Plant Communications
Research Article

CellPress
Partner Journal

High-quality genome assembly and pan-genome studies facilitate genetic discovery in mung bean and its improvement

Changyou Liu[1,6], Yan Wang[1,6], Jianxiang Peng[2,6], Baojie Fan[1], Dongxu Xu[3], Jing Wu[4], Zhimin Cao[1], Yunqing Gao[3], Xueqing Wang[1], Shutong Li[3], Qiuzhu Su[1], Zhixiao Zhang[1], Shen Wang[1], Xingbo Wu[5], Qibing Shang[3], Huiying Shi[1], Yingchao Shen[1], Bingbing Wang[2,*] and Jing Tian[1,*]

[1]Institute of Cereal and Oil Crops, Hebei Academy of Agricultural and Forestry Sciences/Hebei Laboratory of Crop Genetics and Breeding, Shijiazhuang 050035, China

[2]Biobin Data Sciences, Changsha 410221, China

[3]Zhangjiakou Academy of Agricultural Sciences, Zhangjiakou 075300, China

[4]Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[5]Tropical Research and Education Center, Department of Environmental Horticulture, University of Florida, 18905 SW 280th St, Homestead, FL 33031, USA

[6]These authors contributed equally to this article.

***Correspondence: Bingbing Wang (bwang@biobin.com.cn), Jing Tian (nkytianjing@163.com)**

https://doi.org/10.1016/j.xplc.2022.100352

ABSTRACT

Mung bean is an economically important legume crop species that is used as a food, consumed as a vegetable, and used as an ingredient and even as a medicine. To explore the genomic diversity of mung bean, we assembled a high-quality reference genome (Vrad_JL7) that was ~479.35 Mb in size, with a contig N50 length of 10.34 Mb. A total of 40,125 protein-coding genes were annotated, representing ~96.9% of the genetic region. We also sequenced 217 accessions, mainly landraces and cultivars from China, and identified 2,229,343 high-quality single-nucleotide polymorphisms (SNPs). Population structure revealed that the Chinese accessions diverged into two groups and were distinct from non-Chinese lines. Genetic diversity analysis based on genomic data from 750 accessions in 23 countries supported the hypothesis that mung bean was first domesticated in south Asia and introduced to east Asia probably through the Silk Road. We constructed the first pan-genome of mung bean germplasm and assembled 287.73 Mb of non-reference sequences. Among the genes, 83.1% were core genes and 16.9% were variable. Presence/absence variation (PAV) events of nine genes involved in the regulation of the photoperiodic flowering pathway were identified as being under selection during the adaptation process to promote early flowering in the spring. Genome-wide association studies (GWASs) revealed 2,912 SNPs and 259 gene PAV events associated with 33 agronomic traits, including a SNP in the coding region of the SWEET10 homolog (jg24043) involved in crude starch content and a PAV event in a large fragment containing 11 genes for color-related traits. This high-quality reference genome and pan-genome will provide insights into mung bean breeding.

Key words: mung bean, long-read sequencing, *de novo* assembly, pan-genome, gene PAV, GWAS

Liu C., Wang Y., Peng J., Fan B., Xu D., Wu J., Cao Z., Gao Y., Wang X., Li S., Su Q., Zhang Z., Wang S., Wu X., Shang Q., Shi H., Shen Y., Wang B., and Tian J. (2022). High-quality genome assembly and pan-genome studies facilitate genetic discovery in mung bean and its improvement. Plant Comm. **3**, 100352.

INTRODUCTION

Mung bean (*Vigna radiata* L.) is a self-pollinated, fast-growing diploid legume crop species (2n = 2x = 22). As an inexpensive source of carbohydrates, protein (~27% of the dry seed content), folic acid, and iron, mung bean is economically important and widely cultivated, mainly in Asia (Kang et al., 2014). Mostly used as a food (grains), a vegetable (sprouts), and an ingredient

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

(paste), mung bean is also a well-known medicine in China for reducing body heat because of its rich content of flavonoids, vitexin (VITE), and isovitexin (ISOVITE) (Cao et al., 2011). Although high-quality genomic resources are available for many legume crop species, such as soybean (Shen et al., 2019), cowpea (Lonardi et al., 2019), and lima bean (Garcia et al., 2021), and long-read sequencing has enabled high-quality, chromosome-scale assemblies for many plant species (Michael and VanBuren, 2020), there is only a draft genome sequence available for mung bean. This sequence was assembled from Illumina short-read sequences, covering 80% of the estimated genome size, with 239 of 2748 scaffolds organized into 11 pseudochromosomes (Kang et al., 2014).

Recently, several studies on the genetic diversity of mung bean and genome-wide association studies (GWASs) of mung bean germplasm have been performed (Schafleitner et al., 2015; Noble et al., 2017; Breria et al., 2020; Reddy et al., 2020; Sokolkova et al., 2020). A common limitation of these studies was that they were either based on genotyping by sequencing (GBS) or diversity array technology, which revealed only a few thousand markers and was thus insufficient for large-scale gene mining. Whole-genome resequencing of mung bean, as has been applied to soybean (Zhou et al., 2015; Fang et al., 2017), pigeon pea (Varshney et al., 2017), chickpea (Varshney et al., 2019b), and common bean (Wu et al., 2020a), is greatly needed to better understand the genetic variation, nucleotide diversity, population structure, and key genes that govern important agronomic traits.

Owing to variation among different individuals, a single reference genome cannot contain all possible genetic information (Golicz et al., 2016; Hurgobin et al., 2018). The concept of the pan-genome, usually constructed by sequencing dozens to hundreds of individuals, has been proposed to represent the complete genome information of a species (Tettelin, 2005; Golicz et al., 2016). Pan-genomes have been assembled for many plant species, including rice (Qin et al., 2021), rapeseed (Song et al., 2020), soybean (Liu et al., 2020; Torkamaneh et al., 2021), chickpea (Varshney et al., 2021), and pepper (Ou et al., 2018a). Two pan-genome studies have been published for soybean: a graphic pan-genome constructed by *de novo* assembly of 26 representative wild and cultivated accessions using long-read sequencing (Liu et al., 2020) and a pan-genome comprising 204 cultivated soybeans (PanSoy) constructed using next-generation sequencing short reads (Torkamaneh et al., 2021). PanSoy explores the extent of genetic variation in cultivated soybean. The pigeon pea pan-genome was assembled as the first pan-genome of orphan legumes (Zhao et al., 2020). No pan-genome for mung bean has been constructed to date, hindering the progress of genetic discovery.

In this study, we assembled a high-quality reference genome and pan-genome of Chinese mung bean germplasm via deep sequencing of a high-yield variety and 217 accessions. Important agronomic traits, such as yield components, grain composition, morphology, and insect resistance, were measured for these accessions and associated with their genome sequences. Significant single-nucleotide polymorphisms (SNPs) and candidate genes were identified for almost all of the studied traits. These results lay a solid foundation for genomic breeding of mung bean.

RESULTS

Genome assembly and annotation

A high-yielding and early-maturing mung bean variety widely cultivated in China (Jilv 7 [JL7]) was sequenced using the PacBio Sequel II platform (long reads, ~52.83 Gb, 110.21× genome coverage) and Illumina paired-end (PE) read technology (short reads, ~61.28 Gb, 127.85× genome coverage). The estimated genome size was ~479.35 Mb, and the heterozygosity rate was ~0.056%, as calculated by the K-mer (k = 31) method (Supplemental Figure 1); it was ~11.7% smaller than the estimated size of the VC1973A genome (~543 Mb) (Kang et al., 2014). The *de novo*-assembled genome (Vrad_JL7) had a size of 475.19 Mb, which was ~99.13% of the estimated size, with an N50 of 10.34 Mb, and the largest contig was 30.20 Mb (Table 1). Approximately 98.72% (~469.11 Mb) of the genome was anchored onto 11 pseudomolecules according to Hi-C sequencing data (~33.5 Gb) (Figure 1A and Supplemental Figure 2). There were 259 gaps (~0.11 Mb) and 518 contigs (~6.08 Mb) that remained unanchored in the genome (Table 1). Pseudomolecules were named according to those of the VC1973A genome (Kang et al., 2014). Approximately 53.45% of the mung bean genome was composed of repetitive elements (Table 1). Long terminal repeat (LTR) retrotransposons accounted for 33.05% of the genome, and DNA transposons accounted for 4.25%. The LTR/Gypsy and LTR/Copia elements constituted 16.65% and 13.77%, respectively (Supplemental Table 1), which differed from the 25.2% and 11.3% in the VC1973A genome (Kang et al., 2014).

After masking the repetitive regions of the genome, a total of 40 125 protein-coding genes (17 pseudogenes were excluded) and 42 986 transcripts were annotated by combining *ab initio* gene prediction, RNA sequencing (RNA-seq), and protein homology evidence using the BRAKER2 pipeline (Bruna et al., 2021). Among them, 29 114 (72.56%) genes had either RNA-seq or homology evidence (Supplemental Data 1). Possible functions for protein-coding genes were annotated, and 81.60% (32 754) could be assigned functions via the Kyoto Encyclopedia of Genes and Genomes (KEGG) (37.45%), Gene Ontology (GO) (39.83%), Pfam (58.41%), SwissProt (60.26%), or NCBI nonredundant (NR) protein (81.38%) databases. For RNA-encoding genes, 5830 noncoding RNA genes of various types were also predicted using Barrnap and Infernal (version 1.1.4) (Nawrocki and Eddy, 2013) and by searching the Rfam database (Supplemental Table 2).

The overall quality of the Vrad_JL7 assembly was very high. The completeness was >98%, as revealed by Illumina PE read mapping and Benchmarking Universal Single-Copy Orthologs (BUSCO) assessment using eudicots_odb10 (Table 1; Supplemental text; Supplemental Figure 4). Many evaluation scores were better for the Vrad_JL7 assembly than for the VC1973A assembly, including a nearly four-fold greater contig N50 length, fewer gaps, and an LTR assembly index (LAI) of 15.67, demonstrating the reference quality of the assembly (Table 1). The BUSCO completeness for the predicted protein-coding genes was 96.9%, a 20% increase compared with that of the VC1973A assembly (Table 1; Supplemental Figure 4). These results demonstrated that the Vrad_JL7 genome assembly

Genomic feature	Vrad_JL7	VC1973A v1	VC1973A v2
Total assembly size, Mb	475.35	430.88	475.7
No. contigs	632	25 922	1511
Largest contig	30.20 Mb	734.56 kb	12.73 Mb
Contig N50	10.34 Mb	48.83 kb	2.8 Mb
Scaffold N50, Mb	43.79	1.52	47.1
Percentage anchored to chromosomes	98.72	73.09	89.92
No. gaps	259	96 874	1047
Length of gaps, Mb	0.11	33.56	1.81
GC content, %	33.45	33.16	33.27
Complete BUSCOs (genome), %	98.02	96.82	91.36
Complete BUSCOs (protein), %	96.90	81.60	80.01
LAI	15.67	7.86	14.65
Intact LTR-RTs	2725	734	2458
Repetitive sequences, %	53.45	50.10	52.79
Protein-coding genes	40 125	22 427	30 958

Table 1. Summary of assembly and annotations of the Vrad_JL7 and VC1973A genomes

and its annotations were more complete and of higher quality than those of the two versions of VC1973A (Kang et al., 2014; Ha et al., 2021), and it was by far the best quality mung bean genome.

Comparative genomic and evolutionary analysis

To identify evolutionary features of the mung bean genome, the sequences of annotated genes from 12 eudicot plant species were compared with those of the Vrad_JL7 gene set. A total of 32,253 orthogroups/gene families were formed, including 432 single-copy groups. A total of 35 059 (87.4%) mung bean coding genes clustered into gene families, 1532 of which (including 5482 genes) were specific to mung bean (Figure 1B; Supplemental Table 3). The functions of these specific genes were enriched in starch and sucrose metabolism; biosynthesis of amino acids; ribosome biogenesis in eukaryotes; glycine, serine, and threonine metabolism; and more (Figure 1C). These results are consistent with the rich starch and protein content of mung bean grains. Further analysis revealed that 2218 gene families had undergone expansion events and 1093 had undergone contraction events in the mung bean genome (Supplemental Figure 5). The functions of these expanded families were significantly enriched (adjusted $P < 0.05$) in various biological processes and pathways related to the characteristic features of mung bean or enabling adaptation to the environment, including plant–pathogen interactions, biosynthesis of isoflavonoids and terpenoids, and metabolism of unsaturated fatty acids (Supplemental Figure 6; Supplemental Data 2 and 3). Tandem duplication contributed significantly to gene family expansion (Supplemental text). Functions of the contracted gene families were mainly enriched in amino acid and starch metabolism. Interestingly, some contracted family members also had functions in the metabolism of flavonoids or fatty acid pathways (Supplemental Figure 8). Given that there are many different steps in biological processes, the unique, expanded, and contracted gene families likely fine-tuned the path selection in different steps, leading to specific features such as the rich content of certain types of flavonoids (including VITE and ISO-VITE) in mung bean.

Large-scale genome comparison revealed a perfect one-to-one collinear relationship between the chromosomes of mung bean (Vrad_JL7) and adzuki bean (*Vang*) (Figure 1D), including Vrad_JL7 chromosome 3 corresponding to *Vang* chromosome 5. Chromosome 3 of VC1973A v1, however, was probably incorrectly assembled and should be part of chromosome 4 instead (Supplemental Figure 9A). Most chromosomes of cowpea (*Vung*) had good collinearity with mung bean and adzuki bean, with the exception of chromosome 5 of cowpea, which was split into two chromosomes in mung bean and adzuki bean (Figure 1D), suggesting that the event occurred after the divergence of the ancestor of cowpea and the common ancestor of mung bean and adzuki bean.

Population genomic analysis

Genetic diversity in mung bean was evaluated in depth by resequencing 217 accessions: 24 Chinese breeding lines (CBLs), 165 Chinese landraces (CLRs), and 28 non-Chinese lines (NCLs). The average sequencing depth was 12.28× and ranged from 7.06× to 15.31× (Supplemental Table 4). A total of 2 229 343 SNPs, 230 025 short insertions and deletions (indels) (<15 bp), and 56 545 structural variations (SVs) (39 228 large indels [>15 bp], 1991 copy-number variations, 13 937 translocations, and 1389 inversion events) were identified. Their distributions in the genome are summarized in Supplemental Table 5 and Supplemental Table 6 and the Supplemental text. Overall, the different types of variants had similar density distribution patterns in the genome (Figure 2A).

An unrooted tree comprising 207 mung bean accessions (after removing 10 accessions with ambiguous sources) was constructed based on the core SNPs, dividing these accessions into

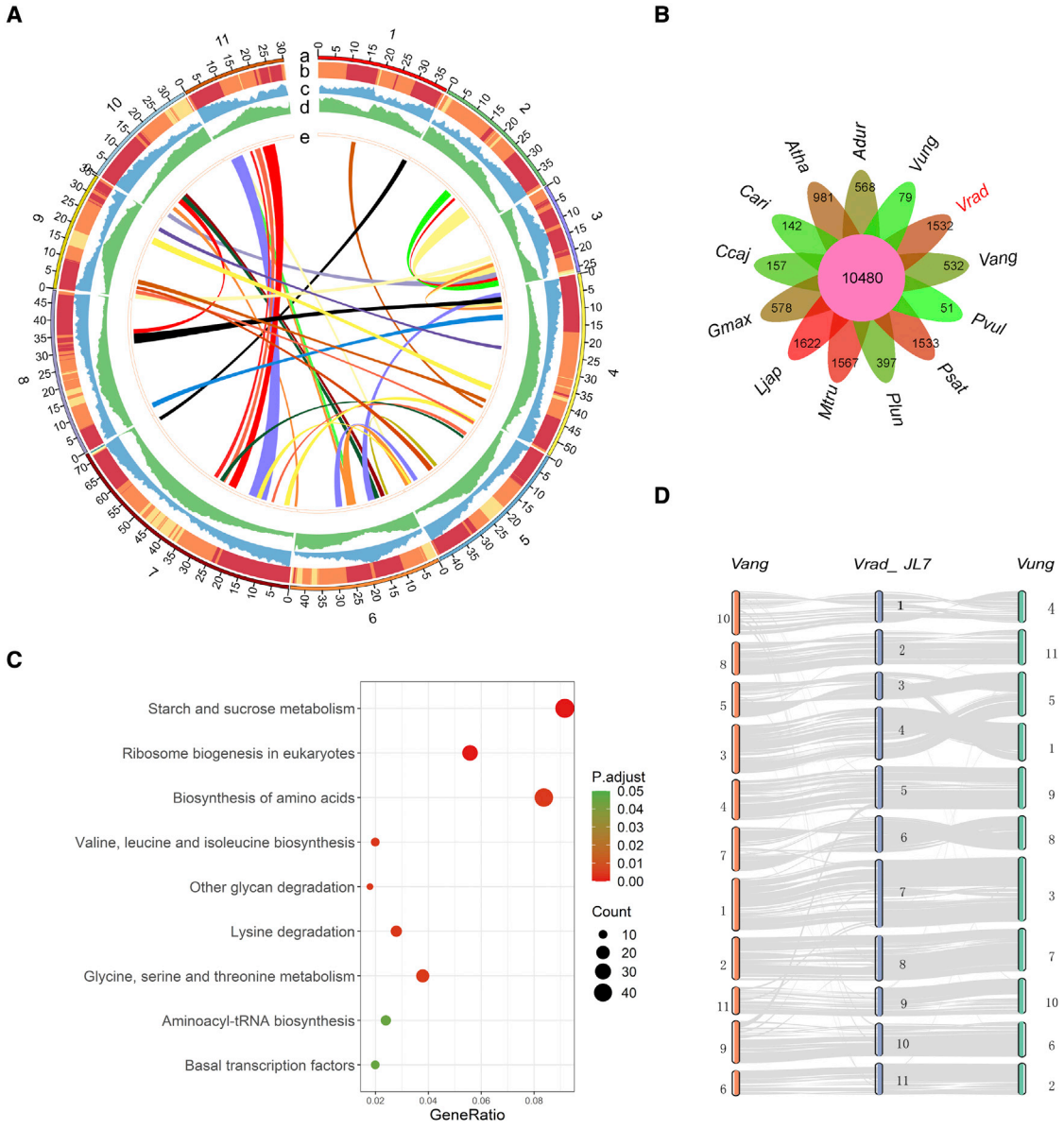


Figure 1. Genome assembly of mung bean and comparative genomic analysis of members of the Leguminosae.

(A) Landscape of genomic features of mung bean. The circles represent, from outermost to innermost, (a) pseudochromosomes, (b) GC content, (c) distribution of genes, (d) DNA transposon and retrotransposon density, and (e) intragenome collinear blocks.

(B) Orthologous gene families among 12 species of Leguminosae and *Arabidopsis thaliana* identified by OrthoFinder (Emms and Kelly, 2019). The numbers represent the gene families identified for each species.

(C) KEGG pathway enrichment of specific gene families in mung bean.

(D) Genomic collinearity between *Vigna radiata* (*Vrad_JL7*), *Vigna angularis* (*Vang*), and *Vigna unguiculata* (*Vung*).

3 groups. The NCL group contained mainly NCL lines plus 3 CBL/CLR lines with a close relationship to NCL lines. The vast majority of CBL and CLR lines clustered within either group 1 or group 2, where group 1 members were genetically closer to those of the NCL Group (Figure 2B). Population structure (K = 3) and principal component analysis (PCA) supported the same grouping structure (Figure 2C). The tree topology based on SNPs was very similar to the results of hierarchical clustering analysis based on gene presence/absence variation (PAV) results (Mantel statistic r: 0.9612; significance: 1e−4) (Supplemental Figure 10). All of the accessions from south China clustered in group 1, and most of the accessions from north China were in group 2 (Figure 2D). The NCL group had the highest nucleotide diversity ($\theta\pi$ = 1.63×10^{-3}) compared with those of group 1 (1.34×10^{-3}) and group 2 (1.02×10^{-3}) or those of south China (1.40×10^{-3}) and north China (1.37×10^{-3}). Pairwise F_{ST} between the NCL group and group 2 was the highest (0.283) compared with those of the NCL group versus group 1 (0.114) and group 1 versus group 2 (0.143), and the F_{ST} of the NCL group versus north China (0.151) was the highest compared with those of NCL group versus south China (0.088) and north China versus south China (0.069). These results indicate that the genetic basis narrowed during the

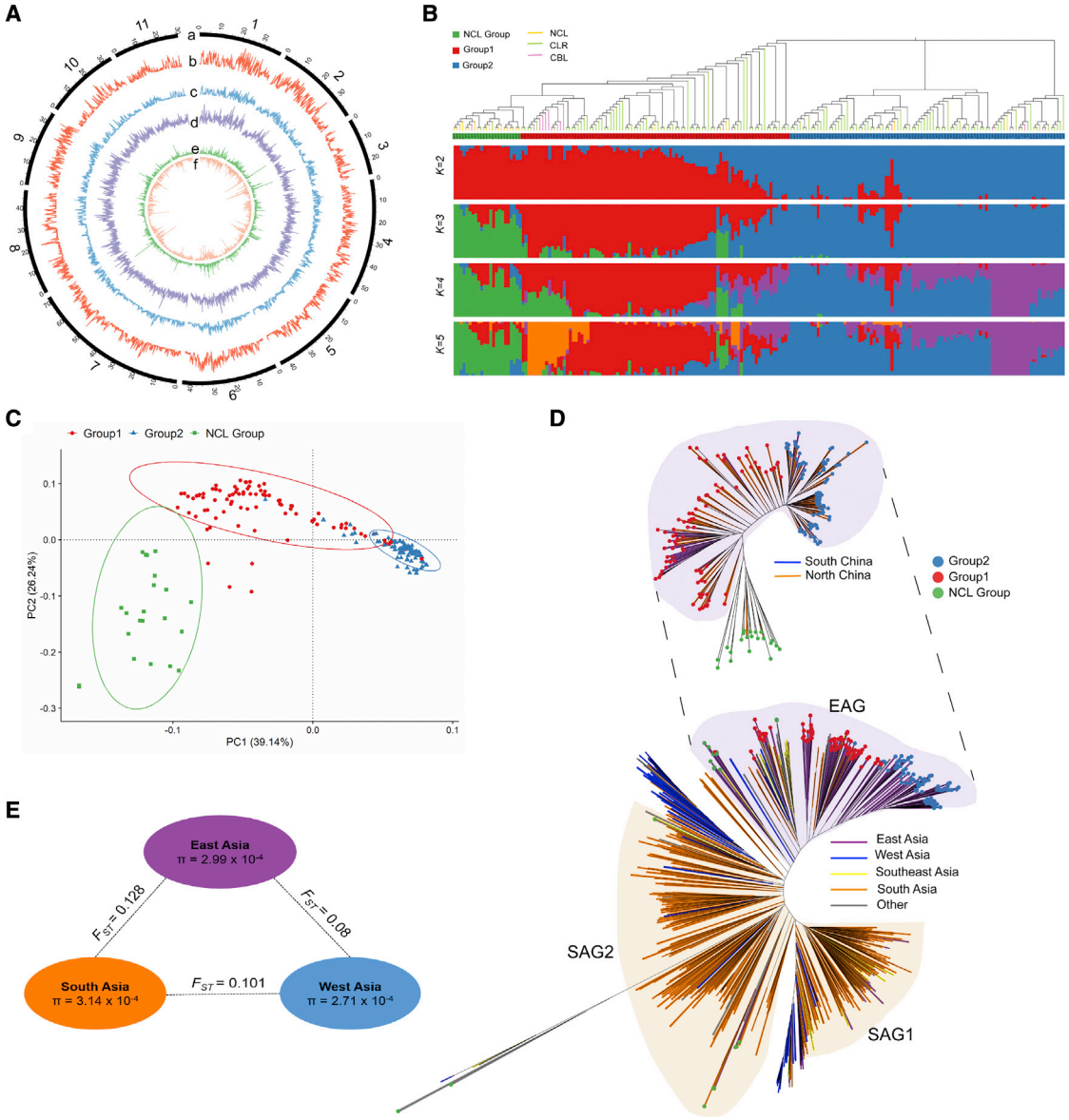


Figure 2. Population genomic analysis of mung bean.

(A) Atlas of variation of 217 accessions. The circles represent, from outermost to innermost, (a) pseudochromosomes; (b–d) SNP, indel, and SV density; and (e and f) nonsynonymous and synonymous SNPs.

(B) Unrooted tree and ADMIXTURE (K = 2–5) plot of 207 accessions inferred from core SNPs. The green, red, and blue strips of the tree represent the NCL group, group 1, and group 2, respectively, and the orange, green, and purple branches represent NCLs, CLRs, and CBLs, respectively.

(C) PCA plot of the first 2 eigenvectors of 207 accessions.

(D) Unrooted tree of 207 and 750 accessions. The red, green, and blue at the tips of the tree represent group 1, group 2, and the NCL group, respectively. The branch colors are shown in the legend.

(E) F-statistics (F_{ST}) and nucleotide diversity (π) of different subgroups in Asia.

process of mung bean adaptation, probably owing to migration from outside China to south China and then to north China.

To view the worldwide diversity landscape of mung bean, public GBS data of 533 accessions from 22 countries were collected and compared with the 217 accessions (Supplemental Figure 11A; Supplemental Data 4). A total of 5671 SNPs common to all 750 accessions from 23 countries were identified and used to construct an unrooted tree. As shown in Figure 2D, accessions from east Asia (~83.82% from China and ~2.90% from South Korea) were clustered into one group (east Asia group [EAG]), and accessions from south Asia (~92.46% from India) were clustered into two groups (south Asia group [SAG]1 and SAG2). Other accessions, including 49 from Iran (west Asia), were closely clustered with either SAG1 or SAG2. Wild mung bean accessions formed an outer group relatively closely related to the members of SAG2. There were two subgroups in the east Asia cluster, corresponding to group 1 and group 2 of the Chinese accessions. The $\theta\pi$ value for south Asia accessions was 3.14×10^{-4}, 5% higher than 2.99×10^{-4} for east Asia. The F_{ST} value for east Asia

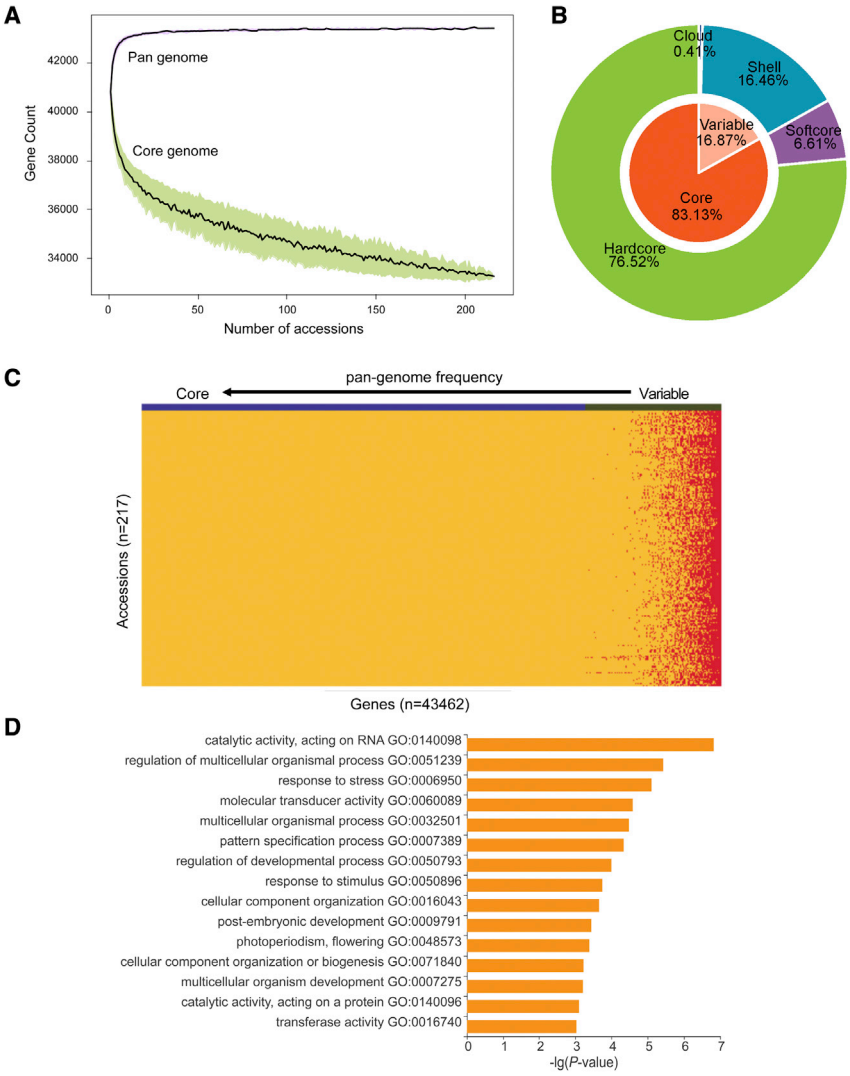


Figure 3. Pan-genome of Chinese mung bean germplasm.

(A) Simulations of the increase in pan-genome size and the decrease in core genome size based on 100 random combinations of each given number of accessions. The pan- and core-genome curves were fitted using data points from all random sub-samples and are indicated by solid black lines. The upper and lower edges of the purple and green areas correspond to the maximum and minimum numbers of genes, respectively.

(B) Composition of the mung bean pan-genome.

(C) Landscape of gene PAVs. The genes are sorted by their occurrence, with the highest frequency of occurrence on the left and the lowest on the right.

(D) GO term enrichment of variable genes.

versus south Asia was 0.128, much higher (60%) than that for east Asia versus west Asia (0.08) and higher (27%) than that for west Asia versus south Asia (0.101) (Figure 2E). These results seemed to support the hypothesis that mung bean originated and was domesticated in south Asia (India) (Fuller, 2007) and spread worldwide from there, including to west Asia and then to east Asia, probably through the Silk Road. After many years of adaptation and selection, mung bean varieties in China have formed two distinct groups.

