frontiers in Plant Science

ORIGINAL RESEARCH
published: 04 June 2021
doi: 10.3389/fpls.2021.655254

Development of GBTS and KASP Panels for Genetic Diversity, Population Structure, and Fingerprinting of a Large Collection of Broccoli (*Brassica oleracea* L. var. *italica*) in China

*Yusen Shen[1], Jiansheng Wang[1], Ranjan K. Shaw[1], Huifang Yu[1], Xiaoguang Sheng[1], Zhenqing Zhao[1], Sujuan Li[2] and Honghui Gu[1]**

[1] Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [2] Central Laboratory of Zhejiang Academy of Agricultural Sciences, Hangzhou, China

OPEN ACCESS

Edited by:
Sanghyeob Lee,
Sejong University, South Korea

Reviewed by:
Sung-Chur Sim,
Sejong University, South Korea
Jundae Lee,
Jeonbuk National University, South Korea

***Correspondence:**
Honghui Gu
guhh2199@163.com

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *18 January 2021*
Accepted: *21 April 2021*
Published: *04 June 2021*

Citation:
Shen Y, Wang J, Shaw RK, Yu H, Sheng X, Zhao Z, Li S and Gu H (2021) Development of GBTS and KASP Panels for Genetic Diversity, Population Structure, and Fingerprinting of a Large Collection of Broccoli (Brassica oleracea L. var. italica) in China. Front. Plant Sci. 12:655254. doi: 10.3389/fpls.2021.655254

Broccoli (*Brassica oleracea* var. *italica*) is one of the most important and nutritious vegetables widely cultivated in China. In the recent four decades, several improved varieties were bred and developed by Chinese breeders. However, the efforts for improvement of broccoli are hindered by limited information of genetic diversity and genetic relatedness contained within the available germplasms. This study evaluated the genetic diversity, genetic relationship, population structure, and fingerprinting of 372 accessions of broccoli representing most of the variability of broccoli in China. Millions of SNPs were identified by whole-genome sequencing of 23 representative broccoli genotypes. Through several stringent selection criteria, a total of 1,167 SNPs were selected to characterize genetic diversity and population structure. Of these markers, 1,067 SNPs were genotyped by target sequencing (GBTS), and 100 SNPs were genotyped by kompetitive allele specific PCR (KASP) assay. The average polymorphism information content (PIC) and expected heterozygosity (gene diversity) values were 0.33 and 0.42, respectively. Diversity analysis revealed the prevalence of low to moderate genetic diversity in the broccoli accessions indicating a narrow genetic base. Phylogenetic and principal component analyses revealed that the 372 accessions could be clustered into two main groups but with weak groupings. STRUCTURE analysis also suggested the presence of two subpopulations with weak genetic structure. Analysis of molecular variance (AMOVA) identified 13% variance among populations and 87% within populations revealing very low population differentiation, which could be attributed to massive gene flow and the reproductive biology of the crop. Based on high resolving power, a set of 28 KASP markers was chosen for DNA fingerprinting of the broccoli accessions for seed authentication and varietal identification. To the best of our knowledge, this is the first comprehensive study to measure diversity and population structure of a large collection of broccoli in China and also the first application of GBTS and KASP techniques in genetic characterization of broccoli. This work broadens the understanding of diversity, phylogeny, and population structure of a large collection of broccoli, which may enhance future breeding efforts to achieve higher productivity.

Keywords: broccoli, whole-genome sequencing, GBTS, KASP, genetic diversity, population structure, fingerprinting

INTRODUCTION

Broccoli (*Brassica oleracea* var. *italica*) is an economically important vegetable widely grown in many countries and is gaining popularity as a human diet due to its rich nutritional value as well as presence of anti-cancer glucosinolates in its florets (Wang et al., 2012). It was originated in the Mediterranean basin and introduced to China in the 20th century (Branca et al., 2005). In recent years, broccoli farming in China has made tremendous progress with about 80,000 ha of cultivated area in 2019. Additionally, many improved broccoli cultivars were produced and developed by Chinese breeders during the recent four decades, and China has become an important exporter of broccoli. However, most of these new cultivars are derived from a core collection of Japanese germplasm (Li et al., 2019) representing a narrow genetic background. For genetic improvement and conservation of crop species, it is vital to study the extent of genetic diversity and population structure prevailing in the species. Previously, only few studies were carried out to decipher the genetic diversity of broccoli using molecular markers (Tonguç and Griffiths, 2004; Lu et al., 2009; Huifang et al., 2011; Ciancaleoni et al., 2014; Stansell et al., 2018). However, a major concern is that, in most of the studies, there was poor representation of samples, or a limited number of samples was used raising the apprehension that the actual diversity present in broccoli has been underestimated. So, there is an urgent need to study the extent of molecular diversity present in broccoli to increase its productivity and quality. In the present investigation, a representative set of 372 broccoli accessions representing most of the variability were used for diversity analysis.

As the breeding of broccoli varieties are gaining momentum, many improved cultivars and modern varieties are being developed by both public and private organizations flooding the seed market. Consequently, it is difficult to distinguish the similar broccoli varieties in the seed trading market. The International Union for the Protection of New Varieties of Plants (UPOV) has established the "distinctness, uniformity, and stability" (DUS) testing for new varieties before registration (Jones et al., 2013). In this scenario, DNA fingerprinting can help in improving the knowledge of varieties, which are otherwise difficult to be distinguished phenotypically (Jones and Mackay, 2015). Additionally, DNA fingerprinting of crop species can also act as an insurance for the plant breeders to safeguard their important varieties and parental lines.

The molecular markers can assess the genetic diversity of crop species based on the variation of DNA that arises from substitution, insertion, and/or deletion in the chromosomes. Over the past three decades, several different DNA marker technologies, including amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs), single nucleotide polymorphisms (SNPs), and multiple nucleotide polymorphisms (MNPs) have been widely applied in DNA fingerprinting, genetic diversity, population structure analysis, and marker-assisted breeding (Tian et al., 2015). SSRs are routinely used for fingerprinting because of their high level of polymorphism (Rauscher and Simko, 2013; Pandino et al., 2015; Ouyang et al., 2018; Zheng et al., 2019). However, SSRs have some disadvantages; for instance, the throughput of detection is low, and the data integration or comparison between different detection platforms is difficult (Tian et al., 2015). On the contrary, SNPs are abundantly present in any given species (Deschamps and Campbell, 2010; Trebbi et al., 2011) and have been used in many genetic studies, including germplasm characterization (genetic diversity, genetic relationship, and population structure), quantitative trait loci (QTL) mapping, and marker-assisted selection (Ertiro et al., 2015). SNP array-based marker sets could also be used for fingerprinting (Tian et al., 2015; Thomson et al., 2017; Xu et al., 2017; Ellis et al., 2018), though this may not be their primary purpose. Recent emergence of next-generation sequencing technologies has made SNP discovery easy, rapid, cost effective, and with high throughput.

In recent times, several SNP genotyping platforms have been developed based on various technological principles, including TaqMan (Livak et al., 1995), genotyping by target sequencing (GBTS) (Guo et al., 2019), and kompetitive allele specific PCR (KASP) (KBiosciences[1]). GBTS is mainly divided into multiplex PCR-based target sequencing and probe-in-solution-based target sequencing. Nowadays, this platform contains several main technologies, such as AmpliSeq (Li et al., 2017), NimbleGen (Krasileva et al., 2017), SureSelect (Jiang et al., 2014), GenoBaits, and GenoPlexs (Guo et al., 2019). As a newly developed genotyping platform, GBTS combines the advantages of the fixed chips with the flexibility and low cost of GBS. As one of the targeted sequence-capture strategies, GBTS has been successfully utilized in maize (Guo et al., 2019), cucumber (Yang J. et al., 2019; Zhang et al., 2020), and pepper (Du et al., 2019). KASP assays developed by LGC Biosearch Technologies, United Kingdom, has

[1] www.kbioscience.co.uk

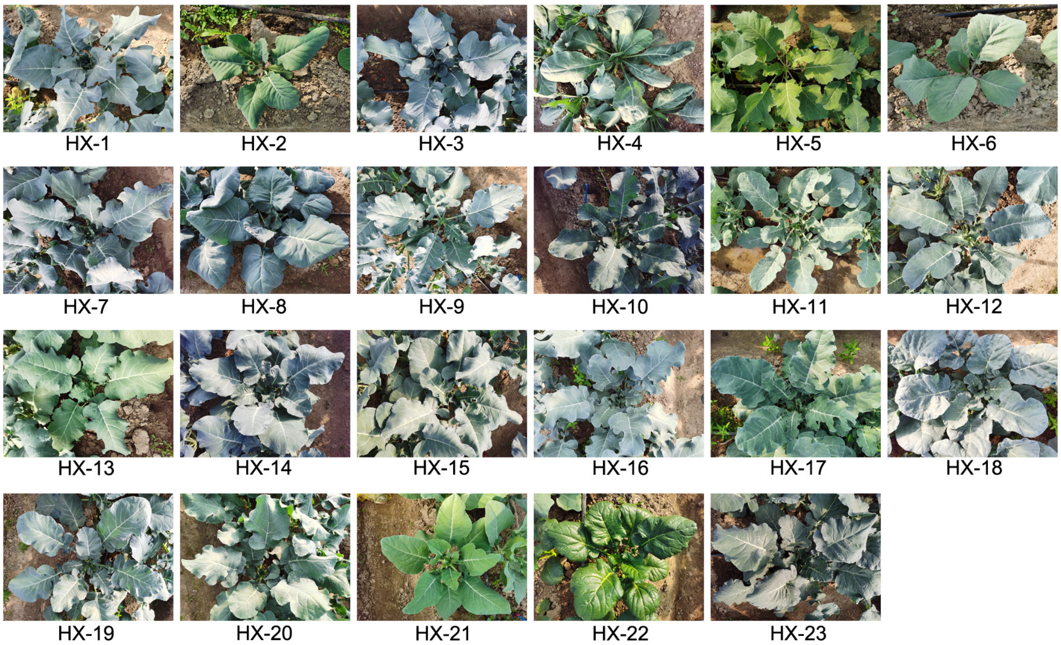

FIGURE 1 | Twenty-three diverse broccoli lines selected for whole-genome sequencing.

emerged as a powerful tool and is based on allele-specific oligo extension and fluorescence resonance energy transfer (FRET) for signal generation and performs bi-allelic scoring of SNPs, insertions, and deletions (InDels) at specific loci (Hao et al., 2017). Compared with fixed chip array, KASP is a more flexible and cost-effective technology, especially for a small number of markers to genotype large number of samples. KASP assays have been used in various crops, such as rice (Steele et al., 2018; Yang G. et al., 2019), wheat (Rasheed et al., 2016), and maize (Ertiro et al., 2015). In some horticultural plants, KASP markers were used primarily for fine mapping of the candidate QTLs (Rett-Cadman et al., 2019; Zhu et al., 2019; Liao et al., 2020). To the best of our knowledge, the development of GBTS or KASP panels for diversity, relatedness, structure analysis, and fingerprinting studies has not been reported in broccoli.

In this backdrop, the present investigation was carried out with the following objectives: (i) to identify genome-wide SNPs through whole-genome sequencing (WGS) of a set of diverse genotypes of broccoli; (ii) to develop a GBTS platform suitable for genotyping broccoli accessions; (iii) to develop a KASP platform suitable for fingerprinting of broccoli accessions for varietal identification; (iv) to decipher the genetic diversity, genetic relationship, and population structure of 372 broccoli accessions.

