Deciphering PDH1's role in mung bean domestication: a genomic perspective on pod dehiscence

Article *in* The Plant Journal · February 2024

DOI: 10.1111/tpj.16680

13 authors, including:

Shuai Li
Qingdao Agricultural University
15 PUBLICATIONS **596** CITATIONS

SEE PROFILE

Liyang Chen
Purdue University
14 PUBLICATIONS **182** CITATIONS

SEE PROFILE

Xutong Wang
Huazhong Agricultural University
41 PUBLICATIONS **928** CITATIONS

SEE PROFILE

the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2024) doi: 10.1111/tpj.16680

Deciphering PDH1's role in mung bean domestication: a genomic perspective on pod dehiscence

Shuai Li[1,*,†], Yaling Li[2,†], Hong Zhu[3,†], Liyang Chen[4], Huiying Zhang[1], Lijie Lian[2], Miaomiao Xu[1], Xilong Feng[2], Rui Hou[1], Xiaolin Yao[1], Yifan Lin[2], Huaying Wang[5] and Xutong Wang[2,*]

[1]*College of Life Sciences, Qingdao Agricultural University, Qingdao 266109, China,*

[2]*National Key Laboratory of Crop Genetic Improvement, College of Plant Science and Technology, Huazhong Agricultural University, No. 1 Shizishan Road, Hongshan District, Wuhan, Hubei 430070, China,*

[3]*College of Agronomy, Qingdao Agricultural University, Qingdao 266109, China,*

[4]*Department of Agronomy, Purdue University, West Lafayette, Indiana 47907, USA, and*

[5]*Northeast Normal University, Changchun 130024, China*

Received 12 September 2023; revised 2 January 2024; accepted 29 January 2024.

*For correspondence (e-mail li2014shuai@qau.edu.cn; xtwang@mail.hzau.edu.cn).

†These authors contributed equally to this work.

SUMMARY

Mung bean (*Vigna radiata*) stands as a crucial legume crop in Asia, contributing to food security. However, our understanding of the underlying genetic foundation governing domesticated agronomic traits, especially those linked to pod architecture, remains largely unexplored. In this study, we delved into the genomic divergence between wild and domesticated mung bean varieties, leveraging germplasm obtained from diverse sources. Our findings unveiled pronounced variation in promoter regions (35%) between the two mung bean subpopulations, suggesting substantial changes in gene expression patterns during domestication. Leveraging transcriptome analysis using distinct reproductive stage pods and subpopulations, we identified candidate genes responsible for pod and seed architecture development, along with Genome-Wide Association Studies (GWAS) and Quantitative Trait Locus (QTL) analysis. Notably, our research conclusively confirmed *PDH1* as a parallel domesticated gene governing pod dehiscence in legumes. This study imparts valuable insights into the genetic underpinnings of domesticated agronomic traits in mung bean, and simultaneously highlighting the parallel domestication of pivotal traits within the realm of legume crops.

Keywords: *Vigna radiata*, domestication, genomic divergence, pod dehiscence, parallel selection.

INTRODUCTION

Domestication is a human-driven selection process that refines crops' agronomic traits for optimal cultivation. As a result, cultivated varieties exhibit distinct morphological and physiological traits compared to their wild progenitors, and this phenomenon is known as domestication syndrome, indicating genomes of domesticated crops have been significantly modified by human selection (Hammer, 1984; Purugganan, 2019). In modern agriculture, numerous elite varieties with diverse environmental adaptation and high yield have been developed based on domesticated crop germplasm. Multiple crop varieties have been released for production in recent decades (Somta et al., 2009), signifying the importance of selected agronomic traits during domestication in influencing agricultural production. Consequently, understanding the genetic basis of domesticated agronomic traits is crucial for advancing modern crop improvement (Abbo et al., 2009, 2014).

Parallel domestication is a widespread phenomenon observed in multiple crops with close genetic relationships, and several parallel domesticated agronomic traits have been identified in different plant species. For instance, the loss of seed dormancy is a component of domestication syndrome in crops, and the soybean seed dormancy *G* gene and its orthologs in rice and tomato demonstrate parallel selection during domestication (Wang et al., 2018). Growth habit is regulated by *TERMINAL FLOWER 1* (*TFL1*) and its orthologs in legumes, such as soybean, mung bean, and common bean, with causative mutations in coding sequences or promoters leading to changes in growth habit from indeterminate to

determinate (Li et al., 2018). The flowering time gene *Heading Date 1* (*Hd1*), a member of the *CONSTANS* gene family, undergoes parallel domestication in cereals, including sorghum, foxtail millet, and rice (Liu et al., 2015). Sorghum *Shattering 1* (*Sh1*), encoding a YABBY transcription factor, controls seed shattering, and its orthologs *OsSh1* in rice and *ZmSh1* in maize have similar functions, indicating that seed shattering has undergone parallel selection in cereal domestication (Lin et al., 2012). Furthermore, Common bean *PvMYB26* and its orthologs in adzuki bean, cowpea, and mung bean are involved in pod shattering resistance, and selected during domestication (Di Vittori et al., 2021; Lin et al., 2023; Takahashi et al., 2020). *Pdh1*, encoding a dirigent-like protein, is involved in pod shattering regulation and undergoes parallel domestication in soybean and common bean (Funatsuki et al., 2014; Parker et al., 2020). These cases suggest identifying the parallel domestication mechanism offers a valuable approach to investigate key genes responsible for similar agronomic traits across crops with close genetic relationships.

Legumes, including mung bean, share a common origin and exhibit many similar domesticated traits (Abbo et al., 2014; Lin et al., 2023; Zhang et al., 2022). Advances in genomic technologies have enabled extensive analysis of domestication traits in legumes like cowpea, soybean, common bean, and mung bean (Kang et al., 2014; Liu et al., 2020, 2022; Pan et al., 2023; Schmutz et al., 2010). Mung bean, an important legume crop, is rich in proteins and vitamins and has a domestication history dating back approximately 8000 generations in South Asia (Fuller, 2007; Fuller & Harvey, 2006; Keatinge et al., 2011; Lambrides & Godwin, 2007; Ong et al., 2023; Sangiri et al., 2007). The release of the mung bean genome and the construction of a pan-genome have facilitated the study of its agronomic traits (Kang et al., 2014; Liu et al., 2022). However, the genetic basis of traits like pod dehiscence remains unclear, despite the identification of several QTLs related to domestication traits (Isemura et al., 2012; Schafleitner et al., 2015).

In this study, we focus on analyzing the genomic divergence in mung bean germplasm to identify candidate genes controlling pod development traits, with an emphasis on the role of *VrPDH1* in pod dehiscence. Additionally, we explore the parallel domestication of pod development-related traits across legume crops and aim to enhance the understanding of mung bean domestication, specifically in the context of pod architecture.