Pan-genome and PAV analysis

De novo assemblies of all 217 mung bean accessions revealed a total of ~86 Gb contigs, with an average assembly size of ~397 Mb and an average contig N50 of ~3380 bp (Supplemental Table 7). After removing contamination and redundant sequences, a total of ~287.73 Mb of non-reference sequences, consisting of 288 128 contigs, were considered to be additional genome space of mung bean. A total of 3337 additional evidence-supported protein-coding genes were annotated, with an average gene length of 545 bp, much shorter than that of genes in the reference genome (4351 bp), indicating that most of these genes were partial or fragmented. Altogether, the assembled size of the mung bean pan-genome constructed in this study was ~762.92 Mb (including Vrad_JL7), and the total number of annotated genes was 43 462.

A "map-to-pan" strategy (Hu et al., 2017) was used to identify the PAV of genes. Genes with more than 20% of the coding DNA sequence (CDS) region covered by more than two-fold read depth were considered to have "presence" in the accession; the others, "absence." Similar to that of PanSoy (Torkamaneh et al., 2021), the mung bean pan-genome constructed mainly of Chinese accessions seemed closed, as most genes were included when randomly sampling 100 accessions from the collection (Figure 3A). Based on gene frequency, 33 258 (76.5%) genes were defined as hardcore genes, 2872 (6.6%) as softcore genes, 7154 (16.5%) as shell genes, and 178 (0.4%) as cloud genes (Figure 3B; Supplemental Data 5). The high proportion of core gene (83%) content was similar to proportions in the PanSoy (90.6%) (Torkamaneh et al., 2021), tomato (74.2%) (Gao et al., 2019), and pigeon pea (86.6%) pan-genomes (Zhao et al., 2020). Each accession had 15%–20% variable genes (including shell and cloud) (Figure 3C). The average length of variable genes on the reference genome was 2438 bp, with an average of 3 exons, which was significantly shorter than that of core genes (4562 bp), probably owing to fewer exons (3.01 exons per gene) in the soft genes than in the core genes (5.54 exons per gene).

The functions of core genes involved basic biological processes, such as cellular processes, metabolic processes, catalytic activity, and binding (Supplemental Figure 12). The functions of variable genes were enriched in many regulatory and environmental response processes, including catalytic activity, responses to stress and stimuli, regulation of multicellular organismal and developmental processes, and photoperiodic flowering (Figure 3D). The variable genes are likely to have played important roles in enabling mung bean to adapt to

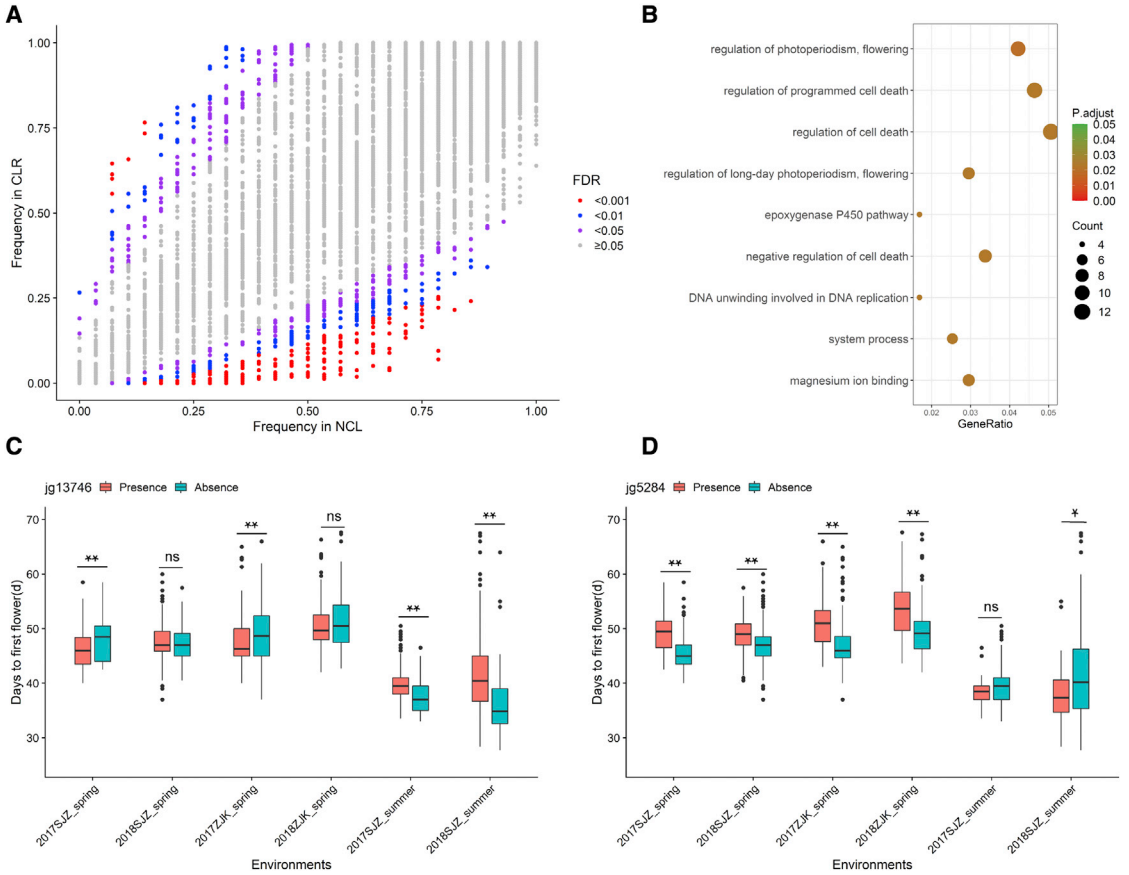


Figure 4. Gene PAV under selection during the adaptation process of mung bean from NCLs to CLRs.

(A) Scatterplots showing gene occurrence frequencies in NCLs and CLRs.

(B) KEGG pathway enrichment of genes that have undergone significant expression changes during adaptation.

(C and **D)** Examples of presence/absence variants of 2 genes (jg13746 and jg5284) that regulate flowering period in different environments. ns, *, and ** indicate statistical significance levels of $P \geq 0.05$, <0.05, and <0.01, respectively.

different environments from the tropics to the temperate zones. Interestingly, the distribution of variable genes was not even, and certain regions, such as 3.0–4.5 Mb on chromosome 2 and 23.0–29.5 Mb on chromosome 1, were hot spots for variable gene clustering (Supplemental Figure 13).

Gene PAV under selection during adaptation

Gene counts and frequencies in the NCLs, CLRs, and CBLs were compared in detail (see Supplemental text). To identify the gene PAV under selection during the adaptation process, the frequency of each gene in the NCLs was plotted against its frequency in the CLRs. As shown in Figure 4A, 809 genes showed significant differences in frequency between the two groups (Fisher's exact test, false discovery rate < 0.05). Among them, 215 genes were considered favorable, as their frequencies increased from NCLs to CLRs, and 412 genes were considered unfavorable. Genes that showed opposite significant changes in frequency from NCLs to CLRs and from CLRs to CBLs were not considered further. Genes related to flowering regulation and programmed cell death were significantly enriched in both favorable and unfavorable genes during adaptation (Figure 4B).

Nine PAV events for genes related to flowering regulation were identified, three of which (jg13350, jg13746, and Pang80812) were present in most CLRs and CBLs. The phenotypic data revealed that the presence of these genes could promote early flowering in spring but was associated with late flowering in summer (Figure 4C). The other six genes (jg1521, jg5273, jg5274, jg5281, jg5284, and Pang68295) were absent from most CLRs and CBLs and present in most NCLs. Their association with flowering phenotype was exactly the opposite of that in the previous case. The absence, instead of the presence, of these genes was associated with early flowering in spring but late flowering in summer (Figure 4D). Data from different environments largely supported the above observations, the signatures of increased linkage disequilibrium (LD), and the loss of genetic diversity in the regions surrounding the PAVs (Supplemental Figure 14), demonstrating that genes that promote early flowering in spring were selected during the adaptation process. No functional groups were found to be significantly enriched for the PAV genes during the improvement process (from CLRs to CBLs), although several genes encoding glucosidases could be candidates for selection.

SNP- and gene PAV-based GWAS of agronomic traits

Phenotypic data for 33 agronomic traits in 217 mung bean accessions were collected in Shijiazhuang (SJZ) for 2 seasons (spring and summer) over 2 years (2017 and 2018) and in Zhangjiakou

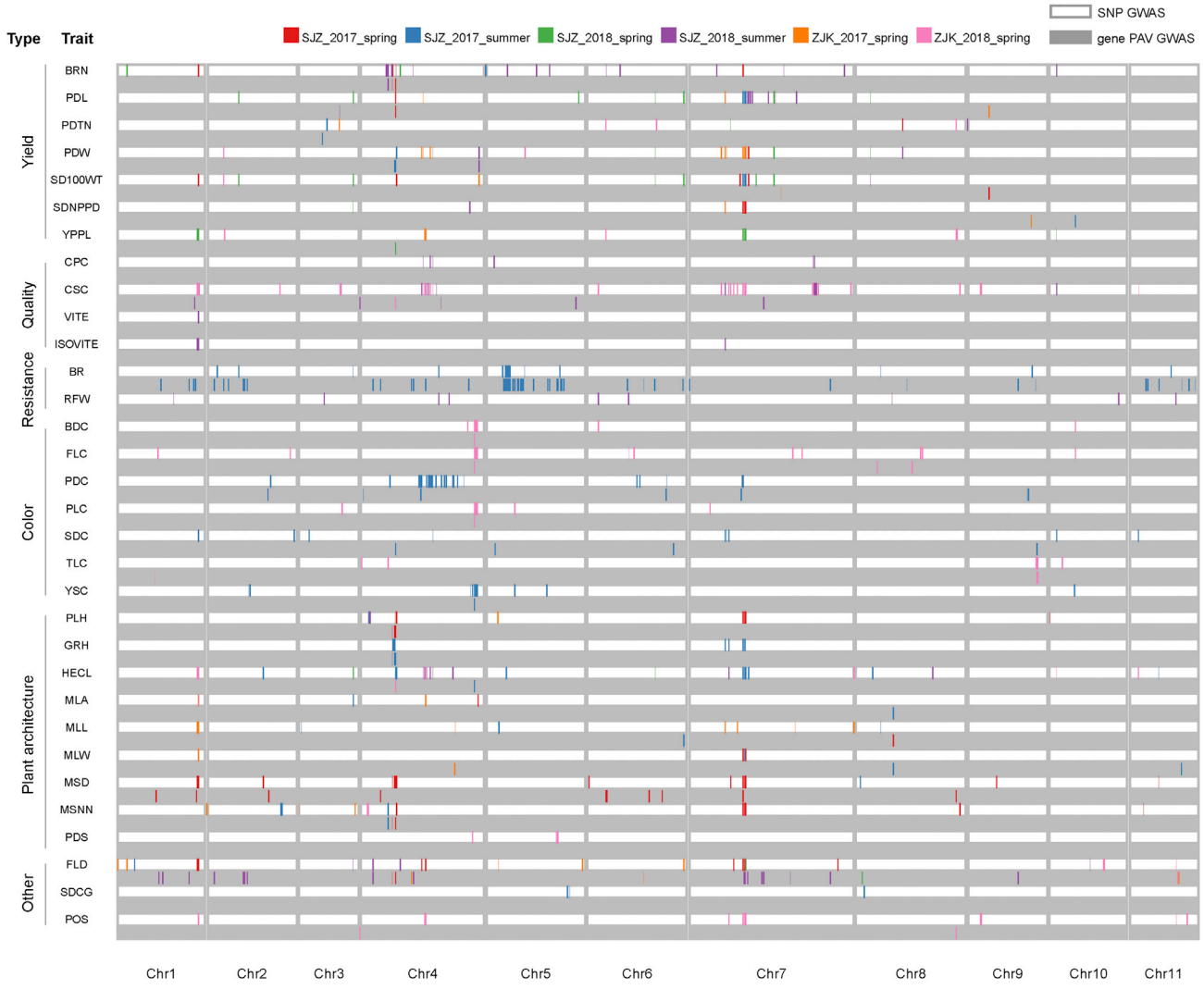


Figure 5. Summary of SNP GWAS and gene PAV GWAS results.
Distribution of all STAs and GPTA events identified by a SNP GWAS (white box) and a gene PAV GWAS (gray box) within the Vrad_JL7 genome for 32 traits of 6 types under 6 different environments. The abbreviations of the traits are shown in Supplemental Table 8.

(ZJK) for 1 season (spring) over 2 years (2017 and 2018). Overall, data for the same trait were strongly correlated across different environments and years. The average broad-sense heritability for traits observed multiple times was 0.72, ranging from 0.3 to 0.98 (Supplemental Table 8). Interestingly, some traits had strong correlations with others. For instance, the average Pearson correlation coefficient was 0.94 for VITE and ISOVITE; 0.89 for maximum leaf length, maximum leaf width, and maximum leaf area; and 0.78 for pod length (PDL), pod width, and 100-seed weight; and the average Spearman correlation coefficient was 0.85 for bud color (BDC), flower color (FLC), petiole color (PLC), and young stem color (YSC) (Supplemental Figure 15). It would not be surprising to discover genes involved in these traits showing pleiotropic effects.

SNP-trait association sites (STAs) were identified for all but one trait (trilobal leaf shape) in at least one environment via GWAS (Figure 5). A total of 2912 STAs were identified for each individual environment, and 1790, 43, and 23 of them were classified as robust, consistent, and stable, respectively (Supplemental Table 9; Supplemental Data 6). Although most STAs were located in intergenic regions, 35% of them were located in the transcribed regions of 248 genes (Supplemental text). Some STAs were shared among highly correlated traits, as was the case for 17 common STAs identified for maximum leaf length and maximum leaf width and 14 STAs for pod width and PDL. Notably, some genomic regions appeared to be hot spots for STAs, as they were associated with multiple traits that were not highly correlated. For example, the terminal region of chromosome 1 (35.499–35.986 Mb) contained 285 STAs for 14 traits, and the middle region of chromosome 7 (16.915–16.993 Mb) contained 132 STAs for 8 traits, including yield-, quality-, and plant architecture–related traits (Figure 5; Supplemental Data 6). These hot-spot genomic regions significantly associated with multiple traits, called "agro-islands," were also found on multiple chromosomes of chickpea (Plekhanova et al., 2017; Varshney et al., 2019a; Sokolkova et al., 2021) and pigeon pea (Varshney et al., 2017); these regions may be closely linked to

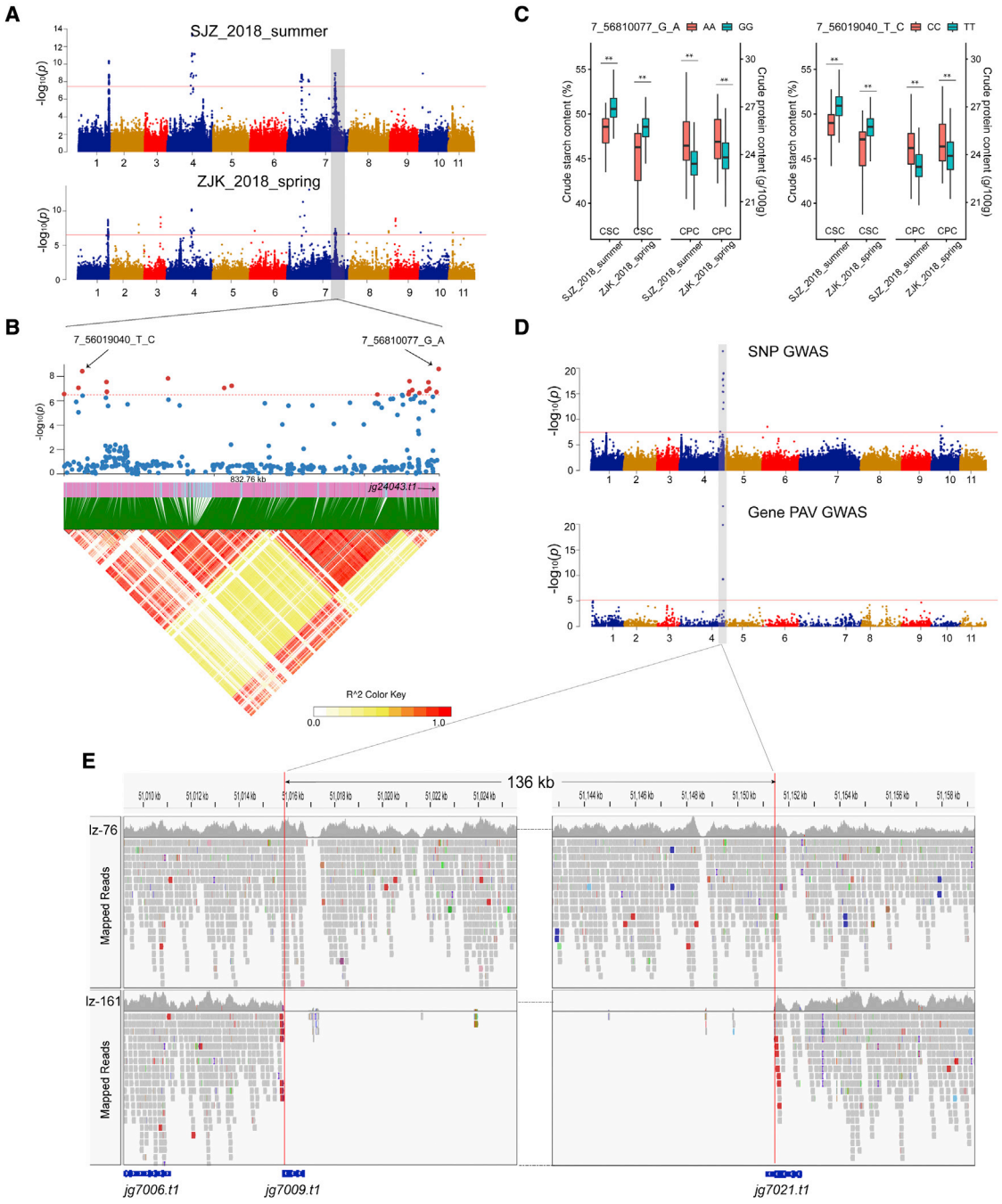


Figure 6. GWAS of CSC and color-related traits.

(A) Manhattan plot of GWAS of CSC traits in 2 different environments (SJZ_2018_summer and ZJK_2018_spring).

(B) LD heatmap showing the regions surrounding the strong peaks (chromosome 7: 55.979–56.812 Mb) identified by a SNP GWAS.

(C) Relationship between the alleles of the 2 STAs (7_56810077_G_A and 7_56019040_T_C) and the CSC and CPC in 2 environments. ** indicates a statistical significance level of $P < 0.001$.

(D) Manhattan plots of SNP-based GWAS and gene PAV-based GWAS for color-related traits, including BDC, FLC, PLC, and YSC.

(E) Presence/absence variants of 136 kb on chromosome 4 according to Integrative Genomics Viewer (IGV). Sample lz-76 (variety: Jilv7) contains this segment, and the color is purple. Sample lz-161 (variety: VC973A) is missing this segment, and the color is green or yellow.

domestication selection. The availability of "islands" in the mung bean genome (e.g., chromosomes 1, 4, and 7) facilitates the identification of breeding targets to reintroduce genetic diversity lost in modern breeding programs owing to domestication and crop improvement.

In addition to SNP GWAS, gene PAV GWAS was conducted by treating gene PAVs as genotyping data. A total of 391 gene PAV-trait association (GPTA) events were identified for 29 of 33 traits in multiple environments, corresponding to 259 unique genes (Figure 5; Supplemental Table 10; Supplemental Data 7).

Although most GPTA genes were associated with only a single trait, 52 genes were related to multiple traits. Approximately 60 GPTA genes were in close proximity (less than 100 kb) to STAs, half of which actually had a distance of less than 10 kb. As SVs would normally be coupled with SNPs in surrounding regions, consistent GPTA and STA events would be very helpful for identifying genes that regulate associated traits. However, reliable GPTA events not related to STA would be complementary to the SNP method for identifying candidate genes, as was reported in previous studies (Song et al., 2020).

Candidate genes associated with important agronomic traits

The contents of VITE and ISOVITE are important grain quality traits. The two traits were highly correlated (r = 0.94); thus, it was not surprising to identify common STAs for these two traits. Two regions were identified for regulation of ISOVITE, one on the terminus of chromosome 1 (35 556 181–35 673 741, 117 kb) and another on chromosome 7. The region on chromosome 1 was also identified for VITE. Twelve candidate genes were annotated in this region. None of them appeared to participate directly in the flavonoid biosynthesis pathway (KEGG map00941). Two genes (jg15859 and jg15860) were annotated as encoding glutamine synthetase, which could be used as a substrate for VITE/ISOVITE synthesis.

For other quality-related traits, a stable STA (7_56810077_G_A) on chromosome 7 was identified for crude starch content (CSC) (Figure 6A). The SNPs were located in the coding region of jg24043, a homolog of the soybean SWEET10 gene. This soybean homolog was reported to be a key gene that regulates seed size, oil content, and protein content (Wang et al., 2020). Interestingly, the most significant SNP (7_56019040_T_C) associated with crude protein content (CPC) was found to have strong LD with jg24043 (Figures 6B and 6C), indicating that the same gene may regulate CSC and CPC concurrently. Another STA identified for CSC was located on chromosome 3 between the jg9561 and jg9562 genes, and both were annotated as β-amylase, supporting their function in starch content.

Significant STA and GPTA events from the same region were identified for several color-related traits, including BDC, FLC, YSC, and PLC (Figure 6D). Eleven consecutive genes (jg7009–jg7020) and part of jg7021 exhibited PAV patterns among the mung bean population. It is very likely that the entire segment of 136 kb on chromosome 4 (51 015 805–51 151 503) containing these genes was missing in some accessions (Figure 6E). For 185 accessions that contained this segment, a purple color appeared in the buds, flowers, young stems, and petioles. For the other 32 accessions that lacked the segment, the corresponding tissues were either green or yellow. Several MYB90-like genes were annotated in the segment (Supplemental Figure 20). The soybean homolog R gene (Glyma.09G235100) plays an important regulatory role in the synthesis of anthocyanins during the process of seed coloring (Gao et al., 2021).

For yield-related traits, a reliable candidate region on chromosome 4 was identified for branch number by consistent STA and GPTA signals in multiple environments, narrowing candidate genes to one of five showing PAV events (Supplemental text). A homolog of NRT1/PTR FAMILY 2.13 could be a candidate gene for PDL, and a homolog of WUSCHEL-related homeobox 3 could be a candidate for yield per plant (YPPL) (Supplemental text). Many other candidate regions/genes were also identified, including a main region on chromosome 5 related to bruchid resistance (BR); two candidate genes (jg3587 and jg35209) for flowering date, which were homologous to the candidate genes Glyma03g01540 and Glyma09g33340/Glyma09g33350 identified via a soybean GWAS (Mao et al., 2017); and important regions for plant height, growth habit, pod shattering, and more. These candidate genes are valuable for further genetic discovery, and molecular markers could be developed to assist in the selection of target phenotypes for mung bean breeding.

DISCUSSION

Genome assembly at the chromosome scale is very important for crop genetics and breeding. In this study, we sequenced and assembled a high-quality mung bean genome (Vrad_JL7) through a combination of long-read and short-read sequencing technologies and annotated 40 125 protein-coding genes through the integration of *ab initio* prediction, RNA-seq, and homology data. We also assembled the first mung bean pan-genome from 217 accessions to understand the entire genome space and discovered genes related to important traits through SNP-based and PAV-based GWAS. The Vrad_JL7 assembly had improved completeness and continuity compared with those of the previously published VC1973A genome and the improved version (VC1973A_v2) (Supplemental Figure 9A; Table 1). It also displayed very good synteny with the genomes of adzuki bean and cowpea (Figure 1D). The annotated gene set was comprehensive, as indicated by the high protein BUSCO score (96.9%). Overall, among all of the available mung bean genome datasets, the Vrad_JL7 assembly and annotation had by far the highest quality.

The recent VC1973A genome (VC1973A_v2) assembled contigs/scaffolds under the guidance of a genetic map composed of 1100 molecular markers (Ha et al., 2021). Although the N50 of contigs and scaffolds seemed to be improved in the VC1973A_v2 assembly, the completeness was reduced by ∼5% (the BUSCO score decreased from 96.82% to 91.36%) (Table 1), and correctness was also compromised, as none of the VC1973A_v2 pseudochromosomes exhibited one-to-one collinearity with the adzuki bean and cowpea genomes (Supplemental Figure 9B). We noticed that some genomic regions of mung bean, cowpea, and adzuki bean seemed to be inverted in the comparison, which could be caused by misassembly in Vrad_JL7. For instance, the density distribution of genes and transposable elements suggested that some segments on chromosomes 1, 2, 3, 6, and 10 may need to be reversed in direction. These segments should be validated, and the assembly could be improved further by integrating genetic and/or optical physical maps. We also noticed that some gene models could be improved. Although 31 848 of 42 986 (74.09%) gene models had either RNA-seq or homology support, some were short (shorter than 200 nt), and others could be partial, fragmented, repetitive, or chimeric (Supplemental Data 1). More experimental evidence and manual curation are necessary to improve the genome annotation. Nonetheless, potential false-positive genes, such as

small open reading frames without homology evidence, did not have much influence on related conclusions in this study (functional enrichment analysis), as their functions could not be assigned.

Genetic variation was assessed in depth for the first time by resequencing 217 mung bean accessions mainly collected from China. Most Chinese accessions were clustered into two groups based on SNP and gene PAV data and were correlated with their geographic distribution (in terms of latitude). There were no clear genomic features distinguishing landraces from cultivars in the population structure analysis. This suggested that the improvement process in mung bean had been relatively slow, and few genomic regions were selected and fixed in modern breeding lines. Interestingly, we found that some landraces showed phenotypes similar to those of wild accessions, such as an indeterminate growth habit and pod shattering. These feralized landraces had similar levels of genetic diversity as other landraces. This phenomenon was probably caused by the de-domestication process, similar to that which occurred during the origin of weedy rice (Qiu et al., 2017). Because of the limited number of wild accessions collected in this study, we were not able to thoroughly analyze the genetics of mung bean domestication. As most of the 217 materials were from landraces and cultivars collected in China, which covered only a portion of global diversity, the assembled pan-genome could more appropriately be called the pan-genome of Chinese mung bean germplasm rather than the pan-genome of mung bean. Nevertheless, this study still provides new insights into the characteristics and genetic diversity of the mung bean genome.

Mung bean was proposed to have originated and been domesticated in India (Fuller, 2007; Kang et al., 2014), but there is no comprehensive population genetic evidence. For the first time, we compared the population diversity on a large scale by combining public GBS data from 533 accessions with resequencing data from 217 accessions. Although only a few thousand SNPs were common between the two datasets, nucleotide diversity ($\theta\pi$) (Figure 2E), LD decay (Supplemental Figure 11B), and other indicators supported the hypothesis that mung bean was introduced from south Asia to west Asia and then to east Asia, probably through the Silk Road. Although the diversity level of Southeast Asia accessions was also higher (3.18×10^{-4}) than that for east and west Asia, this was likely to be caused by collection of accessions from other regions by the Asian Vegetable R&D Center, located in Southeast Asia. After a long-term adaptation process, several distinct groups were formed among the mung bean population associated with their geographic regions, including two groups corresponding to south Asia and one group corresponding to east Asia. Accessions from west Asia and Southeast Asia were largely grouped together with accessions from south Asia. Within the east Asia group, the subgroup structure was very similar to the topology of trees constructed for the 217 accessions alone by the use of the large-scale SNP or PAV data. This demonstrated that the overall structure for all of the accessions based on the common SNP set was correct and that east Asia (mainly China) accessions had formed distinct population features during cultivation and improvement over the past 2000 years. It also strongly suggested that the improvement of mung bean could benefit greatly from the exchange of germplasm across different geographic regions.

Many genomic regions associated with agronomic traits were identified in this study, including a large region with important signals for BR on chromosome 5: 9 456 956–11 579 877 (~2.12 Mb) containing 156 genes. One gene, jg26964, was annotated as gibberellin-regulated protein 14, which corresponded to *Vradi05g03730* in the VC1973A genome and was located upstream of the BR marker W02a4 (Kang et al., 2014). Other previously studied candidate genes, such as genes encoding RD22 proteins and polygalacturonase-inhibiting proteins, were also included in this region (Liu and Fan, 2018; Kaewwongwal et al., 2020). Because neither VC1973A nor Vrad_JL7 shows resistance to bruchids, the two reference genomes may not contain functional resistance genes. We further used the PAV gene data to identify possible candidate genes that are likely to be missing in the reference genome. Of 217 accessions, 8 showed the bruchid resistance phenotype. Five pangenes (Pang34265, Pang44622, Pang57772, Pang58608, and Pang64254) were identified as possible candidate genes based on their frequencies and were also identified in our gene PAV-based GWAS. Among them, Pang58608 and Pang64254 were annotated as resistant specific protein-3, making them very good candidates for further verification. Taken together, the high-quality Vrad_JL7 genome, genetic variation map, and pan-genome constructed in this study lay a good foundation for gene discovery and breeding in mung bean.

METHODS

Sample collection and plant materials

In this study, 217 representative germplasms were selected from 589 accessions in the Chinese mung bean germplasm bank based on their phenotypes, the results of cluster analysis using simple sequence repeat markers, and their geographic distribution. The accessions were planted in SJZ (114.48E, 38.03N) and ZJK (114.88E, 40.82N), Hebei Province, China, for phenotypic observations. Detailed information for each accession is given in Supplemental Table 4. The seed coats of wild accessions were cut open with a knife on the opposite side of the hilum to promote germination. To reduce the environmental impact, we planted in SJZ in the spring and summer seasons of 2017 and 2018, respectively. However, ZJK was planted only in the spring season of the 2 years.