MATERIALS AND METHODS

Candidate Broccoli Genotypes for Sequencing

Based on our previous study on phylogenetic relationship, a set of 23 diverse genotypes were selected for WGS because of their genetic distinctness. Among the 23 diverse broccoli genotypes, 10 were double haploid (DH) lines, and the others were advanced-generation inbred lines. These materials (**Figure 1**) are representative of many agronomical important traits of broccoli, such as head color, head shape, leaf angle, stem height, flowering time, etc. Detailed information of major morphological features (according to DUS) and source of these 23 diverse broccoli genotypes are given in **Supplementary Table 1**.

Broccoli Accessions for Phylogenetic Study and Fingerprinting

To analyze the phylogenetic relationship, genetic variability, population structure, and also to carry out the fingerprinting for varietal identification, a set of 372 accessions were used. Most of these accessions (367) were provided by 11 broccoli breeding groups in China, and five hybrids (widely planted in China) were from the Sakata Seed Corporation of Japan. The 372 broccoli accessions comprised of 102 breeding lines (including DH lines, improved lines, and the parents of commercial hybrids), 248 landraces, and 22 commercial hybrids, could nearly cover all agro-ecological zones of China, where broccoli is widely cultivated.

Whole-Genome Sequencing

Genomic DNA of 23 broccoli genotypes was extracted from young leaf tissues by a modified cetyl trimethyl ammonium bromide (CTAB) method (Hanania et al., 2004). The DNA quality and quantity were determined by a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, United States). The paired-end WGS libraries were constructed using Illumina's sample preparation kit (Illumina Inc., United States) as per the manufacturer's instructions. Genomic DNA (1 μg) was fragmented using Covaris S2 System (Covaris, Inc., United States) followed by adaptor ligation using the Kapa Hyper Prep kit (Kapa Biosystems). The ligated products were purified, and the enrichment of adaptor-modified fragments was performed by PCR. The profiles of the enriched libraries were assessed for size and quality using a high-sensitivity lab chip kit on an Agilent Bioanalyzer (Agilent Technologies, United States). A paired-end cluster generation kit (Illumina Inc., United States) was used for cluster generation on paired-end flow cells. Cluster generation was carried out by hybridization of WGS libraries with a concentration ranging from 10 to 40 nmol/μl, which again diluted to a stock solution of 4 nmol of molecules on the oligonucleotide-coated surface of the flow cell. Isothermal amplification of the libraries was carried out to generate clonal DNA clusters, and the libraries were sequenced on an Illumina HiSeq2500 instrument (Illumina, United States) with an average read length of 2 × 250 bp.

Discovery of Genome-Wide Single Nucleotide Polymorphisms

The original image data generated by the Illumina sequencing machine were converted into sequence data via base calling (Illumina pipeline CASAVA v1.8.2), and the raw reads were then subjected to quality control (QC) procedure and adapter trimming to remove the unusable reads. The available draft genome sequence of HDEM broccoli with a genome size of 630 Mb (Belser et al., 2018) was downloaded, and the filtered sample reads were mapped to the reference genome using the BWA software (Li and Durbin, 2009) with default parameters. SNP discovery and filtration were carried out using SAMtools v0.1.19 (mpileup and varFilter) (Li et al., 2009) with default parameters. The raw SNP sets were called by SAMtools with the parameters as "-q 1 -C 50 -m 2 -F 0.002 -d 1000." This set of SNPs was filtered by using the following criteria: (1) The mapping quality >20; (2) the depth of the variate position >4; (3) window size for occurrences of one SNP: 100 bp. The software ANNOVAR was used for functional annotation of variants (Wang et al., 2010), and the UCSC known genes were used for gene and region annotations (Hsu et al., 2006).

Selection of Single-Nucleotide Polymorphisms for the Genotyping by Target Sequencing Platform and Genotyping

After SNPs were called for 23 broccoli genotypes, the quality of each SNP was assessed based on minor allele frequency (MAF) and polymorphic information content (PIC) using PowerMarker v3.25 software (Liu and Muse, 2005). Several filtering steps were performed to carefully select the SNPs for the GBTS platform. The filtration of SNPs was carried out by using the following criteria: (1) select the SNPs having missing data points of <5%; (2) SNPs with the MAF values of >0.2; (3) merge the adjacent SNPs to get MNPs, and the regions for each MNP should be less than 150 bp within more than one SNP located.

The GBTS library construction consisted of two rounds of PCR. In the first round, the target SNPs in plant DNA samples were amplified and captured using a multiplexed PCR panel. In the second round, a unique barcode was added to the captured product for each DNA sample. First, the multiplexed PCR was performed in 30-μl reactions including 20 ng of DNA template, 10 μl of GenoPlexs 3 × T Master Mix (Molbreeding Biotechnology Company, Shijiazhuang, China). The PCR conditions were as follows: 95°C for 5 min, then 16 cycles of 95°C for 30 s and 60°C for 4 min, and an extension at 72°C for 5 min. The PCR products were purified by magnetic bead suspension and 75% alcohol. Second, PCR was conducted in 30-μl reactions consisting of 11 μl of purified PCR product from the first round, 10 μl of GenoPlexs 3 × T Master Mix, and 1 μl of sequencing connector with barcode sequence. The PCR conditions were as follows: 95°C for 5 min, then nine cycles of 95°C for 30 s, 60°C for 20 s, and 72°C for 30 s, and an extension at 72°C for 5 min. The second round of PCR products was purified with 100 μl of 75% alcohol and 23 μl of Tris-HCl buffer (10 mM, pH 8.0–8.5). Thereafter, the constructed library was sequenced using Illumina HiSeq X Ten platform.

To better characterize the 372 broccoli accessions, the genotyping data were further filtrated as follows: (1) removing the SNPs having missing data points of >20%, (2) removing SNPs with the MAF values of <0.05, and (3) removing the SNPs with the PIC values of <0.2.

Design of Kompetitive Allele Specific PCR Assay and Genotyping

The available SNPs were filtered based on several criteria to choose the best loci for KASP marker development. The filtration of SNPs was carried out by removing (1) SNPs having missing data points of >20%, (2) SNPs with the MAF values of <0.05, (3) other variations located within a distance of 50 bp upstream and downstream of the target SNPs, and (4) SNPs with the PIC values of <0.2. Among the SNPs amenable to KASP assay development, SNPs were further selected for designing of KASP primers by (1) removing the primer sequence with GC content lower than 0.3, (2) based on the loci present within the exonic, intergenic, and upstream/downstream of the functional genes, and (3) loci selected based on physical position, which are evenly distributed across the chromosomes.

For the selected KASP loci, primers were designed, and genotyping of the 372 broccoli accessions was carried out. Basically, the KASP assay contains three components: KASP assay mix, KASP master mix, and template DNA. The KASP assay mix (72X) contains two different, allele-specific, competing forward primers with unique tail sequences at the 5′ end (allele-1 tail has FAM-labeled oligo sequence and allele-2 tail has HEX-labeled oligo sequence) and one common reverse primer. KASP master mix (2X) contains FAM and HEX-specific FRET (fluorescence resonance energy transfer) cassette, ROX passive reference dye, KASP Taq DNA polymerase (specially modified for allele-specific PCR), dNTPs, and $MgCl_2$ in an optimized buffer solution, which is universal to every KASP genotyping assay. The genomic DNA was extracted from 100 mg of young leaf tissue from each broccoli accession by a modified cetyl trimethyl ammonium bromide (CTAB) method (Hanania et al., 2004). The DNA quality and quantity were determined by a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, United States), and DNA concentration of 10–30 ng/ml was used for KASP genotyping. The PCR amplification was performed in 96-well microplates containing a final reaction volume of 10.14 μl. The PCR mixture in each well contained 5 μl of 10 ng/μl of genomic DNA, 5 μl of 2X KASP master mix, and 0.14 μl of KASP assay mixture. After dispensing the reaction mixture of 10.14 μl into a 96-well microplate, PCR was carried out using the KASP thermal cycling program of three steps [Initial activation: 94°C for 15 min; 10 touchdown cycles at 94°C for 20 s, and 61–55°C for 60 s (dropping 0.6°C per cycle); and finally, 26 cycles at 94°C for 20 s and 55°C at 60 s]. After the completion of PCR, the microplate was read with FRET-capable plate detector (Molecular Devices, Sunnyvale, CA, United States) for fluorescence detection of PCR product. Furthermore, the fluorescence readings were analyzed using KlusterCaller^{TM} Version 3.4.1.36 software (LGC Biosearch Technologies, United Kingdom) for visualization of allelic discrimination of each genotype.

Genetic Diversity and Population Structure Analysis

The allelic data of the 372 broccoli accessions genotyped by GBTS and KASP panel were integrated. All the SNP markers were analyzed, and various genetic diversity summary statistics including observed heterozygosity (Ho), expected heterozygosity (He) also called gene diversity (GD), minor allele frequency (MAF), and polymorphic information content (PIC) were measured using the PowerMarker software v3.25 (Liu and Muse, 2005). The genotyping data were used to calculate the pairwise genetic distance matrix and to construct a phylogenetic tree by the neighbor-joining method using MEGA v10.0.4 software (Kumar et al., 2018), with 1,000 bootstrap replications to understand the relationship among the genotypes. Principal component analysis (PCA) was performed using the prcomp function in R language (v4.0.2) to visualize the overall representation of diversity in broccoli accessions.

The population structure in the broccoli accessions was assessed to identify the optimal number of subpopulations. The STRUCTURE v2.3.4 software package (Pritchard et al., 2000; Hubisz et al., 2009) was used for structure analysis based on Bayesian clustering approach. To determine the population structure, 10 runs for each *K* value from 1 to 10 was performed using 10 iterations for each *K*. For each run, a burn-in period of 50,000 was used to minimize the effect of the starting configuration, which was followed by an additional 100,000 iterations using a model with admixture (genotype might have mixed ancestry) and correlated allele frequencies. The likelihood of optimal *K* value was calculated, and the *K* value corresponding to the highest likelihood was interpreted as the number of subpopulations in the samples. The most probable *K* value was determined from the uppermost level of population structure, detected using an *ad hoc* statistic ΔK based on the rate of change in the log probability of data between successive *K* values (Evanno et al., 2005). Using the web-based STRUCTURE HARVESTER v0.6.94 (Earl and VonHoldt, 2012), the *ad hoc* statistic ΔK was calculated.

Population Differentiation and Genetic Diversity Indices

Analysis of molecular variance (AMOVA) was computed using GenAlEx v6.502 (Peakall and Smouse, 2012) to estimate the variance components among and within the subpopulations generated by STRUCTURE HARVESTER. Population differentiation test such as the Wright's fixation index (Fst) (Wright, 1965), which measures the amount of genetic variance of the populations was estimated using GenAlEx v6.502, and significance was tested based on 1,000 bootstraps. Gene flow (Nm) among the population was calculated using the formula, Nm = 0.25 (1 - Fst)/Fst (Slatkin and Barton, 1989). Nei's genetic distance and several other genetic indices such as number of loci with private allele, number of different alleles (Na), number of effective alleles (Ne), Shannon's information index (I), observed heterozygosity (Ho), and expected heterozygosity (He) were also calculated using GenAlEx v6.502.

RESULTS

Whole-Genome Sequencing and Discovery of Single-Nucleotide Polymorphisms

Twenty-three diverse broccoli genotypes were used for WGS, and paired-end sequencing libraries were generated for each sample. The libraries were of good quality according to sizing profiles generated by high-sensitivity lab chip kit (Agilent Technologies, United States), and the fragment size of the libraries ranged from 450 to 750 bp. Sequencing of these libraries in the Illumina HiSeq2500 instrument yielded 13.1 to 18.7 Gb of raw data per sample (**Table 1**). A total of 346.18 Gb of raw data with an average of 15 Gb were generated for 23 broccoli genotypes. After quality filtering and adapter trimming, 345.57 Gb of high-quality data were available for further processing (**Table 1**). On average, 99.8% of raw data were able to pass the filtration process indicating the high quality of sequencing.