RESULTS

Genomic divergence and selection in mung bean

To investigate genomic variations within the mung bean population, we collected 228 whole-genome sequencing (WGS) data, including 217 accessions from a previously published database with 215 domesticated mung bean (DMs) and 2 wild mung bean (WMs) (Liu et al., 2022), and 11 new samples comprising 9 WMs and 2 DMs (Table S1). The geographical distribution of these accessions underscored the diversity within the mung bean population, providing a comprehensive basis for understanding genomic divergence (Figure 1a). To decipher the variation loci between the two subpopulations, we mapped all WGS data to Vrad_JL7 genome sequences with an average mapping rate of 99% in both DMs and WMs (Liu et al., 2022). Our analysis identified 23.2 million INDELs and 16.1 million SNPs in the population, demonstrating considerable genetic variation.

To understand the relationship between the distribution of variations and their effects on the genome, we assigned the variations to intragenic features and intergenic regions. In the mung bean genome, the proportions of various genomic features are as follows: intergenic regions account for 24%, 5′ untranslated regions (5′ UTRs) for 2%, 3′ untranslated regions (3′ UTRs) for 4%, coding DNA sequences (CDS) for 9%, promoters for 18.2%, terminators for 18.2%, and introns for 25%. We observed that 35% of the variations were present in gene promoters, while 4% were in coding sequence (CDS) regions (Figure 1b; Table S2). This suggests significant changes in gene expression patterns due to the divergence in promoter sequences between the two subpopulations. Moreover, approximately 1.7% of these variations resulted in nonsynonymous changes in CDS regions, highlighting alterations in protein sequences within the mung bean population (Figure 1b; Table S2). We then grouped the variations into four categories: ‘Fix’, denoting differences between WMs and DMs; ‘Cultivar’, representing unique variations in WMs with polymorphism in DMs; ‘Wild’, signifying unique variations in DMs with polymorphism in WMs; and ‘Share’, indicating polymorphism in both subpopulations (Figure 1b; Table S2). The ‘Wild’ category showed a higher percentage than the ‘Cultivar’ category in each genomic region, implying that DMs lost significant polymorphism due to artificial selection during domestication from WMs (Figure 1b; Table S2).

To identify selective sweep regions in mung bean, which reflect potential domesticated genes in those regions, we calculated the nucleotide diversity (π) in each subpopulation and the genetic distance (F_{st}) between the two subpopulations. We found that the π value in DMs (0.9×10^{-3}) was significantly lower than that in WMs (3×10^{-3}), while the average F_{st} value was around 0.5, reflecting high genetic divergence between the two subpopulations and indicating that domesticated mung bean was under strong selection constraints (Figure 1c). A total of 306 selective sweep regions were identified based on the high nucleotide diversity rate, which includes five adaptation selective sweeps (AS) and 301 domesticated selective sweeps (DS) (Figure 1d; Table S3).

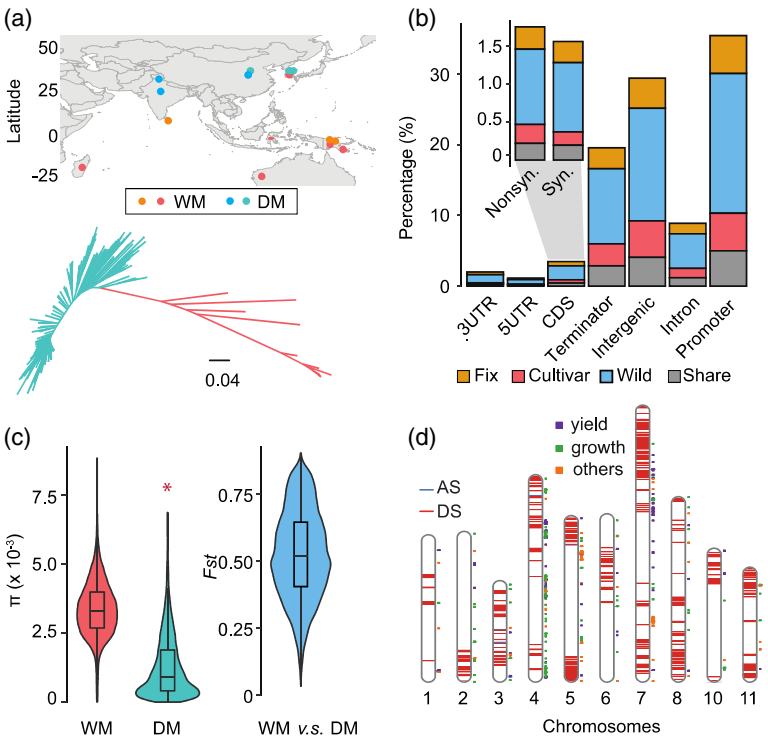

Figure 1. The characteristics of population genomics of mung bean. (a) The geographic distribution of mung bean accessions sequenced in this study, with WM and DM representing wild and domesticated mung bean, respectively. The red/green color represents the accessions used for DNA resequencing, while other colors represent accessions only used for transcriptome analysis. The phylogenetic tree of the mung bean population, with green and red branches indicating WM and DM subpopulations, respectively. (b) The percentage of four types of genomic feature SNPs: Fixed, cultivar, wild, and Share. The definitions are described in the main text. The 'Promoter' is defined as the 2 kilobase (kb) region upstream of the Transcription Start Site (TSS), whereas the 'Terminator' refers to the 2 kb region downstream of the Transcription Termination Site (TTS). (c) The comparison of π (left panel) and F_{st} (right panel) distributions between WM and DM subpopulations. Using the K-S test, the red asterisk indicates significant differences with $P < 0.001$. (d) The distribution of selective sweep (adaptation selective sweeps, AS; domesticated selective sweeps, DS) and GWAS signals in the chromosomes of mung bean, represented by bands and dots, respectively.

We then mapped these regions to GWAS loci and highlighted 157 GWAS loci associated with DSs (Table S4), suggesting artificial selection on these genes, notably for traits like pod length and shattering (Liu et al., 2022).

Distinctive pod development traits in wild and domesticated mung bean

Significant divergence in pod development traits was observed between wild mung beans (WMs) and domesticated mung beans (DMs) (Isemura et al., 2012). This divergence is highlighted by the shattering-resistant phenotype in DMs, as opposed to the shattering-susceptible pods in WMs, a crucial trait for seed dispersal in the wild accessions. These phenotypic differences are substantiated by numerous pod-related agronomic trait loci found within the selective sweep regions (Table S4). In analyzing the pod traits of seven representative genotypes, including three WMs and four DMs from diverse geographic regions (Figure 2a; Table S5), we noted considerable diversity. Specifically, pod length, seed hundred-grain weight, seed length, seed width, and seed height were greater in DMs compared to WMs. In contrast, the seed number per pod showed no significant difference between the two groups (Figure 2a,b). This phenotypic variation offers valuable insights into the selection pressures exerted during domestication, particularly regarding pod shattering and size. To delve deeper into the genetic underpinnings of these phenotypic differences, we conducted a transcriptome analysis of developing pods across six stages (R1–R6) in the selected genotypes (Figure 2a; Table S6). The goal was to identify differentially expressed genes that could be responsible for the observed differences in pod traits, with a particular focus on genes related to pod shattering resistance.

