PAB
Pesquisa Agropecuária Brasileira

ISSN 1678-3921

Journal homepage: www.embrapa.br/pab

For manuscript submission and journal contents, access: www.scielo.br/pab

Genetics/ Original Article

# Construction of SNP-based KASP fingerprints and population genetic analysis of 40 core mung bean germplasm resources

Ruihan Chen
Nanyang Normal University, College of Life Sciences, Nanyang, Henan, China.
E-mail: 964234903@qq.com

Jiafei Liu
Nanyang Normal University, College of Life Sciences, Nanyang, Henan, China.
E-mail: 1522374281@qq.com

Ying Li
Nanyang Normal University, College of Life Sciences, Nanyang, Henan, China.
E-mail: lying06070501@163.com

Weili Hu
Nanyang Academy of Sciences, Institute of Agricultural Sciences, Nanyang, Henan, China.
E-mail: huweili2007@126.com

Maomao Jia
Nanyang Academy of Sciences, Institute of Agricultural Sciences, Nanyang, Henan, China.
E-mail: 2285811183@qq.com

Xu Zhu
Nanyang Academy of Sciences, Institute of Agricultural Sciences, Nanyang, Henan, China.
E-mail: zhuxu315@126.com

Qingping Zhao
Nanyang Normal University, College of Life Sciences, Nanyang, Henan, China.
E-mail: 1075064181@qq.com

Lingling Cao
Nanyang Normal University, College of Life Sciences, Nanyang, Henan, China.
E-mail: 20192003@nynu.edu.cn

Shuqiong Yang(✉)
Nanyang Normal University, College of Life Sciences, Nanyang, Henan, China.
E-mail: sqyang@nynu.edu.cn

✉ Corresponding author

Received
June 3, 2025

Accepted
December 4, 2025

**How to cite**
CHEN, R.; LIU, J.; LI, Y.; HU, W.; JIA, M.; ZHU, X.; ZHAO, Q.; CAO, L.; YANG, S. Construction of SNP-based KASP fingerprints and population genetic analysis of 40 core mung bean germplasm resources. **Pesquisa Agropecuária Brasileira**, v.61, e04253, 2026. DOI: https://doi.org/10.1590/S1678-3921.pab2026.v61.04253.

**Abstract** – The objective of this work was to construct a DNA fingerprinting database of 40 core mung bean (*Vigna radiata*) germplasm using the single nucleotide polymorphism (SNP)-based kompetitive allele specific PCR (KASP) markers identified by genotyping-by-sequencing. After a rigorous filtering, 2,987,425 high-quality SNPs with a high ratio of transitions to transversions (1.97) were obtained, which were mainly synonymous or nonsynonymous single-nucleotide variants, located in intergenic regions and 1.0-kb regions upstream of the transcriptional start sites. The principal component analysis, population structural analysis, and genetic evolutionary tree analysis indicated that the 40 evaluated germplasms were best classified into three groups, which, however, were inconsistent with their geographical origins, indicating that the genetic backgrounds of these germplasms are complex and diverse. Through stringent selection criteria, 25,611 of these high-quality SNPs were successfully converted into KASP markers. The mean values of polymorphism information content, minor allele frequency, and heterozygosity of these KASP markers were 0.366, 0.421, and 0.04, respectively. Finally, 50 core KASP markers were filtered out, and eight of these were verified as able to effectively construct the genetic fingerprints of the 40 core mung bean germplasms.

**Index terms**: *Vigna radiata*, DNA fingerprinting, genetic diversity, genotyping-by-sequencing, germplasm identification.

## Construção de impressões genéticas digitais por KASP baseada em SNPs e análise genética populacional de 40 germoplasmas nucleares de feijão-mungo

**Resumo** – O objetivo deste trabalho foi construir um banco de dados de impressões digitais de DNA de 40 germoplasmas nucleares de feijão-mungo (*Vigna radiata*), tendo-se utilizado os marcadores de PCR competitiva alelo-específica (KASP), baseada em polimorfismos de nucleotídeo único (SNPs), identificados por genotipagem por sequenciamento. Após filtragem rigorosa, foram obtidos 2.987.425 SNPs de alta qualidade, com uma alta proporção de transições para transversões (1,97), principalmente variantes de nucleotídeo único (SNVs) sinônimas ou não sinônimas, localizadas em regiões intergênicas e em regiões de 1,0 kb acima do sítio de início da transcrição. A análise de componentes principais, a análise estrutural da população e a análise da árvore evolutiva genética indicaram que os 40 germoplasmas avaliados foram mais bem classificados em três grupos, que, no entanto, foram inconsistentes com suas origens geográficas, o que indica que os antecedentes genéticos desses germoplasmas são complexos e diversos. Por meio de critérios de seleção

rigorosos, 25.611 desses SNPs de alta qualidade foram convertidos com sucesso em marcadores KASP. Os valores médios do conteúdo de informação de polimorfismo, da frequência do alelo menor e da heterozigosidade desses marcadores KASP foram 0,366, 0,421 e 0,04, respectivamente. Por fim, 50 marcadores KASP nucleares foram filtrados, e oito destes foram verificados como capazes de construir efetivamente impressões genéticas digitais dos 40 germoplasmas nucleares de feijão-mungo.

**Termos para indexação**: *Vigna radiata*, impressão genética digital, diversidade genética, genotipagem por sequenciamento, identificação de germoplasma.

## Introduction

Mung bean (*Vigna radiata* L.) has become a widely cultivated legume crop species mainly in Asia because of its short growth period, long sowing period, strong stress resistance, symbiotic nitrogen fixation, and soil improvement ability (Chen et al., 2020; Mwangi et al., 2021).

Molecular markers have been widely used in genetic diversity, kinship, and fingerprinting studies of crop varieties or lines, at different growth and development stages, providing a great convenience for genetic breeding (Li et al., 2023; Huang et al., 2024; Wang et al., 2024). Single nucleotide polymorphisms are recognized as the most mainstream third-generation molecular markers, which are widely used for crop variety identification and map construction because of their wide quantitative distribution, high polymorphisms, and good stability (Kumar et al., 2020; Hsu et al., 2022; Dube et al., 2023). Until present, some high-throughput and low-cost SNP genotyping platforms have been developed, such as the GoldenGate and Infinium platforms, TaqMan by Life Technologies, and Kompetitive Allele Specific PCR (KASP) system (Xing et al., 2024). As a new type of high-throughput SNP genotyping technology, KASP displayed various advantages of high accuracy and site adaptability, relatively low cost, and suitable for high-throughput detection of a large number of samples. KASP genotyping has a variety of applications in genetic diversity and germplasm genotyping analysis for crops (Xing et al., 2024; Liu et al., 2025; Shen et al., 2025).

The construction of DNA fingerprints based on SNP marker technology is important for ensuring varietal specificity and utility for varieties or species authenticity issues. Over the past two decades, breeders have analyzed the genetic diversity of mung bean separately, using RAPDs, SSRs, and SNPs, which have constructed core germplasm resources to lay the foundation for the genetic breeding of mung bean (Zhao et al., 2020). Recently, the SNP markers discovered by genotyping-by-sequencing (GBS) have been successfully used to identify genetic relationships, construct fingerprint databases and genetic maps, isolate and clone diverse genes of *Triticum aestivum*, *Cymbidium ensifolium*, *Populus deltoides*, *Glycine max*, and other species (Jia et al., 2024; Liu et al., 2025; Sharma & Singh, 2025; Shen et al., 2025). However, thus far, no SNP-based KASP fingerprint database for larger mung bean germplasms, mainly from China, has been reported.

The objective of this work was to construct a DNA fingerprinting database of 40 core mung bean germplasms, using the SNP-based KASP markers produced by genotyping-by-sequencing.

## Materials and Methods

Forty core mung bean gemplasms with diversified phenotypic traits were provided by Nanyang Normal University, Nanyang City, Henan Province (Table 1). For each germplasm, three vigorous seedlings were randomly selected, and a total of 5.0 g young leaf samples were immediately frozen, using liquid nitrogen, and stored at -80°C for DNA extraction.

