Agricultural Sciences, 2024, 15, 256-273
https://www.scirp.org/journal/as
ISSN Online: 2156-8561
ISSN Print: 2156-8553

Analysis on Genetic Diversity of 40 Flowering Cherry Cultivars and Construction of Molecular ID Based on SSR Markers

Chaoren Nie[1], Xiaoguo Xu[2], Xiaoqin Zhang[1], Hongbing Sun[1], Jingya Yu[1], Wensheng Xia[1*]

[1]Wuhan Institute of Landscape Architecture, Wuhan, China
[2]Wuhan Landscape Ecology Group Co., Ltd., Wuhan, China
Email: niechaoren@bjfu.edu.cn, niechaoren@163.com, *1870978491@qq.com

How to cite this paper: Nie, C.R., Xu, X.G., Zhang, X.Q., Sun, H.B., Yu, J.Y. and Xia, W.S. (2024) Analysis on Genetic Diversity of 40 Flowering Cherry Cultivars and Construction of Molecular ID Based on SSR Markers. *Agricultural Sciences*, **15**, 256-273.
https://doi.org/10.4236/as.2024.152015

Received: December 6, 2023
Accepted: February 17, 2024
Published: February 20, 2024

Abstract

Studying on the genetic diversity and genetic relationship of flowering cherry cultivars is extremely important for germplasm conservation, cultivar identification and breeding. Flowering cherry is widely cultivated as an important woody ornamental plant in worldwide, especially Japan, China. However, owning to the morphological similarity, many cultivars are distinguished hardly in non-flowering season. Here, we evaluated the genetic diversity and genetic relationship of 40 flowering cherry cultivars, which are mainly cultivated in China. We selected 13 polymorphicprimers to amplify to allele fragments with fluorescent-labeled capillary electrophoresis technology. The population structure analysis results show that these cultivars could be divided into 4 subpopulations. At the population level, N_a and N_e were 6.062, 4.326, respectively. H_o and H_e were 0.458 and 0.670, respectively. The Shannon's information index (I) was 1.417. The Pop3, which originated from *P. serrulata*, had the highest H_o, H_e, and I among the 4 subpopulations. AMOVA showed that only 4% of genetic variation came from populations, the 39% variation came from individuals and 57% ($p < 0.05$) came from intra-individuals. 5 polymorphic SSR primers were selected to construct molecular ID code system of these cultivars. This analysis on the genetic diversity and relationship of the 40 flowering cherry cultivars will help to insight into the genetic background, relationship of these flowering cherry cultivars and promote to identify similar cultivars.

Keywords

Flowering Cherry, SSR, Genetic Relationship, Molecular ID, Identifying Cultivars

1. Introduction

Flowering cherry refers mainly to ornamental plants in the subgenus *Cerasus*, which belongs to the genus *Prunus* of the family Rosaceae [1]. They are extremely important woody flowering plants in early spring owing to their beautiful blossoms, high ornamental characteristics and wide adaptability and play a key role in urban and rural landscaping. The subgenus *Cerasus* consisting of more than 150 species worldwide distributes mainly in the subtropical, warm temperate, and temperate regions of the northern hemisphere [2]. China is one of the centres of origin and diversity for subg. *Cerasus*, with 48 species and 10 varieties, accounting for nearly one-third of subg. *Cerasus* species [3] [4]. There are more than 300 cultivars in worldwide [5], about 70 cultivars are grown in China [6].

In order to identify and preserve flowering cherry germplasm resources, it is necessary to understand their genetic relationships and population structures and develop a molecular ID code system. Amplification with simple sequence repeats (SSR) molecular markers is one of the most widely used methods for studying genetic diversity and population structure [7]. Over the past two decades, many studies have been conducted on flowering cherry cultivars using SSR markers. Tsuda *et al.* assessed the genetic structures of 12 natural populations of *C. jamasakura* with 10 nuclear SSR markers [8]. Yueliang Lv *et al.* analysed the populations of 13 *P. campanulata* cultivars, also using SSR [9]. Qiong Zhang used 24 SSR primers to analyse 96 members of the subg. *Cerasus* [10]. Shuri Kato used SSR markers to study the origin of Japanese flowering cherry [11], and Tao Fu identified 11 wild flowering cherries [12]. Jie Chen analysed the genetic diversity and structure of *P. serrulata* and Pei Wu examined the genetic characteristics of flowering cherries based on SSR molecular markers [13] [14]. However, there has been little research on the genetic structures of flowering cherry cultivars, or the development of a molecular ID code system. Capillary electrophoresis technology based on Next-Generation Sequencing (NGS) technology is currently being used to detect fragments amplified using SSR markers, provides results that are more accurate, sensitive, and efficient than conventional methods, and is better suited to the analysis of large numbers of samples. This approach has also been used to construct DNA fingerprints, or molecular ID code systems [15]. These methods have been applied widely to crops, fruits, ornamental plants, and other species.

In this study, we assessed the genetic structures and diversity of 40 flowering cherry cultivars and developed a molecular ID code system using SSR fluorescent-labeled capillary electrophoresis technology. This will lay a foundation for germplasm resources protection, cultivar identification.