The number of reads per sample ranged from 87.7 to 124.9 million (**Table 1**). The Hiseq sequencing platform yielded a total of 2,303.9 million paired reads with an average of 100 million reads. The high-quality filtered reads were mapped to the broccoli reference genome. Out of a total of 2,303.9 million reads, 2,272.9 million reads were successfully aligned to the broccoli reference genome. The mapping percentage of each sample ranged from 98.1% to 98.98%. On average, 98.65% of the reads could be successfully mapped to the reference genome. The sequence depth of mapping reads ranged from 25 to 34.59× with an average of 28× (**Table 1**). A million numbers of SNPs were detected between the sample and reference genome ranging from 899,926 to 1,908,908 (**Table 2**). Among the identified SNPs, the number of transitions (Ti) was more than transversions (Ts) across the samples, and the average Ti/Ts was found to be 1.43.

TABLE 1 | Statistics on whole-genome sequencing of 23 genotypes.

Genotypes	Number of raw reads	Total data (Gb)	Mapped reads	Alignment (%)	Average depth (×)
HX-1	97,192,414	14.60	96,008,663	98.78	26.81
HX-2	96,297,844	14.46	94,695,611	98.34	27.55
HX-3	91,090,564	13.68	89,886,851	98.68	25.45
HX-4	88,358,380	13.27	86,911,004	98.36	25.99
HX-5	88,810,900	13.34	87,126,622	98.10	25.55
HX-6	91,782,854	13.79	90,186,403	98.26	26.09
HX-7	101,924,026	15.31	100,625,993	98.73	28.57
HX-8	100,947,244	15.16	99,489,024	98.56	27.92
HX-9	90,268,680	13.55	89,119,556	98.73	25.21
HX-10	95,782,732	14.38	94,745,032	98.92	26.33
HX-11	104,574,752	15.70	103,463,996	98.94	28.30
HX-12	107,531,638	16.15	106,399,003	98.95	29.43
HX-13	120,392,562	18.08	118,738,641	98.63	32.79
HX-14	100,768,972	15.13	99,663,458	98.90	28.48
HX-15	98,685,420	14.82	97,357,254	98.65	27.40
HX-16	104,594,330	15.71	102,981,890	98.46	29.10
HX-17	113,117,330	16.99	111,587,466	98.65	31.07
HX-18	124,943,550	18.77	123,081,201	98.51	34.59
HX-19	94,680,260	14.22	93,710,734	98.98	25.77
HX-20	96,140,460	14.45	94,838,250	98.65	27.65
HX-21	100,730,344	15.14	99,606,256	98.88	28.38
HX-22	87,695,482	13.19	86,434,061	98.56	25.07
HX-23	107,536,214	16.18	106,246,250	98.80	30.15
Average	100,167,259	15.05	98,821,879	98.65	27.98

TABLE 2 | Statistics on single nucleotide polymorphisms (SNPs) identified between the samples and the reference.

Genotypes	Total SNPs identified	Transition (Ti)	Trans version (Ts)	Ti/Ts	Genic SNPs	Intergenic SNPs
HX-1	1,158,081	681,828	476,253	1.43	265,391	669,826
HX-2	1,738,501	1,022,217	716,284	1.43	403,819	990,748
HX-3	1,201,122	705,991	495,131	1.43	265,879	708,587
HX-4	1,712,276	1,006,504	705,772	1.43	396,290	980,145
HX-5	1,908,908	1,121,167	787,741	1.42	418,966	1,125,604
HX-6	1,707,755	1,002,927	704,828	1.42	398,259	972,794
HX-7	1,192,357	701,175	491,182	1.43	260,217	708,394
HX-8	1,448,231	851,295	596,936	1.43	318,736	856,671
HX-9	1,201,811	707,932	493,879	1.43	270,738	702,699
HX-10	946,171	557,839	388,332	1.44	211,313	555,371
HX-11	991,691	585,706	405,985	1.44	227,813	573,127
HX-12	970,147	571,855	398,292	1.44	215,391	570,691
HX-13	1,321,936	778,187	543,749	1.43	283,100	790,635
HX-14	1,066,173	626,568	439,605	1.43	244,623	618,534
HX-15	1,205,647	709,834	495,813	1.43	268,203	711,509
HX-16	1,415,974	832,819	583,155	1.43	310,570	838,402
HX-17	1,291,865	759,962	531,903	1.43	288,518	759,049
HX-18	1,422,938	838,731	584,207	1.44	313,859	836,882
HX-19	899,926	529,813	370,113	1.43	218,272	504,722
HX-20	1,421,281	835,484	585,797	1.43	302,010	857,048
HX-21	1,163,649	684,612	479,037	1.43	259,640	683,677
HX-22	1,289,277	757,110	532,167	1.42	289,306	760,051
HX-23	1,206,976	711,004	495,972	1.43	280,159	693,236
Average	1,299,248	764,372	534,875	1.43	291,786	759,496

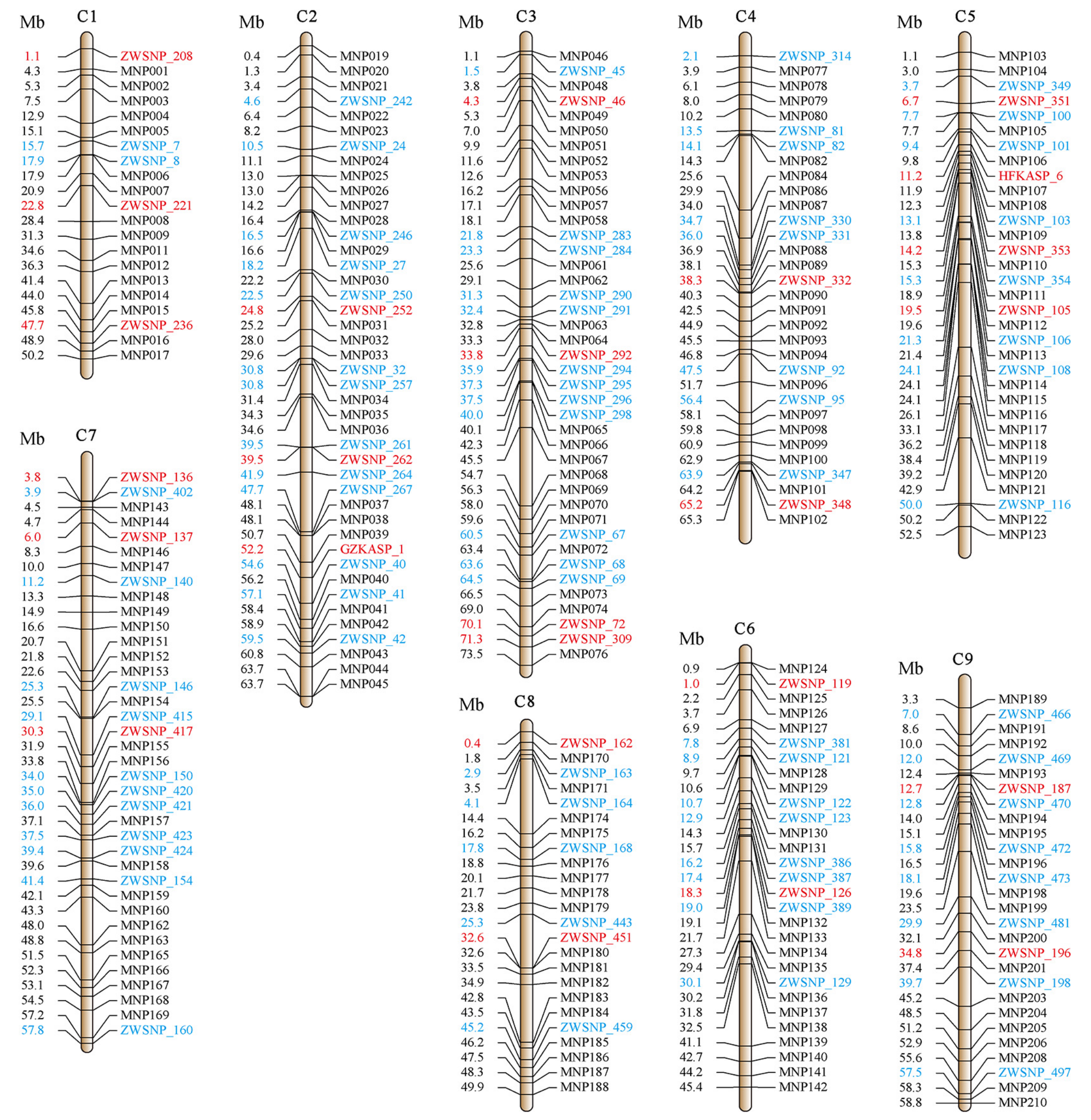

FIGURE 2 | Distribution of all markers among broccoli chromosomes. The horizontal lines perpendicular to a chromosome represent the markers developed in this study, among which the black ones were developed by the GBTS panel, the blue and red ones were developed by the KASP panel, and the red ones were core KASP markers selected for fingerprinting of broccoli accessions.

The intergenic SNPs were almost more than double the number of genic SNPs (**Table 2**).

Genotyping by Target Sequencing and Kompetitive Allele Specific Panel Development

The putative SNPs identified through the re-sequencing of 23 broccoli genotypes were further filtered for the GBTS and KASP panel developments, respectively.

For the GBTS panel, 210 MNPs consisting of 1,332 SNPs were selected, with an average of 6.3 SNPs within each MNP. The utility of these SNPs in deciphering the genetic relationship and diversity of 372 broccoli accessions was assessed by calculating the missing data, MAF, and PIC values. As a result, 1,067 (80.1%) SNPs were retained for further analysis (**Supplementary Table 2**). These SNPs were located in 189 MNP regions and evenly distributed across the nine chromosomes (**Figure 2**).

For the KASP panel, a total of 13,621 SNPs with the PIC values between 0.2 and 0.5 were retained after filtration. Of these, only 8,768 (64.4%) SNPs could be successfully designed as KASP markers. From this set, 2,515 SNPs were from exonic (including non-synonymous, stop-gain and stop-loss SNPs), intronic, intergenic regions, and from the upstream or downstream of the functional genes. Based on their functional role in regulating important agronomic traits and also based on the physical position and uniform genomic distribution, finally, 500 SNPs were selected for KASP marker development. Of the

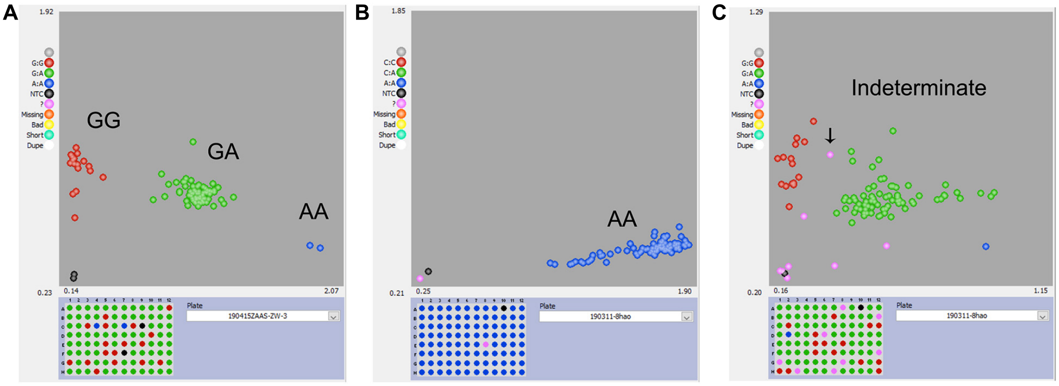

FIGURE 3 | Representative results of KASP genotyping assay. **(A)** KASP marker with good polymorphism that was successfully developed. **(B)** KASP marker with no polymorphism that should be discarded. **(C)** KASP marker with more than 10% missing values or had ambiguous SNP-calling that should be discarded.

targeted 500 KASP markers, only 347 (69.4%) could be genotyped successfully with high-quality genotype clusters (**Figure 3A**), while clear clustering was not obtained for the remaining KASP markers. Out of 347 scorable KASP markers, 54 KASPs were monomorphic (**Figure 3B**), so the remaining 293 KASP markers were used for genotyping of the 372 broccoli accessions. Based on the genotyping data, KASPs having missing data points of more than 10% or showing ambiguous SNP call (**Figure 3C**) were removed. Finally, 100 KASPs with clear genotype cluster and even representation of the chromosomes (**Figure 2**, **Supplementary Table 2**, and **Supplementary Table 3**) were called successfully with high confidence and selected for further analysis.

