



OPEN ACCESS

EDITED BY

Ambika Baldev Gaikwad,
Indian Council of Agricultural Research
(ICAR), India

REVIEWED BY

Changlong Wen,
Beijing Vegetable Research Center, China
Gung Pyo Lee,
Chung-Ang University, Republic of Korea

*CORRESPONDENCE

Zhuming Tan
✉ tanzm@hnkly.cn

†These authors have contributed equally to
this work

RECEIVED 23 July 2024

ACCEPTED 17 December 2024

PUBLISHED 08 January 2025

CITATION

Shen B, Shen A, Tan Y, Liu L, Li S and Tan Z
(2025) Development of KASP markers,
SNP fingerprinting and population genetic
analysis of *Cymbidium ensifolium* (L.) Sw.
germplasm resources in China.
Front. Plant Sci. 15:1460603.
doi: 10.3389/fpls.2024.1460603

COPYRIGHT

© 2025 Shen, Shen, Tan, Liu, Li and Tan. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Development of KASP markers, SNP fingerprinting and population genetic analysis of *Cymbidium ensifolium* (L.) Sw. germplasm resources in China

Baoming Shen[†], Airong Shen[†], Yun Tan, Lina Liu,
Sainan Li and Zhuming Tan*

Institute of Biodiversity, Hunan Academy of Forestry, Changsha, China

Cymbidium ensifolium (L.) Sw. is a valuable ornamental plant in the genus *Cymbidium*, family Orchidaceae, with high economic and ecological significance. However, the lack of population genetic information and molecular markers has hindered the development of the sales market and genetic breeding of *C. ensifolium* despite the abundance of commercial cultivars available. In this study, we aimed to develop a set of single nucleotide polymorphism (SNP) markers to distinguish the main cultivated *C. ensifolium* cultivars in China and provide technical support for domestic cultivar protection, registration, and market rights protection. A total of 1,280,516 high-quality loci were identified from 10,021,591 SNPs obtained by sequencing 50 *C. ensifolium* commercial cultivars using double digest restriction site-assisted DNA sequencing technology. A total of 7,599 SNPs were selected for competitive allele-specific PCR (KASP) primer design, and 4,360 were successfully designed as KASP markers. Population structure analysis revealed that the 50 commercial cultivars were best divided into four populations, with some correlation between the group distribution and the morphological and geographical characteristics of the germplasm. Using the genotyping results from 28 KASP markers screened from the cultivars, a minimum set of 11 markers was identified that could distinguish 83 *C. ensifolium* commercial cultivars completely, with the remaining 17 markers serving as extended markers. The average PIC value of the 11 markers was 0.345, which was considered medium polymorphism. DNA fingerprints were constructed for the 83 cultivars on the basis of the 11 KASP markers, providing a new approach for mapping DNA fingerprints in *C. ensifolium* cultivars with high efficiency, accuracy, and low cost compared with traditional methods.

KEYWORDS

Cymbidium ensifolium, SNP, KASP, genetic diversity, fingerprint

Introduction

Cymbidium ensifolium (L.) Sw. is a perennial evergreen herbaceous plant in the Orchidaceae family. *C. ensifolium* has a karyotype of $2N = 2X = 40$ with chromosomes of different lengths and the genome size is 3.62 Gb (Ai et al., 2021; Li et al., 2002). It is an important species in traditional Chinese *cymbidium* due to its abundant flower colors, diverse flower types, diverse leaf colorations and unique historical and cultural significance (Cao et al., 2022). *C. ensifolium* has significant ornamental, cultural, medicinal and economic value (Jimoh et al., 2022; Yang and Zhu, 2015). There is a wide variety of *C. ensifolium*, with significant price differences among different varieties (Huang, 2012). However, since *C. ensifolium* varieties are distinguished mainly by flower type and color and some varieties have minimal differences, identification becomes challenging, especially during nonflowering periods (Chen et al., 2022). Traditional morphological identification methods are often insufficient, leading to inaccurate identification and causing discrepancies in orchid trading. This situation adversely affects the robust and orderly development of the orchid market. Additionally, the lack of accurate and reliable identification methods has resulted in confusion between identical *C. ensifolium* varieties with different names and different varieties with the same name, which hampers the identification, conservation, improvement, cultivation, and utilization of *C. ensifolium* genetic resources (Huang, 2012). Therefore, there is an urgent need to establish a simple, stable, and reliable identification method for *C. ensifolium* varieties.

DNA molecular markers have advantages such as short cycle times, minimal environmental influence, and high-throughput detection, providing new means for variety identification (Ramesh et al., 2020). Currently, several molecular marker technologies, including Random Amplified Polymorphic DNA (RAPD), which is based on ITS and cpDNA fragments, Inter-Simple Sequence Repeat (ISSR), Simple Sequence Repeat (SSR) markers, and fluorescent SSR markers, have been applied for resource identification of *C. ensifolium* varieties (Li et al., 2014; Jin, 2019; Li, 2014; Wang et al., 2021a; Ai et al., 2019; Hu et al., 2008). However, RAPD is a dominant marker that cannot distinguish between heterozygous and homozygous genotypes, limiting its use in genetic analysis and genetic map construction (Congiu et al., 2000). ITS and limited cpDNA fragment-based markers can only differentiate a small number of varieties because of the limited number of polymorphic sites (Sivu et al., 2022). SSR markers are limited in quantity and detection throughput, have higher detection costs, and require time-consuming and labor-intensive data interpretation, as well as subjective misjudgment of band patterns due to human factors, which restrict their use in a wider range of variety identification work (Zhang et al., 2022b).

Single nucleotide polymorphisms (SNPs) refer to DNA sequence polymorphisms caused by variations such as substitutions and inversions of individual bases in the genomic DNA sequence. SNPs have advantages such as large quantity, wide distribution, allelic dimorphism, and stable inheritance (Li et al., 2023a). These characteristics make SNP-based molecular markers the latest generation of markers. With the emergence of various high-throughput SNP detection platforms, they can effectively

compensate for the technical limitations of SSR markers. The construction of a DNA fingerprint map based on SNP markers is highly important for variety specificity identification, assessment of genetic variations, authenticity verification, identification of seed purity, and other characteristics (Nguyen et al., 2020; Josia et al., 2021; Wang et al., 2021b, 2015; Su et al., 2019). This method has been designated as one of the recommended marker methods by the International Union for the Protection of New Varieties of Plants (UPOV) and the General Guidelines for DNA Identification of Plant Varieties Using DNA markers (NY/T 2594-2016) (Button, 2008). The development of SNP molecular markers can be based on different DNA sources, such as SNP markers developed on the basis of specific genes, EST-SNP markers developed on the basis of express sequence tag (EST) sequences, GSS-SNP markers developed on the basis of gene survey sequence (GSS), and genomic SNPs developed on the basis of whole-genome data (Addison et al., 2020; Wang et al., 2019, 2020; Ashwath et al., 2023). Among these, genomic SNPs provide the most accurate identification but are relatively expensive. With the development of sequencing technology and the popularization of next-generation sequencing (NGS), the cost of sequencing has greatly decreased (van Dijk et al., 2014). Using transcriptome sequencing (RNA-seq), restriction site-associated DNA sequencing (RAD-seq) and double digest restriction site-associated DNA sequencing (ddRAD-seq), more abundant SNP loci can be obtained (Wang et al., 2015; Baird et al., 2008; Magbanua et al., 2023). Among these, RAD-seq and ddRAD-seq can identify SNP loci with broader coverage than can RNA-seq.

There are also various methods for detecting SNP molecular markers, such as direct sequencing, TaqMan probe, amplification refractory mutation system PCR (ARMS-PCR; also known as allele-specific PCR), molecular beacon, high-resolution melting analysis (HRM) technology, cleaved amplified polymorphic sequence (CAPS), SNaPshot, kompetitive allele-specific PCR (KASP), gene chip, and mass spectrometry (Zhao et al., 2017; Kovalchuk and Arkhipova, 2023; Zongze et al., 2018; Gerasimova et al., 2013; Gomes et al., 2018; Amanullah et al., 2022; Pradhan et al., 2023; Guan et al., 2023; Franklin et al., 2020; Wang et al., 2016). Among these, KASP technology has the advantages of low cost, high throughput, time and labor savings, and convenience (He et al., 2014). It has become one of the main methods for SNP genotyping and insertion/deletion (InDel) detection internationally and has been successfully applied in genetic typing and variety breeding of grain crops and economic crops, such as wheat, rice, maize, strawberry, grape, broccoli, cotton, tobacco, and peach (Liu et al., 2023; Tang et al., 2022; Chen et al., 2021; Yang et al., 2020; Wang et al., 2022; Shen et al., 2021; Zhao et al., 2021; Wang et al., 2021b; Fleming et al., 2022). However, there are currently no reports on the application of SNP markers developed on the basis of KASP technology for variety identification, fingerprint map construction, and systematic classification in orchids such as *Cymbidium*.

Therefore, we utilized KASP technology to screen a set of SNP markers that can distinguish *C. ensifolium* germplasm resources in China. In this study, 50 *C. ensifolium* germplasm resources were subjected to Illumina NovaSeq sequencing (Supplementary Table 1). The results were compared with those of the reference genome to identify SNP core markers and interpret their genetic relationships,

genetic diversity, and population structure. Additionally, a DNA fingerprint of 83 *C. ensifolium* varieties was created to effectively differentiate between different varieties of *C. ensifolium* (Supplementary Table 2). These results provide a scientific foundation and data reference for genetic diversity analysis, variety identification, and molecular breeding of *C. ensifolium*.

Materials and methods

Plant materials

Wild *C. ensifolium* resources are included in the National Key Protected Wild Plant List in China, and it is prohibited by law to collect them from the wild. This study adhered to all relevant institutional, national, and international guidelines and laws. No prior permission was required to conduct research on this species. All the plants used in this study are well-known commercial cultivars in China cultivated in pots in a greenhouse under controlled temperature and lighting conditions, a greenhouse on day/night temperatures of 30/23°C under a 14 h light/10 h dark photoperiod. The plant material used in this study was identified by Prof. Zhuming Tan (see author list) and stored at the laboratory of the Hunan Academy of Forestry. To encompass more genetic diversity and obtain a broader range of SNP markers, the selected materials represented diverse morphological variations of *C. ensifolium*, including leaf, sepal, lip, and petal color, size and shape, leaf color and shape, sepal size and shape, lip size and shape, and petal size and shape (Figure 1). Five vigorous plants were randomly selected for each germplasm resource, and young leaf samples were collected, frozen at -80°C, and stored for DNA extraction.

For the preliminary screening of KASP markers, 50 representative *C. ensifolium* cultivars were used (Supplementary Table 1). The rescreening and fingerprint construction using KASP markers included 83 *C. ensifolium* cultivars (Supplementary Table 2).

DNA extraction, library construction and simplified genome sequencing

After the leaf samples were ground, genomic DNA was extracted using a D312 Universal Plant DNA Kit (Genepioneer, China), and the quality and concentration of the DNA were measured, and the purity ($OD_{260\text{ nm}}/OD_{280\text{ nm}}=1.8-2.0$) was determined using a NanoDrop2000 UV spectrophotometer (Thermo Fisher, Waltham, MA, USA). A paired-end library with a length range of 300-500 bp was subsequently constructed using digest restriction site-associated DNA sequencing library construction of qualified sample DNA (Peterson et al., 2012). First, 500 ng of genomic DNA was incubated with 0.6 U EcoRI (NEB), T4 DNA ligase (NEB), ATP (NEB), and EcoRI connectors (including index sequences of differentiated samples) at 37°C for 3 h and annealed at 65°C for 1 h. The restriction enzyme NlaIII (NEB) and the NlaIII connector were then added and incubated for 3 h at 37°C. After the reaction, the endonuclease was inactivated at 65°C for 30 min in a polymerase chain reaction (PCR) amplifier. The 400-600 bp digested products were recovered using a MiniBEST Agarose Gel DNA Extraction Kit (Takara, China) and quantified using a NanoDrop2000 UV spectrophotometer (Thermo Fisher, Waltham, MA, USA). After 50 samples were mixed in equal quantities, the Illumina NovaSeq 6000 PE150 platform was sequenced and used to construct a DNA library of the mixed products.

