nature genetics

Article

https://doi.org/10.1038/s41588-026-02644-5

Graph-based pan-genome reveals structural variations associated with agronomic traits in mung bean

Received: 25 February 2025

Accepted: 22 May 2026

Published online: 10 July 2026

Check for updates

Honglin Chen[1,12] ✉, Longsheng Xing[2,12], Chaonan Guan[1,3,12], Yang Liu[1,12], Tianxiao Chen[1], Rong Cao[1], Qiannan Song[1], Rutwik Barmukh[4], Reyazul Rouf Mir[4,5], Liangliang Hu[1], Bin Zhou[6], Gaoling Luo[7], Dongxu Xu[8], Fengxiang Yin[9], Xingxing Yuan[10], Suhua Wang[1], Xin Chen[10], Lixia Wang[1], Chengzhi Jiao[11] ✉, Huilong Du[2] ✉, Meiliang Zhou[1] ✉, Rajeev K. Varshney[4] ✉ & Xuzhen Cheng[1] ✉

Mung bean (*Vigna radiata*) is a globally important legume crop valued for its short growing cycle, nitrogen-fixing capacity and high nutritional value, particularly in developing countries. Here we report a comprehensive graph-based pan-genome assembled from 11 genetically diverse global accessions. The framework captures 75,268 gene families (50.86% core, 35.19% dispensable and 13.95% private) and 66,862 nonredundant structural variants. Integrating these structural variants and single nucleotide polymorphisms, genome-wide association studies across five environments identified candidate genes for 20 agronomic traits, underscoring the pivotal roles of these variants in driving mung bean domestication and improvement. Mechanistically, we demonstrate that a 68-bp promoter insertion in *VrTIFY6B* and a 136-bp promoter deletion in *VrPGIP1* regulate flavonoid content and confer bruchid resistance, respectively. These genomic resources and actionable functional variants provide a powerful toolkit to accelerate mung bean improvement through marker-assisted breeding, genomic selection and genome editing to address global food security.

Mung bean (*Vigna radiata*, 2*n* = 22) is a crucial pulse crop cultivated widely in Asia for its economic value, nutritional benefits and environmental adaptability[1]. As a self-pollinating species within the *Vigna* genus of the Fabaceae family, it represents a principal global grain legume food crop[2]. Its seeds and sprouts are rich in protein, essential amino acids, vitamins and bioactive compounds such as flavonoids, conferring substantial nutritional and medicinal properties[3]. Owing to its nitrogen-fixing ability, short life cycle, broad adaptability and low water requirements, mung bean is integral to sustainable agriculture[4]. Globally, it is cultivated on more than 7.2 million hectares, yielding 6 million tons annually, with India and China being the primary producers[5,6]. Archeological and genetic evidence indicates domestication occurred in India approximately 3,500 years ago from its wild ancestor, *V. radiata* var. *sublobata*[7]. Extensive genetic diversity characterizes the Chinese mung bean germplasm[5], providing valuable resources for research on its domestication and improvement.

Structural variations (SVs), including insertions, deletions, inversions and translocations, drive genetic diversity and profoundly influence gene regulation and phenotypic traits[8]. Pan-genomes integrate population-scale genomic data, facilitating the systematic identification of these variants. Indeed, recent pan-genome studies in various crop plants, including rice[9], soybean[10], wheat[11], chickpea[12], barley[13], foxtail millet[14] and broomcorn millet[15,16], have established that SVs play critical roles in crop domestication, trait determination[17] and

A full list of affiliations appears at the end of the paper. ✉e-mail: chenhonglin@caas.cn; chengzhi.jiao@pku-iaas.edu.cn; huilongdu@hbu.edu.cn; zhoumeiliang@caas.cn; Rajeev.Varshney@murdoch.edu.au; chengxuzhen@caas.cn

https://doi.org/10.1038/s41588-026-02644-5

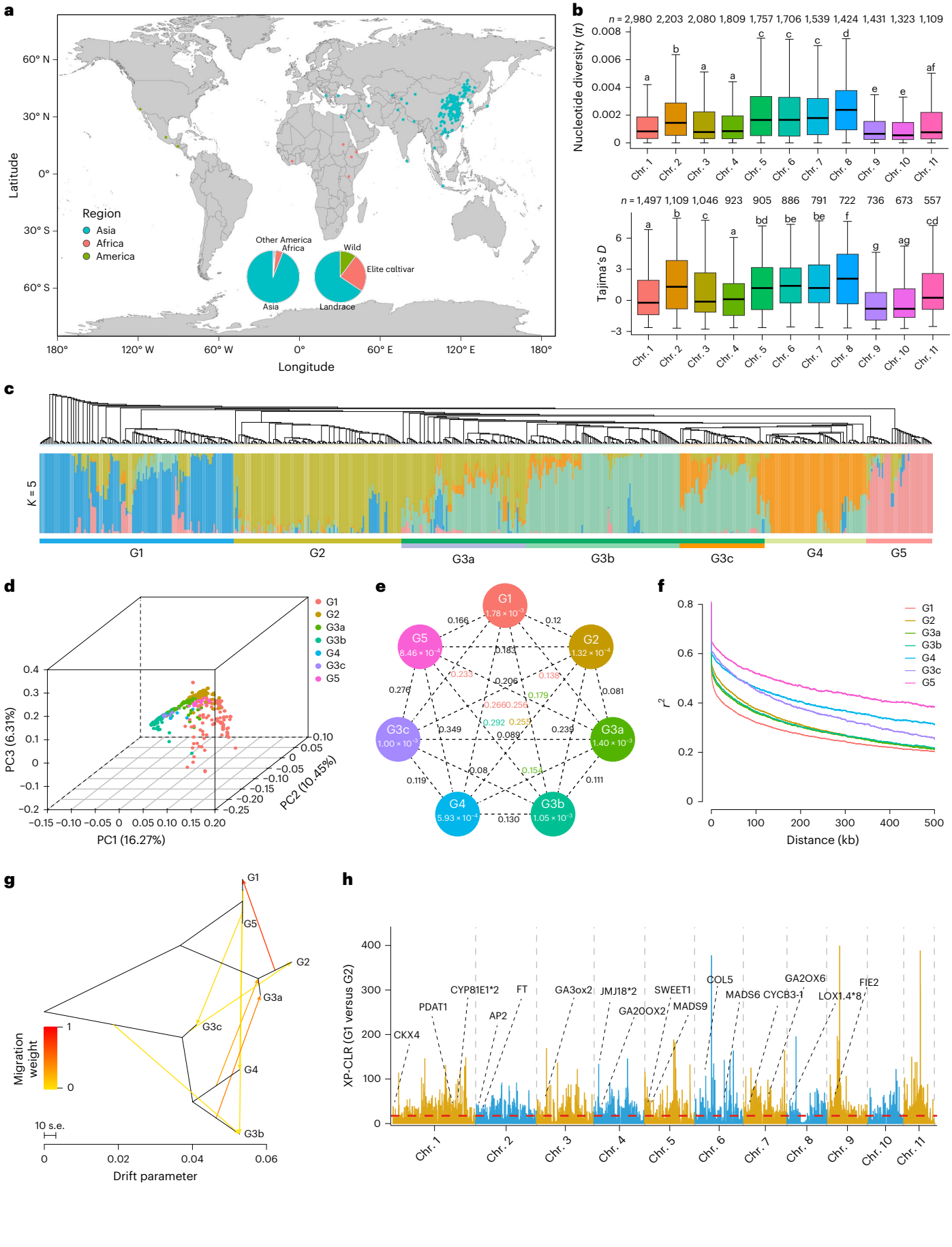


Fig. 1 | Population genomics of 580 mung bean accessions. a, Geographic distribution of the 580 mung bean accessions. Pie charts depict the proportional distribution of accessions across continents (left) and ecotypes (right). **b**, Genome-wide distribution of π and Tajima's D along the 11 mung bean chromosomes. Center line: median; box boundaries: upper and lower quartiles; whiskers: 1.5× interquartile range (IQR). Different letters indicate significant differences between comparison groups (d.f. = 19,360, $P < 2.2 \times 10^{-16}$). **c**, Phylogenetic tree (top) and population structure (bottom) of the 580 accessions. Population structure was inferred using ADMIXTURE at $K = 5$. **d**, PCA of the 580 accessions. **e**, Nucleotide diversity (π, shown within each circle) and population differentiation (F_{ST}, dotted lines) among mung bean subgroups. **f**, LD decay analysis for different mung bean subgroups. **g**, Gene flow analysis between different mung bean subgroups based on TreeMix. **h**, Selective sweep analysis between two subgroups of mung bean accessions (G1 versus G2) based on XP-CLR score. The map in **a** was generated using the R package ggplot2 (ref. 47).

genetic improvement[13]. Despite its widespread cultivation, genetic research on mung bean has lagged behind that of other legumes. Although foundational genomic resources exist, including a single linear reference genome for accession VC1973A[1] and a subsequent pan-genome built using a map-to-pan strategy[18], these approaches remain constrained in resolving complex SVs and accurately capturing the complete gene repertoire of the species. Although population-scale short-read sequencing data have elucidated the mung bean population structure and the genetic basis of several key agronomic traits[18,19], these single nucleotide polymorphism (SNP)-centric approaches struggle inherently to capture larger structural changes. Consequently, a comprehensive, high-resolution view of mung bean's pan-genomic diversity and the functional impact of SVs on key agronomic traits remains elusive.

Here we report chromosome-level de novo genome assemblies for 11 representative wild and cultivated mung bean accessions to construct a graph-based pan-genome for this species. This comprehensive resource enabled the discovery of a vast catalog of SVs and the investigation of their impact on agronomically important traits. By integrating whole-genome resequencing data from 580 germplasm accessions, we elucidated the complex population structure and evolutionary history of mung bean. Furthermore, our pan-genome-enabled genome-wide association studies (GWAS) based on both SVs and SNPs pinpointed key regulatory genes associated with important agronomic traits, highlighting the critical role of previously hidden structural variations in mung bean adaptation and providing actionable targets for its genetic improvement.

Results

Genetic diversity and population structure of 580 mung bean accessions

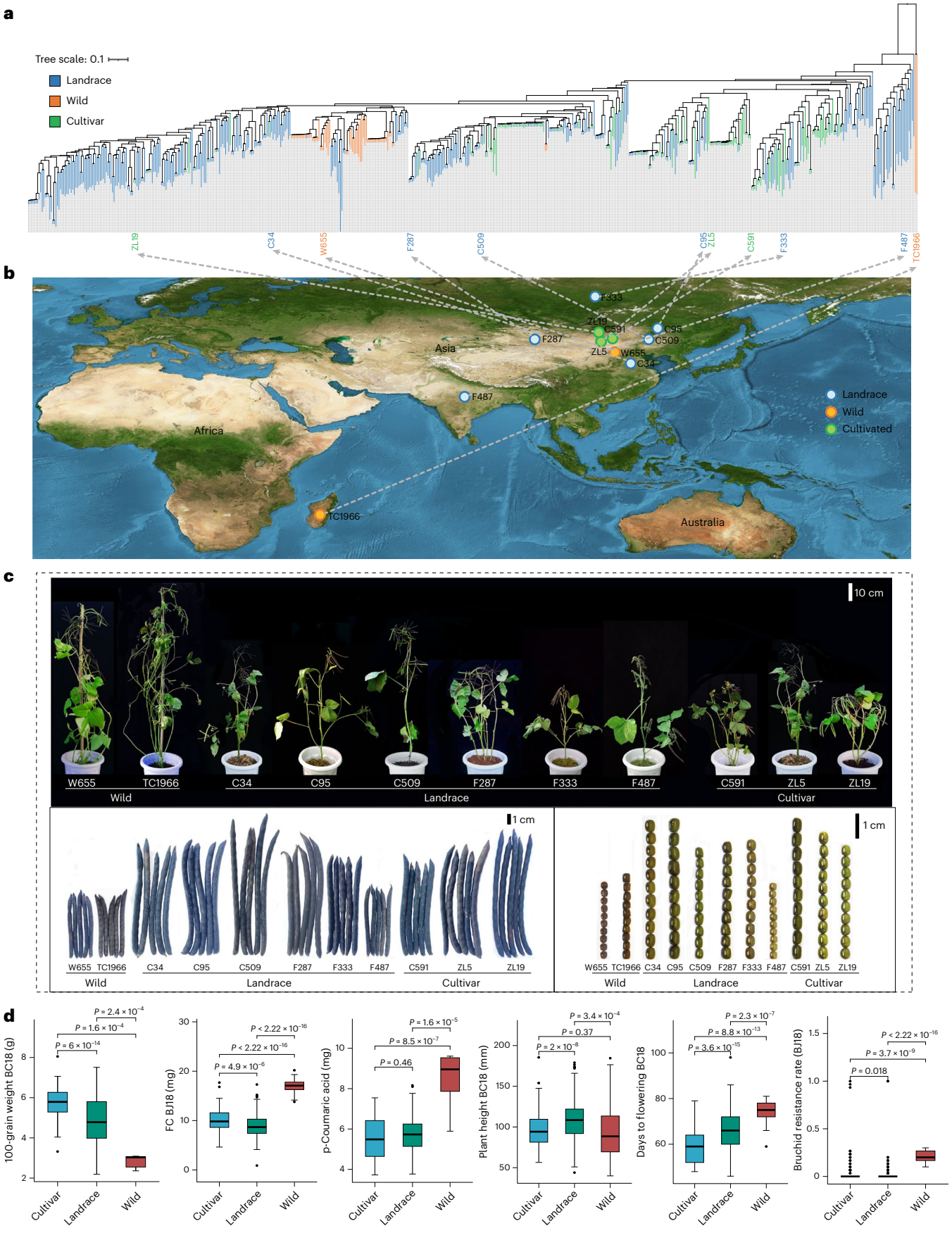
To define the global genetic landscape of mung bean, we performed whole-genome resequencing of 580 accessions, comprising 426 landraces, 97 elite cultivars and 57 wild accessions (Fig. 1a and Supplementary Table 1). The collection represents the species' global footprint, with most accessions (93.97%) originating from Asia, followed by Africa (4.31%) and other regions (1.72%). This effort yielded a comprehensive catalog of 6,222,546 SNPs and 713,751 indels. Among these, 278,679 SNPs and 12,344 indels mapped to coding regions. Notably, 151,653 SNPs were predicted to alter amino acid sequences, extend transcripts or generate premature stop codons, whereas 8,372 indels induced frameshift mutations. We further examined the genome-wide distribution of nucleotide diversity (π) and Tajima's D across chromosomes (Fig. 1b). Genetic diversity was distributed unevenly, with chromosomes (chrs.) 2, 5, 6, 7 and 8 exhibiting higher diversity (π: 1.2×10^{-3} to 1.6×10^{-3}) than the remaining chromosomes (π: 5.0×10^{-4} to 5.3×10^{-4}) (Fig. 1b), suggesting differential chromosomal contributions during domestication and adaptation to new habitats.