Genotyping of 372 Broccoli Accessions for Assessment of Genetic Diversity

Key descriptive statistics for measuring the utility and informativeness of SNP markers were assessed by analyzing the genetic diversity of the 372 broccoli accessions (**Supplementary Table 4**). Of the 1,167 SNPs, 1,067 SNPs were from the GBTS panel, and 100 SNPs were from the KASP panel. The MAF values of all SNPs were 0.11–0.50, with a mean of 0.33 (**Figure 4A**). Most of the SNPs (87%) have MAF values more than 0.2, which could be considered as markers with normal allele frequency. Interestingly, not a single SNP had MAF of <0.05. The PIC values of the individual SNPs varied and ranged from 0.18 to 0.37 with a mean value of 0.33 (**Figure 4A**). The maximum percentage of SNPs (46%) was having PIC values ranging from 0.35 to 0.37. The observed heterozygosity for the KASP marker ranged from 0.00 to 0.93 with an average of 0.23 (**Figure 4B**). The expected heterozygosity (gene diversity) ranged from 0.20 to 0.50 with an average of 0.42 (**Figure 4A**).

Genetic Relationship and Population Structure

The pairwise genetic distance matrix ranged from 0.0004 to 0.6418 with a mean of 0.3894 (**Supplementary Table 5**), indicating the presence of considerable diversity in the broccoli accessions, though few were closely related.

Based on the genotyping data of 1,167 SNP markers in 372 broccoli accessions (**Supplementary Table 3**), a phylogenetic tree was constructed by the neighbor-joining method using the MEGA software. Cluster analysis revealed that all the 372 broccoli accessions were clustered into two major groups (**Figure 5**) with many sub-groups within the major groups, but importantly, the grouping was not supported by high bootstrap values, indicating weak grouping of broccoli accessions. The phylogenetic tree depicting the genetic relationship of 372 broccoli accessions were as expected based on the breeding history, and accessions sharing the pedigree were placed in close proximity. The cluster I containing 98 accessions were mostly improved varieties, and the cluster II contained 274 accessions, which were mostly landraces. PCA based on the pairwise genetic distance matrix complemented the cluster analysis results, and the samples were grouped into two clusters (**Figure 6**). The first and second axes of PCA captured only 8.5% and 12.2% of the overall variance, respectively. This again explained the weak grouping of the broccoli accessions in the present study.

To further verify the results of phylogenetic and PCA analyses, the population structure of 372 broccoli accessions were studied by STRUCTURE v2.3.4 (Pritchard et al., 2000; Hubisz et al., 2009). The number of clusters (K) of the accessions was estimated by setting the number of clusters (K) from 1 to 10 with 10 replications for each K. The average logarithm of the probability of likelihood [LnP(D)] and standard deviations for different number of sub-populations (K = 1 to 10) are presented in **Supplementary Table 6**. The most likely number of clusters (K) was selected by comparing the logarithmized probabilities of the LnP(D) and ΔK data (Yang G. et al., 2019). In this study, the LnP(D) showed continuous gradual increase with the increase in K (**Figure 7A**), making it difficult to assume the best value of K. Remarkably, the number of clusters (K) was plotted against ΔK, which showed a sharp peak at $K = 2$ (**Figure 7B**). Hence, based on the *ad hoc* statistic ΔK method, we decided to choose the value of $K = 2$ for our analysis, which clearly indicated the presence of two subpopulations within the broccoli accessions. A total of 125 accessions were assigned to subpopulation-I and 247 accessions to subpopulation-II (**Figure 7C** and

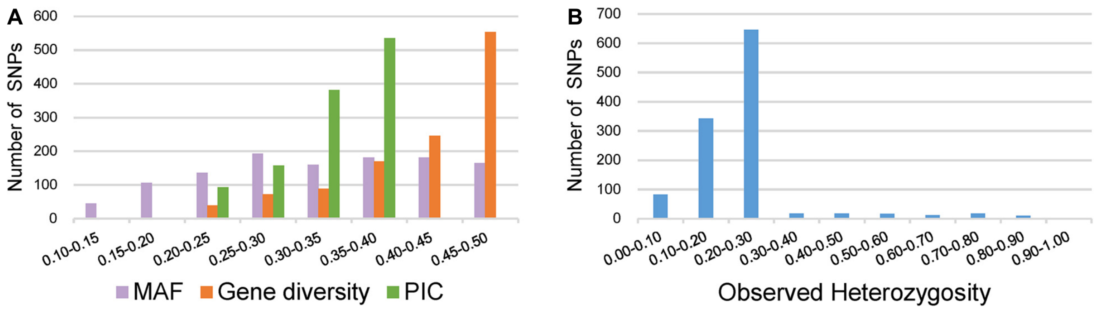

FIGURE 4 | Distribution of genetic diversity for 1,167 markers in the 372 broccoli accessions. **(A)** Minor allele frequency (MAF), gene diversity (GD), and polymorphic information content (PIC). **(B)** Observed heterozygosity.

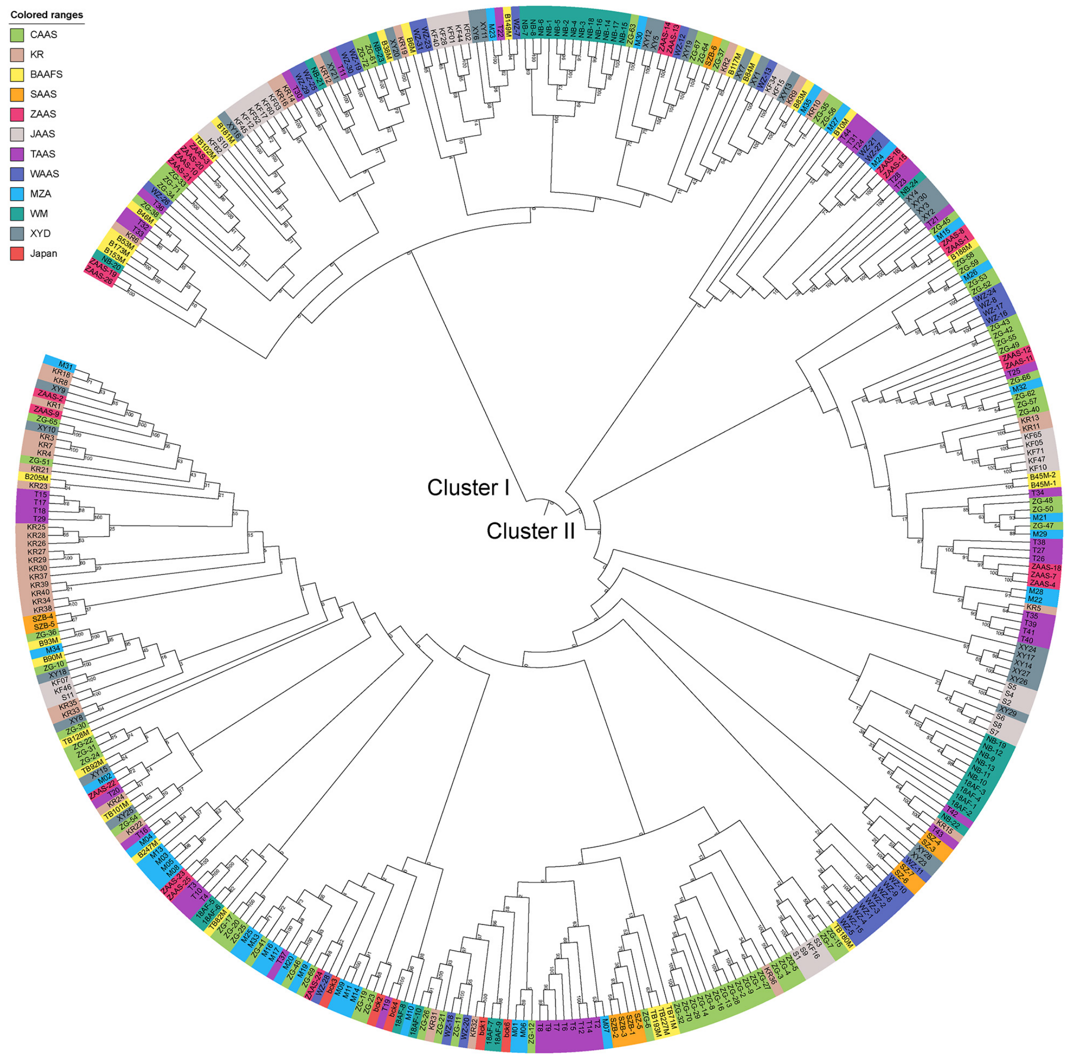

FIGURE 5 | Phylogenetic tree constructed using the neighbor-joining method based on 372 broccoli accessions. The colors in outer circle represent different sources of the accessions. CAAS, Chinese Academy of Agricultural Sciences; KR, Tianjin Kerun Vegetable Research Institute; BAAFS, Beijing Academy of Agriculture and Forestry Sciences; SAAS, Shanghai Academy of Agricultural Sciences; ZAAS, Zhejiang Academy of Agricultural Sciences; JAAS, Jiangsu Academy of Agricultural Sciences; TAAS, Taizhou Academy of Agricultural Sciences; WAAS, Wenzhou Academy of Agricultural Sciences; MZA, Zhejiang Mitsuo Seed Co., Ltd.; WM, Ningbo Weimeng Seed Co., Ltd.; XYD, Zhenjiang Xinyuanda Horticulture Technology Co., Ltd.; Japan, Sakata Seed Corporation, Japan.