Identification of candidate genes responsible for pod and seed architecture development during mung bean domestication

In a previous study, it was reported that most yield and pod architecture-related traits were found on chromosome 7 (Liu et al., 2022). Additionally, two major QTLs related to pod length (*Pdl5.1.1*+) and pod shattering (*Pdr4w5.1.1*–) were previously identified (Isemura et al., 2012). After mapping the four markers they used to the reference genome, it was discovered that *Pdl5.1.1*+ and *Pdr4w5.5.1*– also located on chromosome 7, overlapping with the GWAS signals controlling PDL and POS traits (Figure 3a). These regions were likely to contain key genes that control these two traits. To identify the candidate genes for PDL, the expression patterns of six genes within the region defined by the boundary of the GWAS signals were analyzed (Figure 3b). Two genes were excluded due to their low expression levels in both WMs and DMs across the six reproductive stages (Figure 3b). The other four genes were examined, and *jg21606* and *jg21607* had significantly higher expression levels in DMs than in WMs, specifically in the R3 stage (Table S6). *jg21602* and *jg21603* did not exhibit significant differences between WMs and DMs, but

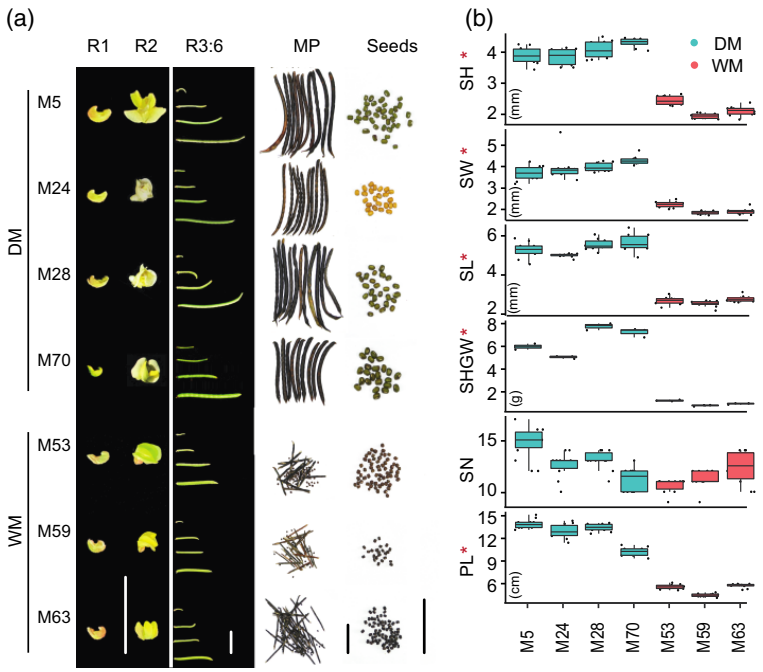

Figure 2. Morphological difference between wild and domesticated mung bean. (a) The phenotype of flower bud (R1), the open flower (R2), beginning pod (R3), full pod (R4), beginning seed pod (R5), and full seed pod in four DMs and three WMs (left panel), the mature pods (MP), and the seeds (right panel). (b) The statistics of agronomic traits, including pod length (PL), seed number (SN), seed hundred-grain weight (SHGW), seed length (SL), seed width (SW), and seed height (SH), are compared between WMs (in red) and DMs (in green). The scale bars indicate a length of 4 cm.

their expression pattern was correlated with middle pod developmental stages (R3-R5) (Figure 4a,b). Interestingly, the ortholog genes of *jg21602* and *jg21603* in *Arabidopsis thaliana* (*AT5G67170* and *AT2G23360*, respectively) are highly expressed in siliques (Rhee et al., 2003). Only one candidate SNP in *jg21602*, belonging to the 'Wild' category, and one candidate SNP in *jg21603*, belonging to the 'Share' category, was found to be under strong selection in DMs after checking genotyping information in 228 mung bean accessions (Figure 4a,b). These two genes potentially contribute to the architecture development of the pods.

To identify genes responsible for controlling pod and seed architectures in mung bean, we analyzed GWAS loci for the pod shattering trait (POS). Previous research reported one SNP located in an intergenic region (Liu et al., 2022), and we identified a candidate gene named *jg21742* by extending the flanking sequence (Figure 3c). *Jg21742* is the ortholog of *PDH1* in mung bean, which has been shown to control the shattering trait in soybean (*Glycine max*) and common bean (*Phaseolus vulgaris*) (Funatsuki et al., 2014; Parker et al., 2020). Our analysis revealed that *jg21742* was expressed around 30-fold higher in WMs than in DMs, particularly in the R3 stage, similar to *GmPDH1* in soybean (Figure 4c) (Funatsuki et al., 2014). Moreover, we found that one nonsynonymous SNP in *jg21742* was fixed in both WMs and DMs subpopulations and overlapped with one DS region, suggesting that *jg21742* was a strong candidate gene for controlling the POS trait in mung bean (Figure 4c). In addition to POS, we also identified *jg24043* as a candidate gene for controlling crude starch content (CSC) (Figure 4d). Our analysis revealed that *jg24043* which is highly expressed in DMs in the R6 stage, was consistent with the stage of starch accumulation in seeds. The causative SNP for *jg24043* exhibited polymorphism in DMs, with 25.2% of 'A' and 74.8% of 'G' alleles, indicating strong artificial selection on the CSC traits (Figure 4d).

PDH1, a parallel domesticated gene controlling pod shattering trait in several legumes

To assess whether *VrPDH1* (*jg21742*) is associated with pod shattering in mung bean, we examined the phenotype and coding sequences of 50 DMs and 17 WMs (Table S7). We identified six haplotypes of the mung bean *VrPDH1* gene, which included seven SNPs in the CDS region, six of which caused amino acid changes. We further categorized the haplotypes into two groups, *VrPDH1* (hap3, hap4, hap5, and hap6) and *vrpdh1* (hap1, hap2), respectively (Figure 5a). Notably, one DM accession presented the 'hap3' type, although this type mainly existed in WMs (Figure 5a). The distinctive patterns between WMs and DMs, except for one accession in the haplotype network of *VrPDH1/vrpdh1*, suggested that genetic introgression or gene flow occurred in the domestication process of mung bean (Figure 5b; Table S7). Since 'hap3' resulted in the SR