The genomic DNA extraction was performed using the FlaPure Plant DNA Extraction Kit (RE711-50, Genesand Biotech Co., Beijing, China), and the quality and concentration of DNA were determined using a NanoDrop2000 UV spectrophotometer (Thermo Fisher, Massachusetts, USA). The ultrasonic DNA fragments were purified and end-repaired, a single "A" nucleotide was added to the 3' ends of the blunt fragments, then the sequencing adapters were ligated to the A-tailed fragments. Libraries were pooled, size-selected (200 ~ 500 bp) in a 1% agarose gel electrophoresis, column-purified using a PCR purification kit, and amplified for 12 cycles using Phusion DNA polymerase (NEB, Ipswich, MA, USA). The average fragment size was estimated in a Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA), using a DNA1000 chip followed by a second column purification, and library quantification

**Table 1.** Geographical origins and morphological characteristics of 40 core mung bean (*Vigna radiata*) germplasms.

| Grouping | Simple ID | Variety | Stem color | Leaf type | Seed color | Seed shape | 100-seed weight | Origins |
|---|---|---|---|---|---|---|---|---|
| G01 | G017 | Yinggelvdou | Purple | Ovate | Green | Short cylindrical | 5.5 | Shanxi |
| G01 | G020 | Lvdou | Purple | Ovate | Green | Long cylindrical | 5.6 | Shanxi |
| G01 | G052 | Chuandilong | Purple | Ovate | Green | Long cylindrical | 4.8 | Shanxi |
| G01 | G060 | Yanglvdou | Purple | Triangular | Green | Short cylindrical | 4.8 | Shanxi |
| G01 | G086 | Lvdou | Purple | Ovate | Green | Spherical | 6.4 | Neimenggu |
| G01 | G091 | Xiaolvdou | Purple | Ovate | Green | Spherical | 5.1 | Neimenggu |
| G01 | G118 | Lvdou | Purple | Ovate | Green | Spherical | 7.3 | Jilin |
| G01 | G120 | Xiaolvdou | Purple | Triangular | Green | Spherical | 7.2 | Jilin |
| G01 | G141 | Huanglvdou | Green | Ovate | Yellow | Spherical | 6.2 | Henan |
| G01 | G237 | Zhonglv13 | Green | Triangular | Black | Short cylindrical | 6.1 | Beijing |
| G01 | G009 | Lvdou | Purple | Ovate | Green | Short cylindrical | 4.2 | Beijing |
| G01 | G056 | Xiaolvdou | Purple | Triangular | Green | Short cylindrical | 5.9 | Shanxi |
| G01 | G071 | Huanglvdou | Purple | Ovate | Yellow | Spherical | 4.9 | Shanxi |
| G01 | G075 | Yuanlvdou | Purple | Ovate | Green | Long cylindrical | 5.7 | Neimenggu |
| G01 | G095 | Xiaolvdou | Purple | Ovate | Brown | Short cylindrical | 4.5 | Neimenggu |
| G01 | G097 | Dabajiao | Purple | Ovate | Brown | Long cylindrical | 5.3 | Neimenggu |
| G01 | G126 | Yicuojia | Purple | Ovate | Green | Long cylindrical | 6.7 | Liaoning |
| G01 | G132 | Minglvdou | Purple | Triangular | Green | Long cylindrical | 4.6 | Shandong |
| G01 | G238 | Zhonglv3 | Purple | Ovate | Green | Spherical | 4.5 | Beijing |
| G01 | G260 | Xiaomingli | Purple | Ovate | Green | Long cylindrical | 5.7 | Shandong |
| G01 | G276 | Jilv6 | Purple | Ovate | Blue cyan | Long cylindrical | 7.6 | Jilin |
| G01 | M001 | Yeslvdou | Purple | Triangular | Brown | Short cylindrical | 3.2 | Henan |
| G01 | M101 | Lvdou | Green | Ovate | Green | Short cylindrical | 7.5 | Henan |
| G01 | G025 | Lvdou | Purple | Triangular | Green | Short cylindrical | 5.5 | Shanxi |
| G01 | G033 | Lvdou | Purple | Ovate | Green | Short cylindrical | 5.2 | Shanxi |
| G01 | G041 | Lvdou | Purple | Triangular | Yellow | Short cylindrical | 5.6 | Shanxi |
| G01 | G066 | Huanglvdou | Purple | Ovate | Yellow | Short cylindrical | 5.2 | Shanxi |
| G01 | G072 | Xiaolvdou | Purple | Triangular | Green | Short cylindrical | 6.1 | Neimenggu |
| G01 | G105 | Jinlvdou | Purple | Triangular | Yellow | Short cylindrical | 5.3 | Neimenggu |
| G01 | G139 | Zhaijiaolvdou | Purple | Ovate | Green | Short cylindrical | 5.2 | Henan |
| G02 | G159 | V3476 | Green | Triangular | Green | Long cylindrical | 6.2 | Philippines |
| G02 | G161 | VC2917 | Green | Ovate | Green | Long cylindrical | 5.3 | WorldVege |
| G02 | G214 | BERKENx109897 | Purple | Ovate | Green | Spherical | 6.2 | Australia |
| G02 | G215 | Heilvdou | Purple | Ovate | Blue cyan | Long cylindrical | 6.6 | Henan |
| G02 | G219 | Jilv9 | Purple | Ovate | Black | Short cylindrical | 6 | Hebei |
| G02 | G253 | Huanghuashenglvdou | Purple | Triangular | Yellow | Short cylindrical | 5.9 | Hubei |
| G02 | G254 | Lvdou | Purple | Ovate | Green | Spherical | 5.9 | Hubei |
| G03 | M025 | Lvdou | Green | Ovate | Green | Long cylindrical | 9.3 | Henan |
| G03 | G249 | Lvdou | Green | Ovate | Green | Long cylindrical | 5.9 | Anhui |
| G03 | G286 | Sulv16 | Green | triangular | Green | Long cylindrical | 8.1 | Jiangsu |

was performed using PicoGreen (Invitrogen, Carlsbad, CA, USA).

Sample sequencing was performed using the Illumina NovaSeq 6000 PE150 platform (Illumina, San Diego, CA, USA), and the quality control of original sequencing reads was performed using fastqc (Li et al., 2024). Clean reads were obtained by using fastp (-q5 -n5) to remove reads with an unknown base number N < 5, a length of bases < 50%, quality value < 5, connector sequences, and other low-quality sequences. BWA-MEM software (-M -R) (Li & Durbin, 2009) compared clean reads to Vrad_JL7 genome (Liu et al., 2022). Insert size, coverage depth, and variation of each sample were counted and detected by comparing these clean read positions to Vrad_JL7 genome (Liu et al., 2022). The generated same files were converted to bam format using Samtools v1.9 (Li et al., 2009). The duplicate tags were detected using MarkDuplicates from Picard (2025), the high-quality reads were retained for subsequent analysis.

The GATK HaplotypeCaller (2025) was used to call original SNPs and InDels; SNPs were further separated using SelectVariants (2025). SNPs without other variant loci at 50 bp before and after were filtered by python scripts, then the bi-allelic SNPs were filtered using the VariantFiltration (2025) with the average sequencing depth ≥5×, Q≥ 30, minor information integrity ≥ 0.90, MAF≥ 0.05, QD< 2.0, MQ< 40.0, FS> 60.0, SOR> 6.0, MQRankSum < -12.5, and ReadPosRankSum < -8.0. The genome and structure annotation files were used to annotate the variable loci using SnpEff software (eff mode) (Pariasca-Tanaka et al., 2015).

PCA was assessed using SMARTPCA program (–make grm – autosome) implemented in EIGENSOFT software (Gong et al., 2016). The maximum likelihood (ML) tree of 40 mung bean germplasms was constructed using FastTree software (Price et al., 2010). Admixture software was used to analyze the population genetic structure (Jaganathan et al., 2015).

