Advancing Grain Legumes Domestication and Evolution Studies with Genomics

Hailin Zhang\(^1\), Martin Mascher\(^1\), Shahal Abbo\(^3\) and Murukarthick Jayakodi\(^1\)*

\(^1\)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstraße 3, Gatersleben, Seeland 06466, Germany
\(^2\)German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstraße 4, Leipzig 04103, Germany
\(^3\)The Levi Eshkol School of Agriculture, The Hebrew University of Jerusalem, POB 12, Rehovot 7610001, Israel

*Corresponding author: E-mail, jayakodi@ipk-gatersleben.de
(Received 19 January 2022; Accepted 9 May 2022)

Grain legumes were domesticated in parallel with cereals in several regions of the world and formed the economic basis of early farming cultures. Since then, legumes have played a vital role in human and animal diets and in fostering agrobiodiversity. Increasing grain legume cultivation will be crucial to safeguard nutritional security and the resilience of agricultural ecosystems across the globe. A better understanding of the molecular underpinnings of domestication and crop evolution of grain legumes may be translated into practical approaches in modern breeding programs to stabilize yield, which is threatened by evolving pathogens and changing climates. During recent decades, domestication research in all crops has greatly benefited from the fast progress in genomic technologies. Yet still, many questions surrounding the domestication and diversification of legumes remain unanswered. In this review, we assess the potential of genomic approaches in grain legume research. We describe the centers of origin and the crucial domestication traits of grain legumes. In addition, we survey the effect of domestication on both above-ground and below-ground traits that have economic importance. Finally, we discuss open questions in grain legume domestication and diversification and outline how to bridge the gap between the preservation of historic crop diversity and their utilization in modern plant breeding.

Keywords: Anti-nutrition • domestication • domestication syndrome • genomics • grain legume • nutrition • symbiosis

Introduction

Legumes (Fabaceae) are economically the most important to global agriculture after the grasses (Poaceae) (Smykálová et al. 2015). The grain legumes are members of the family Fabaceae harvested as dry seeds, while other members are grown for vegetables, forage and other uses. Based on the growth seasons, grain legumes are also referred to as cool-season legumes such as pea (Pisum sativum), lentil (Lens culinaris) and faba bean (Vicia faba) and warm-season legumes like soybean (Glycine max), common bean (Phaseolus vulgaris), peanut (Arachis hypogaea) and Vigna spp. Since grain legumes are rich in protein, dietary fiber, vitamins and minerals (Mudryj et al. 2014), they have been used for human nutrition and livestock feeding for millennia. Legumes fix atmospheric nitrogen through symbiotic rhizobial bacteria (Oldroyd et al. 2011), reducing the need for inorganic nitrogen (N) fertilizers. In rotation systems, they can improve the yield of other crops such as cereals (Cernay et al. 2018). Against the backdrop of climate change and global food demand, increasing grain legume production is crucial to safeguard food and nutritional security without losing agro-biodiversity. Herein, plant domestication refers to the transformation of wild taxa to forms amenable to profitable human husbandry. Understanding the genetic basis of plant domestication can boost modern crop adaptation and improvement (Abbo et al. 2009, 2014). Over several decades, untangling the impact of natural selection and the evolutionary history of domesticated plants has been a major research focus of geneticists (Purugganan 2019). In the Near East, the Neolithic founder legumes include pea, lentil, better vetch (Vicia ervilia), chickpea (Cicer arietinum) and presumably faba bean, which were first domesticated together with the founder cereals approximately 10,500 years ago (Abbo et al. 2005, 2009, Gopher et al. 2021).

Hitherto, domestication has been studied more intensely in cereals than in legumes (Abbo et al. 2009). On the one hand, the lack of genomic resources has posed an obstacle for evolutionary research; on the other hand, some grain legumes such as better vetch are considered as neglected or ‘orphan’ crops and have received little attention from breeders and researchers. Classically, two complementary lines of evidence including archeological records and genetic data are routinely applied to study domestication. Archaeobotany probes plant remains from multiple archeological sites with regard to their origin in time and space to infer the timing of domestication and to discover crop dispersal routes (Caracuta et al. 2016, Mascher et al. 2016).
However, the archeological record is always fragmentary, posing obstacles to the accurate interpretation of the origin and spread of domesticated crops (Kantar et al. 2017). By contrast, the genetic approach employs top-down and bottom-up methods to trace the genetic footprints of domestication and diversification using extant individuals (Ross-Ibarra et al. 2007). The top-down method uses quantitative trait locus (QTL) mapping and genome-wide association studies (GWASs), both of which require phenotypes. In addition, the bottom-up approach is to identify selection signatures by querying extant genetic variation data, which does not need phenotypes.

Over the years, increases in throughput capacity and the reduction in the cost of DNA sequencing have led to the widespread adoption of genomic approaches in domestication and crop evolution studies. In this review, we outline a workflow to study legume domestication using genomics. We then describe the key domestication traits in grain legumes and their centers of origin. We review how genomics has been employed to discover genetic determinants of domestication and diversification traits in grain legumes. Additionally, we review the effects of domestication on grain legume metabolic and symbiosis traits. Finally, we discuss the differences between grain legumes and cereals and provide future directions to study grain legume domestication and evolution and how this knowledge can be translated into improved crop varieties.

Roadmap to Study Legume Domestication with Genomics

Genomics plays an essential role in identifying genes involved in the domestication and diversification of crops (Schreiber et al. 2018). Typically, domestication studies can benefit from reference genome assembly, genome-wide genotyping of large germplasm collections comprising wild germplasm, landraces and elite genotypes and biparental mapping populations derived from wild x domesticated genotypes (Fig. 1). A reference genome assembly is a prerequisite for discovering molecular markers across chromosomes and performing genetics and population genomic analysis. De novo genome assembly for single or multiple domesticated and wild accessions is now realistic for many crops, owing to the cost reduction and the development of highly accurate long-read sequencing technology such as PacBio HiFi (Wenger et al. 2019). Genotyping can employ reduced representation approaches such as genotyping-by-sequencing (GBS) and whole-genome shotgun sequencing (WGS) that affords orders of magnitude more markers than GBS. Reduced representation approaches may provide sufficient resolution for a top-down approach, where biparental mapping populations or diversity panels are used for QTL mapping and GWAS, respectively. However, GBS has less power in tracking local evolutionary signatures in bottom-up approaches. In the past, it was often necessary to strike compromises between panel size and sequence coverage for WGS in crops with large genomes, i.e. to decide whether to sequence more samples at low coverage or fewer samples at high coverage. Also, gene-centric complexity reduction methods such as exome capture and RNA-seq were commonly applied. This dilemma persists in principle, but the cost decreases in recent years have made medium- to high-coverage WGS of panels comprising hundreds of diverse accessions feasible even in species with large genomes and rendered exome capture cost-ineffective as it entails much higher cost for library preparation than WGS. Advances in genotype imputation tools and improved statistical methods permit the sequencing of hundreds to thousands of genotypes at low sequence coverage (1×–5× coverage). The missing data points in the variant matrix can be imputed with computational tools such as BEAGLE (Browning et al. 2021) or IMPUTE2 (Howie et al. 2009).

We propose that population genomic studies aiming to answer key questions in crop evolution should generate WGS data for wild and domesticated (landraces and elites) cultivars. With these genomic data in hand, analyses can shed light on the crop’s center of diversity, determine whether it had one or multiple genetically independent origins, and unravel dispersal routes. The comparisons of wild and landrace populations will pinpoint the genomic regions that were selected ever since domestication until the onset of modern plant breeding; the comparisons between landrace and elite populations can expose the regions targeted by breeders in the past century. Population genomic analysis can also identify key events and trajectories in the demographic history of crops (Doebley et al. 2006). These include (but are not limited to) an approximate timing of domestication and split times of subpopulations. Various demographic models such as the site frequency spectrum in ‘fastsimcoal2’ (Excoffier et al. 2021), joint frequency spectrum in ‘dadi’ (Gutenkunst et al. 2009), pairwise sequentially Markovian coalescent model (PSMC), etc. (Terhorst et al. 2017) have been developed to infer population size history. One caveat is that PSMC and similar methods may require high sequence coverage (~10–30×) (Nadachowska-Brzyska et al. 2016) to resolve heterozygosity. In the case of self-pollinating species, it must be noted that PSMC and similar tools were not designed with inbreeding and homozygous genomes in mind. Care must be taken when interpreting hybrids between populations: e.g., recent crop-wild hybrid derivatives (Campbell et al. 2016) may be mistaken as ancestral forms as they are expected to occupy intermediate positions between wild and domesticated forms.