Raw data quality control

Raw data quality control was conducted on the original sequenced reads. Fastp software (version: 0.20.0, <https://github.com/OpenGene/fastp>) was used to eliminate reads with an unknown base number $N<5$, a quality value <5 , connector sequences, and other low-quality sequences to obtain clean data. Burrows-Wheeler Aligner (BWA) (version: 0.7.17-R1188, <https://github.com/lh3/bwa>) was then used to align the sequenced reads with the reference genome (Li and

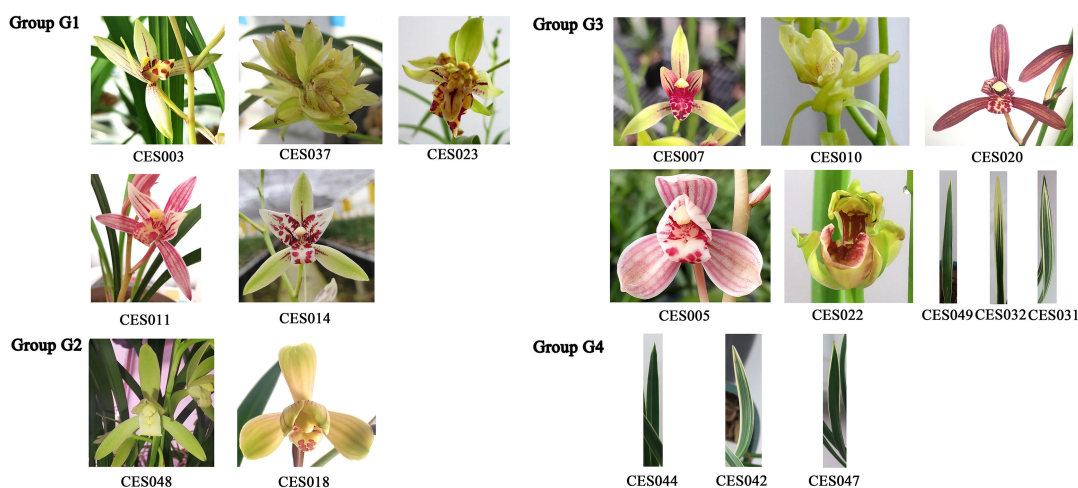


FIGURE 1
Representative flower and leaf characteristics of the tested *C. ensifolium* varieties.

Durbin, 2009). The reference genome used was the GWHBCII000000000 genome Fasta (*C. ensifolium*) (download website: <https://ngdc.cncb.ac.cn/gwh/Assembly/20686/show>), with parameters set as $-M -R [51]$. The insert size and coverage depth of each sample were determined, and variation was detected by comparing the positions of the clean reads in the reference genome (Li et al., 2023a). SAMtools software (version: v1.9, <https://github.com/samtools/samtools>) was used to statistically analyze the sequencing depth, genome coverage, insertion fragment length, and other information of each simplified genome sample (Li et al., 2009).

SNP analysis

SNP analysis was primarily performed using the Genome Analysis Toolkit (GATK) software package (version: 4.1.4.1, <https://github.com/broadinstitute/gatk>). Using the positioning of clean reads in the reference genome, GATK was used to detect SNPs and obtain the final SNP site set, followed by SNP statistics (McKenna et al., 2010). The main detection process included the following steps: (1) Picard's (version: 0.7.17-r1188, <https://github.com/broadinstitute/gatk>) Mark Duplicate tool was used to remove duplicates and mask the effects of PCR duplication on the results from BWA alignment; (2) variant calling was performed using GATK, including SNP and InDel; and (3) variant quality score recalibration (VQSR) was performed using GATK. (4) GATK was used to filter the obtained variants (filtering out sites with $QD < 2.0$, $MQ < 40.0$, $FS > 60.0$, $SOR > 6.0$, $MQRankSum < -12.5$, $ReadPosRankSum < -8.0$), selecting reliable variant results, followed by statistical analysis and bar chart plotting of the variant types by GATK. SNP site annotation was implemented with SnpEff (version: 4.3t, <http://pcingola.github.io/SnpEff>).

Population SNP filtering

To obtain high-quality SNPs for population genetic analysis of *C. ensifolium*, a series of standards were applied for preliminary screening. These standards included criteria such as average sequencing depth $\geq 5\times$, minor allele frequency (MAF) ≥ 0.05 , information integrity ≥ 0.70 , SNP quality value $Q \geq 30$, and number of alleles of 2 (Li et al., 2023a). The SNP density distribution across each chromosome was subsequently calculated and visualized. The average values of the population genetic indices in 50 *C. ensifolium* samples were also calculated and statistically analyzed.

Genetic diversity and population structure analysis

Genetic diversity and population structure analyses were performed using GCTA software (version: 1.92.1, <http://cnsngomics.com/software/gcta/#Overview>) for principal component analysis (PCA) on the basis of the high-quality SNPs obtained. The maximum likelihood (ML) method in RAxML

software (version 8.2.12, <https://github.com/stamatak/standard-RAxML/>) was used to construct an evolutionary tree of 50 *C. ensifolium* samples. Admixture software (version: 1.3.0, <http://software.genetics.ucla.edu/admixture/>) was used to analyze the population genetic structure.

SNP site filtering and KASP primer design

Due to the difficulty in distinguishing between samples in this study, a combination of conventional filtering and manual selection of difficult-to-distinguish varieties was used in the SNP site selection strategy. After considering both filtering criteria, the selection criteria for SNP sites used in KASP primer development in this study were defined as follows: first, conserve the sequence flanking the SNP site on the DNA chain of the chromosome for more than 50 bp; second, retain markers with an average depth of 5X or greater, with the SNP being a biallelic gene; third, trim 100 bp sequences upstream and downstream of the SNP marker, then use blast software (version: 2.10.1+, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) to align the sequence against the reference genome and remove markers with multiple alignment positions; fourth, retain markers with a polymorphic information content (PIC) greater than 0.020, (the low PIC threshold is due to manual selection considerations), and the PIC calculation formula refers to the method of Zhang, et al., where P_i and P_j are the frequency of occurrence of the two alleles of SNP in all varieties tested, and l is the number of samples (Zhang et al., 2020).

$$PIC=1-\sum_{i=0}^l P_i^2-\sum_{i=1}^{l-1}\sum_{j=i+1}^l 2P_i^2 P_j^2$$

For each selected SNP marker, primer design was conducted using Primer3 software (version: 2.4.0, <https://sourceforge.net/projects/primer3/files/latest/download>). Sequences 100 bp upstream and downstream of candidate SNPs were used for KASP marker design. Two allele-specific primers and one universal primer were designed for each KASP target site. The primer design parameters were set as follows: GC content $< 60\%$, melting temperature (T_m) between 57 and 63°C, and PCR product size not exceeding 120 bp. Further optimization of primer design results can be conducted on the basis of actual verification results. Only SNP sites with successful primer design were considered to be qualified KASP markers and were used for downstream analysis. The distribution statistics of the markers with successful primer design were also based on the annotation results.

Verification of SNP locus authenticity and KASP genotyping

To save costs, 28 randomly selected SNP sites with successful KASP primer design were preliminarily screened using Perl scripts, with 5 samples selected for first-generation sequencing validation. The distribution of developed KASP markers on the chromosomal was visualized using Tertools-II (version: 2.138) (Chen et al., 2023).

Primers were designed, conventional sequencing was performed, and the authenticity of the SNP site was verified

TABLE 1 Primers used for the initial screening validation of candidate SNP loci.

| Number | SNP Position | Forward primer (5'-3') | Reverse primer (5'-3') | Product length/bp |
|--------|----------------|----------------------------|----------------------------|-------------------|
| 1 | Chr1-55623523 | AGCGTGGTGGGATTGAACTT | ACGAAAGCACGACGTGTACA | 242 |
| 2 | Chr1-73862970 | GATTGTCCTATTCAAGTCTCATTAGA | ACCTTTCTTGGCTTGGCAA | 164 |
| 3 | Chr1-171445320 | ACCAAAGTGTGAGGCAACCA | CCCTCGACCGGTAAGTTCAC | 196 |
| 4 | Chr10-25134624 | CGGAGCAAACCTCATCACCT | GCAACACACCAGAGGTACCA | 191 |
| 5 | Chr12-53446226 | CTCTCGAGGAGCTATTGCAG | GCCCACTAAAGTGCAAAGCA | 495 |
| 6 | Chr12-63867683 | AGAAGCGATCAAGAACCGGG | AGCATGTCTCACACTGGT | 278 |
| 7 | Chr12-75662664 | TTAGATGGGCCAGGCTTGTG | GGTGCATGGTGAGTTCAGCT | 547 |
| 8 | Chr13-6407723 | GGATGGCTCTGCTTCAGTT | CCTAACTACGCCCTTCGCAA | 364 |
| 9 | Chr13-34339702 | TCAGTTTCGGACGAAGGTCG | ATGTTTGTGCCACGCGTTTC | 271 |
| 10 | Chr16-43018260 | CTATGGGCGTAGAAGTGGTGT | AGTGTCAAATTCGCGCAAC | 367 |
| 11 | Chr16-66004509 | GGCGCATCCAAAGAACAAGG | CCTCTAATATGATCTTGAAGTGCACC | 308 |
| 12 | Chr18-11319799 | CCAAGAACCCTCATGTCCC | TGAGCCATGATAATTGCAAGGT | 381 |
| 13 | Chr18-27339793 | TCGCTTATCAGAAGTGCCACA | TGCTGTTTACTGCCATCCA | 532 |
| 14 | Chr19-8308837 | TGGCAGAAACCTATAGGCCT | TTCCCTGCCAAGCTCAATGT | 177 |
| 15 | Chr2-12538713 | ACGGTTGCTCGGCTTTTCTA | ACCATCCAAACCGTAAGCTTCT | 329 |
| 16 | Chr2-77710898 | GTCCCAAAGACAAGTCCCA | GGCTTGCTAATACGTGGCAC | 221 |
| 17 | Chr2-114570305 | TGCAAGCAAGATGTCTCCGT | CCACGCACTTACAGGAGGTT | 244 |
| 18 | Chr2-140839794 | CGGAGCTTCAATAGTTAACAACAGT | TCTTTCTGATGCAGGGCTGG | 475 |
| 19 | Chr20-7711682 | TGGTTCCACCTTGCCTGTT | CTGATGGCTCCTGGTGTGTT | 208 |
| 20 | Chr20-9861983 | TGAATATACATTTCAACCATGGAGTT | TTTCCCTTGCAGACTTGC | 237 |
| 21 | Chr3-67201341 | TCGCTCCTTCGCTTGTTCAT | CAAGCAGCCCTTGGTATTGA | 549 |
| 22 | Chr3-153114895 | GCCAAGCCTGAGGTGTTAGA | TGTGAGTCAGCTGCACAGT | 190 |
| 23 | Chr4-57124129 | CTCCACCACCTGCATCCATT | TGGATATCCATATGAAGGTGTTGCA | 440 |
| 24 | Chr4-103337174 | TCCTGGTTCCCTGATCCCAT | GATAATGGGGCCGAGAAGG | 471 |
| 25 | Chr5-47997902 | TGCATCCTAACAAGACTTGCA | ACAGTCAGTTAGAGCGCAACA | 500 |
| 26 | Chr8-90715849 | GAGTCAGGTTTCCAACGCG | AACATGTCGCCGTACAACCT | 335 |
| 27 | Chr9-22416891 | GGACTTCAAGGAAGTGGAGGA | AGAAAAGAAACCGCGCAAAC | 511 |
| 28 | Chr9-24025081 | ACCTTGTTGGCCGCTTTTA | AAGCCGAGCTGCAATAAACCC | 431 |

Each SNP location information consists of the number of chromosomes where the SNP site is located plus the location of the corresponding chromosome.