Based on 1,938,724 high-quality SNPs, phylogenetic and population structure analyses identified five primary groups ($K = 5$; G1–G5), which further resolved into seven subpopulations when accounting for G3's topology and geography (Fig. 1c and Supplementary Fig. 1). These subpopulations were corroborated by principal component analysis (PCA) (Fig. 1d and Supplementary Fig. 2). Notably, groups G1, G2, G3b, G4 and G5 remained stable across varying K values (Supplementary Fig. 3). Geographically, G1–G5 accessions span South, Southeast and West Asia, and throughout China. Furthermore, genetic diversity was highest in G1 ($\pi = 1.78 \times 10^{-3}$), and lowest in G2 ($\pi = 1.32 \times 10^{-4}$), with population fixation statistics (F_{ST}) showing moderate differentiation between G1 and others (0.12–0.266) (Fig. 1e). Subgroups within China exhibited weaker divergence, particularly G3b versus G3c ($F_{ST} = 0.08$) and G2 versus G3a ($F_{ST} = 0.081$), consistent with the genetic structure and PCA results (Fig. 1c,d). In contrast, G5 displayed the strongest differentiation from G4 (0.349), G3b (0.292) and G3c (0.276), probably reflecting geographic isolation. The rapid linkage disequilibrium (LD) decay observed in G1 (G5 > G4 > G3c > G2 > G3a ≥ G3b > G1) points to it as the potential center of origin (Fig. 1f). Similarly, G3b showed rapid LD decay among the Chinese subpopulations, consistent with its wild status and lower domestication level. Conversely, the slower decay in G5, G4 and G3c suggests intensive selection. TreeMix analysis revealed eight potential gene flow events (Fig. 1g and Supplementary Fig. 4), including directional gene flow from the center of origin (G1) to G3b and from G2 to G1. These events highlight the bidirectional gene exchange shaping mung bean's global genetic structure.

To evaluate selection pressures during mung bean domestication, we conducted cross-population composite likelihood ratio (XP-CLR) analyses, identifying 4,849–4,862 selective signals across subgroup comparisons (Supplementary Tables 2–5). Notably, 2,615 and 2,667 genes within selective regions in G1 versus G2 and G2 versus G3b, respectively, were associated with differentiation, whereas 2,458 and 3,019 genes in G3b versus G4 and G3b versus G5 were implicated in crop improvement (Fig. 1h and Supplementary Fig. 5). Functional enrichment revealed that genes related to pectinesterase activity, cell wall modification and enzyme inhibition were enriched in G3b versus G4, whereas genes involved in flavonoid and sesquiterpene metabolism were enriched in G3b versus G5, reflecting distinct human selection on these pathways (Supplementary Table 6). Furthermore, 552 genes were under selection in both G3b versus G4 and G3b versus G5, including known candidates governing seed weight (*PP2C* homolog)[20], flowering time (*SKIP14*)[21], disease resistance (*ASP*)[22] and flavonoid metabolism (*FLS*)[23]. These results highlight the genetic drivers of agronomic trait improvement from wild to cultivated mung bean.

Fig. 2 | Phenotypic diversity and selection of 11 representative mung bean accessions. a, Phylogenetic tree of the 580 accessions, with the 11 accessions selected for de novo genome assembly highlighted. **b**, Geographic distribution of the 11 selected accessions. **c**, Representative images showing variation in plant height, pod length and seed size for the 11 accessions. **d**, Comparison of key agronomic traits among wild, landrace and elite cultivars. Boxplots illustrate the distribution of 100-grain weight, FC, *p*-coumaric acid content, plant height, days to flowering and bruchid resistance. The number of samples in cultivar, landrace and wild in boxplots is 140, 381 and 59, respectively. BC18 and BJ18 indicate Baicheng 2018 and Beijing 2018, respectively. In boxplots: center line: median; box edges 25th and 75th percentiles; whiskers: 1.5× IQR. Statistical significance was assessed using a two-sided Wilcoxon rank-sum test. Credit: Basemap adapted from NASA Earth Observatory (https://assets.science.nasa.gov/content/dam/science/esd/eo/images/bmng/bmng-base/july/world.200407.3x5400x2700.jpg and https://urldefense.com/v3/__https:/assets.science.nasa.gov/content/dam/science/esd/eo/images/bmng/bmng-base/july/world.200407.3x5400x2700.jpg__;!!NLFGqXoFfo8MMQ!smctPKMOAEROBwBjI240zdEYq70sj73Jmf5PmLKpzUdf0OAQy9Aam7wzRWOlB7Mz3jbhnKGcXogXbEJTltWli7zVpZwIm9aNFA$).

https://doi.org/10.1038/s41588-026-02644-5



https://doi.org/10.1038/s41588-026-02644-5

Table 1 | Quality assessment and annotation of 11 mung bean genome assemblies

Accession	Assembly size (Mb)	Contig N50 (Mb)	Scaffold N50 (Mb)	Gap number	Chromosome anchoring rate (%)	Gene number	TE content (%)	BUSCO completeness (%)	Short-read mapping rate (%)
C34	499.85	18.33	44.54	50	99.06	34,318	53.45	98.51	99.03
C95	496.42	13.55	45.73	60	98.98	34,311	53.25	98.45	99.13
C509	508.75	15.43	44.40	95	97.12	34,741	54.02	98.39	99.07
C591	503.43	21.02	45.06	27	99.06	34,032	50.09	98.45	99.47
F287	499.17	14.98	44.29	43	98.35	34,508	53.03	98.45	99.11
F333	500.51	17.54	45.78	67	99.30	34,159	53.40	98.45	99.05
F487	502.55	16.71	44.78	77	99.62	34,856	52.70	98.57	98.84
ZL19	497.36	16.41	45.24	44	99.10	34,278	53.65	98.51	99.12
TC1966	494.66	16.79	46.44	49	99.37	34,095	53.25	98.57	97.61
W655	498.15	6.77	44.77	125	98.83	34,164	53.37	98.33	98.95
ZL5	485.46	7.36	44.55	210	98.10	35,882	51.16	98.51	99.54

High-quality genome assemblies of 11 genetically diverse mung bean accessions

To investigate the mung bean genetic diversity, we selected 11 representative accessions from the 580-accession panel, guided by phylogenetic and population structure analyses (Fig. 2a,b and Supplementary Table 7). This set comprised six cultivated accessions (C34, C95, C509, C591, ZL19 and ZL5) and five wild accessions (F287, F333, F487, TC1966 and W655). These accessions exhibited extensive phenotypic diversity across key agronomic traits, including flowering time, plant height, pod length, grain size, 100-grain weight, seed color, flavonoid and *p*-coumaric acid content, and bruchid resistance (Fig. 2c,d). This diversity underscores the necessity of a pan-genome to comprehensively elucidate the underlying genetic basis.

We generated high-quality genome assemblies for the 11 mung bean accessions by integrating Pacific Biosciences (PacBio) HiFi long-read sequencing, high-throughput chromosome conformation capture (Hi-C) and Illumina short-read sequencing (Supplementary Tables 8 and 9). The final genome assemblies achieved an average contig N50 length of 15.00 Mb (range: 13.55–21.02 Mb), with the exceptions of W655 (6.77 Mb) and ZL5 (7.36 Mb) (Table 1). Notably, 97.12–99.63% of the sequences were anchored to 11 chromosomes (Table 1). The haploid genome sizes ranged from 485.46 Mb in ZL5 to 508.75 Mb in C509. The Illumina short reads showed an average alignment rate of 99%, and more than 98% of Benchmarking Universal Single-Copy Orthologues (BUSCO) were identified in all accessions, indicating high assembly completeness (Supplementary Fig. 6). Furthermore, the LTR Assembly Index (LAI) confirmed reference-level quality (LAI > 10) for all assemblies (Supplementary Table 10).

Transposable elements (TEs) constituted 50.09–54.02% of the assembled sequences (Table 1 and Supplementary Tables 11 and 12), predominantly comprising LTR/Other, LTR/*Gypsy*, DNA transposons and LTR/*Copia* elements (Supplementary Fig. 7). Furthermore, we predicted 34,032 to 35,882 protein-coding genes, achieving BUSCO scores of 96.6–98.0% (Table 1, Supplementary Fig. 8 and Supplementary Table 10). Transcriptome data supported 76.81% to 83.61% of predicted exons, and 90.75% to 93.75% of genes received functional annotations (Supplementary Table 13).

The mung bean pan-genome reveals extensive gene content variation

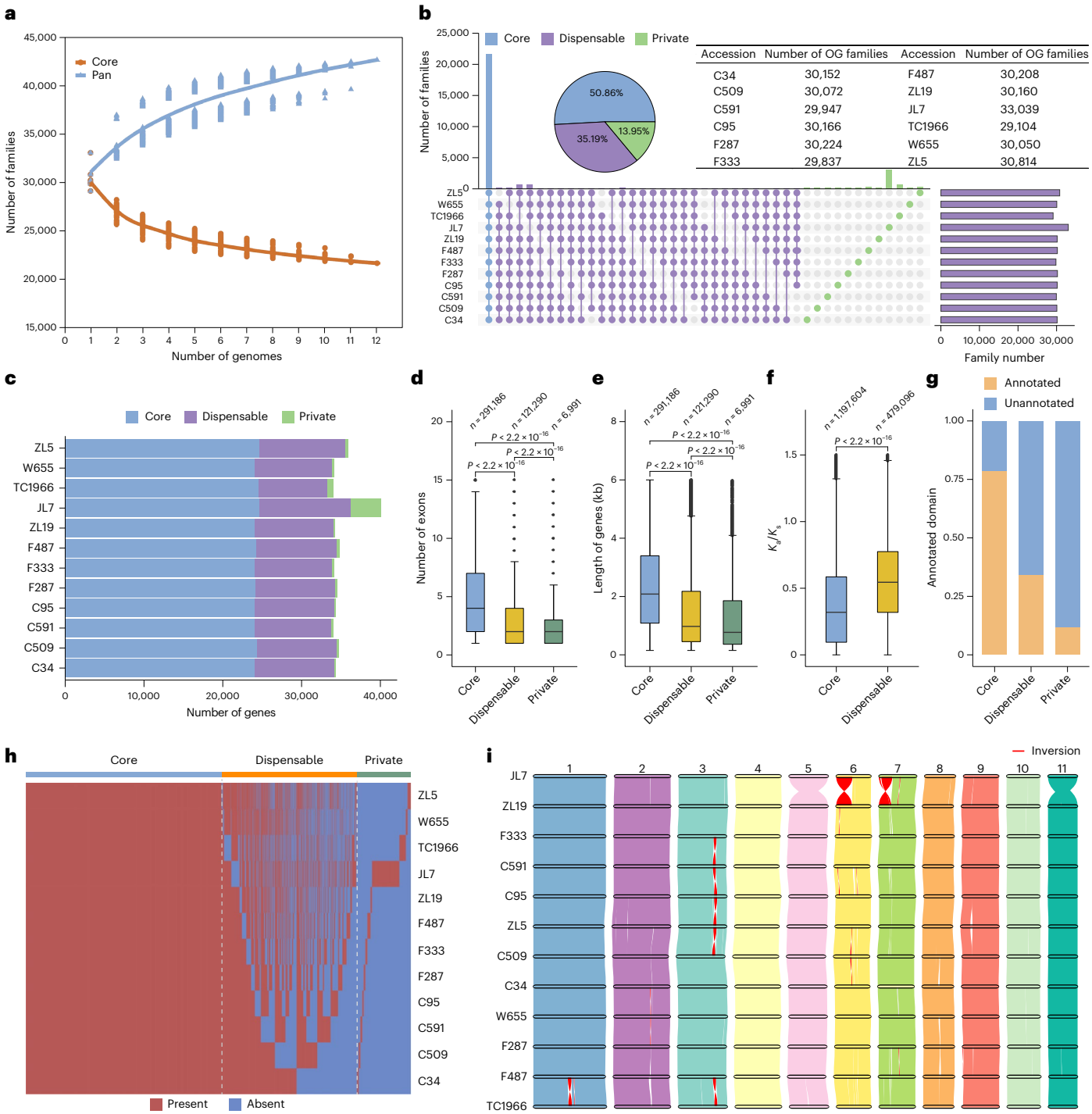
To capture the complete gene repertoire of mung bean, we constructed a pan-genome using protein-coding genes (Methods). The resulting pan-genome comprises 75,268 gene families, categorized into core (50.86%), dispensable (35.19%; present in two or more but not all accessions) and private genes (13.95%; unique to a single accession) (Fig. 3a,b). Compared with the reference cultivar ZL19, the 15,239 newly identified gene families were enriched significantly in pathways mediating environmental adaptation, including fungal response, growth regulation, insect defense and hypoxia response ($P < 10^{-4}$; Supplementary Fig. 9 and Supplementary Table 14).