For KASP marker design, upstream and downstream 100 bp sequences of candidate SNPs were aligned to Vrad_JL7 genome for removing non-specific sequences; these unique sequences were selected with the polymorphism information content (PIC) value > 0.35. For each KASP target site, one universal primer and two allele-specific primers were designed as follows: GC content < 60%, 55°C ≤ Tm ≤ 62°C, and PCR product size ≤ 120 bp. Based on the PIC value and distribution frequency, core markers with a high-detection rate and significant polymorphism amounts which could distinguish all germplasms were screened out. For fingerprinting, the optimal combination of markers was calculated using SNPT software (2025). The genotypes of the optimal combination of markers were heatmapped using RStudio (2025). Each row of the heatmap represents one SNP locus, and each column represents one sample where pure genotypes are A/A = green, C/C = yellow, G/G = purple, T/T = blue; heterozygous genotypes are grey; and the no call genotypes are NA = white.

## Results and Discussion

Genome sequencing of 40 core mung bean germplasms (Table 1) yielded 298.94 Gb of clean data, with an average sequencing depth of 16 ×, and Q30 reached more than 95.09% per sample (Table 2). The obtained clean reads were mapped to Vrad_JL7 genome, and the average mapping efficiency reached 99.50% (Table 3). Among 40 core germplasms, a total of 2,987,425 SNP variants were detected and asymmetrically distributed on 11 chromosomes (Figure 1 A). In particular, C/T (18.465%), G/A (18.462%), A/G (14.697%), and T/C (14.653%) transitions were the dominant types, followed by the transversions of T/A (4.993%), A/T (4.990%), C/A (4.672%), G/T (4.670%), T/G (3.919%), A/C (3.917%), C/G (3.287%), and G/C (3.276%) (Figure 1 B). The resulting transition/transversion ratio of 1.97 was higher than those obtained in previous studies on cigar tobacco, cauliflower, and radish, indicating a higher genomic stability of mung bean (Wang et al., 2021; Yang et al., 2022; Xing et al., 2024).

For the distribution of SNPs, 34.58% were located in intergenic regions, from which 3.36% were located in exons, 8.05% were located in introns, 32.67% were located in 1-kb region upstream of transcriptional start site, 18.95% were located in 1-kb region downstream of transcription termination site, 0.82% and 1.40% were located in 5′ and 3′ UTRs, and 0.04% were located in splice junctions (Figure 1 C). Functional annotation of SNPs in the exonic regions showed that 97,757 nonsynonymous SNVs were predicted to cause changes in encoded amino acid, and 68,471 synonymous SNVs were predicted to not change amino acid sequences (Figure 1 D). The ratio of nonsynonymous SNVs to

**Table 2.** Evaluation of sample sequencing data of 40 core mung bean (*Vigna radiata*) germplasms.

| Sample ID | Read number | Base number | Q30 (%) | Q20 (%) | Average Q |
|---|---|---|---|---|---|
| G009 | 23,574,419 | 7,072,325,700 | 93.10 | 97.61 | 35.84 |
| G017 | 27,288,414 | 8,186,524,200 | 92.59 | 97.38 | 35.74 |
| G020 | 27,844,116 | 8,353,234,800 | 93.18 | 97.67 | 35.85 |
| G025 | 22,294,538 | 6,688,361,400 | 93.43 | 97.76 | 35.90 |
| G033 | 21,817,844 | 6,545,353,200 | 92.92 | 97.55 | 35.81 |
| G041 | 24,368,190 | 7,310,457,000 | 93.70 | 97.87 | 35.94 |
| G052 | 24,404,724 | 7,321,417,200 | 93.29 | 97.69 | 35.87 |
| G056 | 22,237,251 | 6,671,175,300 | 93.07 | 97.61 | 35.83 |
| G060 | 22,679,664 | 6,803,899,200 | 93.22 | 97.67 | 35.86 |
| G066 | 26,806,617 | 8,041,985,100 | 93.04 | 97.59 | 35.83 |
| G071 | 24,218,617 | 7,265,585,100 | 93.00 | 97.56 | 35.82 |
| G072 | 25,981,244 | 7,794,373,200 | 93.45 | 97.77 | 35.90 |
| G075 | 23,720,717 | 7,116,215,100 | 93.04 | 97.58 | 35.83 |
| G086 | 22,900,506 | 6,870,151,800 | 93.31 | 97.71 | 35.88 |
| G091 | 23,528,372 | 7,058,511,600 | 92.61 | 97.37 | 35.75 |
| G095 | 26,322,586 | 7,896,775,800 | 93.34 | 97.73 | 35.88 |
| G097 | 31,017,760 | 9,305,328,000 | 94.09 | 98.10 | 36.02 |
| G105 | 30,077,470 | 9,023,241,000 | 93.50 | 97.88 | 35.92 |
| G118 | 28,458,420 | 8,537,526,000 | 94.94 | 98.46 | 36.18 |
| G120 | 26,885,625 | 8,065,687,500 | 94.64 | 98.35 | 36.13 |
| G126 | 26,975,701 | 8,092,710,300 | 94.80 | 98.42 | 36.16 |
| G132 | 25,844,029 | 7,753,208,700 | 94.78 | 98.40 | 36.15 |
| G139 | 29,244,781 | 8,773,434,300 | 95.09 | 98.50 | 36.20 |
| G141 | 26,677,726 | 8,003,317,800 | 94.87 | 98.43 | 36.16 |
| G159 | 25,546,335 | 7,663,900,500 | 94.56 | 98.25 | 36.10 |
| G161 | 26,708,988 | 8,012,696,400 | 94.29 | 98.19 | 36.06 |
| G214 | 30,324,939 | 9,097,481,700 | 94.23 | 98.15 | 36.05 |
| G215 | 23,223,020 | 6,966,906,000 | 94.76 | 98.38 | 36.14 |
| G219 | 22,392,417 | 6,717,725,100 | 94.45 | 98.26 | 36.09 |
| G237 | 26,743,229 | 8,022,968,700 | 94.92 | 98.44 | 36.17 |
| G238 | 26,229,527 | 7,868,858,100 | 94.09 | 98.10 | 36.02 |
| G249 | 22,832,867 | 6,849,860,100 | 94.63 | 98.33 | 36.12 |
| G253 | 21,362,310 | 6,408,693,000 | 93.16 | 97.65 | 35.85 |
| G254 | 25,901,030 | 7,770,309,000 | 92.97 | 97.55 | 35.81 |
| G260 | 21,048,126 | 6,314,437,800 | 93.15 | 97.63 | 35.85 |
| G276 | 21,182,033 | 6,354,609,900 | 93.53 | 97.79 | 35.91 |
| G286 | 21,132,230 | 6,339,669,000 | 92.76 | 97.46 | 35.78 |
| M101 | 20,875,215 | 6,262,564,500 | 93.28 | 97.70 | 35.87 |
| M025 | 21,947,730 | 6,584,319,000 | 92.71 | 97.43 | 35.76 |
| M001 | 23,846,863 | 7,154,058,900 | 93.19 | 97.66 | 35.85 |