Dense genomic marker data from WGS enable us to infer footprints of selection (Hermisson and Pennings 2017, Alachiotis and Pavlidis 2018). The signals most commonly sought for are selective sweeps, evolutionary events that gave rise to large blocks (hundreds of kilobases to tens of megabases) of linked polymorphisms that are fixed or segregating at high frequency (haplotype blocks) in the domesticated form. In the wild progenitor, by contrast, blocks of linked variants are shorter and more diverse, with many of them segregating at low or medium frequencies (Hermisson and Pennings 2017). Ancient DNA research, also known as archeogenomics, allows us to access directly past genetic diversity that may have been lost
during evolution and to support time estimates of the domestication process and expose molecular diversity in key genes during the domestication episode or of ancient/extinct crop forms (Mascher et al. 2016). The key conceptual drawback of ancient DNA studies is the haphazard nature of their design: it requires a fair share of good luck to unearth ancient plant remains that contain enough DNA for molecular analysis. Many regions of the world, e.g., wet temperate climate zones or the tropics, are not conducive to the preservation of DNA. Even if well-preserved ancient samples are available, technical challenges need to be overcome. The level of DNA degradation is a major limitation in ancient DNA and thus the extracted DNA molecules tend to be short (<50 bp) (Sawyer et al. 2012). In addition, deamination of cytosine to uracil occurs frequently at the ends of these fragments, which impedes the analysis of sequenced fragments while also confirming the authenticity of the material (Sawyer et al. 2012). Nevertheless, efficient protocols for ancient DNA extraction and subsequent sequencing library preparation have been developed (Gamba et al. 2016, Gansauge et al. 2017). If sufficient amounts of ancient DNA sequence have been obtained, a reference genome sequence and dense genome-wide marker data (WGS or exome capture) of a diversity panel of extant genotypes are needed to properly contextualize ancient sequences. Besides, complementary datasets such as transcriptome sequencing (RNA-Seq), epigenetic data and the inclusion of representative genome assemblies from wild and domesticated gene pools can assist in the discovery of alleles or genes related to domestication and crop improvement.

**Domestication of Legumes**

**Domestication traits**

Domesticated plants are often distinguished from their wild progenitors by several morphological, biochemical and developmental traits, collectively referred to as the domestication syndrome (Harlan et al. 1973, Hammer 1984). The genetic makeup of crop plants continues to evolve post domestication; hence, not all phenotypic differences between wild and domesticated forms are by necessity domestication traits. Successive genetic changes after domestication episodes are common in crop plants when they colonize new environments or adapt to new needs of their cultivators, giving rise to new phenotypes, some of them as striking as interspecific differences. The concept of ‘crucial domestication traits’ was proposed (Abbo et al. 2014) to distinguish pristined domestication traits from crop evolution (improvement-related) traits. Often, crucial domestication traits are critical for profitable cultivation and show a clear dimorphism between wild and domesticated plants. On the other hand, improvement-related traits mostly display quantitative phenotype variations among wild and domesticated gene pools (Abbo et al. 2014). Such improvement footprints are the consequence of post-domestication crop diversification. In grain legumes, pod shattering/dehiscence, seed dormancy and seed size are frequently viewed as domesticated traits. But, seed dormancy is the only trait that is determined to be a crucial domestication trait in the cool-season Near Eastern legumes (Abbo et al. 2014). A certain degree of pod shattering was observed quantitatively not only in cool-season legumes such as pea, lentil, chickpea and lupin (Lupinus spp.) but also in warm-season legumes including soybean, common bean and Vigna spp. (Parker et al. 2021b). Therefore, pod indehiscence is likely to be an improvement trait that arose after domestication. Likewise, a continuum of seed sizes is seen in wild and domesticated grain legumes of the Near East, indicating that seed size is an improvement phenotype that was shaped by farmers’ preferences after domestication (Abbo et al. 2014). Other improvement traits of grain legumes involve phenology, shape of grains, nutritional quality and taste (Vaz Patto et al. 2015).

**Domestication centers**

Grain legume domestication took place independently in several distinct geographical locations (Fig. 2), also known as ‘domestication centers’ and dispersed via human migration.
Fig. 2 The origin and domestication of legume species. Marked regions indicate approximate locations of leguminous crop origin and domestication.

...exchange of ideas), e.g. in the case of common bean that was introduced to Europe from Mesoamerica (Gepts 2003, Angioi et al. 2010). Approximately 10,000 years ago, the Neolithic founder legumes including lentil, pea, chickpea and bitter vetch were domesticated in a region near the upper reaches of the Tigris and Euphrates rivers in current southeastern Turkey/northern Syria, part of the so-called Fertile Crescent (Southwest Asia) (Lev-Yadun et al. 2000, Abbo and Gopher 2017). Common bean was initially thought to originate from Northern Peru and Ecuador (South America). Recent genomic evidence supports two independent parallel domestication events (one in Mesoamerica and the other in the Andes), which both occurred between 8000 and 10,000 years ago (Gaut 2014, Rendón-Anaya et al. 2017, Singh et al. 2019). Similarly, two domestication processes were proposed for lima bean, one in Mesoamerica and one in the Andes. Application of GBS approach in wild and domesticated lima bean supported both domestication scenarios (Chacón-Sánchez and Martínez-Castilho 2017). The origin of soybean had been a matter of debate for a long time, but the development of genomic data traced its origin to the Huang-Huai Valley in Central China, the region between the Great Wall and Qinling Mountains, including both sides of the Yellow River (Han et al. 2016, Sedivy et al. 2017), where soybean was domesticated between 5000 and 9000 years ago. It is commonly held that mungbean (Vigna radiata), a crop with very short life cycle (70–80 days from seed to seed) originated in India 4000–6000 years ago (Fuller 2007) and then spread across Asia and Africa (Kim et al. 2015). In 1929, German breeders began to domesticate narrow-leafed lupin (Lupinus angustifolius) as the wild lupins have high levels of alkaloids that make their seeds toxic for direct human and livestock consumption. Intriguingly, within only 40 years’ time, narrow-leafed lupin was transformed into an established crop (Hondelmann 1984, Wang et al. 2021) and is now grown as a protein-rich feed crop in Europe and Australia.

Charting Grain Legume Domestication by Genomics

Genetic mapping of domestication traits (top-down approach)

Pod shattering. The loss of pod dehiscence or shattering is one of the key traits that humans selected for in domesticated grain legumes. Pod shattering is imperative for seed dispersal in the wild but impedes harvest and lowers net yields under a cultivation regime. Environmental factors such as high temperatures and low relative humidity also influence pod shattering (Zhang et al. 2018). Unraveling the genetic basis of pod shattering in grain legumes has long been a major breeding and research target. To date, pod shattering is the best characterized trait associated with domesticated grain legumes (Table 1). Among grain legumes, pod shattering has been studied in the greatest detail in soybean. Funatsuki and his colleagues (Funatsuki et al. 2014) first cloned the major-effect Pod dehiscence 1 (Pdh1) gene, which encodes for a dirigent-type protein. However, the lack of genomic tools hampered the deployment of this gene in breeding to develop pod indehiscient varieties in the past. The construction of the soybean reference genome sequence (Schmutz et al. 2010) and the advent of next-generation sequencing enabled the development of a high-density genotyping assay, the SoySNP50K SNP chip (Song et al. 2013). Subsequently, trait-associated molecular markers have been identified to breed shatter-resistant soybean varieties by genomics-assisted breeding (Miranda et al. 2019). Additionally, the SHATTERING1-5 (SHAT1-5) gene was identified to be involved in the reduction of pod dehiscence in...
soybean (Dong et al. 2014). Recently, an additional gene NST1A (a paralog of SHAT1-S, also known as NST1B) was detected using a GWAS scan (Zhang and Singh 2020). Moreover, the shattering-related candidate gene (Glyma09g06290) was identified by GWAS across multiple environments, and diagnostic markers for the gene were also developed to introgress it into shattering varieties (Hu et al. 2019).