(Table 1). For successfully validated sites, KASP primers were redesigned using the real first-generation sequencing results with Primer3 (version: 2.4.0, <https://sourceforge.net/projects/primer3/files/latest/download>). The primers were synthesized by Sangon Biotech (Shanghai) Co., Ltd., with FAM- or VIC-tails (FAM-tail: 5'-GAAGGTGACCAAGTTCATGCT-3'; VIC-tail: 5'-GAAGTTCGGAGTCAACGGATT-3') (Supplementary Excel 1). Subsequently, KASP genotyping experiments were subsequently conducted.

The KASP assay was performed in a 5 μ L PCR system/condition comprising 2.5 μ L of 2X KASP Master mix (JasonGen, China), 1.25 μ L of primer mixture, and 1.25 μ L of DNA at a concentration of 10–20 ng/mL. The PCR program included 10 min at 95°C, 10 touchdown cycles of 95°C for 20 s and 61–55°C for 60 s (decreasing by 0.6°C per cycle),

and 27 cycles of 95°C for 20 s and 55°C for 60 s. Following PCR, fluorescence data were read and analyzed via the CFX Connect™ Real-Time System (Bio-Rad, USA). If the genotyping results were unsatisfactory, PCR optimization was performed with an additional 3 cycles of 95°C for 20 s and 55°C for 60 s.

Fingerprint construction

For the obtained SNP genotyping results, the optimal combination of markers was calculated for fingerprinting using a Perl program. The genotypes of the optimal combination of markers were heatmapped for fingerprinting, with each row representing one SNP locus and each column representing one

sample. The genotypes were color-coded as follows: AA=green, AG=light pink, CC=yellow, CT=grew, GT=dark red, TT=blue, AC=pink, AT=orange, CG=light blue, and GG=purple; and no call genotypes were designated as NN=white.

Results

Simplified genome sequencing and reference genome alignment

The sequencing of 50 *C. ensifolium* samples on the Illumina NovaSeq 6000 PE150 yielded a total of 64.89 Gb of clean data, with an average sequencing data volume of 1.30 GB per sample. The Q30 scores ranged from more than 89.79% to 94.26%, with an average exceeding 92.76%. The average Q score was between 35.24 and 36.07, and the GC content ranged from 33.86% to 35.17% (Supplementary Table 3). The clean reads obtained were mapped to the reference genome, with an average mapping efficiency of 99.50%. The genome average cover depth was 6.6X, the genome coverage at 5X averaged 1.92%, and that at 10X averaged 0.95% (Supplementary Table 4).

Selection, identification and annotation of high-quality SNPs

Following sequencing, GATK software detected numerous SNP variants in the 50 *C. ensifolium* samples, resulting in a total of 10,021,591 SNPs, with each sample ranging from 691,307 to 1,612,362 (Supplementary Table 5). The most common types of SNP variations were C>T (2,034,818) and G>A (2,037,584), followed by A>G (1,289,885) and T>C (1,286,472), with the least common types being C>G (233,292) and G>C (233,361) (Supplementary Figure 1A, Supplementary Excel 3). A total of 1,280,516 filtered SNPs distributed across 20 main chromosomes were obtained for population genetic analysis (Supplementary Figure 1B, Table 2). An SNP distribution map was then created for the 20 main chromosomes on the basis of the number and density of SNPs (Figure 2). Chromosome GWHBCII00000001 had the highest number of SNPs (67,609), whereas GWHBCII000000020 had the lowest number of SNPs (14,535). The SNPs were evenly distributed on the 20 main chromosomes, with an average of 322 SNPs per MB (data not shown).

Genetic relationships and population structure analysis

Admixture software was used to analyze the population structure of the 50 *C. ensifolium* samples. The population was

divided into four subgroups (G1, G2, G3, G4) on the basis of the lowest cross-validation error rate at K=4 (Figure 3A). G2 included 5 varieties, which primarily descended from a common ancestor and predominantly represented purebred varieties. G4 comprised 8 varieties, predominantly hybrids, including several leaf art varieties. The G1 and G3 subgroups were more diverse, encompassing a variety of flower colors, patterns, and leaf art varieties. G1 comprised 14 varieties, with only CES011 descending from 3 ancestors, while the rest originated from a common ancestor. G3 included 23 varieties, with 16 having a single ancestral origin (Figures 3B, D).

PCA was conducted using GCTA software on high-quality SNPs from the 50 *C. ensifolium* samples, resulting in the samples being discriminated into four groups on the basis of the first two components, accounting for 36.99% and 6.79% of the total variation, respectively (Figure 3C).

An evolutionary tree was also constructed using RAxML software, which clustered the 50 *C. ensifolium* samples into four groups (Figure 4), which was consistent with the PCA and population structure analysis results and was supported by high bootstrap values. The clusters presented a certain degree of similarity in terms of phenotype and origin (Supplementary Table 1).

Development and verification of highly polymorphic KASP SNP markers

To develop high-quality and polymorphic KASP SNP markers that could effectively differentiate the tested samples, stringent criteria were applied, including SNP conservation and uniqueness on the chromosome DNA strands, high sequencing depth, biallelic polymorphisms, and high PIC values. A total of 7,599 SNPs were screened from 10,021,591 SNPs, with 4,360 successfully designed as KASP markers (Supplementary Excel 2). Among these KASP markers, 428 were located in exonic regions, accounting for 9.8% (Supplementary Figure 2).

To validate the practicality of the selected SNP markers, we utilized a Perl program to screen 28 loci from the 4,360 SNPs mentioned above for verification (Supplementary Excel 4, Supplementary Figure 3). The first-generation sequencing results indicated that all 28 loci were indeed present (Table 1). On the basis of the actual sequencing results, we subsequently optimized the primer design for KASP and conducted KASP genotyping experiments (Supplementary Table 3). Among the 50 simplified sequencing samples, most samples were well genotyped and effectively distinguished, but two groups (CES001, CES002, CES019, CES048; CES035, CES043) could not be differentiated (data not shown). Despite multiple rounds of screening, these two

TABLE 2 Average population genetic index values for the 50 *C. ensifolium* varieties genome simplification sequencing samples within the population.

| Population | Sites | Polymorphic | HetObs | HetExp | Pi | Fis | PIC | Ne | MAF |
|------------|-----------|-------------|--------|--------|-------|-------|-------|-------|-------|
| All | 1,280,516 | 1,280,516 | 0.203 | 0.270 | 0.270 | 0.249 | 0.225 | 1.423 | 0.186 |

The population genetic index values are averages. Population, population ID; Sites, number of SNP markers within the population; Polymorphic, number of polymorphic markers; HetObs, observed heterozygosity; HetExp, expected heterozygosity; Pi, nucleotide polymorphism within the population; Fis, average inbreeding coefficient of the population; PIC, polymorphic information content; Ne, effective number of alleles; MAF, minimum allele frequency.

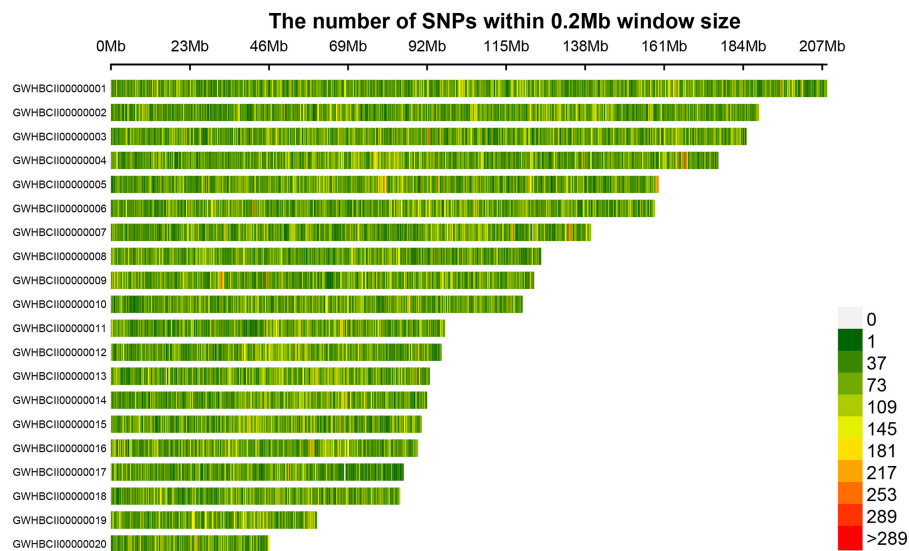


FIGURE 2 SNP density distribution on each chromosome. The horizontal axis represents the chromosome length, and the vertical axis represents the chromosome number. Different colors represent the number of SNPs in different regions.

