Development of a core set of single nucleotide polymorphism markers for genetic diversity analysis and cultivar fingerprinting in cowpea

Xinyi Wu | Baogen Wang | Suqi Wu | Sujuan Li | Yan Zhang | Ying Wang | Yanwei Li | Jian Wang | Xiaohua Wu | Zhongfu Lu | Guojing Li

1Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, Hangzhou, China
2Public Laboratory, Zhejiang Academy of Agricultural Sciences, Hangzhou, China
3Institute of Vegetables, Guangdong Academy of Agricultural Sciences, Guangzhou, China
4State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Zhejiang Academy of Agricultural Sciences, Hangzhou, China

Correspondence
Guojing Li, Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China.
Email: ligj@zaas.ac.cn

Funding information
National Key R&D Program of China, Grant/Award Number: 2017YFE0114500; Key R&D Program of Guangdong Province, Grant/Award Number: 2020B020220002; Major Science and Technology Project of Plant Breeding in Zhejiang Province, Grant/Award Number: 2016C02051-7-3

Abstract
Cowpea is an important legume crop worldwide. However, the diversity and relationship of current accessions remain unclear due to a lack of robust genetic information. A set of 40,089 single nucleotide polymorphisms (SNPs) adapted to the Kompetitive Allele Specific PCR (KASP) SNP genotyping platform were developed from the existing Illumina Cowpea iSelect Consortium Array for cowpea germplasm genetic diversity assessment and variety identification. Using the genotypic data of 299 cowpea accessions obtained from this SNP assay, a precore pool of 434 SNPs and 50 informative core SNPs was selected and validated for use in future genetic diversity analyses of cowpea germplasm. By further genotyping 75 commercial cultivars using the set of 50 core SNPs, pairwise genotype alignment and dendrogram analysis both showed that each cultivar could be uniquely identified by this core set of markers. The KASP markers developed in this study provide a flexible marker resource for both basic and applied research of cowpea. The precore SNP pool and the core SNP set also provide valuable tools for genetic variation assessment and variety rights protection for cowpea breeders.

KEYWORDS
cowpea, cultivar identification, genetic diversity, KASP, SNP

1 INTRODUCTION

Cowpea (Vigna unguiculata L. Walp.) (2n = 2x = 22, family Fabaceae) is an economically important warm season legume crop in the tropics and subtropics (Singh, 2014). Originated in Africa, cowpea was domesticated into two main cultivated subspecies, grain-type common cowpea (V. unguiculata L. Walp. ssp. unguiculata) and vegetable-type asparagus bean (V. unguiculata L. Walp. ssp. sesquipedalis). Common cowpea is mainly distributed in Africa, southern Europe, the southern United States, and Central and South America; its dry grains are used as a staple food for human consumption and fodder for livestock. Asparagus bean, also known as long bean or yardlong bean, is mainly cultivated in East and South Asian countries such as China, Japan, Thailand, and Malaysia; its tender green pods are used as vegetables (Timko et al., 2007). At present, almost all cowpea varieties in the seed markets are inbreds developed by conventional breeding. An increasing number of new varieties are being released to the market that have very similar phenotypes. Thus,
traditional distinctness, uniformity, and stability (DUS) testing faces a big challenge to distinguish them due to the limited number of traits evaluated and inconsistency of these traits under different environmental conditions (Jamali et al., 2019). Thus, to facilitate the rapid release of cowpea varieties and at the same time secure breeders' intellectual rights, new technologies are urgently needed to assist cowpea breeding and intellectual right protection.

With the advancement of crop genomic resources, molecular markers have been widely used in genetic variation assessment, molecular breeding, DNA fingerprinting, and cultivar rights protection (Gao et al., 2016; Su et al., 2018; Varshney et al., 2008; Yang et al., 2019; Zhang et al., 2012). In cowpea, various molecular markers, such as amplified fragment length polymorphism (AFLP) (Boukar et al., 2004; Coulibaly et al., 2002), simple sequence repeat (SSR) or microsatellite (Xu et al., 2010), and single nucleotide polymorphism (SNP) markers (Muchero et al., 2009; Xu et al., 2011), have been developed. Among them, the Illumina Cowpea iSelect Consortium Array (Muñoz-Amatriain et al., 2017) is an important landmark. This array contained 51,128 SNPs derived from whole genome sequences (WGS) of 37 diverse cowpea accessions. These SNPs are uniformly distributed across the cowpea genome and have high representation of cowpea genetic variation. Thus, they provide an ideal resource for cowpea molecular breeding and new variety protection. This SNP array has been successfully used in linkage map construction, genetic diversity analysis of cowpea germplasm, and QTL dissection for important traits (Muñoz-Amatriain et al., 2017; Wu et al., 2018; Xu et al., 2017).

However, this 51k iSelect array has not been widely accepted by world cowpea breeders. Its relatively high cost and the co-segregation of most markers on a specific genetic map make it difficult for breeders to obtain useful SNPs. For most cowpea researchers and breeders, an ideal SNP genotyping platform should be highly adaptable and cost-effective. To date, multiple high-throughput SNP genotyping platforms have been developed, including Illumina GoldenGate (Fan et al., 2003), Infinium (Steemers & Gunderson, 2010), TaqMan technology (Livak et al., 1995), and Kompetitive Allele Specific PCR (KASP) platforms (KBiosciences, www.kbioscience.co.uk). Among them, the Illumina GoldenGate and Infinium platforms require the design and ordering of special beadchips (such as Cowpea iSelect Consortium Arrays); the price for a single chip is expensive, and the SNPs anchored on the chip are fixed. TaqMan and KASP technologies are both flexible and compatible SNP genotyping platforms based on PCR assays. Unlike TaqMan, KASP genotyping requires only two allele-specific PCR primers and one other locus-specific PCR primer, which should be more cost-effective. Thus, KASP technology has been widely used for SNP genotyping in many crops, including wheat, rice, common bean, chickpea, and Brassica rapa (Cortés et al., 2011; Gao et al., 2016; Hiremath et al., 2012; Li et al., 2019; Yang et al., 2019).

To promote the application of KASP-SNP markers in cowpea breeding, in the present study, we converted all 51,128 SNPs from the Illumina Cowpea iSelect Consortium Array into KASP-SNPs and then constructed a set of core SNPs under the KASP genotyping platform. The core SNP set in this study will provide a useful resource for genetic variation assessment of cowpea germplasm resources, DNA fingerprinting of cowpea varieties, and rapid identification of seed quality.

## 2 MATERIALS AND METHODS

### 2.1 Plant materials

Seventy-five cowpea cultivars were used for core SNP screening and DNA fingerprinting. These cultivars were developed by commercial breeding companies or public vegetable research institutions in China.

### 2.2 KASP primer development

The sequence information of SNPs in the Cowpea iSelect Consortium Assay was obtained from Muñoz-Amatriain et al. (2017). Based on the SNP sequences, two allele-specific forward primers and one common reverse primer were designed for each SNP by an independent primer design program following the LGC proprietary Kraken™ software system (https://www.lgcgroup.com/cn/our-science/genomics-solutions/genotyping/kasp-genotyping-chemistry/reagents/kbd/#. WubvXf4h1hg).

### 2.3 Core SNP identification

The genotypic data of 299 cowpea accessions (84 grain-type cowpea and 215 vegetable-type cowpea) identified by the Cowpea iSelect Consortium Assay (Xu et al., 2017) were retrieved to screen core SNPs following three steps. First, SNPs that were difficult to convert into KASP markers or that were not anchored onto the consensus map (Muñoz-Amatriain et al., 2017) or the “ZZ” linkage map (Xu et al., 2017) were filtered out. Second, a precore SNP pool was picked with a computer script based on the following criteria: (1) remove SNPs with the same genotypes in the 299 accessions and keep only one representative SNP; (2) randomly sample a panel of 150 SNPs 10,000,000 times from the remaining markers and ensure that these SNPs were evenly distributed on the cowpea genome; (3) calculate the genetic distance of the 299 accessions using 150 SNPs in each panel if the genetic distance between two accessions appeared to be 0 for all SNP panels, which means 150 SNPs did not have enough power to distinguish all the accessions; and (4) repeat Steps 2 and 3, increasing the SNPs from 151 by one SNP each time, and calculate the genetic distance of the 299 accessions until no genetic distances of 0 appear between each pair of accessions. Third, the best discrimination power for all SNPs in the precore pool was determined using a Perl script described in Du et al. (2019). SNPs with a high discriminatory power of up to 97% across all varieties were considered informative core SNP sets. Finally, all the informative
SNPs based on the consensus map and the “ZZ” linkage map were combined into a core SNP set.

2.4 Population structure validation

To validate the resolution of the core SNP markers, the genotype data for the precore SNP pools and the informative core SNP set of 299 cowpea accessions were retrieved again for genetic diversity analysis. Population structure was analyzed using Structure 2.3.4 software under the admixture model with a burn-in period of 1000 followed by 1000 Markov chain Monte Carlo replications. Five independent runs each were performed with the number of clusters (K) varying from 1 to 10. The optimal K for subgrouping was calculated using STRUCTURE HARVESTER (Earl & vonHoldt, 2012). In addition, an unrooted phylogenetic tree was constructed using Tassel 5.0 based on the neighbor-joining method.

2.5 DNA extraction and KASP genotyping

Genomic DNA was extracted from leaves of 2-week-old seedlings using the cetyl trimethyl ammonium bromide (CTAB) method (Doyle & Doyle, 1987). DNA quality and concentration were measured with a NanoDrop 2000 ultraviolet (UV) spectrophotometer (Thermo Scientific, Waltham, MA), and working solutions were prepared at a concentration of 10 ng/μl. KASP genotyping was performed using an IntelliQube workstation (LGCGenomics Ltd., Hoddesdon, UK). PCR was performed in a 1.6-μl volume containing 0.8-μl DNA template and 0.8-μl mixture of 2x Master mix and Primer mix. The PCR program included 36 cycles following this protocol: 15 min at 95°C; 10 touchdown cycles of 20 s at 94°C and 60 s at 61–55°C (with the annealing temperature for each cycle reduced by 0.6°C per cycle); and 26 cycles of 20 s at 94°C and 60 s at 55°C. The genotype signals were detected in an IntelliQube workstation.

2.6 DNA fingerprinting

Seventy-five varieties were genotyped using the core SNP markers. The genotype of each variety was proposed as its DNA fingerprint. Pairwise genotype alignment by a local script was used to distinguish one variety from others, and neighbor-joining tree analysis was conducted to determine their genetic distances using Tassel 5.0.

3 RESULTS

3.1 KASP primer design

Among the 51,128 SNPs from the Cowpea iSelect Consortium Assay, 50,577 were converted into KASP markers, from which, 40,089 high-quality KASP markers were retained after filtering the primers with GC content <30% or >60%. Then, 598 randomly selected KASP primers were synthesized and tested in different populations, including RIL, F2, and natural populations. Over 96% of these KASP markers exhibited clear amplification products (Figure 1).

3.2 Core-SNP identification

The 299 cowpea accessions were previously genotyped with 49,194 SNPs using the 51,128 SNP assay (Xu et al., 2017). These data were filtered using a missing call rate ≤20%, heterozygosity call rate ≤20% and MAF (minor allele frequency) ≥ 0.01, and 30,211 high-quality SNPs remained that were used to develop the Core-SNP set. Among them, 23,506 SNPs were successfully converted into KASP markers; 20,161 of them were anchored to the consensus map (Muñoz-Amatriain et al., 2017) and were used for core SNP identification. Based on the calculated genetic distance (materials and methods), a minimal set of 188 SNPs, which could distinguish 86.7% of the 299 accessions, was selected as a precore SNP pool. Through further analysis of the full genotypes of the accessions that could not be distinguished, we found some landraces, although with different names, with identical genotypes. Alternatively, some accessions only varied at one SNP locus, and one of them had a missing genotype at this locus. After removing these redundant accessions, only one representative accession was retained. Finally, 24 informative SNPs with a high discriminatory power of up to 100% were proposed as a core SNP set (Figure 2). Following a similar strategy, using the ZZ linkage map as a reference, 6296 available SNPs were selected, and 250 SNPs were used to construct a precore SNP pool. Then, 26 informative SNPs were identified as a second core SNP set (Figure 2). To improve the resolution of the core SNPs, two core SNP sets were combined, and the 50 informative SNPs were finally used as the cowpea core SNP set. These 50 SNPs were distributed on 11 linkage groups, and the SNP number varied from 3 to 7 on each chromosome (Table 1 and Figure 3). In addition, the two precore SNP pools were also combined, and a new precore pool of 434 SNPs was developed after removing four overlapping SNPs between the two pools.

3.3 Genetic diversity validation using different SNP panels

To validate whether the core SNP set had equal power for genetic diversity analysis with the whole Cowpea iSelect Consortium Assay, population structure analysis was conducted in the diversity panel of 299 accessions using 434 precore SNPs and 50 informative core SNPs. As shown in Figure 4, both delta K peaks appeared at K = 2 when using 50 SNPs or 435 SNPs, which was consistent with the results in Xu et al. (2017) with 30,211 SNPs, suggesting that 50-SNPs or 434-SNPs sets have equal power to divide the 299 accessions into two subpopulations. The neighbor-joining tree analysis also displayed similar results using different SNP sets (Figure 4). These results
suggest that the 50 core SNPs can fully represent the whole Cowpea iSelect Consortium Assay in cowpea genetic diversity assessment.

3.4 Discriminatory power of the core SNP set in commercial cultivars

To assess the discriminatory power of the core SNPs, 75 commercial cultivars collected from the seed market were genotyped using the first set of 26 core SNPs. Forty-six cultivars (63%) were distinguished by these SNP genotypes. After integrating the genotyping information from the second set of 24 core SNPs, each cultivar could be distinguished from the other 74 cultivars by pairwise genotype alignment. The neighbor-joining tree analysis clearly distinguished any two of the 75 cultivars by their genetic distance (Figure 5). We also found that the 75 cultivars were distinguished completely when the 43rd SNPs were added, indicating that the discriminatory power of the core SNPs increased along with the number of SNPs used. This phenomenon was also found in pumpkin, where the 96, 48, 24, and 12 core SNP sets were able to identify 85.2%, 63.2%, 24.2%, and 4.9% of the 223 pumpkin accessions, respectively (Nguyen et al., 2020), again indicating a positive correlation between the germplasm discrimination power and the core SNP number. For each cultivar, the genotypes of the 50 core SNPs can be used for their unique DNA fingerprint or applied for rapid seed purity tests.

4 DISCUSSION

With the advantages of high polymorphism, codominant inheritance, high density, and high throughput, SNP markers have been recognized as ideal DNA markers. However, there is a large gap between the development of SNP markers and their application in practical breeding processes because SNP genotyping requires complex detection technologies and has high costs. Recently, multiple cost-effective and flexible core SNP sets based on the KASP technology have been developed.
| SNP   | LG | Pos | Primer_AlleleFAM | Primer_AlleleHEX | Primer_Common | Allele FAM | Allel HEX |
|-------|----|-----|------------------|------------------|--------------|------------|----------|
| 2_24217 | 1  | 12.89 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | TCTTTTCTATCAGATGTCTGA | A | C |
| 2_18117 | 1  | 25.31 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | CGACACTCGTGCCCTAGACC | T | C |
| 2_45212 | 1  | 35.55 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | AAGGTGACAGAAACACTCG | G | T |
| 2_44510 | 1  | 36.08 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | ATGCTACTTTTTAATCAGAG | G | A |
| 2_19015 | 1  | 36.70 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | GTGCTCTTGACACACTTTGCA | C | T |
| 2_19577 | 1  | 63.60 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | CTTCTTTGTGATTGATTAAG | T | C |
| 2_17551 | 1  | 63.80 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | CAGTATGGACACACACATAGTATTATTGATTAAG | C | A |
| 2_18599 | 2  | 0.35 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | AAAGATACCGGTTTGCACGAAGTC | A | G |
| 2_17367 | 2  | 16.24 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | ATAGTATACCGTCAATTTGCAAT | T | G |
| 2_34437 | 3  | 26.81 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | GAAATGCGAGTTGGAATCCTCACC | C | G |
| 2_27363 | 2  | 47.40 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | CTTATGATGAAATGCAATTTTGAACGAAGTC | A | G |
| 2_31738 | 2  | 63.64 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | AAACAGACTTTTGAATACCTCCTGA | G | A |
| 1_0147 | 3  | 3.24 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | AGATTCACACCTCTGGAATCTCCTGGA | G | T |
| 2_07774 | 3  | 34.51 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | ATCAAATTCAACAGTGAATCTCCTGGA | T | G |
| 2_26511 | 3  | 38.72 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | TAAAGGTTAATGAAATGCAATTTTGAACGAAGTC | A | G |
| 2_39823 | 3  | 70.99 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | AAACAGACCTTTTGAATACCTCCTGA | G | T |
| 2_46216 | 3  | 86.18 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | TTTTTTGTATTAATGACCATTTTTTGAATGCAATTTTGAACGAAGTC | A | G |
| 2_31947 | 3  | 101.58 | GAAGGTGACCAAGTTCATGC TGCACAGTACAAACACAGAA | GAAGGTGCGAGTCAACCGGATT GCAAGTGACAAACACAGAC | ATCAAATTCAACAGTGAATCTCCTGGA | T | G |
| SNP    | LG | Pos  | PrimerALLEL能力 | PrimerALLEL能力 | PrimerCommon | Allele FAM | Allel HEX |
|--------|----|------|----------------|----------------|--------------|------------|-----------|
| 2_30135 | 3  | 115.07 | GAAGGTGACCAAGTTCATG | CTTTAACATCTTTAGCATAA | AGCTTGGTATAAATGTGATCATCAGA | A | G |
| 2_44933 | 4  | 6.63  | GAAGGTGACCAAGTTCATGCTTTGGT | CCACCTCTCTGGTACGTTT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_13432 | 4  | 15.16 | GAAGGTGACCAAGTTCATGTTTGTAGGCGGCTTACGTTTG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_08923 | 4  | 57.38 | GAAGGTGACCAAGTTCATGCTGCCTTATGTAAATGACCATAG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_16199 | 4  | 62.36 | GAAGGTGACCAAGTTCATGCTGCCCTTGTATGGACCCGATTCAGTTG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_11960 | 5  | 2.94  | GAAGGTGACCAAGTTCATGCTCATATAAAGCTTCAAAAC | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_09514 | 5  | 21.23 | GAAGGTGACCAAGTTCATGCTGCCTTATGTAAATGACCATAG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_44215 | 5  | 70.00 | GAAGGTGACCAAGTTCATGCTGCCCTTGTATGGACCCGATTCAGTTG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_19571 | 6  | 11.01 | GAAGGTGACCAAGTTCATGCTGCCCTTGTATGGACCCGATTCAGTTG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_24363 | 6  | 19.80 | GAAGGTGACCAAGTTCATGCTGCCCTTGTATGGACCCGATTCAGTTG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_18460 | 6  | 20.45 | GAAGGTGACCAAGTTCATGCTGCCCTTGTATGGACCCGATTCAGTTG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_17089 | 7  | 19.10 | GAAGGTGACCAAGTTCATGCTGCCCTTGTATGGACCCGATTCAGTTG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_17453 | 7  | 28.26 | GAAGGTGACCAAGTTCATGCTGCCCTTGTATGGACCCGATTCAGTTG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_21597 | 7  | 31.49 | GAAGGTGACCAAGTTCATGCTGCCCTTGTATGGACCCGATTCAGTTG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_00124 | 7  | 42.34 | GAAGGTGACCAAGTTCATGCTGCCCTTGTATGGACCCGATTCAGTTG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_14673 | 7  | 51.36 | GAAGGTGACCAAGTTCATGCTGCCCTTGTATGGACCCGATTCAGTTG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 1_0248  | 7  | 59.16 | GAAGGTGACCAAGTTCATGCTGCCCTTGTATGGACCCGATTCAGTTG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_01512 | 8  | 3.85  | GAAGGTGACCAAGTTCATGCTGCCCTTGTATGGACCCGATTCAGTTG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
| 2_07412 | 8  | 32.57 | GAAGGTGACCAAGTTCATGCTGCCCTTGTATGGACCCGATTCAGTTG | GAAGGTGACCAAGTTCATGCTTTGGT | CCAATTGTGAGCAGGCAACTT | T | C |
**TABLE 1** (Continued)

| SNP     | LG | Pos  | Primer_AleleFAM | Primer_AleleHEX | Primer_Common |Allele FAM| Allel HEX|
|---------|----|------|-----------------|-----------------|--------------|----------|----------|
| 1_0649  | 8  | 45.49| GAAGGTGACCAAGTTCTGCTGTGCTTACGTGAAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGTTGAAAGAAGT...
developed in rice, *Brassica rapa*, pepper and pumpkin (Du et al., 2019; Li et al., 2019; Nguyen et al., 2020; Yang et al., 2019). With high representation and resolution, core SNP sets have been widely used in genetic diversity analysis, marker-assisted selection, and seed intellectual rights protection. In the current study, we developed a set of 434 precore SNP pools and a set of 50 core SNPs based on KASP technology and successfully used them for cowpea genetic diversity assessment and DNA fingerprinting.

Compared with the reported core SNP set from resequencing, such as genotyping by sequencing (GBS) in other crops (Du et al., 2019; Nguyen et al., 2020), the core SNP set in our study was directly derived from the existing Cowpea iSelect Consortium Assay. A total of 40,089 high-quality primers, with a 78.4% conversion rate, were obtained from the 51,128 SNPs, which is close to the 81.5% conversion rate in maize (Semagn et al., 2014). A total of 598 KASP primers were synthesized and validated in different
experiments, and identical genotypes were observed between the SNP assay and KASP markers at the same loci; over 96% of SNPs were polymorphic in at least one mapping population. With the cowpea IT97K-499-35 reference genome released, the physical position of these KASP markers was determined on the genome, and up to 62% SNPs in the iSelect array were found to be associated with genes (Lonardi et al., 2019), making them very useful resources for cowpea genetic mapping and gene/QTL cloning.

Assessment of genetic diversity in crop germplasm is essential for successful crop improvement (Varshney et al., 2007). In cowpea, the genetic diversity of different germplasm collections in Ghana, China, the United States, Korea, and Ethiopia has been evaluated using different SNP markers (Egbadzor et al., 2014; Ketema et al., 2020; Muñoz-Amatriain et al., 2017; Seo et al., 2020; Xu et al., 2017). Most studies aimed to evaluate the extent and pattern of genetic diversity in their own research collections and developed core collections as subsets. However, there is almost no research on how many markers and what kind of markers are suitable for cowpea genetic diversity assessment. In the current study, we re-evaluated 299 cowpea accessions using 50 informative core SNPs and 434 precore SNPs. Population structure and neighbor-joining tree analyses also showed similar results to previous evaluation results using the whole Cowpea iSelect Consortium Assay, suggesting that these 50 core SNPs have strong representation of the 51,128 SNPs in cowpea (Muñoz-Amatriain et al., 2017). These results also indicate that probably only 50 core SNPs are needed to analyze genetic dissimilarity among the accessions in different areas.

DUS testing is a standard system for plant variety protection (Jamali et al., 2019). However, as an increasing number of commercial cultivars are released, the phenotypic differences among different accessions become narrower. This makes it difficult to discriminate cultivars with similar genetic backgrounds with DUS testing. Therefore, a DNA-based system with molecular markers has been considered an alternative to improve the efficiency and accuracy of DUS testing (Jamali et al., 2019). In cowpea, SSR markers have been widely used to construct DNA fingerprints for cultivar protection (Danso et al., 2018; Lu et al., 2010; Mendes et al., 2015; Ragul et al., 2018). However, SSR marker resolution depends on the discrimination of electrophoretic bands, which may affect the genotyping accuracy. The relatively low throughput of SSR methods also limits their application to construct DNA fingerprints for a single cultivar or a limited number of cultivars at a single time. Moreover, SNP genotyping based on the KASP method outperforms the SSR approaches in both throughput and precision. In this study, based on the 50 core SNPs, DNA fingerprinting for 75 commercial cultivars was performed, and each cultivar could be distinguished from the other 74 cultivars, indicating that these 50 core SNPs could be used as a standard marker set for cowpea cultivar protection and registration, and they can also be used in rapid seed purity tests.

In conclusion, 40,089 KASP markers were developed, and a core set of 50 highly informative, representative, and evenly distributed SNP markers was identified in cowpea, providing a large flexible marker resource for both basic and applied research, such as genetic variation mining, germplasm management, parental material selection, variety identification, and seed purity tests.
ACKNOWLEDGMENTS
This study was supported by the National Key R&D Program of China (2017YFE0114500), the Key R&D Program of Guangdong Province (2020B020220002), and the Major Science and Technology Project of Plant Breeding in Zhejiang Province (2016C02051-7-3). We thank Jian Zhang (Beijing Academy of Agricultural and Forestry Sciences) and Ye Tao (Biozeroon Biotech, Shanghai) for technical assistance. We also thank Dr. Yiqun Weng from University of Wisconsin-Madison for editing the manuscript.

CONFLICT OF INTEREST
None declared.

ETHICAL APPROVAL
This study does not require any ethical approval.

DATA AVAILABILITY STATEMENT
The genotypic data on the mini core collection that support the findings of this study have been provided in a publication by Xu et al. (2017).

ORCID
Guojing Li https://orcid.org/0000-0001-5011-176X

REFERENCES
Boukar, O., Kong, L., Singh, B. B., Murdoch, L., & Ohm, H. W. (2004). AFLP and AFLP-derived SCAR markers associated with Striga gesnerioides resistance in cowpea. Crop Science, 44, 1259–1264. https://doi.org/10.2135/cropsci2004.1259

Cortés, A. J., Chavarro, M. C., & Blair, M. W. (2011). SNP marker diversity in common bean (Phaseolus vulgaris L.). Theoretical and Applied Genetics, 123(5), 827–845. https://doi.org/10.1007/s00122-011-1630-8

Coulibaly, S., Pasquet, R. S., Papa, R., & Gepts, P. (2002). AFLP analysis of the phenetic organization and genetic diversity of cowpea [Vigna unguiculata (L) Walp] reveals extensive gene flow between wild and domesticated types. Theoretical and Applied Genetics, 104, 258–266. https://doi.org/10.1007/s001220010740

Danso, B. A., Dzidzienyo, D. K., Prempeh, R. N. A., & Quain, M. D. (2018). DNA fingerprinting and assessment of genetic diversity among 22 cowpea [Vigna unguiculata (L) Walp] varieties grown in Ghana. Open Access Library Journal, 5(12), 1. e4938-4. https://doi.org/10.4236/oalib.1104938

Doyle, J. J. & Doyle, J. L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin, 19, 11–15.

Du, H., Yang, J., Chen, B., Zhang, X., Zhang, J., Yang, K., Geng, S., & Wen, C. (2019). Target sequencing reveals genetic diversity, population structure, core-SNP markers, and fruit shape-associated loci in pepper varieties. BMC Plant Biology, 19, 578. https://doi.org/10.1186/s12870-019-2122-2

Earl, D. A. & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Egbadzor, K. F., Ofori, K., Yeboah, M., Ababage, L. M., Opoku-Agyeman, M. O., Danquah, E. Y., & Offei, S. K. (2014). Diversity in 113 cowpea [Vigna unguiculata (L) Walp] accessions assessed with 458 SNP markers. Springerplus, 3, 541. http://www.springerplus.com/content/3/1/541, https://doi.org/10.1186/2193-1801-3-541

Fan, J. B., Oliphant, A., Shen, R., Kermani, B. G., Garcia, F., Gunderson, K. L., & Deloukas, P. (2003). Highly parallel SNP genotyping. Cold Spring Harbor Symposia on Quantitative Biology, 68(2), 69–78. https://doi.org/10.1101/sqb.2003.68.69

Gao, L., Jia, J., & Kong, X. (2016). A SNP-based molecular barcode for characterization of common wheat. PLoS One, 11(3), e0150947. https://doi.org/10.1371/journal.pone.0150947

 Hiremati, P. J., Kumar, A., Pemetsa, R. V., Farmer, A., Schlueter, J. A., Chamarthi, S. K., & Upadhyaya, H. D. (2012). Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. Plant Biotechnology Journal, 10(6), 716–732. https://doi.org/10.1111/j.1467-7652.2012.00710.x

Jamali, S. H., Cockram, J., & Hickey, L. T. (2019). Insights into deployment of DNA markers in plant variety protection and registration. Theoretical and Applied Genetics, 139, 1911–1929. https://doi.org/10.1007/s00122-019-02684-1

Jetema, S., Tesfaye, B., Keneni, G., Fenta, B. A., Assefa, E., Greliche, N., Machuka, E., & Yao, N. (2020). DARTSeq SNP-based markers revealed high genetic diversity and structured population in Ethiopian cowpea [Vigna unguiculata (L) Walp] germplasms. PLoS One, 15(10), e0239122. https://doi.org/10.1371/journal.pone.0239122

Li, P., Su, T., Wang, H., Zhao, X., Wang, W., Yu, Y., Zhang, D., Wen, C., Yu, S., & Zhang, F. (2019). Development of a core set of KASP markers for assaying genetic diversity in Brassica rapa subsp. chinensis Makino. Plant Breeding, 138(3), 1–16. https://doi.org/10.1111/pbr.12686

Livak, K. J., Flood, S., Marmaro, J., Giusti, W., & Deetz, K. (1995). Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. Genome Research, 4(6), 357–362. https://doi.org/10.1101/gr.4.6.357

Lonardi, S., Muñoz-Amatiain, M., Liang, Q., Shu, S., Wanamaker, S., Lo, S., … Close, T. J. (2019). The genome of cowpea (Vigna unguiculata (L) Walp). Plant Journal, 98, 767–782. https://doi.org/10.1111/tpj.14349

Lu, Z., Xu, P., Wang, B., Liu, Y., Wu, X., Hu, T., & Li, G. (2010). A rapid and precise SSR-based procedure for seed quality surveying of Vigna unguiculata ssp sesquipedalis (L)Verd. Acta Agriculturae Zhejiangensis, 22(6), 727–730. http://www.zjnyxb.cn/CN/Y2010/V22/I6/727

Mendes, R. F. M., Britto, F. B., Filho, R. F. F., & Lima, P. S. C. (2015). Genetic variability of indigenous cowpea genotypes as determined using inter-simple sequence repeats sequences repeats. African Journal of Biotechnology, 14(36), 2652–2661. https://doi.org/10.5897/AJB2015.14673

Muchero, M., Diop, N. N., Bhat, P. R., Fenton, R. D., Wanamaker, S., Pottorff, M., Hearne, S., Cisse, N., Fatokun, C., Ehlers, J. D., Roberts, P. A., & Close, T. J. (2009). A consensus genetic map of cowpea [Vigna unguiculata (L) Walp.] and synteny based on EST-derived SNPs. Proceedings of the National Academy of Sciences USA, 106, 18159–18164. https://doi.org/10.1073/pnas.0905886106

Muñoz-Amatiain, M., Mirebrahim, H., Xu, P., Wanamaker, S. I., Luo, M., Alhakami, H., … Close, T. J. (2017). Genome resources for climate-resilient cowpea, an essential crop for food security. Plant Journal, 89, 1042–1054. https://doi.org/10.1111/tpj.13404

Nguyen, N. N., Kim, M., Jung, J., Shim, E., Chung, S., Park, Y., Lee, G. P., & Sim, S. (2020). Genome-wide SNP discovery and core marker sets for assessment of genetic variations in cultivated pumpkin (Cucurbita spp.). Horticulture Research, 7, 1211–1210. https://doi.org/10.1038/s41438-020-00342-9

Ragul, S., Manivannan, N., Mahalingam, A., Prasad, V. B. R., & Narayanan, S. L. (2018). SSR marker based DNA fingerprinting for cowpea varieties of Tamil Nadu [Vigna unguiculata (L) Walp.]. International Journal of Current Microbiology and Applied Sciences, 7(4), 641–647. https://doi.org/10.20546/ijcmas.2018.704.072

WU ET AL.