**Table 3.** Statistics of sample map results of 40 core mung bean (*Vigna radiata*) germplasms.

| Simple ID | Total reads | Genome size | Cover size | Cover bases | Mapped reads |
|---|---|---|---|---|---|
| G009 | 47,148,838 | 475,345,115 | 450,800,488 | 7,044,466,950 | 46,963,113 (99.61%) |
| G017 | 54,576,828 | 475,345,115 | 452,891,446 | 8,134,794,450 | 54,231,963 (99.37%) |
| G020 | 55,688,232 | 475,345,115 | 453,812,550 | 8,335,155,600 | 55,567,704 (99.78%) |
| G025 | 44,589,076 | 475,345,115 | 450,597,417 | 6,674,461,650 | 44,496,411 (99.79%) |
| G033 | 43,635,688 | 475,345,115 | 451,944,879 | 6,526,128,900 | 43,507,526 (99.71%) |
| G041 | 48,736,380 | 475,345,115 | 452,060,400 | 7,285,110,000 | 48,567,400 (99.65%) |
| G052 | 48,809,448 | 475,345,115 | 443,823,495 | 7,302,820,950 | 48,685,473 (99.75%) |
| G056 | 44,474,502 | 475,345,115 | 449,854,705 | 6,637,264,350 | 44,248,429 (99.49%) |
| G060 | 45,359,328 | 475,345,115 | 448,473,206 | 6,788,175,750 | 45,254,505 (99.77%) |
| G066 | 53,613,234 | 475,345,115 | 448,655,351 | 7,995,794,250 | 53,305,295 (99.43%) |
| G071 | 48,437,234 | 475,345,115 | 451,006,148 | 7,248,456,150 | 48,323,041 (99.76%) |
| G072 | 51,962,488 | 475,345,115 | 453,120,498 | 7,773,748,650 | 51,824,991 (99.74%) |
| G075 | 47,441,434 | 475,345,115 | 452,155,426 | 7,084,035,150 | 47,226,901 (99.55%) |
| G086 | 45,801,012 | 475,345,115 | 452,481,635 | 6,846,651,000 | 45,644,340 (99.66%) |
| G091 | 47,056,744 | 475,345,115 | 449,299,234 | 7,031,404,650 | 46,876,031 (99.62%) |
| G095 | 52,645,172 | 475,345,115 | 447,383,249 | 7,878,674,550 | 52,524,497 (99.77%) |
| G097 | 62,035,520 | 475,345,115 | 454,278,871 | 9,272,801,100 | 61,818,674 (99.65%) |
| G105 | 60,154,940 | 475,345,115 | 453,193,898 | 9,000,900,000 | 60,006,000 (99.75%) |
| G118 | 56,916,840 | 475,345,115 | 452,448,634 | 8,517,967,800 | 56,786,452 (99.77%) |
| G120 | 53,771,250 | 475,345,115 | 450,981,313 | 8,048,484,600 | 53,656,564 (99.79%) |
| G126 | 53,951,402 | 475,345,115 | 450,775,662 | 8,076,573,450 | 53,843,823 (99.80%) |
| G132 | 51,688,058 | 475,345,115 | 452,964,855 | 7,736,809,650 | 51,578,731 (99.79%) |
| G139 | 58,489,562 | 475,345,115 | 448,685,389 | 8,752,159,650 | 58,347,731 (99.76%) |
| G141 | 53,355,452 | 475,345,115 | 452,084,692 | 7,982,408,250 | 53,216,055 (99.74%) |
| G159 | 51,092,670 | 475,345,115 | 455,956,903 | 7,158,604,650 | 47,724,031 (93.41%) |
| G161 | 53,417,976 | 475,345,115 | 456,663,036 | 7,989,445,950 | 53,262,973 (99.71%) |
| G214 | 60,649,878 | 475,345,115 | 458,409,219 | 8,937,229,800 | 59,581,532 (98.24%) |
| G215 | 46,446,040 | 475,345,115 | 459,267,598 | 6,942,817,500 | 46,285,450 (99.65%) |
| G219 | 44,784,834 | 475,345,115 | 461,932,393 | 6,697,775,550 | 44,651,837 (99.70%) |
| G237 | 53,486,458 | 475,345,115 | 456,508,621 | 7,994,700,150 | 53,298,001 (99.65%) |
| G238 | 52,459,054 | 475,345,115 | 458,308,415 | 7,840,173,900 | 52,267,826 (99.64%) |
| G249 | 45,665,734 | 475,345,115 | 453,442,238 | 6,829,180,800 | 45,527,872 (99.70%) |
| G253 | 42,724,620 | 475,345,115 | 455,382,677 | 6,385,736,700 | 42,571,578 (99.64%) |
| G254 | 51,802,060 | 475,345,115 | 455,260,315 | 7,749,223,500 | 51,661,490 (99.73%) |
| G260 | 42,096,252 | 475,345,115 | 453,411,428 | 6,299,560,050 | 41,997,067 (99.76%) |
| G276 | 42,364,066 | 475,345,115 | 451,594,559 | 6,338,908,200 | 42,259,388 (99.75%) |
| G286 | 42,264,460 | 475,345,115 | 451,963,687 | 6,320,329,500 | 42,135,530 (99.69%) |
| M101 | 41,750,430 | 475,345,115 | 452,724,225 | 6,249,852,150 | 41,665,681 (99.80%) |
| M025 | 43,895,460 | 475,345,115 | 449,587,032 | 6,567,358,350 | 43,782,389 (99.74%) |
| M001 | 47,693,726 | 475,345,115 | 452,874,083 | 7,134,176,850 | 47,561,179 (99.72%) |

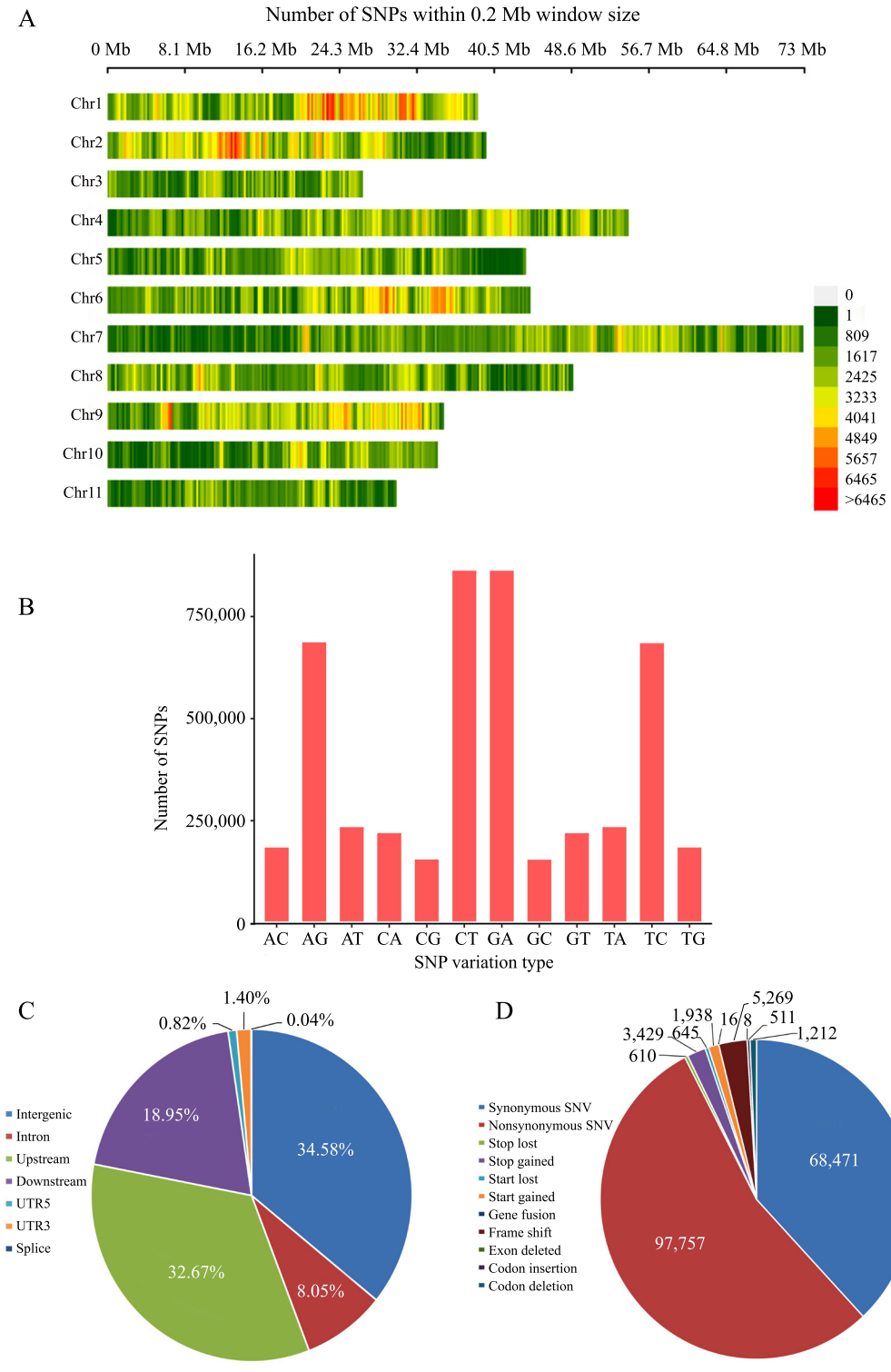


**Figure 1.** Analysis of SNPs in 40 mung bean (*Vigna radiata*) genomes. A, SNP density distribution on each chromosome, for which the horizontal axis represents the chromosome length, and the vertical axis represents the chromosome number; and different colors represent the number of SNPs in different regions. B, Statistics of the variant SNP types, in which the horizontal axis represents the different types of SNP mutations, and the vertical axis represents the number of mutations. C, The positions of the SNPs in the gene structures. D, Annotations of the SNPs in the exons.

synonymous SNVs was 1.43. In addition, 3,429 SNPs were predicted to leading the early gain of a stop codon (stop gain), and 610 SNPs leading to the loss of the stop codon (stop loss), which was significantly higher than these same SNPs contained in sugarcane and cigar tobacco (Wang et al., 2021; Zhang et al., 2022).