In common bean (dry beans), a major QTL PvPdh1 (ortholog of soybean Pdh1) associated with shattering resistance was identified (Parker et al. 2020). Subsequently, molecular markers have also been developed to help introgress it into elite varieties for reduced shattering (Parker et al. 2021a). Additionally, a large-effect QTL PvMYB26 was mapped using a biparental population derived from a cross between the Andean snap bean cv. Midas and the wild Mesoamerican bean (G12873) (Rau et al. 2019). Recently, this locus was narrowed to a 22.5-kb region using an introgression line population and following transcriptome sequencing pinpointed the potential candidate for pod indehiscent (Di Vittori et al. 2020). Similar efforts were undertaken to identify QTL for pod shattering in other warm-season legumes, including CPShat3 and CPShat5 in cowpea (Vigna unguiculata) (Lo et al. 2018) and VaMYB26 in adzuki bean (Vigna angularis) (Takahashi et al. 2020). In each case, the availability of a reference genome has facilitated the discovery of QTL and candidate genes for pod shattering. Less is known about the genetic regulation of pod shattering in cool-season legumes. In pea, Dpo or Dpo1 was found to be a major factor regulating pod shattering (Ladizinsky 1979). The ortholog of Pdh1 in soybean and common bean may be involved in pod shattering also in chickpea (Aguilar-Benitez et al. 2020). Two major mutations (pod1 and pod2) associated with pod dehiscence occur in grass pea (Lathyrus sativus) (Talukdar 2011). Genetic mapping of pod shattering has not yet been done in the Vicia tribe, a research gap likely attributable to a lack of reference genomes for Vicia species. Very recently, reference genomes were generated for important grain legumes including pea and lentil (Table 2) (Kreplak et al. 2019, Ramsay et al. 2021). These new genomic resources are expected to help genetic fine mapping and cloning genes. Parallel patterns of selection have been discovered for pod shattering among grain legume species (Table 1) (Parker et al. 2021b). The orthologous gene of pod shattering can be searched in other grain legumes and examined for pod shattering resistance (Yundaeng et al. 2019). Such an evolutionary informed gene characterization may underpin the reduction of yield losses in other minor legumes such as tepary bean (Phaseolus acutifolius) and hyacinth bean (Lablab purpureus) that lack resources and research. Hence, a comparative genomics study at the phylogenetic level can illuminate whether other domestication traits have evolved in parallel.

Seed dormancy. Seed dormancy is a crucial domestication trait in certain crop plants. In the wild, dormant seeds delay germination and promote survival and fitness via the buildup of soil seed banks (Finch-Savage and Footitt 2017). Seed dormancy in a domesticated crop may cause asynchronous germination, lower crop performance and ultimately, severely limits yields (Abbo et al. 2008, 2011). The physiological mechanisms and the molecular genetics behind seed dormancy are not well-studied in grain legumes. Physical dormancy, namely impermeable seed coats, seems to play a bigger role than physiological dormancy, i.e. inhibition of embryo development (Martin 1946). Major-effect loci controlling seed dormancy were identified in lentil (Ladizinsky 1985), blue lupine (L. angustifolius) (Forbes and Wells 1968), cowpea (Kongjaimun et al. 2012), rice bean (Vigna umbellata) (Isemura et al. 2010) and mungbean (V. radiata) (Isemura et al. 2012). In pea and common vetch, two QTLs were mapped for seed dormancy (Donnelly et al. 1972, Weeden 2007). Nonetheless, these studies have used a limited number of molecular markers from low-throughput and polymerase chain reaction-based simple-sequence repeats that restrict fine mapping to narrow genetic intervals and subsequent causal gene discovery. But in common bean, the development of reference genome and resequencing data allowed mapping the seed dormancy locus within a 118-kb genomic interval with a potential candidate gene (Soltani et al. 2021). The first cloned seed dormancy gene in legumes was GmHs1-1, a soybean gene that encodes a calcineurin-like metallophosphoesterase transmem-

---

**Table 1 Pod shattering genes in grain legumes.**

| Legumes    | Gene/QTL     | Gene product                                           | Chromosome | Cloned? | References |
|------------|--------------|--------------------------------------------------------|------------|---------|------------|
| Soybean    | Pdh1         | Dirigent-type protein                                   | Gm16       | Yes     | (Funatsuki et al. 2014) |
| Soybean    | SHAT1-S      | NAC transcription factor                                | Gm16       | Yes     | (Dong et al. 2014) |
| Soybean    | NST1A        | NAC transcription factor                                | Gm07       | No      | (Zhang and Singh 2020) |
| Soybean    | Glyma09g06290| bHLH (basic helix-loop-helix) gene family              | Gm09       | No      | (Hu et al. 2019) |
| Common bean| PvPdh1       | Dirigent-like gene                                      | Pv03       | No      | (Parker et al. 2020) |
| Common bean| PpMYB26      | MYB transcription factor                                | Pp05       | No      | (Rau et al. 2019) |
| Cowpea     | CPShat3 (VuNAC002) | NAC domain transcription factor                          | Vu03 and Vu05 | No | (Lo et al. 2018) |
| Adzuki bean| VaMYB26      | MYB domain protein                                      | Va07       | No      | (Takahashi et al. 2020) |
| Pea        | Dpo or Dpo1  | Unknown                                                 | LG III     | No      | (Weeden et al. 2002) |
| Lentil     | Pi           | Unknown                                                 | LG IV      | No      | (Ladizinsky 1979) |
| Chickpea   | PDH1         | Dirigent-like protein                                   | LG6        | No      | (Aguilar-Benitez et al. 2020) |
| Grass pea  | pod1 and pod2| NAC transcription factor                                | Ltr VII    | No      | (Talukdar 2011) |
|            |              |                                                        |            |         |             |

NAC: NAM (no apical meristem), ATAF1/2, and CUC2 (cup-shaped cotyledon).
MYB: Myeloblastosis.
| Common name                  | Latin name                  | Haploid genome size (Mb) | Sequencing technologies          | Assembly size (Mb) | Ploidy level | Reproduction system | References                      |
|-----------------------------|-----------------------------|--------------------------|----------------------------------|--------------------|--------------|--------------------|---------------------------------|
| Adzuki bean                 | V. angularis               | 612 Mb                   | Illumina + 454 XLR               | 443                | 2n = 2x = 14  | Selfing            | (Kang et al. 2015)              |
| Andean lupin, pearl lupin   | Lupinus mutabilis          | 1 Gb                     | 454XLR, Pacbio                   | 523                | 2n = 2x = 48  | Selfing            | (Sakai et al. 2015)            |
| Bambara                     | Vigna subterranee          | 550 Mb                   | Illumina                         | 512                | 2n = 2x = 22  | Selfing            | (Chang et al. 2019)            |
| Beach peas/beach cowpea     | Vigna marina               | 365 Mb                   | Illumina                         | 365.6              | 2n = 2x = 22  | Selfing            | (Singh et al. 2019)            |
| Bitter vetch                | V. ervilia                 | 4.2 Gb                   | Illumina + Pacbio                | 519.8              | 2n = 2x = 16  | Selfing            | (Jain et al. 2013)             |
| Black gram                  | Vigna mungo                | 574 Mb                   | Illumina + 10X Genomics + HIC    | 499                | 2n = 2x = 22  | Selfing            | (Pootakham et al. 2020)        |
| Blue lupin (narrow-leaved)  | L. angustifolius           | 924 Mb                   | Illumina                         | 598                | 2n = 2x = 40  | Selfing            | (Yang et al. 2013)             |
| Chickpea                    | C. arietinum (cultivated)  | 740 Mb                   | Illumina                         | 609                | 2n = 2x = 16  | Selfing            | (Hane et al. 2017)             |
| Common bean/kidney beans/dry beans | P. vulgaris | 587 Mb | Illumina + 454 XLR + Sanger | 474.3              | 2n = 2x = 14  | Selfing            | (Schmutz et al. 2014)        |
| Common vetch                | V. sativa                  | 1.8 Gb                   | Illumina + Pacbio                | 549.6              | 2n = 2x = 14  | Selfing            | (Vlasova et al. 2016)          |
| Cowpea/yardlong bean        | V. unguiculata             | 613 Mb                   | Illumina + BioNano + HIC         | 519.4              | 2n = 2x = 14  | Selfing            | (Shirasawa et al. 2021)        |
| Faba beans/broad bean       | V. faba                    | 13 Gb                    | Pacbio + BioNano                 | 632.4              | 2n = 2x = 12  | Mixed mating       | (Xia et al. 2019)              |
| Grass pea                   | L. sativus                 | 6.2 Gb                   | Illumina + 454 XLR               | 623                | 2n = 2x = 14  | Mixed mating       | (Emmrich et al. 2020)          |
| Horse gram                  | Macarystoma uniflorus      | 343.6 Mb                 | Illumina                         | 529.2              | 2n = 2x = 14  | Selfing            | (Shirasawa et al. 2021)        |
| Hyacinth bean               | L. purpureus               | 367 Mb                   | Illumina + Pacbio                | 395                | 2n = 2x = 14  | Selfing            | (Chang et al. 2019)            |
| Jack bean                   | Canavalia gladiata         | 5 Gb                     | Illumina + 454 XLR               | 3760               | 2n = 2x = 14  | Selfing            | (Ramsay et al. 2021)           |
| Lentil                      | L. culinaris (cultivated)  | 4 Gb                     | Illumina + 10X Genomics + HIC    | 475                | 2n = 2x = 14  | Selfing            | (Lonardi et al. 2019)          |
| Lens ervoides (wild)        | Phaseolus lunatus          | 622 Mb                   | Illumina                         | 623                | 2n = 2x = 22  | Selfing            | (Garcia et al. 2021)           |
| Lima bean                   | Vigna aconitifolia         | 1.1 Gb                   | Illumina                         | 1052               | 2n = 2x = 14  | Selfing            | (Kang et al. 2014)             |
| Mungbean/Green gram         | V. radiata                 | 579 Mb                   | Illumina + 454 XLR               | 475                | 2n = 2x = 14  | Selfing            | (Chen et al. 2016)             |
| Pea/dry pea                 | P. sativum                 | 4.4 Gb                   | Illumina + Pacbio + BioNano + HIC| 3920               | 2n = 2x = 14  | Selfing            | (Kreplak et al. 2019)          |
| Peanut                      | Arachis duranensis         | 1.1 Gb                   | Illumina + Pacbio                | 1025               | 2n = 2x = 14  | Selfing            | (Bertioli et al. 2016)         |
| Peanut                      | Arachis ipaensis           | 1.4 Gb                   | Illumina                         | 1051               | 2n = 2x = 14  | Selfing            | (Lu et al. 2018)               |
| Peanut                      | A. hypogea (cultivated)    | 2.7 Gb                   | Illumina + Pacbio + BioNano + HIC| 1391               | 2n = 2x = 14  | Selfing            | (Bertioli et al. 2016)         |
Table 2 (Continued)