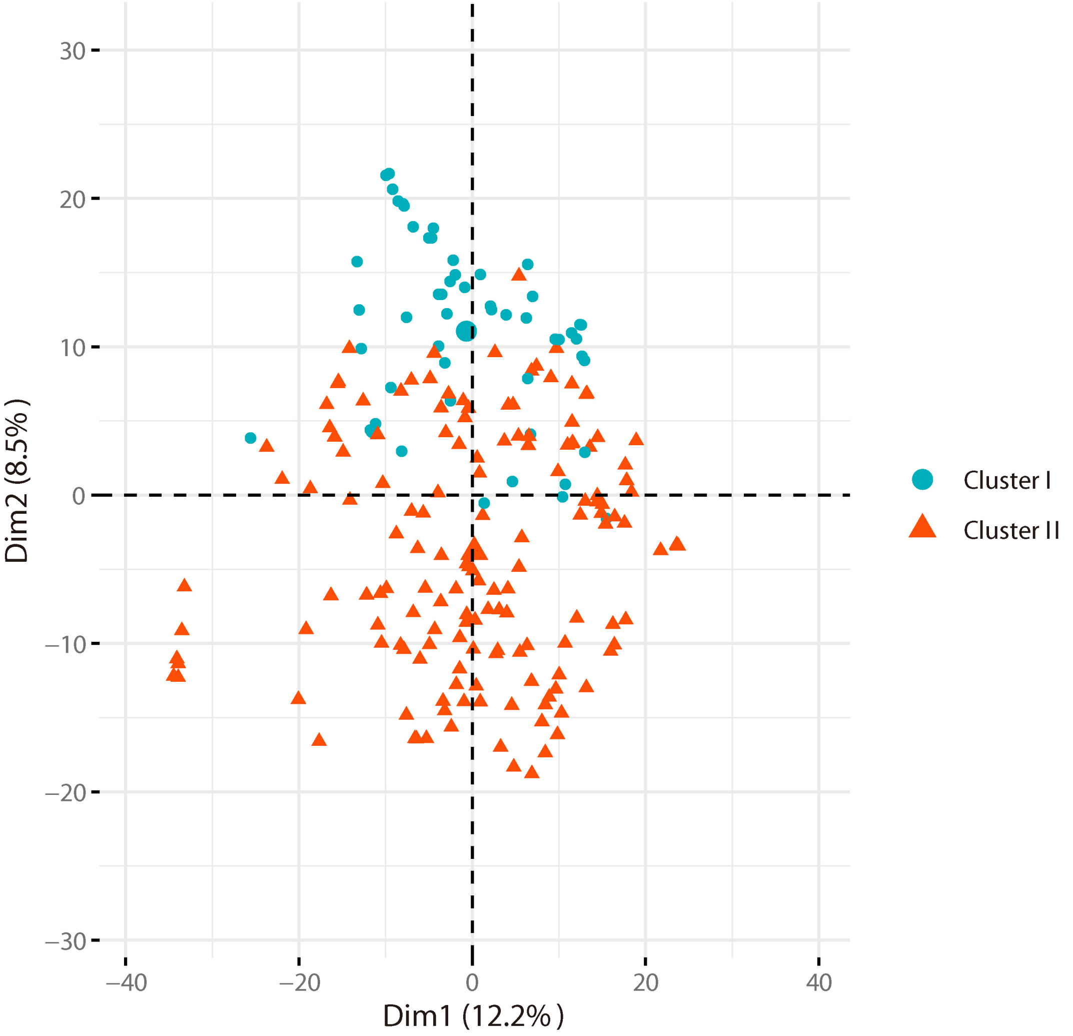

FIGURE 6 | Principal component analysis (PCA) of broccoli accessions. Axes-1 (12.2%) and Axes-2 (8.5%) separate the genotypes into two groups.

Supplementary Figure 1). The bar plot also exhibited wide admixture in the broccoli accessions (**Supplementary Figure 1**), and the two subpopulations did not show any association with the geographic origin of the materials. The STRUCTURE result was in accordance with the results of the phylogenetic and the principal component analyses, which also showed the presence of weak grouping, indicating extensive exchange of the broccoli accessions by breeders. In addition, there was a small peak observed at $K = 5$ (**Figure 7C**), which might indicate another informative population structure. Therefore, the STRUCTURE results at both $K = 2$ and $K = 5$ were subject to the following population genetic analyses.

Genetic Differentiation of Population and Gene Flow

To estimate the genetic variation among and within the two subpopulations identified in STRUCTURE, analysis of molecular variance (AMOVA) was carried out using GenAlEx v6.502. The F_{ST} value among the two populations (POP1 and POP2) was 0.126, supporting low population differentiations. The population differentiation based on AMOVA revealed that only 13% ($p < 0.001$) of the total variation was found among the subpopulations, while the rest (87%) was within the populations (**Supplementary Figure 2**). In addition, the overall gene flow among the population was estimated to be high (Nm = 1.741) (**Table 3**). Nei's standard genetic distance between the two subpopulations was 0.103. The AMOVA and pairwise population F_{ST} analyses were done based on the population structure at $K = 5$, and the results are shown in **Supplementary Tables 7**, **8**.

Allelic Pattern in the Subpopulations

The subpopulation I (I = 0.596, He = 0.411 and uHe = 0.413) shows higher diversity than the subpopulation II (I = 0.570, He = 0.388 and uHe = 0.388) (**Table 4**). The percentage of polymorphic loci per population (PPL) for the two populations

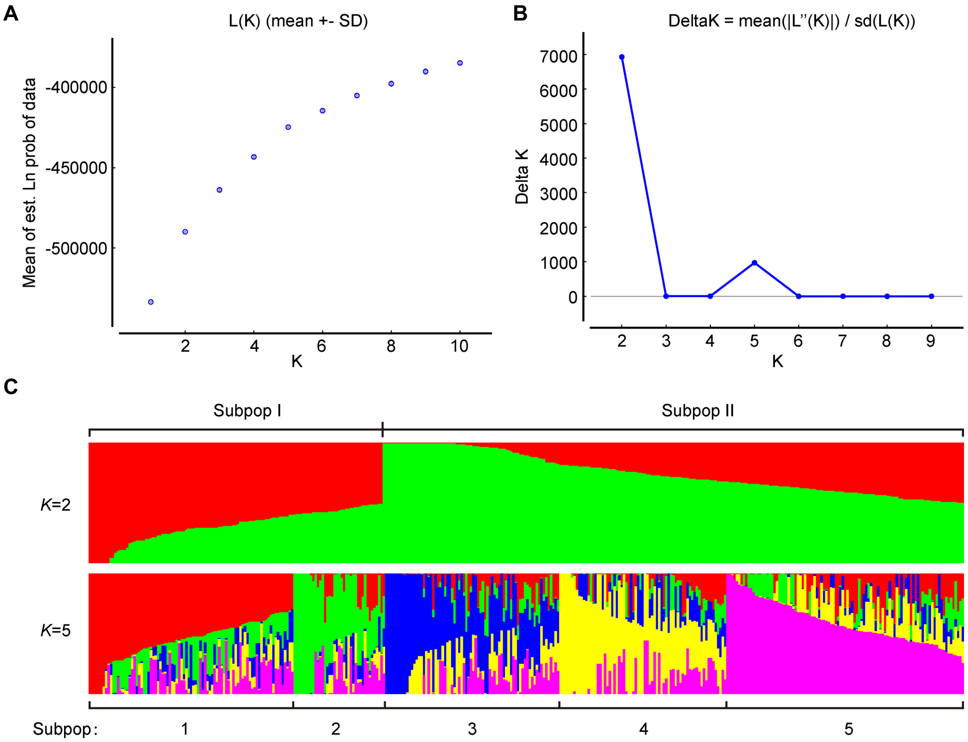

FIGURE 7 | Population structure of 372 broccoli accessions based on 1,167 markers. **(A)** Log probability of data, L(*K*) averaged over the replicates. **(B)**Δ*K* values plotted as the number of subpopulations. **(C)** Subpopulations (*K* = 2 and *K* = 5) inferred using structure analysis.

TABLE 3 | Analysis of molecular variance (AMOVA) depicting the genetic variation among and within two subpopulations of broccoli.

		df	SS	MS	Est. Var.	%
Source	**Among populations**	1	11,370.844	11,370.844	33.547	13%
	Within populations	742	173,385.349	233.673	233.673	87%
	Total	743	184,756.194		267.220	100%
F-statistics	**Fst**	0.126				
	Nm	1.741				

was 100%. The grand mean value of different alleles (Na) and number of effective alleles (Ne) of the two subpopulations were 1.740 and 1.676, respectively. The grand mean value of I, He, and uHe for the two populations were 0.583, 0.399, and 0.401, respectively (**Table 4**).

Selection of Core Kompetitive Allele Specific PCR Markers for Fingerprinting of Broccoli Accessions

To build a rapid and cost-effective way of varietal identification, we selected 28 KASP markers from the genotyping database for the fingerprinting of every accession (**Supplementary Tables 2, 3**). These KASP markers were highly effective in distinguishing the 372 examined accessions, as the markers are evenly distributed across the nine chromosomes (**Figure 2**). For each accession, the genotype-based KASP barcode was used to generate a corresponding 2D barcode using an online tool[2]. This barcode can be scanned to obtain the information used for creating the 2D barcode. **Supplementary Figure 3** depicts a barcode of a representative variety of broccoli used in the present study.

DISCUSSION

Whole-Genome Sequencing and Single-Nucleotide Polymorphism Discovery

In the current study, Illumina platform was chosen to undertake the whole-genome sequencing of 23 diverse broccoli genotypes and discover genome-wide SNPs. The Illumina platform was preferred for WGS due to its throughput, cost effectiveness, and indexing capabilities. The WGS data were generated with a high

[2] www.cli.im

TABLE 4 | Mean of different genetic parameters in each of the two subpopulations.

Population	N	Na	Ne	I	Ho	He	uHe	F	PPL
Pop1	124.675	2.000	1.740	0.596	0.247	0.411	0.413	0.394	100.00%
Pop2	246.045	2.004	1.676	0.570	0.223	0.388	0.388	0.427	100.00%
Mean	185.360	2.002	1.708	0.583	0.235	0.399	0.401	0.411	100.00%

N, the number of samples; Na, number of different alleles; Ne, number of effective alleles = 1/(Sum p_i^2); I, Shannon's Information Index = -1 × Sum[p_i × Ln (p_i)]; Ho, observed heterozygosity = number of Hets/N; He, expected heterozygosity = 1 - Sum p_i^2; uHe, unbiased expected heterozygosity = [2N/(2N - 1)] × He; F, Fixation Index = (He - Ho)/He = 1 - (Ho/He); PPL, percentage of polymorphic loci, where p_i is the frequency of the ith allele for the population and Sum p_i^2 is the sum of the squared population allele frequencies.

coverage (approximately 28×), and a total of 346.18 Gb of raw data for 23 broccoli genotypes were generated. Sequencing with high coverage gives confidence for SNP calling as low coverage sequencing will make SNP calling difficult with high genotyping errors (Heffelfinger et al., 2014). The raw data were of high quality as, on average, more than 99% were able to pass the filtration process with an error rate of as low as 0.03.

The samples reads were aligned to the reference genome of broccoli (HDEM), and the alignment percentage ranged from 98.1 to 98.98 with an average of 98.65%. The high mapping percentage signifies high quality of the sample reads. Among the SNPs, transition SNPs were more frequent than the transversions, which is a common phenomenon in *Brassica* species (Park et al., 2010; Huang et al., 2013). This happens due to synonymous mutations in protein-coding sequences (Guo et al., 2017) and suggested that transition mutations are better tolerated than transversion mutations during natural selection (Luo et al., 2017). The average transition/transversion ratio was observed to be slightly lower (1.43) indicating a high level of sequence divergence (Yang and Yoder, 1999). The genic SNPs identified were twofold lesser than the intergenic SNPs, and this could be possible as intergenic regions evolve faster and accumulate higher polymorphism than the genic regions, which are mostly conserved (Guo et al., 2007). The intergenic region could be a great source of SNPs for genetic studies, which has been poorly exploited in several crops.

Genotyping by Target Sequencing and Kompetitive Allele Specific PCR Panel Development

In this study, a total of 1,067 markers were selected for the GBTS panel development. This GBTS panel with larger numbers of markers can effectively be used for genetic diversity and population structure analysis. In maize, a series of high-quality GBTS panels, including 20 K, 10 K, 5 K, and 1 K SNP loci were developed, making it an affordable genotyping platform for maize marker-assisted breeding (Guo et al., 2019). Based on the targeted sequence-capture strategy, 382 key cucumber varieties were genotyped by 122 SSRs using target SSR-seq method (Yang J. et al., 2019), and 261 cucumber varieties were genotyped by 163 SNPs using target SNP-seq method (Zhang et al., 2020). In pepper, 92 SNPs were used to detect polymorphisms across 271 commercial pepper varieties (Du et al., 2019). In the present study, 210 multiple-SNP (MNP) regions consisting of 1,332 SNPs were selected from the whole-genome sequencing of the 23 broccoli genotypes. A total of 372 broccoli accessions were genotyped by these SNPs, and 1,067 (80.1%) SNPs were selected for genetic diversity and population structure analysis of the broccoli accessions.

