Mol. Sci.* **2022**, *23*, 2671.
105. Jain, M.; Koren, S.; Miga, K.H.; Quick, J.; Rand, A.C.; Sasani, T.A.; Tyson, J.R.; Beggs, A.D.; Dilthey, A.T.; Fiddes, I.T.; Malla, S. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* **2018**, *36*, 338–45.
106. Wenger, A.M.; Peluso, P.; Rowell, W.J.; Chang, P.C.; Hall, R.J.; Concepcion, G.T.; Ebler, J.; Functammasan, A.; Kolesnikov, A.; Olson, N.D.; Töpfer, A. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **2019**, *37*, 1155–62.