With respect to publicly available data, we downloaded three genotyping-by-sequencing datasets of the mung bean core germplasms from the NCBI BioProject database under accession numbers PRJNA645721, PRJNA609409, and PRJNA664607, corresponding to 296 (Sokolkova et al., 2020), 144 (Reddy et al., 2020), and 93 accessions (Wu et al., 2020b), respectively (Supplemental Data 4). In total, 750 mung bean accessions from 23 countries were used for the diversity analysis.

Genome sequencing and assembly

We selected the elite mung bean variety JL7 for whole-genome sequencing and assembly. Genomic DNA was extracted, and libraries were constructed and sequenced on the Illumina NovaSeq platform. The raw reads were preprocessed to remove adaptors and low-quality bases. The K-mer distribution was estimated using KMC (version 3.0) (Kokot et al., 2017) with the parameters "-k31 -t16 -m64 -ci1 -cs10000," and the genome size was estimated with GenomeScope (version 2.0) (Vurture et al., 2017).

For *de novo* assembly of the Vrad_JL7 genome, we used long-read sequencing based on the PacBio Sequel II platform. High-molecular-weight (HMW) DNA was used to construct a DNA library with a ~20-kb insert size, and the library was subsequently sequenced on the PacBio Sequel II sequencing platform at Novogene (Beijing, China). Flye (version

2.8.3-b1695) (Kolmogorov et al., 2019) was used for *de novo* assembly, and BWA-MEM (version 0.7.17) (Li and Durbin, 2009) was used to map Illumina PE reads to the assembled sequence. The genome was then polished using Pilon (version 1.24) (Walker et al., 2014). Hi-C sequencing library construction and sequencing (150-bp PE) were completed on the Illumina HiSeq platform. After removing low-quality raw reads, sequences were mapped to the assembled genome with Juicer (default parameters) (Durand et al., 2016a), and the 3D-DNA pipeline (https://github.com/aidenlab/3d-dna) was then used to obtain the chromosome-scale assembly. Finally, we used Juicebox Assembly Tools (version 1.11.9) (Durand et al., 2016b) to manually correct errors and visualize the results of the assembly.

For genome assessment, the Illumina PE reads were mapped to Vrad_JL7 using BWA-MEM, and the coverage of short reads on the genome was calculated using SAMtools (version 1.13) and Mosdepth (version 0.3.1) (Pedersen and Quinlan, 2018). Then, 2326 single-copy orthologs of dicot species (eudicots_odb10 database) were evaluated using BUSCO (version 3.1.0) (Simao et al., 2015). Finally, the LAI was calculated and evaluated with LTR_retriever (Ou et al., 2018b).

Genome annotation

We constructed a transposable element library by *ab initio* predictions of repeated sequences using RepeatModeler followed by RepeatMasker (version 4.0.9) (Tarailo-Graovac and Chen, 2009) to search for repeat sequence annotations. The BRAKER2 pipeline (Bruna et al., 2021), which integrates RNA-seq and protein homology-based methods, was executed to predict protein-coding genes. First, clean RNA-seq reads (~21.18 Gb from mung bean flower tissue) were mapped to the genome using HISAT2 (version 2.10.2) to obtain transcriptome mapping data. Second, all of the protein sequences from OrthoDB (version 10.0) (Kriventseva et al., 2019) were downloaded and mapped to the genome assembly with ProtHint (version 2.6.0). We then used GeneMark-EP+ (version 4.65) (Bruna et al., 2020) to integrate the two types of data. The final gene structure was predicted by Augustus (version 3.4.0), and the untranslated regions (UTRs) were predicted by GUSHR (version 1.0) (Stanke et al., 2008). BUSCO was used with eudicots_odb10 to evaluate the quality of the annotation. We also predicted rRNA using Barrnap (version 0.9) (https://github.com/tseemann/barrnap) and tRNA and other noncoding RNAs using Infernal (version 1.1.4) (Nawrocki and Eddy, 2013) by searching the Rfam (version 14.1) database (Kalvari et al., 2021). Functional annotations were assigned by homology searching against public databases, including the SwissProt, NR, and Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups databases, using DIAMOND ($E \leq 1e^{-6}$). The GO and KEGG annotations were assigned by transferring the annotation data from the Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups. Pfam and InterPro motif annotations were assigned using InterProScan (version 5.36).

Comparative genome analyses

Protein sequences from the following 12 eudicot genomes were downloaded: *Vigna angularis*, *Phaseolus vulgaris*, *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana* from Ensembl Plants (http://plants.ensembl.org/index.html); *Vigna unguiculata*, *Cajanus cajan*, *Cicer arietinum*, and *Arachis duranensis* from the NCBI Reference Sequence Database; *Pisum sativum* from https://urgi.versailles.inra.fr/download/pea/; *Phaseolus lunatus* from https://data.jgi.doe.gov/refine-download/phytozome?organism=Plunatus&expanded=563; and *Lotus japonicus* from https://drive.google.com/drive/folders/1yMR4fiRKt7fWZ6yTxlB7lsJoClQT9Q_0. Orthologous genes were identified using OrthoFinder (version 2.37) (Emms and Kelly, 2019). An ultrametric tree was constructed using r8s (version 1.81) (Sanderson, 2002), and TimeTree (http://www.timetree.org/) was used to calibrate divergence times. According to the evolutionary tree results with divergence times and gene family clustering by CAFE (version 4.2) (De Bie et al., 2006), the number of gene family members of each branch's ancestors was estimated by a birth mortality model to predict the contraction and expansion of gene families of species relative to their ancestors ($P < 0.05$). GO and KEGG enrichment analyses were performed using clusterProfiler (version 3.14.0) (Yu et al., 2012). Genome collinearity analysis was subsequently performed using MCScan (Python version, https://github.com/tanghaibao/jcvi/), and syntenic blocks were visualized using the dotplot script in the jcvi package. WGDdetector (version 1.00) (Yang et al., 2019) was used to identify whole-genome duplications (WGDs) with the synonymous mutation rate (Ks) method.

Pan-genome construction and analysis

The pan-genome was constructed following the method used for the tomato pan-genome (Gao et al., 2019). First, the genomes of 217 accessions were individually assembled using MaSuRCA (Zimin et al., 2013). Then, each assembled contig was aligned to Vrad_JL7 using QUAST (Gurevich et al., 2013), and contigs longer than 500 nt with an identity less than the threshold were extracted as non-reference sequences. We set the identity threshold to 90% after comparing the relationship between the length and identity of all contigs. Second, we aligned the non-reference sequences to the NT database and removed sequences ($E < 1e^{-5}$) that showed good homology to microorganisms, animals, and other non-Fabales plants. The clean non-reference sequences obtained were then clustered using CD-HIT (version 4.8.1) (Li and Godzik, 2006) to remove redundancy using an identity threshold of 90%. The non-redundant contigs were further aligned to the reference genome using BLAST to ensure that no single contig showed good identity to the reference. Finally, the non-redundant and non-reference contigs were merged with the reference mung bean genome as the mung bean pan-genome. The same annotation pipelines were used to annotate the gene structure and function of the pan-genome.

We used a map-to-pan strategy to identify core and variable genes. A total of 217 mung bean accessions were aligned to the reference genome using BWA-MEM and then Mosdepth (version 0.3.1) (Pedersen and Quinlan, 2018) to estimate CDS coverage. Genes with 2× coverage on at least 20% of the entire CDS region were considered present; otherwise, they were considered absent. By estimating the frequency of PAVs among all 217 accessions, we divided the genes into two categories: core (absence rate <0.05) and variable genes (absence rate ≥0.05). Four subcategories, hardcore, softcore, shell, and cloud, were defined as genes present in 100%, >99%, 1%–99%, and <1% of all accessions, respectively. To estimate the size of the pan-genome and core genome, samples were randomly picked (n = 1–217), and the process was iterated 100 times. Wilcoxon's test ($P < 0.05$) was used to denote the level of significance for differences in gene numbers in different subgroups.

Resequencing and variant calling

Genomic DNA was extracted from 217 mung bean accessions and sequenced on the Illumina NovaSeq platform. The clean reads were mapped to Vrad_JL7 using BWA-MEM with default parameters. Picard (version 2.18.17, http://broadinstitute.github.io/picard/) was used to remove PCR duplicates. Genetic variants, including SNPs and short indels (<15 bp), were detected using the Genome Analysis Toolkit (GATK) (version 3.8.1). SNPs were filtered with the following parameters: QD < 2.0, MQ < 40.0, FS > 60.0, SOR >3.0, MQRankSum < −12.5, and ReadPosRankSum < −8.0, and indels were removed with the parameters QD < 2.0, FS > 200.0, MQ < 40.0, SOR >10.0, and ReadPosRankSum < −20.0. Based on the filtered SNP set, we defined a core SNP/indel set by removing SNPs/indels with more than 2 alleles, a >10% missing rate, and a minor allele frequency of <5%. DELLY software (version 0.8.3) (Rausch et al., 2012) was used to identify SVs, including large indels (>15 bp), duplicate copy-number variations, inversions, and translocations. For public GBS data, the GATK best-practice pipeline described above and NGSEP (version 3.0.2) (Perea et al., 2016) with default parameters were used to call SNPs, and their results were consistent. A total of 5671 SNPs (40% missing rate filtered and imputed using Beagle [version 5.2]; Browning et al., 2018)

were ultimately merged from 750 accessions with GATK. According to the gene models of the Vrad_JL7 assembly, the genetic variants identified above were annotated using SnpEff (version 5.0e) (Cingolani et al., 2012), and the density along each chromosome was determined with VCFtools (version 0.1.17) using 500-kb sliding windows (Danecek et al., 2011).

Population genetic diversity and structure analysis

Based on the SNP data, population structure was determined using ADMIXTURE (version 1.3) (Alexander et al., 2009) with a block-relaxation algorithm. PCA was performed using the smartpca function in EIGENSOFT (version 6.1.4) (Price et al., 2006) with default parameters, and the first two eigenvectors were plotted. Neighbor-joining trees were constructed with the R package ape, and iTOL (https://itol.embl.de/) was used to visualize the trees. Population structure analysis of genic PAVs was performed using the hclust function in R. Weir and Cockerham's estimator of F_{ST} was used to measure genetic differentiation among multiple subpopulations. The values for F_{ST} and genome-wide diversity ($\theta\pi$) were calculated using VCFtools.

Phenotyping and GWAS

Thirty-three agronomic traits were observed for mung bean accessions planted in the field (2 seasons in SJZ and 1 season in ZJK for 2 consecutive years): 19 quantitative traits and 14 discrete traits. Fifteen traits (mostly quantitative traits) were recorded in all 6 fields, and another 15 traits were recorded only once. For traits collected more than three times, the average values were used in downstream analysis. Detailed information for each trait is given in Supplemental Table 8. GWASs based on SNP and gene PAV data were performed using a mixed linear model in the genome-wide efficient mixed-model association algorithm (GEMMA, version 0.98.1) (Zhou and Stephens, 2012). The first three principal components from PCA and a genomic relationship matrix were used to correct the population structure and random polygenic effect. Significant signals were identified using a $P < 0.05$ threshold and applying an adjusted Bonferroni test. STAs and GPTAs that passed the threshold were then evaluated for stability, consistency, and robustness based on standards used by Varshney et al. (2019a, 2019b). STAs/GPTAs with more than 15% phenotypic variation explained were considered robust; STAs/GPTAs identified for more than one location were considered stable; and STAs/GPTAs identified across >1 year/season were defined as consistent. The LD blocks surrounding GWAS signals were further evaluated using LDBlockShow (Dong et al., 2021).

ACCESSION NUMBERS

These sequence data have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA802060 and the National Genomics Data Center BioProject database (https://ngdc.cncb.ac.cn/bioproject/) under accession number PRJCA008996 (GSA: CRA006590). The genome assembly sequences, gene annotation, and supplemental data can be accessed from the FigShare database (https://doi.org/10.6084/m9.figshare.19583446).

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

This research was supported by the National Key R&D Program of China (2019YFD1000700/2019YFD1000702), the China Agricultural Research System (CARS-08-G3), the Key Research and Development Program of Hebei (21326305D), the Hebei Agriculture Research System (HBCT2018070203), and the Hebei Talent Project.

AUTHOR CONTRIBUTIONS

J.T. and C.L. designed the experiments. C.L., Y.W., J.T., B.F., D.X., Z.C., Y.G., X.W., S.L., Q.S., Z.Z., S.W., Q.S., H.S., and Y.S. carried out the phenotyping. J.P., C.L., B.W., and X.W. carried out the sequencing and data analysis. J.T., C.L., J.P., B.W., J.W., and Y.W. wrote the manuscript. All of the authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Jinfeng Chen at the Institute of Zoology, Chinese Academy of Sciences, and Dr. Jun Yang at the Shanghai Chenshan Botanical Garden for their useful comments on the data analysis and manuscript.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

Received: February 14, 2022
Revised: May 31, 2022
Accepted: June 22, 2022
Published: June 26, 2022

REFERENCES

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. **19**:1655–1664. https://doi.org/10.1101/gr.094052.109.

Breria, C.M., Hsieh, C.H., Yen, T.B., Yen, J.Y., Noble, T.J., and Schafleitner, R. (2020). A SNP-based genome-wide association study to mine genetic loci associated to salinity tolerance in mungbean (Vigna radiata L.). Genes **11**:759. https://doi.org/10.3390/genes11070759.

Browning, B.L., Zhou, Y., and Browning, S.R. (2018). A one-penny imputed genome from next-generation reference panels. Am. J. Hum. Genet. **103**:338–348. https://doi.org/10.1016/j.ajhg.2018.07.015.

Brůna, T., Lomsadze, A., and Borodovsky, M. (2020). GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genom Bioinform **2**:lqaa026. https://doi.org/10.1093/nargab/lqaa026.

Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom Bioinform **3**:lqaa108. https://doi.org/10.1093/nargab/lqaa108.

Cao, D., Li, H., Yi, J., Zhang, J., Che, H., Cao, J., Yang, L., Zhu, C., and Jiang, W. (2011). Antioxidant properties of the mung bean flavonoids on alleviating heat stress. PLoS One **6**:e21071. https://doi.org/10.1371/journal.pone.0021071.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly **6**:80–92. https://doi.org/10.4161/fly.19695.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics **27**:2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics **22**:1269–1271. https://doi.org/10.1093/bioinformatics/btl097.

Dong, S.S., He, W.M., Ji, J.J., Zhang, C., Guo, Y., and Yang, T.L. (2021). LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. Brief Bioinform. **22**:bbaa227. https://doi.org/10.1093/bib/bbaa227.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016a). Juicer provides a one-click System for analyzing loop-resolution Hi-C experiments. Cell Syst **3**:95–98. https://doi.org/10.1016/j.cels.2016.07.002.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016b). Juicebox provides a visualization System for Hi-C contact maps with unlimited zoom. Cell Syst **3**:99–101. https://doi.org/10.1016/j.cels.2015.07.012.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. **20**:238. https://doi.org/10.1186/s13059-019-1832-y.

Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., Hu, G., Zhou, Z., Yu, H., Zhang, M., et al. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. Genome Biol. **18**:161. https://doi.org/10.1186/s13059-017-1289-9.

Fuller, D.Q. (2007). Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World. Ann. Bot. **100**:903–924. https://doi.org/10.1093/aob/mcm048.

Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D.M., Burzynski-Chang, E.A., Fish, T.L., Stromberg, K.A., Sacks, G.L., et al. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. Nat. Genet. **51**:1044–1051. https://doi.org/10.1038/s41588-019-0410-2.

Gao, R., Han, T., Xun, H., Zeng, X., Li, P., Li, Y., Wang, Y., Shao, Y., Cheng, X., Feng, X., et al. (2021). MYB transcription factors GmMYBA2 and GmMYBR function in a feedback loop to control pigmentation of seed coat in soybean. J. Exp. Bot. **72**:4401–4418. https://doi.org/10.1093/jxb/erab152.

Garcia, T., Duitama, J., Zullo, S.S., Gil, J., Ariani, A., Dohle, S., Palkovic, A., Skeen, P., Bermudez-Santana, C.I., Debouck, D.G., et al. (2021). Comprehensive genomic resources related to domestication and crop improvement traits in Lima bean. Nat. Commun. **12**:702. https://doi.org/10.1038/s41467-021-20921-1.

Golicz, A.A., Batley, J., and Edwards, D. (2016). Towards plant pangenomics. Plant Biotechnol. J **14**:1099–1105. https://doi.org/10.1111/pbi.12499.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics **29**:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

Ha, J., Satyawan, D., Jeong, H., Lee, E., Cho, K.H., Kim, M.Y., and Lee, S.H. (2021). A near-complete genome sequence of mungbean (Vigna radiata L.) provides key insights into the modern breeding program. Plant Genome **14**:e20121. https://doi.org/10.1002/tpg2.20121.

Hu, Z., Sun, C., Lu, K.C., Chu, X., Zhao, Y., Lu, J., Shi, J., and Wei, C. (2017). EUPAN enables pan-genome studies of a large number of eukaryotic genomes. Bioinformatics **33**:2408–2409. https://doi.org/10.1093/bioinformatics/btx170.

Hurgobin, B., Golicz, A.A., Bayer, P.E., Chan, C.K., Tirnaz, S., Dolatabadian, A., Schiessl, S.V., Samans, B., Montenegro, J.D., Parkin, I.A.P., et al. (2018). Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid Brassica napus. Plant Biotechnol. J **16**:1265–1274. https://doi.org/10.1111/pbi.12867.

Kaewwongwal, A., Liu, C., Somta, P., Chen, J., Tian, J., Yuan, X., and Chen, X. (2020). A second VrPGIP1 allele is associated with bruchid resistance (Callosobruchus spp.) in wild mungbean (Vigna radiata var. sublobata) accession ACC41. Mol. Genet. Genomics **295**:275–286. https://doi.org/10.1007/s00438-019-01619-y.

Kalvari, I., Nawrocki, E.P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., Griffiths-Jones, S., Toffano-Nioche, C., Gautheret, D., Weinberg, Z., et al. (2021). Rfam 14: expanded coverage of metagenomic, viral and microRNA families. Nucleic Acids Res. **49**:D192–D200. https://doi.org/10.1093/nar/gkaa1047.

Kang, Y.J., Kim, S.K., Kim, M.Y., Lestari, P., Kim, K.H., Ha, B.K., Jun, T.H., Hwang, W.J., Lee, T., Lee, J., et al. (2014). Genome sequence of mungbean and insights into evolution within Vigna species. Nat. Commun. **5**:5443. https://doi.org/10.1038/ncomms6443.

Kokot, M., Długosz, M., and Deorowicz, S. (2017). KMC 3: counting and manipulating k-mer statistics. Bioinformatics **33**:2759–2761. https://doi.org/10.1093/bioinformatics/btx304.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. **37**:540–546. https://doi.org/10.1038/s41587-019-0072-8.

Kriventseva, E.V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F.A., and Zdobnov, E.M. (2019). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. **47**:D807–D811. https://doi.org/10.1093/nar/gky1053.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics **22**:1658–1659. https://doi.org/10.1093/bioinformatics/btl158.

Liu, C.Y., Su, Q.Z., Fan, B.J., Cao, Z.M., Zhang, Z.X., Wu, J., Cheng, X.Z., and Tian, J. (2018). Genetic mapping of bruchid resistance gene in mungbean V1128. Acta Agron. Sin. **44**:1875. https://doi.org/10.3724/SP.J.1006.2018.01875.

Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., Zhou, G.A., Zhang, H., Liu, Z., Shi, M., et al. (2020). Pan-genome of wild and cultivated soybeans. Cell **182**:162–176.e13. https://doi.org/10.1016/j.cell.2020.05.023.

Lonardi, S., Muñoz-Amatriaín, M., Liang, Q., Shu, S., Wanamaker, S.I., Lo, S., Tanskanen, J., Schulman, A.H., Zhu, T., Luo, M.C., et al. (2019). The genome of cowpea (Vigna unguiculata [L.] Walp.). Plant J. **98**:767–782. https://doi.org/10.1111/tpj.14349.

Mao, T., Li, J., Wen, Z., Wu, T., Wu, C., Sun, S., Jiang, B., Hou, W., Li, W., Song, Q., et al. (2017). Association mapping of loci controlling genetic and environmental interaction of soybean flowering time under various photo-thermal conditions. BMC Genomics **18**:415. https://doi.org/10.1186/s12864-017-3778-3.

Michael, T.P., and VanBuren, R. (2020). Building near-complete plant genomes. Curr. Opin. Plant Biol. **54**:26–33. https://doi.org/10.1016/j.pbi.2019.12.009.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics **29**:2933–2935. https://doi.org/10.1093/bioinformatics/btt509.

Noble, T.J., Tao, Y., Mace, E.S., Williams, B., Jordan, D.R., Douglas, C.A., and Mundree, S.G. (2017). Characterization of linkage disequilibrium and population structure in a mungbean diversity panel. Front. Plant Sci. **8**:2102. https://doi.org/10.3389/fpls.2017.02102.

Ou, L., Li, D., Lv, J., Chen, W., Zhang, Z., Li, X., Yang, B., Zhou, S., Yang, S., Li, W., et al. (2018a). Pan-genome of cultivated pepper (Capsicum) and its use in gene presence–absence variation analyses. New Phytol. **220**:360–363. https://doi.org/10.1111/nph.15413.

Ou, S., Chen, J., and Jiang, N. (2018b). Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res. **46**:e126. https://doi.org/10.1093/nar/gky730.

Pedersen, B.S., and Quinlan, A.R. (2018). Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics **34**:867–868. https://doi.org/10.1093/bioinformatics/btx699.

Perea, C., De La Hoz, J.F., Cruz, D.F., Lobaton, J.D., Izquierdo, P., Quintero, J.C., Raatz, B., and Duitama, J. (2016). Bioinformatic analysis of genotype by sequencing (GBS) data with NGSEP. BMC Genomics **17** (*Suppl 5*):498. https://doi.org/10.1186/s12864-016-2827-7.

Plekhanova, E., Vishnyakova, M.A., Bulyntsev, S., Chang, P.L., Carrasquilla-Garcia, N., Negash, K., Wettberg, E.V., Noujdina, N., Cook, D.R., Samsonova, M.G., et al. (2017). Genomic and phenotypic analysis of Vavilov's historic landraces reveals the impact of environment and genomic islands of agronomic traits. Sci. Rep. **7**:4816. https://doi.org/10.1038/s41598-017-05087-5.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. **38**:904–909. https://doi.org/10.1038/ng1847.

Qin, P., Lu, H., Du, H., Wang, H., Chen, W., Chen, Z., He, Q., Ou, S., Zhang, H., Li, X., et al. (2021). Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. Cell **184**:3542–3558.e16. https://doi.org/10.1016/j.cell.2021.04.046.

Qiu, J., Zhou, Y., Mao, L., Ye, C., Wang, W., Zhang, J., Yu, Y., Fu, F., Wang, Y., Qian, F., et al. (2017). Genomic variation associated with local adaptation of weedy rice during de-domestication. Nat. Commun. **8**:15323. https://doi.org/10.1038/ncomms15323.

Rausch, T., Zichner, T., Schlattl, A., Stutz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics **28**:i333–i339. https://doi.org/10.1093/bioinformatics/bts378.

Reddy, V.R.P., Das, S., Dikshit, H.K., Mishra, G.P., Aski, M., Meena, S.K., Singh, A., Pandey, R., Singh, M.P., Tripathi, K., et al. (2020). Genome-wide association analysis for phosphorus use efficiency traits in mungbean (Vigna radiata L. Wilczek) using genotyping by sequencing approach. Front. Plant Sci. **11**:537766. https://doi.org/10.3389/fpls.2020.537766.

Sanderson, M.J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics **19**:301–302. https://doi.org/10.1093/bioinformatics/19.2.301.

Schafleitner, R., Nair, R.M., Rathore, A., Wang, Y.W., Lin, C.Y., Chu, S.H., Lin, P.Y., Chang, J.C., and Ebert, A.W. (2015). The AVRDC - the World Vegetable Center mungbean (Vigna radiata) core and mini core collections. BMC Genomics **16**:344. https://doi.org/10.1186/s12864-015-1556-7.

Shen, Y., Du, H., Liu, Y., Ni, L., Wang, Z., Liang, C., and Tian, Z. (2019). Update soybean Zhonghuang 13 genome to a golden reference. Sci. China Life Sci. **62**:1257–1260. https://doi.org/10.1007/s11427-019-9822-2.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Sokolkova, A., Burlyaeva, M., Valiannikova, T., Vishnyakova, M., Schafleitner, R., Lee, C.R., Ting, C.T., Nair, R.M., Nuzhdin, S., Samsonova, M., et al. (2020). Genome-wide association study in accessions of the mini-core collection of mungbean (Vigna radiata) from the World Vegetable Gene Bank (Taiwan). BMC Plant Biol. **20**:363. https://doi.org/10.1186/s12870-020-02579-x.

Sokolkova, A.B., Bulyntsev, S.V., Chang, P.L., Carrasquila-Garcia, N., Cook, D.R., von Wettberg, E., Vishnyakova, M.A., Nuzhdin, S.V., and Samsonova, M.G. (2021). The search for agroislands in the chickpea genome. Biophysics **66**:395–400. https://doi.org/10.1134/s0006350921030192.

Song, J.M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D., Wang, B., Lu, S., Zhou, R., et al. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. Nat. Plants **6**:34–45. https://doi.org/10.1038/s41477-019-0577-7.

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics **24**:637–644. https://doi.org/10.1093/bioinformatics/btn013.

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics **Chapter 4**:Unit 4 10. https://doi.org/10.1002/0471250953.bi0410s25.

Tettelin, H., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., DeBoy, R.T., et al. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae implications for the microbial pan-genome. Proc. Natl. Acad. Sci. USA **102**:0506758102.

Torkamaneh, D., Lemay, M.A., and Belzile, F. (2021). The pan-genome of the cultivated soybean (PanSoy) reveals an extraordinarily conserved gene content. Plant Biotechnol. J. **19**:1852–1862. https://doi.org/10.1111/pbi.13600.

Varshney, R.K., Saxena, R.K., Upadhyaya, H.D., Khan, A.W., Yu, Y., Kim, C., Rathore, A., Kim, D., Kim, J., An, S., et al. (2017). Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. Nat. Genet. **49**:1082–1088. https://doi.org/10.1038/ng.3872.

Varshney, R.K., Pandey, M.K., Bohra, A., Singh, V.K., Thudi, M., and Saxena, R.K. (2019a). Toward the sequence-based breeding in legumes in the post-genome sequencing era. Theor. Appl. Genet. **132**:797–816. https://doi.org/10.1007/s00122-018-3252-x.

Varshney, R.K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H.D., Yang, W., Bajaj, P., Cubry, P., Rathore, A., Jian, J., et al. (2019b). Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. Nat. Genet. **51**:857–864. https://doi.org/10.1038/s41588-019-0401-3.

Varshney, R.K., Roorkiwal, M., Sun, S., Bajaj, P., Chitikineni, A., Thudi, M., Singh, N.P., Du, X., Upadhyaya, H.D., Khan, A.W., et al. (2021). A chickpea genetic variation map based on the sequencing of 3, 366 genomes. Nature **599**:622–627. https://doi.org/10.1038/s41586-021-04066-1.

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., Schatz, M.C., and Berger, B. (2017). Genome Scope: fast reference-free genome profiling from short reads. Bioinformatics **33**:2202–2204. https://doi.org/10.1093/bioinformatics/btx153.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One **9**:e112963. https://doi.org/10.1371/journal.pone.0112963.

Wang, S., Liu, S., Wang, J., Yokosho, K., Zhou, B., Yu, Y.C., Liu, Z., Frommer, W.B., Ma, J.F., Chen, L.Q., et al. (2020). Simultaneous changes in seed size, oil content and protein content driven by selection of SWEET homologues during soybean domestication. Natl. Sci. Rev. **7**:1776–1786. https://doi.org/10.1093/nsr/nwaa110.

Wu, J., Wang, L., Fu, J., Chen, J., Wei, S., Zhang, S., Zhang, J., Tang, Y., Chen, M., Zhu, J., et al. (2020a). Resequencing of 683 common bean genotypes identifies yield component trait associations across a north-south cline. Nat. Genet. **52**:118–125. https://doi.org/10.1038/s41588-019-0546-0.

Wu, X., Islam, A.S.M.F., Limpot, N., Mackasmiel, L., Mierzwa, J., Cortés, A.J., and Blair, M.W. (2020b). Genome-wide SNP identification and association mapping for seed mineral concentration in mung bean (Vigna radiata L.). Front. Genet. **11**:656. https://doi.org/10.3389/fgene.2020.00656.

Yang, Y., Li, Y., Chen, Q., Sun, Y., and Lu, Z. (2019). WGDdetector: a pipeline for detecting whole genome duplication events using the genome or transcriptome annotations. BMC Bioinf. **20**:75. https://doi.org/10.1186/s12859-019-2670-3.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS **16**:284–287. https://doi.org/10.1089/omi.2011.0118.

Zhao, J., Bayer, P.E., Ruperao, P., Saxena, R.K., Khan, A.W., Golicz, A.A., Nguyen, H.T., Batley, J., Edwards, D., and Varshney, R.K. (2020). Trait associations in the pangenome of pigeon pea (Cajanus cajan). Plant Biotechnol. J. **18**:1946–1954. https://doi.org/10.1111/pbi.13354.

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. **44**:821–824. https://doi.org/10.1038/ng.2310.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y., Ma, Y., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat. Biotechnol. **33**:408–414. https://doi.org/10.1038/nbt.3096.

Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., and Yorke, J.A. (2013). The MaSuRCA genome assembler. Bioinformatics **29**:2669–2677. https://doi.org/10.1093/bioinformatics/btt476.

Plant Communications, Volume *3*

Supplemental information

High-quality genome assembly and pan-genome studies facilitate genetic discovery in mung bean and its improvement

Changyou Liu, Yan Wang, Jianxiang Peng, Baojie Fan, Dongxu Xu, Jing Wu, Zhimin Cao, Yunqing Gao, Xueqing Wang, Shutong Li, Qiuzhu Su, Zhixiao Zhang, Shen Wang, Xingbo Wu, Qibing Shang, Huiying Shi, Yingchao Shen, Bingbing Wang, and Jing Tian

Supplemental Information for

Title: High-quality genome assembly and pan-genome studies facilitate genetic discovery in mungbean and its improvement

Changyou Liu[1#], Yan Wang[1#], Jianxiang Peng[2#], Baojie Fan[1], Dongxu Xu[3], Jing Wu[4], Zhimin Cao[1], Yunqing Gao[3], Xueqing Wang[1], Shutong Li[3], Qiuzhu Su[1], Zhixiao Zhang[1], Shen Wang[1], Xingbo Wu[5], Qibing Shang[3], Huiying Shi[1], Yingchao Shen[1], Bingbing Wang[2*], Jing Tian[1*]

[1]Institute of Cereal and Oil Crops, Hebei Academy of Agricultural and Forestry Sciences/Hebei Laboratory of Crop Genetic and Breeding, Shijiazhuang 050035, China

[2]Biobin Data Sciences Co., Ltd, Changsha 410221, China

[3]Zhangjiakou Academy of Agricultural Sciences, Zhangjiakou 075300, China

[4]Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[5]Tropical Research and Education Center, Department of Environmental Horticulture, University of Florida. 18905 SW 280th St, Homestead, FL 33031, USA

[#] These authors contributed equally.

[*]Corresponding authors: Bingbing Wang: bwang@biobin.com.cn; Jing Tian: nkytianjing@163.com

Supplemental Text

Genome assembly quality assessment

To assess the quality of the genome assembly, three different methods were used. First, Illumina PE reads were mapped to the assembled genome. The mapping rate for 127.85-fold Illumina reads was 99.66%, 98.73% were properly paired, and only 0.05% of the reads were singletons. The sequence mapping coverage was normally distributed, without heterozygosity and repeated peaks (Figure S3). The assembled sequence covered the whole genome, and there was no heterozygous region assembled into two haplotypes or the repeated region collapsed into one sequence. Second, by using BUSCO analysis, we estimated the genome completeness to be 98.02% (Table 1; Figure S4). Finally, the LAI, which exploited the completeness of the assembled LTR as an indicator of the quality of the whole genome assembly, was utilized. The LAI score of the assembly was 15.67 (Table 1), which was of reference genome level according to developer classification standards (Ou *et al.*, 2018b).

Duplications contributing to the expanded gene family

Many expanded genes resulted from tandem duplication. For instance, among the 41 genes annotated as an expanded family enriched in the isoflavonoid synthesis pathway, 35 (85%) were clustered into seven clusters (Data S3). Among them, six genes encoding isoflavone methyltransferase were clustered within a 50 kb region on Chr. 4 (51,555,368-51,607,654), of whose gene structure was similar. Three of these genes were expressed in flowers according to the RNA-seq data (Figure S7). Moreover, there were 12 genes annotated as isoflavonoid malonyl transferase or phenolic glucoside malonyl transferase clustered in an 80 kb region immediately downstream of the previous methyltransferase cluster. These examples indicated that tandem duplication played an important role in gene family expansion in mungbean.

Genome duplication can also lead to gene family expansion. As shown in Figure 1A, two large blocks on pseudomolecules Chr. 7 and Chr. 11, corresponding to approximately 5.2 Mb and 3.4 Mb, respectively, showed good collinearity. There is another example involved the two blocks of 4.3 Mb and 3.2 Mb between Chr. 2 and Chr. 3. There were a total of 39 synteny blocks over 1 Mb in size, containing thousands of duplicated genes.

Divergence of Legumes

Phylogenetic analysis revealed that the Phaseoleae and other species in the Papilionoideae family diverged ~50.51 million years ago (Mya). Three species from the genera *Vigna*, namely, mungbean (*Vigna radiata*), adzuki bean (*Vigna angularis*), and cowpea (*Vigna unguiculata*), were closely related. The divergence of mungbean from adzuki bean, cowpea, and soybean occurred ~5.7 Mya, ~7.5 Mya, and ~20.2 Mya, respectively (Figure S5), the findings of which were similar to those a previous study on the divergence time between mungbean and soybean (Kang *et al.*, 2014; Lavin *et al.*, 2005) but different from the previously estimated divergence time (~8 Mya) between mungbean and adzuki bean (Kang *et al.*, 2015).

Distribution of mungbean SNPs

A total of 9.89 billion PE clean reads and 1.48 Tb of sequence data were generated by sequencing 217 mungbean accessions. When mapping the sequence reads to the Vrad_JL7 genome, we found that 99.11% of reads from each accession could be mapped to and covered 93.36% of the Vrad_JL7 genome on average (Table S4). A total of 2,229,343 high-quality SNPs were identified among these accessions, 13.85% of which were distributed in the gene region rather than the intergenic region. For SNPs located in the gene regions, most were located in introns (~11.25% of the total proportion of SNPs), and only a small portion were present in the exons of the protein-coding region (~2.60%). We annotated 61,507 nonsynonymous SNPs, 49,828

synonymous SNPs, and 2,603 SNPs that cause stop or start codon changes. The distribution of SNPs on each chromosome of the genome is correlated with chromosome length and gene density. Chr. 11 (95,596) and Chr. 3 (101,950) had the fewest, and Chr. 7 (291,312), Chr. 4 (264,479) and Chr. 6 (261,463) had the most (Table S5).

Gene numbers of different accessions in the mungbean pan-genome

The gene numbers of the three categories were significantly different from each other, with CBLs having the most genes (40,971), CLRs having the second most (40,862) and NCLs having the fewest (40,787). The trend seemed opposite to the trend in the tomato pan genome (Gao *et al.*, 2019). Two wild accessions (TC1966 and ACC41) of NCLs from Madagascar and Australia were outliers, with only 39,573 and 38,862 genes, respectively. Their morphologies were also quite distinct from those of other accessions. Twenty-eight genes unique to the two wild accessions were identified, including those encoding a LRR receptor-like serine threonine-protein kinase (Pang23039), chalcone-flavanone isomerase (Pang46397), and galacturonosyltransferase (Pang24618), indicating that wild mungbean accessions are good sources for improving important traits, such as disease resistance and flavanone content.

Distribution of STAs in the mungbean genome

As shown in Figure 5, the distribution of the STAs identified in this study was not even; Chr. 4 had the most STAs (1,431 STAs from 24 traits), and Chr. 9 had the fewest, and only 24 STAs were found from five traits. A total of 1,894 STAs were in intergenic regions, 299 in the 5’ UTR or 3’ UTR of the genes, and 1,224 in at least one gene region (exonic, intronic, start codon, intron splice sequences). A total of 248 genes (270 transcripts) contained STAs (Data S6). For different types of traits, a total of 1,384 STAs were identified for seven yield-related traits, of which 596 were in the

gene region, causing 89 nonsynonymous mutations in transcripts. The numbers of total STAs and genic and nonsynonymous STAs for eight plant architecture-related traits were 655, 232, and 95, respectively; for seven colour-related traits, there were 433, 139, and 58, respectively; for four quality-related traits, there were 274, 101, and 47, respectively; for two biotic stress resistance traits, there were 610, 190, and 45, respectively; and for three other important agronomic traits, there were 131, 56, and 19, respectively.

Candidate genes associated with yield-related traits

Among all traits, BRN had the most STAs (1,026) identified, 947 of which were located on Chr. 4: 12-13 Mb. The same region was identified in multiple environments and by multiple methods (SNP-based and gene PAV-based) (Figure S16). Thirteen STAs and five GPTAs were consistent across different environments, thus narrowing the region to 180 kb, with GPTAs in the middle surrounded by STAs (Figure S17). Twenty-three genes with possible functions in disease resistance, sugar metabolism, DNA repair, etc., were found in this region, five of which (jg4298, jg4300-jg4303) revealed PAV events. The functions of the five genes were disease resistance related, hexokinase or unknown. The presence/absence of these five genes in different mungbean accessions was strongly associated with BRN, indicating that one of them could regulate BRN in mungbean (Figure S18). On Chr. 10, seven STAs were mapped to the coding region of gene jg30665, a homologue of transcription factor TCP2, which was found to play important roles in regulating the morphogenesis of shoot organs and ovule development in Arabidopsis (Wei *et al.*, 2015).

For other yield-related traits, 197 STAs and 16 GPTAs were identified for PDL, PDW or SD100WT, 30 of them (23 STAs and 7 GPTAs) were the same for at least 2 traits, and 7 STAs were the same for all three yield traits. One STA associated with PDL (7_27965615_A_C) was located at the promoter region of jg22573, which was

annotated as a NRT1/PTR FAMILY 2.13 like protein. The average PDL of accessions with the CC allele at this site was significantly shorter than that of accessions with the AA allele (Figure S19). The Arabidopsis homologue AT1G69870 was expressed mainly in the pods, siliques and petals, indicating that this gene is a good candidate for pod development in mungbean. For YPPL, several consistent STAs were identified on region Chr. 4 (29,263,225-29,697,203) containing 31 candidate genes, which seemed to belong to a single LD block. Among them, jg5489 was annotated as the WUSCHEL-related homeobox 3 gene. The rice homologue gene (WOX3) is related to growth and development (Dai *et al.*, 2007), and the Norway spruce homologue gene (PaWOX3) regulates lateral organs (Alvarez *et al.*, 2015).

Supplemental Figures

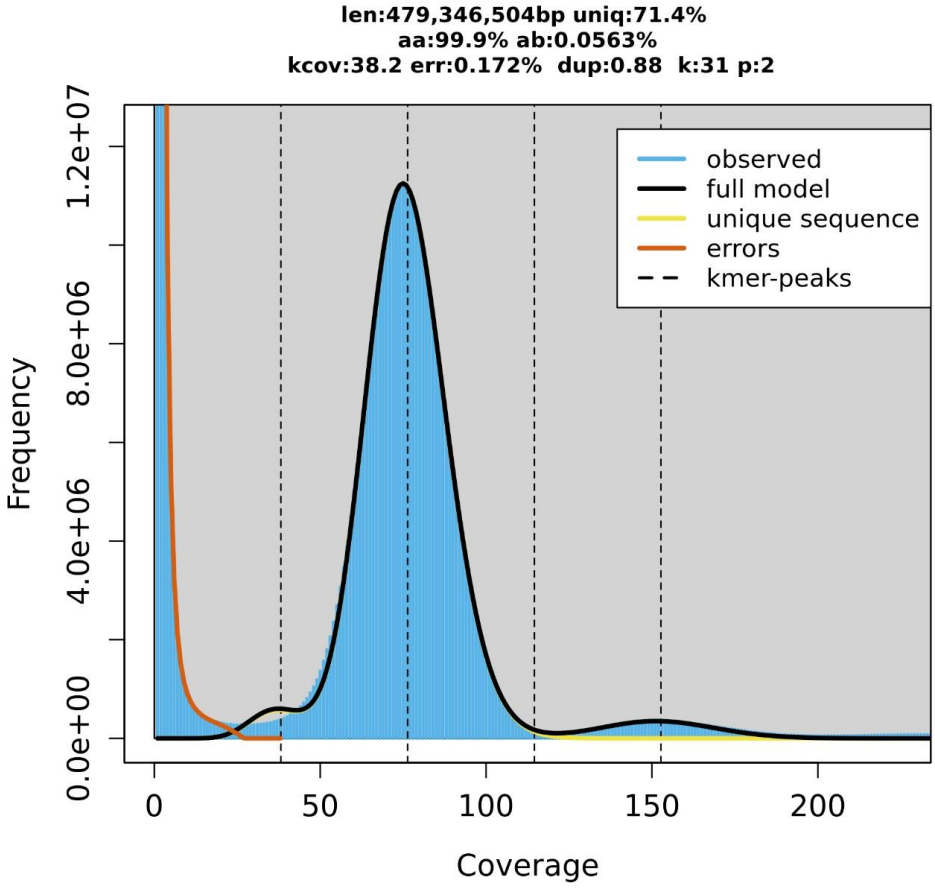
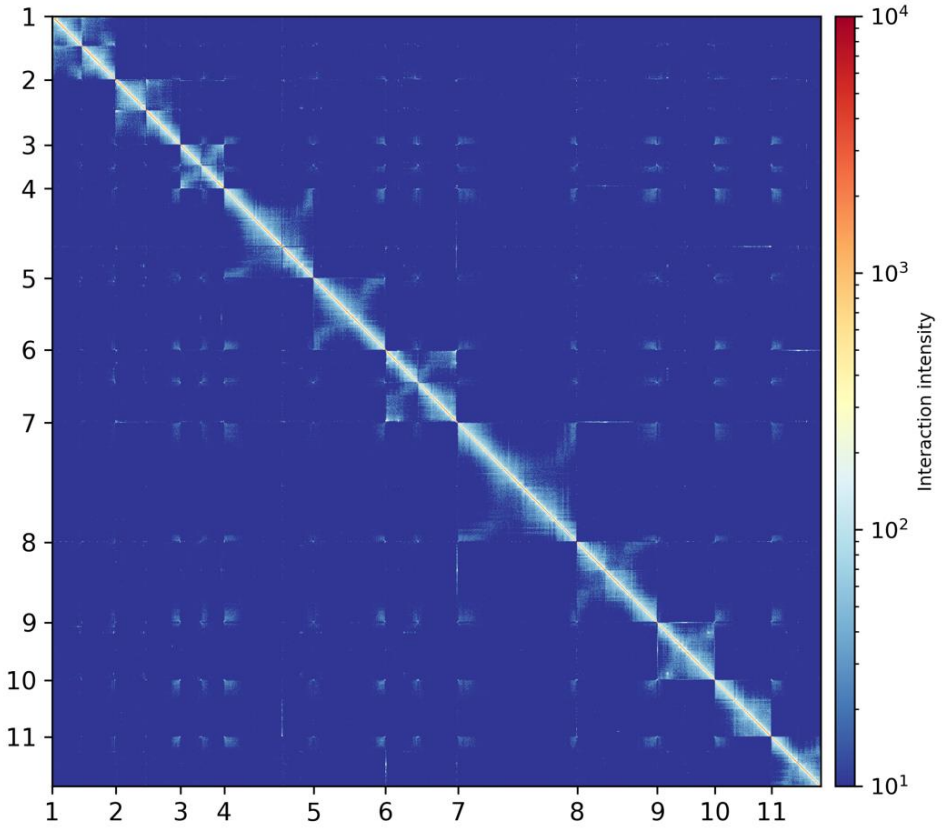


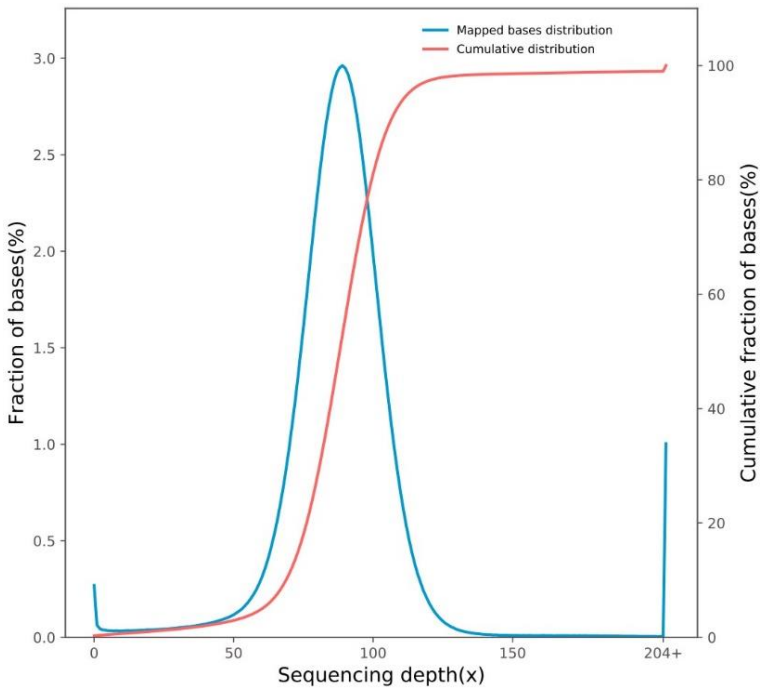
Figure S1. Genome survey of mungbean Jilv7. The k-mer frequencies were generated from raw read data to estimate the genome size, abundance of repetitive elements and rate of heterozygosity via GenomeScope (v2.0) (Vurture *et al.*, 2017).



143

144 **Figure S2.** Hi-C heatmap of chromosome interactions.

145



146

147 **Figure S3.** Coverage depth of the Illumina short reads mapped to the Vrad_JL7

148 genome.

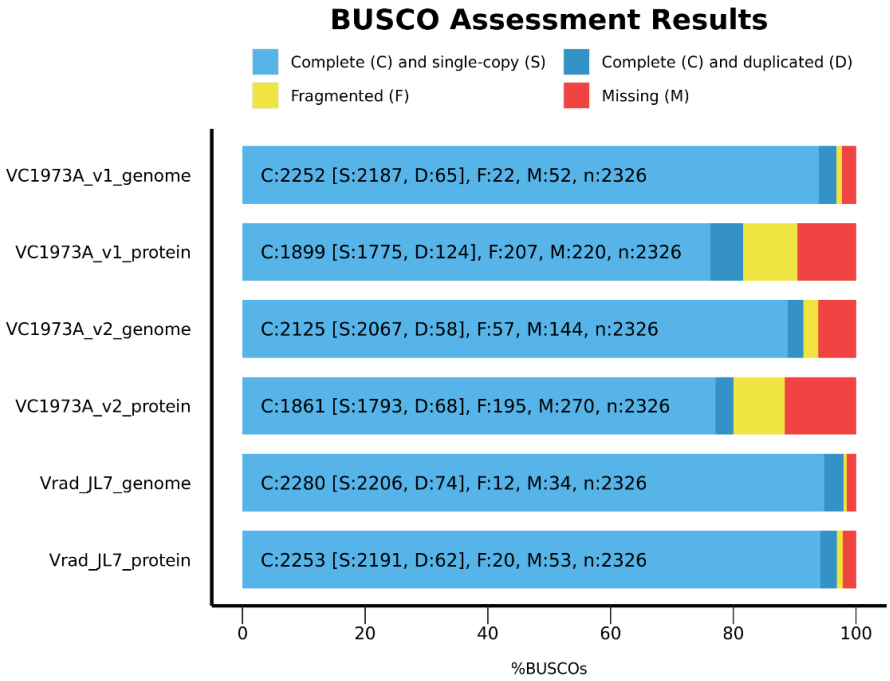


Figure S4. BUSCO comparison of the genome and protein sequence data of Vrad_JL7, VC1973A version 1 (VC1973A_v1), and VC1973A version 2 (VC1973A_v2).

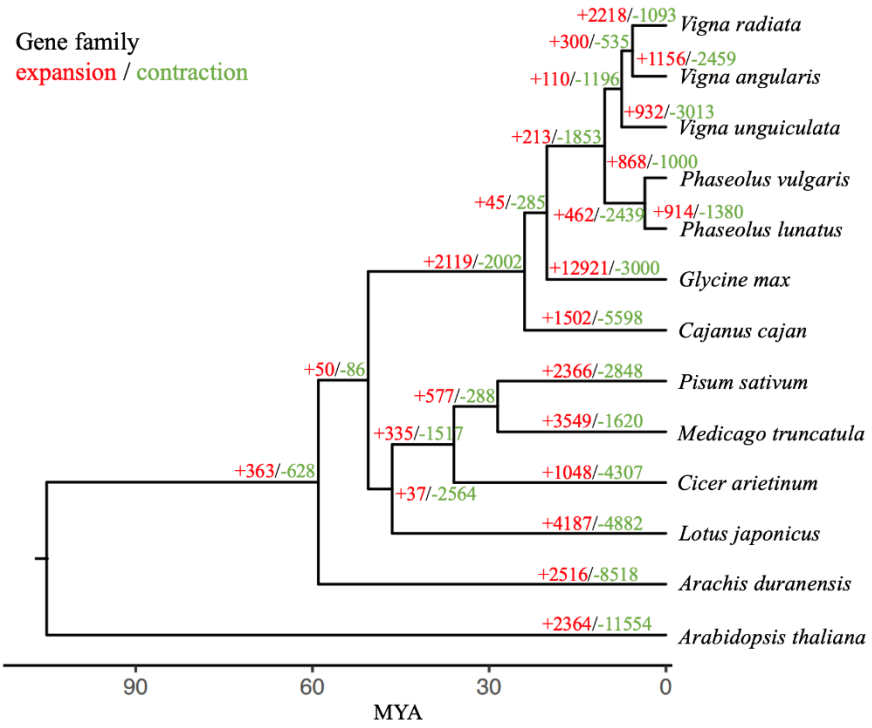


Figure S5. Phylogenetic tree comprised mungbean and 12 other species. The number on each branch indicates the number of genes within the expanded (red)/contracted (green) gene families in each plant species.

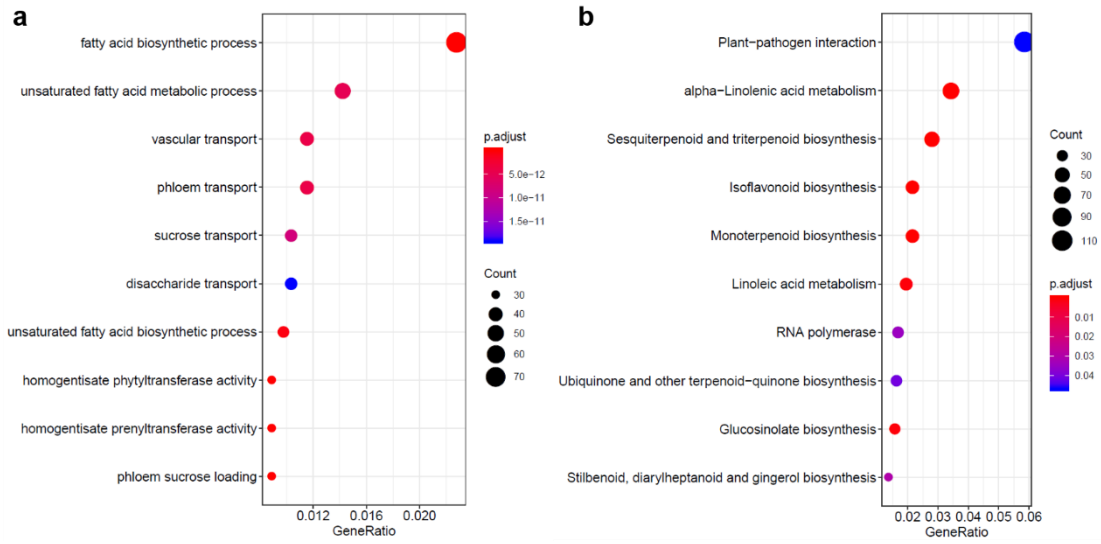


Figure S6. GO (a) and KEGG pathway (b) enrichment of expanded gene families in mungbean.

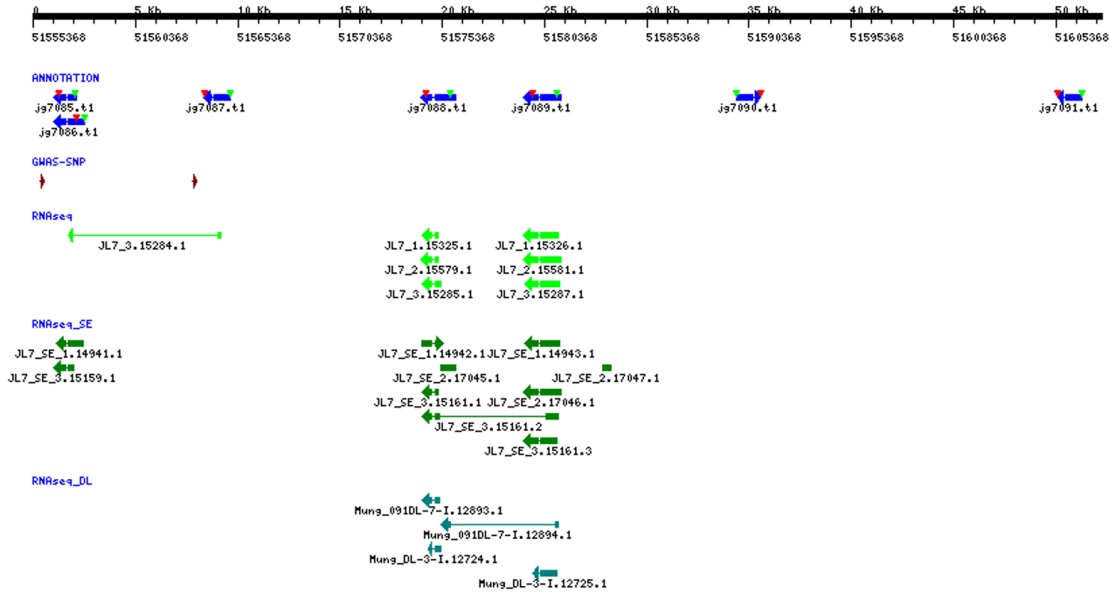


Figure S7. Gene structure of the 50 kb region of Chr. 4 (51,555,368-51,607,654) in Vrad_JL7 and its supporting transcriptomic evidence. The left ANNOTATION track represents annotated genes. The RNAseq, RNAseq_SE and RNAseq_DL tracks represent transcripts from different tissues. The green and red triangles represent the start and end positions), respectively, of the ORF.

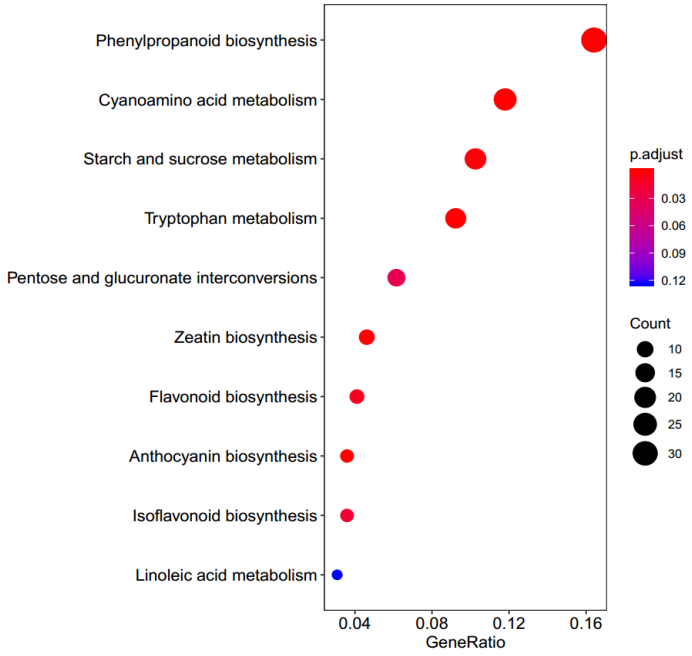


Figure S8. KEGG pathway enrichment of contracted gene families in mungbean.

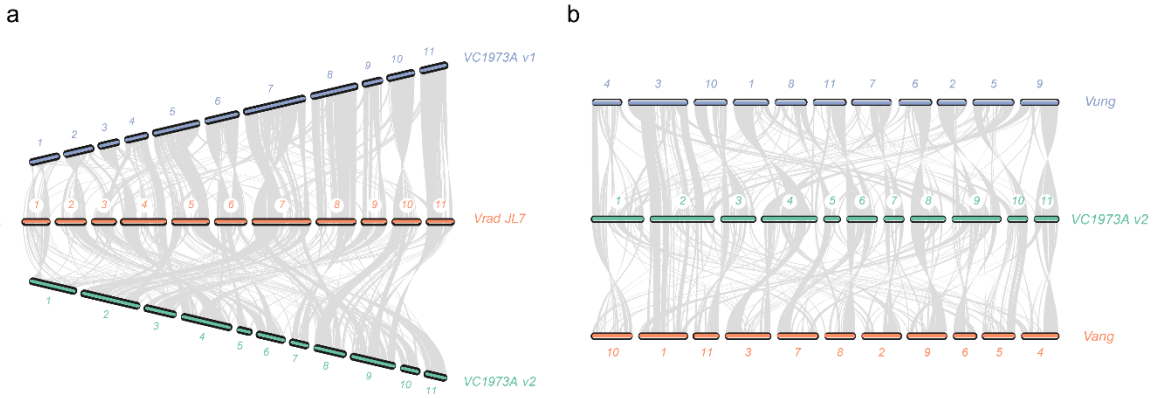


Figure S9. a Chromosome collinearity between Vrad_JL7 and VC1973A (version 1 and version 2). b Chromosome collinearity between VC1973A version 2 and adzuki bean (*Vang*) and cowpea (*Vung*).