For the fingerprinting of the broccoli lines, few KASP markers will be more cost effective and flexible. KASP assay has been widely used by plant breeders for genetic diversity study, genetic mapping, and genetic purity test in several major crops including pigeon pea (Saxena et al., 2012), chickpea (Hiremath et al., 2012), maize (Jagtap et al., 2020), cotton (Islam et al., 2015), tomato (Devran et al., 2016), peanut (Zhao et al., 2017), wheat (Rasheed et al., 2016), and rice (Pariasca-Tanaka et al., 2015; Cheon et al., 2018; Steele et al., 2018; Yang G. et al., 2019). In the present study, out of 28,220 filtered SNPs, 13,621 SNPs with PIC values between 0.2 and 0.5 were retained for KASP assay design. However, only 8,768 (64.4%) SNPs could be designed as KASP markers successfully. This low conversion rate may have occurred probably due to the presence of duplicate loci, paralogous sequence, or incorrect primer design (Semagn et al., 2014; Steele et al., 2018; Jagtap et al., 2020). Optimizing the PCR conditions may improve the rate of successful KASP assay design.

Assessment of the Genetic Diversity

A total of 372 broccoli accessions represents a wide variability and comprised of several improved lines from China, elite breeding lines, parents of commercial hybrids, landraces, and a few commercial hybrids. All these accessions were genotyped by 1,167 SNPs developed from the GBTS and KASP panels. The PIC values and expected heterozygosity (also called gene diversity) are two parameters in measuring genetic diversity in any population. In the present study, the PIC values of the SNPs ranged from 0.18 to 0.37 with a mean value of 0.33. In a previous study by Li et al. (2019) who evaluated the genetic diversity of 95 broccoli genotypes of China with SSR markers, the PIC values of the markers ranging from 0.48 to 0.99, with an average of 0.79, which was considerably high were found. The average PIC value of 0.33 in the present study could be considered high due to the bi-allelic nature of the SNPs, which restrict the range of PIC values from 0 to 0.5 (Eltaher et al., 2018). The expected heterozygosity (gene diversity) ranged from 0.20 to 0.50 with a mean value of 0.42. The mean value of expected heterozygosity revealed the prevalence of low to moderate level of diversity in broccoli. This study is comprised of broccoli accessions provided by 11 breeding groups in China, so good representation of diversity of broccoli was expected. The use of a representative set of germplasm is necessary to get the proper estimate of gene diversity in a species (Senthilvel et al., 2017). Though for the first time, a large collection of broccoli accessions with a representative set of accessions was used to estimate genetic diversity of broccoli in China, high level of diversity was not observed.

The MAF value of the SNPs ranged from 0.11 to 0.50, with a mean of 0.33. MAF threshold dramatically affects population structure inference, and inference of population structure is sensitive to MAF (Linck and Battey, 2019). SNPs with low MAF detected through NGS analysis tend to be less polymorphic than SNPs with higher MAF. In genetic diversity studies, it is desirable to maximize the number of polymorphic markers by selecting SNPs with moderate to high MAF. In this context, 87% SNPs have the MAF values more than 0.2, and importantly, not a single marker has MAF < 0.05. So, the set of SNP markers used in the present study is ideal for structure analysis and association studies.

Genetic Relatedness and Structure

The pairwise genetic distance matrix ranged from 0.0004 to 0.6418 with a mean of 0.3894. Most of the accessions were found considerably distant, though a few accessions, mostly improved cultivars, were found genetically similar indicating the accessions having a common breeding history. Maximum genetic distance of 0.6418 was found between the broccoli accession ZAAS-8 and KR8. Interestingly, the genetic distance between several broccoli accessions belonging to different breeding groups was reported to be as low as 0.0004 (M31 and KR18). This could be due to the massive exchange of materials between different broccoli breeding groups in China resulting in the development of genetically similar broccoli cultivars. Recently, Li et al. (2019) reported high similarity coefficients ranging from 0.6909 to 0.8969 with an average of 0.7809 among the 95 broccoli genotypes in China collected from around the world. Lu et al. (2009) also reported a high similarity ranging from 0.8421 to 0.9330 between several broccoli genotypes. Several authors (Tonguç and Griffiths, 2004; Louarn et al., 2007) also reported similar close relationship of broccoli genotypes and suggesting the prevalence of narrow genetic base in broccoli (Hu and Quiros, 1991).

For phylogenetic analysis, all 372 broccoli accessions were clustered into two major groups with many sub-groups within the major groups, but the grouping was not supported by high bootstrap values indicating the prevalence of weak grouping patterns. Low bootstrap values suggest recombination or gene flow between different "branches" of the phylogenetic tree. Again, low bootstrap values explain that the members on the branch should not be divided into separate groups as it seems. The accessions that share the pedigree and were having similar breeding history were placed in close proximity as expected. Group I contained 98 accessions and comprised of improved varieties. Group II contained 274 accessions and represents mostly landraces and few improved accessions. Five Japanese F_1 hybrids were grouped in cluster II and were placed closely with several accessions such as M14, T19, and KR 32. As mentioned earlier, broccoli was introduced into China from Japan four decades back (Li et al., 2019). Most of the modern broccoli cultivars of China might have been derived from the Japanese lines. So not surprisingly, the genetic distance between the Japanese F_1 hybrids used in the present study and several improved varieties were found to be very low (**Supplementary Table 5**). The phylogenetic analysis again confirms the lineage of the Chinese broccoli varieties to the Japanese broccoli. Though from 1990 to the early 21st century, many broccoli cultivars were bred and developed by Chinese institutes (Li et al., 2019), still most of the broccoli cultivars widely planted in China are closely related and have a narrow genetic background. So, there is a necessity to broaden the genetic base of Chinese broccoli by incorporating diverse germplasm into the breeding program. However, the landraces used in the present study were genetically more diverse and could be used in the breeding program. PCA results corroborated with the findings of phylogenetic tree indicated that the samples were grouped into two clusters, but the variance captured by the first and second axes of PCA was only 8.5% and 12.2%, respectively, again explaining weak grouping.

All 372 broccoli accessions were divided into two sub-populations by the STRUCTURE analysis, which was in agreement with the phylogenetic tree and PCA results. Subpopulation-I contained 125 accessions, and subpopulation-II contained 247 accessions. Most of the broccoli accessions in cluster I of the phylogenetic tree were part of sub-population I. Similarly, most of the broccoli accessions in cluster II of the phylogenetic tree were placed in sub-population II (**Supplementary Figure 1**). On the whole, STRUCTURE and phylogenetic analyses agreed with each other with minor exceptions. This small discrepancy between the two methods of grouping is expected as the cluster analysis in neighbor joining tree assigned a fixed branch position to each accession, while STRUCTURE analysis resulted in a sub-population membership percentage, and the highest percentage was used to assign individuals to groups for easy interpretation (Wang et al., 2009), but importantly, the Fst value between the two populations was found to be low (0.126), suggesting the existence of weak genetic structure in the panel of 372 broccoli accessions. The two subpopulations also did not show any association with the geographic origin of the materials, indicating continuous exchange of broccoli parental lines among the breeders in China, and a close relationship exist among the broccoli accessions. Similar results of weak grouping coincided with the PCA and phylogenetic tree analyses. In the two inferred groups (**Supplementary Figure 1**), the samples were observed with potential admixture.

Population structure study is important for association mapping studies, and testing of population structure is conducted first to identify true marker–trait association. Absence of strong structure is a desirable feature for association analysis to avoid spurious marker–trait associations (Flint-Garcia et al., 2003). In this context, the panel of 372 broccoli accessions used in the present study is an ideal population for association study to identify associations between several agronomically important traits and marker alleles in broccoli. Furthermore, genetic diversity and population structure assessment study will assist the breeders in developing a core collection of broccoli for breeding and molecular studies.

Genetic Differentiation of Populations

Fst value is a measure of population differentiation due to genetic structure. According to Wright (1943), the population differentiation could be considered high if the Fst value is greater than 0.25. The genetic differentiation between the two subpopulations was measured, and the overall Fst value was found to be low (0.126). Nei's pairwise genetic distance among the two subpopulations was also found to be low (0.103). A low level of differentiation among the two subpopulations could have happened due to extensive gene flow, which acts as a powerful force to decrease population differentiation. Basically, gene flow value (Nm) greater than 1 is "strong enough" to prevent population differentiation (Slatkin and Barton, 1989). The low population differentiation of the present study was supported by sufficient gene flow (Nm = 1.741), suggesting a high genetic exchange between the populations. Low Fst value also coincided with AMOVA results where majority of the total variation (87%) was due to the "within the populations" variations, while only 13% of the total variation was due to "among-subpopulation" variations (**Supplementary Figure 2**). Majority of the genetic variations are harbored by individuals of the present study. These results again confirmed the finding of phylogenetic and STRUCTURE analysis, which showed the presence of weak genetic structure in the broccoli accessions.

Fingerprinting of Broccoli Accessions

Due to a short breeding history of broccoli in China and also due to close relationship between certain varieties, it becomes difficult to distinguish between different cultivars using only morphological characters. Often, spurious seeds are easily mixed with the genuine ones in the broccoli seed market, which eventually harm the interest of consumers and breeders. Additionally, cytoplasmic male sterility (CMS) lines are widely used for the development of commercial F_1 hybrids of broccoli. There is every chance that the CMS lines may get contaminated by foreign pollen, mechanical admixture, artificial mislabeling, or by mixing with other heterogeneous seeds during seed production. Loss of parental line purity can cause irreparable damage to the breeders. To solve these problems, it is essential to confirm the seed authenticity and purity of the parental lines before hybrid development and distribution.

In this context, DNA fingerprinting could act as an insurance for the breeders in safeguarding the elite varieties and germplasm. DNA fingerprinting have advantages in identifying varieties and are free from environmental impact, compared with customary field inspection (Tommasini et al., 2003). Molecular markers, especially SNPs, are widely used for cultivar identification in many crops (Yoon et al., 2007; Jung et al., 2010; Ganopoulos et al., 2013). In the present study, we have investigated the use of KASP assays for varietal identification of broccoli. Though we have developed 100 KASP assays, to make the fingerprinting cost effective, 28 KASP markers were chosen based on high PIC value and high discriminating power. The fingerprinting result showed that 28 KASP markers are sufficient to distinguish the 372 broccoli accessions, and these markers are considered as core markers for fingerprinting (**Supplementary Table 3**). The SNP genotype of each broccoli accession was used to generate a corresponding 2D barcode using an online tool (see text footnote 2) and could be used as a reference genetic barcode for each genotyped broccoli accession. This barcode contained fingerprinting information of a large collection (372) of broccoli cultivars in China and can ensure the authenticity of varieties. As SNPs are bi-allelic, allele identification is always comparable among different genotyping platforms. The broccoli SNP barcode database constructed in this study will allow several broccoli breeding groups working in China to use this set of SNPs for varietal identification.

Furthermore, the KASP markers developed in our study can be utilized in genetic purity of the broccoli F_1 hybrids. For hybrid purity, three to five KASPs will be adequate to ascertain the genetic purity. The markers can be selected from the 100 KASPs according to the parental genotypes of a specific hybrid of broccoli.