As expected, core genes predominated across the pan-genome (ranging from 24,075 to 24,693 among the 12 accessions) (Fig. 3c). The number of dispensable genes remained relatively stable, ranging from 8,754 to 11,603. Compared with dispensable and private genes, core genes exhibited more exons (Fig. 3d), a larger average gene size (Fig. 3e), higher expression levels (Supplementary Fig. 10) and lower nonsynonymous substitution rate (K_a)/synonymous substitution rate (K_s) (K_a/K_s) ratios (Fig. 3f), indicating stronger evolutionary constraints. These results align with findings from pan-genome studies of other species including cucumber[24], potato[25] and pea[26]. Furthermore, more than 75% of the core genes harbored Pfam domains, whereas only 30% of the dispensable genes and less than 10% of the private genes did (Fig. 3g). The core genes were enriched primarily in conserved pathways, such as circadian rhythm and glycosyltransferase activity. In contrast, dispensable and private genes were enriched in secondary metabolism pathways, including isoflavonoid biosynthesis, cytochrome P450 activity and sesquiterpenoid biosynthesis (Supplementary Fig. 11), highlighting secondary metabolic diversification among mung bean accessions. Similarly, gene presence/absence variations (PAVs) across mung bean accessions exhibited an uneven genomic distribution (Fig. 3h). Genome-wide synteny analysis revealed an overall 1:1 syntenic relationship among accessions, while identifying several large inversions between cultivated accessions (for example, JL7 versus ZL19) and between wild and cultivated accessions (for example, F333 versus C591) (Fig. 3i).

The landscape of structural variations in mung bean

To establish a consensus set of SVs in mung bean, we performed SV calling across 11 accessions using ZL19 as the reference genome, yielding a total of 66,862 nonredundant SVs (≥50 bp). Insertions (32,836) and deletions (32,716) constituted most of the SVs, whereas translocations (958) and inversions (352) were relatively rare (Fig. 4a,b and Supplementary Table 15). Most insertions and deletions were less than 5 kb in size (Supplementary Fig. 12), and their genomic distribution was uneven, exhibiting notable enrichment on chrs. 5, 6 and 7 (Fig. 4b). To validate SV quality, we manually curated 80 randomly selected SVs ranging from 1.3 kb to 5.0 Mb. Specifically, we validated 40 SVs (1–7 kb) by mapping PacBio HiFi reads to the assemblies, confirming high identification accuracy (Supplementary Fig. 13 and Supplementary Table 16). Furthermore, we utilized Hi-C reads from two accessions (F487 and

https://doi.org/10.1038/s41588-026-02644-5



Accession	Number of OG families	Accession	Number of OG families
C34	30,152	F487	30,208
C509	30,072	ZL19	30,160
C591	29,947	JL7	33,039
C95	30,166	TC1966	29,104
F287	30,224	W655	30,050
F333	29,837	ZL5	30,814

Fig. 3 | Mung bean pan-genome structure and gene family analysis. a, Variations in the number of gene families between the pan-genome and the core genome. The numbers of pan-gene families and core gene families were stated 50 times when using one genome and 100 times when using more than one genome. **b**, Composition of the pan-genome and individual genomes. The pie chart indicates the proportion of core, dispensable and private gene families. The UpSet plot shows the number of gene families in different combinations of mung bean germplasms. The histogram on the right side displays the number of gene families in each genome. Meanwhile, the number of gene families that are identified in each accession was presented in the upper right table. **c**, Distribution of pan-gene categories (core, dispensable, private) within each of the 12 mung bean accessions. **d**–**f**, Comparison of core, dispensable and private pan-genes for number of exons per gene (**d**), gene length (**e**) and K_a/K_s ratios (**f**): center line: median; box boundaries: upper and lower quartiles; whiskers: 1.5× IQR. Statistical significance was determined by a two-sided Wilcoxon rank-sum test. **g**, Percentage of genes in each pan-gene category that contain a recognizable protein domain. **h**, Heatmap of gene PAVs across the 12 accessions. **i**, Macrosynteny among the 12 mung bean genomes. Inversions are highlighted in red.

TC1966) to validate 40 large SVs (100 kb to 5.0 Mb) at 5-kb resolution (Supplementary Fig. 13 and Supplementary Table 16).

We explored SV distribution patterns further. Notably, the number of pan-SVs (SVs present in at least one accession but not shared by all accessions) increased rapidly with additional accessions, whereas core SVs (shared by all accessions) decreased markedly, suggesting that most SVs are accession-specific (Fig. 4c, top). SVs were categorized into core, softcore, dispensable and private groups (Fig. 4c, bottom).

Across accessions, core and softcore SVs represented a minor fraction (0.96–2.52%), whereas dispensable SVs predominated (62.32–89.39%). Notably, the proportion of private SVs was substantially higher in wild accessions F487 (36.32%) and TC1966 (80.38%) than in domesticated lines (8.09–18.53%), reflecting extensive genetic differentiation (Fig. 4c, bottom). Deletions and insertions accounted for most SVs (>98%) (Fig. 4d, top), and their total length was greater in wild than in domesticated accessions (Fig. 4d, bottom). These findings underscore the extensive SV diversity in mung bean and suggest that accession-specific SVs probably emerged independently during domestication and adaptation.

Given that a substantial proportion of SVs originate from TEs[7], we investigated their colocalization. Approximately 73.88% (48,414 of 65,535) of SVs overlapped with TEs—a proportion significantly higher than that observed for TEs in the reference genome (50.5%; $P < 0.001$; Supplementary Fig. 14). LTR retrotransposons, particularly *Gypsy* and *Copia*, as well as DNA transposons, were prevalent in SV regions (Fig. 4e), suggesting that TE activity drives SV formation in mung bean. Although most SVs reside in intergenic regions, a substantial fraction overlaps with coding regions (Fig. 4f). Notably, SVs were more prevalent upstream than downstream of genes, implying their potential roles in transcriptional regulation (Fig. 4f and Supplementary Fig. 15). The upstream, downstream and gene body regions were the primary gene-related segments affected by SVs (Fig. 4g), further supporting their regulatory relevance to nearby genes. Collectively, these findings reveal extensive genetic variation in mung bean germplasm.

SVs have a widespread impact on gene expression and phenotypic variation

Chromosomal distribution analyses indicated that SVs were distributed unevenly across chromosomes, with relatively few detected on chrs. 4, 9 and 10, whereas chrs. 5, 6 and 7 exhibited dense SV accumulations (Fig. 5a). This pattern closely matches the chromosomal distribution of pan-SVs (Fig. 4b). Across the genome, we identified 25 SV hotspot regions, defined as the top 5% of regions with the highest SV frequencies (Fig. 5b and Supplementary Fig. 16). Notably, the distribution of SVs on chrs. 5, 6 and 7 exhibited accession-specific uneven patterns (Fig. 5b).

We compared the expression of genes overlapping SVs (SV-genes) with those lacking SV overlap (non-SV-genes). Based on RNA sequencing (RNA-seq) data from five tissues (Supplementary Table 9), SV-genes exhibited significantly lower expression levels than non-SV-genes ($P < 2.2 \times 10^{-16}$; Supplementary Fig. 17), supported by their distinct expression distribution between the two groups (Fig. 5c). SVs within coding regions induced the most pronounced changes in gene expression, followed by those in upstream regions (Fig. 5d). These SV-genes were enriched significantly in pathways related to plant–pathogen interactions, terpenoid biosynthesis, protein processing and transporter functions (Fig. 5e). Moreover, SV hotspot regions encompassed genes mediating plant–pathogen interactions, secondary metabolite biosynthesis (for example, isoflavonoids), and retinol metabolism (Supplementary Fig. 18).

Several examples illustrate SV-mediated gene regulation. For example, *Vr05g06070*, which encodes a cytochrome P450 involved in secondary metabolite biosynthesis, harbored a 170-bp deletion within its gene body in four accessions (C34, C509, F287 and W655) (Fig. 5f), significantly downregulating its expression based on RNA-seq data (Fig. 5g). Similarly, a 913-bp insertion was detected upstream of *Vr02g06340*, which encodes an E3 ubiquitin-protein ligase involved in plant immunity[27], in accessions C34, F287, TC1966 and W655 (Fig. 5h), with its presence or absence significantly altering expression (Fig. 5i). In accession TC1966, a 543-bp promoter deletion was observed in *Vr02g28470*, which encodes LysM DOMAIN RECEPTOR-LIKE KINASE 3 (LYK3)—an entry receptor in rhizobial nodulation factor signaling[28]. This promoter deletion altered the expression of *Vr02g28470*, leading to the formation of smaller nodules in TC1966 (Supplementary Fig. 19). Furthermore, a 10.54-kb deletion in four accessions (W655, C34, C591 and F287) resulted in the loss of *Vr08g18660*, encoding a TMV resistance protein (Supplementary Fig. 20a), consequently abolishing its expression in these four accessions (Supplementary Fig. 20b). Moreover, a 14.70-kb deletion was found in four accessions (F333, C591, JL7 and ZL5). This deletion SV led to the loss of the last six exons of *Vr09g09330* (encoding a cellulose synthase-like protein) and caused a significant reduction in its expression level (Supplementary Fig. 20c,d). These findings demonstrate that SVs can modulate gene expression, thereby contributing to phenotypic diversity and environmental adaptability among mung bean accessions.

Pan-genome GWAS reveals SVs and SNPs controlling key agronomic traits

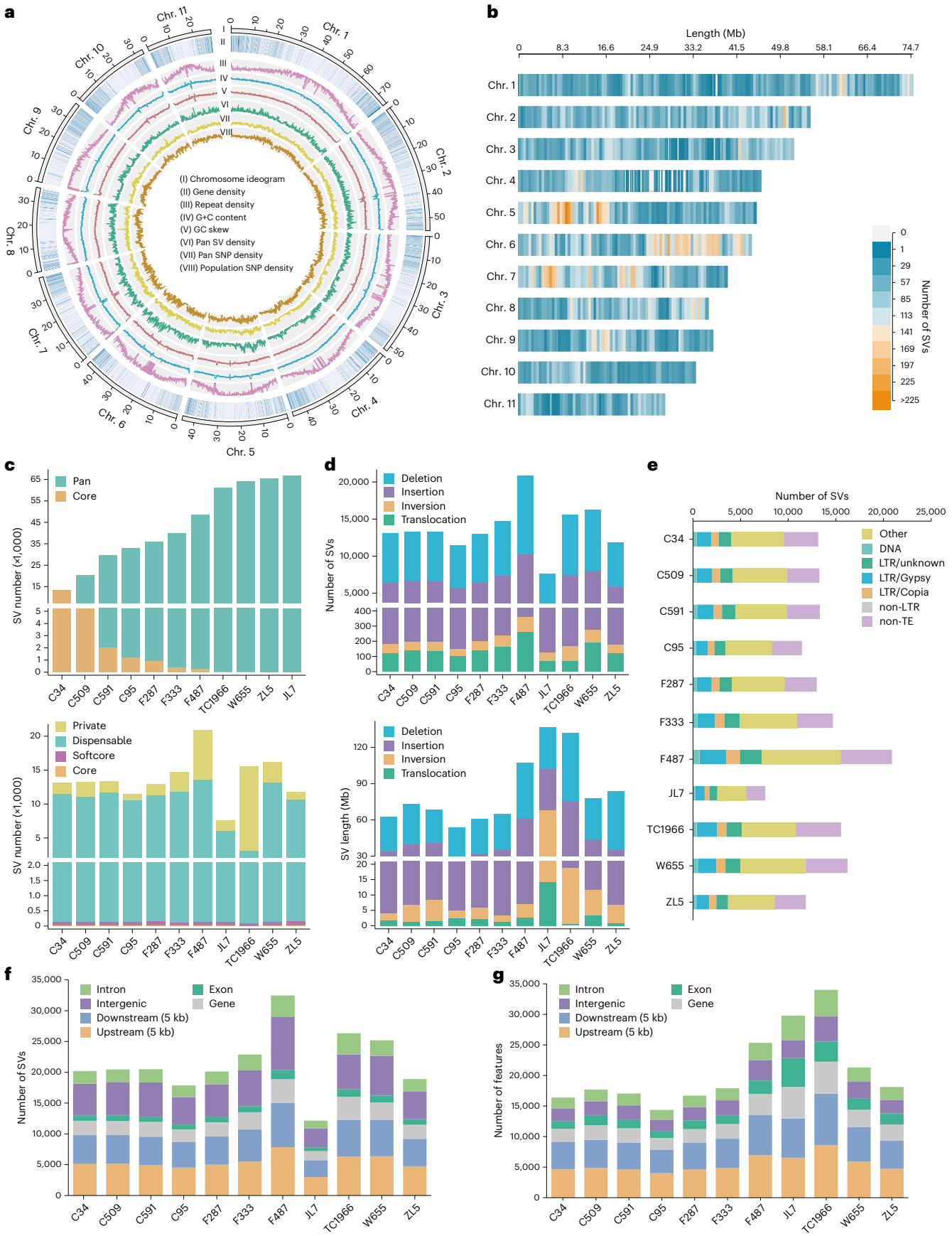
Previous studies indicate that SVs act frequently as causative variations driving trait diversity[29–32]. Therefore, we constructed a mung bean graph-based pan-genome by integrating the linear genome sequence of ZL19 and the 64,415 nonredundant SVs (excluding translocation SVs) derived from representative mung bean genomes. This graph-based genome improves the accuracy of short-read mapping to SV regions, thereby facilitating SV genotyping in the mung bean population. To dissect the genetic architecture of complex traits, we measured 20 agronomic traits across 580 accessions and conducted GWASs using 1,938,724 SNPs and 37,196 SVs (Supplementary Tables 19–21 and Supplementary Notes 1 and 2). Here we highlight three key traits, namely flavonoid content (FC), bruchid resistance and 100-grain weight, which are associated closely with the domestication syndrome and play crucial roles in mung bean improvement. Overall, 17.61% of GWAS signals overlapped between SNPs and SVs (within a 1-Mb flanking region), whereas 73.94% and 8.45% of the signals were identified uniquely by SNPs and SVs, respectively (Supplementary Table 22). We summarized all SVs associated significantly with these main phenotypic traits (Supplementary Table 23), which included both small- and large-scale SVs ranging from 50 to 30,714 bp.