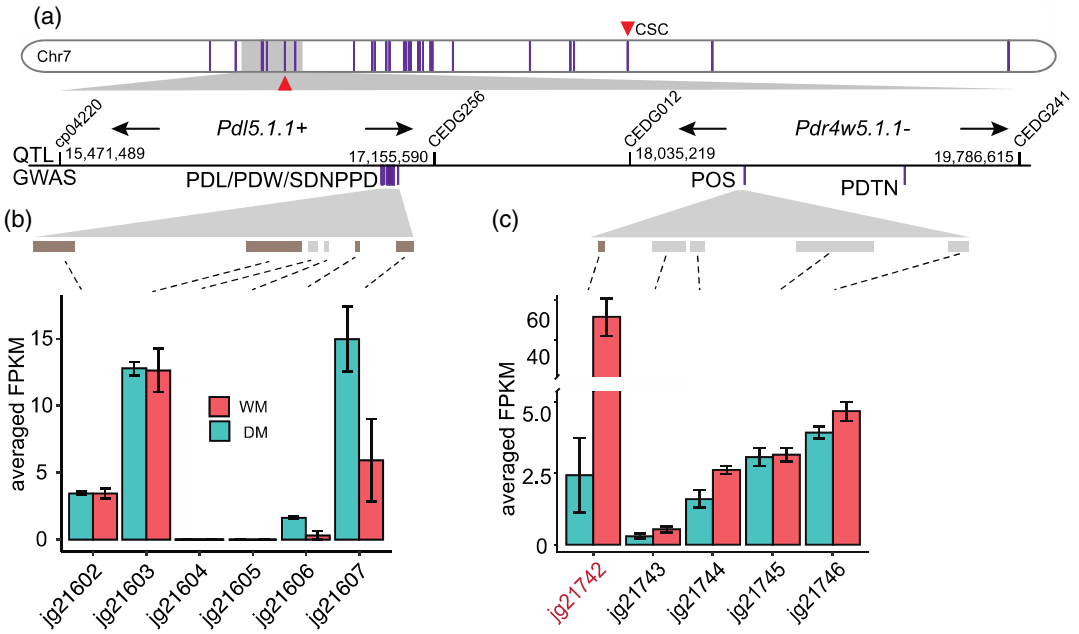

Figure 3. Identification of candidate genes associated with pod and seed architecture in mung bean. (a) Mapping location of two QTLs for pod length and pod dehiscence on chromosome 7. The purple color bars in the graph represent the GWAS signal as reported in Liu et al. (2022). The markers cp04220/CEDG256 and CEDG012/CEDG241 were used to map the QTLs for *Pdl5.1.1*+ and *Pdr4w5.1.1*, respectively. The abbreviations CSC, PDL, PDW, POS, SDNPPD, and PDTN stand for crude starch content, pod length, pod width, pod shattering, number of seeds per pod, and total number of pods, respectively. (b, c) Gene expression levels of candidate regions controlling pod length (b) and pod dehiscence (c) in WMs and DMs.

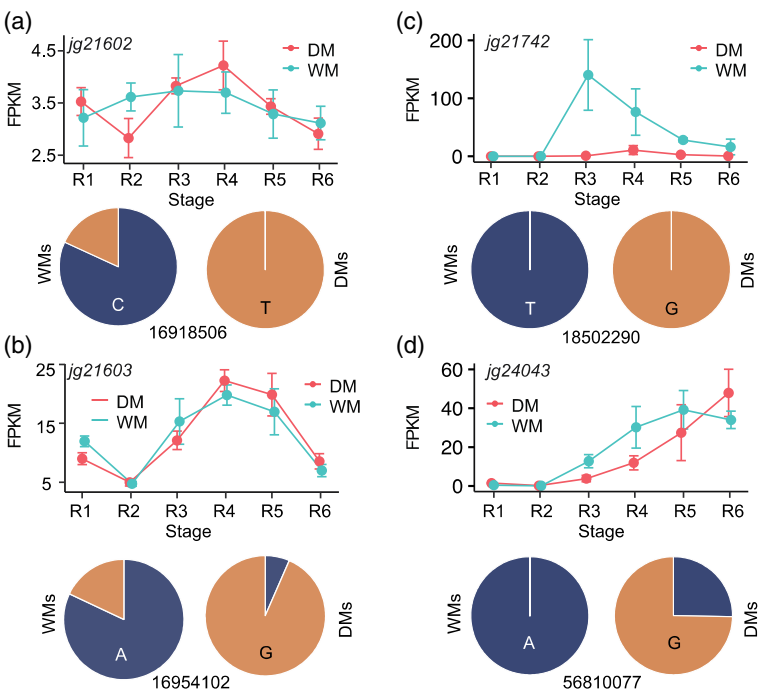

Figure 4. Differential expression and allele frequency of candidate genes in WMs and DMs. Upper panel, Expression levels of the four candidate genes, *jg21602* (a), *jg21603* (b), *jg21742* (c), and *jg24043* (d), in six developmental pods in WMs and DMs. Lower panel, the allele frequency of potentially causative mutations under domestication selection of the four candidate genes, *jg21602* (a), *jg21603* (b), *jg21742* (c), and jg24043 (d).

phenotype rather than SS, *VrPDH1* should be a potential major controlling gene, although there may be others yet to be discovered. To gain further insight into the impact of the causative mutants in the third and ninth amino acids of the *VrPDH1* protein sequence, we utilized the *VrPDH1* protein sequence to predict the structure, transmembrane probability, and signal peptide location (Figure 5c). Our analysis revealed that the initial 20 amino acids of *VrPDH1* formed a signal peptide with the potential to transmembrane, indicating that the variation in the signal peptide may potentially affect the protein's subcellular localization (Figure 5c). While *PDH1* has only been functionally cloned in soybean and common bean, studies on other legume species such as Lima bean and chickpea have reported that *PDH1* may be potentially associated with pod-shattering traits during domestication (Aguilar-Benitez et al., 2020; Funatsuki et al., 2014; Garcia et al., 2021; Parker et al., 2020). Sequence alignment showed that the protein sequence of PDH1 was conserved among these three species, and the major flexible region was within the signal peptide sequence in the N-terminal (Figure 5c).

To investigate the function of *VrPDH1/vrpdh1* in pod dehiscence regulation, the coding sequences of *VrPDH1/vrpdh1* were fused with the cauliflower mosaic virus 35S promoter and transformed into non-dehisced soybean genotype Williams 82. To confirm this, we analyzed the expression of *VrPDH1/vrpdh1* in the transgenic plants and found that all the transgenic lines with *VrPDH1* haplotype or domesticated *vrpdh1* haplotype showed higher expression levels than that in Williams 82 (Figure 6a). The transgenic lines with *VrPDH1* haplotypes had an increased shattering phenotype, with approximately 45%, 40%, and 40% of dehisced pods in three lines, respectively. While the transgenic lines with *vrpdh1* haplotypes showed similar dehisced pods to non-transgenic soybean Williams 82, with approximately 5% of dehisced pods in each transgenic line, indicating that *VrPDH1* is responsible for pod dehiscence in mung bean (Figure 6b, c). Taken together, these findings suggest that *PDH1* is a parallel domesticated gene in several legumes.

DISCUSSION

In this study, we analyzed the genetic basis underlying important agronomic traits during mung bean domestication and identified candidate genes controlling pod and seed architecture. Our findings not only provide insights into the domestication process of mung bean but also reveal parallel domestication of key traits in legumes with close genetic relationships.

The process of domestication has a significant effect on genome sequence divergence. During domestication, important agronomic traits are selected and can greatly modify the genomes of domesticated crops, which in turn impacts agricultural production (Tian et al., 2021). Therefore, understanding the genetic basis of domesticated agronomic traits is critical for modern crop improvement. However, the domesticated footprint of mung bean has not been well studied. Fortunately, the release of a high-quality chromosome genome sequence of mung bean (JL7) has provided an opportunity to investigate the genomic divergence between wild and domesticated mung bean (Liu et al., 2022). Our study has highlighted significant genomic divergence between wild and domesticated mung bean varieties, emphasizing the profound impact of domestication on genome evolution. This divergence, especially notable in gene promoters and coding sequences, underscores the substantial modifications that domestication has imposed on the genome, influencing key agronomic traits (Tian et al., 2021; Figure 1). Such genomic evolution has crucial implications for crop breeding and elucidates mung bean's domestication history. In addition to genome sequence divergence, the process of domestication also affects the divergence of gene expression (Swanson-Wagner et al., 2012). Transcriptome data is not only applied as a valuable resource to study gene regulation in different stages or subpopulations but also can be used to identify candidate genes based on gene expression patterns. This study aimed to identify candidate domesticated genes controlling pod and seed architectures in mung bean by utilizing time-series transcriptome data related to different reproductive stages and subpopulations.