2. Materials and Methods

2.1. Plant Materials

All the cultivars used in this study could be classified into five taxonomic groups

[3], hereafter referred to as Populations (Table 1). And they were taken from the Anshan nursery at the Wuhan Institute of Landscape Architecture, 3 - 5 young leaves of each cultivar were collected at random from three individuals on May 2021. All samples were enclosed in plastic bags and stored at −80˚C in the laboratory prior to DNA extraction.

2.2. DNA Extraction and PCR Amplification

Genomic DNA were extracted using the Rapid Plant Genome DNA Isolation Kit (B518231, Sangon Biotech, Shanghai, China) according to the production instruction. The DNA concentration was measured with a NanoDrop 2000 and then was stored at −20˚C. We collected 38 SSR primers from some studies on *Subgenes Cerasus*. These primers were shown in Supplementary Table S1. These primers were labeled with fluorescent dyes (FAM, HEX, or TAM). The amplified products were analyzed by polyacrylamide gel electrophoresis, and after gel imaging, selecting polymorphic primers according to the bands quality.

Table 1. List of plant materials used in the study.

Number	cultivar	taxon	Number	cultivar	taxon
1	*P. cerasiodes* var. *rubea*	pop1	21	*P. subhirtella* "*Ujou-shidare*"	pop4
2	*P. campanulata* "*Yangming*"	pop2	22	*P. kanzakura* "*Praecox*"	pop2
3	*P. campanulata* "Kanhizakura-plena"	pop2	23	*P. pseudocerasus* "*Keio-zakura*"	pop1
4	*P. serrulata* "Hongye"	pop3	24	*P. campanulata* "*Ryukyu-hizakura*"	pop2
5	*P. pseudocerasus* "Introsa"	pop1	25	*P. kanzakura* "*Yokohama-hizakura*"	pop2
6	*P. subhirtella* "Plena Rosea"	pop4	26	*P.* "*Youkou*"	pop2
7	*P. serrulata* "Speciosa"	pop3	27	*P. jamasakura* "*Sendaiya*"	pop5
8	*P. sieboldii* "Beni-yutaka"	pop3	28	*P.* ×*subhirtella* "*Autumnalis*"	pop4
9	*P.* "Yoshino-shidare"	pop4	29	*P. jamasakura* "*Imperialis*"	pop5
10	*P. conradinae*	pop1	30	*P. serrulata* "*Taihaku*"	pop3
11	*P. serrulata* "Hisakura"	pop3	31	*P. siebildii* "*Caespitosa*"	pop4
12	*P. campanulata* "Feihan"	pop2	32	*P. serrulata* "*Kouka*"	pop3
13	*P. yedoensis* "Somei-yoshino"	pop4	33	*P. serrulata* "Grandiflora"	pop3
14	*P. serrulata* "Superba"	pop3	34	*P. serrulata* "Benitemari"	pop3
15	*P. serrulata* "Albo-rosea"	pop3	35	*P. serrulata* "Senriko"	pop3
16	*P. serrulata* "Sekiyama"	pop3	36	*P. serrulata* "Sphaerantha"	pop3
17	*P. campanulata*	pop2	37	*P. serrulata* "Arasiyama"	pop3
18	*P. kanzakura* "Tairyo-zakura"	pop2	38	*P. jamasakura* "Ichihara"	pop5
19	*P. serrulata* "Yeabeni-ohshima"	pop3	39	*P. serrulata* "Imose"	pop3
20	*P. serrulata* "Mollis"	pop3	40	*P. subhirtella* "Yeabeni-higan"	pop4

PCR was performed using 25 μL of reaction solution, the detail as follow: 0.5 μL 10 mM dNTP, 2.5 μL Taq buffer, 0.5 μL of each primer at 20 μmol·L^{-1}, 1 μL genomic DNA, 2.5 μL 25 mM $MgCl_2$, 1.0 U Taq DNA polymerase, and 17.8 μL double-distilled H_2O. Amplification was carried with the following parameters: an initial denaturation at 95˚C for 3 min, ten cycles of denaturation at 95˚C for 30 s, annealing at 60˚C for 35 s, and extension at 72˚C for 30 s. This was followed by 20 cycles of denaturation at 95˚C for 30 s, annealing at 60˚C for 30 s, and extension at 72˚C for 30 s, with a final extension at 72˚C for 6 min. The PCR products were detected with ABI 3730xl DNA Analyser (Applied Biosystems, USA).

2.3. Data Processing

2.3.1. Cluster and Non-Metric Multidimensional (NMDS) Analysis

Cluster analysis based on Nei' coefficient using UPGMA was performed with MVSP ver. 3.2. NMDS analysis was conducted using the package vegan (2.5.6) in R [16].