CONCLUSION

This study provides comprehensive information about the genetic diversity of broccoli accessions in China based on a large population of 372 broccoli accessions. Genome-wide SNPs were discovered by WGS of 23 diverse broccoli genotypes, and the SNPs were converted into GBTS and KASP panels for genotyping of a large set of broccoli accessions. Evaluation of genetic diversity demonstrated that low to moderate genetic diversity prevails in broccoli cultivars widespread in China, and few accessions are in close relationship. This indicated the prevalence of a narrow genetic base among broccoli accessions and the need to broaden the gene pool by adding diverse genotypes into the breeding program. Phylogenetic and PCA analyses based on 1,167 SNPs revealed the presence of two groups but did not show strong groupings. The STRUCTURE results also suggested the presence of two subpopulations with weak genetic structures. The population differentiation was found to be low indicating extensive gene flow. A set of 28 KASP markers was chosen for DNA fingerprints of the broccoli accessions for varietal identification. The SNP genotype of each broccoli accession was used to generate a 2D barcode containing the fingerprinting information of a large collection (372) of broccoli accessions in China. The KASP markers developed in our study could also be used for seed authenticity and purity evaluation of broccoli cultivars. To our knowledge, this is the first study to measure diversity and population structure of a large collection of broccoli in China and also the first application of GBTS and KASP techniques in broccoli for genetic studies. The information generated in the present study will assist in the selection of suitable genotypes for breeding in developing new cultivars as well as physiological and molecular studies in broccoli.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA681704.

AUTHOR CONTRIBUTIONS

HG designed the experiments, supervised the study, and revised the manuscript. YS and RS analyzed the experimental data and wrote the manuscript. JW and ZZ collected the accessions and selected the core germplasms for re-sequencing. HY, XS, and SL designed the KASP markers and performed the research. All authors discussed the results and contributed to the final manuscript.

FUNDING

Funding was provided by the National Key Research and Development Program of China (2017YFE0114500 and 2017YFD0101805), the National Science Foundation of China (NSFC) (32002039), the Natural Science Foundation of Zhejiang Province (LQ20C150001), the Science and Technology Department of Zhejiang Province (2016C02051-5), the Zhejiang Academy of Agricultural Sciences for Subject construction (A1), and the National Major Scientific Joint Research of Improved (Broccoli) Varieties of China (2018–2022).

ACKNOWLEDGMENTS

We thank the breeding groups in China for kindly providing the DNA of broccoli accessions, including the Chinese Academy of Agricultural Sciences, the Tianjin Kerun Vegetable Research Institute, the Beijing Academy of Agriculture and Forestry Sciences, the Shanghai Academy of Agricultural Sciences, the Jiangsu Academy of Agricultural Sciences, the Taizhou Academy of Agricultural Sciences, the Wenzhou Academy of Agricultural Sciences, the Zhejiang Mitsuo Seed Co., Ltd., the Ningbo Weimeng Seed Co., Ltd., and the Zhenjiang Xinyuanda Horticulture Technology Co., Ltd.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.655254/full#supplementary-material

Supplementary Figure 1 | Bar plot of the STRUCTURE result.

Supplementary Figure 2 | Percentage of molecular variance in the population.

Supplementary Figure 3 | Barcode of a representative variety of broccoli.

Supplementary Table 1 | Phenotype of the 23 whole-genome sequencing genotypes.

Supplementary Table 2 | Information of the GBTS and KASP markers used in this study.

Supplementary Table 3 | Genotyping data of 372 broccoli accessions using GBTS and KASP markers.

Supplementary Table 4 | The MAF, PIC value, gene diversity, and observed heterozygosity of the markers.

Supplementary Table 5 | Genetic distance across the 372 broccoli accessions.

Supplementary Table 6 | Mean and standard deviations of the logarithm of the probability of data for different number of sub population (*K*) tested in broccoli.

Supplementary Table 7 | AMOVA depicting the genetic variation among and within five subpopulations of broccoli.

Supplementary Table 8 | Results of pairwise population Fst analysis when *K* = 5.

REFERENCES

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F.-C., Falentin, C., et al. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* 4, 879–887. doi: 10.1038/s41477-018-0289-4

Branca, F., Bahcevandziev, K., Perticone, V., and Monteiro, A. (2005). Sources of resistance to downy mildew (*Peronospora parasitica* (Pers. ex Fr.) Fr.) in Sicilian germplasm of cauliflower and broccoli. *Biodivers. Conserv.* 14, 841–848. doi: 10.1007/s10531-004-0652-9

Cheon, K., Baek, J., Cho, Y., Jeong, Y., Lee, Y., Oh, J., et al. (2018). Single nucleotide polymorphism (SNP) discovery and kompetitive allele-specific PCR (KASP) marker development with Korean Japonica rice varieties. *Plant Breed. Biotechnol.* 6, 391–403. doi: 10.9787/PBB.2018.6.4.391

Ciancaleoni, S., Chiarenza, G. L., Raggi, L., Branca, F., and Negri, V. (2014). Diversity characterisation of broccoli (*Brassica oleracea* L. var. *italica* Plenck) landraces for their on-farm (*in situ*) safeguard and use in breeding programs. *Genet. Resour. Crop Evol.* 61, 451–464. doi: 10.1007/s10722-013-0049-2

Deschamps, S., and Campbell, M. A. (2010). Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. *Mol. Breed.* 25, 553–570. doi: 10.1007/s11032-009-9357-9

Devran, Z., Göknur, A., and Mesci, L. (2016). Development of molecular markers for the *Mi-1* gene in tomato using the KASP genotyping assay. *Hortic. Environ. Biotechnol.* 57, 156–160. doi: 10.1007/s13580-016-0028-6

Du, H., Yang, J., Chen, B., Zhang, X., Zhang, J., Yang, K., et al. (2019). Target sequencing reveals genetic diversity, population structure, core-SNP markers, and fruit shape-associated loci in pepper varieties. *BMC Plant Biol.* 19:578. doi: 10.1186/s12870-019-2122-2

Earl, D. A., and VonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Ellis, D., Chavez, O., Coombs, J., Soto, J., Gomez, R., Douches, D., et al. (2018). Genetic identity in genebanks: application of the SolCAP 12K SNP array in fingerprinting and diversity analysis in the global in trust potato collection. *Genome* 61, 523–537. doi: 10.1139/gen-2017-0201

Eltaher, S., Sallam, A., Belamkar, V., Emara, H. A., Nower, A. A., Salem, K. F. M., et al. (2018). Genetic diversity and population structure of $F_{3:6}$ nebraska winter wheat genotypes using genotyping-by-sequencing. *Front. Genet.* 9:76. doi: 10.3389/fgene.2018.00076

Ertiro, B. T., Ogugo, V., Worku, M., Das, B., Olsen, M., Labuschagne, M., et al. (2015). Comparison of kompetitive allele specific PCR (KASP) and genotyping by sequencing (GBS) for quality control analysis in maize. *BMC Genomics* 16:908. doi: 10.1186/s12864-015-2180-2

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 357–374. doi: 10.1146/annurev.arplant.54.031902.134907

Ganopoulos, I., Tsaballa, A., Xanthopoulou, A., Madesis, P., and Tsaftaris, A. (2013). Sweet cherry cultivar identification by high-resolution-melting (hrm) analysis using gene-based SNP markers. *Plant Mol. Biol. Report* 31, 763–768. doi: 10.1007/s11105-012-0538-z

Guo, C., McDowell, I. C., Nodzenski, M., Scholtens, D. M., Allen, A. S., Lowe, W. L., et al. (2017). Transversions have larger regulatory effects than transitions. *BMC Genomics* 18:394. doi: 10.1186/s12864-017-3785-4

Guo, X., Wang, Y., Keightley, P. D., and Fan, L. (2007). Patterns of selective constraints in noncoding DNA of rice. *BMC Evol. Biol.* 7:208. doi: 10.1186/1471-2148-7-208

Guo, Z., Wang, H., Tao, J., Ren, Y., Xu, C., Wu, K., et al. (2019). Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. *Mol. Breed.* 39:37. doi: 10.1007/s11032-019-0940-4

Hanania, U., Velcheva, M., Sahar, N., and Perl, A. (2004). An improved method for isolating high-quality DNA from *Vitis vinifera* nuclei. *Plant Mol. Biol. Report* 22, 173–177. doi: 10.1007/BF02772724

Hao, X., Yang, T., Liu, R., Hu, J., Yao, Y., Burlyaeva, M., et al. (2017). An RNA sequencing transcriptome analysis of Grasspea (*Lathyrus sativus* L.) and development of SSR and KASP markers. *Front. Plant Sci.* 8:1873. doi: 10.3389/fpls.2017.01873

Heffelfinger, C., Fragoso, C. A., Moreno, M. A., Overton, J. D., Mottinger, J. P., Zhao, H., et al. (2014). Flexible and scalable genotyping-by-sequencing strategies for population studies. *BMC Genomics* 15:979. doi: 10.1186/1471-2164-15-979

Hiremath, P. J., Kumar, A., Penmetsa, R. V., Farmer, A., Schlueter, J. A., Chamarthi, S. K., et al. (2012). Large scale development of cost—effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnol. J.* 10, 716–732. doi: 10.1111/j.1467-7652.2012.00710.x

Hsu, F., Kent, J. W., Clawson, H., Kuhn, R. M., Diekhans, M., and Haussler, D. (2006). The UCSC known genes. *Bioinformatics* 22, 1036–1046. doi: 10.1093/bioinformatics/btl048

Hu, J., and Quiros, C. F. (1991). Identification of broccoli and cauliflower cultivars with RAPD markers. *Plant Cell Rep.* 10, 505–511. doi: 10.1007/BF00234583

Huang, S., Deng, L., Guan, M., Li, J., Lu, K., Wang, H., et al. (2013). Identification of genome-wide single nucleotide polymorphisms in allopolyploid crop *Brassica napus*. *BMC Genomics* 14:717. doi: 10.1186/1471-2164-14-717

Hubisz, M. J., Falush, D., Stephens, M., and Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* 9, 1322–1332. doi: 10.1111/j.1755-0998.2009.02591.x/full

Huifang, Y., Zhenqing, Z., Xiaoguang, S., Jiansheng, W., and Honghui, G. (2011). Evaluation of genetic diversity in self-incompatible broccoli DH lines assessed by SRAP markers. *African J. Biotechnol.* 10, 12561–12566. doi: 10.5897/AJB11.822

Islam, M. S., Thyssen, G. N., Jenkins, J. N., and Fang, D. D. (2015). Detection, validation, and application of genotyping-by-sequencing based single nucleotide polymorphisms in upland cotton. *Plant Genome* 8, 1–10. doi: 10.3835/plantgenome2014.07.0034

Jagtap, A. B., Vikal, Y., and Johal, G. S. (2020). Genome-wide development and validation of cost-effective KASP marker assays for genetic dissection of heat stress tolerance in Maize. *Int. J. Mol. Sci.* 21:7386. doi: 10.3390/ijms21197386

Jiang, L., Liu, X., Yang, J., Wang, H., Jiang, J., Liu, L., et al. (2014). Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. *BMC Genomics* 15:1105. doi: 10.1186/1471-2164-15-1105

Jones, H., and Mackay, I. (2015). Implications of using genomic prediction within a high-density SNP dataset to predict DUS traits in barley. *Theor. Appl. Genet.* 128, 2461–2470. doi: 10.1007/s00122-015-2601-2

Jones, H., Norris, C., Smith, D., Cockram, J., Lee, D., O'Sullivan, D. M., et al. (2013). Evaluation of the use of high-density SNP genotyping to implement UPOV Model 2 for DUS testing in barley. *Theor. Appl. Genet.* 126, 901–911. doi: 10.1007/s00122-012-2024-2