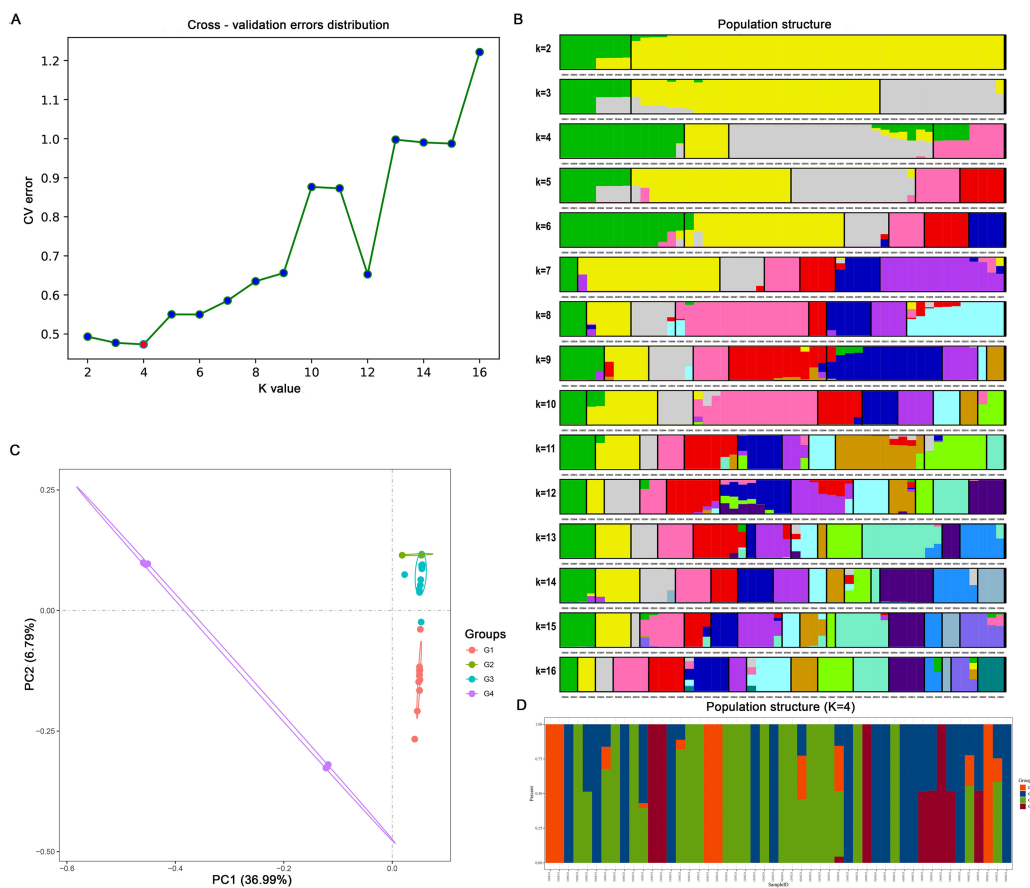
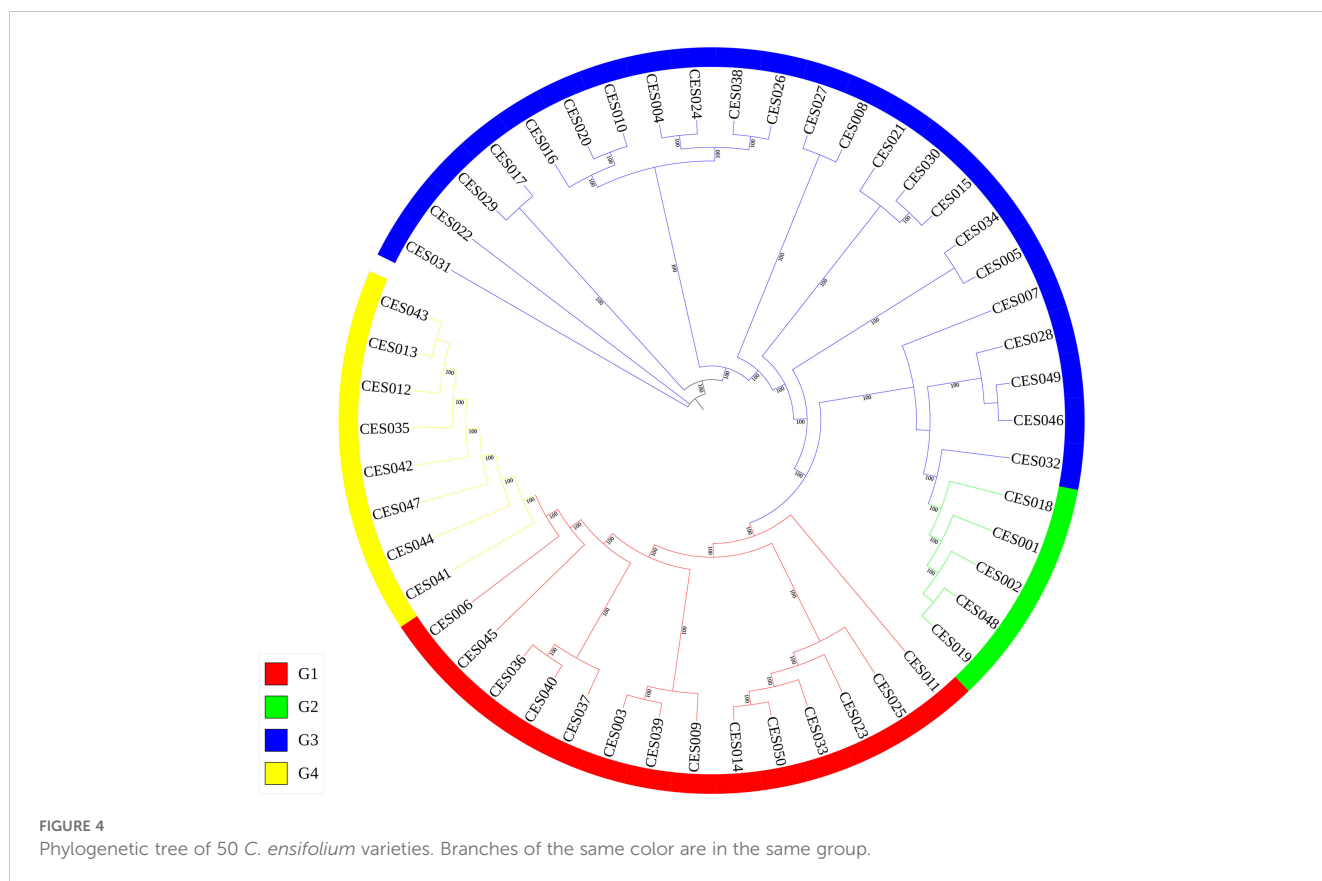


FIGURE 3 Bioinformatic analysis of 50 *C. ensifolium* varieties based on single nucleotide polymorphisms (SNPs). **(A)** The cross-validation error rate corresponding to different K values. **(B)** Population structure of 50 *C. ensifolium* varieties at different K values. The K value represents the cross-validation error rate. **(C)** A two-dimensional diagram of principal component analysis (PCA). **(D)** Population structure of 50 *C. ensifolium* samples when K = 4.



groups could not be distinguished, making the development of KASP markers difficult. Considering the cost of developing markers, this study did not continue to develop KASP primers to differentiate these similar varieties. To verify the utility of the selected KASP markers, we selected another 37 cultivated varieties of *C. ensifolium* for KASP genotyping. The results revealed that four markers (Chr12-75662664, Chr16-66004509, Chr18-11319799, Chr3-153114895) had only one genotype, whereas the other 24 markers were polymorphic and could effectively distinguish the other 37 cultivated varieties of *C. ensifolium* (Figures 5A, B; Supplementary Excel 5). These findings indicate that the selected KASP markers have good usability. Due to the large number of *C. ensifolium* varieties, by expanding the screening range these four KASP primers with consistent genotypes may still exhibit polymorphisms, thus supporting the potential value of the markers developed in this study.

Construction of a DNA fingerprint

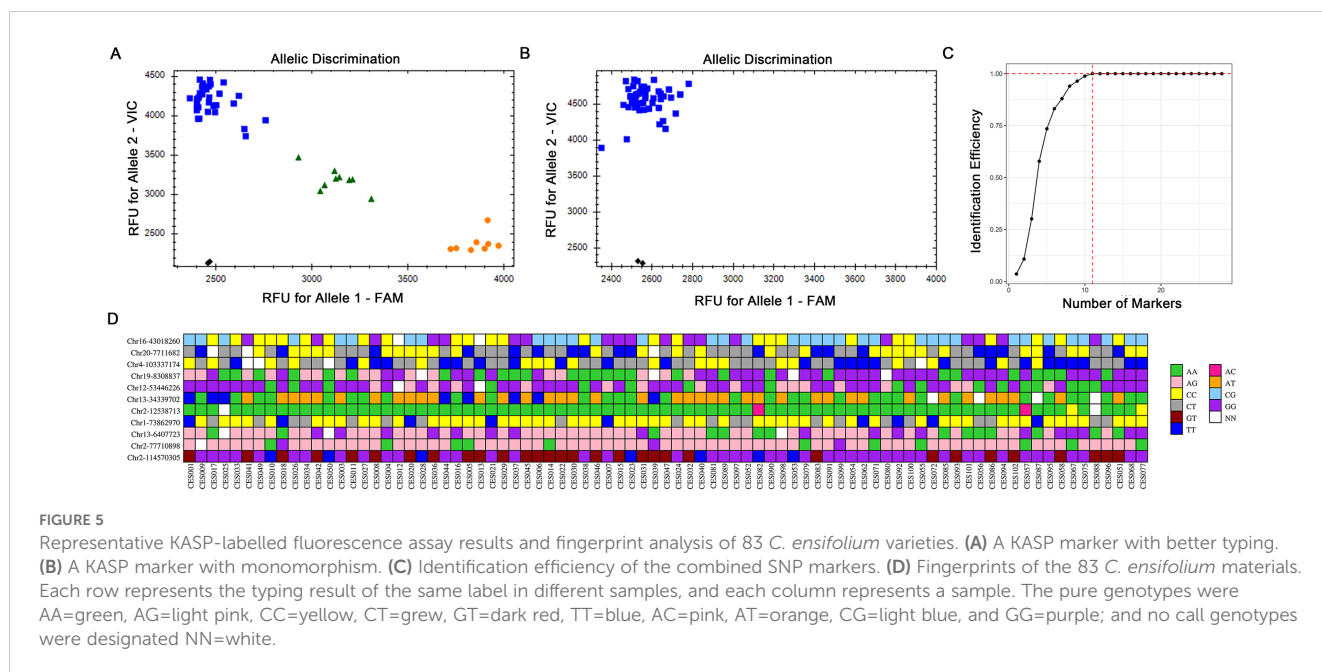
To protect the genetic resources of *C. ensifolium* varieties, we constructed a fingerprint map for all tested varieties on the basis of the KASP marker genotyping results, with one variety selected from each undistinguished group as a representative (CES001, CES035). A good fingerprint map requires a minimum number of markers to distinguish the maximum number of varieties to achieve simplicity, efficiency, and economy. By selecting 11 SNPs from the 28 obtained SNP loci that could distinguish 83 *C. ensifolium* varieties, we

constructed a highly simplified SNP combination distributed across seven chromosomes (Figure 5C). Using these 11 SNPs, we successfully distinguished all 83 varieties, and there was at least one different SNP between each of the two samples. The PIC values of the 11 core KASP SNPs ranged from 0.19 to 0.37, and 90% > 0.33, indicating medium genetic diversity (Supplementary Excel 4). The combination of these 11 SNPs effectively distinguished the *C. ensifolium* varieties, and the fingerprint map constructed on the basis of the genotyping results further confirmed this distinction (Figure 5D). The remaining 17 KASP SNP markers can be used as backup markers when expanding sample testing to complement and improve the fingerprint map (Supplementary Excel 4).

Discussion

SNP-based genetic relationships among the *C. ensifolium*

C. ensifolium is widely distributed in southern China and is highly diverse in terms of flower color, flower type, leaf color, leaf shape, and plant morphology. As a traditional Chinese *Cymbidium* species, *C. ensifolium* is highly valued for its historical, economic, and ornamental significance (Cao et al., 2022; Ai et al., 2021). In the past, wild types with unique traits were sought from the field to breed better varieties of *C. ensifolium*. However, due to the emphasis on the protection of orchid species in China, the collection of wild genetic resources has been prohibited by law. Therefore, breeding new



varieties from existing varieties has become the primary means of creating new orchid varieties. Current studies have focused mainly on the genetic control of specific traits in plants, such as flower color, leaf color, and drought resistance, providing references for future breeding techniques (Ai et al., 2023; Li et al., 2023b; Mei et al., 2024). Before new varieties are bred, it is essential to determine the genetic and kinship relationships between breeding materials (Chen et al., 2020). However, few studies have investigated the genetic and kinship relationships among commercial *C. ensifolium* varieties at the genomic level.

In this study, 50 mainstream commercial varieties of *C. ensifolium* were selected, and using ddRAD-seq technology, a series of polymorphic SNP markers were obtained through strict screening criteria. Genetic and kinship analyses of the 50 commercial varieties of *C. ensifolium* on the basis of SNP sequences were conducted. PCA and population structure analysis indicated that the 50 *C. ensifolium* varieties could be divided into 4 groups, which was also supported by the results of the maximum likelihood phylogenetic tree. The selected *C. ensifolium* varieties in this study overlap with those selected by Ai et al., but the clustering results are different, which may be due to different population constructions and segments selected for building the phylogenetic tree (Ai et al., 2019). Among the 50 selected *C. ensifolium* varieties, most originated from Taiwan, such as *C. ensifolium* var. 'Qi Xiannv', *C. ensifolium* var. 'Shizhang Hong', *C. ensifolium* var. 'Lishan Shiwang', *C. ensifolium* var. 'Hong Niang', *C. ensifolium* var. 'Jin He', *C. ensifolium* var. 'Fushan Qidie', and *C. ensifolium* var. 'Baodao Xiannv', which were all classified in Group 1 (G1), whereas those originating from Sichuan were mostly classified in Group 3 (G3). This finding indicates that the grouping in this study is consistent with the original sources of the varieties. Even though varieties from different regions have undergone long-term domestication, they still retain their own distinct features (Lin et al., 2022). Within the G2 group, most individuals are from

the southeastern coastal regions of China. In addition to the hybrid variety *C. ensifolium* var. 'Zhu Jin', the other varieties also share some common phenotypic characteristics, such as flowers with light-colored sepals and labellums that are solid in color without spots. On the other hand, the G4 group consists mainly of hybrid species. Additionally, some cultivars were not grouped together with others from the same geographical origin, indicating widespread genetic exchange among cultivated varieties.