SVs and SNPs in *VrTIFY6B* regulate FC

Flavonoids are essential bioactive compounds in mung bean, with FC representing a key agronomic trait. Integrated SV/SNP-GWAS analyses identified *Vr06g21880* (*VrTIFY6B*, encoding a TIFY6B protein) as associated significantly with FC (Fig. 6a,b and Supplementary Tables 24 and 25). *VrTIFY6B* overlapped with a domestication sweep between the G2 and G3b groups (Fig. 6c), where FC levels were significantly higher in G3b than in G2 ($P < 0.0001$; Fig. 6d). These results indicate that artificial selection during domestication reduced FC. Haplotype analysis based on SNPs revealed three haplotypes with distinct FC differences (Fig. 6e–g). Hap1 exhibited approximately a tenfold higher expression level than Hap2, correlating positively with increased FC levels (Fig. 6k), and its expression peaked during the middle stage of seed development (Fig. 6l). Hap1 occurred at low frequencies in G3b and G3c, whereas Hap2 was present predominantly in G1 and G3a (Supplementary Fig. 22a). Among wild accessions, Hap0 and Hap1 were

Fig. 4 | Landscape of SVs in the mung bean pan-genome. a, Circos plot of the reference genome features and pan-genome SV and SNP features. Tracks from outermost to innermost represent features I–VII as indicated. **b**, Genome-wide distribution and density of pan-SVs across the 11 chromosomes. Color intensity is proportional to SV density. **c**, Distribution of core and pan-SVs (top) and four SVs categories (bottom) in different mung bean germplasms. **d**, Number (top) and total length (bottom) of insertions, deletions, inversions and translocations across the 11 accessions. Four types of SV included insertions, deletions, inversions and translocations. **e**, Proportional overlap of SVs with main TE superfamilies. **f**,**g**, Genomic context of SVs. Number of SVs overlapping various genomic features (**f**). Number of genomic features affected by SVs (**g**). Features are categorized based on their position relative to protein-coding genes.

https://doi.org/10.1038/s41588-026-02644-5



https://doi.org/10.1038/s41588-026-02644-5

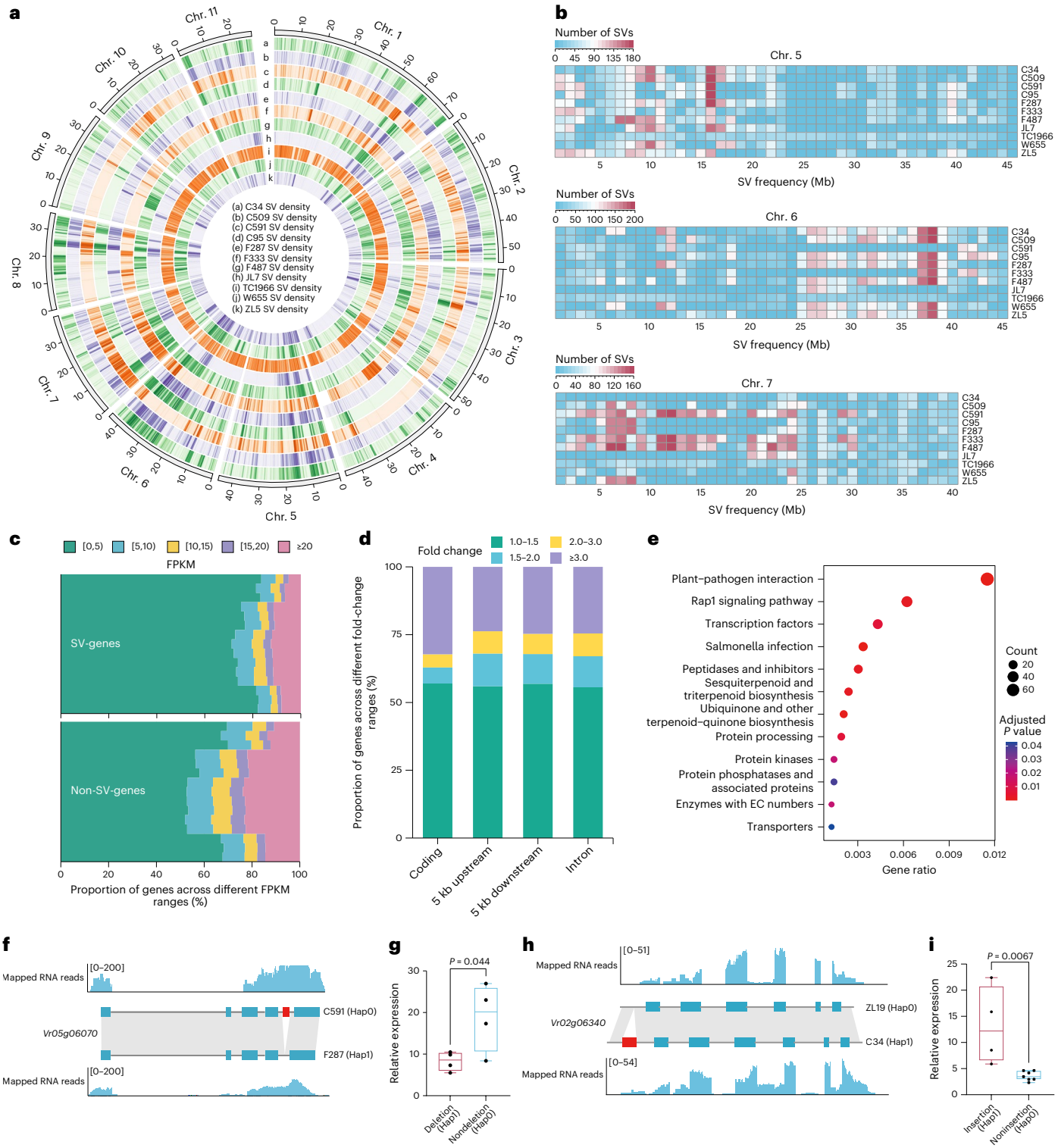
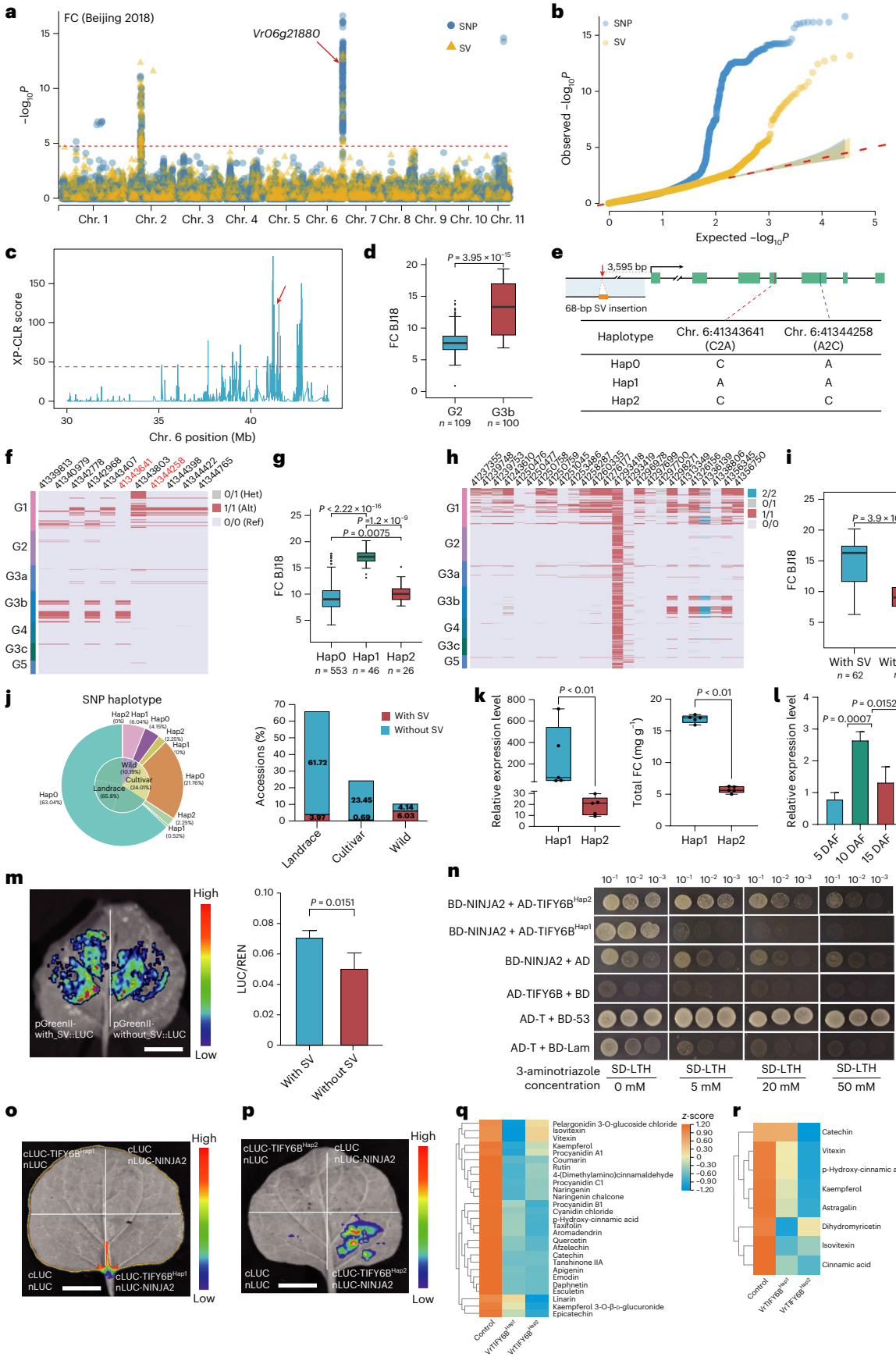


Fig. 5 | SV hotspots and their impact on gene expression. a, Circos plot showing the density of SVs in each of the 11 accessions relative to the ZL19 reference genome. Tracks from outermost to innermost represent features a–k as indicated. **b**, Heatmap of SV frequency on chrs. 5, 6 and 7. Each cell represents the SV count within a 1-Mb genomic interval. The color gradient indicates low and high SV frequency from blue to red, respectively. **c**, Expression level distributions for SV-genes (top) and non-SV-genes (bottom). **d**, Fold change in gene expression between accessions with and without SVs in the coding region, intron, 5 kb upstream of the start codon and 5 kb downstream of the stop codon of protein-coding genes. **e**, Top enriched functional categories for SV-genes. Statistical significance was assessed using a one-sided hypergeometric test, and multiple comparisons were corrected using the Benjamini–Hochberg method. **f**,**g**, A 170-bp deletion in the body of *Vr05g06070* (**f**) is associated with reduced gene expression (**g**). **h**,**i**, A 913-bp insertion upstream of *Vr02g06340* (**h**) is associated with altered gene expression (**i**). In **g** and **i**, statistical significance was assessed using a two-tailed Student's *t*-test.

https://doi.org/10.1038/s41588-026-02644-5



https://doi.org/10.1038/s41588-026-02644-5

Fig. 6 | GWAS identifies *VrTIFY6B* as a candidate gene for FC. a, Manhattan plot from the SNP/SV-GWAS for FC. The candidate gene *Vr06g21880* (*VrTIFY6B*) is located at the peak on chr. 6; dashed red line: Bonferroni-adjusted significance threshold ($P = 1 \times 10^{-5}$). **b**, Quantile-quantile (QQ) plot for the GWAS. **c**, Cross-population composite likelihood ratio (XP-CLR) analysis between G2 and G3b groups on chr. 6, highlighting *VrTIFY6B* within a selective sweep. **d**, FC in the G2 and G3b population groups. Significant difference was determined using a two-sided Wilcoxon rank-sum test. **e**, Gene model for *VrTIFY6B*, showing exons (green boxes), introns (black lines) and the location of key variants. **f**, Haplotype block structure based on SNPs associated with FC. Alt, homozygous for the alternate allele; Het, heterozygous; Ref, homozygous for the reference allele. **g**, The three main SNP-based haplotypes (Hap0, Hap1, Hap2) show significant differences in FC. Significant difference was determined using a two-sided Wilcoxon rank-sum test. **h**, Haplotype block structure based on a 68-bp insertion SV. **i**, FC in individuals with or without the 68-bp insertion. Significant difference was determined using a two-sided Wilcoxon rank-sum test. **j**, Frequency of SNP- and SV-based haplotypes in wild, landrace and cultivated accessions. **k**, Relative expression of *VrTIFY6B* and FC for accessions carrying Hap1 and Hap2. Significant difference was determined using a two-sided Wilcoxon rank-sum test; $n = 5$ biological replicates. **l**, Expression profile of *VrTIFY6B* at three seed development stages (5, 10 and 15 days after flowering (DAF)). Error bar: mean ± s.d.; $n = 3$ biological replicates. Significant difference was tested using a two-tailed Student's *t*-test. **m**, Transcription activity of *VrTIFY6B* promoters with or without the 68-bp insertion SV, measured by a dual-luciferase assay. LUC/REN, firefly luciferase/renilla luciferase. Scale bar: 1 cm. Error bar: mean ± s.d.; $n = 3$ biological replicates. Significant difference was tested using a two-tailed Student's *t*-test. **n**, Y2H assay showing the interaction between VrTIFY6B protein variants (Hap1, Hap2) and the corepressor VrNINJA. BD and AD denote the GAL4 DNA-binding domain and GAL4 activation domain, respectively. BD-NINJA2 represents *VrNINJA2* fused to BD, whereas AD-TIFY6B^{Hap1} and AD-TIFY6B^{Hap2} represent the two VrTIFY6B variants fused to AD. BD and AD alone indicate the corresponding empty-vector controls. AD-T plus BD-53 was used as the positive interaction control, whereas AD-T plus BD-Lam was used as the negative interaction control. Serially diluted yeast cells were grown on synthetic dropout medium lacking leucine, tryptophan and histidine (SD-LTH), supplemented with 0, 5, 20 or 50 mM 3-aminotriazole (3AT). **o**,**p**, LCI assay in *N. benthamiana* leaves showing the interaction between different variants of *VrTIFY6B* and *VrNINJA2*. The color scale indicates relative luminescence signal intensity, ranging from low (blue) to high (red). Representative images are shown in **o** and **p**. Scale bar: 1 cm. **q**,**r**, FC in transgenic hairy roots of buckwheat (**q**) and mung bean (**r**) overexpressing *VrTIFY6B*^{Hap1} and *VrTIFY6B*^{Hap2}. The color scales represent *z*-score-normalized flavonoid contents. In **d**, **g**, **i** and **k**: center line: median; box boundaries: upper and lower quartiles; whiskers: 1.5× IQR.

detected at frequencies of 4.15% and 6.04%, respectively; in cultivated accessions, Hap0 accounted for 21.76% and Hap2 for 2.25%, whereas landraces contained all three haplotypes (Fig. 6j). SV-based haplotype analysis further demonstrated that SVs were strongly associated with FC (Fig. 6e,h). A 68-bp insertion located 3,595 bp upstream of *VrTIFY6B* differentiated accessions based on FC levels (Fig. 6i), and was distributed widely among wild, cultivated and landrace accessions (Fig. 6j). Across all mung bean accessions, the 'noninsertion' genotype was most prevalent, whereas the insertion was detected in G1, G2, G3a, G3b and G3c (Supplementary Fig. 22b). Further analysis indicated that a nonsynonymous mutation in *VrTIFY6B* was linked tightly with the 68-bp insertion SV (Supplementary Fig. 22c,e). Notably, individuals carrying both SNP Hap1 and the insertion SV exhibited higher FC than those carrying only the insertion SV (Supplementary Fig. 22d). These results indicate that both SNPs and SVs contribute synergistically to variation in FC, consistent with their broader role in domestication and crop improvement[33].