Similar domesticated agronomic traits have been identified in different legume species as they evolved from a common origin (Purugganan, 2019). Although the underlying mechanisms are still unclear, parallel domesticated genes have become increasingly sufficient evidence (Lin et al., 2012; Wang et al., 2018). Untangling the impact of human selection on domesticated legume crops' agronomic traits is a major research focus for geneticists (Ambika et al., 2022). Our study provides essential information for further characterization of mung bean domestication and sheds light on the parallel domestication mechanism in legume crops. In addition, the pods of cultivated mung bean are distinct from their wild accessions (Figure 2), such as pod length, shatter resistance, and domesticated genes responsible for these agronomic traits have been identified in this study. For instance, the expression of *jg21602* and *jg21603*, located in pod length QTL (Figure 3), are correlated with middle pod developmental stages (Figure 4), and their homologous genes in *Arabidopsis* are highly expressed in siliques, indicating that these two genes might be responsible for pod length in mung bean domestication. *Jg24043*, a member of the *SWEET10* gene family, controls CSC, an important trait in seed quality, associated with a defined DS (Liu et al., 2022; Zhang et al., 2020). The causative SNP causes an amino acid change and the 'A' allele is fixed in wild mung bean,

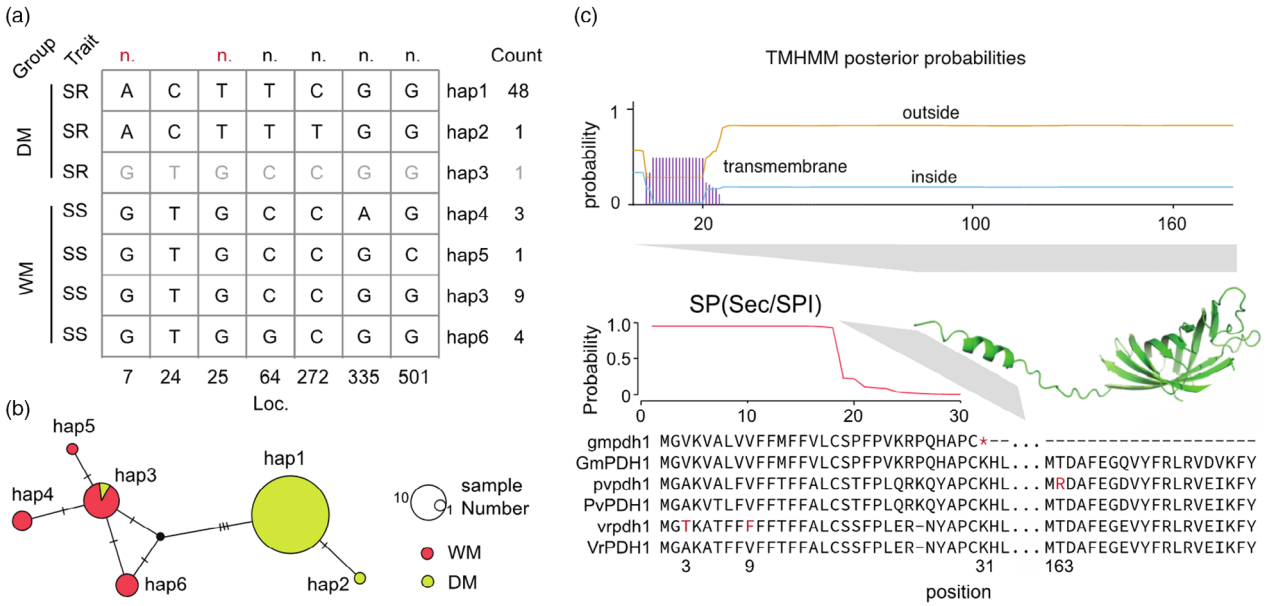

Figure 5. Analysis of *VrPDH1* haplotypes and protein structure in legume species. (a) The frequency and contribution of six haplotypes in mung bean populations. SR and SS indicate shatter resistance and shatter susceptibility, respectively. (b) The haplotype network of *VrPDH1* in two subpopulations, with red and green representing WMs and DMs, respectively. (c) The protein structure and sequence alignment of VrPDH1 among representative legume species. Gm, soybean; Pv, common bean; Vr, mung bean.

while the 'G' allele is related to high CSC but low crude protein content (CPC) in domesticated mung bean, indicating artificial selection on CSC traits (Figure 4d) (Liu et al., 2022). Moreover, we found that *VrPDH1* is a parallel domesticated gene controlling the shattering trait in legumes (Figure 5). We confirm through haplotype analysis and transgenic experiments that *VrPDH1* is responsible for pod dehiscence in mung bean (Figure 5). Furthermore, we predicted that variation in the signal peptide sequence in the N-terminal region of the *VrPDH1* protein may affect its subcellular localization, which is important for its function in pod dehiscence regulation (Figure 5c). These findings provide crucial insights into the genetic basis of important agronomic traits in mung bean, which can be used as potential targets for breeding efforts to improve yield and quality. Additionally, the identification of parallel domesticated genes, such as *VrPDH1*, in different legume species highlights the evolutionary conservation of genetic mechanisms underlying domestication.

Our findings offer vital insights into the genetic basis of key agronomic traits in mung bean, with significant implications for the genetic improvement of this and other legume crops. Identifying genes such as *VrPDH1* that control pod development traits lays the groundwork for future breeding programs aimed at enhancing yield and quality in legumes. Overall, this study not only enhances our understanding of mung bean domestication but also provides a model for studying parallel domestication mechanisms in legumes. These insights are invaluable for developing more effective breeding strategies and advancing the genetic improvement of legume crops.

EXPERIMENTAL PROCEDURES

To investigate the domestication history of mung bean, nine wild mung bean and six cultivated mung bean accessions were collected. These plants were grown in the field in Qingdao, China (see Table S1). To analyze the traits related to pod and seed development, we measured pod length, seed number, seed hundred-grain weight, seed length, seed width, seed height, and pod dehiscence using pods and seeds harvested from the field. For transcriptome analysis, we collected a total of 42 samples using six distinct developmental pods, including flower bud (R1), the open flower (R2), beginning pod (R3), full pod (R4), beginning seed pod (R5), and full seed pod (R6). Samples were taken from three wild and four cultivated mung bean plants grown in the field.