2.3.2. Genetic Diversity and Differentiation

GenAIEx 6.502 was used to assess the genetic diversity parameters per locus and population [17] [18]. We also conducted analysis of molecular variance (AMOVA, 1000 permutations) among the populations with the same software. We also analysed the following parameters: number of alleles (N_a), number of effective alleles (N_e), observed heterozygosity (H_o), expected heterozygosity (H_e), number of migrants (N_m), Shannon's information index (I), coefficient of genetic differentiation (F_{ST}), Fixation Index (F), and inbreeding coefficient within individuals (F_{IS}).

2.3.3. Population Structure Analysis

Population structure was analysed in Structure (version 2.3.4) software, Parameter settings were as follows: burn-in 10,000 iterations. Clustering number from 2 to 10, repeat times 15. The results were submitted to the online program Structure Harvester [19]. The optimal K value was calculated with the method developed by Evanno [20].

2.3.4. Construction of Molecular ID Code System

Coding was done according to the amplified fragment size. The detailed steps are as follows: 1) the amplified fragments were ordered by size; 2) codes consisted of combination of two Arabic numerals (*i.e.*, 01~99), Missing values were represented by 00. If the value exceed 99, it was encoded using two English letters (*i.e.*, aa - zz); and 3) the selected SSR primers were arranged in a fixed order. Molecular ID codes were then produced by sequentially combining all codes from the amplified fragments for each cultivar.

3. Results

3.1. Selecting the Polymorphic Primers

By literatures, 38 pairs of SSR primers of flowering cherry or similar species were

collected, and then 8 representative samples from the tested plant materials were used to PCR amplify for polymorphic primers selection. By gel imaging, the stable amplification, clear bands 13 polymorphism were pick out from 38 primers, The detail as Figure 1, these primers information also shown in Supplementary Table S1.

3.2. Cluster Analysis

As Figure 2 showing, the genetic similarities range from 0.87 to 0.20, with a mean of 0.494. These cultivars could be divided into four groups at the genetic similarity value of 0.26; the groups were designated A, B, C, and D. Group A was consisted of seven cultivars, with genetic similarities between 0.87 and 0.47. Five of the seven cultivars belonged to *P. subhirtella*, and the other two belonged to *P. yedoensis* "*Somei-yoshino*" and *P. pseudocerasus* "*Introsa*". There were21 cultivars in Group B, including 18 cultivars of *P. serrulata* and three of *P. jamasakura*. The genetic similarity valuein Group B ranged from 0.86 - 0.36, and were higher than Group A. This group was characterised by highly diverse flower colours, inflorescences, petal numbers, and flower types. *Prunus serrulata* "*Sekiyama*" and *P. serrulata* "*Hongye*" were found to be closely related, with the similarity value of 0.86. Group C comprised nine cultivars with genetic similarities ranging from 0.55 - 0.34; this group was mainly composed of cultivars and hybrid progenies of *P. campanulata.* Group D included *P. conradinae*, *P. cerasoides var. rubea*, and *P. pseudocerasus* "*Keio-zakura*".

3.3. Non-Metric Multidimensional (NMDS) Analysis

The stress value in the NMDS analysis was 0.22, indicating a poor fit according to Kruskal J B's standard of division stress value [21]. In Figure 3(a), The linear fit was poor (R^2 = 0.775) but the non-metric fit was high (R^2 = 0.95), suggesting that non-metric analyses were appropriate. In Figure 3(b), the sizes of the sample points with bubbles were representative of the fitness value; larger bubbles indicated poorer fit. Figure 3(c) showed the clustering of the cultivars in the two-dimensional space of nmds1 and nmds2. As the cluster analysis, the NMDS

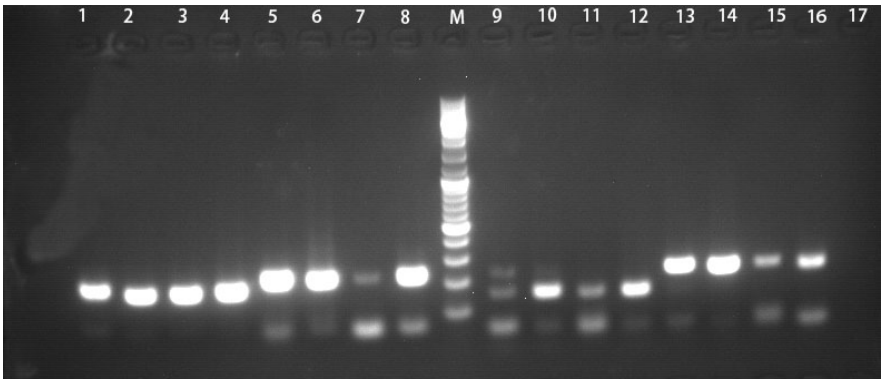


Figure 1. Partial screen of SSR primer. M = DNA Marker; 1 - 4 = PTCR1; 5 - 8 = CPSCT029; 9 - 12 = BPPCT005; 13 - 16 = BPPCT0037.

analysis indicated that the cultivars separated into four groups when K = 4. Group 1 included *three accessions*: *P. conradinaes* (10), *P. cerasoides var. Rubea* (1), and *P. pseudocerasus* "*Keio-zakura*" (23). Since these belong to three different species in Subgen. *Cerasus*, their spatial distributions were scattered in Figure 3(c). Group 2 included nine cultivars of bell flower cherry (*P. campanulata*). The Groups 3 and 4 cultivars were native to Japan and belonged to *P.serrulata* and *P. subhirtella*, respectively, indicating a narrow genetic background.