Hybrid breeding involves selecting parental varieties with a large genetic distance and distant kinship, which aids in selecting new varieties with excellent traits (Geng et al., 2021; Cui et al., 2023). The evolutionary tree in this study shows clear genetic distances and kinship relationships among the 50 *C. ensifolium* varieties. For example, *C. ensifolium* var. 'Zhaojun Xue', *C. ensifolium* var. 'Da Jiangjun', *C. ensifolium* var. 'Huangjin Xianzi', and *C. ensifolium* var. 'Danxia Xiannv' are located on the same branch and have large genetic distances from the remaining varieties, providing theoretical references for the selection of parental varieties in *C. ensifolium* breeding.

Cultivar identification and the advantages of SNP molecular markers

C. ensifolium is a highly valuable ornamental and economic orchid, and through generations of breeding, there are currently an estimated thousand varieties of *C. ensifolium* available on the market (<https://www.hmlan.com/auction/search-101005.htm?q=%BD%A8%C0%BC&noex=>) (accessed on 5 June 2024). Identifying the correct variety of orchids to purchase online during nonflowering periods has become a challenge for many orchid enthusiasts. The traditional classification of *C. ensifolium* varieties is based mainly on morphological characteristics. Wang used 14 quantitative traits and 11 qualitative traits to describe 39

Cymbidium varieties in detail, providing a phenotypic basis for the identification of these varieties (Wang, 2020). Nevertheless, identifying varieties during nonflowering periods remains difficult. To address the shortcomings of morphological identification, researchers have also attempted to use various molecular methods for *C. ensifolium* variety identification. For example, Wang et al. successfully differentiated 9 orchid species, including *C. ensifolium*, using ALFP technology (Wang and Wang, 2014). Wang et al. also distinguished 85 *C. ensifolium* cultivars using 19 ISSR primers (Wang et al., 2011). Hu et al. differentiated 38 *C. ensifolium* varieties using 18 RAPD primers with genetic distances ranging from 0.0420 to 0.5385 [12]. Li et al. developed 55 genic-SSR polymorphic markers from the total RNA of *C. ensifolium* var. Tiegusu and distinguished 9 *Cymbidium* species and 12 *C. ensifolium* cultivars using evolutionary tree construction. The genetic distance ranged from 0.016 to 0.618 (Li et al., 2014). Although these methods can be used to differentiate *C. ensifolium* cultivars, they have limitations such as inconvenient operation, high cost, inability to distinguish genotypes, and low number of polymorphic sites, highlighting the urgent need for a new, simple, accurate, and efficient molecular identification method.

SNP molecular marker technology has numerous advantages, including a large quantity, wide distribution, allelic dimorphism, and stable inheritance (Li et al., 2023a). This technology has been successfully applied in various crops, such as rice, grape, potato, cotton, radish, and honeysuckle (Xing et al., 2024; Wang et al., 2022; Li et al., 2023a; Morales et al., 2020; Gazendam et al., 2022; Kuang et al., 2016). It is one of the marker methods recommended by the UPOV and the general guidelines for the identification of plant varieties using DNA markers (NY/T 2594-2016) (Button, 2008). However, its application in *Cymbidium* species, particularly *C. ensifolium*, is relatively limited (Yang et al., 2023).

In this study, using the ddRAD-seq technology on 50 *C. ensifolium* cultivars, a total of 10,021,591 SNP loci were obtained, surpassing the number of polymorphic markers obtained by Li et al. using genic-SSR markers in *C. ensifolium* (Li et al., 2014). On average, 964,159 SNP loci were developed per cultivar, whereas 334,967 SNP loci were obtained per *Cymbidium sinense* (Jack. ex Andr.) Willd. cultivar using SLAF-seq technology (Yang et al., 2023). These results indicate the presence of a rich SNP locus population in orchids, providing a valuable resource for developing SNP markers. After stringent filtering, 4,360 highly polymorphic SNP loci were selected, and a set of 11 SNP markers were identified that could effectively distinguish 83 *C. ensifolium* cultivars. This identification efficiency is significantly greater than that of traditional *C. ensifolium* classification methods such as AFLP, ISSR, RAPD, and genic-SSR (Wang and Wang, 2014; Wang et al., 2021a, 2011; Hu et al., 2008; Li et al., 2014).

The main advantages of the KASP genotyping technique for identifying varieties

The Laboratory of Government Chemists (LGC) developed a high-throughput genotyping technique based mainly on SNPs

based on the principle of KASP (Dipta et al., 2024). Overall, the KASP genotyping technique offers several advantages for identifying varieties. Compared with other SNP genotyping technologies, it is cost-effective, with lower material expenses per reaction (Ayalew et al., 2019; Yuan et al., 2014). Additionally, it is a simple and gel-free assay that can be easily performed using regular qPCR instruments, reducing labor costs. Compared with AFLP and SSR technology, the automatic genotyping technique, which is based on fluorescence differences, also minimizes the effect of error when the gel electrophoresis image is read (Wang and Wang, 2014; Sun et al., 2023). The design principle behind KASP primers allows quick and accurate genotyping on the basis of SNP polymorphisms, making it a valuable tool for quality control and QTL mapping (Kumar et al., 2022; Zeng et al., 2022). In conclusion, the KASP genotyping technique is a reliable and efficient method for identifying genetic variations in plant varieties.

Due to the numerous advantages of the KASP genotyping technique, KASP assays have been developed for genotyping analysis in a variety of plant species. For example, in conventional and hybrid rice, Tang et al. developed 48 KASP markers, and the 48 KASP markers had a 100% discrimination rate in 53 conventional indica varieties and 193 hybrid varieties (Tang et al., 2022). In cabbage, Li et al. selected 442 KASP SNP markers among 50 resequenced genotypes on the basis of high polymorphism information content, high minor allele frequency, wide average distribution and low heterozygosity. Using the KASP genotyping data, the genetic similarity among three kinds of inbred lines (spring cabbage, autumn cabbage and winter cabbage) was analyzed, and the heterotic groups within each ecotype were classified. Seven heterotic groups were identified for spring cabbage (77), six for autumn cabbage (70), and five for winter cabbage (97) (Li et al., 2020). In *Brassica rapa*, Hong et al. developed 100 accession-specific markers as accession-specific KASP markers. Using the results of their validation experiments, they successfully distinguished the accession-specific markers in individual accessions in test populations from noncore or commercial cultivars (Hong et al., 2022). In addition, KASP markers play important roles in genotyping and variety identification in other plants, such as apple, coffee, and wheat (Winfield et al., 2020; Patterson et al., 2017; Roncallo et al., 2019; Akpertey et al., 2020; Zhang et al., 2023b; Kumar et al., 2022). However, little research has been conducted on SNP markers and KASP marker development in *C. ensifolium*.

In our study, using ddRAD genome simplified sequencing data of 50 *C. ensifolium* commercial varieties, we obtained 10,021,591 SNP sites. After a series of strict screening criteria, we obtained 7,599 SNP sites with high polymorphism information content, high minor allele frequency, and a wide average distribution. Among these, 4,360 SNP markers were successfully converted into KASP markers, accounting for 57.4% of all 7,599 SNP markers. The conversion efficiency of KASP markers was lower than that of Cabbage (88.4%), *Capsicum annuum* L. (88.2%), and rice (94.8%) (Zhang et al., 2023b; Li et al., 2020; Yang et al., 2019). In terms of the conversion efficiency of all SNP sites into KASP sites, this study revealed an efficiency of 0.04%, which was lower than that of maize (2.42%) (Chen et al., 2021). These findings indicate that the development of KASP SNP markers for *C. ensifolium* is relatively

challenging. These KASP SNP markers provide a rich resource database of polymorphic sites for the identification and consistency testing of *C. ensifolium* varieties and even orchid varieties.

Identification and validation of 83 *C. ensifolium* germplasm resources by SNP fingerprints

Germplasm resources, as important biological resources, constitute the genetic basis for breeding high-quality and unique new varieties (Wang et al., 2021b; Li et al., 2023a). With the development of technology, the types and quantity of discovered germplasm resources are becoming increasingly diverse and abundant. Conducting DNA fingerprinting on germplasm resources is a good choice for effective protection and efficient utilization of many germplasm resources. DNA fingerprinting was first proposed by the geneticist Alec Jeffreys from the University of Leicester in 1985. DNA fingerprinting uses isolated human microsatellite DNA as a gene probe, hybridizes it with enzyme-cut fragments of human nuclear DNA, and obtains hybrid bands composed of alleles from multiple loci of different lengths. These fingerprints are unique to each individual, similar to human fingerprints (Jeffreys, 2013). The construction of a DNA fingerprint map can provide each germplasm with a unique identity, assist in accurate identification of germplasm resources, and play an important role in variety specificity and authenticity, seed purity identification, improved resource utilization efficiency, and protection of the intellectual property of plant breeding (Xing et al., 2024; Shen et al., 2021; Yang et al., 2022; Tian et al., 2021). SNP molecular markers are widely used for constructing DNA fingerprint maps in various plants, such as maize, cucumber, honeysuckle, cigar tobacco, radish, and red raspberry (Mannino et al., 2023; Zhang et al., 2024; Wei et al., 2024; Li et al., 2023a; Wang et al., 2021b; Tian et al., 2021; Xing et al., 2024; Clare et al., 2023; Zhang et al., 2022a). The accuracy of SNP fingerprinting has been validated in cucumber through distinctness, uniformity, and stability (DUS) testing, further demonstrating the accuracy and practicality of SNP fingerprinting (Zhang et al., 2022a). However, there have been no reports of SNP fingerprints of *C. ensifolium* germplasm resources.

SNP molecular markers are widely used for gene identification, germplasm characterization, and variety fingerprinting (Yang and Zhu, 2015; Chen et al., 2022; Magbanua et al., 2023; Zhao et al., 2017). However, it has been reported that only a small proportion of SNP loci can be selected and genotyped successfully. The authenticity of SNP sites in simplified floral genome sequencing is low, making the validation of SNP site authenticity even more important for constructing a fingerprint map (Liu et al., 2022a). Our study also revealed a low authenticity of SNP sites in simplified sequencing sites, leading to difficulties in KASP primer development. Using 50 simplified sequencing varieties, KASP primers were developed using Perl programs, resulting in 28 KASP markers. Most *C. ensifolium* varieties could be easily distinguished, but there were still two groups that could not be differentiated even after multiple KASP primer developments, namely, CES001, CES002, CES019, and CES048 and CES035, and CES043. This may be due to the limited number of 28 KASP-SNP markers and the limited genomic variations that can be

revealed, necessitating the development of more *C. ensifolium* KASP-SNP markers. Additionally, the simplified genome data used for marker development cover only approximately 2% of the genome, limiting the variation that can be revealed. Furthermore, the two groups may be closely related, with minimal genetic differences (Mei, 2023). Therefore, one representative variety (CES001 and CES035) was selected from each group for fingerprint construction. Furthermore, 37 commercially cultivated *C. ensifolium* varieties were added to validate the utility of the KASP markers. These 28 markers were successfully used for genotyping 83 commercial *C. ensifolium* varieties.

The construction of a molecular fingerprint map of germplasm resources requires the use of a minimal number of primers to differentiate the maximum number of germplasms. Therefore, selecting appropriate primers is an important prerequisite for constructing a molecular fingerprint map (Wang et al., 2021b; Li et al., 2023a). Based on the principle of using the minimum number of markers to differentiate the maximum number of varieties, 11 KASP markers were selected from 28 KASP markers (PIC: 0.195-0.375; MAF: 0.125-0.479; HE: 0.219-0.499), which could efficiently differentiate all 83 *C. ensifolium* commercial varieties with at least one genotypic difference. On average, one SNP locus could identify 8 varieties, which is higher than the SNP marker differentiation efficiency developed in Chinese flowering cabbage (18 core SNP markers could completely differentiate all 89 cabbage varieties) and melon (40 core SNP markers efficiently differentiate 99% of the 259 commercial melon varieties) (Ren et al., 2023; Zhang et al., 2023a). However, it is lower than the SNP marker differentiation efficiency developed in radish, which has reached 24 accessions per SNP locus (15 core SNP markers could completely differentiate all 356 radish varieties) (Xing et al., 2024). This may be attributed to the relatively high quantity and quality of the genomic database used for radish SNP marker development (Xing et al., 2024). The development of radish markers has demonstrated the great potential and advantages of using SNP loci to differentiate varieties, prompting us to continuously expand the genomic database of *C. ensifolium* to develop more efficient SNP markers.

It is noted that there are some varieties that cannot be distinguished when using SNP loci to differentiate species, which has been observed in crops like melon and cigar tobacco (Zhang et al., 2023a; Wang et al., 2021b). In a study on cigar tobacco, Yanyan et al. utilized 47 core KASP markers to differentiate 216 cigar tobacco germplasm resources, and found that some varieties could not be distinguished. Through phenotype analysis, these were identified as synonyms (Wang et al., 2021b). The phenomenon of synonyms also exists in the *C. ensifolium* market, where vendors may alter the phenotype of orchids through physical or chemical treatments to create fake new varieties, and rename existing varieties arbitrarily (Ning, 2022; Li, 1994). This confusion in commercial varieties is detrimental to the conservation and breeding of orchids, and SNP marker-based variety identification can greatly regulate the *C. ensifolium* variety market. In cases where differentiation is not possible, DUS determination (NY/T 2441-2013 Guidelines for the conduct of tests for distinctness, uniformity and stability *Cymbidium*) can be used for further confirmation of varieties.

The PIC value is considered the most important indicator of the usefulness of molecular markers. Markers with high PIC values are usually highly polymorphic, whereas markers with low PIC values are considered less polymorphic. Furthermore, markers with PIC values >0.5 are usually considered highly polymorphic, those with PIC values of $0.25 < \text{PIC} \leq 0.5$ are usually considered medium polymorphic, and those with PIC values ≤ 0.25 are usually considered lowly polymorphic (Liu et al., 2022b). The 11 core markers required for fingerprint map construction had a PIC range of 0.195–0.375, with 90% > 0.336 indicating medium polymorphism, demonstrating the practicality and reliability of fingerprint map construction. Therefore, the 11 KASP markers are clearly reliable, effective, and accurate in detecting 83 *C. ensifolium* germplasm resources.

However, these 83 *C. ensifolium* commercial varieties constitute only a small portion of the total commercial varieties of *C. ensifolium*. To distinguish total commercial varieties, especially similar varieties, more SNP markers with greater discriminatory ability are necessary. Although only 28 out of 4,360 KASP markers were verified, verifying the remaining markers could identify more high-quality markers. These 83 commercial varieties include various typical shapes of *C. ensifolium* commercial varieties, as well as newly bred hybrid varieties in China in recent years. Additionally, utilizing SNP molecular markers and fingerprints comprehensively can effectively improve the identification capabilities of *C. ensifolium* varieties containing wild varieties, which is essential for wild resource protection, and further research is needed in this area.

Conclusion

This study provides comprehensive information about the genetic diversity of *C. ensifolium* commercial cultivars in China on the basis of a population of 50 *C. ensifolium* commercial cultivars. A series of SNPs were discovered by ddRAD-seq of 50 diverse *C. ensifolium* commercial cultivars, and the SNPs were converted into KASP panels for the genotyping of a large set of *C. ensifolium* commercial cultivars. Phylogenetic and PCA analyses revealed that the 50 *C. ensifolium* commercial cultivars were divided into four well-separated clusters, and a correlation was observed between the group distribution and the geographical origin of the *C. ensifolium* germplasm. A set of 28 KASP SNP markers was screened, and a minimum set of 11 KASP SNP markers (with 90% PIC values >0.336) was verified to distinguish 83 *C. ensifolium* commercial cultivars completely. The 11 SNP genotypes of each *C. ensifolium* variety were used to generate SNP fingerprints of a major collection (83) of cultivated *C. ensifolium* varieties in China. The KASP markers developed in this study could also be utilized for evaluating the variety authenticity of *C. ensifolium* cultivars. This is the first study to measure the diversity and population structure of a large collection of *C. ensifolium* in China on the basis of SNPs from simplified genome sequencing and the first application of KASP techniques in *C. ensifolium* for genetic studies. This is also the first study to construct a fingerprint chart of *C. ensifolium* commercial cultivars using SNP markers. The information generated in this study will aid in the selection of suitable genotypes for the breeding of new cultivars and provide a scientific basis and technical support

for the protection and identification of new *C. ensifolium* cultivars and wild germplasm resources.

Data availability statement

The datasets generated for this study can be found in the NCBI database (BioProject accession number: PRJNA1127271). The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

BS: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Visualization, Writing – original draft, Writing – review & editing. AS: Data curation, Formal analysis, Methodology, Writing – review & editing. YT: Resources, Validation, Writing – review & editing. LL: Resources, Validation, Writing – review & editing. SL: Resources, Validation, Writing – review & editing. ZT: Methodology, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Hunan Provincial Forestry Ecological Protection, Restoration and Development Special Fund Project (Xiangcai Zihuan Zhi (2022) No. 64; Xiangcai Zihuan Zhi (2023) No. 26; Xiangcai Zihuan Zhi (2023) No. 72).

Acknowledgments

We would like to thank the Department of Wildlife Conservation, The Forestry Department of Hunan Province, for providing project financial support and the staff of Nanjing Genepioneer Biotechnologies Co., Ltd., for their support in simplified genome sequencing and SNP analyses and verification. We also thank the peer reviewers for their suggestions.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2024.1460603/full#supplementary-material>

SUPPLEMENTARY EXCEL 1

Information on 28 *C. ensifolium* KASP marker primers.

SUPPLEMENTARY EXCEL 2

Information on 4360 *C. ensifolium* SNP loci.

SUPPLEMENTARY EXCEL 3

Statistical table of the number of variations in the SNP type.

SUPPLEMENTARY EXCEL 4

Genetic diversity information of 28 KASP markers in *C. ensifolium* materials.

SUPPLEMENTARY EXCEL 5

KASP genotyping results of 28 markers for 83 *C. ensifolium* varieties.

References

- Addison, C. K., Angira, B., Kongchum, M., Harrell, D. L., Baisakh, N., Linscombe, S. D., et al. (2020). Characterization of haplotype diversity in the BADH2 aroma Gene and development of a KASP SNP assay for predicting aroma in U.S. Rice. *Rice (N Y)* 13, 47. doi: 10.1186/s12284-020-00410-7
- Ai, Y., Chen, L., Xie, T. X., Chen, J., Lan, S. R., and Peng, D. H. (2019). Construction of core collection of *Cymbidium ensifolium* cultivars based on SSR fluorescent markers. *Acta Hort. Sin.* 46, 1999–2008. doi: 10.16420/j.issn.0513-353x.2019-0327
- Ai, Y., Li, Z., Sun, W.-H., Chen, J., Zhang, D., Ma, L., et al. (2021). The *Cymbidium* genome reveals the evolution of unique morphological traits. *Horticulture Res.* 8, 255. doi: 10.1038/s41438-021-00683-z
- Ai, Y., Zheng, Q. D., Wang, M. J., Xiong, L. W., Li, P., Guo, L. T., et al. (2023). Molecular mechanism of different flower color formation of *Cymbidium ensifolium*. *Plant Mol. Biol.* 113, 193–204. doi: 10.1007/s11103-023-01382-0
- Akpertey, A., Padi, F. K., Meinhardt, L., and Zhang, D. (2020). Effectiveness of single nucleotide polymorphism markers in genotyping germplasm collections of *Coffea canephora* using KASP assay. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.612593
- Amanullah, S., Osae, B. A., Yang, T., Li, S., Abbas, F., Liu, S., et al. (2022). Development of whole genome SNP-CAPS markers and preliminary QTL mapping of fruit pedicel traits in watermelon. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.879919
- Ashwath, M. N., Lavale, S. A., Santhoshkumar, A. V., Mohapatra, S. R., Bhardwaj, A., Dash, U., et al. (2023). Genome-wide association studies: an intuitive solution for SNP identification and gene mapping in trees. *Funct. Integr. Genomics* 23, 297. doi: 10.1007/s10142-023-01224-8
- Ayalaw, H., Tsang, P. W., Chu, C., Wang, J., Liu, S., Chen, C., et al. (2019). Comparison of TaqMan, KASP and rhAmp SNP genotyping platforms in hexaploid wheat. *PLoS One* 14, e0217222. doi: 10.1371/journal.pone.0217222
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3, e3376. doi: 10.1371/journal.pone.0003376
- Button, P. (2008). The international union for the protection of new varieties of plants (UPOV) recommendations on variety denominations. *Acta Hort.* 799, 191–200. doi: 10.17660/ActaHort.2008.799.27
- Cao, H., Li, H., Chen, X., Zhang, Y., Lu, L., Li, S., et al. (2022). Insight into the molecular mechanisms of leaf coloration in *Cymbidium ensifolium*. *Front. Genet.* 13. doi: 10.3389/fgene.2022.923082
- Chen, M. K., Chen, L., Sun, W. H., Ma, S. H., Lan, S. R., Peng, D. H., et al. (2022). Genetic diversity analysis and core collection of *Cymbidium ensifolium* germplasm resources. *Acta Hort. Sin.* 49, 175–186. doi: 10.16420/j.issn.0513-353x.2021-0312
- Chen, Y., Tang, C. L., Li, H., Qian, L., Luo, S. K., and Zhou, X. R. (2020). Analysis of affinity of *Bauhinia variegata* germplasm resources based on ISSR markers. *Southwest China J. Agric. Sci.* 33, 26–31. doi: 10.16213/j.cnki.scjas.2020.1.005
- Chen, Z., Tang, D., Ni, J., Li, P., Wang, L., Zhou, J., et al. (2021). Development of genetic KASP SNP markers from RNA-Seq data for map-based cloning and marker-assisted selection in maize. *BMC Plant Biol.* 21, 157. doi: 10.1186/s12870-021-02932-8
- Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Xu, J., et al. (2023). TBtools-II: A “one for all, all for one” bioinformatics platform for biological big-data mining. *Mol. Plant* 16, 1733–1742. doi: 10.1016/j.molp.2023.09.010
- Clare, S. J., King, R. M., Hardigan, M., Dossett, M., Montanari, S., Chagné, D., et al. (2023). Development of KASP fingerprinting panel for clonal identification in red raspberry (*Rubus idaeus* L.). *Plant Breed.* 142, 798–808. doi: 10.1111/pbr.13141
- Congiu, L., Chicca, M., Cella, R., Rossi, R., and Bernacchia, G. (2000). The use of random amplified polymorphic DNA (RAPD) markers to identify strawberry varieties: a forensic application. *Mol. Ecol.* 9, 229–232. doi: 10.1046/j.1365-294x.2000.00811.x
- Cui, X. Q., Huang, C. Y., Deng, J. L., Li, X. M., Li, X. L., and Zhang, Z. B. (2023). SNP markers development and genetic relationship analysis of *Dendrobium* germplasms using SLAF-seq technology. *Biotechnol. Bull.* 39, 141–148. doi: 10.13560/j.cnki.biotech.bull.1985.2022-1386
- Dipta, B., Sood, S., Mangal, V., Bhardwaj, V., Thakur, A. K., Kumar, V., et al. (2024). KASP: a high-throughput genotyping system and its applications in major crop plants for biotic and abiotic stress tolerance. *Mol. Biol. Rep.* 51, 508. doi: 10.1007/s11033-024-09455-z
- Fleming, M. B., Miller, T., Fu, W., Li, Z., Gasic, K., and Saski, C. (2022). Ppe.XapF: High throughput KASP assays to identify fruit response to *Xanthomonas arboricola* pv. pruni (Xap) in peach. *PLoS One* 17, e0264543. doi: 10.1371/journal.pone.0264543
- Franklin, R. N., Karim, N., Goecker, Z. C., Durbin-Johnson, B. P., Rice, R. H., and Parker, G. J. (2020). Proteomic genotyping: Using mass spectrometry to infer SNP genotypes in pigmented and non-pigmented hair. *Forensic Sci. Int.* 310, 110200. doi: 10.1016/j.forsciint.2020.110200
- Gazendam, I., Mojabelo, P., and Bairu, M. W. (2022). Potato cultivar identification in south Africa using a custom SNP panel. *Plants (Basel)* 11, 1546. doi: 10.3390/plants11121546
- Geng, X., Qu, Y., Jia, Y., He, S., Pan, Z., Wang, L., et al. (2021). Assessment of heterosis based on parental genetic distance estimated with SSR and SNP markers in upland cotton (*Gossypium hirsutum* L.). *BMC Genomics* 22, 123. doi: 10.1186/s12864-021-07431-6
- Gerasimova, Y. V., Ballantyne, J., and Kolpashchikov, D. M. (2013). Detection of SNP-containing human DNA sequences using a split sensor with a universal molecular beacon reporter. *Methods Mol. Biol.* 1039, 69–80. doi: 10.1007/978-1-62703-535-4_5
- Gomes, S., Castro, C., Barrias, S., Pereira, L., Jorge, P., Fernandes, J. R., et al. (2018). Alternative SNP detection platforms, HRM and biosensors, for varietal identification in *Vitis vinifera* L. using F3H and LDOX genes. *Sci. Rep.* 8, 5850. doi: 10.1038/s41598-018-24158-9
- Guan, S., Li, W., Jin, H., Zhang, L., and Liu, G. (2023). Development and validation of a 54K genome-wide liquid SNP chip panel by target sequencing for dairy goat. *Genes (Basel)* 14, 1122. doi: 10.3390/genes14051122
- He, C., Holme, J., and Anthony, J. (2014). SNP genotyping: the KASP assay. *Methods Mol. Biol.* 1145, 75–86. doi: 10.1007/978-1-4939-0446-4_7
- Hong, S., Choi, S. R., Kim, J., Jeong, Y. M., Kim, J. S., Ahn, C. H., et al. (2022). Identification of accession-specific variants and development of KASP markers for assessing the genetic makeup of *Brassica rapa* seeds. *BMC Genomics* 23, 326. doi: 10.1186/s12864-022-08567-9
- Hu, W., Huang, R. Z., Pan, X. H., Li, J. F., and Sun, D. (2008). RAPD analysis of thirty-eight *Cymbidium ensifolium* cultivars. *Acta Hort. Sin.* 2, 289–294. doi: 10.16420/j.issn.0513-353x.2008.02.012
- Huang, H. X. (2012). *Study on the economic and the cultural value in Chinese orchids* (Fujian: Fujian Agriculture and Forestry University) 2012.
- Jeffreys, A. J. (2013). The man behind the DNA fingerprints: an interview with Professor Sir Alec Jeffreys. *Invest. Genet.* 4, 21. doi: 10.1186/2041-2223-4-21
- Jimoh, T. O., Costa, B. C., Chansrinyom, C., Chaotham, C., Chanvorachote, P., Rojsittisak, P., et al. (2022). Three new dihydrophenanthrene derivatives from *Cymbidium ensifolium* and their cytotoxicity against cancer cells. *Molecules* 27, 2222. doi: 10.3390/molecules27072222
- Jin, F. (2019). *Analysis of Cymbidium ensifolium transcriptomic database, and the development and application of Genic-SSR markers* (Hubei: Hubei University) 2019.
- Josia, C., Mashingaidze, K., Amelework, A. B., Kondwakwenda, A., Musvosvi, C., and Sibiyi, J. (2021). SNP-based assessment of genetic purity and diversity in maize hybrid breeding. *PLoS One* 16, e0249505. doi: 10.1371/journal.pone.0249505
- Kovalchuk, S. N., and Arkhipova, A. L. (2023). Development of TaqMan PCR assay for genotyping SNP rs211250281 of the bovine agpat6 gene. *Anim. Biotechnol.* 34, 3250–3255. doi: 10.1080/10495398.2022.2077742

- Kuang, M., Wei, S. J., Wang, Y. Q., Zhou, D. Y., Ma, L., Fang, D. D., et al. (2016). Development of a core set of SNP markers for the identification of upland cotton cultivars in China. *J. Integr. Agr.* 15, 954–962. doi: 10.1016/S2095-3119(15)61226-6
- Kumar, S., Pradhan, A. K., Kumar, U., Dhillion, G. S., Kaur, S., Budhlakoti, N., et al. (2022). Validation of Novel spot blotch disease resistance alleles identified in unexplored wheat (*Triticum aestivum* L.) germplasm lines through KASP markers. *BMC Plant Biol.* 22, 618. doi: 10.1186/s12870-022-04013-w
- Li, X. H. (1994). How to identify fake “precious orchids. *China Flower Penjing* 10, 32. Available at: https://kns.cnki.net/kcms2/article/abstract?v=Mw9fkKjKJppdeRanRxBKkpcuurTbTis0ud67JHxjK7GQKMEZn6JsxD55mENX3vaGslc0b9n2q2EMw61oQFnu7EhvyOy0rdaiy2dKkF4dTTSRHMg62A2Nu2S-oMDMVkp_ss1hrpuEi0f-lkSIV7W1jyTcbrq6EHt4LqkSfHkJ4U3XqMTmY541zO2NvU&uniplatform=NZKPT&language=CHS
- Li, Y. (2014). Population genetic structure of *Cymbidium ensifolium* from Jianxi based on ITS and cpDNA molecular markers (Nanchang: Nanchang University) 2014.
- Li, J., Chang, X., Huang, Q., Liu, P., Zhao, X., Li, F., et al. (2023a). Construction of SNP fingerprint and population genetic analysis of honeysuckle germplasm resources in China. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1080691
- 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, Y. G., Guo, W. H., and Wu, B. J. (2002). Studies on karyotypes of four species of *Cymbidium* in China. *Acta Botanica Boreali-Occidentalia Sin.* 22, 1438–1444. Available at: https://kns.cnki.net/kcms2/article/abstract?v=Mw9fkKjKJqfpZL1mxF5S5v5nXFxt3f9KXcKGMaNKkW45jj3fDke0Eyqt3adfypl1uwellenSXSJkpcmfWXXRs2N5rFr-PBXOgZtcU2tdoqRZgZj3GS9X0lyguQGfvel7q2SG9-okgrVlFHVsaAgYO_2fv8qi7Zn2uisEqgf-lidS8pd_zZeHouFnliul&uniplatform=NZKPT&language=CHS
- 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, X., Jin, F., Jin, L., Jackson, A., Huang, C., Li, K., et al. (2014). Development of *Cymbidium ensifolium* genic-SSR markers and their utility in genetic diversity and population structure analysis in cymbidiums. *BMC Genet.* 15, 124. doi: 10.1186/s12863-014-0124-5
- Li, X., Liu, L., Sun, S., Li, Y., Jia, L., Ye, S., et al. (2023b). Transcriptome analysis reveals the key pathways and candidate genes involved in salt stress responses in *Cymbidium ensifolium* leaves. *BMC Plant Biol.* 23, 64. doi: 10.1186/s12870-023-04050-z
- Li, Z., Yu, H., Li, X., Zhang, B., Ren, W., Liu, X., et al. (2020). Competitive allele-specific PCR (KASP) genotyping and heterotic group classification of 244 inbred lines in cabbage (*Brassica oleracea* L. var. capitata). *Euphytica* 216, 106. doi: 10.1007/s10681-020-02640-8
- Lin, L., Wang, F., Wu, M., and Wang, S. (2022). ddRAD sequencing-based scanning of genetic variants in *Sargassum fusiforme*. *J. Mar. Sci. Eng.* 10, 958. doi: 10.3390/jmse10070958
- Liu, X. F., Li, X. M., Zhang, G. W., Zhang, P. J., Liu, G., Huang, M., et al. (2022a). Population genetic analysis of *Jacaranda mimosifolia* by RAD high throughput sequencing technique. *J. Trop. Subtrop Bot.* 30, 613–622. doi: 10.11926/jtsb.4517
- Liu, G., Liu, D., Zhang, A., Liu, H., Mía, M. S., Mullan, D., et al. (2023). Identification of KASP markers and candidate genes for drought tolerance in wheat using 90K SNP array genotyping of near-isogenic lines targeting a 4BS quantitative trait locus. *Theor. Appl. Genet.* 136, 190. doi: 10.1007/s00122-023-04438-3
- Liu, Y., Zheng, Y. Q., Li, C. H., Lin, F. R., and Huang, P. (2022b). Genomic characteristics and population genetic variation of *Dalbergia cultrata* Graham ex Benth in China. *For. Res.* 35, 44–53. doi: 10.13275/j.cnki.lykxyj.2022.004.005
- Magbanua, Z. V., Hsu, C. Y., Pechanova, O., Arick, M. 2nd, Grover, C. E., and Peterson, D. G. (2023). Innovations in double digest restriction-site associated DNA sequencing (ddRAD-Seq) method for more efficient SNP identification. *Anal. Biochem.* 662, 115001. doi: 10.1016/j.ab.2022.115001
- Mannino, G., Kunz, R., and Maffei, M. E. (2023). Discrimination of green coffee (*Coffea arabica* and *Coffea canephora*) of different geographical origin based on antioxidant activity, high-throughput metabolomics, and DNA RFLP fingerprinting. *Antioxidants (Basel)* 12, 1135. doi: 10.3390/antiox12051135
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Mei, X. (2023). Identification of cowpea germplasm resources by KASP-SNP molecular markers and screening of cowpea regeneration genotypes (Wuhan: Jiangnan University) 2023.
- Mei, H., Zhang, X., Zhao, F., Ruan, R., and Fu, Q. (2024). Integrated metabolome and transcriptome analysis provides insight into the leaf color change of *Cymbidium ensifolium*. *Acta Physiologiae Plantarum* 46, 50. doi: 10.1007/s11738-024-03671-7
- Morales, K. Y., Singh, N., Perez, F. A., Ignacio, J. C., Thapa, R., Arbelaez, J. D., et al. (2020). An improved 7K SNP array, the C7AIR, provides a wealth of validated SNP markers for rice breeding and genetics studies. *PLoS One* 15, e0232479. doi: 10.1371/journal.pone.0232479
- Nguyen, N. N., Kim, M., Jung, J. K., Shim, E. J., Chung, S. M., Park, Y., et al. (2020). Genome-wide SNP discovery and core marker sets for assessment of genetic variations in cultivated pumpkin (*Cucurbita* spp.). *Hortic. Res.* 7, 121. doi: 10.1038/s41438-020-00342-9
- Ning, H. J. (2022). Classification of species and varieties of Chinese orchid (Beijing: Beijing Forestry University) 2022.
- Patterson, E. L., Fleming, M. B., Kessler, K. C., Nissen, S. J., and Gaines, T. A. (2017). A KASP genotyping method to identify northern watermilfoil, Eurasian watermilfoil, and their interspecific hybrids. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00752
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One* 7, e37135. doi: 10.1371/journal.pone.0037135
- Pradhan, A. K., Budhlakoti, N., Chandra Mishra, D., Prasad, P., Bhardwaj, S. C., Sareen, S., et al. (2023). Identification of novel QTLs/Defense genes in spring wheat germplasm panel for seedling and adult plant resistance to stem rust and their validation through KASP marker assays. *Plant Dis.* 107, 1847–1860. doi: 10.1094/PDIS-09-22-2242-RE
- Ramesh, P., Mallikarjuna, G., Sameena, S., Kumar, A., Gurulakshmi, K., Reddy, B. V., et al. (2020). Advancements in molecular marker technologies and their applications in diversity studies. *J. Biosci.* 45, 123. doi: 10.1007/s12038-020-00089-4
- Ren, H. L., Xu, D. L., Zhang, J., Zou, J. W., Li, G. G., Zhou, X. Y., et al. (2023). Establishment of SNP fingerprinting and identification of Chinese flowering cabbage varieties based on KASP genotyping. *Acta Hort.* 50, 307–318. doi: 10.16420/j.issn.0513-353x.2021-1046
- Roncallo, P. F., Beaufort, V., Larsen, A. O., Dreisigacker, S., and Echenique, V. (2019). Genetic diversity and linkage disequilibrium using SNP (KASP) and AFLP markers in a worldwide durum wheat (*Triticum turgidum* L. var. durum) collection. *PLoS One* 14, e0218562. doi: 10.1371/journal.pone.0218562
- Shen, Y., Wang, J., Shaw, R. K., Yu, H., Sheng, X., Zhao, Z., et al. (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. doi: 10.3389/fpls.2021.655254
- Sivu, A. R., Pradeep, N. S., Pandurangan, A. G., Dwivedi, M. D., and Pandey, A. K. (2022). Phylogenetic relationships of Indian *Mimemyclon* L. (Melastomataceae) based on nrDNA ITS and cpDNA rbcL sequence data. *J. Genet.* 101, 8. doi: 10.1007/s12041-021-01343-w
- Su, J., Zhang, F., Chong, X., Song, A., Guan, Z., Fang, W., et al. (2019). Genome-wide association study identifies favorable SNP alleles and candidate genes for waterlogging tolerance in chrysanthemums. *Hortic. Res.* 6, 21. doi: 10.1038/s41438-018-0101-7
- Sun, N., Chen, J., Wang, Y., Hussain, I., Lei, N., Ma, X., et al. (2023). Development and utility of SSR markers based on Brassica sp. whole-genome in triangle of U. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1259736
- Tang, W., Lin, J., Wang, Y., An, H., Chen, H., Pan, G., et al. (2022). Selection and validation of 48 KASP markers for variety identification and breeding guidance in conventional and hybrid rice (*Oryza sativa* L.). *Rice (N Y)* 15, 48. doi: 10.1186/s12284-022-00594-0
- Tian, H., Yang, Y., Wang, R., Fan, Y., Yi, H., Jiang, B., et al. (2021). Screening of 200 core SNPs and the construction of a systematic SNP-DNA standard fingerprint database with more than 20,000 maize varieties. *Agriculture* 11, 597. doi: 10.3390/agriculture11070597
- van Dijk, E. L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet.* 30, 418–426. doi: 10.1016/j.tig.2014.07.001
- Wang, H. L. (2020). Evaluation of phenotypic traits and genetic diversity of 39 *Cymbidium ensifolium* (Guangxi: Guangxi University) 2020.
- Wang, H. L., Bu, Z. Y., Zeng, Y. H., and Long, Q. Y. (2021a). Study on genetic diversity of *Cymbidium ensifolium* germplasm based on ISSR marker. *Chin. J. Trop. Crops* 42, 2526–2534. doi: 10.3969/j.issn.1000-2561.2021.09.012
- Wang, F. Q., Fan, X. C., Zhang, Y., Liu, C. H., and Jiang, J. F. (2020). Application and prospect of SNP molecular markers in crop variety identification. *J. Plant Genet. Resour* 21, 1308–1320. doi: 10.13430/j.cnki.jpgr.20200309002
- Wang, F. Q., Fan, X. C., Zhang, Y., Sun, L., Liu, C. H., and Jiang, J. F. (2022). Establishment and application of an SNP molecular identification system for grape cultivars. *J. Integr. Agr.* 21, 1044–1057. doi: 10.1016/S2095-3119(21)63654-7
- Wang, Q., Fu, L., Zhang, X., Dai, X., Bai, M., Fu, G., et al. (2016). Expansion of a SNaPshot assay to a 55-SNP multiplex: Assay enhancements, validation, and power in forensic science. *Electrophoresis* 37, 1310–1317. doi: 10.1002/elps.201500353
- Wang, H.-Z., Lu, J.-J., Hu, X., and Liu, J.-J. (2011). Genetic variation and cultivar identification in *Cymbidium ensifolium*. *Plant Systematics Evol.* 293, 101–110. doi: 10.1007/s00606-011-0429-z
- Wang, Y., Lv, H., Xiang, X., Yang, A., Feng, Q., Dai, P., et al. (2021b). Construction of a SNP fingerprinting database and population genetic analysis of cigar tobacco germplasm resources in China. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.618133
- Wang, Y., Shahid, M. Q., Ghouri, F., Ercisli, S., and Baloch, F. S. (2019). Development of EST-based SSR and SNP markers in *Gastrodia elata* (herbal medicine) by sequencing, *de novo* assembly and annotation of the transcriptome. *3 Biotech.* 9, 292. doi: 10.1007/s13205-019-1823-4
- Wang, B., Tan, H. W., Fang, W., Meinhardt, L. W., Mischke, S., Matsumoto, T., et al. (2015). Developing single nucleotide polymorphism (SNP) markers from transcriptome sequences for identification of longan (*Dimocarpus longan*) germplasm. *Hortic. Res.* 2, 14065. doi: 10.1038/hortres.2014.65