Genome resequencing and phylogenic analysis

To investigate the genomic variations between wild and domesticated mung bean, we selected 11 accessions for whole-genome resequencing. The DNA was extracted using the CTAB protocol, and the MGIEasy PCR-Free DNA library preparation kit was used to construct the libraries suitable for the DNBSEQ-T7 platform. After sequencing and quality control, we generated 150 bp paired-end reads. To perform phylogenetic analysis, 217 public mung bean accessions were downloaded from the EBI-ENA (Liu et al., 2022). We processed all sequencing reads as follows: first, Cutadapt (version 3.5.0) was used to remove potential adaptors and low-quality reads (Martin, 2011). Then, we mapped the clean reads to the mung bean reference genome (JL7, https://figshare.com/articles/dataset/Mungbean_Vigna_radiata_genome_assembly_and_population_resequencing/19583446) using BWA (version 0.7.17-r1188) and filtered out PCR duplicates and multiple mapped

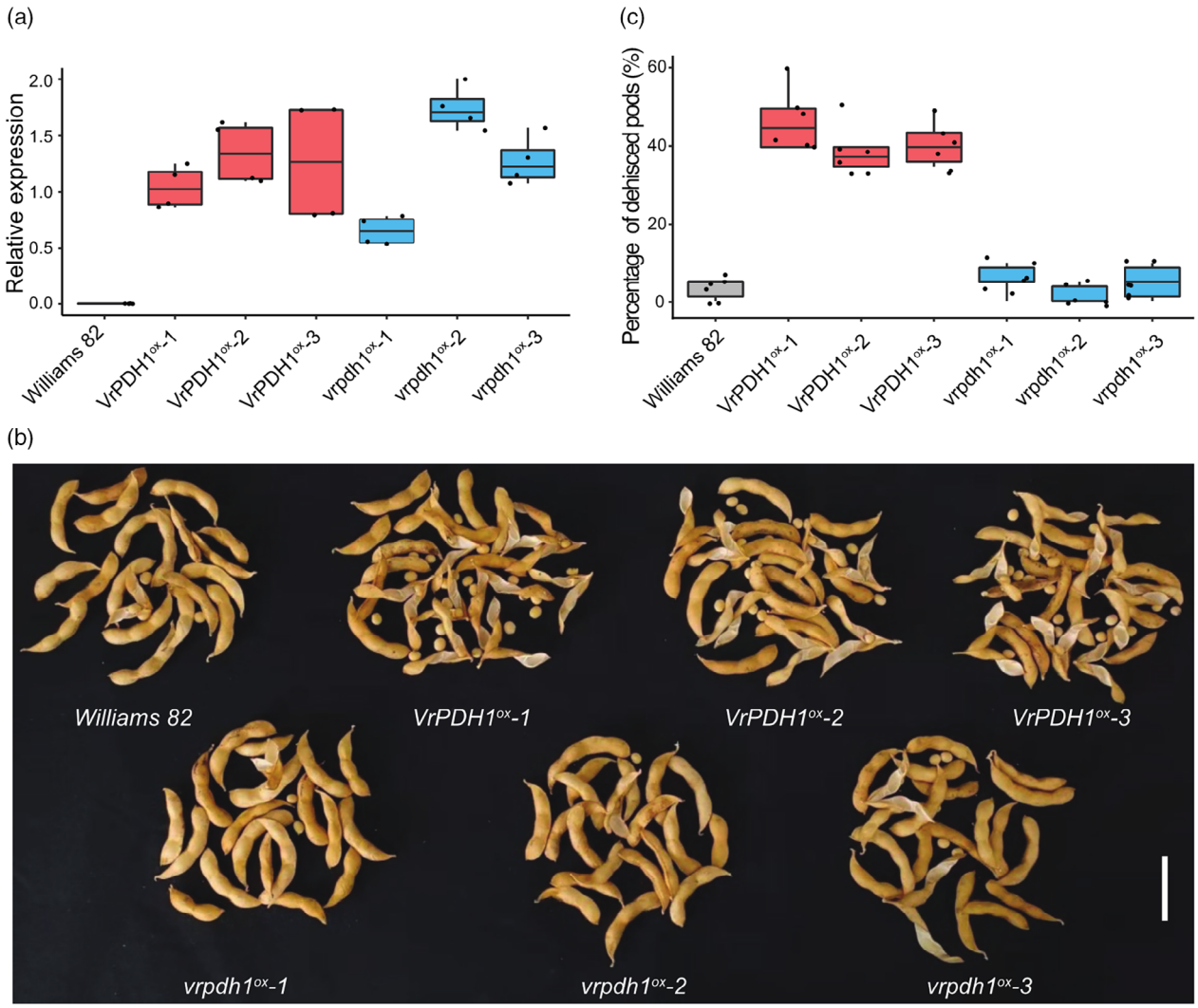

Figure 6. Function evaluation of *VrPDH1* in controlling pod dehiscence in mung bean. (a) The expression levels of *VrPDH1/vrpdh1* in transformed plants. (b) The pod dehiscence phenotype in transgenic soybean lines. (c) The percentage of dehisced pods in transgenic soybean lines.

reads using Samtools (version 1.15.1-41-gc7acf84) (Danecek et al., 2021; Li, 2013). Next, the GATK pipeline was used to generate the confidential SNPs and INDELs for variation and evolution analysis. We kept the SNPs if they were bi-allelic and had an MAF >0.01 (McKenna et al., 2010).

To construct the phylogenetic tree, we randomly selected 50K SNPs and converted the VCF file to FASTA format using an in-house Perl script, which served as the input file for MEGA software (version 11.0.10) (Tamura et al., 2021). The NJ tree and p-distance were used to generate the phylogenetic tree in MEGA, and the Evolview web tools to illustrate the phylogenetic tree (Subramanian et al., 2019). To annotate the effects of SNPs on the genome, the snpEff program was used (version 5.1d) to annotate the effect with specific gene features based on the JL7 genome annotation file (gff3) (Cingolani et al., 2012). We calculated population genetic parameters, such as π and F_{st}, using VCFtools (version 0.1.16) (Danecek et al., 2011). The selective sweep regions were defined based on the high nucleotide diversity rates and divided into two groups: domesticated selective sweeps (DS), with higher π value in wild and higher F_{st} value, and adaption selective sweeps (AS), with higher π value in domesticated mung bean than wild, and higher F_{st} value. We performed all statistical analyses using the base package in R language (version 4.2.3, https://www.R-project.org/).

Transcriptome sequencing and data processing

To analyze the transcriptome of mung bean samples, we extracted RNA and constructed RNA-seq libraries according to previously described methods (Li et al., 2021). The clean reads from each library were mapped to the JL7 mung bean reference genome sequence using HISAT2 (version 2.2.1) with default parameters (Kim et al., 2019). To obtain gene expression data, we extracted reads uniquely mapped to individual genes using SAMtools with the parameter (−Q 30) and quantified mRNA abundance using StringTie, normalizing to FPKM (Pertea et al., 2015).

Plasmid construction and transformation

To obtain the full coding sequences of *VrPdh1* and *vrpdh1*, specific primers containing *Xho*I and *Xba*I enzyme site sequences

were designed and used to amplify the DNA from JP226875 (a wild accession from East Sepik, Papua New Guinea) and JP229096 (a cultivar variety from Sukhothai, Thailand). The resulting PCR products and pPTN1171 vector were digested with *Xho*I and *Xba*I and then ligated using T_4 ligase (Ping et al., 2014). The resulting plasmids were transformed into *Agrobacterium tumefaciens* strain EHA101, and the transgenic soybean Williams 82 was obtained using *Agrobacterium*-mediated cotyledonary node transformation method as described (Song et al., 2013). For soybean shattering phenotype analysis, the R8 stage maturity pods of T_2 transgenic *VrPDH1/vrpdh1* soybean lines and Williams 82 were harvested from the field and transferred into 37°C oven. The dehisced pod was analyzed 50 days later. A list of all primers used in this study can be found in Table S8.