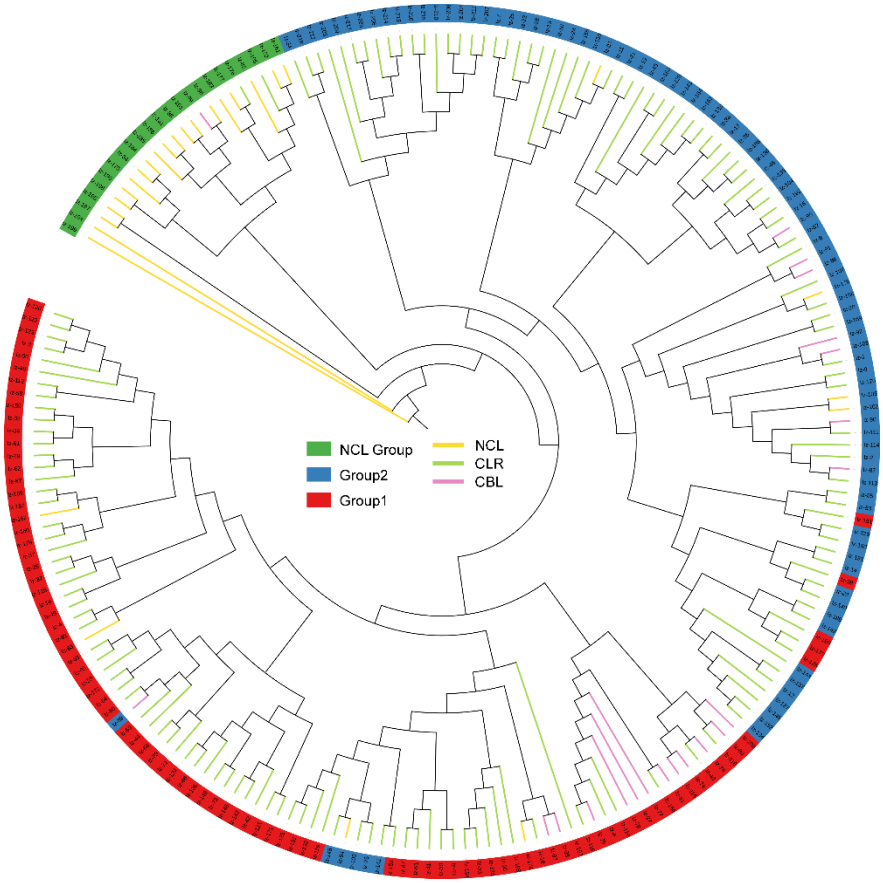


Figure S10. Hierarchical clustering results of 217 mungbean accessions based on the PAV gene.

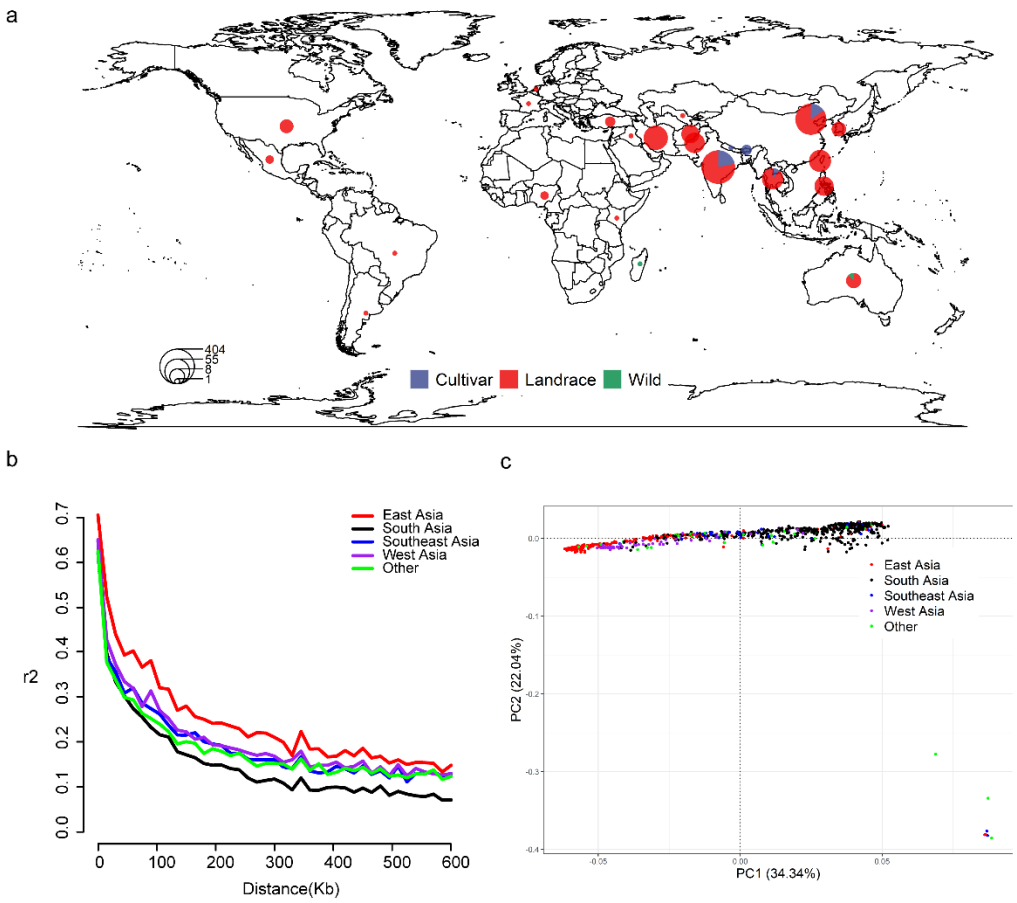
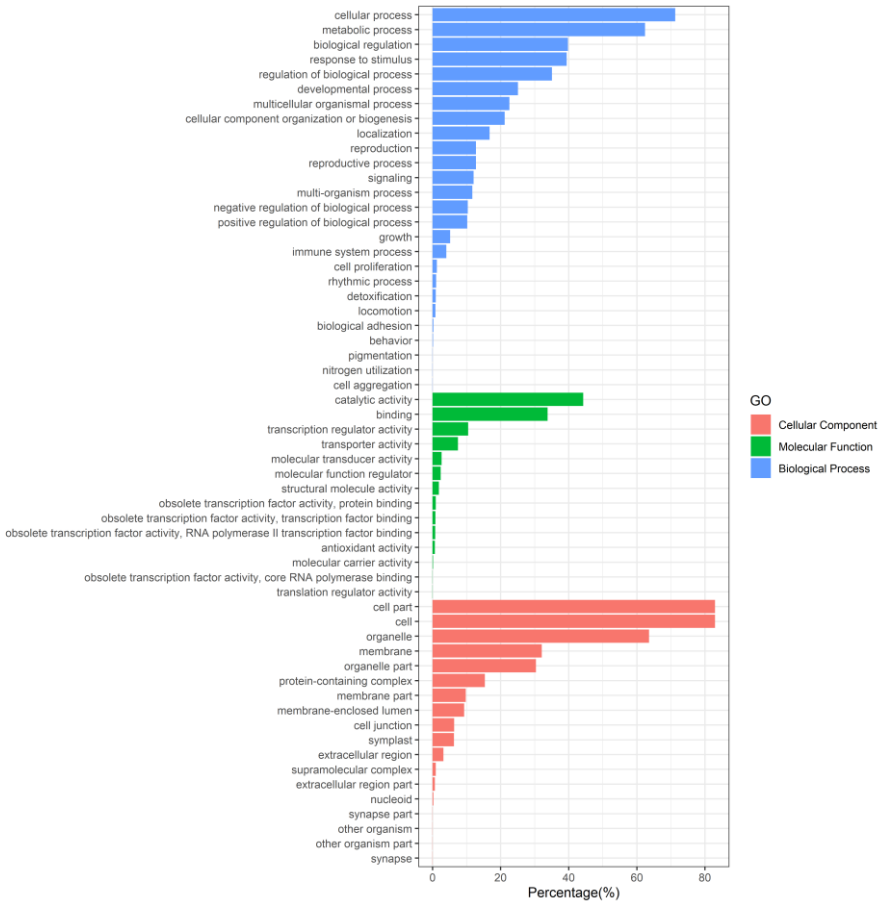
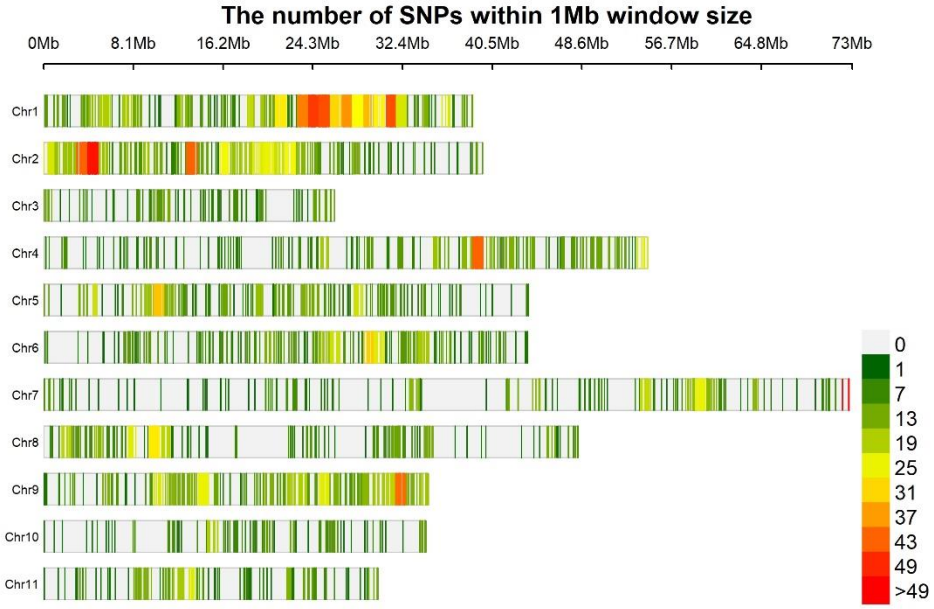


Figure S11. Information on 750 mungbean germplasm accessions from around the world. a. Geographical distribution; b. LD decay of different subpopulations; c. PCA.



182

183 **Figure S12.** GO annotations of the core genes of the mungbean pan-genome.



184

Figure S13. Density distribution of variable genes in the pan-genome of mungbean across 11 chromosomes.

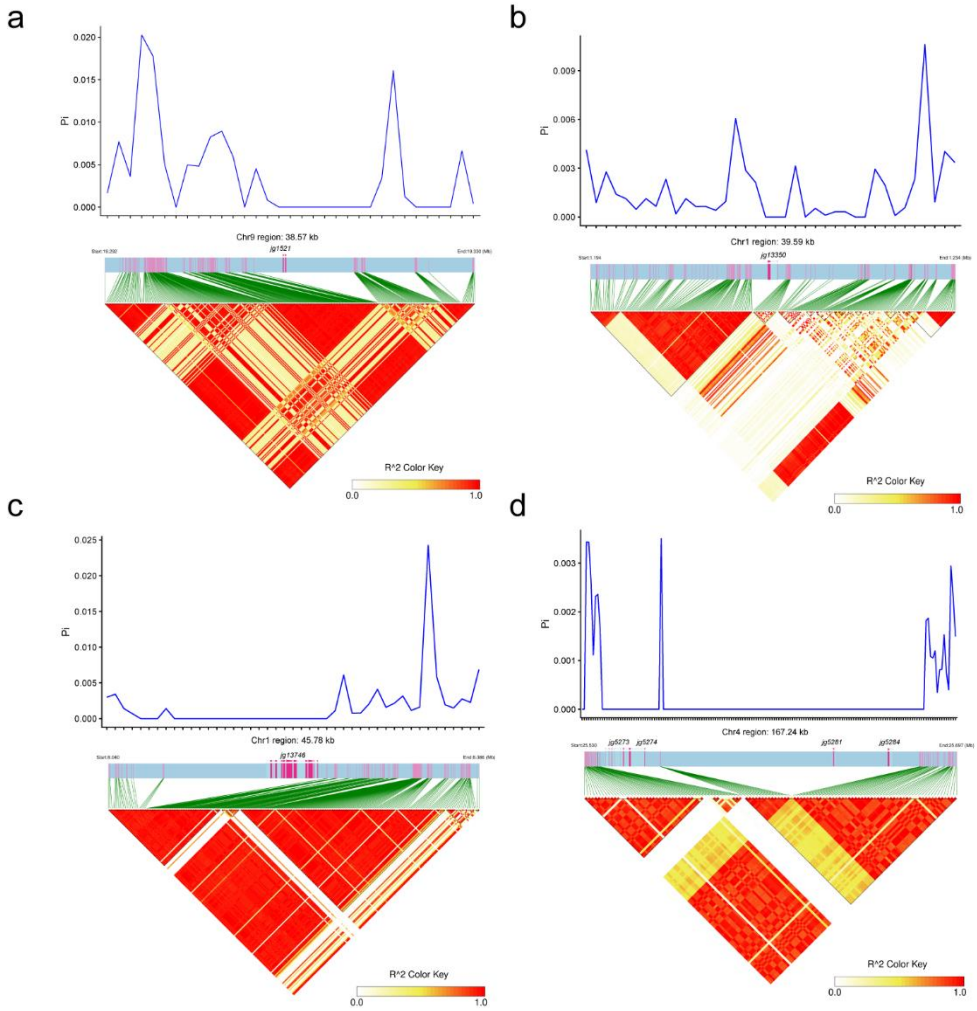
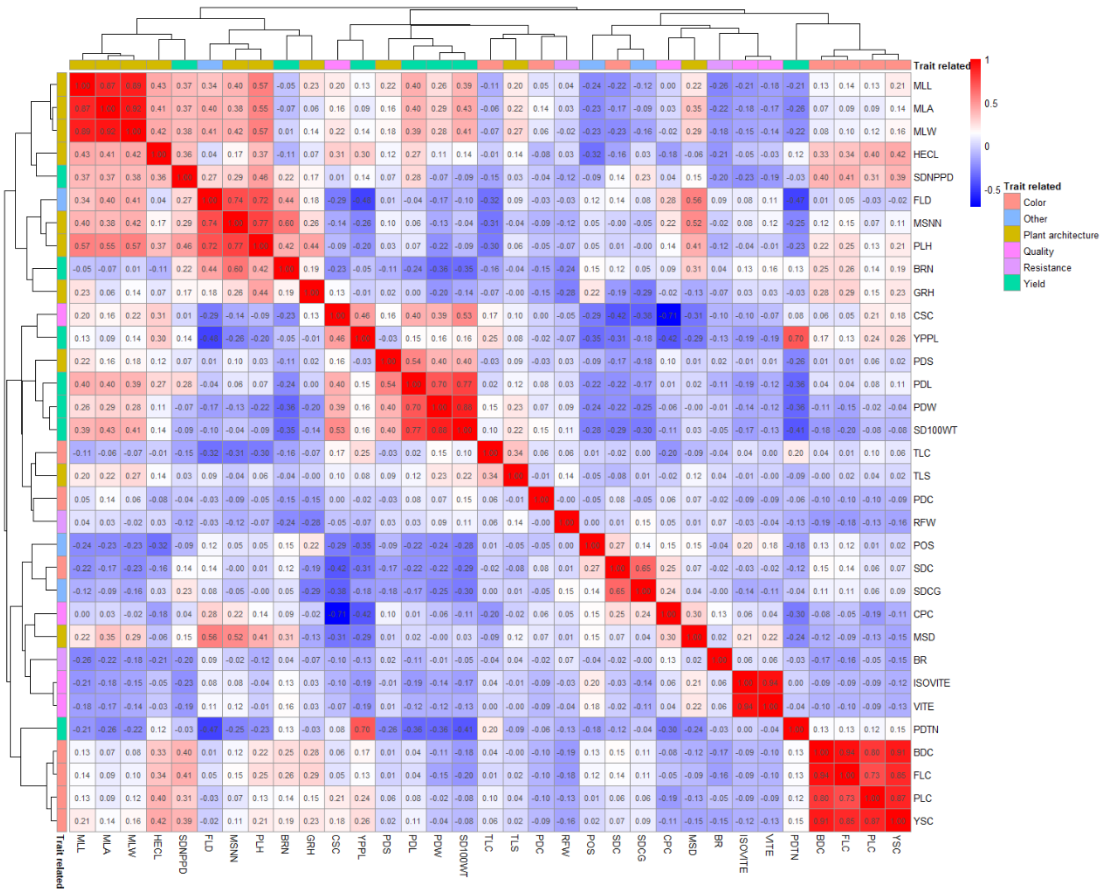


Figure S14. LD and genetic diversity around PAVs associated with flowering regulation during adaptation. Included are seven genes from the reference genome, namely, (a) jg1521; (b) jg13350; (c)jg13746; (d) jg5273, jg5274, jg5281 and jg5284.

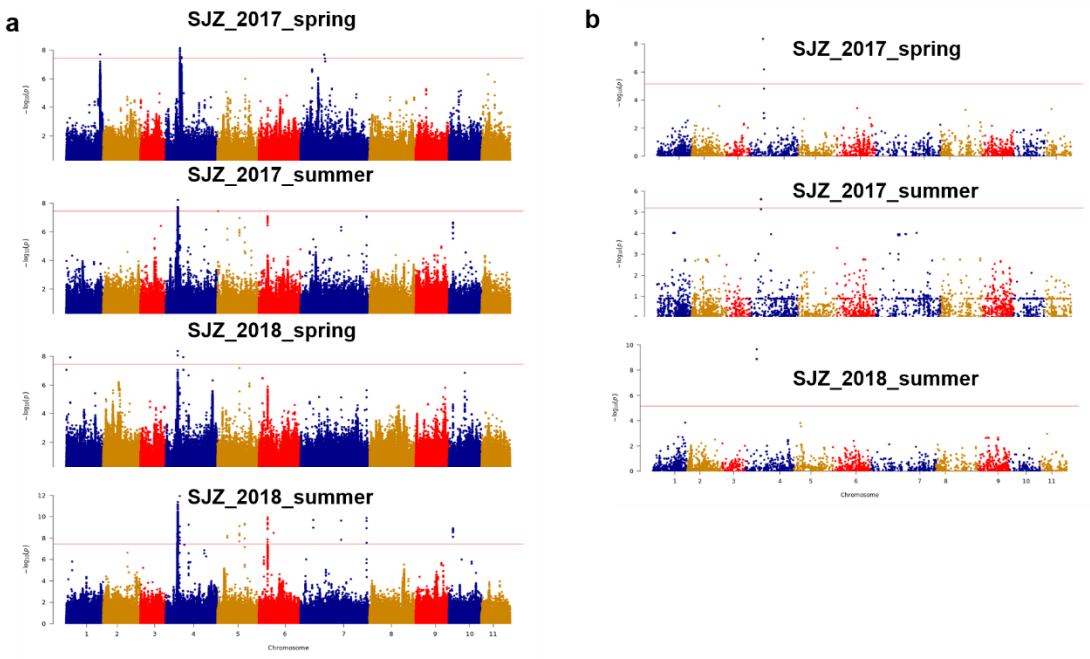


193

194 **Figure S15.** Heatmap of average phenotypic correlations in all the tested

195 environments.

196



197

Figure S16. An example region (Chr. 4: 12~13 Mb) corresponding to the BRN trait that was identified in multiple environments and by (a) a SNP GWAS and (b) a gene PAV GWAS.

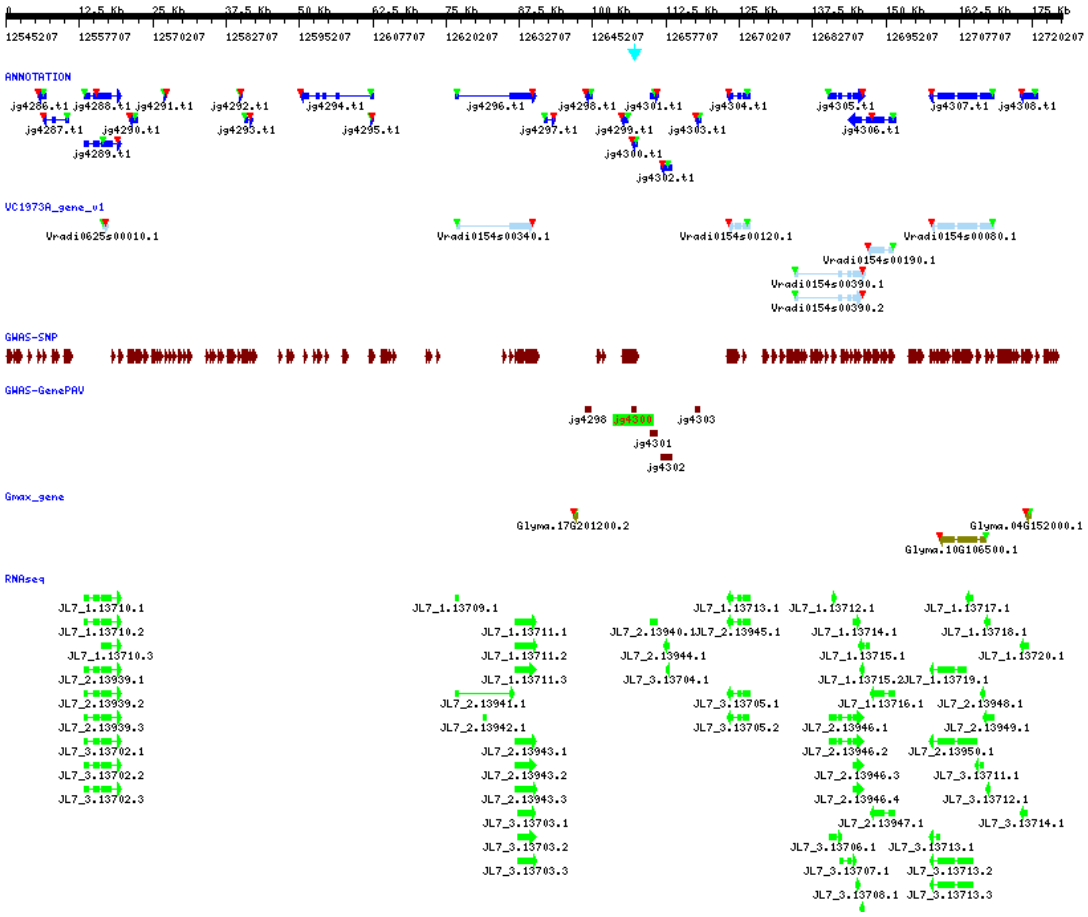


Figure S17. STA and GPTA events for the BRN trait in a 180 kb region. In the track name shown on the left, ANNOTATION stands for the Vrad_JL7 predicted gene. VC1973A_gene_v1 represents the VC1973_v1 predicted gene. The GWAS-SNPs represent STAs, and each arrow represents an STA. GWAS-GenePAV represents the GPTA genes, Gmax_gene represents the corresponding soybean homologous genes, and RNAseq represents the transcript evidence supporting the Vrad_JL7 gene. The green and red triangles represent the start and end positions, respectively, of the ORF.

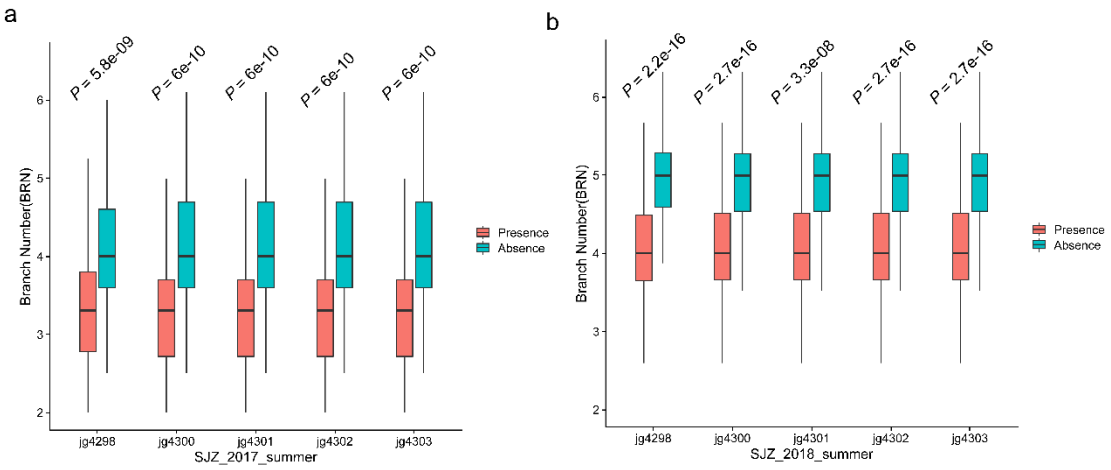


Figure S18. Relationships between the presence/absence of five genes and BRN in different environments. The *P* value is the significance level calculated according to a t test.

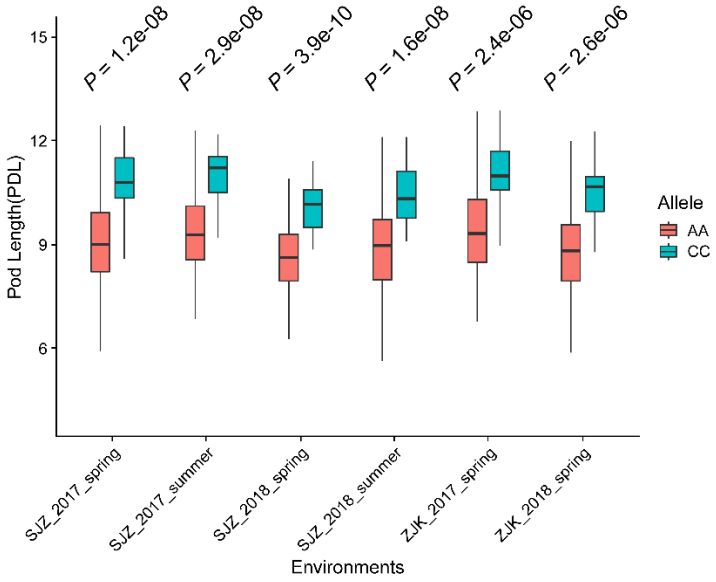


Figure S19. Relationships between different alleles of 7_27965615_A_C and PDL in different environments. The *P* value is the significance level calculated according to a t test.

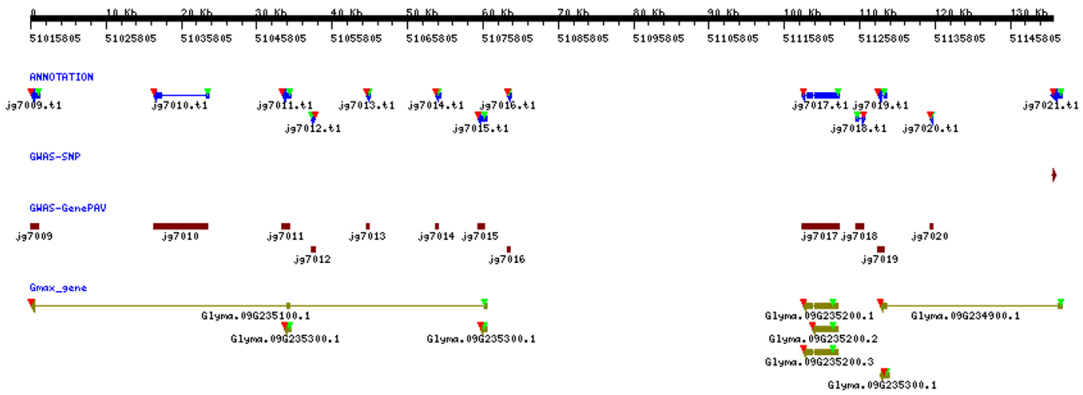


Figure S20. Gene PAV for colour-related traits and its corresponding soybean homologues. In the track name shown on the left, ANNOTATION stands for the Vrad_JL7 predicted gene. GWAS-SNPs represent STAs, GWAS-GenePAV represents the GPTA gene, and Gmax_gene represents the corresponding soybean homologous gene. The green and red triangles represent the start and end positions, respectively, of the ORF.

Supplemental Tables

Table S1 Classification and annotation of repetitive sequences in Vrad_JL7 genome.

TE Classification		Copies	Length (bp)	Percentage of genome (%)
Retrotransposon	Total	249623	151659937	31.91
	LTR/Copia	118364	65,468,906	13.77
	LTR/Gypsy	101,662	79,136,575	16.65
	LTR/Unknown	44,433	11,413,863	2.40
	LTR/Cassandra	6,553	1,085,282	0.23
	LINE/L1	2409	818,921	0.17
	LINE/RTE-BovB	970	188,195	0.04
	LINE/I-Jockey	158	62,309	0.01
	SINE	565	96,361	0.02
DNA transposon	Total	45187	20210548	4.25
	CMC-EnSpm	18,772	11,296,708	2.38
	MULE-MuDR	15,496	5,814,773	1.22
	hAT-Ac	6,610	2,118,849	0.45
	hAT-Tag1	1,561	401,633	0.08
	PIF-Harbinger	1,317	519,528	0.11
	hAT	1,096	241,623	0.05
	TcMar-Mariner	801	133,277	0.03
	hAT-Tip100	652	196,714	0.04
	Crypton-H	346	100,356	0.02
	IS3EU	295	57,010	0.01
	hAT-Charlie	195	45,481	0.01
	TcMar-Stowaway	188	23,927	0.01
Unknown		250881	66,173,679	13.92
Total		718,854	254,081,323	53.45

Table S2 The number of non-coding RNAs in Vrad_JL7 genome.

Type	Number
tRNA	681
catalytic intron	59
small nucleolar RNA	410
spliceosomal RNA	85
microRNA	147
18S rRNA	123
28S rRNA	124
5S rRNA	4,078
5.8S rRNA	123

Table S3 Gene family statistics of 13 species.

Species Name	Number of genes	Number of genes in orthogroups	Number of orthogroups containing species	Number of species-specific orthogroups	Number of genes in species-specific orthogroups
Vigna radiata (Vrad)	40,125	35,059	19,335	1,532	5,482
Arachis duranensis (Adur)	34,553	32,969	16,415	568	4,966
Arabidopsis thaliana (Artha)	27,628	24,706	14,294	981	4,734
Cicer arietinum (Cari)	24,962	24,591	16,036	142	824
Cajanus cajan (Ccaj)	29,199	28,501	16,936	157	478
Glycine max (Gmax)	55,897	49,377	18,622	578	1,795
Lotus japonicus (Ljap)	39,734	34,093	18,396	1,622	5,194
Medicago truncatula (Mtru)	50,444	43,509	19,007	1,567	7,948
Phaseolus lunatus (Plun)	43,997	41,861	17,645	397	6,708
Pisum sativum (Psat)	44,756	39,036	18,252	1,533	9,649
Phaseolus vulgaris (Pvul)	28,134	27,203	17,428	51	152
Vigna angularis (Vang)	33,860	31,866	17,555	532	2,921
Vigna unguiculata (Vung)	28,314	27,874	16,524	79	296

Table S4 The 217 mungbean accessions and their sequenced information in our study.