A two-dimensional PCA showed that 40 mung bean germplasms were divided into three well-separated groups, while each group contained germplasms from different geographical origins (Figure 2 A). The population structure analysis of these 40 germplasms showed that the lowest cross-validation error rate is K = 3; three obtained groups haven't been completely classified according to their geographical origins (Figure 2 B-D). A maximum likelihood tree showed that 40 germplasms were also clustered into three groups with a high bootstrap value, which were not also completely classified by their geographical origins (Figure 2 E). These clustering results of PCA, population structure analysis, and phylogenetic tree were basically consistent and complemented one another, all of them indicating that the genetic background of these 40 germplasms were complex and diverse. Meanwhile, similar results were also observed by Lavanya et al. (2008), who report that 54 mung bean accessions were divided into two major groups and four subgroups, using random amplified polymorphic DNA (RAPD) markers, which also did not show any significant correlation between their genetic divergence and geographical distribution. The diversity evaluation differentiated mung bean accessions between cultivated (157) and wild (1) forms into nine subgroups, which also showed that no geographical distinctions existed among mung bean accessions collected in China (Chen et al., 2015). According to these morphological diversity analysis, the stem color was the most stable while the 100-seed weight was the most variable trait among these 40 core germplasms. These observed morphological characteristics did not rhyme with the genetic clustering or geographical origins of 40 core germplasms, which may be due to the complex genetic backgrounds and smaller environmental selection of 40 germplasms.

A total of 25,611 high-quality SNPs were successfully transformed into KASP markers with 46.54% transformation rate, which displayed good polymorphism and variety discrimination ability. KASP markers were distributed as follows: about 7,493 (29.26%) in the intergenic region; 7,147 (27.91%) in the intron region; 4,199 (16.40%) in the exon region; and 3,008 (11.74%) and 3,000 (11.71%) were distributed in the gene regulatory region at 2 kb upstream and downstream of the gene, respectively (Figure 3 A). The PIC values of 25,611 SNPs ranged from 0.350 to 0.375 (mean value = 0.366); MAF values ranged from 0.346 to 0.500 (mean value = 0.421); and observed heterozygosity ranged from 0 to 1.00 (mean value = 0.04) (Figure 3 B, C, D). A combination of chromosomal location, PIC, MAF, observed heterozygosity, and missing values of 40 mung bean genotypes were used to obtain 50 high-quality core SNP-based KASP markers. Owing to the biallelic nature of the SNP, the PIC value of SNPs was limited from 0 to 0.5 (Eltaher et al., 2018). In comparison with the mean values of genomic SSRs (0.40) and EST-based SSRs (0.30) markers in mung bean, the mean PIC value (0.421) of 50 core KASP markers was considered the highest one (Chen et al., 2015). The threshold of MAF significantly affected fingerprinting and inferred population structure; SNPs with higher MAF values tended to be more polymorphic (Linck & Battey, 2019). Sixty-six percent of core KASP markers had MAF values between 0.4 and 0.5 (mean value = 0.445), suggesting the polymorphic practicality of these core markers used in the present study. According to the saturation curve analysis of marker identification efficiency, eight core candidate KASP markers were finally selected for their higher PIC (0.351 ~ 0.375), higher MAF (0.350 ~ 0.500), lower observed heterozygosity (0 ~ 0.375), and no missing values (Figure 4 A). Detailed information on eight core KASP markers (marker name, location, variant type, and primer sequences) was provided (Table 4, Figure 4 B). Based on these eight newly developed core KASP markers, fingerprints of 40 mung bean germplasms were constructed (Figure 4 C). Different from most previous studies focusing on morphology and quality evaluation of different mung bean germplasms, the SNP fingerprinting using KASP technology provided a more accurate, rapid, convenient, and efficient method for the diversity identification of mung bean germplasms (Chen et al., 2020; Basnet et al., 2024).