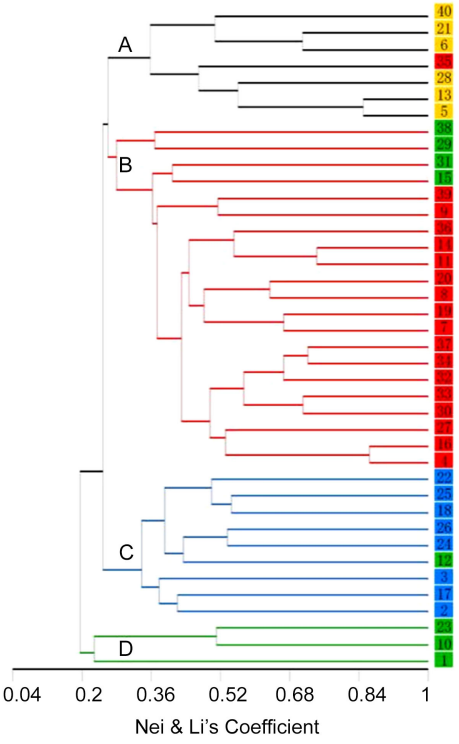
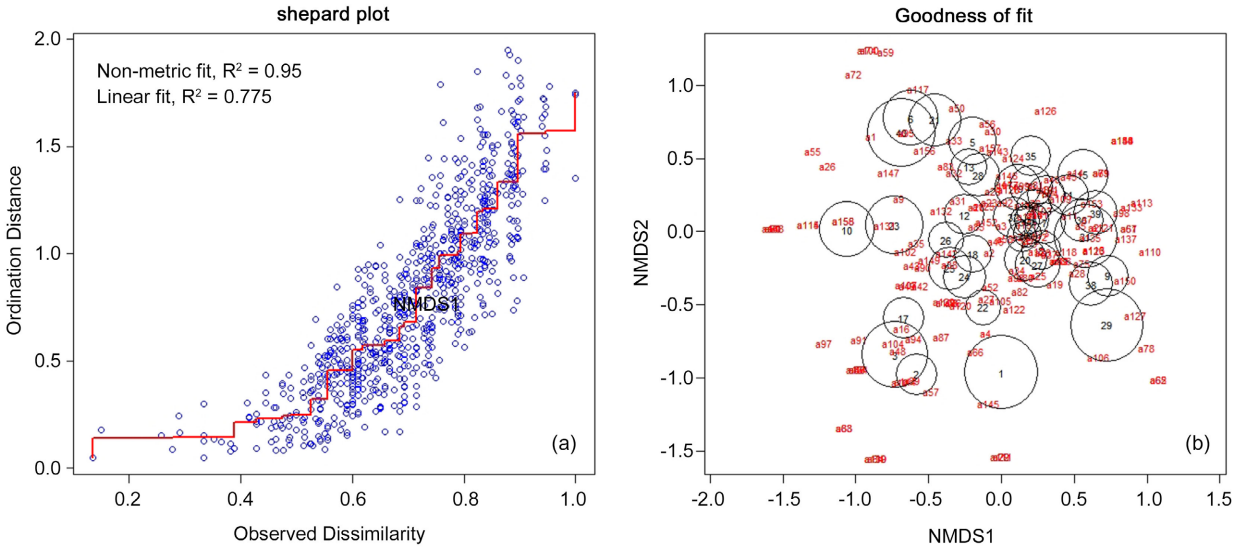


Figure 2. Cluster of 40 flowering cherry cultivars based on Nei & Li's Coefficient.