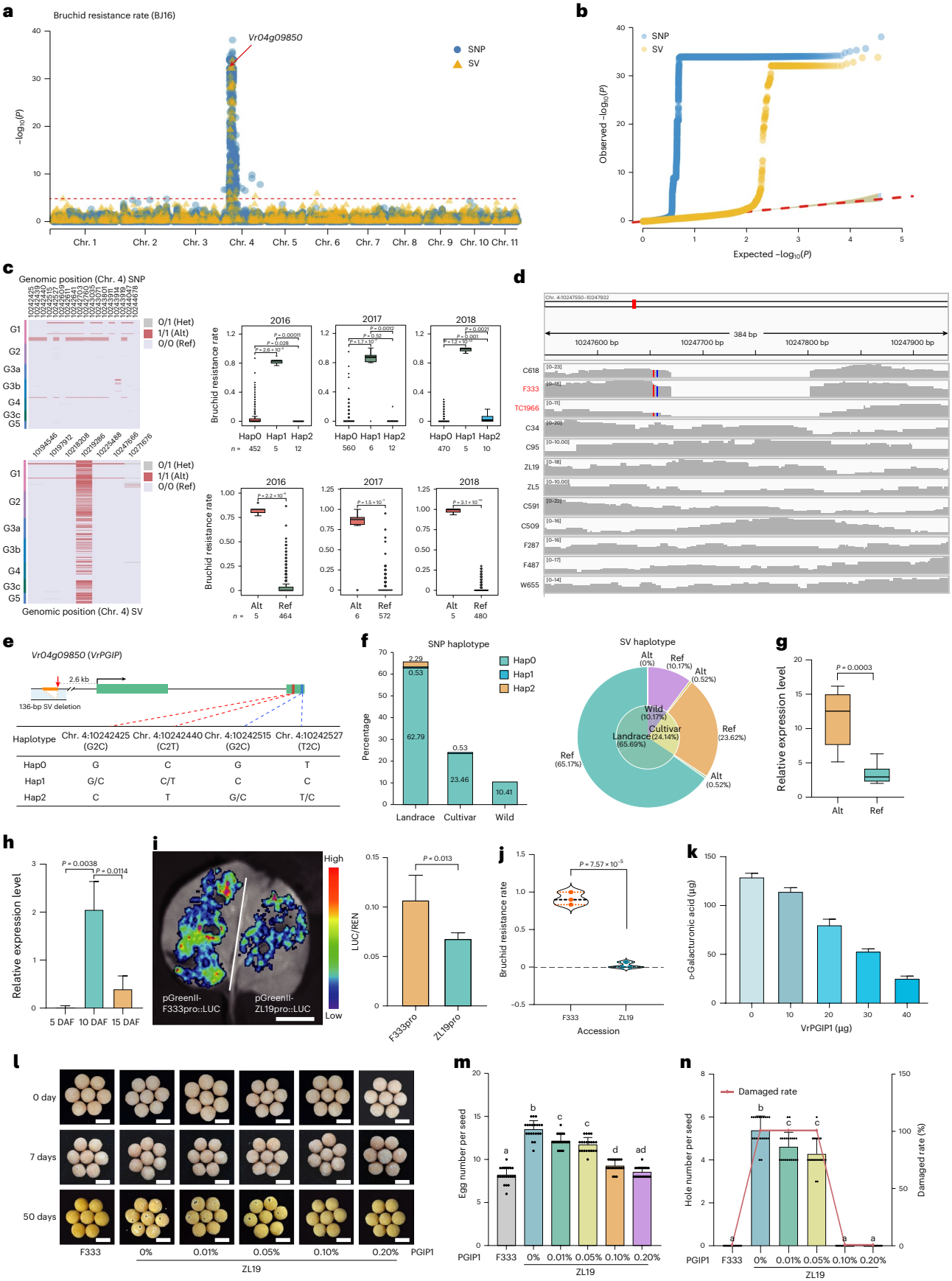
We further explored the functional impact of these variations. Transient activation assays revealed that the *VrTIFY6B* promoter variant 'with_SV' drove higher luciferase (*LUC*) expression compared to the 'without_SV' variant (Fig. 6m), indicating that the SV modulates *VrTIFY6B* transcriptional activity. Given that the SNPs within the coding exons of *VrTIFY6B* lead to nonsynonymous substitutions (Fig. 6e), we investigated their effects on downstream protein interactions. The TIFY motif—a critical component and repressor of the jasmonic acid signaling pathway—binds to domain C of its corepressor NINJA[33,34], thereby repressing flavonoid production[35,36]. Yeast two-hybrid (Y2H) and luciferase complementation imaging (LCI) assays demonstrated that *VrTIFY6B*^{Hap2} interacted more strongly with *VrNINJA* than with *VrTIFY6B*^{Hap1} in vitro (Fig. 6n–p). As mung bean and Tartary buckwheat share a similar flavonoid profile, the regulatory role of *VrTIFY6B* was examined in both species. Overexpression of *VrTIFY6B* in mung bean and Tartary buckwheat confirmed its regulatory role in flavonoid biosynthesis (Fig. 6q–r), consistent with previous findings that TIFY proteins can inhibit jasmonic acid synthesis[37]. The fact that both species harbor comparable flavonoid compounds, together with evidence that buckwheat JAZ proteins (also known as TIFY proteins) suppress flavonoid biosynthesis by regulating the *FtMYB11* repressor[38], highlights the conserved function of *VrTIFY6B* in flavonoid metabolism across species. Collectively, these findings demonstrate that SVs and SNPs contribute jointly to variation in FC through transcriptional and protein-level modulation, consistent with their broader role in shaping domestication-related traits and crop improvement[30].

A promoter deletion in *VrPGIP1* confers bruchid resistance

Bruchid resistance represents a critical agronomic trait for mung bean breeding and genetic improvement. Integrated SV- and SNP-GWAS analyses pinpointed *Vr04g09850* (*VrPGIP1*, encoding a polygalacturonase-inhibiting protein) as the primary candidate gene (Fig. 7a,b and Supplementary Tables 26 and 27). Haplotype analysis incorporating both SNPs and SVs within *VrPGIP1* revealed two haplotypes governing trait variation across the mung bean population (Fig. 7c, left). Bruchid resistance levels differed significantly between accessions harboring these two haplotypes (Fig. 7c, right). Specifically, a 136-bp deletion located 2.6 kb upstream of *VrPGIP1* was present in highly resistant accessions but was absent in susceptible accessions

Fig. 7 | GWAS identifies a promoter deletion in *VrPGIP1* that enhances bruchid resistance. a, Manhattan plot from the SNP- and SV-GWAS for bruchid resistance; dashed red line: Bonferroni-adjusted significance threshold ($P = 1 \times 10^{-5}$). **b**, QQ plot representation for GWAS analysis of bruchid resistance rate. **c**, Haplotype analysis and corresponding bruchid resistance levels for haplotypes defined by SNPs and a 136-bp deletion SV. Statistical test was performed using a two-sided Wilcoxon rank-sum test. **d**, Integrative Genomics Viewer screenshot showing read depth across the *VrPGIP1* promoter, confirming the 136-bp deletion in the resistant accessions F333 and TC1966. **e**, Gene model for *VrPGIP1*, indicating the location of the 136-bp promoter deletion and key exonic SNPs. **f**, Frequency of the main haplotypes in wild, landrace and cultivated accessions. **g**, Relative expression level of *VrPGIP1* in accessions with (Alt) and without (Ref) the promoter deletion. Statistical test was performed using a two-tailed Student's *t*-test. **h**, Expression profile of *VrPGIP1* during seed development; error bar: mean ± s.d.; $n = 3$ biological replicates. Significant difference was tested using a two-tailed Student's *t*-test. **i**, Transcription activity of the *VrPGIP1* promoters from a resistant (F333, with deletion) and susceptible (ZL19, without deletion) accession. Scale bar: 1 cm. Error bar: mean ± s.d.; $n = 3$ biological replicates. Significant difference was tested using a Student's *t*-test. **j**, Bruchid resistance assay comparing F333 and ZL19. Statistical test was performed using a two-tailed Student's *t*-test. **k**, In vitro assay showing the polygalacturonase-inhibiting activity of the VrPGIP1 protein. Error bars: mean ± s.d.; $n = 3$ biological replicates. **l**–**n**, Efficacy of purified VrPGIP1 protein in inhibiting bruchid development in artificial seeds at 7 days (**l**) and 50 days (egg number per seed (**m**); hole number per seed (**n**)) postinfestation. In **c** and **g**, center line: median; box boundaries: upper and lower quartiles; whiskers: 1.5× IQR. In **m** and **n**, error bars: mean ± s.d. Different letters indicate significant differences between examined groups (one-way analysis of variance followed by Tukey's multiple comparisons test, $P < 2 \times 10^{-16}$).

https://doi.org/10.1038/s41588-026-02644-5



(Fig. 7d,e). Hap0 was ubiquitous in wild accessions, whereas cultivated accessions carried Hap0 and Hap1. Landrace accessions and the G1 subgroup contained all three haplotypes, with Hap0 being the most prevalent (Fig. 7f). Similarly, the G3b subgroup contained Hap0 and Hap2, whereas all other subgroups comprised exclusively Hap0 (Supplementary Fig. 23a). Notably, the 136-bp deletion was specific to the G1 subgroup and absent in wild accessions (Fig. 7f and Supplementary Fig. 23b). Genome assemblies confirmed the presence of the deletion in highly resistant germplasms in F333 and TC1966 (Fig. 7d and Supplementary Fig. 24). Expression analysis indicated that this deletion modulated *VrPGIP1* expression (Fig. 7g), which peaked during the middle stage of seed development (Fig. 7h). To validate this regulatory effect, transient activation assays in leaves revealed that the promoter harboring the deletion drove higher *LUC* expression, correlating with increased bruchid resistance (Fig. 7i,j). Furthermore, functional assays confirmed that VrPGIP1 inhibits polygalacturonase activity (Fig. 7k), and resistance was enhanced in a dose- and time-dependent manner (Fig. 7l–n). Collectively, these results demonstrate that the 136-bp deletion in the *VrPGIP1* promoter enhances bruchid resistance in mung bean by upregulating its transcription.

Discussion

Mung bean is an important warm-season legume crop in Asia[19,39], but a single reference genome cannot capture its full genetic diversity. Previous pan-genomes constructed using a map-to-pan strategy represented SV loci as single haplotypes[18]. Graph-based pan-genomes provide a more comprehensive and accurate representation of genetic diversity by capturing several haplotypes and complex structural variations within a single framework[40,41]. Similar approaches in crops such as cucumber[24], *Brassica oleracea*[42] and foxtail millet[14] have driven the identification of key agronomic trait variation sites. This study provides a graph-based pan-genome for mung bean, establishing a new paradigm for genomics-assisted breeding in this legume, paralleling recent high-resolution genomic advances in closely related *Vigna* species such as adzuki bean[43]. Integrating resequencing data from 580 accessions, this resource enables the dissection of the genetic architecture of key agronomic traits and facilitates the identification of causal variants for precision breeding.

Population genetic variation analyses revealed the roles of natural selection, genetic drift and gene flow in shaping the species' evolutionary history[44]. Our analysis traced the spread of mung bean and identified G1 as the primary ancestral progenitor subpopulation, aligning with the historical migration routes of mung bean from South Asia to Southeast Asia, West Asia and East Asia[45]. The differences in plant development, yield and quality traits between wild and elite cultivars probably reflect selection for locally adapted and agronomically preferred traits. Selective sweep regions associated with mung bean domestication and improvement harbor potential candidate genes crucial for mung bean breeding.