No.	Sample ID	Accession	Country	China Region	Province PRC	Categories	Group in Phylogenetic Tree	Clean Reads	Clean Base (Gb)	GC(%)	Q20(%)	Q30(%)	Mapped Reads	Mapped Rate(%)	Coverage Rate(%)	Mean Depth(X)
1	lz-1	Lvfeng 2	China	North	Heilongjiang	CLR	Group2	43,466,324	6.52	35	95	93	43,392,425	99.16	92.88	11.60
2	lz-2	Lvzi 5	China	North	Heilongjiang	CLR	Group2	50,785,548	7.62	34	96	94	49,908,099	97.66	93.13	13.48
3	lz-3	Lvzi 12	China	North	Heilongjiang	CLR	Group1	45,038,980	6.76	34	97	95	45,092,931	99.71	95.26	12.22
4	lz-4	Lvzi 17	China	North	Heilongjiang	CLR	Group1	43,475,982	6.52	34	97	95	43,540,326	99.64	94.72	11.86
5	lz-5	L2L173	China	North	Liaoning	CLR	--	44,509,506	6.68	36	97	94	44,307,746	98.95	93.41	12.02
6	lz-6	L2L177	China	North	Liaoning	CLR	Group1	45,178,974	6.78	36	97	95	44,983,620	99.16	95.10	12.22
7	lz-7	C001	China	North	Neimenggu	CLR	Group2	45,438,034	6.82	36	97	95	45,262,661	98.99	92.81	12.07
8	lz-8	C028	China	North	Neimenggu	CLR	Group2	41,956,054	6.29	35	97	95	41,824,620	99.10	92.91	11.29
9	lz-9	C029	China	North	Neimenggu	CLR	Group2	42,933,762	6.44	35	97	93	42,808,273	99.09	93.05	11.67
10	lz-10	C037	China	North	Neimenggu	CLR	Group2	45,022,496	6.75	35	97	94	44,905,646	99.14	93.18	11.98
11	lz-11	C046	China	North	Neimenggu	CLR	Group2	41,629,482	6.24	35	97	94	41,647,440	99.39	92.73	11.44
12	lz-12	C0072	China	North	Hebei	CLR	Group2	39,696,202	5.95	35	97	94	39,763,766	99.49	92.78	10.73
13	lz-13	C0427	China	North	Shanxi	CLR	Group2	46,107,244	6.92	35	96	93	46,204,495	99.55	92.58	12.58
14	lz-14	C0817	China	North	Liaoning	CLR	Group2	44,377,156	6.66	35	97	94	44,261,444	99.10	92.87	11.87
15	lz-15	C0967	China	North	Shandong	CLR	Group2	42,121,770	6.32	35	97	94	42,198,861	99.52	92.68	11.46
16	lz-16	C1018	China	North	Shandong	CLR	Group1	47,286,220	7.09	35	97	94	47,390,349	99.63	93.43	12.98
17	lz-17	C1065	China	North	Shandong	CLR	Group2	45,486,440	6.82	35	98	95	45,212,427	98.80	92.73	12.19
18	lz-18	C1257	China	North	Shandong	CLR	Group1	40,420,486	6.06	35	97	94	40,414,331	99.35	93.02	10.90
19	lz-19	Resource 6	China	North	Shanxi	CLR	Group2	44,899,414	6.73	35	97	94	44,848,881	99.22	92.72	12.24
20	lz-20	Resource 8	China	North	Shanxi	CLR	Group2	43,503,312	6.53	36	97	95	43,594,623	99.55	92.56	11.53
21	lz-21	Resource 9	China	North	Shanxi	CLR	Group2	45,014,248	6.75	36	96	93	44,859,661	98.97	92.49	11.99
22	lz-22	Resource 10	China	North	Shanxi	CLR	Group2	45,512,888	6.83	36	96	93	45,352,830	99.01	92.45	12.09
23	lz-23	Resource 11	China	North	Shanxi	CLR	Group2	46,162,346	6.92	35	97	95	46,083,206	99.23	93.35	12.74
24	lz-24	Resource 62	China	North	Shanxi	CLR	Group2	44,650,416	6.7	35	96	93	44,495,021	99.02	92.39	11.97
25	lz-25	Resource 66	China	North	Shanxi	CLR	Group2	44,472,598	6.67	35	97	94	44,329,783	99.06	92.89	12.15
26	lz-26	Yulv 1	China	North	Shaanxi	CLR	Group1	50,798,050	7.62	37	95	93	50,469,116	98.76	93.27	13.08
27	lz-27	Zhonglv 4	China	North	Beijing	CBL	Group1	47,003,870	7.05	38	96	93	46,642,987	98.63	94.06	12.70
28	lz-28	Lvzi 5	China	North	Shaanxi	CLR	Group2	40,395,574	6.06	35	96	94	40,388,184	99.36	92.27	10.99
29	lz-29	C1451	China	North	Henan	CLR	Group1	46,247,228	6.94	35	96	94	46,323,026	99.64	93.95	12.69
30	lz-30	C1463	China	North	Henan	CLR	Group1	49,367,470	7.41	34	96	94	49,415,669	99.54	93.55	13.52
31	lz-32	C1467	China	North	Henan	CLR	Group1	50,723,580	7.61	37	96	94	50,681,680	99.35	93.74	13.07
32	lz-33	C1475	China	North	Henan	CLR	Group1	47,061,114	7.06	34	97	94	47,133,081	99.57	93.89	12.96
33	lz-34	C1498	China	North	Henan	CLR	Group1	48,296,466	7.24	36	97	95	47,784,426	98.41	94.51	13.06
34	lz-35	C1509	China	North	Henan	CLR	Group1	43,469,472	6.52	36	97	94	43,211,041	98.83	93.92	11.75
35	lz-36	C1518	China	North	Henan	CLR	Group2	46,539,950	6.98	35	97	95	46,312,567	98.91	92.97	12.53
36	lz-37	C1527	China	North	Henan	CLR	Group1	42,995,068	6.45	36	97	94	42,893,510	99.17	93.56	11.47
37	lz-38	C1551	China	North	Henan	CLR	Group2	42,780,744	6.42	35	97	94	42,632,480	99.01	92.76	11.64
38	lz-39	C1556	China	North	Henan	CLR	Group2	45,838,458	6.88	37	97	94	44,596,447	96.59	92.96	11.98
39	lz-40	C1559	China	North	Henan	CLR	Group1	40,702,758	6.11	35	97	94	40,724,122	99.90	97.62	11.03
40	lz-41	C1567	China	North	Henan	CLR	Group1	41,264,612	6.19	37	97	94	40,887,267	98.38	92.78	10.73
41	lz-42	C1569	China	North	Henan	CLR	Group1	42,785,238	6.42	35	96	94	42,657,769	99.10	93.27	11.75
42	lz-43	C1580	China	North	Henan	CLR	Group1	40,531,242	6.08	35	97	96	40,570,530	99.45	92.91	11.08
43	lz-44	C1596	China	North	Henan	CLR	Group1	46,807,996	7.02	35	96	94	46,782,452	99.36	93.84	12.56
44	lz-45	C0882	China	North	Shandong	CLR	Group2	44,987,314	6.75	35	97	95	44,463,799	98.15	92.97	12.17
45	lz-46	C0886	China	North	Shandong	CLR	Group2	44,632,032	6.69	36	97	93	44,386,241	98.77	92.66	11.92
46	lz-47	C0891	China	North	Shandong	CLR	Group1	42,799,746	6.42	37	96	93	42,570,475	98.73	93.22	11.31
47	lz-48	C0907	China	North	Shandong	CLR	Group1	41,929,644	6.29	35	97	96	41,987,905	99.62	94.00	11.40
48	lz-49	C1009	China	North	Shandong	CLR	Group1	45,021,412	6.75	35	97	93	44,906,786	99.18	94.14	12.25
49	lz-50	C1052	China	North	Shandong	CLR	Group1	42,387,000	6.36	35	97	95	42,058,235	98.58	92.65	11.50
50	lz-51	C1061	China	North	Shandong	CLR	Group1	44,927,748	6.74	36	97	93	44,835,101	99.07	93.54	12.04
51	lz-52	C1084	China	North	Shandong	CLR	Group2	43,570,452	6.54	35	97	93	43,592,695	99.39	92.83	11.81
52	lz-53	AHYL2014-06	China	South	Anhui	CLR	Group2	47,548,812	7.13	35	97	94	47,573,106	99.39	92.70	12.88
53	lz-54	AHYL2014-07	China	South	Anhui	CLR	Group2	46,634,836	7	35	96	94	46,724,757	99.58	93.44	12.68
54	lz-55	AHYL2014-08	China	South	Anhui	CLR	Group2	48,383,560	7.26	35	97	94	48,166,064	98.97	92.63	13.21
55	lz-56	AHYL2014-19	China	South	Anhui	CLR	Group1	44,916,950	6.74	34	96	93	44,842,169	99.39	95.05	12.45
56	lz-57	LD109	China	South	Hubei	CLR	Group1	46,490,176	6.97	36	97	94	46,505,024	99.47	94.49	12.64
57	lz-58	LD127	China	South	Hubei	CLR	Group1	45,660,750	6.85	34	97	94	45,612,659	99.39	94.51	12.57
58	lz-59	LD128	China	South	Hubei	CLR	Group1	45,334,186	6.8	34	97	94	45,358,887	99.46	93.76	12.49
59	lz-60	LD135	China	South	Hubei	CLR	Group1	44,967,106	6.75	35	96	93	44,991,803	99.47	93.96	12.42
60	lz-61	LD142	China	South	Hubei	CLR	Group1	42,930,862	6.44	36	97	94	42,829,436	99.16	94.49	11.45
61	lz-62	LD163	China	South	Hubei	CLR	Group1	43,969,424	6.6	35	96	94	43,977,994	99.39	94.04	11.98
62	lz-63	LD175	China	South	Hubei	CLR	Group1	44,192,210	6.63	35	97	94	44,175,020	99.42	94.07	12.34
63	lz-64	LD235	China	South	Hubei	CLR	Group1	45,608,446	6.84	35	97	94	44,382,470	96.74	93.69	12.29
64	lz-65	LD251	China	South	Hubei	CLR	Group1	47,183,460	7.08	36	97	93	46,132,233	97.19	93.78	12.39
65	lz-66	LD264	China	South	Hubei	CLR	Group1	45,317,814	6.8	34	97	94	45,283,103	99.33	93.73	12.61
66	lz-67	LD274	China	South	Hubei	CLR	Group1	46,818,886	7.02	34	97	94	46,809,132	99.50	94.94	13.06
67	lz-68	Lvzi 1	China	South	Chongqing	CLR	Group1	42,904,172	6.44	34	97	95	42,958,543	99.49	93.08	11.85
68	lz-69	Lvzi 2	China	South	Chongqing	CLR	Group1	46,065,000	6.91	36	97	95	46,189,297	99.63	93.72	12.30
69	lz-70	Lvzi 3	China	South	Chongqing	CLR	Group1	47,033,276	7.05	34	97	95	47,096,672	99.54	93.79	12.99
70	lz-71	Lvzi 5	China	South	Chongqing	CLR	Group1	48,932,012	7.34	34	97	95	48,862,400	99.30	94.32	13.57
71	lz-72	Lvzi 10	China	South	Chongqing	CLR	--	47,627,124	7.14	35	96	93	47,623,382	99.39	93.46	13.03
72	lz-73	Lvdou 1	China	South	Hainan	CLR	Group1	42,032,416	6.3	35	97	94	42,117,168	99.54	95.13	11.39
73	lz-74	Lvdou 5	China	South	Hainan	CLR	--	45,629,496	6.84	35	97	95	45,549,292	99.26	93.94	12.40
74	lz-75	Jilv 2	China	North	Hebei	CBL	Group1	42,045,876	6.31	36	97	95	41,782,271	98.85	95.17	11.40
75	lz-76	Jilv 7	China	North	Hebei	CBL	Group1	45,326,608	6.8	35	97	95	45,321,625	99.79	97.29	12.67

76	lz-77	Jilv 9	China	North	Hebei	CBL	Group1	49,526,020	7.43	35	97	95	49,489,617	99.42	94.69	13.63
77	lz-78	Jilv 10	China	North	Hebei	CBL	Group1	43,673,484	6.55	35	97	96	43,156,933	98.42	94.92	11.87
78	lz-79	Jilv 11	China	North	Hebei	CBL	Group1	43,042,892	6.46	36	97	95	42,975,337	99.39	95.47	11.49
79	lz-80	Baolv 942	China	North	Hebei	CBL	Group1	40,922,226	6.14	38	97	95	40,446,125	98.20	93.40	10.51
80	lz-81	Bao 956-6	China	North	Hebei	CBL	Group1	43,528,846	6.53	35	97	95	43,551,020	99.51	95.22	11.88
81	lz-82	aoding White Pc	China	North	Hebei	CLR	Group1	43,810,862	6.57	36	97	95	43,461,534	98.80	95.60	11.80
82	lz-83	Zhonglv 5	Other	--	--	NCL	NCL Group	42,418,378	6.36	35	97	96	42,475,812	99.65	93.99	11.82
83	lz-84	Zhonglv 8	Other	--	--	NCL	NCL Group	40,638,588	6.1	35	97	95	40,709,667	99.58	93.54	11.27
84	lz-85	Zhonglv 11	Other	--	--	NCL	NCL Group	57,156,902	8.57	35	97	95	57,208,537	99.52	93.88	15.75
85	lz-86	Zhonglv 12	Other	--	--	NCL	NCL Group	46,395,580	6.96	36	97	95	46,266,125	99.02	93.33	12.47
86	lz-87	Zhonglv 14	China	North	Beijing	CBL	Group2	53,962,430	8.09	37	97	95	54,059,365	99.39	93.27	14.23
87	lz-88	Bailv 9	China	North	Liaoning	CBL	Group2	54,509,064	8.18	35	97	96	54,668,949	99.67	93.00	14.64
88	lz-89	Bailv 11	China	North	Liaoning	CBL	Group1	46,631,102	6.99	36	96	93	46,665,214	99.49	94.29	12.56
89	lz-90	Ji Lv 5	China	North	Jilin	CBL	Group2	47,951,366	7.19	36	97	96	47,772,872	99.02	93.29	12.98
90	lz-91	Ji Lv 6	China	North	Jilin	CBL	Group2	41,248,540	6.19	35	97	95	41,301,385	99.44	92.67	11.13
91	lz-92	Ji Lv 7	China	North	Jilin	CBL	Group2	59,770,998	8.97	37	96	94	59,840,665	99.43	93.30	15.71
92	lz-93	Jinlv 1	China	North	Shanxi	NCL	Group2	49,590,486	7.44	36	96	94	49,646,274	99.38	92.87	13.27
93	lz-94	Jinlv 3	China	North	Shanxi	NCL	Group1	42,658,574	6.4	35	96	94	41,871,283	97.55	93.40	11.37
94	lz-95	Jinlv 4	China	North	Shanxi	CBL	Group2	48,135,684	7.22	35	96	94	48,212,296	99.51	92.66	13.10
95	lz-96	Weilv 4	Other	--	--	NCL	NCL Group	51,227,522	7.68	35	97	95	51,333,007	99.47	92.97	13.86
96	lz-97	Weilv 7	China	North	Shandong	CBL	Group1	51,639,854	7.75	35	96	94	51,709,660	99.51	93.89	14.17
97	lz-98	Weilv 8	China	North	Shandong	CBL	Group1	50,000,816	7.5	37	96	94	50,020,921	99.19	93.02	13.46
98	lz-99	Weilv 2116	China	North	Shandong	CBL	Group2	44,666,692	6.7	35	95	92	44,653,309	99.36	93.81	12.07
99	lz-100	Weilv 2118	China	North	Shandong	CBL	--	48,026,628	7.2	35	96	94	47,937,177	99.18	93.14	12.97
100	lz-101	Weilv 9002-341	China	North	Shandong	CBL	--	49,235,476	7.39	36	96	94	49,199,569	99.17	92.98	13.15
101	lz-102	Liaolv 6	Other	--	--	NCL	Group2	42,781,128	6.42	35	96	94	25,551,068	99.23	92.39	6.97
102	lz-103	Liaolv 10	Other	--	--	NCL	Group2	41,228,160	6.18	35	96	94	41,233,253	99.40	93.03	11.14
103	lz-104	Yingge	China	North	Hebei	CLR	Group2	45,523,150	6.83	35	97	95	44,202,794	96.54	93.12	12.20
104	lz-105	Dayinggelv 985	China	North	Jilin	CLR	--	51,735,212	7.76	36	97	95	51,369,888	98.24	91.18	13.60
105	lz-106	Daxinchangming	China	North	Zhejiang	CLR	Group2	46,083,956	6.91	35	97	95	46,015,973	99.21	94.29	12.67
106	lz-107	Minglv 1	China	South	Anhui	CLR	Group2	49,496,708	7.42	35	97	95	49,556,022	99.50	93.32	13.38
107	lz-108	Nenlv 1	China	North	Heilongjiang	CBL	Group2	40,128,088	6.02	37	97	96	40,149,990	99.34	92.94	10.67
108	lz-109	Yulin	China	North	Shaanxi	CLR	Group1	42,858,576	6.43	35	97	95	42,900,981	99.56	93.62	11.41
109	lz-110	Nanyang	China	North	Henan	CLR	Group1	41,777,892	6.27	35	97	96	41,717,047	99.40	95.23	11.58
110	lz-111	Xilv 1	China	North	Shaanxi	CLR	Group2	43,018,396	6.45	36	97	95	42,998,744	99.32	93.19	11.52
111	lz-112	Qingshuihe	China	North	Neimenggu	CLR	Group1	45,737,030	6.86	35	97	95	45,712,493	99.46	94.90	12.64
112	lz-113	Henan black	China	North	Henan	CLR	Group2	50,168,524	7.53	35	97	95	50,098,852	99.30	93.34	13.74
113	lz-114	Bobai	China	South	Anhui	CLR	Group2	44,893,944	6.73	36	96	94	44,985,299	99.58	93.10	12.00
114	lz-115	Sukang 2	China	North	Jiangsu	CBL	Group1	47,154,200	7.07	35	97	95	47,185,226	99.65	95.91	13.04
115	lz-116	HBM0001	China	North	Hebei	CLR	Group1	48,531,040	7.28	35	97	94	48,552,849	99.63	95.31	13.34
116	lz-117	HBM0005	China	North	Hebei	CLR	--	46,337,342	6.95	35	97	95	46,423,706	99.55	93.90	12.73
117	lz-118	HBM0007	China	North	Hebei	CLR	Group1	40,011,412	6	35	97	95	40,068,581	99.62	94.93	11.07
118	lz-119	HBM0014	China	North	Hebei	CLR	Group1	45,133,746	6.77	36	97	95	44,338,472	97.55	93.28	12.09
119	lz-120	HBM0016	China	North	Hebei	CLR	Group1	44,461,532	6.67	36	97	96	44,384,308	99.31	95.09	11.96
120	lz-121	HBM0018	China	North	Hebei	CLR	Group1	50,396,590	7.56	38	97	95	49,381,194	97.27	93.19	13.09
121	lz-122	HBM0019	China	North	Hebei	CLR	Group1	46,215,870	6.93	35	97	95	46,257,344	99.60	95.31	12.89
122	lz-123	HBM0021	China	North	Hebei	CLR	Group1	47,413,386	7.11	36	97	95	47,473,249	99.56	95.15	12.73
123	lz-124	HBM0026	China	North	Hebei	CLR	Group1	41,047,022	6.16	35	97	96	41,130,806	99.67	94.17	11.26
124	lz-125	HBM0031	China	North	Hebei	CLR	Group1	46,051,838	6.91	36	97	95	46,202,314	99.67	92.97	12.15
125	lz-126	C0073	China	North	Hebei	CLR	Group2	43,194,186	6.48	35	97	95	43,187,336	99.39	93.29	11.71
126	lz-127	C0078	China	North	Hebei	CLR	Group2	45,732,916	6.86	36	97	95	45,824,729	99.50	93.42	12.29
127	lz-128	C0081	China	North	Hebei	CLR	Group1	41,054,800	6.16	36	97	95	40,804,348	98.79	93.46	11.13
128	lz-129	C0113	China	North	Hebei	CLR	Group2	51,428,680	7.71	39	97	95	51,525,244	99.24	92.43	13.36
129	lz-130	C0126	China	North	Hebei	CLR	Group1	41,390,672	6.21	36	97	94	41,416,189	99.29	93.71	11.04
130	lz-131	C0134	China	North	Hebei	CLR	Group1	44,349,482	6.65	36	97	96	44,372,942	99.48	92.79	11.80
131	lz-132	C0162	China	North	Hebei	CLR	Group1	40,622,434	6.09	35	97	96	40,652,789	99.42	92.93	11.14
132	lz-133	C0166	China	North	Hebei	CLR	Group2	48,936,216	7.34	35	97	96	49,019,709	99.50	92.88	13.41
133	lz-134	C0167	China	North	Hebei	CLR	Group2	45,933,526	6.89	35	97	95	46,004,209	99.48	92.95	12.57
134	lz-135	C0189	China	North	Hebei	CLR	Group2	47,952,712	7.19	35	97	95	48,022,702	99.48	92.86	12.92
135	lz-136	C0194	China	North	Hebei	CLR	Group1	45,408,556	6.81	35	97	95	45,479,571	99.65	93.96	12.40
136	lz-137	C0195	China	North	Hebei	CLR	Group2	40,485,012	6.07	36	97	95	40,332,028	98.94	93.00	10.79
137	lz-138	C0196	China	North	Hebei	CLR	Group2	44,456,358	6.67	35	97	96	44,530,238	99.55	92.80	11.85
138	lz-140	C0208	China	North	Hebei	CLR	Group2	42,448,640	6.37	35	97	95	42,140,927	98.68	92.69	11.38
139	lz-141	C0218	China	North	Hebei	CLR	Group1	42,753,008	6.41	35	97	95	42,516,975	98.82	93.68	11.57
140	lz-142	C0223	China	North	Hebei	CLR	Group2	42,090,518	6.31	35	97	94	42,096,582	99.37	92.88	11.42
141	lz-143	C0227	China	North	Hebei	CLR	Group2	46,644,068	7	35	97	95	46,696,558	99.43	92.82	12.64
142	lz-144	C0228	China	North	Hebei	CLR	Group2	50,188,750	7.53	35	97	96	50,044,962	99.10	93.17	13.72
143	lz-145	C0230	China	North	Hebei	CLR	Group2	47,943,158	7.19	35	97	94	47,971,578	99.46	93.02	13.10
144	lz-146	C0249	China	North	Hebei	CLR	Group1	47,215,928	7.08	35	97	96	47,281,243	99.61	93.67	12.91
145	lz-147	C0251	China	North	Hebei	CLR	Group1	44,973,464	6.75	36	97	96	44,902,886	99.19	93.71	11.93
146	lz-148	C0274	China	North	Hebei	CLR	Group1	44,216,116	6.63	35	97	96	44,209,580	99.41	93.43	12.10
147	lz-149	C0279	China	North	Hebei	CLR	Group1	45,472,112	6.82	37	97	95	45,208,735	98.86	93.97	12.45
148	lz-150	C0286	China	North	Hebei	CLR	Group2	46,454,212	6.97	35	97	96	46,595,440	99.65	92.70	12.90
149	lz-151	C0289	China	North	Hebei	CLR	Group2	41,794,182	6.27	35	97	95	41,883,364	99.61	92.96	11.36
150	lz-152	C0292	China	North	Hebei	CLR	Group2	54,709,598	8.21	36	97	95	54,791,347	99.48	92.68	14.84
151	lz-153	C0300	China	North	Hebei	CLR	Group2	55,584,832	8.34	36	97	96	55,726,456	99.58	92.49	14.95
152	lz-154	C0320	China	North	Hebei	CLR	Group1	46,907,098	7.04	35	97	94	46,981,445	99.62	93.88	12.61
153	lz-155	C0350	China	North	Hebei	CLR	Group2	47,106,126	7.07	36	97	94	46,109,914	97.15	92.44	12.45
154	lz-156	C0655	China	North	Neimenggu	CLR	Group1	40,416,326	6.06	35	97	95	40,500,150	99.66	94.62	11.09
155	lz-157	C1039	China	North	Shandong	CLR	Group1	40,341,214	6.05	38	97	95	39,766,846	97.74	92.19	10.56
156	lz-158	C1595	China	North	Henan	CLR	Group1	43,916,248	6.59	35	97	96	43,885,848	99.32	93.25	11.86

157	lz-159	C1710	China	South	Hubei	CLR	Group1	44,417,992	6.66	36	97	94	44,277,920	99.15	94.56	11.79
158	lz-160	C1718	China	South	Hubei	CLR	Group1	49,400,904	7.41	35	97	96	49,492,839	99.57	94.21	13.26
159	lz-161	C3408(VC1973A	Other	--	--	NCL	NCL Group	49,298,068	7.39	35	97	94	49,026,867	98.93	94.58	13.50
160	lz-162	C3419	Other	--	--	NCL	NCL Group	44,432,444	6.66	35	97	96	44,426,240	99.42	94.02	12.19
161	lz-163	C3677	Other	--	--	CLR	Group1	44,195,714	6.63	36	97	94	44,056,876	98.97	92.89	11.82
162	lz-164	C3682	China	North	Hebei	CLR	Group2	45,931,836	6.89	35	97	95	45,999,150	99.56	92.57	12.25
163	lz-165	C3692	China	North	Hebei	CLR	Group1	40,766,134	6.11	35	97	95	40,724,705	99.43	94.73	11.26
164	lz-166	C3733	China	North	Hebei	CLR	--	46,632,448	6.99	36	97	95	46,773,618	99.64	92.80	12.55
165	lz-167	C3739	China	North	Hebei	CLR	Group2	45,575,662	6.84	36	97	94	45,690,885	99.59	92.40	12.19
166	lz-168	C3812	China	North	Jilin	CLR	Group2	43,790,986	6.57	35	97	96	43,895,760	99.63	92.98	11.72
167	lz-169	C3860	China	North	Liaoning	CLR	Group2	47,107,240	7.07	35	97	96	47,158,200	99.50	93.07	12.71
168	lz-170	C3866	China	North	Liaoning	CLR	Group2	50,238,236	7.54	35	97	95	50,196,870	99.38	93.20	13.82
169	lz-171	C4344	China	North	Liaoning	CLR	Group1	43,500,662	6.53	38	97	95	42,373,998	96.67	92.89	11.15
170	lz-172	C4356	China	South	Hunan	CLR	Group1	42,264,742	6.34	35	97	95	42,299,256	99.43	93.56	11.56
171	lz-173	C4435	China	North	Beijing	CLR	Group2	50,065,320	7.51	35	97	95	49,997,901	99.25	93.42	13.54
172	lz-174	C4466	Other	--	--	NCL	NCL Group	49,576,198	7.44	35	97	94	49,377,209	98.94	93.10	13.37
173	lz-175	C4468	Philippines	--	--	NCL	NCL Group	42,360,968	6.35	35	97	95	42,449,418	99.50	92.98	11.58
174	lz-176	C4635	Other	--	--	CLR	Group2	44,751,246	6.71	39	97	94	44,821,287	99.31	92.18	11.46
175	lz-177	C4640	Other	--	--	NCL	NCL Group	43,822,438	6.57	35	97	95	43,835,526	99.51	94.21	11.80
176	lz-178	C4706	Other	--	--	NCL	NCL Group	49,220,634	7.38	35	97	94	49,252,027	99.50	93.18	13.17
177	lz-179	C5879	China	South	Hubei	CLR	Group1	46,938,296	7.04	35	97	94	46,728,157	98.95	95.41	12.85
178	lz-180	C5883	China	North	Neimenggu	CLR	--	43,499,880	6.52	35	97	95	43,491,488	99.42	93.92	11.87
179	lz-181	VC3541B	Other	--	--	NCL	NCL Group	47,276,980	7.09	37	97	94	47,198,959	99.10	93.19	12.35
180	lz-182	VC4503B	Other	--	--	NCL	NCL Group	45,168,670	6.78	37	97	95	44,951,947	98.88	94.13	12.34
181	lz-183	VC5734A	Other	--	--	NCL	NCL Group	42,109,008	6.32	35	97	95	41,905,974	98.83	92.54	11.30
182	lz-184	VC6089A	Other	--	--	NCL	NCL Group	41,952,334	6.29	35	96	93	41,923,212	99.28	92.99	11.46
183	lz-185	VC6144B	Other	--	--	NCL	NCL Group	49,522,792	7.43	35	97	94	49,464,178	99.14	93.31	13.57
184	lz-186	VC6379	Other	--	--	NCL	NCL Group	47,170,246	7.08	35	97	94	47,324,968	99.54	92.46	12.71
185	lz-187	C0185	China	North	Hebei	CLR	Group1	44,747,690	6.71	35	96	94	44,652,836	99.35	94.80	12.35
186	lz-188	C0076	China	North	Hebei	CLR	Group2	42,686,498	6.4	35	96	93	42,724,010	99.42	92.69	11.64
187	lz-189	C0205	China	North	Hebei	CLR	Group2	46,849,028	7.03	35	96	93	46,652,943	98.95	93.12	12.58
188	lz-190	C0229	China	North	Hebei	CLR	Group2	46,096,774	6.91	35	96	93	45,919,591	99.02	93.19	12.57
189	lz-191	C0305	China	North	Hebei	CLR	Group2	49,314,556	7.4	35	96	93	48,732,206	98.20	93.05	13.17
190	lz-192	C0328	China	North	Hebei	CLR	Group2	45,379,086	6.81	35	96	93	45,017,700	98.61	92.90	12.19
191	lz-193	C0337	China	North	Hebei	CLR	Group2	48,269,668	7.24	35	95	92	48,235,359	99.37	93.60	12.88
192	lz-194	TC1966	Madagascar	--	--	NCL	NCL Group	51,900,836	7.79	36	94	92	52,099,523	98.59	85.75	12.80
193	lz-195	V1128	India	--	--	NCL	NCL Group	45,893,882	6.88	35	96	94	45,563,897	98.40	91.34	12.16
194	lz-196	V2802	Philippines	--	--	NCL	NCL Group	45,064,212	6.76	35	96	94	44,955,378	99.02	92.36	12.23
195	lz-197	V2709	India	--	--	NCL	NCL Group	47,791,142	7.17	35	96	94	47,802,108	98.98	90.61	12.71
196	lz-198	V2817	Nigeria	--	--	NCL	NCL Group	43,671,086	6.55	35	97	94	43,421,561	98.71	92.37	11.82
197	lz-199	ACC41	Australia	--	--	NCL	NCL Group	47,081,342	7.06	35	96	95	47,611,107	98.80	83.37	11.35
198	lz-200	Jilv 13	China	North	Hebei	CBL	--	47,327,610	7.1	35	96	93	47,263,319	99.42	94.55	13.04
199	lz-201	VC2917	Other	--	--	NCL	NCL Group	40,382,718	6.06	35	97	95	40,158,846	98.82	93.02	10.92
200	lz-202	Jilv 0509	China	North	Hebei	CBL	Group1	41,969,274	6.3	36	95	93	41,696,404	98.84	93.79	10.78
201	lz-203	L133	China	North	Hebei	CLR	Group2	42,962,392	6.44	36	96	94	41,612,019	96.19	92.59	11.07
202	lz-204	L131	China	North	Hebei	CLR	Group2	46,833,766	7.03	35	96	92	45,355,218	96.20	92.94	12.13
203	lz-205	L177	China	North	Hebei	CLR	Group2	46,847,004	7.03	34	96	92	46,920,719	99.46	92.76	12.91
204	lz-206	L098	China	North	Hebei	CLR	Group2	45,710,840	6.86	35	96	92	45,402,459	98.64	92.64	12.01
205	lz-207	YLD022	China	North	Hebei	CLR	Group2	47,129,600	7.07	34	96	93	46,985,689	99.00	92.83	12.89
206	lz-208	YLD003	China	North	Hebei	CLR	Group2	41,926,054	6.29	34	96	92	41,550,284	98.43	92.88	11.45
207	lz-209	WM11-6	China	North	Tianjin	CLR	Group2	48,324,166	7.25	35	96	93	48,423,399	99.49	92.65	13.00
208	lz-210	WM11-9	China	North	Shandong	CLR	Group2	49,701,174	7.46	34	96	93	49,168,080	98.20	92.69	13.52
209	lz-211	L042	China	North	Hebei	CLR	Group2	49,439,704	7.42	35	96	93	48,552,299	97.53	92.71	13.33
210	lz-212	WM11-7	China	North	Shandong	CLR	Group2	44,404,544	6.66	34	96	93	44,479,861	99.50	93.00	12.18
211	lz-213	Liaoning 27	China	North	Liaoning	CLR	Group2	45,622,616	6.84	36	96	92	45,702,379	99.46	92.47	12.02
212	lz-214	L167	China	North	Hebei	CLR	Group2	46,500,422	6.98	35	96	92	46,479,749	99.24	92.72	12.58
213	lz-215	L174	China	North	Hebei	CLR	Group2	46,019,014	6.9	34	96	92	46,077,933	99.48	92.74	12.40
214	lz-216	L165	China	North	Hebei	CLR	Group2	46,610,112	6.99	35	95	92	46,389,101	98.86	93.00	12.62
215	lz-217	L179	China	North	Hebei	CLR	Group2	44,973,618	6.75	34	96	93	44,972,366	99.26	92.73	12.44
216	lz-218	L161	China	North	Hebei	CLR	Group2	44,405,134	6.66	35	96	93	44,121,490	98.65	92.81	12.04
217	lz-219	Jiangsu wild 19	China	South	Jiangsu	CLR	Group2	44,508,828	6.68	34	96	93	44,606,560	99.50	92.85	12.30

Table S5 Summary of SNP locations: detection of SNP loci classified according to coding (intron or exon) or non-coding and intergenic regions on all 11 chromosomes (Chr. 1 to Chr.11) of the mungbean genome.

	Chr.1	Chr.2	Chr.3	Chr.4	Chr.5	Chr.6	Chr.7	Chr.8	Chr.9	Chr.10	Chr.11	Unplaced	Total
Total	255,066	253,428	101,950	264,479	164,166	261,463	291,312	195,318	225,503	117,205	95,596	3,857	2,229,343
Intergenic	191,382	196,452	69,868	197,759	126,928	200,553	221,064	143,857	169,180	89,556	66,421	3,426	1,676,446
Intron	60,462	52,384	29,563	56,194	33,534	51,058	60,032	43,226	50,219	25,411	26,681	312	489,076
Exon	12,825	10,903	6,772	14,366	7,467	13,668	14,568	10,373	10,032	5,668	6,321	85	113,048
Downstream	104,231	99,316	54,876	122,313	71,544	106,461	129,904	97,295	86,294	48,248	49,275	800	970,557
Upstream	104,542	94,341	55,156	129,952	74,177	113,649	135,805	99,179	88,047	48,458	50,876	798	994,980
Splicing	948	818	647	1,101	649	1,069	1,263	915	828	454	549	12	9,253
UTR_3_PRIME	5,782	4,790	3,518	7,364	3,843	6,517	8,443	5,776	4,323	2,660	3,261	39	56,316
UTR_5_PRIME	3,509	3,368	2,450	4,656	2,734	3,383	4,890	4,338	2,888	1,892	2,450	12	36,570
Number Non-synonymous(Percentage)	6933 (2.72%)	6054 (2.39%)	3522 (3.45%)	7670 (2.90%)	4137 (2.52%)	7487 (2.86%)	7874 (2.70%)	5547 (2.84%)	5608 (2.49%)	3185 (2.72%)	3432 (3.59%)	58 (1.50%)	61507 (2.76%)
Number Synonymous(Percentage)	5609 (2.93%)	4657 (2.37%)	3186 (4.56%)	6538 (3.31%)	3168 (2.50%)	5998 (2.99%)	6482 (2.93%)	4679 (3.25%)	4260 (2.52%)	2374 (2.65%)	2771 (4.17%)	106 (3.09%)	49828 (2.97%)

Table S6 Summary of SNP identified in 217 mungbean genotypes.

Sample ID	Total SNPs	Reference SNPs	Heterozygous SNPs	Homozygous SNPs	Synonymous SNPs	Non-synonymous SNPs	Ratio
lz-1	2,076,105	1,311,057	179,883	585,165	39,845	54,530	1.3686
lz-10	2,120,691	1,319,816	211,395	589,480	40,679	55,656	1.3682
lz-100	2,158,663	1,339,589	173,752	645,322	41,674	56,661	1.3596
lz-101	2,152,114	1,281,000	190,466	680,648	41,456	56,559	1.3643
lz-102	1,815,102	1,171,436	150,615	493,051	36,402	49,605	1.3627
lz-103	2,142,043	1,379,259	176,157	586,627	41,311	56,354	1.3641
lz-104	2,166,271	1,362,390	187,375	616,506	41,622	56,805	1.3648
lz-105	2,119,586	1,144,025	210,884	764,677	41,149	55,733	1.3544
lz-106	2,138,648	1,406,274	227,801	504,573	41,241	55,973	1.3572
lz-107	2,179,130	1,368,614	190,107	620,409	41,799	57,037	1.3646
lz-108	2,074,427	1,308,432	185,845	580,150	40,860	55,630	1.3615
lz-109	2,148,037	1,436,280	169,800	541,957	41,290	56,425	1.3666
lz-11	2,137,856	1,287,037	201,846	648,973	41,171	56,145	1.3637
lz-110	2,142,655	1,716,204	108,145	318,306	41,189	56,176	1.3639
lz-111	2,151,979	1,384,372	179,629	587,978	41,373	56,543	1.3667
lz-112	2,169,256	1,639,514	127,018	402,724	41,640	56,763	1.3632
lz-113	2,190,263	1,408,779	182,701	598,783	41,952	57,261	1.3649
lz-114	2,150,275	1,380,586	182,932	586,757	41,461	56,622	1.3657
lz-115	2,167,601	1,785,121	92,491	289,989	41,601	56,820	1.3658
lz-116	2,169,234	1,774,540	104,686	290,008	41,632	56,884	1.3664
lz-117	2,154,914	1,361,798	167,929	625,187	41,223	56,360	1.3672
lz-118	2,115,352	1,453,567	279,572	382,213	40,735	55,625	1.3655
lz-119	2,131,874	1,376,756	182,326	572,792	41,112	56,117	1.3650
lz-12	2,116,326	1,273,495	197,210	645,621	40,652	55,617	1.3681
lz-120	2,139,643	1,598,613	127,431	413,599	41,105	56,147	1.3659
lz-121	2,118,309	1,334,045	197,865	586,399	41,262	56,414	1.3672
lz-122	2,175,600	1,654,415	120,810	400,375	41,750	56,942	1.3639
lz-123	2,145,198	1,661,902	113,521	369,775	41,388	56,595	1.3674
lz-124	2,137,931	1,488,662	162,846	486,423	41,324	56,303	1.3625
lz-125	2,152,704	1,326,558	194,692	631,454	41,445	56,535	1.3641
lz-126	2,151,094	1,369,275	178,915	602,904	41,408	56,447	1.3632
lz-127	2,149,771	1,247,020	341,592	561,159	41,404	56,567	1.3662
lz-128	2,101,248	1,372,746	170,593	557,909	40,612	55,612	1.3693
lz-129	2,102,939	1,246,278	207,193	649,468	41,362	56,359	1.3626
lz-13	2,116,187	1,265,832	204,209	646,146	40,852	55,763	1.3650
lz-130	2,107,376	1,335,079	253,191	519,106	40,609	55,460	1.3657
lz-131	2,153,195	1,326,552	195,631	631,012	41,623	56,700	1.3622
lz-132	2,138,952	1,263,874	205,239	669,839	41,258	56,191	1.3619
lz-133	2,183,667	1,315,448	202,832	665,387	41,918	57,143	1.3632
lz-134	2,170,461	1,325,765	193,201	651,495	41,601	56,708	1.3631
lz-135	2,177,533	1,272,947	210,668	693,918	41,780	56,972	1.3636
lz-136	2,161,722	1,469,201	168,904	523,617	41,627	56,807	1.3647
lz-137	2,109,279	1,330,256	186,672	592,351	40,717	55,633	1.3663
lz-138	2,153,511	1,280,089	202,931	670,491	41,597	56,643	1.3617
lz-14	2,153,058	1,322,395	190,428	640,235	41,232	56,318	1.3659
lz-140	2,143,057	1,294,910	192,936	655,211	41,420	56,492	1.3639
lz-141	2,136,171	1,134,185	391,601	610,385	41,242	56,125	1.3609

lz-142	2,145,801	1,291,985	192,013	661,803	41,335	56,411	1.3647
lz-143	2,172,408	1,302,715	201,145	668,548	41,721	56,915	1.3642
lz-144	2,190,860	1,355,919	194,530	640,411	41,968	57,207	1.3631
lz-145	2,179,602	1,321,063	196,143	662,396	41,862	57,074	1.3634
lz-146	2,175,195	1,457,955	181,012	536,228	41,773	56,967	1.3637
lz-147	2,122,409	1,354,065	297,176	471,168	40,795	55,910	1.3705
lz-148	2,165,407	1,432,194	171,643	561,570	41,538	56,628	1.3633
lz-149	2,117,025	1,522,057	157,252	437,716	41,743	56,906	1.3632
lz-15	2,145,693	1,280,865	204,092	660,736	41,166	56,135	1.3636
lz-150	2,179,596	1,301,442	207,990	670,164	41,809	56,946	1.3621
lz-151	2,150,305	1,278,814	247,505	623,986	41,350	56,346	1.3627
lz-152	2,191,001	1,337,814	198,194	654,993	41,911	57,287	1.3669
lz-153	2,188,150	1,254,363	220,973	712,814	41,987	57,275	1.3641
lz-154	2,167,802	1,430,975	164,014	572,813	41,590	56,699	1.3633
lz-155	2,139,156	1,252,063	212,372	674,721	41,164	56,175	1.3647
lz-156	2,095,824	1,166,817	566,318	362,689	40,795	55,438	1.3589
lz-157	1,975,391	1,249,747	199,468	526,176	38,944	53,564	1.3754
lz-158	2,152,385	1,302,567	199,210	650,608	41,363	56,446	1.3646
lz-159	2,132,257	1,497,227	168,705	466,325	41,171	56,033	1.3610
lz-16	2,160,945	1,380,432	182,876	597,637	41,467	56,591	1.3647
lz-160	2,169,887	1,466,382	161,270	542,235	41,662	56,830	1.3641
lz-161	2,167,155	1,477,260	138,155	551,740	41,557	56,819	1.3673
lz-162	2,155,048	1,375,443	179,970	599,635	41,472	56,383	1.3595
lz-163	2,104,064	1,281,450	201,158	621,456	40,427	55,360	1.3694
lz-164	2,156,657	1,253,174	221,164	682,319	41,554	56,629	1.3628
lz-165	2,130,647	1,565,009	130,529	435,109	41,255	56,246	1.3634
lz-166	2,113,301	1,359,638	171,414	582,249	41,152	56,223	1.3662
lz-167	2,158,460	1,280,513	208,295	669,652	41,624	56,704	1.3623
lz-168	2,153,312	1,356,675	187,025	609,612	41,539	56,574	1.3619
lz-169	2,170,068	1,360,819	192,769	616,480	41,664	56,790	1.3630
lz-17	2,163,933	1,307,169	195,244	661,520	41,617	56,740	1.3634
lz-170	2,188,951	1,413,630	182,413	592,908	42,011	57,278	1.3634
lz-171	2,058,202	1,316,282	179,912	562,008	40,644	55,355	1.3619
lz-172	2,122,778	1,418,980	160,451	543,347	40,864	55,609	1.3608
lz-173	2,187,093	1,390,978	183,829	612,286	41,862	57,067	1.3632
lz-174	2,171,658	1,311,138	201,657	658,863	41,713	56,770	1.3610
lz-175	2,126,981	1,329,711	160,120	637,150	40,992	55,874	1.3630
lz-176	1,928,289	1,249,056	149,093	530,140	39,391	53,663	1.3623
lz-177	2,129,176	1,588,823	97,200	443,153	41,195	56,220	1.3647
lz-178	2,165,321	1,364,345	196,127	604,849	41,572	56,797	1.3662
lz-179	2,141,023	1,352,898	456,763	331,362	41,123	55,956	1.3607
lz-18	2,100,623	1,347,039	184,661	568,923	40,089	54,943	1.3705
lz-180	2,141,576	1,479,691	147,536	514,349	41,389	56,471	1.3644
lz-181	2,104,253	1,362,439	141,522	600,292	40,998	55,688	1.3583
lz-182	2,103,291	1,442,362	131,016	529,913	40,930	55,969	1.3674
lz-183	2,107,339	1,315,572	160,334	631,433	40,933	55,595	1.3582
lz-184	2,118,470	1,362,191	162,466	593,813	40,991	55,769	1.3605
lz-185	2,141,672	1,390,651	150,724	600,297	41,278	56,157	1.3605
lz-186	2,130,477	1,266,979	169,329	694,169	41,178	56,073	1.3617
lz-187	2,155,910	1,575,719	135,317	444,874	41,558	56,650	1.3632
lz-188	2,151,090	1,286,111	199,990	664,989	41,380	56,420	1.3635

lz-189	2,164,297	1,351,179	184,101	629,017	41,610	56,777	1.3645
lz-19	2,156,397	1,246,184	217,184	693,029	41,388	56,408	1.3629
lz-190	2,164,584	1,377,105	185,724	601,755	41,507	56,658	1.3650
lz-191	2,176,403	1,273,683	204,841	697,879	41,704	56,832	1.3627
lz-192	2,157,696	1,300,559	202,142	654,995	41,416	56,546	1.3653
lz-193	2,164,500	1,346,138	234,224	584,138	41,661	56,786	1.3630
lz-194	1,924,389	1,046,612	150,502	727,275	39,538	52,309	1.3230
lz-195	2,097,011	1,100,655	201,127	795,229	40,429	55,081	1.3624
lz-196	2,121,084	1,272,350	180,027	668,707	40,872	55,626	1.3610
lz-197	2,102,089	1,069,446	178,327	854,316	40,847	55,264	1.3530
lz-198	2,111,709	1,266,826	180,545	664,338	40,619	55,412	1.3642
lz-199	1,848,317	992,519	107,963	747,835	38,386	49,940	1.3010
lz-2	2,144,689	1,372,783	178,381	593,525	40,934	55,909	1.3658
lz-20	2,118,844	1,290,344	194,597	633,903	41,083	56,009	1.3633
lz-200	2,164,324	1,643,838	131,428	389,058	41,592	56,708	1.3634
lz-201	2,083,483	1,301,153	132,070	650,260	40,664	55,085	1.3546
lz-202	2,069,878	1,588,739	112,244	368,895	40,353	54,905	1.3606
lz-203	2,096,054	1,177,585	205,200	713,269	40,908	55,632	1.3599
lz-204	2,164,658	1,267,567	203,455	693,636	41,504	56,647	1.3649
lz-205	2,178,391	1,214,944	211,508	751,939	41,728	56,790	1.3610
lz-206	2,149,696	1,180,586	214,774	754,336	41,308	56,252	1.3618
lz-207	2,179,268	1,228,009	212,858	738,401	41,753	56,880	1.3623
lz-208	2,159,377	1,265,629	200,969	692,779	41,458	56,521	1.3633
lz-209	2,167,051	1,200,006	205,232	761,813	41,609	56,596	1.3602
lz-21	2,144,608	1,221,857	221,094	701,657	41,202	56,153	1.3629
lz-210	2,181,906	1,184,063	226,805	771,038	41,733	56,830	1.3618
lz-211	2,183,894	1,234,585	216,214	733,095	41,778	56,975	1.3638
lz-212	2,163,222	1,242,566	195,717	724,939	41,558	56,565	1.3611
lz-213	2,149,665	1,174,107	218,943	756,615	41,489	56,510	1.3620
lz-214	2,168,468	1,191,539	217,630	759,299	41,618	56,671	1.3617
lz-215	2,167,506	1,233,087	203,097	731,322	41,572	56,721	1.3644
lz-216	2,171,011	1,267,482	192,804	710,725	41,609	56,691	1.3625
lz-217	2,175,809	1,213,386	212,044	750,379	41,672	56,777	1.3625
lz-218	2,162,130	1,224,073	212,191	725,866	41,608	56,687	1.3624
lz-219	2,167,628	1,212,156	202,107	753,365	41,662	56,732	1.3617
lz-22	2,098,569	1,234,346	207,809	656,414	40,576	55,415	1.3657
lz-23	2,172,641	1,359,847	181,291	631,503	41,632	56,899	1.3667
lz-24	2,108,216	1,265,367	201,707	641,142	40,538	55,383	1.3662
lz-25	2,155,582	1,376,322	190,186	589,074	41,384	56,360	1.3619
lz-26	2,100,199	1,403,501	169,115	527,583	40,475	55,332	1.3671
lz-27	2,087,094	1,583,840	138,680	364,574	41,243	56,333	1.3659
lz-28	2,053,683	1,272,739	191,968	588,976	39,846	54,442	1.3663
lz-29	2,110,844	1,454,138	161,981	494,725	40,579	55,301	1.3628
lz-3	2,160,127	1,676,292	109,305	374,530	41,270	56,413	1.3669
lz-30	2,149,399	1,438,155	168,150	543,094	41,056	56,275	1.3707
lz-32	2,075,407	1,389,917	164,381	521,109	40,284	55,093	1.3676
lz-33	2,166,486	1,402,898	165,939	597,649	41,428	56,455	1.3627
lz-34	2,146,306	1,548,108	138,456	459,742	40,696	55,937	1.3745
lz-35	2,108,782	1,449,115	151,750	507,917	40,471	55,335	1.3673
lz-36	2,167,108	1,310,876	192,745	663,487	41,476	56,693	1.3669
lz-37	2,107,556	1,415,699	164,994	526,863	40,423	55,219	1.3660

lz-38	2,143,986	1,316,248	196,131	631,607	41,121	56,175	1.3661
lz-39	2,143,293	1,276,446	192,031	674,816	41,277	56,341	1.3649
lz-4	2,144,332	1,524,823	139,595	479,914	41,120	56,106	1.3644
lz-40	2,133,292	1,622,234	43,054	468,004	40,989	56,072	1.3680
lz-41	2,001,769	1,274,233	171,432	556,104	39,457	54,093	1.3709
lz-42	2,126,832	1,355,851	236,080	534,901	40,583	55,637	1.3709
lz-43	2,118,410	1,312,200	194,217	611,993	40,574	55,323	1.3635
lz-44	2,150,658	1,467,457	162,807	520,394	41,133	56,232	1.3671
lz-45	2,140,419	1,219,092	331,284	590,043	40,858	55,903	1.3682
lz-46	2,113,498	1,260,431	198,381	654,686	41,069	56,035	1.3644
lz-47	2,005,329	1,330,285	158,469	516,575	39,494	54,048	1.3685
lz-48	2,112,548	1,489,546	146,181	476,821	40,448	55,212	1.3650
lz-49	2,120,243	1,460,077	152,367	507,799	40,835	55,670	1.3633
lz-5	2,124,779	1,371,362	169,515	583,902	40,619	55,713	1.3716
lz-50	2,103,302	1,248,995	193,196	661,111	40,751	55,490	1.3617
lz-51	2,087,782	1,369,753	160,251	557,778	40,342	55,105	1.3659
lz-52	2,131,196	1,265,037	195,152	671,007	41,085	56,023	1.3636
lz-53	2,160,892	1,273,281	205,640	681,971	41,365	56,485	1.3655
lz-54	2,159,803	1,408,630	189,676	561,497	41,415	56,483	1.3638
lz-55	2,165,248	1,317,921	192,492	654,835	41,820	57,013	1.3633
lz-56	2,157,369	1,640,793	123,299	393,277	41,395	56,361	1.3615
lz-57	2,134,076	1,583,390	122,282	428,404	40,831	55,860	1.3681
lz-58	2,152,024	1,530,657	149,522	471,845	41,249	56,231	1.3632
lz-59	2,139,352	1,409,276	181,620	548,456	40,970	55,996	1.3668
lz-6	2,128,960	1,739,916	104,854	284,190	41,306	56,299	1.3630
lz-60	2,139,549	1,514,076	156,735	468,738	40,866	55,922	1.3684
lz-61	2,096,140	1,502,472	132,925	460,743	40,245	55,117	1.3695
lz-62	2,111,466	1,403,884	165,669	541,913	40,455	55,225	1.3651
lz-63	2,158,287	1,496,211	155,826	506,250	41,414	56,435	1.3627
lz-64	2,144,031	1,444,128	155,321	544,582	40,982	55,898	1.3640
lz-65	2,138,532	1,450,287	160,915	527,330	40,942	55,956	1.3667
lz-66	2,167,956	1,459,745	168,943	539,268	41,510	56,656	1.3649
lz-67	2,162,616	1,572,227	144,325	446,064	41,480	56,504	1.3622
lz-68	2,148,995	1,298,498	190,543	659,954	41,316	56,368	1.3643
lz-69	2,140,097	1,458,838	165,431	515,828	41,100	56,207	1.3676
lz-7	2,141,808	1,328,505	194,776	618,527	41,397	56,391	1.3622
lz-70	2,160,021	1,383,731	170,172	606,118	41,549	56,448	1.3586
lz-71	2,174,786	1,442,523	164,256	568,007	41,657	56,713	1.3614
lz-72	2,159,193	1,402,968	175,930	580,295	41,486	56,515	1.3623
lz-73	2,064,396	955,179	865,492	243,725	39,953	54,350	1.3603
lz-74	2,158,231	1,501,011	153,733	503,487	41,631	56,663	1.3611
lz-75	2,107,689	1,706,979	105,225	295,485	40,844	55,705	1.3638
lz-76	2,174,080	2,107,792	25,822	40,466	41,780	57,023	1.3648
lz-77	2,179,615	1,696,303	115,415	367,897	41,860	57,123	1.3646
lz-78	2,149,041	1,680,416	116,534	352,091	41,365	56,501	1.3659
lz-79	2,110,273	1,787,824	77,500	244,949	40,755	55,844	1.3702
lz-8	2,142,128	1,298,711	190,211	653,206	41,268	56,254	1.3631
lz-80	1,995,271	1,465,536	168,798	360,937	39,933	54,482	1.3643
lz-81	2,119,250	1,208,553	663,971	246,726	40,959	55,693	1.3597
lz-82	2,143,357	1,776,490	79,099	287,768	41,333	56,570	1.3686
lz-83	2,143,760	1,524,697	136,340	482,723	41,412	56,303	1.3596

lz-84	2,134,048	1,384,906	165,296	583,846	41,255	56,179	1.3618
lz-85	2,180,450	1,439,269	148,911	592,270	41,728	56,905	1.3637
lz-86	2,131,006	1,320,454	140,373	670,179	41,219	55,933	1.3570
lz-87	2,167,592	1,392,102	183,278	592,212	41,780	57,024	1.3649
lz-88	2,188,685	1,308,373	213,535	666,777	41,874	57,070	1.3629
lz-89	2,153,718	1,564,916	141,671	447,131	41,381	56,510	1.3656
lz-9	2,154,723	1,360,444	184,684	609,595	41,452	56,576	1.3649
lz-90	2,178,150	1,394,784	183,330	600,036	41,774	57,020	1.3650
lz-91	2,136,774	1,287,744	207,200	641,830	41,040	55,997	1.3644
lz-92	2,187,752	1,378,446	190,070	619,236	41,936	57,286	1.3660
lz-93	2,168,954	1,307,052	204,769	657,133	41,715	56,857	1.3630
lz-94	2,150,239	1,376,128	179,375	594,736	41,376	56,529	1.3662
lz-95	2,179,948	1,265,719	209,802	704,427	41,859	57,004	1.3618
lz-96	2,164,141	1,323,155	155,316	685,670	41,583	56,513	1.3590
lz-97	2,186,711	1,521,424	156,626	508,661	41,859	57,127	1.3647
lz-98	2,138,922	1,366,137	176,237	596,548	41,316	56,348	1.3638
lz-99	2,130,166	1,477,342	121,441	531,383	41,242	55,987	1.3575

Table S7 Summary of assemblies for mungbean pan genome.