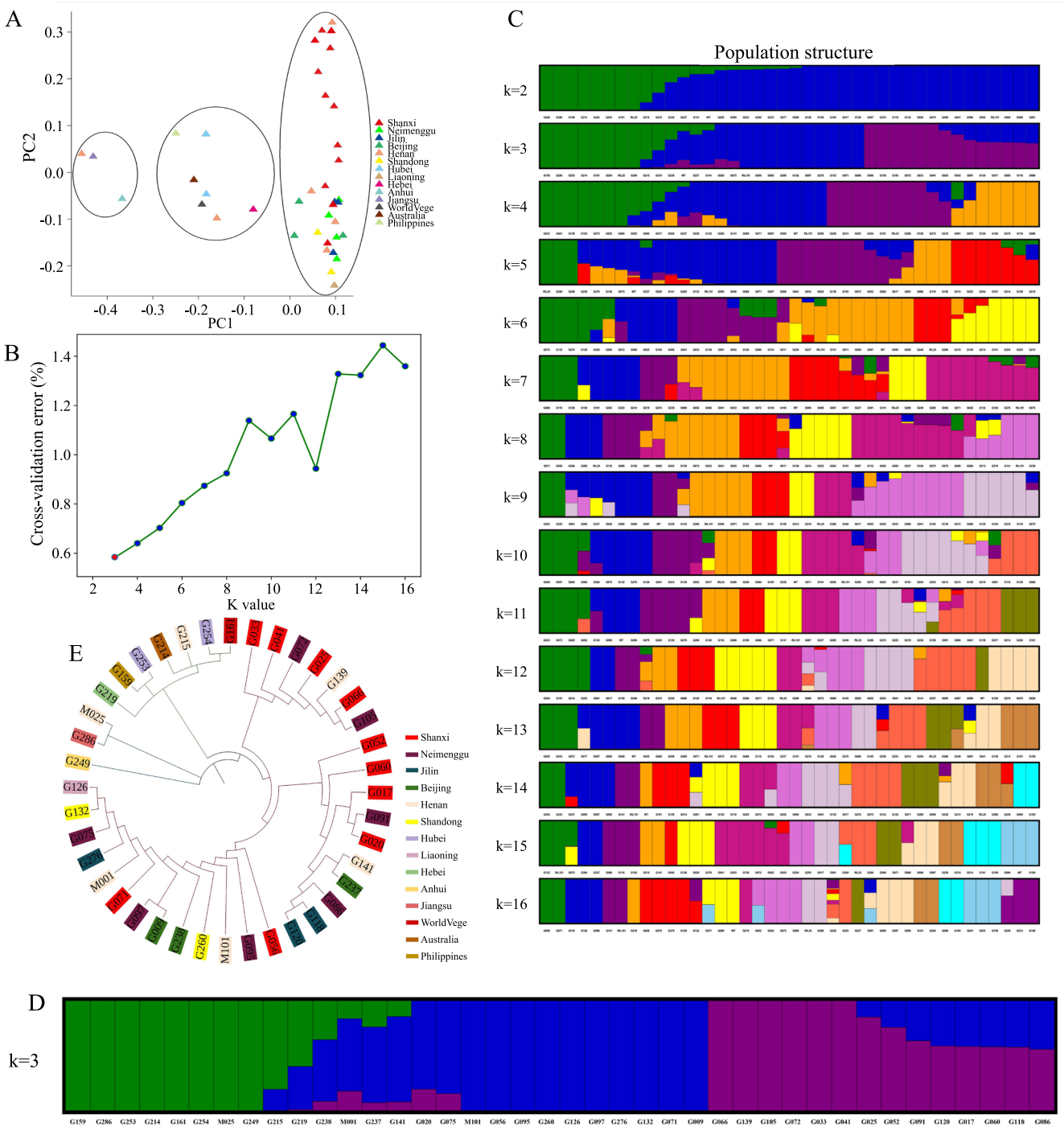


**Figure 2.** Phylogenetic and population genetic analysis of 40 core mung bean (*Vigna radiata*) germplasms based on polymorphic SNP loci. A, a two-dimensional diagram of PCA. B, Cross-validation error rates corresponding to different K values. C, Population structure of 40 core mung bean germplasms at different K values. D, Population structure of 40 core mung bean germplasms at K = 3. E) An unrooted maximum likelihood phylogenetic tree.

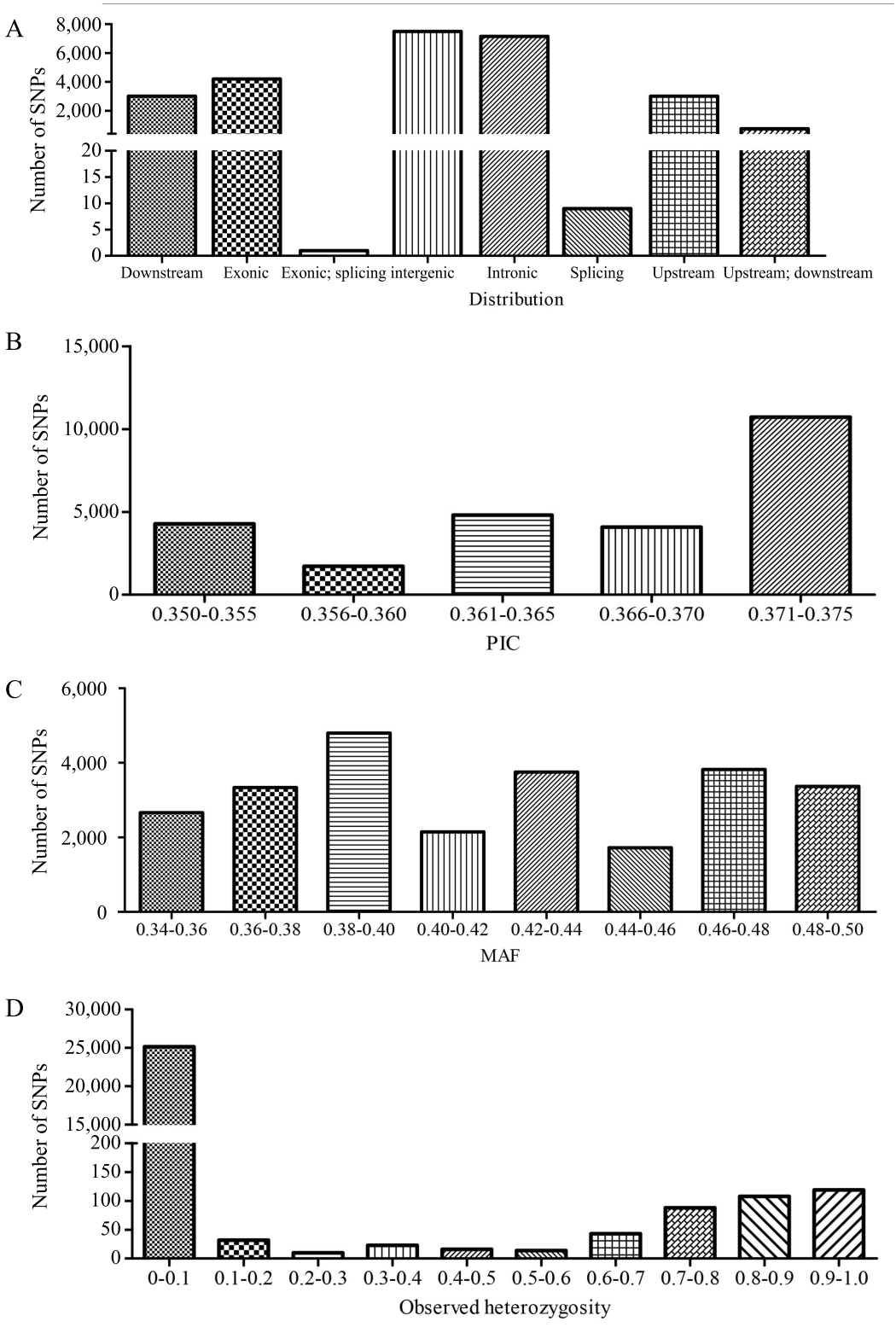


**Figure 3.** Population genetic analysis of mung bean (*Vigna radiata*) germplasms based on SNP loci: A, distribution; B, PIC - polymorphism information content; C, MAF - minor allele frequency; and D, observed heterozygosity values for 25,611 SNP markers based on data from 40 core mung bean germplasms.

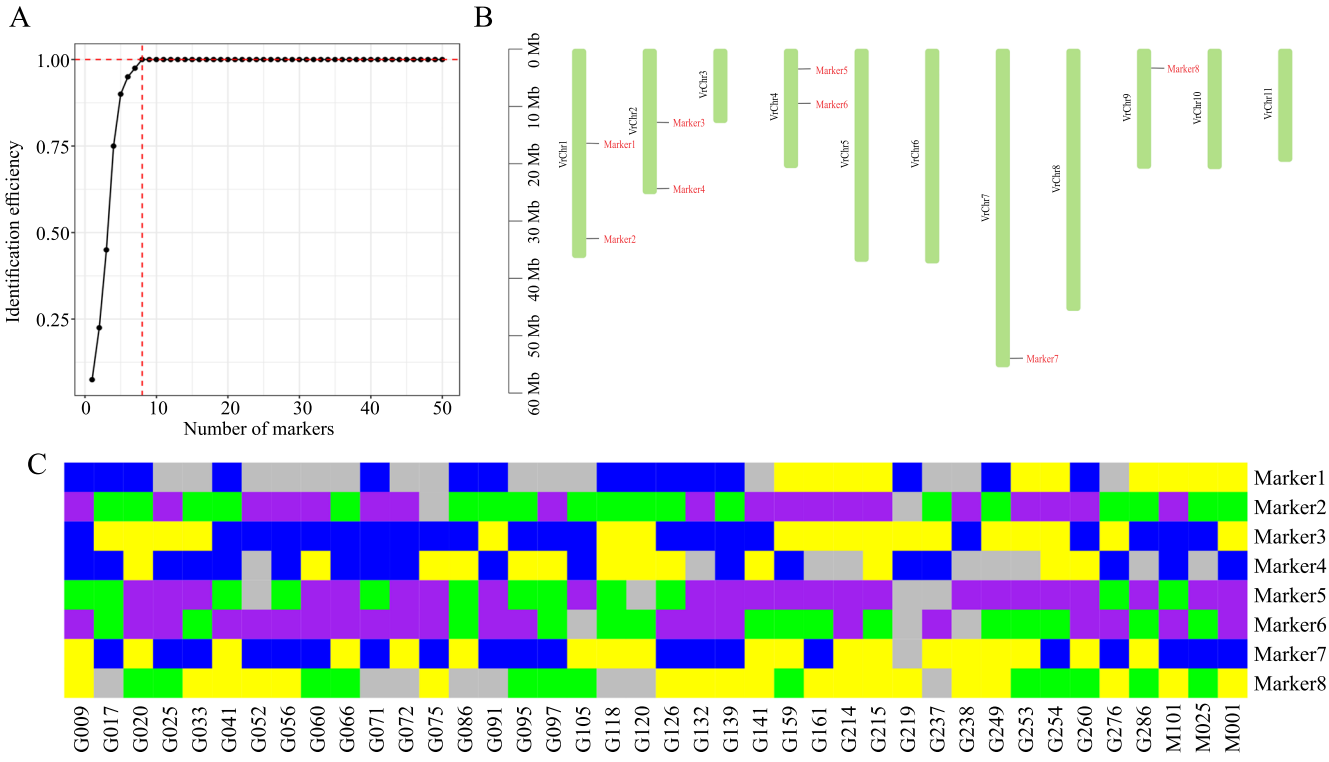


**Figure 4.** Fingerprint analysis of 40 core mung bean (*Vigna radiata*) germplasms: A, identification efficiency of the combined SNP markers; B, distribution of 8 core markers on mung bean chromosomes; C, fingerprints of the 40 core mung bean germplasms. Each row represents a genome, and each column represents a sample.