- Wang, X. Y., and Wang, C. X. (2014). Genetic diversity analysis of *Cymbidium* cultivars by AFLP molecular markers. *J. Agric.* 4, 69–72. Available at: https://kns.cnki.net/kcms2/article/abstract?v=Mw9fkKjKljqOF1QglCf3HBLbbxm3TpLcOazXokW7QsUrVx_lh9ZUvdPhS7seSq-KCUDfJp1ox8PdSl-c7SoZMUueRwQrSWGkedEwt74R6xvuiq7eEnPspbmASgn-ZGd5BQd6Kx4zgkcpjTpsvY2CYjo1_PCtsQ-hDCgEXBk7lt5V-dihdML_oPkv0sDoxM4&uniplatform=NZKPT&language=CHS
- Wei, X. Y., Liu, H., Ma, H., Bie, T. D., and Sun, Y. (2024). Analysis of genetic diversity and fingerprint construction of 96 *Cymbidium* based on ISSR markers. *J. Plant Genet. Resour* 25, 586–599. doi: 10.13430/j.cnki.jpgr.20230810003
- Winfield, M., Burrige, A., Ordidge, M., Harper, H., Wilkinson, P., Thorogood, D., et al. (2020). Development of a minimal KASP marker panel for distinguishing genotypes in apple collections. *PLoS One* 15, e0242940. doi: 10.1371/journal.pone.0242940
- Xing, X., Hu, T., Wang, Y., Li, Y., Wang, W., Hu, H., et al. (2024). Construction of SNP fingerprints and genetic diversity analysis of radish (*Raphanus sativus* L.). *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1329890
- 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 (N Y)* 12, 21. doi: 10.1186/s12284-019-0272-3
- Yang, F., Guo, Y., Li, J., Lu, C., Wei, Y., Gao, J., et al. (2023). Genome-wide association analysis identified molecular markers and candidate genes for flower traits in Chinese orchid (*Cymbidium sinense*). *Hortic. Res.* 10, uhad206. doi: 10.1093/hr/uhad206
- Yang, Y., Lyu, M., Liu, J., Wu, J., Wang, Q., Xie, T., et al. (2022). Construction of an SNP fingerprinting database and population genetic analysis of 329 cauliflower cultivars. *BMC Plant Biol.* 22, 522. doi: 10.1186/s12870-022-03920-2
- Yang, X. F., Zhang, Z. Y., Miao, L. X., Zhang, Y. C., Shen, L., Qin, Q. P., et al. (2020). Preliminary analysis on the molecular differences of fruit colouring between Yuexin strawberry and its somaclonal mutation. *Acta Hortic. Sin.* 47, 1999–2008. doi: 10.16420/j.issn.0513-353x.2019-0996
- Yang, F., and Zhu, G. (2015). Digital gene expression analysis based on *De Novo* transcriptome assembly reveals new genes associated with floral organ differentiation of the orchid plant *Cymbidium ensifolium*. *PLoS One* 10, e0142434. doi: 10.1371/journal.pone.0142434
- Yuan, J., Wen, Z., Gu, C., and Wang, D. (2014). Introduction of high throughput and cost effective SNP genotyping platforms in Soybean. *Plant Genetics Genomics Biotechnol.* 2, 90–94. doi: 10.5147/pggb.2014.0144
- Zeng, Z., Guo, C., Yan, X., Song, J., Wang, C., Xu, X., et al. (2022). QTL mapping and KASP marker development for seed vigor related traits in common wheat. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.994973
- Zhang, Z., Cao, Y., Wang, Y., Yu, H., Wu, H., Liu, J., et al. (2023b). Development and validation of KASP markers for resistance to *Phytophthora capsici* in *Capsicum annuum* L. *Mol. Breed* 43, 20. doi: 10.1007/s11032-023-01367-3
- Zhang, S., Li, Y., Li, Y., Zhang, Y., Hao, Y., Hou, Z., et al. (2024). Development of SSR markers for and fingerprinting of walnut genetic resources. *Forests* 15, 405. doi: 10.3390/f15030405
- Zhang, Y., Wang, J., Yang, L., Niu, J., Huang, R., Yuan, F., et al. (2022b). Development of SSR and SNP markers for identifying opium poppy. *Int. J. Legal Med.* 136, 1261–1271. doi: 10.1007/s00414-022-02810-4
- Zhang, J., Yang, J., Fu, S., Ren, J., Zhang, X., Xia, C., et al. (2022a). Comparison of DUS testing and SNP fingerprinting for variety identification in cucumber. *Hortic. Plant J.* 8, 575–582. doi: 10.1016/j.hpi.2022.07.002
- Zhang, J., Yang, J., Lv, Y., Zhang, X., Xia, C., Zhao, H., et al. (2023a). Genetic diversity analysis and variety identification using SSR and SNP markers in melon. *BMC Plant Biol.* 23, 39. doi: 10.1186/s12870-023-04056-7
- 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, Y., Chen, W., Cui, Y., Sang, X., Lu, J., Jing, H., et al. (2021). Detection of candidate genes and development of KASP markers for Verticillium wilt resistance by combining genome-wide association study, QTL-seq and transcriptome sequencing in cotton. *Theor. Appl. Genet.* 134, 1063–1081. doi: 10.1007/s00122-020-03752-4
- Zhao, Y., Liu, X. D., Zhao, H. K., Yuan, C. P., Qi, G. X., Wang, Y. M., et al. (2017). Comparison of methods for SNP genotyping in soybean. *Mol. Plant Breed* 15, 3540–3546. doi: 10.13271/j.mpb.015.003540
- Zongze, Z., Xin, Y., Awais, A. A., Weiqiang, L., Chunqun, W., Di, W., et al. (2018). Development of a tetra-primer ARMS-PCR for detecting the E198A SNP in the isotype-1 beta-tubulin gene of *Haemonchus contortus* populations in China. *Vet. Parasitol.* 252, 127–130. doi: 10.1016/j.vetpar.2018.01.021