Jung, J., Park, S.-W., Liu, W. Y., and Kang, B.-C. (2010). Discovery of single nucleotide polymorphism in Capsicum and SNP markers for cultivar identification. *Euphytica* 175, 91–107. doi: 10.1007/s10681-010-0191-2

Krasileva, K. V., Vasquez-Gross, H. A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., et al. (2017). Uncovering hidden variation in polyploid wheat. *Proc. Natl. Acad. Sci. U.S.A.* 114, E913–E921. doi: 10.1073/pnas.1619268114

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, L., Fang, Z., Zhou, J., Chen, H., Hu, Z., Gao, L., et al. (2017). An accurate and efficient method for large-scale SSR genotyping and applications. *Nucleic Acids Res.* 45:e88. doi: 10.1093/nar/gkx093

Li, Z., Mei, Y., Liu, Y., Fang, Z., Yang, L., Zhuang, M., et al. (2019). The evolution of genetic diversity of broccoli cultivars in China since 1980. *Sci. Hortic. (Amsterdam)* 250, 69–80. doi: 10.1016/j.scienta.2019.02.034

Liao, N., Hu, Z., Li, Y., Hao, J., Chen, S., Xue, Q., et al. (2020). Ethylene responsive factor 4 is associated with the desirable rind hardness trait conferring cracking resistance in fresh fruits of watermelon. *Plant Biotechnol. J.* 18, 1066–1077. doi: 10.1111/pbi.13276

Linck, E., and Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Mol. Ecol. Resour.* 19, 639–647. doi: 10.1111/1755-0998.12995

Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282

Livak, K. J., Flood, S. J. A., Marmaro, J., Giusti, W., and 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 Res.* 4, 357–362. doi: 10.1101/gr.4.6.357

Louarn, S., Torp, A. M., Holme, I. B., Andersen, S. B., and Jensen, B. D. (2007). Database derived microsatellite markers (SSRs) for cultivar differentiation in *Brassica oleracea*. *Genet. Resour. Crop Evol.* 54, 1717–1725. doi: 10.1007/s10722-006-9181-6

Lu, X., Liu, L., Gong, Y., Zhao, L., Song, X., and Zhu, X. (2009). Cultivar identification and genetic diversity analysis of broccoli and its related species with RAPD and ISSR markers. *Sci. Hortic. (Amsterdam)* 122, 645–648. doi: 10.1016/j.scienta.2009.06.017

Luo, Z., Iaffaldano, B. J., Zhuang, X., Fresnedo-Ramírez, J., and Cornish, K. (2017). Analysis of the first *Taraxacum kok-saghyz* transcriptome reveals potential rubber yield related SNPs. *Sci. Rep.* 7:9939. doi: 10.1038/s41598-017-09034-2

Ouyang, P., Kang, D., Mo, X., Tian, E., Hu, Y., and Huang, R. (2018). Development and characterization of high-throughput EST-based SSR markers for *Pogostemon cablin* using transcriptome sequencing. *Molecules* 23:2014. doi: 10.3390/molecules23082014

Pandino, G., Lombardo, S., Moglia, A., Portis, E., Lanteri, S., and Mauromicale, G. (2015). Leaf polyphenol profile and SSR-based fingerprinting of new segregant *Cynara cardunculus* genotypes. *Front. Plant Sci.* 5:800. doi: 10.3389/fpls.2014.00800

Pariasca-Tanaka, J., Lorieux, M., He, C., McCouch, S., Thomson, M. J., and Wissuwa, M. (2015). Development of a SNP genotyping panel for detecting polymorphisms in *Oryza glaberrima*/*O. sativa* interspecific crosses. *Euphytica* 201, 67–78. doi: 10.1007/s10681-014-1183-4

Park, S., Yu, H.-J., Mun, J.-H., and Lee, S.-C. (2010). Genome-wide discovery of DNA polymorphism in *Brassica rapa*. *Mol. Genet. Genomics* 283, 135–145. doi: 10.1007/s00438-009-0504-0

Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.

Rasheed, A., Wen, W., Gao, F., Zhai, S., Jin, H., Liu, J., et al. (2016). Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. *Theor. Appl. Genet.* 129, 1843–1860. doi: 10.1007/s00122-016-2743-x

Rauscher, G., and Simko, I. (2013). Development of genomic SSR markers for fingerprinting lettuce (*Lactuca sativa* L.) cultivars and mapping genes. *BMC Plant Biol.* 13:11. doi: 10.1186/1471-2229-13-11

Rett-Cadman, S., Colle, M., Mansfeld, B., Barry, C. S., Wang, Y., Weng, Y., et al. (2019). QTL and transcriptomic analyses implicate cuticle transcription factor *SHINE* as a Source of natural variation for epidermal traits in cucumber fruit. *Front. Plant Sci.* 10:1536. doi: 10.3389/fpls.2019.01536

Saxena, R. K., Varma Penmetsa, R., Upadhyaya, H. D., Kumar, A., Carrasquilla-Garcia, N., Schlueter, J. A., et al. (2012). Large-scale development of cost-effective single-nucleotide polymorphism marker assays for genetic mapping in *Pigeonpea* and comparative mapping in *Legumes*. *DNA Res.* 19, 449–461. doi: 10.1093/dnares/dss025

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol. Breed.* 33, 1–14. doi: 10.1007/s11032-013-9917-x

Senthilvel, S., Shaik, M., Anjani, K., Shaw, R. K., Kumari, P., Sarada, C., et al. (2017). Genetic variability and population structure in a collection of inbred lines derived from a core germplasm of *Castor*. *J. Plant Biochem. Biotechnol.* 26, 27–34. doi: 10.1007/s13562-016-0356-8

Slatkin, M., and Barton, N. H. (1989). A Comparison of three indirect methods for estimating average levels of gene flow. *Evolution (N. Y)* 43, 1349–1368. doi: 10.2307/2409452

Stansell, Z., Hyma, K., Fresnedo-Ramírez, J., Sun, Q., Mitchell, S., Björkman, T., et al. (2018). Genotyping-by-sequencing of *Brassica oleracea* vegetables reveals unique phylogenetic patterns, population structure and domestication footprints. *Hortic. Res.* 5:38. doi: 10.1038/s41438-018-0040-3

Steele, K. A., Quinton-Tulloch, M. J., Amgai, R. B., Dhakal, R., Khatiwada, S. P., Vyas, D., et al. (2018). Accelerating public sector rice breeding with high-density KASP markers derived from whole genome sequencing of Indica rice. *Mol. Breed.* 38:38. doi: 10.1007/s11032-018-0777-2

Thomson, M. J., Singh, N., Dwiyanti, M. S., Wang, D. R., Wright, M. H., Perez, F. A., et al. (2017). Large-scale deployment of a rice 6 K SNP array for genetics and breeding applications. *Rice* 10:40. doi: 10.1186/s12284-017-0181-2

Tian, H. L., Wang, F. G., Zhao, J. R., Yi, H. M., Wang, L., Wang, R., et al. (2015). Development of maizeSNP3072, a high-throughput compatible SNP array, for DNA fingerprinting identification of Chinese maize varieties. *Mol. Breed.* 35:136. doi: 10.1007/s11032-015-0335-0

Tommasini, L., Batley, J., Arnold, G., Cooke, R., Donini, P., Lee, D., et al. (2003). The development of multiplex simple sequence repeat (SSR) markers to complement distinctness, uniformity and stability testing of rape (*Brassica napus* L.) varieties. *Theor. Appl. Genet.* 106, 1091–1101. doi: 10.1007/s00122-002-1125-8

Tonguç, M., and Griffiths, P. D. (2004). Genetic relationships of *Brassica* vegetables determined using database derived simple sequence repeats. *Euphytica* 137, 193–201. doi: 10.1023/B:EUPH.0000041577.84388.43

Trebbi, D., Maccaferri, M., de Heer, P., Sørensen, A., Giuliani, S., Salvi, S., et al. (2011). High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum Desf.*). *Theor. Appl. Genet.* 123, 555–569. doi: 10.1007/s00122-011-1607-7

Wang, J., Gu, H., Yu, H., Zhao, Z., Sheng, X., and Zhang, X. (2012). Genotypic variation of glucosinolates in broccoli (*Brassica oleracea* var. *italica*) florets from China. *Food Chem.* 133, 735–741. doi: 10.1016/j.foodchem.2012.01.085

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, 1–7. doi: 10.1093/nar/gkq603

Wang, M. L., Zhu, C., Barkley, N. A., Chen, Z., Erpelding, J. E., Murray, S. C., et al. (2009). Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. *Theor. Appl. Genet.* 120, 13–23. doi: 10.1007/s00122-009-1155-6

Wright, S. (1943). Isolation by distance. *Genetics* 28, 114–138.

Wright, S. (1965). The Interpretation of population structure by F-Statistics with special regard to systems of mating. *Evolution (N. Y)* 19:395. doi: 10.2307/2406450

Xu, C., Ren, Y., Jian, Y., Guo, Z., Zhang, Y., Xie, C., et al. (2017). Development of a maize 55 K SNP array with improved genome coverage for molecular breeding. *Mol. Breed.* 37:20. doi: 10.1007/s11032-017-0622-z

Yang, G., Chen, S., Chen, L., Sun, K., Huang, C., Zhou, D., et al. (2019). Development of a core SNP arrays based on the KASP method for molecular breeding of rice. *Rice* 12:21. doi: 10.1186/s12284-019-0272-3

Yang, J., Zhang, J., Han, R., Zhang, F., Mao, A., Luo, J., et al. (2019). Target ssr-seq: a novel SSR genotyping technology associate with perfect SSRs in genetic analysis of cucumber varieties. *Front. Plant Sci.* 10:531. doi: 10.3389/fpls.2019.00531

Yang, Z., and Yoder, A. D. (1999). Estimation of the transition/transversion rate bias and species sampling. *J. Mol. Evol.* 48, 274–283. doi: 10.1007/PL00006470

Yoon, M. S., Song, Q. J., Choi, I. Y., Specht, J. E., Hyten, D. L., and Cregan, P. B. (2007). BARCSoySNP23: a panel of 23 selected SNPs for soybean cultivar identification. *Theor. Appl. Genet.* 114, 885–899. doi: 10.1007/s00122-006-0487-8

Zhang, J., Yang, J., Zhang, L., Luo, J., Zhao, H., Zhang, J., et al. (2020). A new SNP genotyping technology Target SNP-seq and its application in genetic analysis of cucumber varieties. *Sci. Rep.* 10:5623. doi: 10.1038/s41598-020-62518-6

Zhao, S., Li, A., Li, C., Xia, H., Zhao, C., Zhang, Y., et al. (2017). Development and application of KASP marker for high throughput detection of *AhFAD2* mutation in peanut. *Electron. J. Biotechnol.* 25, 9–12. doi: 10.1016/j.ejbt.2016.10.010

Zheng, X., Cheng, T., Yang, L., Xu, J., Tang, J., Xie, K., et al. (2019). Genetic diversity and DNA fingerprints of three important aquatic vegetables by EST-SSR markers. *Sci. Rep.* 9:14074. doi: 10.1038/s41598-019-50569-3

Zhu, H., Zhai, W., Li, X., and Zhu, Y. (2019). Two QTLs controlling clubroot resistance identified from bulked segregant sequencing in Pakchoi (*Brassica campestris* ssp. *chinensis Makino*). *Sci. Rep.* 9:9228. doi: 10.1038/s41598-019-44724-z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Shen, Wang, Shaw, Yu, Sheng, Zhao, Li and Gu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.