| Common name     | Latin name               | Haploid genome size | Sequencing technologies         | Assembly size (Mb) | Ploidy level | Reproduction system | References                  |
|-----------------|--------------------------|---------------------|----------------------------------|--------------------|--------------|---------------------|-----------------------------|
| Pigeonpea       | *Arachis monticola* (wild) | 852 Mb              | Illumina + Pacbio + BioNano HiC | 2620               | 2n = 2x = 22 | Mixed mating        | (Yin et al. 2018)           |
| Cajanus cajan   |                          |                     |                                   |                    |              |                     |                             |
| Rice bean       | *V. umbellata*           | 562 Mb              | Illumina + Pacbio                | 471                | 2n = 2x = 22 | Selfing            | (Schnitz et al. 2010)       |
| Soybean         | *G. max* (cultivated)    | 1.1 Gb              | Sanger                           | 1115               | 2n = 2x = 40 | Selfing            | (Li et al. 2014)            |
| Glycine soja (wild) |                    |                     | Illumina                         | 813–985            |              |                     | (Qi et al. 2014)            |
| Glycine soja (wild) |            |                     | Illumina                         | 868                |              |                     | (Shen et al. 2018)          |
| Soybean         | *G. soja* (wild)         |                     | Pacbio + BioNano + HiC           | 1025               |              |                     | (Xie et al. 2019)           |
| Soybean         | *G. max* (cultivated)    |                     | Pacbio + BioNano + HiC           | 1013.20            |              |                     | (Y. Liu et al. 2020)        |
| Soybean         | *G. soja* (wild)         |                     | Pacbio + BioNano + HiC           | 992.3–1,059.8      |              |                     | (Liu et al. 2018)           |
| Soybean         | *G. latifolia* (wild)    |                     | 10X Genomics                     | 939.3              |              |                     |                             |
| Tepary bean     | *P. acutifolius*         | 682 Mb              | Pacbio + HiC                     | 5126–661.8         | 2n = 2x = 22 | Selfing            | (Mohgaddam et al. 2021)     |
| White lupin     | *Lupinus albus*          | 451 Mb              | Pacbio + BioNano                 | 451                | 2n = 2x = 50 | Selfing            | (Hufnagel et al. 2020)      |
| Winged beans    | *Psophocarpus tetragonolobus* | 1.2 Gb        | Pacbio + HiC                     | 558.9              |              |                     | (Xue et al. 2020)           |
| Yellow lupin    | *L. luteus*              |                     | Pacbio + HiC                     |                    | 2n = 2x = 18 | Selfing            |                             |
brane protein (Sun et al. 2015). This achievement was facilitated by sequencing data from a mapping population that permit to fine map to a 22-kb region harboring two genes (Sun et al. 2015). Furthermore, with the aid of high-throughput genomic data, the green seed coat gene, also known as G allele, which is also responsible for seed dormancy was cloned, and this gene was found to have undergone parallel selection in different crop families (Wang et al. 2018). Likewise, the resequencing approach has identified two candidate genomic regions spanning 2.4 Mb and 0.74 Mb associated with seed dormancy in peanut (Kumar et al. 2019). The genetic linkage or co-localization of QTL controlling seed dormancy and pod shattering was also reported in lentil (Ladizinsky 1985) and common bean (Soltani et al. 2021), indicating that pleiotropy or tight linkage of several loci could be an important attribute of domestication (Harlan et al. 1973, Meyer and Purugganan 2013).

**Genome scans for footprints of selection (bottom-up approach)**

QTL scans aim at identifying genomic regions controlling traits hypothesized by researchers to be involved in the domestication syndrome. By contrast, bottom-up approaches are hypothesis-free in that they do not require predictions as to which phenotypes were selected for. While selection scans are phenotype-agnostic, the co-location of sweeps and QTL for domestication traits reinforces the confidence in either of them.

In soybean, a strong selective sweep spanning approximately 116 kb was spotted at the SHAT1-5 locus: all soybean landraces shared a single haplotype, highly indicative of strong recent selection (Dong et al. 2014). A similar signature was observed in 100 kb surrounding the Pvdh1 locus in common bean (Parker et al. 2021a). For seed dormancy, a 160-kb selective sweep covering the seed dormancy gene GmHs1-1 was detected in soybean using resequencing data (Sun et al. 2015). Furthermore, resequencing data from a panel of 302 accessions comprising wild germplasm, landraces and improved cultivars demonstrated domestication selection signals as well as post-domestication improvement traits including seed weight, seed coat color and oil content in soybean (Zhou et al. 2015). Zhou et al. (2015) pointed out that sweep scans can refine prior QTL mapping results. For instance, previous studies (Liu et al. 2007) had identified a pod shattering QTL spanning 12 Mb, while the selection scan of wild soybeans versus landraces narrowed down the interval to a 190-kb region containing only 14 genes. Refined intervals can guide the prioritization, validation and subsequent functional characterization of domestication genes. As in QTL mapping, the resolution of sweep scans is determined, among other things, by genome size, sequence diversity and recombination rates. Different contrasts in sweep scans (e.g. wild progenitor vs. landraces and landraces vs. elite cultivars) may give clues about the relative timing of selection. For example, Zhou et al. (2015) showed that seed size was targeted during the domestication process, while seed color was likely shaped by post-domestication variety selection. Selective sweeps can also tag regions involved in traits that are adaptive in both wild and domesticated forms but subject to different selective regimes in either taxon. For instance, Varshney et al. (2019) identified the selection signals possibly related to key biotic and abiotic stress tolerance in resequencing of 429 diverse chickpea accessions (Varshney et al. 2019).

**Effect of Domestication on Secondary Metabolites**

Grain legumes are not only an important source of proteins in the human diet, but they are also rich in secondary metabolites such as polyphenols, alkaloids and saponins (Gupta 1987). In the wild, these play a role in defense against herbivores and pathogens. In the domesticated form, secondary metabolites may either promote human health or act as anti-nutritional factors. Domestication and post-domestication improvement have altered contents of the essential primary and secondary metabolites in legumes. Understanding the underlying genetic factors will boost breeding and inform metabolic engineering to improve the nutritional value of current grain legumes. Genes involved in biosynthetic pathways for health-promoting metabolites that were lost in the domestication bottleneck may be re-introduced from extant wild relatives.