The SV identification revealed 66,862 SVs, which is comparable to the previous mung bean pan-genome[18], but lower than the reported figures for several other crop pan-genomes, such as rice[9,46], soybean[10], tomato[30] and foxtail millet[14]. Crucially, the total number of identified SVs is an intrinsic evolutionary footprint of a species rather than an artifact of genome assembly contiguity. Although the remaining gaps in our current assemblies suggest room for improvement towards telomere-to-telomere (T2T) status, cross-validation with published T2T data confirms that these local gaps exert a negligible impact on global SV quantification. Furthermore, the comparatively lower SV abundance in mung bean is primarily driven by the limited genetic divergence and strong phylogenetic relatedness among the selected accessions—a biological phenomenon similarly observed in the broomcorn millet pan-genome[15], where low nucleotide diversity within the germplasm naturally correlates with reduced structural polymorphism. Consistent with findings in foxtail millet[14], TEs, particularly DNA transposons, contributed substantially to SV formation. SVs can influence gene expression by altering gene sequences, copy numbers, or *cis*-regulatory elements[30]. Our analysis demonstrated that SV-associated genes exhibit lower expression levels, similar to patterns observed in rice[9,46] and foxtail millet[14]. These findings indicate that SVs generally function as negative dosage regulators of gene expression. Mechanistically, SVs may repress expression by disrupting transcription factor binding sites in regulatory regions, triggering nonsense-mediated decay through intragenic frameshifts, or altering local chromatin accessibility. Collectively, these insights highlight the critical role of genomic structural integrity in maintaining proper transcriptional regulation.

A key advance of this study is linking SVs directly to agronomic traits through pan-genome GWAS, uncovering causal variants missed by SNP-based approaches. For FC, a promoter SV and coding SNPs in *VrTIFY6B* synergistically regulate the trait. For bruchid resistance, a promoter deletion in *VrPGIP1* upregulates defense, providing a ready-to-use breeding marker. For traits such as 100-grain weight, SNPs outperformed SVs due to their higher marker density and greater genotyping accuracy. SNPs introduce primarily nonsynonymous mutations in protein-coding genes, whereas SVs can profoundly disrupt gene structure and induce substantial changes in the expression of nearby genes. Therefore, integrating both SNPs and SVs is essential to fully capture the genetic variation. Although stable transformation in mung bean remains challenging, our transient and heterologous assays provide strong evidence for candidate gene function. Future work will focus on optimizing transformation protocols to achieve stable transgenic validation of these key genes.

In summary, this graph-based pan-genome provides a comprehensive understanding of the genomic diversity and structural variation in mung bean, offering insights into the evolutionary mechanisms shaping key agronomic traits. It serves as a valuable resource for accelerating genomics-assisted breeding and improvement programs in mung bean.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-026-02644-5.

References

1. Kang, Y. J. et al. Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat. Commun.* **5**, 5443 (2014).
2. Hou, D. Z. et al. Mung bean (*Vigna radiata* L.): bioactive polyphenols, polysaccharides, peptides, and health benefits. *Nutrients* **11**, 1238–1266 (2019).
3. Tao, M. F. et al. Flavonoids from the mung bean coat promote longevity and fitness in *Caenorhabditis elegans*. *Food Funct.* **12**, 8196–8207 (2021).
4. Rohilla, V. et al. Association mapping for yield attributing traits and yellow mosaic disease resistance in mung bean [*Vigna radiata* (L.) Wilczek]. *Front. Plant Sci.* **12**, 749439 (2021).
5. Chen, H. L. et al. Assessment of genetic diversity and population structure of mung bean (*Vigna radiata*) germplasm using EST-based and genomic SSR markers. *Gene* **566**, 175–183 (2015).
6. Gayacharan et al. in *The Mung Bean Genome* (eds Nair, R. M. et al.) Ch. 2 (Springer, 2020).
7. Kim, S. K., Nair, R. M., Lee, J. & Lee, S. H. Genomic resources in mungbean for future breeding programs. *Front. Plant Sci.* **6**, 626 (2015).
8. Yuan, Y., Bayer, P. E., Batley, J. & Edwards, D. Current status of structural variation studies in plants. *Plant Biotechnol. J.* **19**, 2153–2163 (2021).

9. Qin, P. et al. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* **184**, 3542–3558 (2021).
10. Liu, Y. et al. Pan-genome of wild and cultivated soybeans. *Cell* **182**, 162–176 (2020).
11. Walkowiak, S. et al. Multiple wheat genomes reveal global variation in modern breeding. *Nature* **588**, 277–283 (2020).
12. Varshney, R. K. et al. A chickpea genetic variation map based on the sequencing of 3,366 genomes. *Nature* **599**, 622–627 (2021).
13. Jayakodi, M. et al. The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature* **588**, 284–289 (2020).
14. He, Q. et al. A graph-based genome and pan-genome variation of the model plant *Setaria*. *Nat. Genet.* **55**, 1232–1242 (2023).
15. Chen, J. et al. Pangenome analysis reveals genomic variations associated with domestication traits in broomcorn millet. *Nat. Genet.* **55**, 2243–2254 (2023).
16. Lu, Q. et al. Genomic variation in weedy and cultivated broomcorn millet accessions uncovers the genetic architecture of agronomic traits. *Nat. Genet.* **56**, 1006–1017 (2024).
17. Lye, Z. N. & Purugganan, M. D. Copy number variation in domestication. *Trends Plant Sci.* **24**, 352–365 (2019).
18. Liu, C. et al. High-quality genome assembly and pan-genome studies facilitate genetic discovery in mung bean and its improvement. *Plant Commun.* **3**, 100352 (2022).
19. Han, X. et al. Resequencing of 558 Chinese mungbean landraces identifies genetic loci associated with key agronomic traits. *Front. Plant Sci.* **13**, 1043784 (2022).
20. Lu, X. et al. A PP2C-1 allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Mol. Plant* **10**, 670–684 (2017).
21. Suamuang, S. et al. Identification of quantitative trait loci controlling flowering time in black gram (*Vigna* mungo [L.] Hepper). *Agric. Nat. Resour.* **57**, 43–50 (2023).
22. He, Y. et al. Multiomics analysis reveals the molecular mechanisms underlying virulence in Rhizoctonia and jasmonic acid-mediated resistance in Tartary buckwheat (*Fagopyrum tataricum*). *Plant Cell* **35**, 2773–2798 (2023).
23. Huang, Z., Shen, F., Chen, Y., Cao, K. & Wang, L. Chromosome-scale genome assembly and population genomics provide insights into the adaptation, domestication, and flavonoid metabolism of Chinese plum. *Plant J.* **108**, 1174–1192 (2021).
24. Li, H. et al. Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. *Nat. Commun.* **13**, 682 (2022).
25. Tang, D. et al. Genome evolution and diversity of wild and cultivated potatoes. *Nature* **606**, 535–541 (2022).
26. Yang, T. et al. Improved pea reference genome and pan-genome highlight genomic features and evolutionary characteristics. *Nat. Genet.* **54**, 1553–1563 (2022).
27. Zhang, C. et al. The rice E3 ubiquitin ligase-transcription factor module targets two trypsin inhibitors to enhance broad-spectrum disease resistance. *Dev. Cell* **59**, 2017–2033 (2024).
28. Smit, P. et al. Medicago LYK3, an entry receptor in rhizobial nodulation factor signaling. *Plant Physiol.* **145**, 183–191 (2007).
29. Zhou, Y. et al. Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* **606**, 527–534 (2022).
30. Alonge, M. et al. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* **182**, 145–161 (2020).
31. Zhang, Z. et al. Genome-wide mapping of structural variations reveals a copy number variant that determines reproductive morphology in cucumber. *Plant Cell* **27**, 1595–1604 (2015).
32. Li, N. et al. Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat. Genet.* **55**, 852–860 (2023).
33. Pauwels, L. et al. NINJA connects the co-repressor TOPLESS to jasmonate signalling. *Nature* **464**, 788–791 (2010).
34. Yang, J. et al. The crosstalks between jasmonic acid and other plant hormone signaling highlight the involvement of jasmonic acid as a core component in plant response to biotic and abiotic stresses. *Front. Plant Sci.* **10**, 1349 (2019).
35. Zhang, K. et al. Jasmonate-responsive MYB factors spatially repress rutin biosynthesis in *Fagopyrum tataricum*. *J. Exp. Bot.* **69**, 1955–1966 (2018).
36. Ding, M. et al. FtBPM3 modulates the orchestration of FtMYB11-mediated flavonoids biosynthesis in Tartary buckwheat. *Plant Biotechnol. J.* **19**, 1285–1287 (2021).
37. Chung, H. & Howe, G. A critical role for the TIFY motif in repression of jasmonate signaling by a stabilized splice variant of the JASMONATE ZIM-domain protein JAZ10 in Arabidopsis. *Plant Cell* **21**, 131–145 (2009).
38. Zhou, M. et al. FtSAD2 and FtJAZ1 regulate activity of the FtMYB11 transcription repressor of the phenylpropanoid pathway in *Fagopyrum tataricum*. *New Phytol.* **216**, 814–828 (2017).
39. Graham, P. H. & Vance, C. P. Legumes: importance and constraints to greater use. *Plant Physiol.* **131**, 872–877 (2003).
40. Shi, J. et al. Plant pan-genomics and its applications. *Mol. Plant* **16**, 168–186 (2023).
41. Wang, S. et al. Graph-based pan-genomes: increased opportunities in plant genomics. *J. Exp. Bot.* **74**, 24–39 (2023).
42. Guo, N. et al. A graph-based pan-genome of *Brassica oleracea* provides new insights into its domestication and morphotype diversification. *Plant Commun.* **5**, 100791 (2024).
43. Hu, L. et al. High-resolution genomic resources for trait mapping and precision breeding for adzuki bean (*Vigna angularis*). *Adv. Sci.* **13**, e07157 (2026).
44. Sork, V. L. Gene flow and natural selection shape spatial patterns of genes in tree populations: implications for evolutionary processes and applications. *Evol. Appl.* **9**, 291–310 (2016).
45. Tomooka, N., Lairungreang, C., Nakeeraks, P., Egawa, Y. & Thavarasook, C. Center of genetic diversity and dissemination pathways in mung bean deduced from seed protein electrophoresis. *Theor. Appl. Genet.* **83**, 289–293 (1992).
46. Zhao, Q. et al. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* **50**, 278–284 (2018).
47. Wickham, H. *ggplot2*: *Elegant Graphics for Data Analysis* (Springer, 2016).

[1]Institute of Crop Sciences, Chinese Academy of Agricultural Sciences/China-Uruguay Joint Laboratory on Soybean Research and Innovation, Beijing, China. [2]College of Life Sciences, Institute of Life Science and Green Development, Hebei University, Baoding, China. [3]Institute of Economic Crops, Henan Academy of Agricultural Sciences, Zhengzhou, China. [4]Centre for Crop and Food Innovation, WA State Agricultural Biotechnology Centre, Food

Futures Institute, Murdoch University, Murdoch, Western Australia, Australia. [5]Division of Genetics and Plant Breeding, Faculty of Agriculture (FoA), SKUAST-Kashmir, Kashmir, India. [6]Crop Research Institute of Anhui Academy of Agricultural Sciences, Hefei, China. [7]Institute of Vegetable Research, Guangxi Academy of Agricultural Sciences, Nanning, China. [8]Zhangjiakou Academy of Agricultural Sciences, Zhangjiakou, China. [9]Baicheng Academy of Agricultural Sciences, Baicheng, China. [10]Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China. [11]State Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences, Weifang, China. [12]These authors contributed equally: Honglin Chen, Longsheng Xing, Chaonan Guan, Yang Liu. ✉e-mail: chenhonglin@caas.cn; chengzhi.jiao@pku-iaas.edu.cn; huilongdu@hbu.edu.cn; zhoumeiliang@caas.cn; Rajeev.Varshney@murdoch.edu.au; chengxuzhen@caas.cn

Methods

Plant materials and phenotyping

We selected 580 mung bean accessions from the Center for Crop Germplasm Resources (CCGR), Chinese Academy of Agricultural Sciences, comprising 426 landraces, 57 wild accessions and 97 cultivated accessions, representing 19 countries (Supplementary Table 1). These accessions encompass nearly all regions of mung bean cultivation, providing a comprehensive representation of the species' genetic diversity. Multi-environment trials were conducted from 2014 to 2022 at five geographically distinct field stations: Baicheng (45.44° N, 122.82° E), Beijing (40.23° N, 116.56° E), Zhangjiakou (39.92° N, 114.88° E), Anhui (31.87° N, 117.26° E) and Nanning (23.15° N, 108.28° E). Agronomic traits were recorded systematically for each accession at maturity; detailed measurement criteria and procedures are provided in Supplementary Methods.

Genome sequencing

Eleven mung bean accessions representing main racial groups and geographic regions were selected for de novo genome assembly (Supplementary Table 1). Ten accessions (C34, C509, C591, C95, F287, F333, F487, ZL19, TC1966 and W655) were sequenced using PacBio HiFi and one accession (ZL5) was sequenced using PacBio CLR. High molecular weight (HMW) DNA was extracted from fresh leaves of a single plant per accession using the NucleoBond HMW DNA Kit (Macherey-Nagel). The HiFi SMRTbell libraries were prepared using the SMRTbell Express Template Prep Kit 3.0 (Pacific Biosciences), size-selected (9–50 kb) and sequenced on the Sequel IIe platform (Novogene Co., Ltd.). SMRTlink software (PacBio) was used to process raw polymerase reads and to generate HiFi reads with the parameters '--min-passes=3--min-rq=0.99,' yielding 21.04–31.60 Gb per accession (Supplementary Table 2). For ZL5, CLR libraries were prepared following the standard protocol and sequenced on the PacBio Sequel II platform, producing 56.48 Gb of data (Supplementary Table 2).

For each accession, fresh leaves from a single plant were fixed with a 5% formaldehyde solution, quenched with glycine and used for genomic DNA extraction. Hi-C libraries were prepared using the NEBNext Ultra II DNA PCR-free Library Prep Kit (New England Biolabs) and sequenced on the NovaSeq 6000 platform (Novogene). Hi-C data yields ranged from 56.15 Gb to 73.37 Gb per accession (Supplementary Table 2).

For each accession, total RNA was extracted from five tissues (roots, stems, leaves, flowers and seeds), assessed for integrity using a Fragment Analyzer 5400 (Agilent Technologies), and used to construct libraries with the NEBNext Ultra RNA Library Prep Kit (New England Biolabs). These libraries were sequenced on the NovaSeq 6000 platform, yielding 5.85–8.30 Gb per accession (Supplementary Table 3). Paired-end whole-genome resequencing libraries were constructed and sequenced on the Illumina HiSeq X Ten platform (Novogene).