**Table 4.** KASP primer names, positions, variant types, and sequences for the eight core SNP markers.

| ID | Primer sequences (5' ~ 3') | Variant type |
|---|---|---|
| Marker1 | F: ATCCCCAACCTCGCCCTTAAATC[C\|T]<br>R: AGGTCTGCTGTACTAGTTTCGTG | C/T |
| Marker2 | F: GATCAGAAGCTATACTGTATAAAATCCCAC<br>R: GGTGTTAAAGTCAAACCTGTAGACACAAC[T\|C] | A/G |
| Marker3 | F: AACAGATTGACAATGGTGTTGTTATTCC<br>R: TTTTTCCTCCCACAAAGTATCATTCCATC[G\|A] | C/T |
| Marker4 | F: CCTTGGCTCCGAAATCATCAACA[T\|C]<br>R: TAATAGAGTTGCCCCTCGTGGAA | T/C |
| Marker5 | F: CTTTCTTTTTCCGACAATTTGTGCAGAG[G\|A]<br>R: GGGTAAGGATTTTCCGCTTAATAAGAAT | G/A |
| Marker6 | F: GCCTCACCAATCATTTCAACCTTGTAC[A\|G]<br>R: AGGAGGATGAAAATTCCAATAGTCAGA | A/G |
| Marker7 | F: AAATATAGCTTCTGTTCACACTCTGGG<br>R: AACCATAGCATCCTTCAACTTATGCCA[G\|A] | C/T |
| Marker8 | F: AGGATCGCTTTCAGTGTCTTCAAG<br>R: AAGATGACATTCCGATGTATAGCCCTTT[G\|T] | C/A |

## Conclusions

1. An integrated genotyping-by-sequencing and competitive allele specific PCR (KASP) technology is a powerful tool for investigating population structure and genetic diversity of mung bean (*Vigna radiata*) germplasms.

2. SNP-based KASP fingerprints are efficient for distinguishing 40 core mung bean germplasms.

## References


BASNET, B.; UPRETI, U.; THAPALIYA, K.P. Genotypic variations in postfertility traits and yield components of mung bean (*Vigna radiata* (L.) R. Wilczek) germplasms in Chitwan, Nepal. **Heliyon**, v.10, e39226, 2024. DOI: https://doi.org/10.1016/j.heliyon.2024.e39226.

CHEN, H.; HU, L.; YANG, Y.; HAO, X.; LI, S.; WANG, S.; WANG, L.; CHENG, X. Evaluation and genetic diversity analysis of agronomic traits and bruchid resistance using 481 worldwide mung bean germplasms. **Journal of Plant Genetic Resources**, v.21, p.549-559, 2020. DOI: https://doi.org/10.13430/j.cnki.jpgr.20190717003.

CHEN, H.; QIAO, L.; WANG, L.; WANG, S.; BLAIR, M.W.; CHENG, X. Assessment of genetic diversity and population structure of mung bean (*Vigna radiata*) germplasm using EST-based and genomic SSR markers. **Gene**, v.566, p.175-183, 2015. DOI: https://doi.org/10.1016/j.gene.2015.04.043.

DUBE, S.P.; SIBIYA, J.; KUTU, F. Genetic diversity and population structure of maize inbred lines using phenotypic traits and single nucleotide polymorphism (SNP) markers. **Scientific Reports**, v.13, art.17851, 2023. DOI: https://doi.org/10.1038/s41598-023-44961-3.

ELTAHER, S.; SALLAM, A.; BELAMKAR, V.; EMARA, H.A.; NOWER, A.A.; SALEM, K.F.M.; POLAND, J.; BAENZIGER, P.S. Genetic diversity and population structure of F3:6 Nebraska winter wheat genotypes using genotyping-by-sequencing. **Frontiers in Genetics**, v.9, art.76, 2018. DOI: https://doi.org/10.3389/fgene.2018.00076.

GONG, D.; HUANG, L.; XU, X.; WANG, C.; REN, M.; WANG, C.; CHEN, M. Construction of a high-density SNP genetic map in flue-cured tobacco based on SLAF-seq. **Molecular Breeding**, v.36, art.100, 2016. DOI: https://doi.org/10.1007/s11032-016-0514-7.

HAPLOTYPECALLER. Available at: <https://gatk.broadinstitute.org/hc/en-us/articles/360042913231-HaplotypeCaller>. Accessed on: Nov. 27 2025.

HSU, Y.-M.; WANG, S.-S.; TSENG, Y.-C.; LEE, S.-R.; FANG, H.; HUNG, W.-C.; KUO, H.-I.; DAI, H.-Y. Assessment of genetic diversity and SNP marker development within peanut germplasm in Taiwan by RAD-seq. **Scientific Reports**, v.12, art.14495, 2022. DOI: https://doi.org/10.1038/s41598-022-18737-0.

HUANG, H.; HE, S.; ZHENG, X.; SHI, D.; BAI, P.; ZHAO, Y.; YU, J.; NIU, X. Simple sequence repeat marker-based genetic diversity and chemical composition analysis of ancient *Camellia sinensis* in Jiulong County, Sichuan Province, China. **Genes**, v.15, art.1317, 2024. DOI: https://doi.org/10.3390/genes15101317.

JAGANATHAN, D.; THUDI, M.; KALE, S.; AZAM, S.; ROORKIWAL, M.; GAUR, P.M.; KISHOR, P.B.; NGUYEN, H.; SUTTON, T.; VARSHNEY, R.K. Genotyping-by-sequencing based intra-specific genetic map refines a ‘’QTL-hotspot” region for drought tolerance in chickpea. **Molecular Genetics and Genomics**, v.290, p.559-571, 2015. DOI: https://doi.org/10.1007/s00438-014-0932-3.

JIA, Q.; HU, S.; LI, X.; WEI, L.; WANG, Q.; ZHANG, W.; ZHANG, H.; LIU, X.; CHEN, X.; WANG, X.; CHEN, H. Identification of candidate genes and development of KASP markers for soybean shade-tolerance using GWAS. **Frontiers in Plant Science**, v.15, art.1479536, 2024. DOI: https://doi.org/10.3389/fpls.2024.1479536.

KUMAR, D.; CHHOKAR, V.; SHEORAN, S.; SINGH, R.; SHARMA, P.; JAISWAL, S.; IQUEBAL, M.A.; JAISWAR, A.; JAISRI, J.; ANGADI, U.B.; RAI, A.; SINGH, G.P.; KUMAR, D.; TIWARI, R. Characterization of genetic diversity and population structure in wheat using array based SNP markers. **Molecular Biology Reports**, v.47, p.293-306, 2020. DOI: https://doi.org/10.1007/s11033-019-05132-8.

LAVANYA, G.R.; SRIVASTAVA, J.; RANADE, S.A. Molecular assessment of genetic diversity in mung bean germplasm. **Journal of Genetics**, v.87, p.65-74, 2008. DOI: https://doi.org/10.1007/s12041-008-0009-3.

LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics**, v.25, p.1754-1760, 2009. DOI: https://doi.org/10.1093/bioinformatics/btp324.