**Beneficial compounds**

Bioactive compounds in legumes may have beneficial effects in the prevention and treatment of chronic diseases in humans such as cardiovascular ailments, diabetes, cancer and neurodegenerative diseases (Guadaoui et al. 2020). For example, legumes produce 3,4-dihydroxyphenyl-L-alanine, which is used in the treatment of Parkinson’s disease (Etemadi et al. 2018). An examination of wild and domesticated accessions supported a selection pressure for higher tryptophan levels in chickpea (Kerem et al. 2007). Consistent with intense breeding for high oil content in soybean, genes regulating fatty acid biosynthesis showed footprints of selection in soybean when compared to the wild progenitors (Zhou et al. 2015) and several genes involved in fatty acid biosynthesis were co-located with oil content QTLs (Iqbal et al. 2020). Carotenoids are considered as antioxidants in food systems. A study that investigated the 10 most important grain legumes [peanut, chickpea, soybean, vetch, lentil, lupin, common bean, pea, faba bean (unknown progenitor) and cowpea] and their wild relatives have found, on average, 48% lower carotenoid content, including lutein and zeaxanthin in domesticated legumes (Fernández-Marín et al. 2014). Protein quality is determined by the essential amino acid composition, digestibility and bioavailability of individual amino acids (Vaz Patto 2016). High-quality protein legumes may play an important role in a future human diet less reliant on carbon-intense meat production (Poore and Nemecek 2018). However, little is known about the genetic basis of protein quality (in terms of improved amino acid composition and digestibility) in grain legumes. A large-scale screening of genetic resources is needed to identify natural variation for protein quality and
breed for improved protein content and quality. Genotyping of large germplasm collections has become feasible with contemporary technology like GBS or skim sequencing (Milner et al. 2019). Therefore, efforts should be made to fingerprint diverse gene pools and explore variation for domestication and improvement-related traits to meet current nutritional requirements for human beings and livestock.

**Anti-nutritional compounds**

Some of the secondary metabolites in grain legumes cause indigestibility, making them unpalatable to humans and animals. Such secondary metabolites are known as anti-nutritional factors. Examples are tannins, vicine and convicine (Kumar et al. 2021). The color of the seed coat and seed tissues is primarily controlled by secondary metabolites including beneficial carotenoids, anthocyanin and antinutrient compounds—tannins (Espinosa-Alonso et al. 2006). Changes in seed coat color are commonly observed in domesticated and improved legumes as compared to their wild progenitors. For instance, the wild common beans have colored seed coats, while elite varieties are predominantly black seeded (McCLean et al. 2018). A survey of thousands of wild and domesticated soybeans illustrated that all wild soybeans have black seed coats whereas the domesticated soybeans have a range of seed coat colors including yellow, brown and green (Jeong et al. 2019), but the modern cultivars are predominantly yellow (also called colorless or white). For many decades, selective breeding efforts have targeted to eliminate anti-nutritional factors. The bitter taste caused by alkaloids has been intentionally bred out of legumes. In yellow lupin (Lupinus luteus), the domesticated cultivar has much lower alkaloid content than the wild form (Iqbal et al. 2020). Similarly, faba bean breeding has largely focused on the development of anti-nutrition-free (such as tannins and low vicine and convicine) cultivars. With an integrative analysis of transcriptomes and metabolomes, the key regulatory gene involved in vicine and convicine pathway was disclosed (Björnsdóttir et al. 2021). Phytic acid (myo-inositol-1,2,3,4,5,6-hexakisphosphate) is the major storage form of phosphorus in grain legumes such as soybean and common bean. Its poor digestibility and chelation of mineral cations (such as iron, zinc, potassium, calcium and magnesium) reduce the nutritional availability (Erdman 1981). The early development of genome sequence and adoption of WGS had located the low-phytate genes in soybean (Maroof et al. 2009). Raffinose family oligosaccharides, which cause flatulence in humans and animals, are also abundant in the legume family (Elango et al. 2022). Genomic approaches have also facilitated the development of molecular markers associated with low-raffinose phenotype in soybean (Dierking and Bilyeu 2008). The identification of key genes associated with diverse secondary metabolite pathways is therefore imperative to optimize the target metabolite content in commercial cultivars. The availability of genomic resources accelerates such quality improvements in seed legumes. Yet, the reduction in secondary metabolites in domesticated cultivars might make them susceptible to biotic stresses compared to their wild counterparts. It is also equally important to identify resistance alleles in the exotic material and stack them in modern cultivars.

**Effect of Domestication on Symbiosis**

Legumes are capable of establishing a root-nodule symbiosis with nitrogen-fixing soil bacteria called rhizobia, which greatly increases nitrogen availability for the plant host. This symbiosis plays a crucial role in developing sustainable farming systems. However, domestication and intense artificial selection for high yield with the use of chemical fertilizers may have disrupted the plant–microbe interaction, reducing the rate of bacterial nitrogen fixation or nitrogen uptake by the host (Porter and Sachs 2020, J. Liu et al. 2020). Plant cultivation often includes the use of manure or artificial fertilizers, so that domesticated forms might have become dependent on readily available soil nitrogen (Porter and Sachs 2020). The ancestral symbiosis has been retained but the host’s ability to favor the most efficient bacterial partners might have been compromised (Porter and Sachs 2020). A few small-scale studies have explored the effect of domestication and crop evolution on symbiosis. The investigation of host-range evolution using 80 isolates showed that only 34% of the strains were able to nodulate cultivated legume species while 89% were able to nodulate the other wild species, demonstrating that the domesticated species are loosing their interaction with beneficial bacteria as compared with wild legumes (Mutch and Young 2004). It has been observed that the wild soybean (Glycine soja) can recruit more diverse Bradyrhizobium strains in its rhizosphere than the cultivated soybean (Chang et al. 2019). Similarly, a nodulation experiment in chickpea indicated that the wild progenitor of chickpea (Cicer reticulatum) entered symbiotic relationships with more diverse Mesorhizobium than domesticated chickpea (Kim et al. 2014). This was ascribed to the loss of alleles of nod factor signaling genes under domestication in chickpea (Kim et al. 2014). In contrast, Greenlon et al. (2019) showed that the wild chickpeas had evolved symbiosis with specific bacteria whereas the domesticated chickpea exhibited symbiosis with diverse bacterial strains as they spread to new locations, indicating the co-evolution of rhizobia with chickpea domestication and diversification (Greenlon et al. 2019). Hence, the symbiosis might vary by crop and geography. Genomic approaches should be employed to elucidate pathways and genes associated with legume-rhizobial symbiosis in wild legumes to improve biological nitrogen fixation in domesticated legumes for sustainable agriculture. Furthermore, the latest advancement in sequencing has enabled large-scale metagenomics to understand plant–microbe interactions (Knief 2014).

**Conclusion and Future Perspectives**

Replacing meat-based protein diets with plant-based protein diets is critical to maintain the global food and nutritional sustainability (Shepon et al. 2018). One aspect of this transition
is the increased local production of protein-rich leguminous crops. In that respect, a better understanding of legume domestication and diversification is indispensable. The biological differences between legumes and cereals may have influenced the selection of potential plants by the first farmers (Abbo et al. 2009). Substantial biological differences were spotted between legumes and cereals including seed dormancy, seed dispersal and population structure, suggesting that hunter-gatherers interacted differently with wild legumes than with wild cereals (Abbo et al. 2009, 2009). Deeper insights on cereal and legume domestication will allow us to understand the biological as well as cultural processes associated with domestication and evolution. Additionally, such knowledge promotes crop improvements not only in cereals but also in grain legumes, which is important for achieving a resilient food system. However, to date, a comprehensive study that compares the domestication syndromes between cereal and legume species at a genome-wide level and with single-gene resolution is still missing. We attribute this situation to the lack of investment/research effort or genomic resources available for important grain legumes as compared to important cereals (maize, wheat and barley), all of which have had their genome sequenced in the last decade. For example, the pea and mung bean (Vicia sativa L.) have been targeted for domestication studies to understand the processes associated with domestication and evolution (Varshney et al. 2021). In parallel, the greater investment in major cereals and the development of high-throughput genomics resources enabled innovative solutions to increase crop yield and performance under different biotic and abiotic conditions (Varshney et al. 2021).

In legumes, it has been observed that disease-resistance-related genes were lost as a result of domestication and subsequent crop improvement (Zhou et al. 2015, Varshney et al. 2019). This implies that a reference genome from a single domesticated crop cannot capture all the resistance-conferring alleles and might cause reference bias in estimating genetic variation between wild and domesticated populations. The construction of a pan-genome (the universe of DNA sequences within a species) is advantageous to capture diverse alleles (Jayakodi et al. 2021). The cost reduction and recent development in accurate long-read sequencing technologies have simplified the genome assembly process in several crops (Jayakodi et al. 2021). Moreover, building a pan-genome with the inclusion of domesticated and wild species (i.e. primary gene pool) of a grain legume is profitable as it may capture genes or alleles for disease resistance, legume-rhizobial symbiosis and nutritional quality. Furthermore, pan-genome studies in crop plants demonstrated that past geographic range expansion and recent breeding have been accompanied by large structural variations such as inversions and presence/absence variants (Jayakodi et al. 2020, Hufford et al. 2021). In this vein, the pan-genome will become an important tool for grain legume research in order to discover resistance and adaptation genes in historic germplasm and deploy them in newly developed cultivars. Until now, among legumes, pan-genomes with chromosome-level assemblies have been constructed only for soybean (Y. Liu et al. 2020). Likewise, epigenetic variation, e.g., changes such as changes in DNA methylation or histone methylation cause phenotypic variations during plant evolution (Ding and Chen 2018). To date, comprehensive studies focusing on epigenetic changes associated with domestication are lacking in legumes. Very few efforts have been made to study the epigenetic changes in relation to abiotic stresses (Varotto et al. 2020). Like other research areas, the analysis of ancient DNA in legumes is still in its incipient stage. The discovery of 14,000-year-old faba bean-like seeds in the prehistoric site of el-Wad Terrace, Mount Carmel, Israel (Caracuta et al. 2016), suggests a Levantine origin of that crop. However, the charred state of seeds limits the application of genomics in ancient remains.

Lost beneficial alleles can be brought back from the wild to domesticated accessions via introgression. However, this method has been hampered by linkage drag (co-transfer of undesirable alleles or deleterious alleles). Nonetheless, this problem can be alleviated by deploying molecular maps and utilizing advanced backcross QTL approach with the aid of genomic tools. Pratap et al. (2021) summarized the successful examples of introgression breeding in legumes (Pratap et al. 2021). Nevertheless, the potential of introgression breeding in legumes needs to be better exploited to transfer beneficial genes from the wild to develop superior cultivars. Re-domestication, on the other hand, revives the cultivation of lost species. For instance, the Marama bean (Tylosena esculentum (Burchell) A. Schreiber) and the African locust bean (Parkia biglobosa (Jacq.) R.Br. ex G.Don) have been targeted for domestication with the use of genomes (Smykal et al. 2018). Modern genome-editing technologies such as Clustered Regularly Interspaced Short Palindromic Repeat/ CRISPR associated protein 9 have been shown to be a reliable tool for the rapid and accurate improvement of target traits in different plants (Zaidi et al. 2020). However, the optimization of an efficient transformation system in legumes is the major bottleneck due to genotype dependence and recalcitrance. Still, the success rate is inadequate in important grain legumes such as pea, chickpea, pigeonpea and mungbean (Choudhury and Rajam 2021). It is anticipated that the rapid methodological advancements in functional genomics might alleviate the transformation bottleneck in the future. Thus, a combination of genomics and genome editing offers a unique opportunity to understand crop domestication and evolution and translate the acquired knowledge into improved grain legume varieties.

Data Availability
No new data were generated in support of this review article.

Funding
German Leibniz Association in the frame of the Leibniz Junior Research groups (J118/2021/REPLACE).

Disclosures
The authors have no conflicts of interest to declare.
References

Abbo, S. and Gopher, A. (2017) Near Eastern plant domestication: a history of thought. Trends Plant Sci. 22: 491–511.

Abbo, S., Gopher, A., Rubin, B. and Lev-Yadun, S. (2005) On the origin of Near Eastern founder crops and the ‘Dump-heap Hypothesis’. Genet. Resour. Crop Evol. 52: 491–495.

Abbo, S., Pinnasi van-Oss, R., Gopher, A., Sarangá, Y., Ofner, I. and Peleg, Z. (2014) Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. Trends Plant Sci. 19: 351–360.

Abbo, S., Rachamim, E., Zehavi, Y., Zezak, I., Lev-Yadun, S. and Gopher, A. (2011) Experimental growing of wild pea in Israel and its bearing on Near East plant domestication. Ann. Bot. 107: 1399–1404.

Abbo, S., Sarangá, Y., Peleg, Z., Kerem, Z., Lev-Yadun, S. and Gopher, A. (2009) Reconsidering domestication of legumes versus cereals in the ancient Near East. Q. Rev. Biol. 84: 29–50.

Abbo, S., Zezak, I., Schwartz, E., Lev-Yadun, S. and Gopher, A. (2008) Experimental harvesting of wild peas in Israel: implications for the origins of Near East farming. J. Archaeol. Sci. 35: 922–929.

Aguilar-Benitez, D., Rubio, J., Millán, T., Gil, J., Die, J.V. and Castro, P. (2020) Genetic analysis reveals PDHl as a candidate gene for control of pod dehiscence in chickpea. Mol. Breed. 40: 40.

Alachiotis, N. and Pavlidis, P. (2018) RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. Commun. Biol. 1: 79.

Angioi, S.A., Rau, D., Attene, G., Nanni, L., Bellucci, E., Logozzo, G., et al. (2010) Beans in Europe: origin and structure of the European landraces of Phaseolus vulgaris L. Theor. Appl. Genet. 121: 829–843.

Bertioli, D., Cannon, S.B., Froenicke, L., Huang, G., Farmer, A.D., Cannon, E.K., et al. (2016) The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. Nat. Genet. 48: 438–446.

Björmdottir, E., Nadzieja, M., Chang, W., Escobar-Herrera, L., Mancinotti, D., Angra, D., et al. (2021) VC1 catalyses a key step in the biosynthesis of vicine in faba bean. Nat. Plants 7: 923–931.

Bollington, A., DeLonge, M., Mungra, D., Hayek, M., Saifuddin, M. and McDermid, S.S. (2022) Closing research investment gaps for a global food transformation. Front. Sustain. Food Syst. 5: 794594.

Brownling, B.L., Tian, X., Zhou, Y. and Browning, S.R. (2021) Fast two-stage phasing of large-scale sequence data. Ann. Hum. Genet. 108: 1880–1890.

Campbell, L.G., Teitel, Z. and Mirikit, M.N. (2016) Contemporary evolution and the dynamics of invasion in crop–wild hybrids with heritable variation for two weedy life–histories. Evol. Appl. 9: 697–708.

Caracuta, V., Weinstein-Evron, M., Kaufman, D., Yeshurun, R., Silvent, J. and Boaretto, E. (2016) 14,000-Year-old seeds indicate the Levantine origin of the lost progenitor of faba bean. Sci. Rep. 6: 37399.

Cerny, C., Makowski, D. and Pelzer, E. (2018) Preceding cultivation of grain legumes increases cereal yields under low nitrogen input conditions. Environ. Chem. Lett. 16: 631–636.

Chacón-Sánchez, M.J. and Martínez-Castillo, J. (2017) Testing domestication scenarios of lima bean (Phaseolus lunatus L.) in Mesocamerica: insights from genome-wide genetic markers. Front. Plant Sci. 8: 1551.

Chang, C., Chen, W., Luo, S., Ma, L., Li, X. and Tian, C. (2019) Rhizosphere microbiota assemblage associated with wild and cultivated soybeans grown in three types of soil suspensions. Arch. Agron. Soil Sci. 65: 74–87.

Chen, X., Li, H., Pandey, M.K., Yang, Q., Wang, X., Garg, V., et al. (2016) Draft genome of the peanut A-genome progenitor (Arachis duranensis) provides insights into geoacarpa, oil biosynthesis, and allergens. Proc. Natl. Acad. Sci. USA 113: 6785–6790.

Chen, X., Lu, Q., Liu, H., Zhang, J., Hong, Y., Lan, H., et al. (2019) Sequencing of cultivated peanut, Arachis hypogaea, yields insights into genome evolution and oil improvement. Mol. Plant 12: 920–934.

Choudhury, A. and Rajam, M.V. (2021) Genetic transformation of legumes: an update. Plant Cell Rep. 40: 1813–1830.

Di Vittori, V., Bitocchi, E., Rodriguez, M., Alseelk, S., Bellucci, E., Nanni, L., et al. (2020) Pod indehiscence in common bean is associated to the fine regulation of PvMYB26 and a non-functional accession layer. J. Exp. Bot. 72: 1617–1633.

Dierking, E.C. and Bilyeu, K.D. (2008) Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. Plant Genome 1: 135–145.

Ding, M. and Chen, Z.J. (2018) Epigenetic perspectives on the evolution and domestication of polyploid plant and crops. Curr. Opin. Plant Biol. 42: 37–48.

Doebely, J.F., Gaut, B.S. and Smith, B.D. (2006) The molecular genetics of crop domestication. Cell 127: 1309–1321.

Dong, Y., Yang, X., Liu, J., Wang, B.-H., Liu, B.L. and Wang, Y.Z. (2014) Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. Nat. Commun. 5: 3352.

Donnelly, E.D., Watson, J.E. and McGuire, J.A. (1972) Inheritance of hard seed in Vicia. J. Hered. 63: 361–365.

Elango, D., Rajendran, K., Van der Laan, L., Sebastian, S., Raigne, J., Thaiparambil, N.A., et al. (2022) Raffinose family oligosaccharides: friend or foe for human and plant health? Front. Plant Sci. 13: 829118.

Emmrich, P.M., SARK, A., Najci, I., Kaitkhottill, G.G., Ellis, N., Moore, C., et al. (2020) A draft genome of grass pea (Lathyrus sativus), a resilient diploid legume. bioRxiv 058164.

Erdman, J.W. Jr. (1985) Bioavailability of trace minerals from cereals and legumes. Cereal. Chem. 58: 21–26.

Espinoza-Alonso, L.G., Lygin, A., Widholm, J.M., Valverde, M.E. and Paredes-Lopez, O. (2006) Polyphenols in wild and weedy Mexican common beans (Phaseolus vulgaris L.). J. Agric. Food Chem. 54: 4436–4444.

Eremadi, F., Hashemi, M., Randhir, R., Zandvakili, O. and Ebad, A. (2018) Accumulation of L-DOPA in various organs of faba bean and influence of drought, nitrogen stress, and processing methods on L-DOPA yield. Crop J. 6: 426–434.

Excoffier, L., Marchi, N., Marques, D.A., Matthey-Doret, R., Gouy, A. and Sousa, V.C. (2021) fastsimcoal2: demographic inference under complex evolutionary scenarios. Bioinformatics 37: 4882–4885.

Fernández-Marín, B., Milla, R., Martín-Robles, N., Arc, E., Kranmer, I., Becerril, J.M., et al. (2014) Side-effects of domestication: cultivated legume seeds contain similar tocopherols and fatty acids but less carotenoids than their wild counterparts. BMC Plant Biol. 14: 1599.

Finn-Savage, W.E. and Footit, S. (2017) Seed dormancy cycling and the regulation of dormancy mechanisms to time germination in variable field environments. J. Exp. Bot. 68: 843–856.

Forbes, I. and Wells, H.D. (1968) Hard and soft seededness in blue lupine, Lupinus angustifolius L.: inheritance and phenotype classification. Crop Sci. 8: 195–197.

Fuller, D.Q. (2007) Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World. Ann. Bot. 100: 903–924.

Funatsuiki, H., Suzuki, M., Hirose, A., Inaba, H., Yamada, T., Hajikia, M., et al. (2014) Molecular basis of a shattering resistance boosting global dissemination of soybean. Proc. Natl. Acad. Sci. 111: 17797.

Gamba, C., Hanghej, K., Gauntet, C., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A.S., et al. (2016) Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. Mol. Ecol. Resour. 16: 459–469.

Gansauge, M.-T., Gerber, T., Glooke, I., Korlevi, P., Lippik, L., Nagel, S., et al. (2017) Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. Nucleic Acids Res. 45: e79–e79.

García, T., Duitama, J., Zullo, S.S., Gil, J., Ariani, A., Dohle, S., et al. (2021) Comprehensive genomic resources related to domestication and crop improvement traits in lima bean. Nat. Commun. 12: 702.
Gaut, B.S. (2014) The complex domestication history of the common bean. Nat. Genet. 46: 663–664.

Goetz, P. (2003) Crop domestication as a long-term selection experiment. Plant Breed. Rev. 24: 1–44.

Gopher, A., Lev-Yadun, S. and Abbo, S. (2021) Breaking Ground: Plant Domestication in the Neolithic Levant: The “Core-Area One-Event” Model. Emery and Claire Yass Publications in Archaeology, The Institute of Archaeology, Tel Aviv University.

Greenlon, A., Chang, P.L., Damtew, Z.M., Muleta, A., Carrasquilla-Garcia, N., Kim, D., et al. (2019) Global-level population genomics reveals differential effects of geography and phylogeny on horizontal gene transfer in soil bacteria. Proc. Natl. Acad. Sci. 116: 15200.

Guadalupe, A., Elyadini, M. and Hamal, A. (2020) Legumes as preventive nutraceuticals for chronic diseases. In Sustainable Agriculture Reviews 4/5 pp. 115–136. Springer, Cham.

Gupta, S., Nawaz, K., Parween, S., Roy, R., Sahu, K., et al. (2016) Draft genome sequence of Cicer reticulatum L., the wild progenitor of chickpea provides a resource for agronomic trait improvement. DNA Res. 24: 1–10.

Gupta, Y.P. (1987) Anti-nutritional and toxic factors in food legumes: a review. Plant Foods Human Nutr. 37: 201–228.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H. and Bustamante, C.D. (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5: e1000695.

Ha, J., Satyaway, D., Jeong, H., Lee, E., Cho, K.H., Kim, M.Y., et al. (2021) A near-complete genome sequence of mungbean (Vigna radiata L.) provides key insights into the modern breeding program. Plant Genome 14: e20121.

Hammer, K. (1984) Das Domestikationssyndrom. Die Kulturpflanze 32: 11–34.

Han, Y., Zhao, X., Liu, D., Li, Y., Lightfoot, D.A., Yang, Z., et al. (2016) Domestication footprints anchor genomic regions of agronomic importance in soybeans. New Phytol. 209: 871–884.

Hane, J.K., Ming, Y., Kamphuis, L.G., Nelson, M.N., Garg, G., Atkins, C.A., et al. (2017) A comprehensive draft genome sequence for lupin (Lupinus luteus), an emerging health food: insights into plant–microbe interactions and legume evolution. Plant Biotechnol. J. 15: 318–330.

Harlan, J.R., de Wet, J.M.J. and Price, E.G. (1973) Comparative evolution of cereals. Evolution 27: 311–325.

Hermisson, J. and Pennings, P.S. (2017) Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under standing the patterns and probabilities of selection footprints under rapid adaptation. Methods Ecol. Evol. 8: 700–716.

Hondelmann, W. (1984) The lupin—ancient and modern crop plant. Theor. Appl. Genet. 68: 1–9.

Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5: e1000529.

Hu, D., Kan, G., Hu, W., Li, Y., Hao, D., Li, X., et al. (2019) Identification of loci and candidate genes responsible for pod dehiscence in soybean via genome-wide association analysis across multiple environments. Front Plant Sci. 10: 811.

Hufford, M.B., Seetharam, A.S., Woodhouse, M.R., Choquile, K.M., Ou, S., Liu, J., et al. (2021) De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. Science 373: 655–662.

Hufnagel, B., Marques, A., Soriano, A., Marques, L., Divol, F., Doumas, P., et al. (2020) High-quality genome sequence of white lupin provides insight into soil exploration and seed quality. Nat. Commun. 11: 492.

Iqbal, M.M., Erskine, W., Berger, J.D. and Nelson, M.N. (2020) Phenotypic characterisation and linkage mapping of domestication syndrome traits in yellow lupin (Lupinus luteus L.). Theor. Appl. Genet. 133: 2975–2987.

Isemura, T., Kaga, A., Tabata, S., Somta, P., Srinives, P., Shimizu, T., et al. (2012) Construction of a genetic linkage map and genetic analysis of domestication-related traits in mungbean (Vigna radiata). PLoS One 7: e41304.

Isemura, T., Kaga, A., Tomooka, N., Shimizu, T. and Vaughan, D.A. (2010) The genetics of domestication of rice bean, Vigna unguiculata. Ann. Bot. 106: 927–944.

Jain, M., Misra, G., Patel, R.K., Priya, P., Jhanwar, S., Khan, A.W., et al. (2013) A draft genome sequence of the pulse crop chickpea (Cicer arietinum L.). Plant J. 74: 715–729.

Jayakodi, M., Padmarasu, S., Haberer, G., Bonthala, V.S., Gundlach, H., Monat, C., et al. (2020) The barley pan-genome reveals the hidden legacy of mutation breeding. Nature 588: 284–289.

Jayakodi, M., Schreiber, M., Stein, N. and Mascher, M. (2021) Building pan-genome infrastructures for crop plants and their use in association genetics. DNA Res. 28: daa030.

Jeong, S.C., Moon, J.K., Park, S.K., Kim, M.S., Lee, K., Lee, S.R., et al. (2014) Genetic diversity patterns and domestication origin of soybean. Theor. Appl. Genet. 133: 1179–1193.

Kang, Y.J., Kim, S.K., Kim, M.Y., Lestari, P., Kim, K.H., Ha, B.K., et al. (2014) Genome sequence of mungbean and insights into evolution within Vigna species. Nat. Commun. 5: 5443.

Kang, Y.J., Satyaway, D., Shim, S., Lee, T., Lee, J., Hwang, W.J., et al. (2015) Draft genome sequence of adzuki bean, Vigna angularis. Sci. Rep. 5: 8069.

Kantar, M.B., Nashoba, A.R., Anderson, J.E., Blackman, B.K. and Rieseberg, L.H. (2017) The genetics and genomics of plant domestication. BioScience 67: 971–982.

Kaul, T., Eswaran, M., Thanagar, A., Meyyazhagan, A., Nehra, M., Raman, N.M., et al. (2019) Rice bean (Vigna unguiculata) draft genome sequence: unravelling the late flowering and unpalatability related genomic resources for efficient domestication of this underutilized crop. bioRxiv. 816595.

Kerem, Z., Lev-Yadun, S., Gopher, A., Weinberg, P. and Abbo, S. (2007) Chickpea domestication in the Neolithic Levant through the nutritional perspective. J. Archaeol. Sci. 34: 1289–1293.

Kim, D.H., Kaashyap, M., Rathore, A., Das, R.R., Parupalli, S., Upadhyaya, H.D., et al. (2014) Phylogenetic diversity of Mesorhizobium in chickpea. J. Biosci. 39: 513–517.

Kim, S.K., Nair, R.M., Lee, J. and Lee, S.-H. (2015) Genomic resources in mungbean for future breeding programs. Front Plant Sci. 6.

Kienf, C. (2014) Analysis of plant microbe interactions in the era of next generation sequencing technologies. Front Plant Sci. 5: 216.

Kongajimun, A., Kaga, A., Tomooka, N., Somta, P., Vaughan, D.A. and Srinives, P. (2012) The genetics of domestication of yardlong bean, Vigna unguiculata (L.) Walp. ssp. unguiculata cv.-gr. sesquipedalis. Ann. Bot. 109: 1185–1200.

Kreplak, J., Madou, M.-A., Cápal, P., Novák, P., Labadie, K., Aubert, G., et al. (2019) A reference genome for pea provides insight into legume genome evolution. Nat. Genet. 51: 1411–1422.

Kumar, R., Janila, P., Vishwakarma, M.K., Khan, A.W., Manohar, S.S., Gangurde, S.S., et al. (2019) Whole genome re-sequencing-based QTL-seq identified candidate genes and molecular markers for fresh seed dormancy in groundnut. Plant Biotechnol. J. 18: 992–1003.

Kumar, Y., Basu, S., Goswami, D., Devi, M., Shivarre, U.S. and Vishwakarma, B.K. (2021) Anti-nutritional compounds in pulses: implications and alleviation methods. Legume Sci. e111: 1–13.

Ladizinsky, G. (1985) The genetics of hard seed coat in the genus Lens. J. Hered. 76: 497–502.

Ladizinsky, G. (1985) The genetics of hard seed coat in the genus Lens. J. Hered. 76: 497–502.
Shirasawa, K., Kosugi, S., Sasaki, K., Ghelfi, A., Okazaki, K., Toyoda, A., et al. (2021) Genome features of common vetch (Vicia sativa) in natural habitats. *Plant Direct* 5: e352.

Singh, A.K., Velmurugan, A., Gupta, D.S., Kumar, J., Kesari, R., Konda, A., et al. (2019) Draft genome sequence of a less-known wild Vigna: Beach pea (*V. marina* cv. AN8p-14-03). *Crop J.* 7: 660–666.

Smýkal, P., Coyne, C.J., Ambrose, M.J., Maxted, N., Schaefer, H., Blair, M.W., et al. (2015) Legume crops phylogeny and genetic diversity for science and breeding. *Crit. Rev. Plant Sci.* 34: 43–104.

Smýkal, P., Nelson, M.N., Berger, J.D. and Von Wettberg, E.J.B. (2018) The impact of genetic changes during crop domestication. *Agronomy* 8: 119.

Soltani, A., Walter, K.A., Wiersma, A.T., Santiago, J.P., Quisquiel, M., Chitwood, D., et al. (2021) The genetics and physiology of seed dormancy, a crucial trait in common bean domestication. *BMC Plant Biol.* 21: 58.

Song, Q., Hyten, D.L., Jia, G., Quisquiel, C.V., Fickus, E.W., Nelson, R.L., et al. (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One* 8: e54985.

Spriggs, A., Henderson, S.T., Hand, M.L., Johnson, S.D., Taylor, J.M. and Koltunow, A. (2018) Assembled genomic and tissue-specific transcriptomic data resources for two genetically distinct lines of Cowpea (*Vigna unguiculata* (L.) Walp.). *Gates Open Res.* 2: 7.

Sun, L., Miao, Z., Cai, C., Zhang, D., Zhao, M., Wu, Y., et al. (2015) GmHs1-1, encoding a calcineurin-like protein, controls hard-seededness in soybean. *Nat. Genet.* 47: 939–943.

Takahashi, Y., Kongjaimun, A., Muto, C., Kobayashi, Y., Kumagai, M., Sun, L., Miao, Z., Cai, C., Zhang, D., Zhao, M., Wu, Y., et al. (2015) Genome and transcriptome resource for trait improvement. *Nat. Biotechnol.* 33: 105–143.

Varshney, R.K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H.D., Saxena, R.K., Langridge, P., et al. (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *Nat. Biotechnol.* 31: 631–649.

Varshney, R.K. (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*). *Ann. Bot.* 109: 1017–1025.

Varshney, R.K., Bohra, A., Yu, J., Graner, A., Zhang, Q. and Sorrells, M.E. (2021) Designing future crops: genomics-assisted breeding comes of age. *Trends Plant Sci.* 26: 631–649.

Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.C., Hall, R.J., Conception, G.T., et al. (2019) Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* 37: 1155–1162.

Xu, X., Zhang, Y., Yuan, W., Xu, Y., Muhammad Aslam, M., Miao, R., et al. (2020) The genome evolution and low-phosphorus adaptation in white lupin. *Nat. Commun.* 11: 1069.

Ying, D., Cui, M., Zhang, Y., Zhang, W., Liu, S., et al. (2018) Genome of an allotetraploid wild peanut *Arachis monticola* a de novo assembly. *GigaScience* 7: gyx066.

Yundaeng, C., Somta, P., Amkul, K., Kongjaimun, A., Kaga, A. and Tomooka, N. (2019) Construction of genetic linkage map and genome dissection of domestication-related traits of moth bean (*Vigna aconitofolia*), a legume crop of arid areas. *Mol. Genet. Genomics* 294: 621–635.

Zaidi, S.S.E.A., Mahas, A., Vanderschuren, H. and Mahfouz, M.M. (2020) Engineering crops of the future: CRISPR approaches to develop climate-resilient and disease-resistant plants. *Genome Biol.* 21: 289.

Zhang, J. and Singh, A.K. (2020) Genetic control and geo-climate adaptation of pod dehiscence provides novel insights into soybean domestication. *G3 (Bethesda)* 10: 545.

Zhang, Q., Tu, B., Liu, C. and Liu, X. (2018) Pod morphology, morphogenesis and dehiscing forces in pod dehiscence of soybean (*Glycine max* (L.) Merrill). *Flora* 248: 48–53.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Liu, Y., Li, W., et al. (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33: 408–414.

Zhuan, W., Chen, H., Yang, M., Wang, J., Pandey, M.K., Zhang, C., et al. (2019) The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* 51: 865–876.