Genome assembly and quality assessment

Genome size and heterozygosity were estimated using Jellyfish (v.2.2.6)[48] and GenomeScope2 (ref. 49). PacBio HiFi accessions were assembled using Hifiasm (v.0.16.1)[50] with the Circular Consensus Sequencing reads under default parameters, and primary contigs were used for downstream analyses. The ZL5 accession was assembled using CANU (v.2.2)[51] and HERA[52]. Contigs were scaffolded into pseudochromosomes using 3D-DNA[53] and Juicer[54] based on Hi-C data, with manual correction using the Juicebox[55]. The ZL5 assembly was further polished using Pilon[56] with Illumina short reads. BUSCO (v.5.2.2)[57] was used to assess the assembly completeness using the embryophyta_odb10 (embryophyta, 2020-09-10) dataset, and LAI[58] was calculated using LTR_retriever (v.1.9)[59].

Repeat annotation

TEs were annotated using a combination of ab initio and homology-based methods. Tandem repeat sequences were detected using Tandem Repeat Finder (TRF v.4.07b)[60] and a de novo TE library was generated using RepeatModeler (v.1.0.11) (https://www.repeatmasker.org/RepeatModeler/) with default parameters. Intact LTRs were annotated using LTR_retriever by integrating LTRharvest[61] and LTR_FINDER[62], and combined with the de novo TE library for RepeatMasker (v.4.1.1)[63] annotation. For the homology-based search, repetitive sequences were identified using RepeatMasker against the Repbase library (v.15.02).

Gene annotation

Protein-coding genes were annotated by integrating ab initio, transcriptome-based, and homology-based approaches. AUGUSTUS (v.3.4.0)[64] was performed for ab initio prediction. RNA-seq data from five tissues were processed using Trimmomatic (v0.39)[65] with parameters 'LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36.' The clean reads were aligned to the reference genome using HISAT2[66] and StringTie2[67] was used for transcript assembly with default parameters. Candidate coding regions were identified using TransDecoder (v.5.5.0) (https://github.com/TransDecoder/TransDecoder). De novo transcript assembly was performed with Trinity (v.2.14.0)[68], and gene structures were detected using the Program to Assemble Spliced Alignments[69]. Protein sequences from eight plant species were included. *Arabidopsis thaliana* (GCF_000001735.4), *Medicago truncatula* (GCF_003473485.1), *Glycine max* (GCF_000004515.6), *Vigna angularis* (GCF_001190045.1), *Vigna radiata* (GCF_000741045.1), *Vigna umbellata* (GCF_018835915.1) and *Vigna unguiculata* (GCF_004118075.2) were downloaded from the NCBI RefSeq database; *Medicago sativa* (ZhongmuNo.1) was downloaded from the figshare database. The protein sequences were aligned to each genome using GenomeThreader (v.1.7.1)[70] to generate homology-based annotations. All lines of evidence were integrated using EVidenceModeler (v.1.1.1)[69] to generate a consensus geneset. To ensure high-confidence annotations, gene models overlapping TEs were removed, and only those supported by at least one line of evidence from manually curated homologous proteins, domain hits or RNA-seq data were retained.

Canonical noncoding genes were annotated using RNAmmer[71], Infernal (v.1.1.4) (Rfam database v.14.8)[72] and tRNAscan-SE (v.2.0.9)[73]. Protein-coding genes were annotated functionally by BLASTP search against the SwissProt database ($\leq 1 \times 10^{-5}$). Potential domains were annotated using InterProScan v.5.48-83.0 (ref. 74) with the parameters '-f TSV -t p -iprlookup -goterms -pa.' Gene ontology (GO) terms and Pfam annotations were derived from the InterProScan output. The Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology entries were annotated using KofamScan[75] with default parameters.

Phylogenetic analysis

To construct a phylogenetic tree for the 11 mung bean accessions, orthologous genes were identified using OrthoFinder (v.2.5.4)[76], including those from the closely related legume species *V. angularis*. Single-copy genes were aligned using MAFFT (v.7.490)[77] with default parameters and trimmed using trimAl (v.1.2rev59)[78] with the parameter '-gappyout.' The resulting amino acid alignments were concatenated into super sequences. The best-fit evolutionary model (JTT + I + G + F) was selected using ProtTest (v.3.4.2)[79] with the parameters '-S 1 -all-distributions -all -ncat 8 -F -AIC -BIC -tc 0.5.' A maximum-likelihood tree was constructed using RAxML (v.2.1.11)[80] with 1,000 bootstrap replicates, with *V. angularis* as the outgroup.

Gene-based pan-genome construction

All protein-coding genes from 12 mung bean accessions (11 assembled in this study plus JL7) were clustered using OrthoFinder (v.2.5.4)[76] with default parameters. Gene families were classified as core (present in all genomes), dispensable (2–11 genomes) or private (single genome). Pan- and core-genome sizes were estimated using 2,000 random samplings for each subset size (2–11 accessions). ZL19—a cultivar exhibiting

desirable agronomic traits such as high yield, upright growth and lodging resistance—was chosen as the reference genome. To estimate the K_a, K_s and the K_a/K_s ratio for different types of pan-gene, amino acid sequences were aligned using MUSCLE with default parameters and converted to coding sequence (CDS) alignments using PAL2NAL[81]. KaKs_Calculator (v.2.0)[82] was used to calculate K_a and K_s.

Identification and characterization of SVs

To identify SVs, we used 11 de novo genome assemblies generated in this study alongside the previously reported JL7 genome. We aligned the 11 mung bean genomes to the ZL19 genome using MUMmer (v.4.0.0beta2)[83] with the parameters '-mum -minmatch=50 -mincluster=100.' SVs were detected using SyRI (v.1.5.3)[84] with default parameters. SVs in syntenic regions were classified into four types, including insertions, deletions, inversions and translocations. Insertions and deletions were collectively referred to as PAVs. Only SVs ≥50 bp were retained for downstream analyses.

To identify shared SVs relative to ZL19 across mung bean accessions, we merged SVs of the same type using specific overlap criteria: deletion, inversion and translocation SVs were merged if they overlapped by >90%, and insertion SVs were merged if the overlap was >90% and they were within 5 bp of each other on the reference genome[7]. To identify SV hotspot regions, the density of SV breakpoints in nonoverlapping 1-Mb windows was quantified for each chromosome. All windows were then ranked in descending order of breakpoint density, and the top 5% were designated as SV hotspots. The consecutive hotspot windows were merged into contiguous hotspot regions. To assess the genomic distribution of SVs, we intersected SV coordinates with gene annotations using BEDTools[85]. SVs were classified into five categories based on overlap with (1) CDSs, (2) introns, (3) 5 kb upstream of the start codon, (4) 5 kb downstream of the stop codon and (5) intergenic regions. If an SV overlapped with two or more genomic regions, it was counted in all relevant categories. The proportion of SVs in each category was then calculated.

Variant calling of 580 mung bean accessions

The raw Illumina DNA reads were processed to remove adapter and low-quality sequences using Trimmomatic (v.0.39)[65] with the parameters 'TruSeq3-PE-2.fa:2:30:10:1:TRUE SLIDINGWINDOW:4:20 LEADING:3 TRAILING:3 MINLEN:40.' Clean reads were mapped to the ZL19 reference genome using BWA-MEM[86] with default parameters. SAMtools[87] was used to filter and sort the mapping results, and only the reads with a mapping quality of ≥30 were retained for downstream analysis. Duplicate reads in each alignment file were marked using the MarkDuplicates function in the GATK (v.4.2)[88]. SNPs and small indels were then called from the bam files using the HaplotypeCaller algorithm in GATK. Raw SNPs were then hard filtered using GATK (v.4.2) with parameters 'QD < 2.0 || FS > 60.0 || MQ < 40.0 || SOR > 3.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0,' and raw small indels were filtered with 'QD < 2.0 || FS > 200.0 || ReadPosRankSum < −20.0.' Only biallelic variants were retained for downstream analyses. Subsequently, VCFtools[89] was used for soft filtering, retaining only variants with a minor allele frequency ≥0.01 and a missing genotype rate ≤0.1. SnpEff (v.5.0e)[90] was used to annotate single nucleotide variants with default parameters.

Phylogeny and population structure analysis

The VCF file of the mung bean population was used as input to VCF2Dis (https://github.com/BGI-shenzhen/VCF2Dis) to generate a distance matrix for neighbor-joining tree construction using PHYLIP (v3.697). PCA was conducted in PLINK (v.1.90b6.26)[91] after LD-pruning ('--indep-pairwise 50 10 0.1'). PC1 separated G1 and G3a from G3b, and PC3 further distinguished G1 from G2 and G5. Population structure was inferred using ADMIXTURE (v.1.3.0; $K = 2$–7)[92], revealing subgroup differentiation, including G4 ($K = 4$) and G5 ($K = 5$). LD decay was estimated using PopLDdecay[93] with default parameters. Population genetic indices (F_{ST} and π) were calculated in nonoverlapping 10-kb windows using VCFtools. XP-CLR[94] was used for identifying selective sweeps between two subpopulations.

Construction of the graph-based mung bean pan-genome and SV genotyping

A graph-based mung bean genome was constructed by using SVs identified from the 11 assembled genomes. The vg-construct function within the vg toolkit (v.1.28.0)[95] was performed for integrating PAVs and inversions detected by SyRI into the ZL19 linear reference genome with default parameters, generating a vg-format pan-genome. The complex regions in the raw graph were eliminated using the vg-prune function, and vg-index was used for building indexes based on the pruned graph.

To identify SVs in the 580 accessions, filtered Illumina short reads were mapped to the graph-based reference genome and SVs were genotyped using the pipeline integrated in the vg toolkit (v.1.28.0). Briefly, the clean reads were mapped to the graph pan-genome using vg-map with default parameters. The resulting aligned GAM files were filtered with the vg-filter function to improve precision. Then, vg-pack and vg-snarls were performed sequentially with default parameters to generate a compressed coverage index and snarls. Further, SV genotypes were then called from the short reads using the vg-call function with default parameters. The resulting genotyping data were merged using BCFtools[96] for downstream analyses. Finally, the SV genotypes were filtered using VCFtools with the criteria: minor allele frequency >0.01 and missing rate <0.3.

SNP-based and SV-based GWAS

Genome-wide SNPs were used for GWAS analysis using GEMMA[97]. The kinship was used as a covariance factor in the association model. The linear mixed model was applied for the analysis. The effective number of SNPs was calculated using the Genetic Type I Error Calculator. The Bonferroni-corrected P-value thresholds (SNP, 1.0×10^{-6}; SV, 1.0×10^{-5}) were determined by a uniform threshold of $1/N$, where N is the effective number of SNPs or SVs calculated by the Genetic Type I Error Calculator. The candidate gene region was determined based on the LD between the peak SNP and its neighboring SNPs, with the LD cutoff set at half the maximum LD value for *V. radiata*. The SV-GWAS was performed using the same model.

Geographic map generation

The geographical coordinates of the collection sites for all varieties and phenotypes in this study were plotted on a map using the ggplot2 (ref. 47) package in R (v.4.1.0).

Gene expression analysis

To estimate gene expression levels in 11 mung bean accessions, clean RNA-seq reads from different tissues were aligned to the chromosome-level ZL19 assembly using HISAT2 (ref. 66). The fragments per kilobase of exon model per million mapped fragments (FPKM) values of genes were estimated using StringTie2 (ref. 67) with the parameter '-e.'

Functional enrichment analysis

GO and KEGG enrichment analyses were performed using the clusterProfiler[98] package. GO or KEGG terms with an adjusted P value < 0.05 were considered statistically significantly enriched.

Y2H assay

The CDSs of *VrTIFY6B*^{Hap2}, *VrTIFY6B*^{Hap1}, *VrPPD1* and *FtPPD* were cloned individually into the pGADT7 vector, and VrNINJA2, *VrKIX8*^{Hap0} and *VrKIX8*^{Hap1} were inserted into the pGBKT7 vector via homologous recombination. Pairs of these constructs (or empty vectors) were cotransformed into yeast strain Y2Hgold using a transformation kit

(SK2400-200, Coolaber). Transformants were cultured initially on SD medium without Leu or Trp at 28 °C for 3–4 days. Subsequently, yeast cells were serially diluted tenfold and plated onto SD medium without Leu, Trp and His or SD medium without Leu, Trp, His and Ade, then incubated at 30 °C until colony emergence. All oligonucleotide primers used in vector construction are listed in Supplementary Table 36.

Dual-luciferase assay

For dual-luciferase assays, promoters were cloned into pGreenII 0800-LUC vector. *Agrobacterium tumefaciens* strain GV3101 containing these constructs was cultured overnight at 28 °C. Cultures were then diluted to an optical density at 600 nm of 0.6 in resuspension buffer containing 10 mM $MgCl_2$, 10 mM MES and 100 mM acetosyringone. *Nicotiana benthamiana* leaves were injected with *A. tumefaciens* carrying the construct. The injected leaves were incubated in the dark for 24 h and then exposed to a 48 h of light–dark cycle, after which the injected leaves were detached and sprayed with 1 mM D-luciferin sodium salt and 0.01% (v/v) Triton X-100 in ddH_2O. The luminescence was captured using an LB983 Nightowl II system. Primer sequences are provided in Supplementary Table 36.

Firefly LCI assays

For LCI assays, the full length *VrTIFY6B*^{Hap1} and *VrTIFY6B*^{Hap2} and *VrNINJA2* sequences were amplified with specific primers and cloned into pCAMBIA1300-cLUC and pCAMBIA1300-nLUC vectors. The constructs were transformed into GV3101 and subsequently coinfiltrated into 4-week-old *N. benthamiana* leaves. Luminescence signals were captured using a live imaging system. The primers used in the assay are listed in Supplementary Table 36.