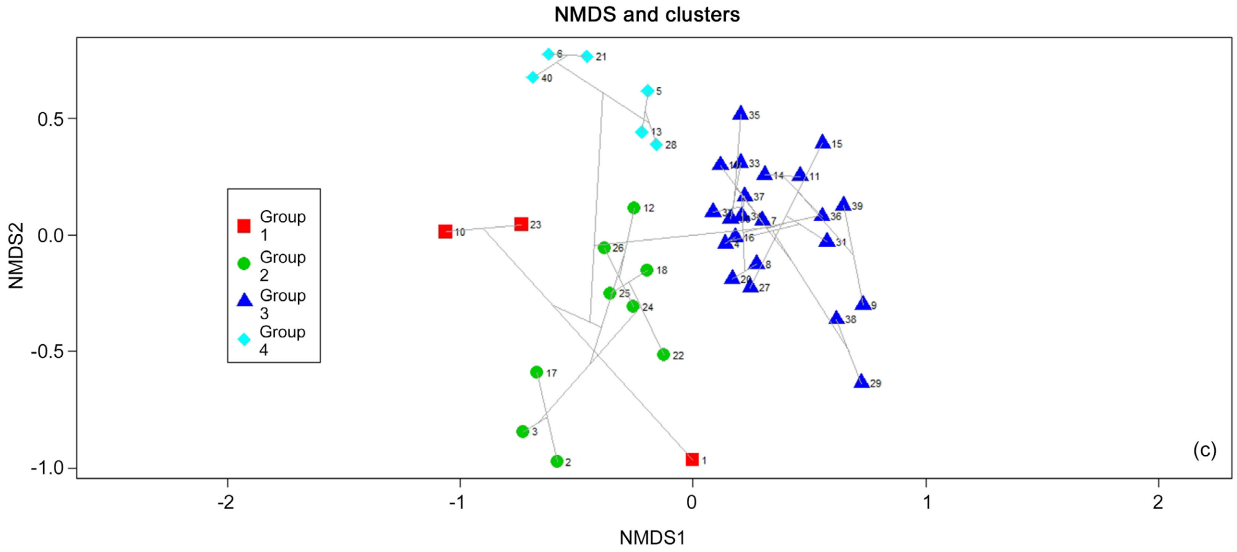


Figure 3. NMDS and clusters analysis of 40 flowering cherry cultivars.

3.4. Genetic Diversity and Differentiation

A total of 167 alleles were obtained from the 40 flowering cherry cultivars using the 13 SSR primers, with an average of 12.8 alleles per locus in Table 2. Among the locuses, the observed heterozygosity (H_o) and expected heterozygosity (H_e) ranged from 0.161 - 0.825 and 0.338 - 0.853, with means of 0.458 and 0.670, respectively. H_e was higher than H_o. The mean Shannon's information index (I) and number of effective alleles (N_e) were 1.91 and 5.74, respectively. The number of migrants (Nm) ranged from 0.016 to 0.763, with a mean of 1.576. The fixation index (F) ranged from 0.712 to 2.631, with a mean of 0.318. Random breeding would produce F values close to zero, whereas larger positive values were indicative of inbreeding. The inbreeding coefficients within the sub-populations (F_{ST}) ranged from 0.087 to 0.252, with a mean of 0.157.

The population-level genetic diversity parameters were shown in Table 3. Among the populations, N_a and N_e were 3.923 - 11.232 and 2.700 - 4.814, with means of 6.062 and 4.326, respectively. H_o ranged from 0.410 to 0.453, and H_e from 0.564 to 0.770, with means 0.458 and 0.670, respectively. H_o was higher than *He* in all populations. *I* value ranged from 1.146 to 1.924, with a mean of 1.417. Population 3, which originated from *P. serrulata*, had higher values for H_o, H_e, and *I* compared to other populations, indicating that this population had higher levels of genetic diversity and differentiation. The *F* values ranged from 0.228 to 0.397, with a mean of 0.321. *F* was expected to be close to zero in situations of random mating, or under Hardy-Weinberg equilibrium.

The F_{ST} value among the five populations was 0.036, suggesting that inter-population variability only accounted for 3.6% of the genetic variation in the samples, whereas the other 96.4% of the variation occurred within the populations. Thus, intra-population variation was the major variation in the samples. The AMOVA assessed variation both populations and individuals (Table 4); the

results indicated that variation between populations accounted for 4%, the39% variation came from individuals and 57% ($p < 0.05$) came from intra-individuals. The results showed that the genetic variation in the 40 cultivars were mainly attributable to variation within individuals; In contrast, the genetic differentiation between the populations was very low.

Table 2. Genetic diversity parameters of 13 SSR locuses.

Locus	*N*	*Na*	*Ne*	*I*	*Ho*	*He*	*uHe*	*F*	*FIS*	*FST*	*Nm*
AM288205	7.600	3.600	2.475	1.039	0.448	0.587	0.650	0.228	0.237	0.129	1.684
CPSCT012	8.000	4.200	3.164	1.215	0.161	0.663	0.732	0.763	0.757	0.161	1.301
DY640364	7.600	3.200	1.705	0.639	0.325	0.338	0.368	0.051	0.036	0.246	0.768
EMPA022	7.800	5.400	4.174	1.378	0.334	0.673	0.736	0.598	0.504	0.217	0.904
EMPA026	7.800	10.000	8.080	2.093	0.806	0.853	0.942	0.060	0.055	0.087	2.631
EMPA027	7.800	6.200	3.924	1.526	0.544	0.741	0.819	0.277	0.265	0.117	1.890
EMPAS02B	7.600	9.600	7.600	2.038	0.825	0.840	0.927	0.016	0.018	0.106	2.114
M13B	8.000	2.200	1.705	0.555	0.200	0.330	0.365	0.217	0.394	0.260	0.712
PBBCT34	8.000	8.400	6.138	1.854	0.717	0.797	0.879	0.089	0.101	0.132	1.651
PCCGA25	8.000	7.200	4.296	1.608	0.367	0.749	0.825	0.519	0.511	0.099	2.285
PCHGMS1	8.000	7.200	5.728	1.708	0.200	0.786	0.863	0.755	0.746	0.130	1.677
PCHGMS3	8.000	7.000	3.939	1.567	0.567	0.727	0.805	0.243	0.220	0.105	2.122
UDP96-018	7.600	4.600	3.316	1.199	0.458	0.628	0.689	0.312	0.271	0.252	0.744
Mean	7.831	6.062	4.326	1.417	0.458	0.670	0.738	0.318	0.317	0.157	1.576

Table 3. Genetic diversity parameters for 5 populations.

population	*N*	*Na*	*Ne*	*I*	*Ho*	*He*	*uHe*	*F*
pop1	4.000	4.769	4.261	1.412	0.462	0.712	0.813	0.401
pop2	8.615	6.692	4.814	1.539	0.410	0.690	0.732	0.397
pop3	17.538	11.231	6.452	1.924	0.532	0.770	0.792	0.305
pop4	6.000	3.923	2.700	1.064	0.423	0.564	0.615	0.228
pop5	3.000	3.692	3.404	1.146	0.462	0.615	0.738	0.276
Mean	7.831	6.062	4.326	1.417	0.458	0.670	0.738	0.321

Table 4. AMOVA of 40 flowering cherry cultivars.

Source of vairiation	df	SS	MS	F	percentage of variation
Among Pops	4	39.272	9.818	1.374	4%
Among Indiv	35	250.028	7.144	2.362**	39%
Within Indiv	40	121.000	3.025	3.246*	57%
Total	79	410.300			100%

3.5. Genetic Structure Analysis

A genetic structure analysis provided detailed information on the level of genome admixture between the populations. The results from Structure and Structure Harvester indicated that the ΔK was optimal at K = 4 (Figure 4). It followed that the optimal number of subpopulations was 4; that was to say, all of the cultivars could be divided into 4 subgroups, we referred to as Population 1-4. As Figure 5, the proportions of the 4 subpopulations were 0.148, 0.214, 0.452, and 0.186, respectively. The H_e values were 0.6635, 0.8553, 0.7589, and 0.6172,

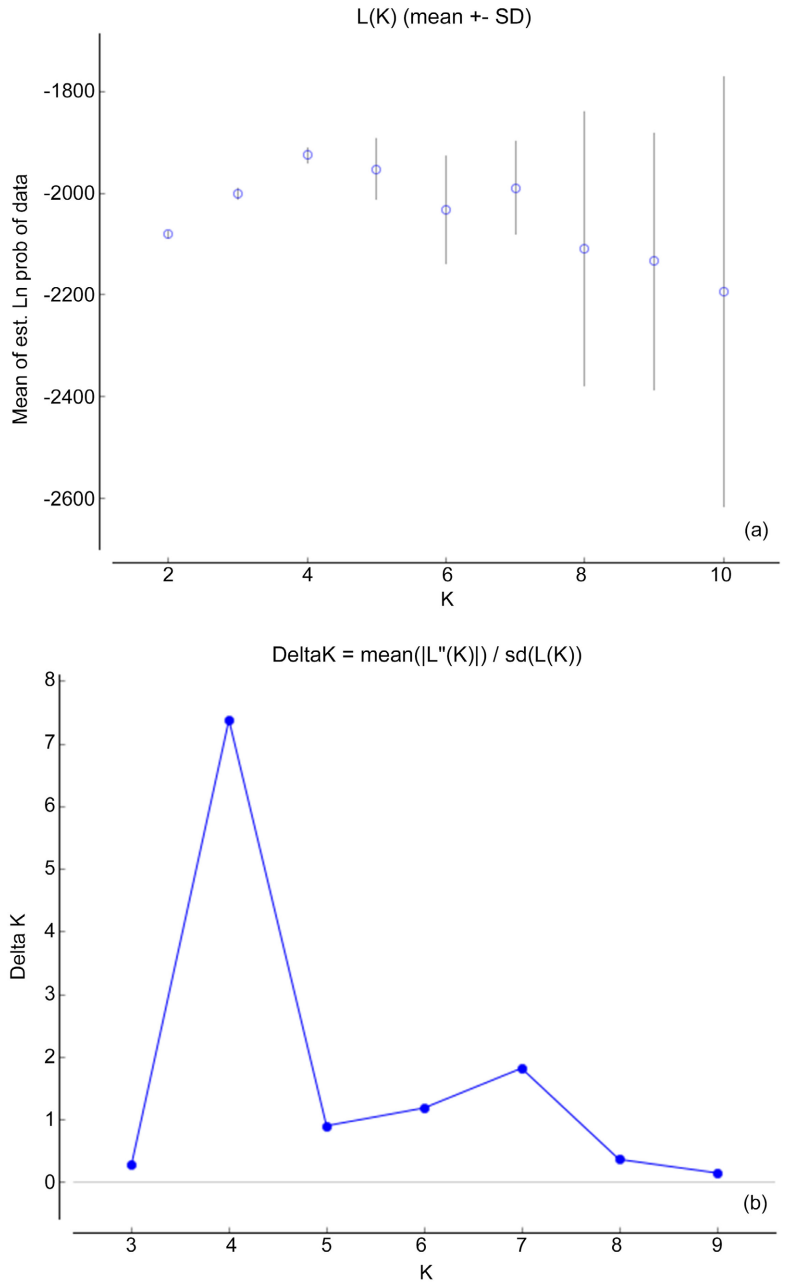


Figure 4. Diagnostic plots of L(K) and ΔK from the STRUCTURE analysis 40 flowering cherry cultivars.

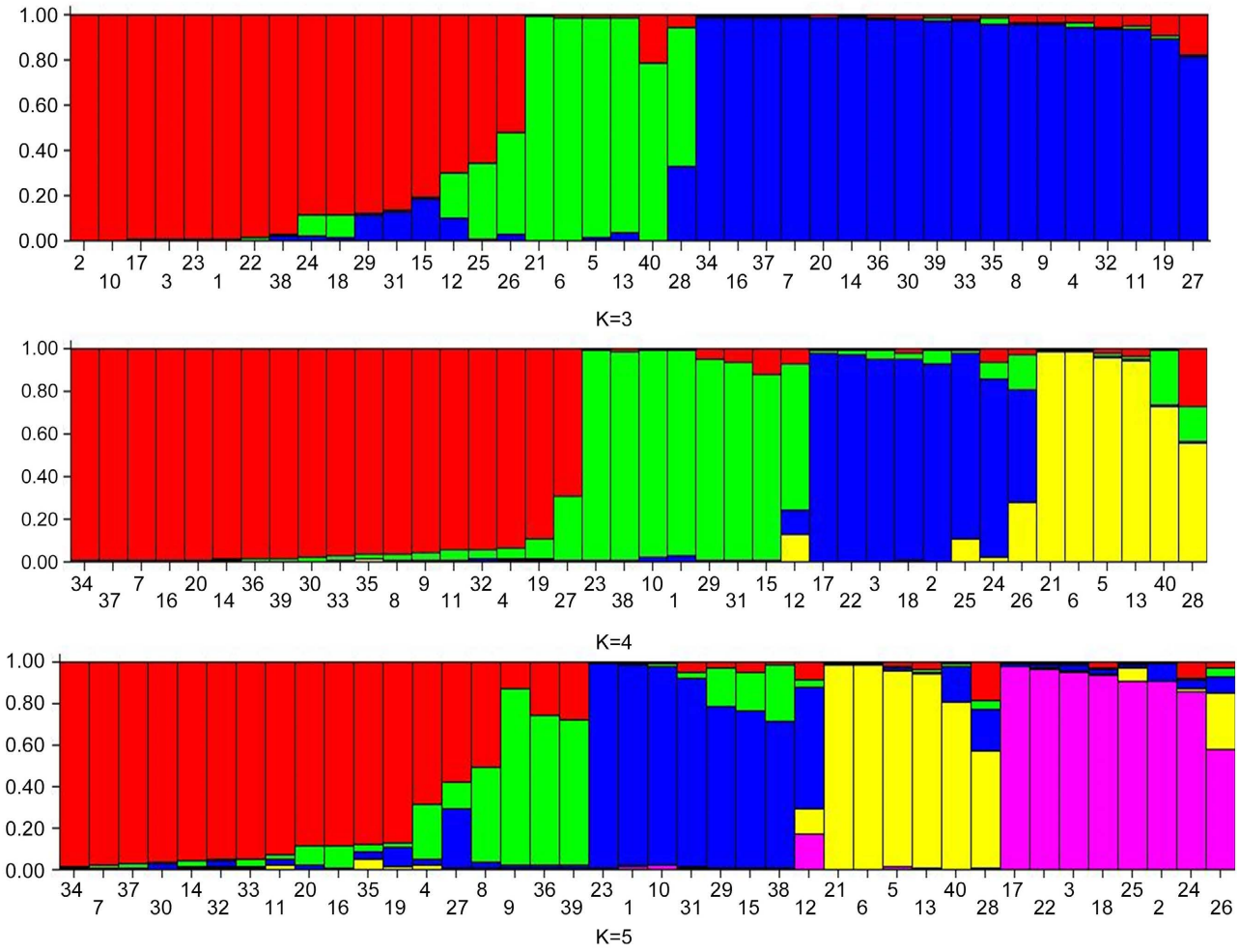


Figure 5. Population genetic structure of 40 flowering cherry cultivars at K = 3, 4, 5.

and the F_{ST} values were 0.1660, 0.0015, 0.1543, and 0.2898, respectively. The results essentially reflected the phylogenetic relationships. Q value ≥ 0.6 indicated relatively pure lines, and Q values < 0.6 indicated admixture. In this study, the Q values for 38 samples were ≥0.6, showing that these lines were relatively pure. Only two individuals were Q values < 0.6, indicating that the two samples were admixed, had a highly diverse genetic composition or were of mixed origin (**Figure 5**).