Sequence alignment and protein structure prediction

To perform sequence alignment and protein structure prediction, DNA was isolated from mung bean genotypes, and PCR was conducted as per the protocol described (Li et al., 2018), followed by sequencing of the PCR products. The sequence alignment of *VrPDH1/vrpdh1* was performed using MUSCLE (version 3.8.1551), and manual identification was done to detect sequence variation (Edgar, 2004). Haplotype network analysis was carried out using DnaSP (Rozas et al., 2017). The protein structure of VrPDH1 was predicted using AlphaFold2, while the transmembrane domain and signal peptide were predicted using TMHMM and SignalP web tools (Almagro Armenteros et al., 2019; Chen et al., 2003; Cramer, 2021). For gene expression analysis, RNA was isolated and quantitative reverse transcription PCR (qRT-PCR) was conducted following the protocol described (Li et al., 2021). Table S8 lists all the primers used in this study.

ACCESSION NUMBERS

All data are available in the main text, supplementary materials, public databases, or referenced studies. The raw DNA-seq and RNA-seq read sequences generated in this study have been deposited in the National Genomics Data Center (https://ngdc.cncb.ac.cn/) under the accession number PRJCA016462. Table S1 provides information about the mung bean accessions used in this study, as well as details about the tissues analyzed. The variations among 228 mung bean germplasm accessions can be accessed from the FigShare databases (https://doi.org/10.6084/m9.figshare.22679665.v1). Follow this link to review the data: https://ngdc.cncb.ac.cn/gsa/s/ltGU2YKr.

AUTHOR CONTRIBUTIONS

XW and SL designed the research; XW, SL, YL, HZ, and HW analyzed the data; LC, HZ, RH, LL, XF, XY, HW, and YL performed the research; XW and SL wrote the manuscript.

ACKNOWLEDGEMENTS

We thank Suk-Ha Lee at Seoul National University for providing the mung bean germplasm. This research was funded by the National Natural Science Foundation of China (grant 31971898 and 32171979), the National Key Research and Development Program of China (grant 2022YFD1201502), the Excellent Youth Fund of Shandong Province (grant ZR2022YQ21), the Taishan Program of Shandong Province (grant tsqn202211190), and Program for Scientific Research Innovation Team of Young Scholar in Universities of Shandong Province (grant 2021KJ029).

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Table S1. Details of mung bean accessions used for sequencing in this study.

Table S2. Statistics of variant effects on the genome in each category.

Table S3. Genome coordinates of selective sweep regions.

Table S4. Regions of selective sweep and corresponding GWAS signals in mung bean (Liu et al., 2022).

Table S5. Detailed information of mung bean accessions used for RNA-seq analysis in this study.

Table S6. Expression levels of mung bean genes in all samples.

Table S7. Detailed information of mung bean accessions used for *VrPDH1/vrpdh1* analysis in this study.

Table S8. The primers used in this study.

REFERENCES

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

Abbo, S., **Saranga, Y.**, **Peleg, Z.**, **Kerem, Z.**, **Lev-Yadun, S.** & **Gopher, A.** (2009) Reconsidering domestication of legumes versus cereals in the ancient near east. *Quarterly Review of Biology*, **84**, 29–50.

Aguilar-Benitez, D., **Rubio, J.**, **Millán, T.**, **Gil, J.**, **Die, J.** & **Castro, P.** (2020) Genetic analysis reveals *PDH1* as a candidate gene for control of pod dehiscence in chickpea. *Molecular Breeding*, **40**, 40.

Almagro Armenteros, J.J., **Tsirigos, K.D.**, **Sønderby, C.K.**, **Petersen, T.N.**, **Winther, O.**, **Brunak, S.** ***et al.*** (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, **37**, 420–423.

Ambika, **Aski, M.S.**, **Gayacharan**, **Hamwieh, A.**, **Talukdar, A.**, **Kumar Gupta, S.** ***et al.*** (2022) Unraveling origin, history, genetics, and strategies for accelerated domestication and diversification of food legumes. *Frontiers in Genetics*, **13**, 932430.

Chen, Y., **Yu, P.**, **Luo, J.** & **Jiang, Y.** (2003) Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. *Mammalian Genome*, **14**, 859–865.

Cingolani, P., **Platts, A.**, **Wang le, L.**, **Coon, M.**, **Nguyen, T.**, **Wang, L.** ***et al.*** (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, **6**, 80–92.

Cramer, P. (2021) AlphaFold2 and the future of structural biology. *Nature Structural & Molecular Biology*, **28**, 704–705.

Danecek, P., **Auton, A.**, **Abecasis, G.**, **Albers, C.A.**, **Banks, E.**, **DePristo, M.A.** ***et al.*** (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Danecek, P., **Bonfield, J.K.**, **Liddle, J.**, **Marshall, J.**, **Ohan, V.**, **Pollard, M.O.** ***et al.*** (2021) Twelve years of SAMtools and BCFtools. *GigaScience*, **10**, giab008.

Di Vittori, V., **Bitocchi, E.**, **Rodriguez, M.**, **Alseekh, S.**, **Bellucci, E.**, **Nanni, L.** ***et al.*** (2021) Pod indehiscence in common bean is associated with the fine regulation of *PvMYB26*. *Journal of Experimental Botany*, **72**, 1617–1633.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

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

Fuller, D.Q. & **Harvey, E.L.** (2006) The archaeobotany of Indian pulses: identification, processing and evidence for cultivation. *Environmental Archaeology*, **11**, 219–246.

Funatsuki, H., **Suzuki, M.**, **Hirose, A.**, **Inaba, H.**, **Yamada, T.**, **Hajika, M. *et al.*** (2014) Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 17797–17802.

Garcia, 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. *Nature Communications*, **12**, 702.

Hammer, K. (1984) Das domestikationssyndrom. *Die Kulturpflanze*, **32**, 11–34.

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.

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. *Nature Communications*, **5**, 5443.

Keatinge, J.D.H., **Easdown, W.J.**, **Yang, R.Y.**, **Chadha, M.L.** & **Shanmugasundaram, S.** (2011) Overcoming chronic malnutrition in a future warming world: the key importance of mungbean and vegetable soybean. *Euphytica*, **180**, 129–141.

Kim, D., **Paggi, J.M.**, **Park, C.**, **Bennett, C.** & **Salzberg, S.L.** (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, **37**, 907–915.

Lambrides, C.J. & **Godwin, I.D.** (2007) Mungbean. In: Kole, C. (Ed.) *Genome mapping and molecular breeding in plants*. Heidelberg, Berlin: Springer, pp. 69–90.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, **1303**, 3997v2.

Li, S., **Ding, Y.**, **Zhang, D.**, **Wang, X.**, **Tang, X.**, **Dai, D. *et al.*** (2018) Parallel domestication with a broad mutational spectrum of determinate stem growth habit in leguminous crops. *Plant Journal*, **96**, 761–771.

Li, S., **Wang, X.**, **Xu, W.**, **Liu, T.**, **Cai, C.**, **Chen, L. *et al.*** (2021) Unidirectional movement of small RNAs from shoots to roots in interspecific heterografts. *Nature Plants*, **7**, 50–59.

Lin, Y.P., **Chen, H.W.**, **Yeh, P.M.**, **Anand, S.S.**, **Lin, J.**, **Li, J. *et al.*** (2023) Demographic history and distinct selection signatures of two domestication genes in mungbean. *Plant Physiology*, **193**, 1197–1212.

Lin, Z., **Li, X.**, **Shannon, L.M.**, **Yeh, C.T.**, **Wang, M.L.**, **Bai, G. *et al.*** (2012) Parallel domestication of the *Shattering1* genes in cereals. *Nature Genetics*, **44**, 720–724.

Liu, C., **Wang, Y.**, **Peng, J.**, **Fan, B.**, **Xu, D.**, **Wu, J. *et al.*** (2022) High-quality genome assembly and pan-genome studies facilitate genetic discovery in mung bean and its improvement. *Plant Communications*, **3**, 100352.

Liu, H., **Liu, H.**, **Zhou, L.**, **Zhang, Z.**, **Zhang, X.**, **Wang, M. *et al.*** (2015) Parallel domestication of the *Heading Date 1* gene in cereals. *Molecular Biology and Evolution*, **32**, 2726–2737.

Liu, Y., **Du, H.**, **Li, P.**, **Shen, Y.**, **Peng, H.**, **Liu, S. *et al.*** (2020) Pan-genome of wild and cultivated soybeans. *Cell*, **182**, 162–176.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, **17**, 10–12.

McKenna, A., **Hanna, M.**, **Banks, E.**, **Sivachenko, A.**, **Cibulskis, K.**, **Kernytsky, A. *et al.*** (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

Ong, P., **Lin, Y.**, **Chen, H.**, **Lo, C.**, **Burlyaeva, M.**, **Noble, T. *et al.*** (2023) Environment as a limiting factor of the historical global spread of mungbean. *eLife*, **12**, e85725.

Pan, L., **Liu, M.**, **Kang, Y.**, **Mei, X.**, **Hu, G.**, **Bao, C. *et al.*** (2023) Comprehensive genomic analyses of *Vigna unguiculata* provide insights into population differentiation and the genetic basis of key agricultural traits. *Plant Biotechnology Journal*, **21**, 1426–1439.

Parker, T.A., **Berny Mier, Y.**, **Teran, J.C.**, **Palkovic, A.**, **Jernstedt, J.** & **Gepts, P.** (2020) Pod indehiscence is a domestication and aridity resilience trait in common bean. *New Phytologist*, **225**, 558–570.

Pertea, M., **Pertea, G.M.**, **Antonescu, C.M.**, **Chang, T.C.**, **Mendell, J.T.** & **Salzberg, S.L.** (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, **33**, 290–295.

Ping, J., **Liu, Y.**, **Sun, L.**, **Zhao, M.**, **Li, Y.**, **She, M. *et al.*** (2014) *Dt2* is a gain-of-function MADS-domain factor gene that specifies semideterminacy in soybean. *The Plant Cell*, **26**, 2831–2842.

Purugganan, M.D. (2019) Evolutionary insights into the nature of plant domestication. *Current Biology*, **29**, R705–R714.

Rhee, S.Y., **Beavis, W.**, **Berardini, T.Z.**, **Chen, G.**, **Dixon, D.**, **Doyle, A. *et al.*** (2003) The *Arabidopsis* information resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Research*, **31**, 224–228.

Rozas, J., **Ferrer-Mata, A.**, **Sánchez-DelBarrio, J.C.**, **Guirao-Rico, S.**, **Librado, P.**, **Ramos-Onsins, S.E. *et al.*** (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, **34**, 3299–3302.

Sangiri, C., **Kaga, A.**, **Tomooka, N.**, **Vaughan, D.** & **Srinives, P.** (2007) Genetic diversity of the mungbean (*Vigna radiata*, Leguminosae) gene-pool on the basis of microsatellite analysis. *Australian Journal of Botany*, **55**, 837–847.

Schafleitner, R., **Nair, R.M.**, **Rathore, A.**, **Wang, Y.W.**, **Lin, C.Y.**, **Chu, S.H. *et al.*** (2015) The AVRDC-The World Vegetable Center mungbean (*Vigna radiata*) core and mini core collections. *BMC Genomics*, **16**, 344.

Schmutz, J., **Cannon, S.B.**, **Schlueter, J.**, **Ma, J.**, **Mitros, T.**, **Nelson, W. *et al.*** (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, **463**, 178–183.

Somta, P., **Seehalak, W.** & **Srinives, P.** (2009) Development, characterization and cross-species amplification of mungbean (*Vigna radiata*) genic microsatellite markers. *Conservation Genetics*, **10**, 1939–1943.

Song, Z.Y., **Tian, J.L.**, **Fu, W.Z.**, **Li, L.**, **Lu, L.H.**, **Zhou, L. *et al.*** (2013) Screening Chinese soybean genotypes for *agrobacterium*-mediated genetic transformation suitability. *Journal of Zhejiang University-SCIENCE B*, **14**, 289–298.

Subramanian, B., **Gao, S.**, **Lercher, M.J.**, **Hu, S.** & **Chen, W.H.** (2019) Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Research*, **47**, W270–W275.

Swanson-Wagner, R., **Briskine, R.**, **Schaefer, R.**, **Hufford, M.B.**, **Ross-Ibarra, J.**, **Myers, C.L. *et al.*** (2012) Reshaping of the maize transcriptome by domestication. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 11878–11883.

Takahashi, Y., **Kongjaimun, A.**, **Muto, C.**, **Kobayashi, Y.**, **Kumagai, M.**, **Sakai, H. *et al.*** (2020) Same locus for non-shattering seed pod in two independently domesticated legumes, *Vigna angularis* and *Vigna unguiculata*. *Frontiers in Genetics*, **11**, 748.

Tamura, K., **Stecher, G.** & **Kumar, S.** (2021) MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, **38**, 3022–3027.

Tian, Z., **Wang, J.W.**, **Li, J.** & **Han, B.** (2021) Designing future crops: challenges and strategies for sustainable agriculture. *Plant Journal*, **105**, 1165–1178.

Wang, M., **Li, W.**, **Fang, C.**, **Xu, F.**, **Liu, Y.**, **Wang, Z. *et al.*** (2018) Parallel selection on a dormancy gene during domestication of crops from multiple families. *Nature Genetics*, **50**, 1435–1441.

Zhang, H., **Goettel, W.**, **Song, Q.**, **Jiang, H.**, **Hu, Z.**, **Wang, M.L. *et al.*** (2020) Selection of *GmSWEET39* for oil and protein improvement in soybean. *PLoS Genetics*, **16**, e1009114.

Zhang, H., **Mascher, M.**, **Abbo, S.** & **Jayakodi, M.** (2022) Advancing grain legumes domestication and evolution studies with genomics. *Plant and Cell Physiology*, **63**, 1540–1553.