Subcellular localization

The CDS of *VrKIX8* was cloned and inserted into the pCAMBIA1300-GFP vector. This recombinant plasmid and a nuclear-localized marker (p2300-H2B-mCherry) were transformed into *A. tumefaciens* strain GV3101. Co-infiltration was performed on *N. benthamiana* leaves. After incubation, fluorescence was detected using a Zeiss LSM900 laser scanning confocal microscope. Primer sequences are provided in Supplementary Table 36.

Quantitative PCR with reverse transcription analysis

Total RNA was extracted using the RNAEasy Plant Tissue Kit (DP452, Tiangen). First-strand cDNA was synthesized with HiScript III RT SuperMix (+gDNA wiper) (R323, v.21.1, Vazyme) following the manufacturer's instructions. Quantitative PCR with reverse transcription was conducted using Taq Pro Universal SYBR qPCR Master Mix (Q712, v.20.1, Vazyme), adhering to the manufacturer's guidelines. Primer sequences are provided in Supplementary Table 36.

Enzyme assays

The *VrPGIP1* CDS was amplified and inserted into the pGEX-4T2 fusion expression vector, which was then transformed into *Escherichia coli* BL21 (DE3) cells. Protein expression was induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside at 18 °C overnight. The fusion proteins were purified by immobilization onto Pierce GST agarose followed by elution with reduced L-glutathione. The polygalacturonase-inhibiting activity of *VrPGIP1* was assessed by quantifying the reduction of reducing sugars released from polygalacturonic acid using a colorimetric assay[99]. Primer sequences are listed in Supplementary Table 36.

Nodulation phenotyping

Mung bean seeds were surface-sterilized with 75% ethanol and 2.5% sodium hypochlorite, then germinated in darkness at 28 °C for 3 days. Seedlings were grown in pots of sterilized vermiculite in a growth chamber. *Bradyrhizobium diazoefficiens* USDA110 was cultured in tryptone yeast medium to an optical density at 600 nm of 0.8. The cells were collected by centrifugation, washed and resuspended in fresh tryptone yeast medium. Two-week-old seedlings were inoculated with 10 ml of the bacterial suspension. Plants were watered subsequently with nitrogen-free Hoagland solution. At 20 days postinoculation, root nodules were counted and photographed, and root fresh and dry weights were measured. The experiment consisted of three biological replicates.

Generation of transgenic hairy roots and *Arabidopsis*

The CDS sequences of *VrTIFY6B*^{Hap1} and *VrTIFY6B*^{Hap2} were amplified by PCR and inserted into the pCAMBIA1302 vector. The *VrTIFY6B*^{Hap1} or *VrTIFY6B*^{Hap2} recombinant plasmids were transformed into *Agrobacterium rhizogenes* strain A4 to generate transgenic hairy roots in buckwheat and mung bean, by infecting hypocotyl segments followed by cocultivation and antibiotic selection to induce root emergence[100]. The *VrKIX8* and *Vr04g06481* CDS sequences were cloned and subsequently inserted into the pCAMBIA1307 vector. The recombinant plasmids were also introduced into *A. tumefaciens* strain GV3101 to produce transgenic *Arabidopsis* using the floral dip method. Each experiment was conducted with three biological replicates. Hairy roots were harvested 2 weeks postinoculation and dried for FC analysis. Primer sequences are provided in Supplementary Table 36.

Bruchid resistance analysis of *VrPGIP1*

The bruchid resistance assay for *VrPGIP1* against *Callosobruchus maculatus* was performed by infesting customized artificial seeds with newly emerged adult bruchids and monitoring insect development[99]. The seed coats of ZL19 and F333 (a high bruchid resistance control) accessions were mechanically removed by decortication and subsequently ground into fine powders. The resulting ZL19 powder served as the matrix for constructing artificial seeds containing varying concentrations (0%, 0.01%, 0.05%, 0.1% and 0.2%) of VrPGIP1. F333 served as the control without VrPGIP1 protein. The number of eggs and emergence holes was recorded at 7 and 50 days after infestation.

Statistics and reproducibility

All computational analyses and statistical tests were executed within the R environment (v.4.1.0). We used two-tailed Student's *t*-tests and Wilcoxon rank-sum tests to assess group differences, with the specific methodology for each experiment detailed in the corresponding figure legends. Crucial statistical parameters, including *P* values, sample sizes (*n*) and numbers of replicates, are documented explicitly throughout the paper to ensure study transparency and reproducibility.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All long-read sequencing, Hi-C sequencing (CRA018804) and transcriptome sequencing data (CRA018850) of *Vigna radiata* genomes generated in this study have been deposited in the Genome Sequence Archive (GSA) at the National Genomics Data Center (NGDC) (https://ngdc.cncb.ac.cn), China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences under the BioProject accession (PRJCA029414). The 580 whole-genome resequencing data (CRA018971) generated in this study have been deposited in the GSA at NGDC under the BioProject accession (PRJCA029425). The whole-genome sequence data reported in this paper have been deposited in the Genome Warehouse (GWH) at the NGDC, Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation, under accession numbers that are publicly accessible at https://ngdc.cncb.ac.cn/gwh/ (GWH: GWHFDOM00000000.1, GWHFDON00000000.1, GWHFDOO00000000.1, GWHFDOP00000000.1,

GWHFDOQ00000000.1, GWHFDOR00000000.1, GWHFDOS00000000.1, GWHFDOT00000000.1, GWHFDOU00000000.1, GWHFDOV00000000.1, GWHFDOW00000000.1). All long-read sequencing data, Hi-C sequencing and transcriptome sequencing data have been deposited in the National Center for Biotechnology Information (NCBI) database under the BioProject accession PRJNA1160184. The 580 NGS resequencing data have been deposited in the NCBI database under the BioProject accession PRJNA1161446. Source data are provided with this paper.

Code availability

All custom code and scripts used in this study are publicly available on GitHub (https://github.com/jackiexls/vigna-pan-genome) and are available via Zenodo at https://doi.org/10.5281/zenodo.19186411 (ref. 101).

References

48. Marcais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770 (2011).
49. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).
50. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
51. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
52. Du, H. & Liang, C. Assembly of chromosome-scale contigs by efficiently resolving repetitive sequences with long reads. *Nat. Commun.* **10**, 5360 (2019).
53. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
54. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
55. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
56. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
57. Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
58. Ou, S., Chen, J. & Jiang, N. Assessing genome assembly quality using the LTR assembly index (LAI). *Nucleic Acids Res.* **46**, e126 (2018).
59. Ou, S. & Jiang, N. LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2018).
60. Benson, G. Tandem repeats finder-a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
61. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinform.* **9**, 18 (2008).
62. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, 265–268 (2007).
63. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **4**, 11–14 (2009).
64. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, 435–439 (2006).
65. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
66. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
67. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
68. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
69. Haas, B. J. et al. Automated eukaryotic gene structure annotation using evidencemodeler and the program to assemble spliced alignments. *Genome Biol.* **9**, 7 (2008).
70. Gremme, G., Steinbiss, S. & Kurtz, S. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **10**, 645–656 (2013).
71. Lagesen, K. et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108 (2007).
72. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
73. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **5**, 955–964 (1997).
74. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
75. Aramaki, T. et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
76. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
77. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
78. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
79. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011).
80. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
81. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, 609–612 (2006).
82. Wang, D., Zhang, Y., Zhang, Z., Zhu, J. & Yu, J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom. Bioinform.* **8**, 77–80 (2010).
83. Marcais, G. et al. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944 (2018).
84. Goel, M., Sun, H., Jiao, W. B. & Schneeberger, K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 277 (2019).
85. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
86. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
87. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

88. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
89. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
90. Cingolani, P. et al. 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 (2012).
91. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
92. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
93. Zhang, C., Dong, S. S., Xu, J. Y., He, W. M. & Yang, T. L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786–1788 (2019).
94. Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective sweeps. *Genome Res.* **20**, 393–402 (2010).
95. Hickey, G. et al. Genotyping structural variants in pangenome graphs using the vg toolkit. *Genome Biol.* **21**, 35 (2020).
96. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
97. Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* **11**, 407–409 (2014).
98. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
99. Zhang, Q. et al. Two polygalacturonase-inhibiting proteins (VrPGIP) of *Vigna radiata* confer resistance to bruchids (*Callosobruchus* spp.). *J. Plant Physiol.* **258**, 153376 (2021).
100. Zhang, K. et al. Resequencing of global Tartary buckwheat accessions reveals multiple domestication events and key loci associated with agronomic traits. *Genome Biol.* **22**, 23 (2021).
101. Xing, L. Scripts and codes used in the pan-genome of *Vigna radiata*. *Zenodo* https://doi.org/10.5281/zenodo.19186411 (2026).

Acknowledgements

We thank L. Mao and Q. He (Institute of Crop Sciences, Chinese Academy of Agricultural Sciences) for their insightful discussions.

Author contributions

H.C. and X. Cheng conceived and designed the research. H.C., L.H., X. Chen, S.W. and L.W. participated in material preparation. L.X. and H.D. contributed to the genome assembly and annotation. H.C., L.X. and C.G. performed genomic variant calling, selective signature identification and genome-wide association studies. L.X., C.G. and R.B. performed population genetics analysis. Y.L., T.C., R.C. and Q.S. performed gene expression, functional enrichment and phenotypic data cleaning. Y.L., Q.S. and T.C. contributed to the functional characterization. L.H., B.Z., G.L., D.X., F.Y. and X.Y. planted the materials and collected phenotypic data at different geographic locations. H.C., C.G., Y.L., L.X. and M.Z. oversaw the integration and conceptualization of results and wrote the paper. H.C., C.G., R.B., R.R.M., C.J., H.D., M.Z. and R.K.V. revised the paper. All authors read, edited and approved the final paper.

Funding

This work was supported by the National Natural Science Foundation of China (grant numbers 32572308, 32241041 and 31401442 to H.C.), the National Key Research and Development Program of China (grant numbers 2018YFD1000703 and 2021YFD1200105 to H.C.), the China Agriculture Research System (CARS-08 to H.C.) and the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences. R.K.V. thanks the Food Futures Institute, Murdoch University and the Grains Research and Development Corporation (project numbers UOQ2403-012RTX and UMU2403-009RTX) for supporting this work in part.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-026-02644-5.

Correspondence and requests for materials should be addressed to Honglin Chen, Chengzhi Jiao, Huilong Du, Meiliang Zhou, Rajeev K. Varshney or Xuzhen Cheng.

Peer review information *Nature Genetics* thanks Hai-Chun Jing and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Rajeev K Varshney

Last updated by author(s): Apr 9, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	None
Data analysis	Jellyfish (v2.2.6), Hifiasm (v0.16.1), CANU (v2.2). BUSCO (v5.2.2), LTR retriever (v1.9). RepeatModeler (v1.0.11). RepeatMasker (v4.1.1), AUGUSTUS (v3.4.0), Trimmomatic (v0.39), TransDecoder (v5.5.0), Trinity (v2.14.0), GenomeThreader (v1.7.1) , EVidenceModeler (v1.1.1) , Infernal (v1.1.4), tRNAscan-SE (v2.0.9), OrthoEinder (v2.5.4), MAEET (v7.490), trimAl (v1.2rev59), ProtTest (v3.4.2), RAxML (v2.1.11), KaKs_Calculator (v2.0), Mummer (v4.0.0beta2), SyRI (v1.5.3), GATK (v4.2), SnpEff (v5.0e), phylip (v3.697), PLINK (v1.90b6.26), ADMIXTURE (v1.3.0), Variation graph (vg) toolkit (v1.28.0), vg toolkit (v1.28.0), XP-CLR (v2.0.0), R (v4.1.0)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All long-read sequencing, Hi-C sequencing (CRA018804), and transcriptome sequencing data (CRA018850) of Vigna radiata genomes generated in this study have been deposited in the Genome Sequence Archive (GSA) at the National Genomics Data Center (NGDC) (https://ngdc.cncb.ac.cn), China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences under the BioProject (PRJCA029414). 580 whole-genome resequencing data (CRA018971) generated in this study have been deposited in the GSA at NGDC under the BioProject (PRJCA029425). The whole genome sequence data reported in this paper have been deposited in the Genome Warehouse (GWH) in NGDC, Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation, under accession numbers that are publicly accessible at https://ngdc.cncb.ac.cn/gwh/ (GWH: GWHFDOM00000000.1, GWHFDON00000000.1, GWHFDOO00000000.1, GWHFDOP00000000.1, GWHFDOQ00000000.1, GWHFDOR00000000.1, GWHFDOS00000000.1, GWHFDOT00000000.1, GWHFDOU00000000.1, GWHFDOV00000000.1, GWHFDOW00000000.1). Additionally, all long-read sequencing data, Hi-C sequencing, and transcriptome sequencing data have been deposited in the National Center for Biotechnology Information database under the BioProject accession PRJNA1160184. 580 NGS resequencing data have been deposited in the NCBI database under the BioProject accession PRJNA1161446.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.* *Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.* *Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size for each experiment is indicated in the Figures and Figure legends. No statistical methods were used to predetermine sample size.
Data exclusions	No data have been excluded from the analyses presented in this study.
Replication	All the experiments were repeated at least 3 biological times, and each biological experiment contained 3 technical replicates, showing the

Replication	similar results.
Randomization	Randomization and covariates were not relevant to our study design as we investigated single factors within each study.
Blinding	For microscopy images and phenotype images, the investigators were blinded to group allocation during data collection and analysis. The relevant controls were processed at the same time as the samples and read in an unbiased manner.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☐	☐	Public health
☐	☐	National security
☐	☐	Crops and/or livestock
☐	☐	Ecosystems
☐	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☐	☐	Demonstrate how to render a vaccine ineffective
☐	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☐	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☐	☐	Increase transmissibility of a pathogen
☐	☐	Alter the host range of a pathogen
☐	☐	Enable evasion of diagnostic/detection modalities
☐	☐	Enable the weaponization of a biological agent or toxin
☐	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	All the mung bean germplasm resources used in this study are sourced from the Nature Crop Germplasm Bank of China.
Novel plant genotypes	In this study, all transgenic hairy roots were obtained by infecting plants with Agrobacterium rhizogenes A4.
Authentication	In this study, all transgenic plants have been detected and identified through PCR.