Assembly	Total length	# contigs	Largest contig	GC (%)	N50 (bp)	N75	L50	L75	# N's per 100 kb	non-reference and non-contamination(bp)
lz-1	382,334,782	234,554	342,934	33.55	1,916	1,188	60,635	121,414	38.61	20,379,488
lz-10	401,201,276	167,468	368,279	33.67	3,277	1,854	35,951	75,800	26.59	14,033,038
lz-100	410,215,212	139,946	249,720	33.39	4,225	2,367	29,050	61,016	29.35	12,295,315
lz-101	379,814,146	223,971	231,857	33.84	2,036	1,219	55,055	112,694	151.59	23,025,387
lz-102	403,651,119	142,810	279,688	33.37	4,019	2,262	29,920	62,907	48.81	11,814,876
lz-103	392,618,771	185,307	167,286	33.59	2,775	1,611	42,595	87,831	49.64	15,261,559
lz-104	412,562,059	148,899	394,848	33.82	3,930	2,217	31,435	65,762	37.61	14,534,156
lz-105	416,831,818	155,186	388,613	33.64	3,810	2,154	32,901	68,471	37.06	20,685,502
lz-106	406,641,627	167,043	336,029	33.42	3,361	1,898	36,022	75,356	39.84	14,820,378
lz-107	412,321,862	132,093	293,875	33.42	4,571	2,548	27,205	56,998	31.03	12,077,446
lz-108	362,294,866	243,996	402,011	34.26	1,653	1,032	63,828	128,911	51.45	26,752,125
lz-109	399,457,104	172,803	392,596	33.43	3,102	1,795	39,122	80,631	44.78	13,301,922
lz-11	393,360,832	184,153	401,697	33.55	2,791	1,617	41,764	86,952	30.11	15,725,351
lz-110	399,033,375	185,997	205,986	33.50	2,800	1,643	43,266	88,682	48.47	11,334,220
lz-111	398,286,063	175,948	402,658	33.56	3,027	1,744	39,688	82,003	52.07	14,625,587
lz-112	408,449,205	146,979	393,197	33.36	3,913	2,231	31,591	65,597	38.70	9,299,389
lz-113	417,516,252	103,291	383,915	33.29	6,274	3,419	19,773	42,081	27.86	10,276,594
lz-114	395,170,868	180,917	352,420	33.65	2,885	1,663	40,830	84,841	24.34	15,313,826
lz-115	407,408,192	163,458	406,115	33.48	3,420	1,963	36,033	74,571	40.70	9,194,037
lz-116	408,556,049	146,256	401,718	33.43	3,964	2,233	31,047	64,826	40.22	8,417,299
lz-117	409,835,907	154,562	246,508	33.39	3,681	2,115	33,824	69,860	37.92	12,652,268
lz-118	389,981,073	206,606	191,692	33.56	2,374	1,395	48,895	100,643	48.56	16,414,490
lz-119	391,096,219	204,630	247,806	33.83	2,418	1,409	47,744	98,826	48.38	19,599,565
lz-12	381,964,860	216,543	392,178	33.70	2,141	1,290	53,051	108,429	32.60	19,864,632
lz-120	396,748,135	197,037	399,995	33.64	2,595	1,506	45,567	94,293	44.99	14,455,380
lz-121	388,541,212	212,737	289,501	34.50	2,268	1,326	50,086	103,681	42.78	22,541,217
lz-122	411,432,454	141,933	403,941	33.35	4,136	2,335	29,988	62,596	38.96	9,072,669
lz-123	394,930,758	194,815	401,857	33.72	2,619	1,523	45,057	93,060	44.51	13,993,061
lz-124	395,974,872	190,495	294,836	33.54	2,699	1,573	43,857	90,736	43.45	14,656,678
lz-125	398,206,068	189,819	401,309	33.52	2,726	1,590	43,758	90,362	41.08	15,996,236
lz-126	398,710,640	178,794	402,252	33.52	2,959	1,713	40,515	83,751	35.09	14,616,051
lz-127	397,346,223	196,529	401,948	33.62	2,616	1,513	44,961	93,333	39.24	17,958,632
lz-128	375,816,011	221,810	402,142	34.07	2,039	1,202	53,282	110,625	57.42	22,039,419
lz-129	369,047,704	214,780	368,116	34.18	2,081	1,229	51,958	106,874	42.54	23,447,022
lz-13	393,370,922	191,199	393,385	33.60	2,669	1,549	43,931	91,037	33.97	16,885,603
lz-130	382,178,245	228,136	229,031	33.74	1,993	1,211	57,184	115,982	54.44	21,176,472
lz-131	395,521,050	179,367	185,340	33.67	2,937	1,680	39,927	83,397	35.47	15,156,947
lz-132	394,704,367	191,024	285,542	33.51	2,678	1,563	44,425	91,391	49.94	16,105,870
lz-133	409,780,384	142,674	372,019	33.40	4,113	2,306	29,776	62,565	25.84	12,852,879
lz-134	404,425,124	166,813	402,077	33.38	3,302	1,891	37,013	76,674	41.14	14,381,363
lz-135	409,804,566	152,791	401,642	33.38	3,744	2,132	33,054	68,702	32.70	13,543,993
lz-136	403,337,512	161,195	404,036	33.55	3,478	1,942	34,124	72,197	37.60	12,725,775
lz-137	380,657,324	227,744	324,425	34.09	1,988	1,191	55,605	114,665	46.14	22,106,827
lz-138	400,527,584	159,538	306,945	33.58	3,522	1,943	33,313	70,897	36.50	14,606,155
lz-14	394,969,119	188,355	353,317	33.52	2,713	1,591	43,518	89,952	25.83	15,569,559

lz-140	394,406,017	182,777	383,819	33.68	2,851	1,643	41,400	85,809	39.39	15,576,729
lz-141	403,079,650	165,472	332,506	33.49	3,403	1,887	34,776	73,630	41.45	15,835,618
lz-142	394,390,168	181,336	164,623	33.53	2,871	1,661	41,256	85,316	53.87	15,556,118
lz-143	403,772,240	162,687	261,088	33.51	3,392	1,953	36,028	74,422	41.99	13,966,207
lz-144	415,671,523	118,216	228,905	33.35	5,328	2,903	23,127	49,261	26.62	11,346,691
lz-145	410,726,348	134,635	233,379	33.43	4,448	2,474	27,235	57,711	30.52	12,187,215
lz-146	410,559,594	135,511	370,706	33.37	4,419	2,441	27,615	58,490	32.68	11,029,457
lz-147	383,706,838	223,486	381,787	33.90	2,071	1,225	53,700	111,254	39.13	22,285,989
lz-148	403,412,342	168,367	394,434	33.39	3,229	1,870	37,848	78,098	33.12	12,863,429
lz-149	382,237,916	197,990	198,945	34.27	2,476	1,413	45,003	94,137	47.68	18,306,291
lz-15	394,933,080	182,666	280,465	33.46	2,823	1,651	41,987	86,724	29.22	15,522,938
lz-150	409,300,828	140,553	207,850	33.45	4,202	2,341	29,216	61,350	35.29	12,806,954
lz-151	399,486,902	165,811	329,843	33.51	3,287	1,868	35,885	75,351	34.85	14,375,530
lz-152	406,442,533	146,802	260,926	33.67	3,955	2,210	30,797	64,554	32.09	13,542,963
lz-153	411,031,262	128,315	257,073	33.57	4,938	2,616	24,296	52,555	27.31	13,739,426
lz-154	405,243,367	155,798	381,035	33.46	3,632	2,043	33,072	69,643	30.02	12,347,773
lz-155	388,314,418	212,061	124,611	33.76	2,261	1,354	51,598	105,144	26.04	19,163,547
lz-156	383,265,209	200,450	403,371	33.89	2,456	1,384	44,751	94,629	50.25	19,280,489
lz-157	328,747,445	251,931	348,770	34.97	1,307	858	70,859	140,636	47.08	36,540,165
lz-158	404,052,427	165,162	403,350	33.57	3,360	1,900	35,858	75,001	41.36	14,649,979
lz-159	392,718,545	205,415	401,803	33.57	2,402	1,427	49,194	100,586	43.91	15,887,770
lz-16	397,130,528	171,379	401,734	33.44	3,083	1,793	38,672	80,168	14.50	13,038,333
lz-160	408,196,572	167,128	248,384	33.43	3,329	1,911	37,023	76,708	35.20	13,208,788
lz-161	410,907,688	143,411	151,508	33.46	4,135	2,296	29,565	62,323	38.74	10,475,779
lz-162	406,506,461	154,891	402,698	33.37	3,662	2,069	33,265	69,542	38.13	12,481,189
lz-163	377,045,125	237,232	305,767	34.01	1,842	1,127	60,184	122,307	48.67	24,178,993
lz-164	401,807,961	150,678	353,718	33.55	3,778	2,095	31,243	66,353	30.32	13,941,774
lz-165	391,046,366	200,021	380,782	33.73	2,498	1,452	46,547	96,376	36.11	14,950,877
lz-166	377,229,488	196,363	284,801	34.29	2,484	1,391	43,576	92,290	32.15	20,903,671
lz-167	394,114,986	179,433	369,754	33.77	2,924	1,675	40,283	83,729	42.15	16,138,706
lz-168	397,984,108	181,382	305,278	33.50	2,899	1,678	41,337	85,440	36.13	15,298,784
lz-169	407,538,707	139,100	287,785	33.45	4,272	2,354	28,350	60,072	34.28	12,437,821
lz-17	410,157,173	127,137	406,612	33.54	4,815	2,647	25,291	53,601	46.92	11,950,640
lz-170	414,326,042	107,716	402,504	33.46	6,126	3,237	19,792	42,796	31.38	10,617,818
lz-171	358,454,002	238,586	353,257	35.32	1,671	1,017	59,993	123,955	33.94	30,036,917
lz-172	388,538,558	221,024	217,190	33.57	2,132	1,289	54,701	111,229	41.37	19,143,841
lz-173	412,551,172	129,868	349,960	33.34	4,661	2,592	26,361	55,651	28.97	11,216,128
lz-174	412,602,299	133,135	291,615	33.34	4,550	2,513	27,005	57,137	35.06	12,754,053
lz-175	399,836,576	179,222	185,568	33.45	2,951	1,722	41,120	84,448	47.90	15,590,890
lz-176	316,149,581	243,291	275,458	35.47	1,292	848	68,484	135,945	48.82	36,243,331
lz-177	397,388,686	180,226	323,989	33.57	2,929	1,684	40,617	84,346	44.55	13,300,679
lz-178	409,554,981	130,559	208,941	33.43	4,672	2,551	25,941	55,230	32.97	11,911,330
lz-179	406,493,503	191,913	191,364	33.66	2,826	1,582	41,954	88,479	40.46	17,075,099
lz-18	370,438,531	238,621	353,867	33.69	1,779	1,105	61,594	124,593	24.42	22,156,904
lz-180	397,298,793	183,479	363,195	33.68	2,856	1,648	41,343	85,996	34.52	14,408,648
lz-181	399,122,888	190,423	401,886	34.00	2,786	1,570	42,202	88,517	48.48	18,341,219
lz-182	382,147,119	215,201	274,784	34.15	2,180	1,284	51,550	106,226	47.51	19,428,776
lz-183	398,434,011	178,377	402,241	33.45	2,959	1,716	40,507	83,724	42.13	16,496,916

lz-184	396,813,361	171,028	249,557	33.53	3,129	1,786	37,584	78,673	37.12	15,320,962
lz-185	409,892,440	158,283	296,276	33.39	3,572	2,044	34,634	71,850	33.47	14,157,237
lz-186	406,077,331	152,308	280,699	33.58	3,823	2,089	30,817	66,112	28.30	16,034,937
lz-187	401,293,639	161,840	213,767	33.63	3,472	1,919	33,733	71,812	35.70	12,152,525
lz-188	395,091,034	183,536	392,899	33.53	2,831	1,636	41,724	86,458	44.94	15,903,603
lz-189	399,192,375	175,162	234,544	33.53	3,059	1,753	38,900	81,085	34.27	14,663,822
lz-19	403,027,103	166,489	347,953	33.43	3,273	1,883	36,840	76,638	27.11	14,406,195
lz-190	400,924,049	166,559	167,769	33.60	3,322	1,862	35,829	75,290	32.01	14,101,741
lz-191	411,157,514	134,097	384,536	33.37	4,475	2,474	27,101	57,610	32.05	12,951,290
lz-192	400,898,943	156,379	349,663	33.49	3,588	1,985	32,300	69,178	28.37	14,038,812
lz-193	402,631,381	153,337	360,717	33.55	3,764	2,054	31,123	66,600	33.16	13,627,068
lz-194	399,829,916	134,297	174,494	34.05	4,857	2,380	22,615	51,400	37.10	37,128,410
lz-195	402,412,143	165,773	404,658	33.52	3,333	1,877	35,843	75,158	38.66	18,973,417
lz-196	398,333,772	174,867	335,984	33.49	3,028	1,763	39,734	81,886	47.65	16,691,202
lz-197	409,278,399	139,208	293,058	33.34	4,225	2,377	29,078	60,900	45.08	18,690,283
lz-198	397,402,756	183,852	293,004	33.70	2,853	1,642	41,477	86,191	44.75	19,255,360
lz-199	410,312,325	137,527	261,413	33.37	4,319	2,406	28,388	59,723	40.32	41,012,973
lz-2	404,435,564	170,346	177,706	33.27	3,174	1,853	38,558	79,438	33.12	13,406,773
lz-20	378,567,248	204,147	331,783	33.96	2,315	1,353	47,649	99,390	24.62	18,896,433
lz-200	412,626,683	125,079	287,361	33.39	4,957	2,705	24,383	52,204	41.01	9,088,218
lz-201	388,656,225	195,445	401,671	33.54	2,537	1,490	45,724	94,339	42.25	16,882,146
lz-202	375,645,016	223,958	295,462	33.71	1,999	1,213	56,078	113,932	55.47	18,728,606
lz-203	385,689,981	188,549	203,754	33.86	2,640	1,539	43,494	89,933	41.95	18,192,123
lz-204	407,268,837	141,511	241,782	33.40	4,118	2,318	29,726	62,149	36.76	13,656,318
lz-205	403,521,936	108,900	174,942	33.17	6,682	3,142	16,485	38,105	35.77	14,087,693
lz-206	404,300,903	145,318	328,200	33.36	3,951	2,224	30,853	64,490	38.69	13,997,616
lz-207	404,113,459	107,455	68,259	33.12	6,814	3,203	16,453	37,722	36.12	13,804,064
lz-208	402,730,155	150,230	306,329	33.36	3,748	2,130	32,391	67,441	39.75	13,205,191
lz-209	410,057,899	130,658	371,507	33.33	4,573	2,558	27,076	56,623	31.37	13,083,660
lz-21	396,383,124	183,353	261,018	33.50	2,826	1,645	41,938	86,753	26.09	16,336,478
lz-210	406,535,439	107,477	96,279	33.18	6,781	3,203	16,511	38,000	31.43	14,025,538
lz-211	405,328,455	107,245	77,162	33.18	6,815	3,204	16,218	37,577	30.99	13,835,343
lz-212	397,222,723	120,257	130,226	33.16	5,671	2,720	19,291	44,062	34.79	14,467,934
lz-213	399,076,139	164,087	383,716	33.60	3,350	1,887	35,210	74,138	30.53	15,693,661
lz-214	408,571,532	133,473	280,971	33.31	4,428	2,493	27,900	58,216	34.51	13,234,552
lz-215	407,832,373	129,161	375,753	33.32	4,643	2,572	26,260	55,417	35.50	12,312,849
lz-216	406,755,092	136,540	392,623	33.30	4,303	2,402	28,221	59,397	35.08	11,981,210
lz-217	403,402,922	115,126	225,595	33.15	6,137	2,928	18,025	41,446	35.62	14,402,331
lz-218	402,257,456	165,450	196,694	33.39	3,306	1,903	36,862	76,193	33.16	14,857,686
lz-219	402,540,681	115,360	180,638	33.14	6,194	2,906	17,681	40,986	35.29	14,723,164
lz-22	386,077,362	208,689	207,477	33.69	2,297	1,365	49,880	102,662	32.03	18,778,104
lz-23	414,150,914	112,338	204,995	33.41	5,675	3,082	21,690	46,191	50.30	10,772,515
lz-24	390,047,165	206,952	222,336	33.53	2,348	1,408	50,292	102,319	30.58	17,107,054
lz-25	400,637,450	165,245	206,775	33.43	3,282	1,889	36,552	76,043	25.75	13,876,727
lz-26	386,318,770	228,048	214,633	33.93	2,030	1,210	55,410	114,443	34.16	20,702,415
lz-27	364,491,694	218,813	402,572	34.18	1,982	1,187	54,104	110,476	39.40	19,263,649
lz-28	367,331,110	237,192	200,297	33.91	1,771	1,099	61,192	123,795	28.07	23,079,827
lz-29	383,177,697	222,514	288,556	33.50	2,067	1,271	56,263	113,400	25.14	16,883,428

lz-3	399,547,241	123,555	207,157	33.16	5,355	2,636	20,696	46,796	39.02	9,902,022
lz-30	383,654,035	141,131	131,860	33.31	4,165	2,133	26,011	57,439	22.69	13,505,415
lz-32	372,719,655	232,890	257,743	33.96	1,871	1,130	58,195	119,185	31.41	21,878,792
lz-33	404,392,914	162,928	385,510	33.27	3,360	1,952	36,510	75,235	24.31	12,517,299
lz-34	399,526,074	206,374	221,570	33.82	2,442	1,432	48,035	99,828	27.74	16,423,963
lz-35	382,243,336	237,994	293,130	34.05	1,872	1,136	59,314	121,806	32.40	22,119,836
lz-36	411,830,201	111,845	395,795	33.39	5,609	3,085	22,064	46,493	60.72	10,850,417
lz-37	381,049,243	233,317	308,256	33.59	1,913	1,171	58,831	119,708	34.14	20,544,735
lz-38	394,825,785	189,126	298,570	33.38	2,682	1,589	44,602	91,302	24.31	15,626,161
lz-39	402,378,294	162,090	402,185	34.08	3,462	1,943	34,561	72,454	74.98	15,618,732
lz-4	391,642,335	135,217	159,990	33.18	4,608	2,298	23,848	53,271	29.94	12,352,016
lz-40	398,177,207	144,404	407,049	33.46	3,911	2,196	30,504	63,898	68.95	4,552,414
lz-41	358,443,659	208,998	266,677	34.45	2,084	1,211	48,977	102,685	109.75	21,811,748
lz-42	387,318,222	219,850	278,517	33.45	2,139	1,290	54,229	110,434	26.95	18,616,654
lz-43	385,867,516	218,832	403,870	33.45	2,138	1,304	54,698	110,526	29.54	18,399,448
lz-44	397,713,563	195,118	402,955	33.38	2,603	1,547	46,523	94,807	27.54	14,549,128
lz-45	394,018,394	206,066	280,404	33.49	2,398	1,429	49,529	101,085	22.53	17,722,413
lz-46	397,556,363	171,015	403,577	34.13	3,171	1,793	37,132	77,834	64.33	15,932,522
lz-47	358,255,379	218,530	401,605	34.36	1,936	1,154	53,467	110,250	111.25	22,326,886
lz-48	376,576,777	239,050	157,578	33.63	1,821	1,140	62,518	124,987	28.33	18,961,245
lz-49	397,701,581	157,395	402,037	33.53	3,501	1,997	34,252	71,117	92.28	11,689,677
lz-5	393,308,238	198,870	297,460	33.69	2,529	1,460	45,334	94,916	30.10	17,326,703
lz-50	394,923,210	159,184	234,678	33.49	3,398	1,955	35,303	72,825	85.20	13,375,301
lz-51	384,391,722	191,522	354,939	33.69	2,591	1,506	44,203	91,369	101.25	16,194,616
lz-52	401,358,163	145,765	373,586	33.37	3,855	2,202	31,343	65,211	67.43	12,821,638
lz-53	393,319,559	184,344	288,592	33.36	2,766	1,628	43,021	88,345	19.56	15,449,888
lz-54	403,161,169	161,894	195,029	33.39	3,388	1,952	35,889	74,419	24.92	12,937,409
lz-55	399,940,387	142,055	348,327	33.89	4,194	2,229	27,519	59,631	22.84	14,089,452
lz-56	398,802,113	179,314	229,312	33.29	2,931	1,718	41,573	85,148	23.00	10,997,208
lz-57	388,236,666	219,459	392,335	33.59	2,153	1,302	54,093	110,050	25.73	16,838,866
lz-58	389,758,928	139,178	171,714	33.22	4,394	2,210	24,716	55,194	21.54	12,615,086
lz-59	402,932,600	171,263	213,353	33.32	3,144	1,832	38,924	80,090	25.62	12,776,127
lz-6	392,479,957	178,330	294,520	33.80	2,963	1,652	38,375	81,620	26.19	11,702,253
lz-60	394,334,508	209,300	401,701	33.35	2,337	1,415	51,451	104,074	28.99	15,467,546
lz-61	377,959,063	244,353	291,954	33.59	1,769	1,108	64,179	128,376	39.86	20,967,325
lz-62	382,993,168	239,715	161,911	33.52	1,860	1,152	62,073	124,576	43.57	21,029,316
lz-63	402,338,494	169,304	279,828	33.35	3,190	1,854	38,199	78,783	28.75	11,836,094
lz-64	397,918,938	197,373	260,171	33.49	2,572	1,525	46,818	95,650	26.87	16,614,562
lz-65	388,036,086	215,519	393,568	33.59	2,211	1,326	52,517	107,192	25.88	18,545,841
lz-66	400,316,693	178,595	303,286	33.38	2,948	1,731	41,363	84,786	22.93	12,967,166
lz-67	402,385,864	121,158	204,686	33.22	5,732	2,729	19,112	44,162	27.79	11,527,307
lz-68	385,587,611	135,127	310,222	33.26	4,534	2,255	23,389	52,908	24.62	14,919,714
lz-69	393,512,802	204,445	160,621	33.56	2,433	1,439	48,870	99,865	39.72	16,787,648
lz-7	394,697,928	172,517	379,962	33.74	3,126	1,732	36,502	77,929	28.35	15,925,892
lz-70	398,556,095	125,980	164,133	33.18	5,383	2,561	20,017	46,341	25.22	13,641,382
lz-71	404,075,208	118,024	76,961	33.15	5,997	2,828	18,359	42,458	25.02	12,313,119
lz-72	405,136,021	154,994	303,599	33.34	3,614	2,074	33,982	70,343	30.11	12,279,650
lz-73	382,955,873	224,274	293,885	33.82	2,069	1,197	52,593	110,307	33.04	24,725,771

lz-74	410,213,801	147,845	396,851	33.56	3,962	2,226	30,980	64,868	41.60	11,814,957
lz-75	386,011,522	217,897	402,572	34.04	2,170	1,285	52,502	107,860	57.00	16,194,291
lz-76	409,450,253	142,250	296,557	33.38	4,105	2,308	29,963	62,657	40.21	4,958,916
lz-77	411,040,230	137,096	379,377	33.40	4,383	2,432	28,100	59,088	36.17	9,064,540
lz-78	403,302,928	159,739	402,389	33.49	3,482	1,984	34,803	72,470	32.65	10,490,512
lz-79	379,861,019	220,166	369,010	33.90	2,093	1,234	52,785	109,289	51.72	17,277,024
lz-8	403,573,595	141,781	402,661	33.47	4,094	2,271	29,359	62,003	55.97	12,293,084
lz-80	339,060,291	228,925	240,400	35.00	1,634	991	58,037	119,538	54.84	28,449,070
lz-81	400,060,388	182,303	350,988	33.48	2,998	1,650	38,569	82,159	41.43	16,332,665
lz-82	396,389,553	191,368	401,937	33.57	2,680	1,568	44,396	91,336	51.83	12,119,494
lz-83	395,145,727	182,559	196,125	33.70	2,851	1,658	41,849	86,143	36.69	13,821,888
lz-84	401,508,068	167,295	293,792	33.48	3,250	1,873	37,224	77,090	38.21	13,713,622
lz-85	424,875,772	81,080	330,983	33.23	8,582	4,623	14,680	31,376	28.23	8,998,379
lz-86	404,474,806	162,820	231,560	33.60	3,460	1,927	34,267	72,529	42.34	14,782,721
lz-87	402,161,222	164,501	294,694	33.59	3,367	1,928	36,035	74,545	44.04	14,633,651
lz-88	411,733,364	100,860	107,811	33.10	7,838	3,563	14,222	33,411	23.59	12,907,950
lz-89	400,184,139	180,727	231,551	33.45	2,938	1,700	40,930	84,662	41.63	13,491,903
lz-9	398,963,636	173,041	150,146	33.45	3,087	1,780	38,616	80,352	26.51	14,026,413
lz-90	413,312,513	144,094	235,669	33.57	4,162	2,308	29,562	62,390	39.14	12,129,709
lz-91	392,945,861	193,296	402,913	33.56	2,617	1,534	45,004	92,695	50.32	17,189,022
lz-92	410,346,684	136,773	238,241	33.45	4,485	2,438	27,037	57,557	40.93	13,100,634
lz-93	403,184,938	165,749	298,896	33.48	3,369	1,879	35,331	74,533	37.27	15,014,787
lz-94	400,737,735	170,934	270,028	33.43	3,141	1,821	38,573	79,631	38.47	13,466,335
lz-95	409,631,144	133,510	400,119	33.41	4,445	2,501	27,530	57,851	29.73	12,892,844
lz-96	414,604,111	131,769	210,303	33.32	4,590	2,577	27,405	57,141	38.02	12,152,300
lz-97	414,935,582	126,043	255,994	33.32	4,895	2,706	25,306	53,440	34.96	10,533,101
lz-98	391,302,611	185,816	347,067	33.68	2,767	1,599	42,431	87,563	77.11	17,223,893
lz-99	400,558,469	170,009	409,601	33.41	3,185	1,824	37,697	78,386	44.11	13,205,594
Mean	396,561,657	172,661	300,986	33.58	3,380	1,888	38,662	80,403	39.14	15,656,462

Table S8 Description of phenotypic traits for the 217 mungbean accssions.

Trait Related	Trait Name	Trait Abbreviation	Type	Year	Location	Season	Average	Stdev	Min.	Max.	Skewness	Kurtosis	Broad-sense Heritability
Yield	Branch number	BRN	Quantitative	2	2	2	4.14	1.22	1.20	7.70	0.28	2.73	0.61
	Pod length(cm)	PDL	Quantitative	2	2	2	9.25	1.40	5.23	13.50	0.15	2.97	0.95
	Total number of pod	PDTN	Quantitative	2	2	2	33.15	20.63	0.30	115.20	1.02	4.16	0.71
	Pod width(mm)	PDW	Quantitative	2	2	2	4.91	0.51	3.48	6.47	0.09	2.90	0.30
	100 seed-weight(g)	SD100WT	Quantitative	2	2	2	5.23	1.21	1.99	8.46	-0.17	2.89	0.98
	Number of seeds per pod	SDNPPD	Quantitative	2	2	2	11.30	1.12	7.90	14.35	-0.26	2.98	0.79
	Yield per plant(g)	YPPL	Quantitative	2	2	2	15.48	9.85	0.14	49.35	0.96	3.73	0.63
Quality	Crude protein content(g/100g)	CPC	Quantitative	1	2	1	23.92	1.34	20.50	27.80	0.30	3.01	0.58
	Crude starch content(%)	CSC	Quantitative	1	2	1	49.20	1.99	43.57	54.36	-0.13	2.95	0.89
	Vitexin(mg/g)	VITE	Quantitative	1	1	1	2.52	0.66	1.44	5.84	1.26	6.34	--
	Isovitexin(mg/g)	ISOVITE	Quantitative	1	1	1	3.24	0.85	1.84	6.61	1.00	4.70	--
Resistance	Bruchid resistance	BR	Discrete	1	1	1	0.04	0.19	0.00	1.00	4.92	25.16	--
	Resistance to fusarium wilt	RFW	Discrete	1	1	1	4.19	2.77	1.00	9.00	0.28	1.74	--
Color	Bud color	BDC	Discrete	1	1	1	1.85	0.36	1.00	2.00	-1.94	4.78	--
	Flower color	FLC	Discrete	1	1	1	2.77	0.57	1.00	3.00	-2.39	7.26	--
	Pod color	PDC	Discrete	1	1	1	0.03	0.18	0.00	1.00	5.29	29.03	--
	Petiole color	PLC	Discrete	1	1	1	1.81	0.39	1.00	2.00	-1.56	3.45	--
	Seed color	SDC	Discrete	1	1	1	0.27	0.85	0.00	4.00	3.50	14.40	--
	Trilobata leaf color	TLC	Discrete	1	1	1	2.09	0.31	1.00	3.00	2.30	8.46	--
	Young stem color	YSC	Discrete	1	1	1	0.82	0.39	0.00	1.00	-1.66	3.76	--
Plant architecture	Plant height(cm)	PLH	Quantitative	2	2	2	69.91	30.15	24.40	172.27	1.42	4.76	0.76
	Growth habit	GRH	Discrete	2	2	1	1.45	0.59	1.00	3.00	0.91	2.84	0.65
	Hypocotyl plus epicotyl length(mm)	HECL	Quantitative	2	2	2	38.98	22.72	1.23	83.53	-0.28	1.86	0.44
	Maximum leaflet area(mm2)	MLA	Quantitative	2	2	2	8594.00	2289.06	3188.12	16288.01	0.43	2.87	0.86
	Maximum leaflet length(mm)	MLL	Quantitative	2	2	2	124.76	18.46	78.53	171.04	0.13	2.49	0.84
	Maximum leaflet width(mm)	MLW	Quantitative	2	2	2	105.93	16.19	62.32	153.92	0.12	2.56	0.94
	Main stem dimeter(mm)	MSD	Quantitative	2	1	2	9.15	1.72	5.46	14.60	0.70	3.00	0.58
	Node number of main stem	MSNN	Quantitative	2	2	2	11.51	1.57	7.33	17.60	0.54	3.66	0.86
	Pod shape	PDS	Discrete	1	1	1	1.28	0.45	1.00	2.00	0.98	1.95	--
	Trilobata leaf shape	TLS	Discrete	1	1	1	2.65	1.23	1.00	9.00	0.65	5.65	--
Other	Days to first flower(d)	FLD	Quantitative	2	2	2	45.01	6.12	27.67	60.33	-0.20	2.78	0.65
	Seed capsule	SDCG	Discrete	1	1	1	0.40	0.49	0.00	1.00	0.40	1.16	--
	Pod Shattering	POS	Discrete	1	1	1	1.37	0.92	1.00	4.00	2.28	6.53	--

Table S9 Summary of SNP-trait association sites (STAs) in different environoment.

Trait	2017SJZ		2018SJZ		2017ZJK	2018ZJK	Total STAs	Consistent(more than 1 year or season)	Stable(more than one location)	Robust (>15% PVE)
	Spring	Summer	Spring	Summer	Spring	Spring				
BRN	14	14	4	994			1013	13		690
PDL	6	4	12	18	7		43	4		14
PDTN	1	1	1	2	1	3	9			1
PDW	5	5	3	17	134	2	158	4	4	113
SD100WT	10	4	11	1	4	1	26	5		15
SDNPPD	4		1	1	54		60			40
YPPL	4	4	161		5	7	169	8	4	113
CPC				9			9			
CSC				71		89	138		22	29
VITE				3			3			
ISOVITE				136			136			1
BR		600					600			546
RFW				10			10			
BDC						16	16			15
FLC						27	27			15
PDC		148					148			83
PLC						36	36			8
SDC		194					194			92
TLC						21	21			6
YSC		31					31			24
PLH	7	7		1	2	3	19	1		5
GRH		181					181			18
HECL	4	43	5	9	19	206	264	22		195
MLA	11	1			17	4	30	3		2
MLL		6			236		242			176
MLW	4	1			17	2	21	3		3
MSD	231	1					232			15
MSNN	9	6			3	3	21			1
PDS						8	8			
TLS										
FLD	89	1	1	10	7	14	119	3	1	23
SDCG		4					4			2
POS						18	18			15

Table S10 Summary of gene PAV-trait association (GPTA) in different environoment.

Trait	2017SJZ		2018SJZ		2017ZJK	2018ZJK	Total	Consistent(more than 1 year or season)	Stable(more than one location)	Robust (>15% PVE)
	Spring	Summer	Spring	Summer	Spring	Spring				
BRN	7	5		6			13	5		7
PDL	5		1	1	3		10			
PDTN		2					2			1
PDW		8		2			10			
SD100WT	2				1		3			
SDNPPD		1			1		2			
YPPL			1				1			
CPC						1	1			
CSC				6		3	9			
VITE										
ISOVITE										
BR		144					144			132
RFW										
BDC						12	12			12
FLC						15	15			12
PDC		8					8			1
PLC						12	12			12
SDC		9					9			1
TLC						6	6			
YSC		12					12			12
PLH	11	1					12			
GRH		10				1	11			2
HECL		10				1	11			
MLA	1	2					2	1		
MLL	1	1					2			
MLW		3			1		4			
MSD	19						19			5
MSNN	7	5					12			
PDS						1	1			
TLS						2	2			
FLD	2		1	25	4	1	33			10
SDCG										
POS						7	7			