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R. 1000 genome project data processing subgroup, the sequence alignment/map format and SAMtools. **Bioinformatics**, v.25, p.2078-2079, 2009. DOI: https://doi.org/10.1093/bioinformatics/btp352.

LI, J.; CHANG, X.; HUANG, Q.; LIU, P.; ZHAO, X.; LI, F.; WANG, Y.; CHANG, C. Construction of SNP fingerprint and population genetic analysis of honeysuckle germplasm resources in China. **Frontiers in Plant Science**, v.14, art.1080691, 2023. DOI: https://doi.org/10.3389/fpls.2023.1080691.

LI, Y.D.; LUO, X.B.; PENG, X.; YANG, G.Q.; JIN, Y.Y.; ZU, G.D.; TIAN, H.; ZHANG, W.P. Development of SNP and InDel markers in radish and their association with phenotypic characters. **Acta Agriculturae Zhejiangensis**, v.36, p.1055-1066, 2024. DOI: https://doi.org/10.3969/j.issn.1004-1524.20230663.

LINCK, E.; BATTEY, C.J. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. **Molecular Ecology Resources**, v.19, p.639-647, 2019. DOI: https://doi.org/10.1111/1755-0998.12995.

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.; TIAN,

J. High-quality genome assembly and pan-genome studies facilitate genetic discovery in mung bean and its improvement. **Plant Communications**, v.3, art.100352, 2022. DOI: https://doi.org/10.1016/j.xplc.2022.100352.

LIU, C.; YAN, J.; ZHANG, Z.; PEI, L.; LI, C.; ZHANG, X.; SHI, S. Genetic variation analysis and development of KASP marker for leaf area and hight in southern-type *Populus deltoides*. **Plants (Basel)**, v.14, art.330, 2025. DOI: https://doi.org/10.3390/plants14030330.

MARKDUPLICATES (PICARD). Available at: <https://gatk.broadinstitute.org/hc/en-us/articles/360037052812-MarkDuplicates-Picard>. Accessed on: Nov. 27 2025.

MWANGI, J.W.; OKOTH, O.R.; KARIUKI, M.P.; PIERO, N.M. Genetic and phenotypic diversity of selected Kenyan mung bean (*Vigna radiata* L. Wilckzek) genotypes. **Journal of Genetic Engineering and Biotechnology**, v.19, art.142, 2021. DOI: https://doi.org/10.1186/s43141-021-00245-9.

PARIASCA-TANAKA, J.; LORIEUX, M.; HE, C.; MCCOUCH, S.; THOMSON, M.J.; WISSUWA, M. Development of a SNP genotyping panel for detecting polymorphisms in *Oryza glaberrima*/*O. sativa* interspecific crosses. **Euphytica**, v.201, p.67-78, 2015. DOI: https://doi.org/10.1007/s10681-014-1183-4.

PRICE, M.N.; DEHAL, P.S.; ARKIN, A.P. FastTree 2 - approximately maximum-likelihood trees for large alignments. **PLoS ONE**, v.5, e9490, 2010. DOI: https://doi.org/10.1371/journal.pone.0009490.

RSTUDIO. Available at: <https://posit.co/products/open-source/rstudio-server/>. Accessed on: Nov. 27 2025.

SELECTVARIANTS. Available at: <https://gatk.broadinstitute.org/hc/en-us/articles/35967603979419-SelectVariants>. Accessed on: Nov. 27 2025.

SHARMA, J.S.; SINGH, L. Developing breeder-friendly KASP markers for rust *R* genes. **Methods in Molecular Biology**, v.2898, p.273-280, 2025. DOI: https://doi.org/10.1007/978-1-0716-4378-5_17.

SHEN, B.; SHEN, A.; TAN, Y.; LIU, L.; LI, S.; TAN, Z. Development of KASP markers, SNP fingerprinting and population genetic analysis of *Cymbidium ensifolium* (L.) Sw. germplasm resources in China. **Frontiers in Plant Science**, v.15, art.1460603, 2025. DOI: https://doi.org/10.3389/fpls.2024.1460603.

SNPT SOFTWARE. Available at: <http://www.shigatox.net/stec/cgi-bin/snpt>. Accessed on: Nov. 27 2025.

VARIANTFILTRATION. Available at: <https://gatk.broadinstitute.org/hc/en-us/articles/360041850471-VariantFiltration>. Accessed on: Nov. 27 2025.

WANG, Y.; LV, H.; XIANG, X.; YANG, A.; FENG, Q.; DAI, P.; LI, Y.; JIANG, X.; LIU, G.; ZHANG, X. Construction of a SNP fingerprinting database and population genetic analysis of cigar tobacco germplasm resources in China. **Frontiers in Plant Science**, v.12, art.618133, 2021. DOI: https://doi.org/10.3389/fpls.2021.618133.

WANG, Z.; ZENG, H.; QIN, X.; LI, Z.; TONG, Y.; LIU, K.; LI, Q. Genetic diversity and fingerprinting of *Elaeagnus angustifolia* based on SSR molecular markers. **Chinese Journal of Biotechnology**, v.40, p.3530-3547, 2024. DOI: https://doi.org/10.13345/j.cjb.240284.

XING, X.; HU, T.; WANG, Y.; LI, Y.; WANG, W.; HU, H.; WEI, Q.; YAN, Y.; GAN, D.; BAO, C.; WANG, J. Construction of SNP fingerprints and genetic diversity analysis of radish (*Raphanus sativus* L.). **Frontiers in Plant Science**, v.15, art.1329890, 2024. DOI: https://doi.org/10.3389/fpls.2024.1329890.

YANG, Y.; LYU, M.; LIU, J.; WU, J.; WANG, Q.; XIE, T.; LI, H.; CHEN, R.; SUN, D.; YANG, Y.; YAO, X. Construction of an SNP fingerprinting database and population genetic analysis of 329 cauliflower cultivars. **BMC Plant Biology**, v.22, art.522, 2022. DOI: https://doi.org/10.1186/s12870-022-03920-2.

ZHANG, H.; LIN, P.; LIU, Y.; HUANG, C.; HUANG, G.; JIANG, H.; XU, L.; ZHANG, M.; DENG, Z.; ZHAO, X. Development of SLAF-sequence and multiplex SNaPshot panels for population genetic diversity analysis and construction of DNA fingerprints for sugarcane. **Genes**, v.13, art.1477, 2022. DOI: https://doi.org/10.3390/genes13081477.

ZHAO, Y.; WANG, Y.; WANG, L.; ZHANG, D. Molecular identification of mung bean accessions (*Vigna radiata* L.) from Northeast China using capillary electrophoresis with fluorescence-labeled SSR markers. **Food and Energy Security**, v.9, e182, 2020. DOI: https://doi.org/10.1002/fes3.182.

## Author contributions

**Ruihan Chen**: formal analysis (lead), writing - original draft (lead); **Jiafei Liu**: formal analysis (equal); **Ying Li**: formal analysis (equal), writing - original draft (equal); **Weili Hu**: formal analysis (equal); **Maomao Jia**: formal analysis (equal); **Xu Zhu**: writing - review & editing (equal); **Qingping Zhao**: writing - review & editing (equal); **Lingling Cao**: writing - review & editing (equal); **Shuqiong Yang**: methodology, review & editing (equal).

**Chief editor**: Edemar Corazza
**Edited by**: Mírian Baptista

## Data availability statement

Data available: genome-sequencing data reported in this study have been deposited in the China National GeneBank Database (https://db.cngb.org/cnsa/), under accession number sub072248.

## Declaration of use of AI technologies

No generative artificial intelligence (AI) was used in this study.

## Conflict of interest statement

The authors declare no conflicts of interest.

## Acknowledgments

This research was supported by the Natural Science Foundation projects of Nanyang Normal University (2019ZX027, 2022PY014, and 2024QN009), Natural Science Foundation of Henan province (232300420213), Key Scientific Research Project of Higher Education of Henan Province (24A210021), and Nanyang Science and Technology Program (23KJGG293 and 23KJGG264), in China.

## Disclaimer/Publisher's note: