BMC Genomics

RESEARCH

Open Access

Development and application of genome-derived SSR markers for genetic diversity analysis, molecular fingerprinting, and core collection construction in bougainvillea (*Bougainvillea* spp.)

Rongcun Gan[1,2], Yu Han[1,2], Dandan Rao[1,2], Tengmin Li[1,2], Erhuan Wu[1,2], Yutong Cui[1] and Yu Chen[1,2*]

Abstract

Bougainvillea (*Bougainvillea* spp.) is an important ornamental shrub, yet its cultivar identification and genetic background remain poorly characterized. To address this, we developed 18 polymorphic SSR markers based on the *B. glabra* 'Formosa' genome and used them to analyze 203 *Bougainvillea* cultivars. These markers exhibited high discriminative power, successfully differentiating all accessions and revealing significant genetic heterogeneity. Population structure analysis indicated that the germplasm primarily derived from two ancestral groups, with most cultivars originating from bud sports. A high-resolution molecular fingerprinting system was established, enabling accurate cultivar identification. A 51-accession core germplasm collection was constructed using a 25% sampling strategy. This approach successfully maintained the original genetic diversity and minimized redundancy. This study provides a reliable SSR marker set and a representative core collection, offering practical tools for *Bougainvillea* cultivar identification, intellectual property protection, breeding programs, and germplasm conservation.

Keywords Bougainvillea, SSR, Population genetic structure, Core collection

Introduction

Bougainvillea (*Bougainvillea* spp.), commonly known as paperflower, is a woody climbing shrub native to South America. It has a cultivation history in China exceeding a century [1]. This genus is characterized by its high ornamental value and has become a vital landscaping species in southern China over the past three decades [2]. Intensive breeding practices, involving natural hybridization, genetic mutation, and artificial selection, have significantly enriched its germplasm diversity and resulted in complex patterns of genetic differentiation among cultivars.

Cultivar identification in *Bougainvillea* relies predominantly on morphological assessment of key ornamental traits, including plant architecture, leaf morphology, bract color, and spine characteristics [3]. However, the limitations of this approach, due to phenotypic plasticity influenced by environmental factors and developmental stage, have become increasingly

*Correspondence:
Yu Chen
cfstuchen@126.com
[1]Hainan Academy of Forestry (Hainan Academy of Mangrove), Haikou 571100, China
[2]Key Laboratory of Tropical Forestry Resources Monitoring and Application of Hainan Province, Haikou 571100, China

apparent. The shortcomings of morphological identification are driving the development of complementary molecular techniques, particularly as the number of cultivars continues to surge. Establishing an integrated identification system that combines morphological and molecular markers would significantly enhance the accuracy of cultivar discrimination, which is crucial for germplasm conservation, breeding innovation, commercial development, and intellectual property protection [4]. Furthermore, unclear genetic relationships among cultivars and an incomplete understanding of their genetic backgrounds significantly constrain breeding programs and commercial cultivation [5], highlighting the urgent need for comprehensive genetic diversity assessments.

Molecular markers provide direct access to DNA-level genetic variation, greatly facilitating plant cultivar identification and genetic diversity research [6]. Previous researches utilizing Inter-Simple Sequence Repeat (ISSR) [7], Random Amplified Polymorphic DNA (RAPD) [8], Random Amplified Polymorphic DNA (RAPD) [9] and other molecular markers [10] have established preliminary fingerprinting systems and characterized the genetic diversity patterns in *Bougainvillea*. Simple Sequence Repeats (SSRs) have become powerful tools for fingerprinting and assessing genetic diversity because of their high abundance, strong reliability, codominance, and technical simplicity. Their successful application has been demonstrated across numerous plant species [11–16].

However, existing research conclusions are primarily based on SSR primers screened from transcriptome data and small sample sizes. These findings are biased, unsuitable for large-scale samples, exhibit weak species specificity, and demonstrate low polymorphism. This hinders a full elucidation of the overall genetic architecture, variation patterns, and population structure within *Bougainvillea*, ultimately compromising the representativeness and generalizability of the conclusions. Utilizing the *Bougainvillea glabra* 'Formosa' genome [17], we screened 192 primer pairs to obtain 18 polymorphic SSR markers. When validated, these markers demonstrated 100% discriminative power among 203 *Bougainvillea* cultivars. This suite of markers provides a solid technical foundation for *Bougainvillea* cultivar identification, legal certification, and traceability systems. Furthermore, the methods for constructing a *Bougainvillea* core collection were explored using SSR marker data, providing a scientific basis for safeguarding, genetic breeding, and efficient application of *Bougainvillea* resources.

Materials and methods

Plant material

For this study, the youngest, newly expanded leaves from 203 *Bougainvillea* cultivars were collected and used as the plant material (Supplementary Table 1).

Genomic DNA extraction

Total genomic DNA was isolated from plant tissues using a magnetic bead-based DNA extraction kit (Solarbio, Beijing, China). DNA integrity and concentration were first assessed through 1% agarose gel electrophoresis (w/v) and subsequently quantified spectrophotometrically using a NanoDrop 8000 (Thermo Fisher Scientific, USA). Stock solutions were diluted to a working concentration of 20 ng/μL and stored at − 20 °C until needed for analysis.

SSR marker development, PCR amplification, and genotyping

For the construction of SSR markers, this study employed MISA 2.0 software to analyze SSR loci in the genomic sequence of *B. glabra* 'Formosa' (the genome sequence was downloaded from NCBI under accession number: GCA_045838725.1), followed by primer design for genotyping the SSR loci using Primer3 software. Initially, 192 SSR primers were obtained and screened for polymorphic information across a representative panel of 10 *Bougainvillea* accessions (B1, B4, B6, B10, B13, B25, B40, B41, B57, B73). Following PCR amplification, fragment analysis was performed using both capillary-based separation (ABI 3730xl Genetic Analyzer, Applied Biosystems, USA) and conventional agarose gel electrophoresis. Based on stringent criteria, including peak quality (≥ 3 alleles per locus), amplification reproducibility, and polymorphism level (moderate to high), 18 information-rich core SSR markers were selected for subsequent analyses (Supplementary Table 2). These 18 polymorphic SSRs were subsequently employed for genotyping a comprehensive collection of 203 *Bougainvillea* cultivars. Diluted PCR products underwent capillary electrophoresis for allele sizing. All oligonucleotide primers were commercially synthesized by Tianyi Huayu Gene Technology Co., Ltd. (Wuhan, China).

Genotypic data processing and fingerprint construction

Allele sizing data were converted into binary matrices (presence = 1, absence = 0) for fingerprint analysis. A cryptographic hash algorithm (SHA-256) was applied to generate fixed-length unique identifiers for each genotype profile. Subsequently, these digital fingerprints were encoded as Quick Response (QR) codes using Python 3.14. The QR code encoding process was implemented with three core Python libraries: pandas

(v2.1.4) for parsing tabular allele sizing data, qrcode (v7.4.2) for QR code construction, and Pillow (v10.1.0) for high-resolution image rendering.

Genetic diversity and population structure analysis

Genetic diversity parameters were calculated using GenAlEx 6.501 [18] to assess the level and distribution of genetic variation within the entire sample set. To investigate the underlying population structure, a Bayesian clustering approach was implemented in Structure 2.3.4 [19]. The analysis was configured with the following parameters: a range of hypothetical populations (K) from 1 to 20 was tested, with an initial burn-in period of 10,000 iterations followed by a Markov Chain Monte Carlo (MCMC) chain length of 100,000. For each K value, 20 independent runs were performed to ensure the robustness of the results. The optimal number of subpopulations (K) was determined by calculating the ad hoc statistic ΔK using the online tool Structure Harvester, which identifies the most likely level of population subdivision. Based on the optimal K value, the output Q matrices were processed. The final population structure visualization was generated using CLUMMP and Distruct software [20] to create a clear graphical representation of the population assignments.

Core collection development

A core collection, representing the maximum genetic diversity of the entire sample set with a minimal number of entries, was established using the R package Corehunter 3.0. The sampling strategy was optimized to maximize the Modified Roger's Distance (MR) between the selected core subset and the remaining samples, ensuring that the core set captures the broadest possible genetic spectrum. Core collections were constructed across multiple, pre-determined sampling proportion gradients (e.g., 5%, 10%, 15%, 20% of the total population). To evaluate the representativeness and genetic capture efficiency of each gradient, various population genetic parameters, including Shannon's information index, were calculated for each extracted subset and compared to the values of the entire collection.

Results

Genomic SSR analysis

Analysis of the *Bougainvillea* genome (NCBI accession number GCA_045838725.1) identified a total of 2,166,729 putative SSRs. Among these potential SSRs, dinucleotide repeats were the most abundant, comprising 1,902,764 (87.82%), followed by trinucleotide repeats (167,173; 7.72%), tetranucleotide repeats (61,081; 2.82%), and mononucleotide repeats (25,821; 1.19%) (Fig. 1A). Analysis showed that among dinucleotide and trinucleotide repeats, AT/AT was the predominant dinucleotide type (805,860), representing 42.35% of the total. The predominant trinucleotide repeat identified as AAT/ATT (32,609), constituted 19.51% of all such repeats (Fig. 1B). In terms of length distribution, SSRs of 30 bp were the most numerous (346,642, 10.14%), followed by those of 12 bp (285,433, 8.35%) and 14 bp (243,915, 7.13%) (Fig. 1C). Based on the criteria of expected product size (100–350 bp) and annealing temperature (50.0 °C to 60.0 °C), 192 pairs of putative SSR primers were randomly selected for subsequent experiments (Supplementary Table 2).

Genetic diversity of *Bougainvillea* populations at different SSR loci

A set of 192 primer pairs was developed based on the genome and used for a preliminary genetic diversity screening across 10 phylogenetically distant *Bougainvillea* cultivars (Supplementary Table 2, Table 1). The results identified 18 polymorphic SSR primers (Fig. 2), which collectively amplified 69 alleles. The number of alleles per locus ranged from 3 to 6, with a mean of 3.833. The PIC values varied from 0.295 to 0.754, averaging 0.529 (Table 1). This preliminary screening confirmed the polymorphism of these 18 SSR primers, supporting their suitability for subsequent experiments. Consequently, these 18 primers were used to assess the genetic diversity in 203 *Bougainvillea* cultivars, to derive allelic variation

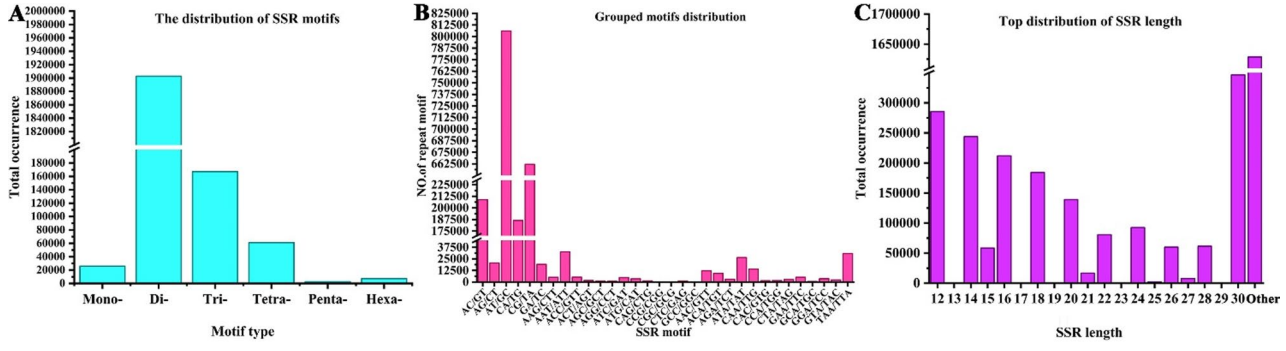

Fig. 1 SSR motif characteristics: (A) Motif frequency; (B) Motif groups; (C) SSR length

Table 1 Analysis of preliminary screening results for 18 SSR primer pairs

Locus	Na	PIC	locus	Na	PIC	locus	Na	PIC
SJMSSR015	3	0.455	SJMSSR055	4	0.686	SJMSSR132	4	0.438
SJMSSR019	6	0.715	SJMSSR068	3	0.300	SJMSSR133	3	0.457
SJMSSR026	3	0.550	SJMSSR077	3	0.450	SJMSSR162	3	0.371
SJMSSR049	3	0.295	SJMSSR085	3	0.475	SJMSSR171	3	0.418
SJMSSR052	4	0.632	SJMSSR088	4	0.489	SJMSSR181	5	0.728
SJMSSR053	5	0.473	SJMSSR091	4	0.599	SJMSSR185	6	0.754
Mean	/						3.833	0.516

Number of observed alleles (Na) and polymorphic information content (PIC)

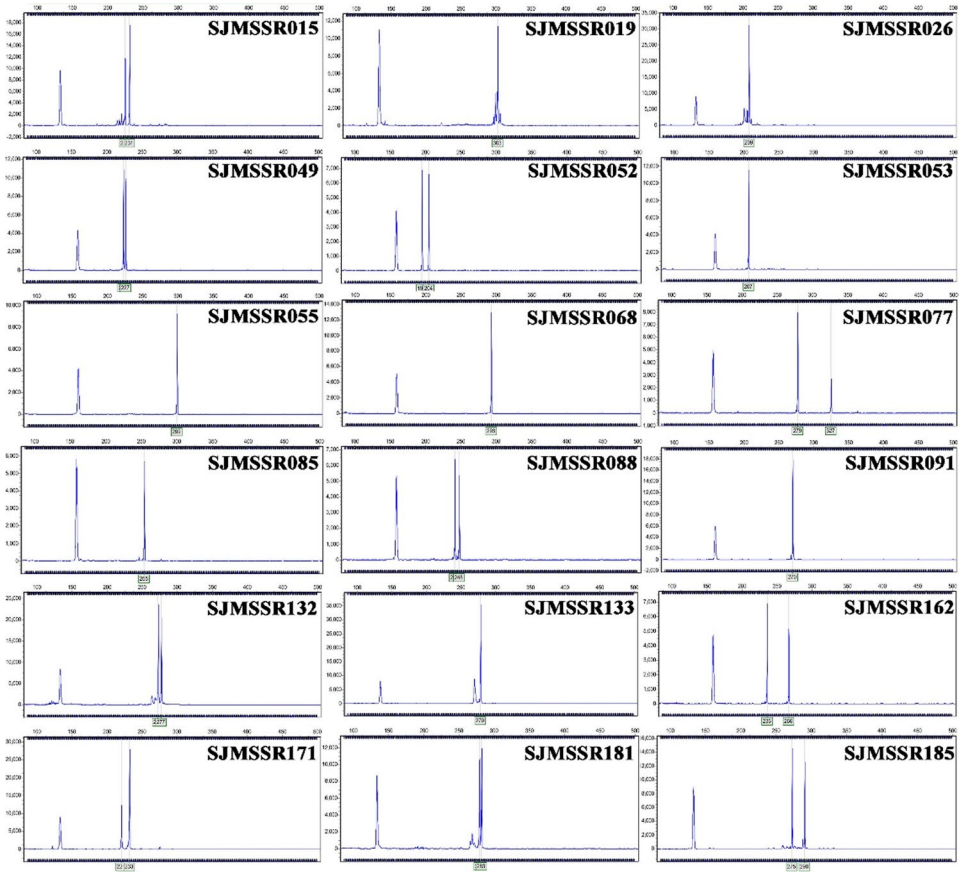

Fig. 2 Polymorphism of 18 SSR primers in *Bougainvillea* cultivar B1. The x-axis represents the size of amplified fragments, and the y-axis denotes the intensity of amplified fragments

and key diversity parameters for each marker. The differences in allelic loci and diversity parameters for each marker were determined. The SSR genotyping dataset generated and analyzed in this study has been deposited in the Zenodo repository and is accessible via the persistent link: https://doi.org/10.5281/zenodo.18491785. The 18 SSR primer pairs amplified a total of 96 Na, ranging from 3 to 11 alleles per locus, with a mean of 5.333 alleles per locus. The effective number of Ne total 48.409, varying from 1.173 to 4.597, with a mean of 2.689 per locus. Genetic diversity indices revealed I ranging from 0.315 to 1.732, with a mean of 1.117. Ho varied between 0.097 and 0.783, averaging 0.441. He ranged between 0.147 and 0.782, averaging 0.576, while PIC varied from 0.14 to 0.752, with a mean of 0.529 (Table 2).

Genetic clustering analysis

Phylogenetic relationships among 203 *Bougainvillea* accessions were assessed using an UPGMA clustering tree constructed via the Neighbor-Joining (NJ) method. The tree reveals 5 distinct groups at a genetic distance of approximately 0.36 (Fig. 3). SSR-V contained only one accession, while SSR-IV, SSR-III, and SSR-II comprised 10, 6, and 56 accessions, respectively. SSR-I was the largest, consisting of 130 accessions. The

Table 2 Analysis of polymorphic bands of 18 pairs of SSR primers

Locus	Na	Ne	I	Ho	He	PIC
SJMSSR015	3	2.387	0.942	0.650	0.581	0.491
SJMSSR019	11	4.597	1.732	0.565	0.782	0.750
SJMSSR026	8	3.375	1.375	0.370	0.704	0.651
SJMSSR049	3	1.173	0.315	0.118	0.147	0.140
SJMSSR052	6	3.197	1.372	0.33	0.687	0.647
SJMSSR053	7	3.228	1.476	0.522	0.69	0.657
SJMSSR055	3	2.487	1.000	0.097	0.598	0.530
SJMSSR068	3	2.085	0.892	0.429	0.520	0.464
SJMSSR077	4	1.429	0.607	0.143	0.300	0.280
SJMSSR085	6	2.962	1.312	0.443	0.662	0.610
SJMSSR088	6	2.408	1.07	0.401	0.585	0.510
SJMSSR091	5	3.125	1.286	0.783	0.68	0.627
SJMSSR132	6	2.658	1.208	0.66	0.624	0.582
SJMSSR133	3	1.518	0.628	0.399	0.341	0.309
SJMSSR162	6	2.064	1.061	0.438	0.516	0.485
SJMSSR171	3	1.88	0.744	0.606	0.468	0.382
SJMSSR181	6	3.252	1.381	0.591	0.693	0.653
SJMSSR185	7	4.584	1.700	0.389	0.782	0.752
Mean	5.333	2.689	1.117	0.441	0.576	0.529

Effective number of alleles (Ne), observed heterozygosity (Ho), expected heterozygosity (He), and Shannon's diversity index (I)

pairwise genetic distances among all accessions ranged from 0.02 to 0.89, with a mean value of 0.46. Analysis of genetic distances within and between groups indicated closer relationships within SSR-I and SSR-II, whereas relatively distant relationships were observed between groups. SSR-V, SSR-IV, and SSR-III exhibit relatively distant genetic relationships with SSR-I and SSR-II.

Population genetic structure analysis

Population genetic structure was determined with a Bayesian model implemented in STRUCTURE to assess genetic backgrounds and potential gene flow. Genotypic data from 18 polymorphic SSR loci across 203 *Bougainvillea* accessions were analyzed. The optimal number of genetic clusters (K) was determined by evaluating the ΔK statistic, which decreased monotonically with increasing K values (Fig. 4). Maximum likelihood analysis identified K = 2 as the most biologically meaningful grouping. Bayesian clustering at K = 2 partitioned the germplasm into two distinct genetic groups: P-I ($n = 71$) and P-II ($n = 132$) (Fig. 5). Among populations defined by SSR analysis, the SSR-I group consisted predominantly of accessions with P-II ancestry, with only minor admixture from P-I. Conversely, the SSR-II group was primarily composed of P-I origin with limited P-II introgression. In contrast, SSR-III, SSR-IV, and SSR-V groups exhibited substantially more pronounced genetic admixture compared to SSR-I and SSR-II.

These genetic composition patterns were further supported by Q-values from STRUCTURE analysis. Each accession was assigned a Q-value based on the inferred K value, where $Q \geq 0.600$ indicates a relatively homogeneous genetic background, and lower values suggest admixed origins. Higher Q-values reflect greater probability of belonging to the corresponding subpopulation. Within P-I, 66 accessions (92.96%) showed $Q \geq 0.600$, while in P-II, 128 accessions (96.97%) exhibited $Q \geq 0.600$ (Fig. 6A). These results demonstrate that most accessions in both clusters maintain relatively pure genetic backgrounds with only minimal gene flow from other genetic groups. PCoA was performed using the sample distance matrix based on the population genetic structure grouping. Each point represents an individual accession, with closer proximity indicating greater genetic similarity. The plot revealed two major clusters along the first two principal coordinates (P-I and P-II), despite some partial overlap between clusters (Fig. 6B).

Molecular fingerprinting and genetic identification in *Bougainvillea*

In this study, PCR amplification was performed on 203 *Bougainvillea* accessions using 18 pairs of core SSR primers. Through systematic comparison of amplification bands, high-resolution molecular fingerprint profiles were successfully constructed for 203 cultivars (Supplementary Table 3, Fig. 7, The supplementary Fig. 1).

Establishment of a core germplasm collection

Core germplasms were sampled at seven proportions: 10%, 15%, 20%, 25%, 30%, 35%, and 40%. The evaluation parameters of the core germplasms and the

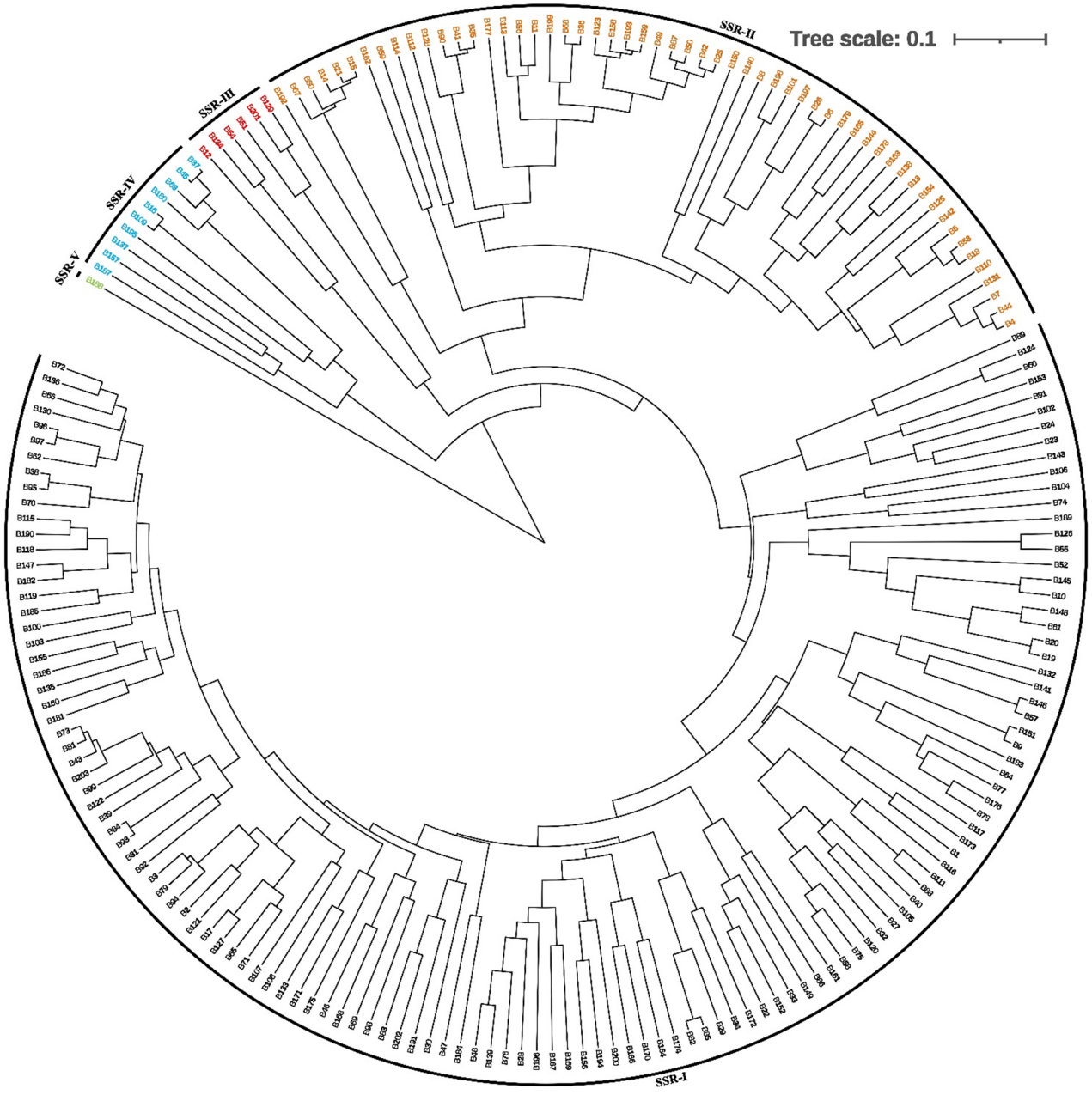

Fig. 3 Phylogenetic tree from SSR data clustering

original germplasms are presented in Table 3. The results revealed that as the sampling proportion increased, both the observed number of alleles and the Shannon's index exhibited an upward trend, with the number of alleles ranging from 4.833 to 5.167 and the Shannon's index ranging from 1.208 to 1.231. Shannon's information index achieved its highest value (1.231) at a 25% sampling proportion. This suggests the proportion successfully captured gene loci from the initial gene pool, and maintained considerable within-species variation. Therefore, 25% was selected as the optimal sampling proportion, ultimately leading to the construction of a primary core germplasm comprising 51 *Bougainvillea* germplasm resources. Analysis of the core and initial germplasms is shown in Table 4. The core collection retained 25.12% of the original materials, with the following retention rates: 96.89% for alleles, 114.91% for effective alleles, 109.03% for Nei's gene diversity, 110.21% for Shannon's index, and 109.83% for polymorphism information content. The core germplasm exceeded both the original and retained collections in effective allele number, Nei's gene diversity, Shannon's index, and PIC. This confirms that the core collection is a highly representative subset. Based on the genetic diversity data obtained using SSR molecular

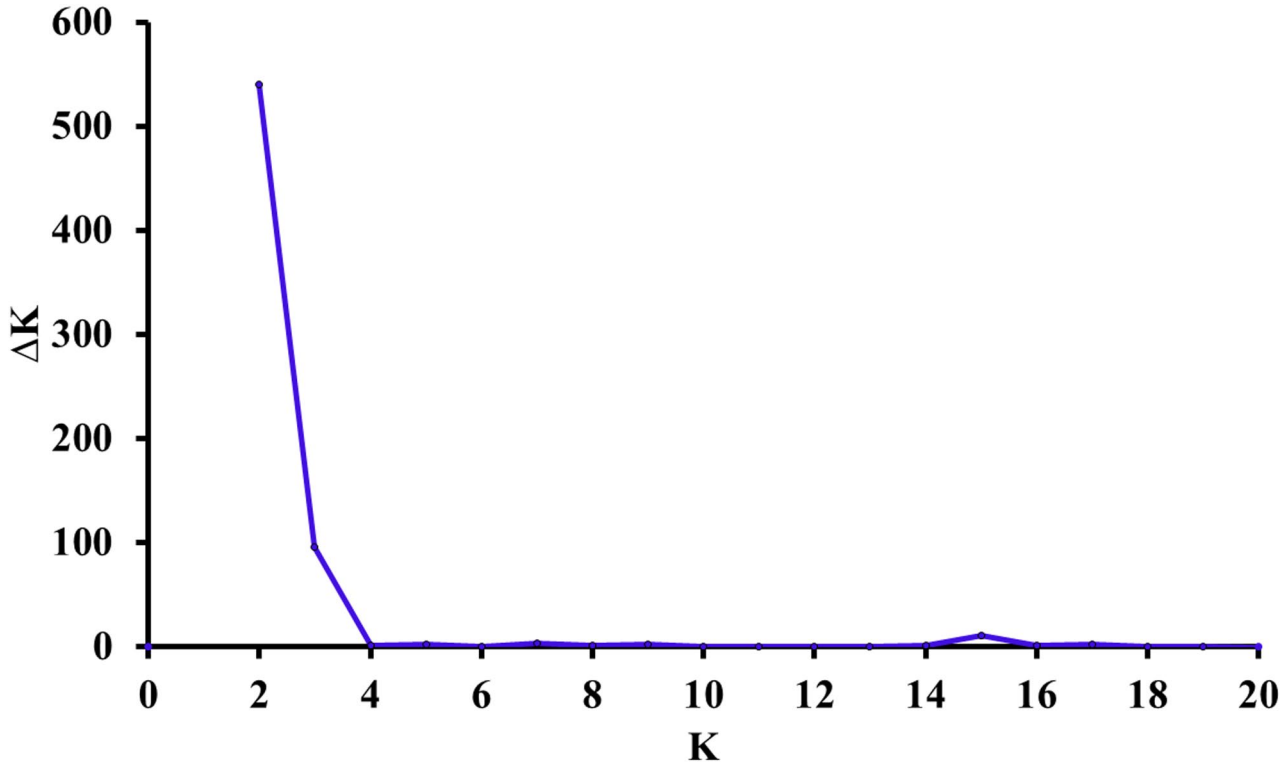

Fig. 4 Determination of the optimal K value for *Bougainvillea* cultivars

markers, a primary core germplasm of 51 *Bougainvillea* germplasms was constructed, with germplasm numbers B3, B7, B8, B9, B10, B12, B14, B16, B29, B42, B47, B52, B59, B63, B69, B74, B88, B89, B91, B102, B104, B105, B106, B107, B112, B114, B124, B125, B126, B128, B134, B137, B138, B143, B146, B150, B156, B161, B164, B165, B172, B173, B177, B188, B189, B191, B192, B195, B196, B197, and B201. Further principal component analysis (PCA) was conducted on the SSR molecular data of the 51 core germplasms and the 203 original *Bougainvillea* germplasms. Figure 8 presents the scatter plot of the principal component scores for the two germplasm collections. It can be observed that the scatter distribution regions and shapes of the core germplasm and the original germplasm exhibited a high degree of overlap. Peripheral individuals were also selected into the core germplasm repository, while only a portion of the lines with high genetic similarity were included, eliminating genetic redundancy and ensuring the representativeness and heterogeneity of the core germplasm. In summary, the finally constructed core germplasm effectively represents the original population, exhibits high heterogeneity, and to a certain extent, eliminates genetic redundancy. The establishment of the *Bougainvillea* core germplasm provides a foundation and guarantee for resource preservation, evaluation, the excavation of specific resources, and efficient utilization. It should be given full attention in future germplasm resource conservation, evaluation, and innovative utilization efforts.

Discussion

The effectiveness of SSR markers in assessing genetic diversity and establishing fingerprint databases has been well established across numerous plant taxa [21–23]. Genetic analysis of 203 *Bougainvillea* cultivars using 18 SSR markers confirmed substantial genetic heterogeneity within the population, consistent with previous findings obtained through RAPD [10] and SRAP [24] markers. Consistent results from diverse analytical approaches confirm the robustness and efficacy of the employed markers for profiling intraspecific variation in *Bougainvillea*. In summary, these 18 SSR markers demonstrated high polymorphism and effectiveness, making them suitable for constructing DNA fingerprints, conducting genetic diversity studies, and performing subsequent analyses. As shown in Tables 1 and 2, Na and PIC values of the 18 SSR primer pairs exhibited considerable variation with the expansion of the sample size. Specifically, the Na value of SJMSSR019 increased from 6 to 11, that of SJMSSR026 rose from 3 to 8, and those of SJMSSR085 and SJMSSR162 increased from 3 to 6. Concurrently, notable shifts were observed in PIC values at multiple loci: for instance, SJMSSR049 decreased from 0.295 to 0.140, whereas SJMSSR053 increased from 0.473 to 0.657. These discrepancies indicate that parameter

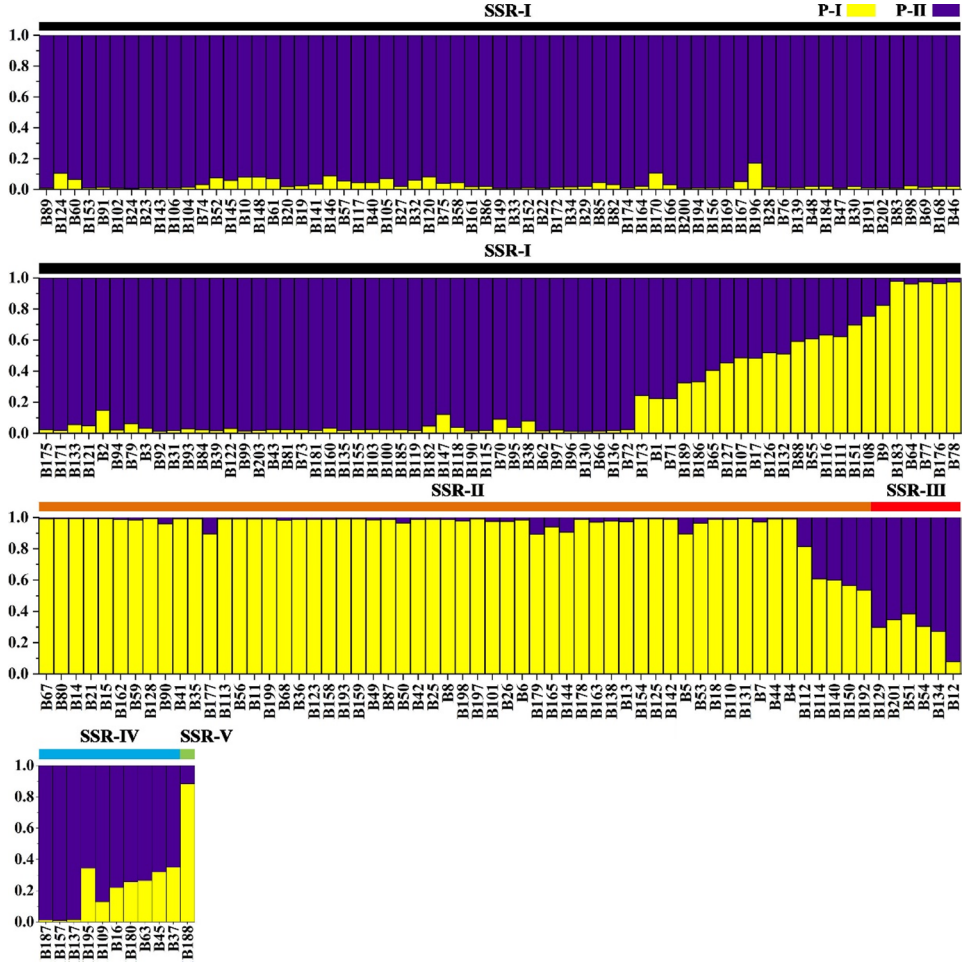

Fig. 5 Analysis of ancestral populations in 203 *Bougainvillea* varieties was performed with SSR data using STRUCTURE v2.3.4

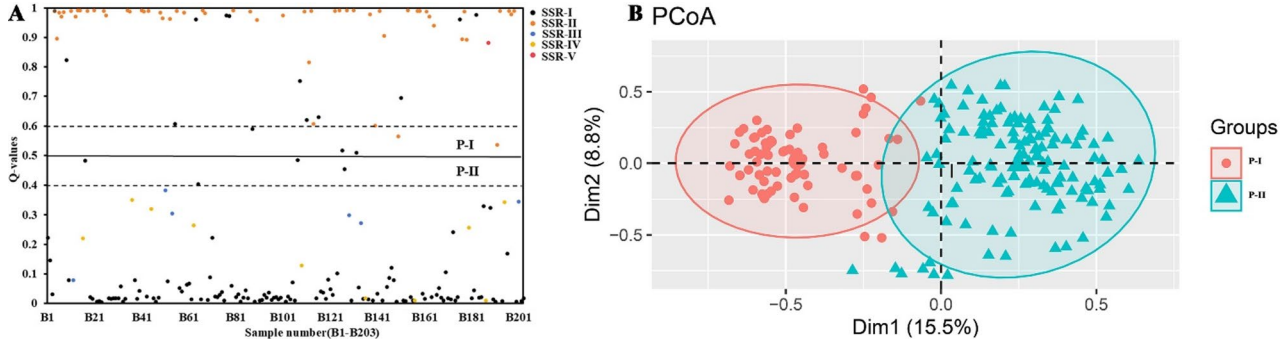

Fig. 6 Population genetic structure of 203 *Bougainvillea* varieties based on SSR markers **A** Q-value matrix; **B** PCoA analysis. Principal coordinate analysis (PCoA)

estimates derived from limited sample sizes are subject to substantial bias, while enlarging the sample size improves data accuracy [25]. Accordingly, an SSR system established based on a large sample set demonstrates higher reliability.

These results of the phylogenetic tree demonstrate that the phylogeny constructed using novel SSR markers accurately reflects the genetic relationships among the 203 accessions. However, these results align with the conclusions of Kumar et al. [26], who classified 50 *Bougainvillea* cultivars into 5 subgroups using 5 pairs of core SSR primers. This result can possibly be attributed to the fact that *Bougainvillea* from both regions were subjected to similar breeding practices, which resulted in analogous clustering patterns. The optimal number of genetic clusters (K) identified in our study differs from the findings

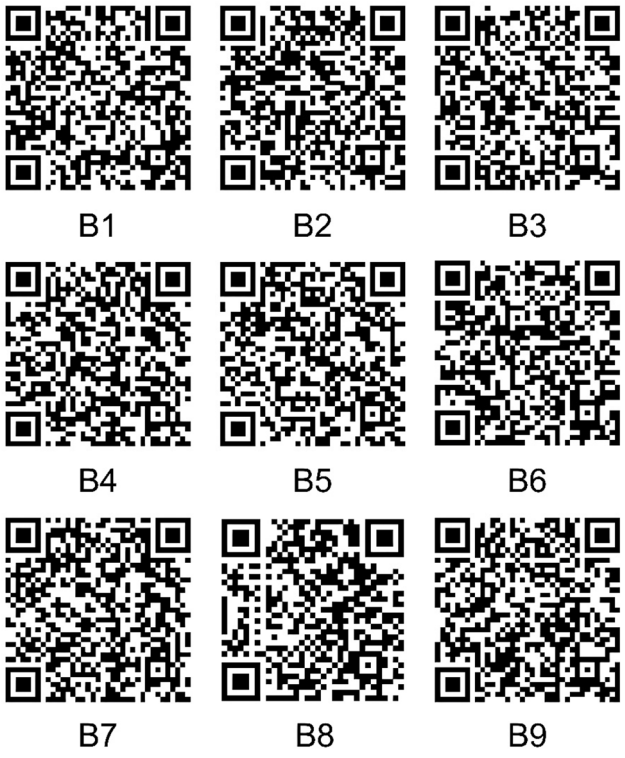

Fig. 7 Fingerprints of B1-B9

reported by Kumar et al. [9]. While their analysis of 125 *Bougainvillea* cultivars indicated an optimal K-value of 7, our study determined a K-value of 2. This discrepancy may be attributed to the following reasons. First, the domestic cultivars examined in our study exhibit relatively homogeneous genetic backgrounds with limited ancestral origins, whereas the germplasm analyzed by Kumar et al. [9] likely encompasses broader genetic diversity resulting from the introduction of exotic species and hybridization among geographically distinct populations [27]. This result of PCoA analysis corroborates the population structure analysis, demonstrating consistent genetic relationships among accessions. Moreover, the clustering pattern derived from PCA allows for better characterization of population differentiation and refines the assignment of genotypes to their populations of origin [28].

In this study, genetic analysis of 203 *Bougainvillea* germplasms showed that the genetic distances among the materials ranged from 0.02 to 0.89, with an average value of 0.46. While the average genetic distance was relatively low, a small number of cultivars exhibited notably high genetic distances. Overall, the genetic distance results were similar to those reported by Kumar et al. [9] (0.09–0.86). This indicates considerable overall genetic diversity within the analyzed materials, a pattern potentially attributable to the close genetic relationships among the cultivars. The observed discrepancies between studies primarily stem from the fact that most *Bougainvillea* cultivars originate from natural bud sports, with only a limited number derived from artificial breeding [29]. Consequently, distantly related individuals predominantly belong to artificially bred lineages, whereas genetically similar natural bud sports collectively form a large, dominant population. Both population genetic analysis and PCoA collectively confirm that the existing *Bougainvillea* germplasm resources are primarily derived from two ancestral groups. The vast majority of cultivars have been developed through bud sports selected from these two ancestral groups, resulting in relatively homogeneous genetic backgrounds. In contrast, only a very small number of accessions were developed through artificial hybridization between different bud sport lines. These hybrid-derived individuals not only exhibited clear genetic admixture but also maintained

Table 3 Assessment of genetic parameter differences in core versus original collections across sampling proportions

Sampling proportion	Sample number	Na	Ne	He	I	PIC
10%	20	4.833	3.118	0.630	1.223	0.583
15%	30	4.944	3.118	0.625	1.223	0.579
20%	41	5.056	3.097	0.625	1.223	0.578
25%	51	5.167	3.090	0.628	1.231	0.581
30%	61	5.167	3.033	0.622	1.220	0.576
35%	71	5.222	2.978	0.617	1.207	0.570
40%	81	5.167	2.945	0.609	1.195	0.564

Table 4 Evaluation of genetic parameters in the core germplasm relative to the original collection

Category	Germplasm amount	Na	Ne	He	I	PIC
Original germplasm	203	5.333	2.689	0.576	1.117	0.529
Core germplasm	51	5.167	3.090	0.628	1.231	0.581
Core germplasm retention rate/%	25.12	96.89	114.91	109.03	110.21	109.83
Reserve germplasm	152	5.278	2.744	0.581	1.134	0.536
Reserve germplasm retention rate/%	74.88%	98.97	102.05	100.87	101.52	101.32

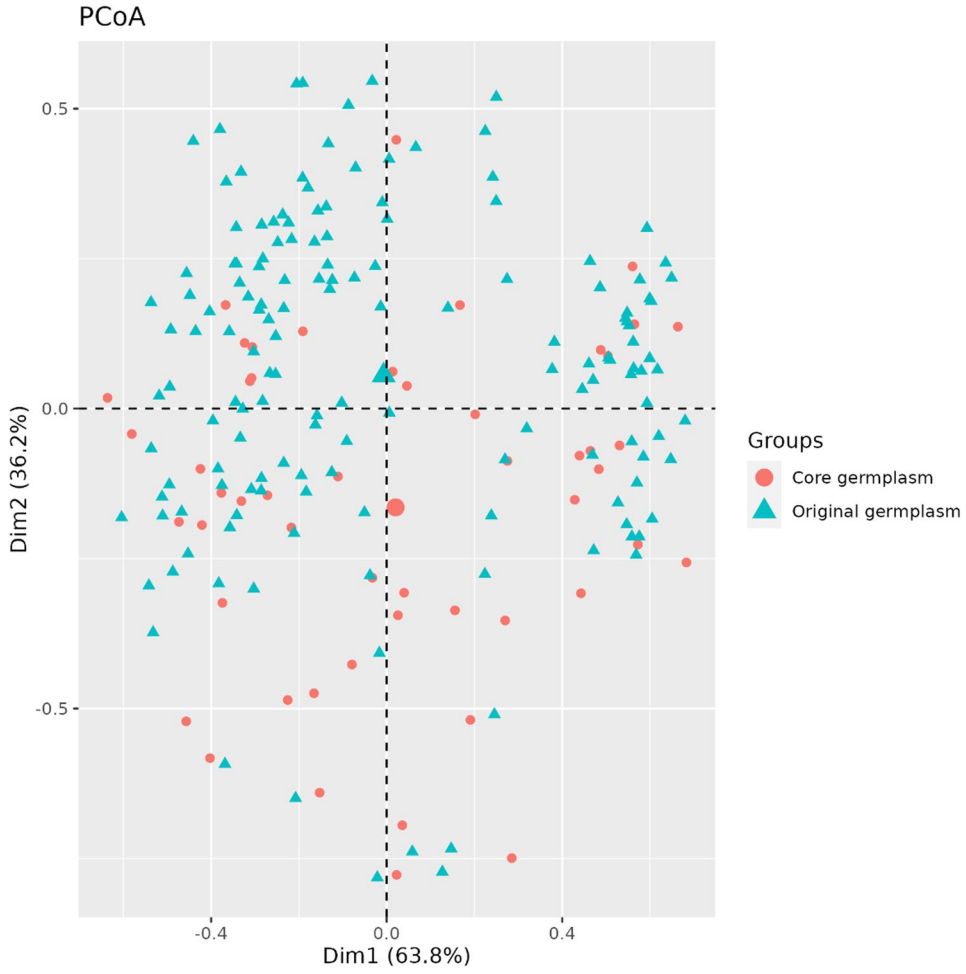

Fig. 8 Scatter plot of principal component scores for core germplasm and original germplasm

relatively large genetic distances from the principal germplasm population.

The fingerprint profiles constructed in this experiment exhibit high specificity and stability, with each cultivar displaying a unique combination of characteristic bands, thereby providing a reliable molecular basis for cultivar identification. As a pivotal biotechnology tool, plant genetic fingerprinting is closely linked to economic value, quality traits, and intellectual property protection, serving as a critical bridge between genotypes and phenotypic traits. By integrating cryptographic hashing and QR code technology, this study transforms conventional laboratory data into portable, scannable digital identifiers. These QR codes enable rapid on-site verification of plant identity without reliance on centralized databases, offering significant potential for large-scale germplasm management, variety rights protection, and traceability systems. The scalability of this approach supports integration into existing agricultural certification frameworks and blockchain-based traceability platforms, facilitating end-to-end connectivity from breeding programs to commercial markets. To date, SSR-based genetic fingerprints have been successfully established in diverse plant species, including *Ipomoea batatas* [30], *Camellia sinensis* [31], *Trifolium repens* [32], *Brassica oleracea* [33], and *Ailanthus altissima* [34]. Future work will focus on developing user-friendly mobile applications for QR code decoding, expanding the database to cover additional cultivars, and exploring integration with smart farming systems to enhance precision agriculture and supply chain transparency.

A priority in managing plant genetic resources is to develop a minimal core library that captures the essential diversity of the entire collection. While core collections can be constructed using phenotypic traits to capture the diversity of the whole collection, such metric data are susceptible to environmental influence. Variations in climate and soil conditions may distort research findings, as noted in [35]. Molecular marker techniques allow for rapid and accurate evaluation of genetic affinities among accessions, thereby drastically reducing the timeframe needed to establish a core collection [36]. When

constructing a core germplasm collection using phenotypic data, key variables including distance and proportion, sampling methods, and clustering approaches require comprehensive consideration [37], as the chosen criteria determine the composition of the collection. The integration of techniques like SSR, SNP, and resequencing enables the development of more precise core collections that capture the diversity present in the source population [38, 39]. A sampling proportion between 10% and 30% is usually adopted when establishing core collections for horticultural crops [40–46]. The sampling proportion in this study was 25%, which is analogous to those used for core collections of Japanese persimmon and litchi [39, 47]. For further research on the core collection, it is essential to enrich multiple aspects such as key phenotypic traits, safeguard against the loss of germplasm with important genetic characteristics, and incorporate a wider range of variation into the core collection, which holds significant importance for comprehensively establishing a core collection of *Bougainvillea*.

Compared with previous studies, the present work covers a broader range of germplasm with improved representativeness. It not only significantly expands the coverage of molecular identification in *Bougainvillea* cultivars but also demonstrates considerable innovation and application potential in resolving genetic diversity, protecting varietal rights, and enhancing breeding adaptability. Furthermore, this research clarified the systematic genetic relationships among 203 *Bougainvillea* cultivars and established a comprehensive molecular identification system for *Bougainvillea*. The findings provide a solid theoretical foundation and technical support for authenticity verification, accurate identification, intellectual property protection, and full-chain traceability management in *Bougainvillea* cultivation. Simultaneously, the *Bougainvillea* core collection forms the basis for the preservation and evaluation of resources, the mining of unique germplasm, and their efficient utilization. It should therefore be given due emphasis in future work concerning germplasm resource conservation, evaluation, and innovative exploitation.

Conclusion

This study successfully developed 18 highly polymorphic and specific core SSR markers based on the *Bougainvillea* genome. Analysis of 203 *Bougainvillea* cultivars using this marker system revealed significant genetic heterogeneity and a population structure primarily derived from two ancestral groups, with most cultivars originating from bud sports and exhibiting relatively homogeneous backgrounds. A comprehensive molecular fingerprinting and identification system was established, enabling 100% discrimination among all tested cultivars. An optimal sampling proportion of 25% was identified. This resulted in a primary core germplasm of 51 accessions that successfully represents the original genetic variation with minimal redundancy. The developed SSR marker set and the established core collection provide a robust technical foundation and valuable genetic resources for cultivar identification, intellectual property protection, breeding innovation, and the conservation and sustainable utilization of *Bougainvillea* germplasm.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12804-w.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Acknowledgements

Not applicable.

Authors' contributions

Y.C., R.G., and Y.H., designed the research. R.G., Y.H., D.R., T.L., E.W., and Y.C., performed the research. All authors analyzed and interpreted the data. R.G., wrote the manuscript. All authors commented on the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the Technical Innovation Project of Hainan Provincial Scientific Research Institutes (KYYSLK2024-002).

Data availability

The genome sequence of *Bougainvillea glabra* 'Formosa' used in this study is available from NCBI under the accession number GCA_045838725.1. The SSR genotyping dataset generated and analyzed in this study has been deposited in the Zenodo repository and is accessible via the persistent link: https://doi.org/10.5281/zenodo.18491785.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 23 January 2026 / Accepted: 26 March 2026
Published online: 07 April 2026

References

1. Zhang D, Xue Y, Feng N, Bai J, Ma D, Sheng Q, et al. Physiological responses and salt tolerance evaluation of different varieties of *Bougainvillea* under salt stress. Plants-Basel. 2024;13:2409. https://doi.org/10.3390/plants13172409.
2. Liang Y, Li D, Sheng Q, Zhu Z. Exogenous salicylic acid alleviates NO_2 damage by maintaining cell stability and physiological metabolism in *Bougainvillea*×buttiana 'Miss Manila' seedlings. Plants-Basel. 2023. https://doi.org/10.3390/plants12183283. 12;3283.

3. Bautista MAC, Zheng Y, Boufford DE, Hu Z, Deng Y, Chen T. Phylogeny and taxonomic synopsis of the genus *Bougainvillea* (Nyctaginaceae). Plants-Basel. 2022;11:1700. https://doi.org/10.3390/plants11131700.
4. Cao Jie, Zhou ZC, Tu JF, Yao JL, Xu F, Wang GY, et al. Genetic diversity and population structure analysis of sand pear (Pyrus pyrifolia) 'Nakai' varieties using SSR and AFLP markers. Not Bot Horti Agrobo. 2019;47:970–9. https://doi.org/10.15835/nbha47311570.
5. Zheng JJ, Su HQ, Pu SS, Chen H, El-Kassaby YA, Yang ZJ, et al. High-yield hybrid breeding of Camellia oleifolia based on ISSR molecular markers. BMC Plant Biol. 2024;24:517. https://doi.org/10.1186/s12870-024-05218-x.
6. Grover A, Sharma PC. Development and use of molecular markers: past and present. Crit Rev Biotechnol. 2016;36:290–302. https://doi.org/10.3109/07388551.2014.959891.
7. Zhang H, Huang T, Zhou Q, Sheng Q, Zhu Z. Complete chloroplast genomes and phylogenetic relationships of *Bougainvill*ea spectabilis and *Bougainvillea* glabra (Nyctaginaceae). Int J Mol Sci. 2023;24:13044. https://doi.org/10.3390/ijms241713044.
8. Ma X, Wang H, Hu Y, Hu Q, Qian R, Zheng J. ddRAD-seq data reveals the significant population structure and potential causes of population divergence in *Bougainvillea*. Sci Rep. 2025;15:34200. https://doi.org/10.1038/s41598-025-15668-4.
9. Kumar PP, Janakiram T, Bhat KV. Microsatellite based DNA fingerprinting and assessment of genetic diversity in bougainvillea cultivars. Gene. 2020;753:144794. https://doi.org/10.1016/j.gene.2020.144794.
10. Rastogi RR, Singh N, Singh S, Pal AK, Roy RK, Rana TS. Assessment of genetic variability in the *Bougainvillea* varieties using morphological and molecular markers. Indian J Exp Biol. 2019;57:408–17.
11. Hu G, Jiang Q, Wang Z, Li Z, Liao W, Shen D, et al. Genetic diversity analysis and core collection construction of the Actinidia chinensis complex (Kiwifruit) based on SSR markers. Agron. 2022;12:3078. https://doi.org/10.3390/agronomy12123078.
12. Zhao Z, Zhang H, Wang P, Yang Y, Sun H, Li J. Development of SSR molecular markers and genetic diversity analysis of Clematis acerifolia from Taihang Mountains. PLoS ONE. 2023;18:e0285754. https://doi.org/10.1371/journal.pone.0285754.
13. Zhu Y, Ma T, Lin Y, Peng Y, Huang Y, Jiang J. SSR molecular marker developments and genetic diversity analysis of Zanthoxylum nitidum. (Roxb) DC Sci Rep. 2023;13:20767. https://doi.org/10.1038/s41598-023-48022-7.
14. Awan FS, Maryam, Jaskani MJ, Sadia B. Gender identification in date palm using molecular markers. Methods Mol Biol. 2017;1638:209–25. https://doi.org/10.1007/978-1-4939-7159-6-18.
15. Neiva DS, Melo Júnior AF, Oliveira DA, Royo VA, Brandão MM, Menezes EV. Acrocomia emensis (Arecaceae) genetic structure and diversity using SSR molecular markers. Genet Mol Res. 2016;15:10. https://doi.org/10.4238/gmr.15017785.
16. Yan H, Qi H, Li Y, Wu Y, Wang Y, Chen J. Assessment of the genetic relationship and population structure in oil-tea Camellia species using simple sequence repeat (SSR) markers. Genes. 2022;13:2162. https://doi.org/10.3390/genes13112162.
17. Lan L, Li H, Xu S, Xu Y, Leng Q, Zhang L, et al. Chromosome-level and haplotype-resolved genome assembly of *Bougainvillea glabra*. Sci Data. 2025;12:107. https://doi.org/10.1038/s41597-024-04333-6.
18. Zhang Q, Jin Z, Hu J, Zhang X, He J, Feng Y, et al. Assessment of genetic diversity in Korean field mouse (Apodemus peninsulae): A study based on microsatellite molecular markers. Mol Biol Rep. 2025;52:234. https://doi.org/10.1007/s11033-025-10328-2.
19. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14:2611–20. https://doi.org/10.1111/j.1365-294X.2005.02553.x.
20. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genet. 2000;155:945–59. https://doi.org/10.1093/genetics/155.2.945.
21. Rabeh K, Mghazli N, Gaboun F, Filali-Maltouf A, Sbabou L, Belkadi B. Genome-wide identification of SSR markers from coding regions for endangered Argania spinosa L. skeels and construction of SSR database: AsSSRdb. Database. 2024;2024:baae118. https://doi.org/10.1093/database/baae118.
22. Li X, Qiao L, Chen B, Zheng Y, Zhi C, Zhang S, et al. SSR markers development and their application in genetic diversity evaluation of garlic (Allium sativum) germplasm. Plant Divers. 2021;44:481–91. https://doi.org/10.1016/j.pld.2021.08.001.
23. Ha YJ, Sa KJ, Lee JK. Identifying SSR markers associated with seed characteristics in perilla (Perilla frutescens L). Physiol Mol Biol Plants. 2021;27:93–105. https://doi.org/10.1007/s12298-021-00933-3.
24. Kumar M, Chaudhary V, Sirohi U, Sharma VR, Naresh RK. Application of molecular markers and their utility in genetic studies of floricultural crops: A review. Int J Agric Environ Biotechnol. 2019;12:229–47. https://doi.org/10.30954/0974-1712.08.2019.7.
25. Stockwell DRB, Peterson AT. Effects of sample size on accuracy of species distribution models. Ecol model. 2002;148:1–13. https://doi.org/10.1016/S0304-3800(01)00388-X.
26. Kumar PP, Janakiram T, Bhat KV, Jain R, Prasad KV, Prabhu KV. Molecular characterization and cultivar identification in *Bougainvillea spp.* using SSR markers. Indian J Agric Sci. 2014;84:1024–30. https://doi.org/10.56093/ijas.v84i9.43474.
27. An M, Deng M, Zheng SS, Jiang XL, Song YG. Introgression threatens the genetic diversity of Quercus austrocochinchinensis (Fagaceae), an endangered oak: a case inferred by molecular markers. Front Plant Sci. 2017;8:229. https://doi.org/10.3389/fpls.2017.00229.
28. van Heerwaarden J, Odong TL, van Eeuwijk FA. Maximizing genetic differentiation in core collections by PCA-based clustering of molecular marker data. Theor Appl Genet. 2013;126:763–72. https://doi.org/10.1007/s00122-012-2016-2.
29. Salam P, Bhargav V, Gupta YC, Nimbolkar PK. Evolution in bougainvillea (*Bougainvillea* Commers.)-a review. J Appl Nat Sci. 2017;9:1489. https://doi.org/10.31018/jans.v9i3.1389.
30. Zhao N, Li H, Zhai H, HE S, Liu Q. SSR fingerprinting of 203 sweetpotato (*Ipomoea batatas* (L.) Lam.) varieties. J Integr Agr. 2018;17:86–93. https://doi.org/10.1016/S2095-3119(17)61687-3.
31. Liu S, Liu H, Wu A, Hou Y, An Y, Wei C. Construction of fingerprinting for tea plant (*Camellia sinensis*) accessions using new genomic SSR markers. Mol Breed. 2017;37:93. https://doi.org/10.1007/s11032-017-0692-y.
32. Ma S, Han C, Zhou J, Hu R, Jiang X, Wu F, et al. Fingerprint identification of white clover cultivars based on SSR molecular markers. Mol Biol Rep. 2020;47:8513–21. https://doi.org/10.1007/s11033-020-05893-7.
33. Xie Q, Zhao Y, Liu Y, Han F, Liu W, Li Z. Genetic diversity and DNA fingerprinting in broccoli carrying multiple clubroot resistance genes based on SSR markers. Appl Sci. 2022;12:4754. https://doi.org/10.3390/app12094754.
34. Zhang M, Zheng C, Li J, Wang X, Liu C, Li X, et al. Genetic diversity, population structure, and DNA fingerprinting of *Ailanthus altissima* var. erythrocarpa based on EST-SSR markers. Sci Rep. 2023;13:19315. https://doi.org/10.1038/s41598-023-46798-2.
35. Zhu H, Song P, Koo DH, Guo L, Li Y, Sun S, et al. Genome wide characterization of simple sequence repeats in watermelon genome and their application in comparative mapping and genetic diversity analysis. BMC Genomics. 2016;17:557. https://doi.org/10.1186/s12864-016-2870-4.
36. Fang H, Liu H, Ma R, Liu Y, Li J, Yu X, et al. Genome wide assessment of population structure and genetic diversity of Chinese Lou onion using specific length amplified fragment (SLAF) sequencing. PLoS ONE. 2020;15:e0231753. https://doi.org/10.1371/journal.pone.0231753.
37. Mahmoodi R, Dadpour MR, Hassani D, Zeinalabedini M, Vendramin E, Micali S, et al. Development of a core collection in Iranian walnut (Juglans regia L.) germplasm using the phenotypic diversity. Sci Hortic. 2019;249:439–48. https://doi.org/10.1016/j.scienta.2019.02.017.
38. Hoshikawa K, Lin Y, Schafleitner R, Shirasawa K, Isobe S, Nguyen DC, Ohsawa R, et al. Genetic diversity analysis and core collection construction for Amaranthus tricolor germplasm based on genome-wide single-nucleotide polymorphisms. Sci Hortic. 2023;307:111428. https://doi.org/10.1016/j.scienta.2022.111428.
39. Sun Q, Bai L, Ke L, Xiang X, Zhao J, Ou L. Developing a core collection of litchi (Litchi chinensis Sonn.) based on EST-SSR genotype data and agronomic traits. Sci Hortic. 2012;146:29–38. https://doi.org/10.1016/j.scienta.2012.08.012.
40. Colmsee C, Keller ERJ, Zanke C, Senula A, Funke T, Oppermann M, et al. The Garlic and Shallot Core Collection image database of IPK presenting two vegetatively maintained crops in the Federal ex situ genebank for agricultural and horticultural crops at Gatersleben, Germany. Genet Resour Crop Ev. 2012;59(7):1407–15. https://doi.org/10.1007/s10722-011-9768-4.
41. da Cunha CP, Resende FV, Zucchi MI, Pinheiro JB. SSR-based genetic diversity and structure of garlic accessions from Brazil. Genetica. 2014;142(5):419–31. https://doi.org/10.1007/s10709-014-9786-1.
42. Gopal J, Kumar V, Kumar R, Mathur P. Comparison of different approaches to establish a core collection of Andigena (Solanum tuberosum Group

Andigena) potatoes. Potato Res. 2013;56(1):85–98. https://doi.org/10.1007/s11540-013-9232-2.
43. Hu J, Wang P, Su Y, Wang R, Li Q, Sun K. Microsatellite diversity, population structure, and core collection formation in Melon Germplasm. Plant Mol Biol Rep. 2015;33(3):439–47. https://doi.org/10.1007/s11105-014-0757-6.
44. Liu J, Wang Z, Li C, Bian Y, Xiao Y. Evaluating genetic diversity and constructing core collections of Chinese Lentinula edodes cultivars using ISSR and SRAP markers. J Basic Microb. 2015;55(6):749–60. https://doi.org/10.1002/jobm.201400774.
45. Taniguchi F, Kimura K, Saba T, Ogino A, Yamaguchi S, Tanaka J. Worldwide core collections of tea (*Camellia sinensis*) based on SSR markers. Tree Genet Genomes. 2014;10(6):1555–65. https://doi.org/10.1007/s11295-014-0779-0.
46. Van Raamsdonk LWD, Wijnker J. The development of a new approach for establishing a core collection using multivariate analyses with tulip as case. Genet Resour Crop Ev. 2000;47(4):403–16. https://doi.org/10.1023/A:1008777422034.
47. Zhang YF, Zhang QL, Yang Y, Luo ZR. Development of Japanese persimmon core collection by genetic distance sampling based on SSR markers. Biotechnol Biotec Eq. 2009;23(4):1474–8. https://doi.org/10.2478/V10133-009-0014-6.

Publisher's note