3.6. Construction of Molecular ID Code System

The size of the fragments amplified using the 13 SSR primers ranged from 111 to 298 bp. These SSR primers PBBCT34, EMPAS02B, EMPA026, PSCHGMS1, and EMPA027 were found to be highly polymorphic and suitable for classifying ten cultivars at least; the PBBCT34, EMPAS02B, and EMPA026primers could classify 21, 26, and 30 cultivars, respectively. All samples could be classified adequately using a combination of these three primers. To allow for the addition of cultivars in the future, we integrated one additional primerEMPA027, into mo-

lecular ID code system (Supplementary Table S2). As an example, the Japanese flowering cherrycultivar *P.* × *yedoensis* "*Somei-yoshino*", the sizes of the fragments obtained from the PBBCT34, EMPAS02B, EMPA026, and EMPA027 primer sets were 208/218 bp, 166/18 2bp, 206/212 bp, and 162/162 bp, respectively. According to Table 5, the corresponding fragment codes were 54/59, 33/41, 53/56, and 31/31, respectively. These were combined sequentially into a 16-bit string, which comprised the cultivar's molecular ID code. In the above example, the final molecular ID code was 5459334153563131. The molecular ID code of 40 flowering cherry cultivars were shown in Table 5.

4. Discussion

4.1. Genetic Relationships

Cluster analysis indicated that our results were generally consistent with morphological classifications [3]. Genetic variation between cultivars within the same lines was low, indicating that these lines had relatively narrow genetic backgrounds. This was attributable to seed production and cross breeding. Conversely, the genetic differences between lines were substantial, with the high diversity due to their wide distribution and numerous cultivars. These results were consistent with the previous findings [10] [13] [22].

Table 5. The molecular ID code of 40 flowering cherry cultivars.

cultivar	moleculer ID code	cultivar	moleculer ID code
P. "Youkou"	5494323762643131	*P. subhirtella* "Ujou-shidare"	5454333845454343
P. sieboldii "Beni-yutaka"	6372293254584040	*P. kanzakura* "Praecox"	5555374049553030
P. campanulata "Kanhizakura-plena"	7070464751543030	*P. pseudocerasus* "Keio-zakura"	6266404345582528
P. serrulata "Hongye"	6487373852562626	*P. campanulata* "Ryukyu-hizakura"	5971324958643146
P. pseudocerasus "Introsa"	5559000000002631	*P.* kanzakura "Yokohama-hizakura"	5559414940513131
P. subhirtella "Plena Rosea"	5461283251513030	*P. campanulata* "*Yangming*"	5765485363642929
P. serrulata "Speciosa"	6363414149582626	*P. jamasakura* "Sendaiya"	6363000051594343
P. serrulata "Taihaku"	6372333852582545	*P. subhirtella* "Autumnalis"	6169333840522547
P. "Yoshino-shidare"	6873292951522929	*P. jamasakura* "Imperialis"	6161343940492727
P. serrulata "Imose"	7390323254573434	*P. conradinae*	5762161655582828
P. serrulata "Hisakura"	5973151753532626	*P. siebildii* "Caespitosa"	6161333847574141
P. campanulata "Feihan"	5464414349642546	*P. serrulata* "Kouka"	6363384358582543
P. yedoensis "Somei-yoshino"	5459334153563131	*P. serrulata* "Grandiflora"	6372333851513943
P. serrulata "Superba"	5972323353532626	*P. serrulata* "Benitemari"	6372383953582626
P. serrulata "Albo-rosea"	5963323852524244	*P. serrulata* "Senriko"	5963333352572525
P. serrulata "Sekiyama"	6387373853572626	*P. serrulata* "Sphaerantha"	5859334649523434
P. campanulata	6464444555603636	*P. serrulata* "Arasiyama"	6363333858582543
P. kanzakura "Tairyo-zakura"	5959404347552626	*P. jamasakura* "Ichihara"	6882334050513939
P. serrulata "Yeabeni-ohshima"	5963334149584747	*P. cerasiodes var.rubea*	6875333545492820
P. serrulata "Mollis"	6263293358584343	*P. subhirtella* "Yeabeni-higan"	5454183255563030

NMDS analysis was used to assess the similarity of data between objects and illustrate the spatial relationships between them [3]. This method could sort, cluster and reduce dimensions. In this study, the close relationships of 40 samples were clearly and vividly illustrated in two-dimensional space. At K = 4, the cluster results were similar to UPGMA.

The results of the population structure analysis were consistent with those of the cluster analysis and the NMDS, with the 40 tested samples dividing into 4 groups. Each method had its own unique advantages; for instance, NMDS could reflect relationships between genetics and geographical distribution in two-dimensional space. However, for individuals with complex genetic backgrounds, the results of the different methods showed differences in clustering and population structure divisions, as demonstrated by the results of Haiwen Zhang [23].

4.2. Analysis of Genetic Diversity and Structure

The genetic diversity among cultivars was found to be very high. Compared to similar species, the mean of Shannon's information index (I = 1.417) was higher than those of P. mahaleb (I = 0.1720) [24] and *P. serrulata* (I = 0.939) [25], but lower than that of *P. pseudocerasus* (I = 1.525) [26]. The high genetic diversity may be associated with out-crossing mating systems, wide geographic distributions, and large difference in climate and habitat conditions among the cultivars. The diversity of the cultivars originating from *Prunus* subg. *Cerasus* was also an important factor.

Population genetic structure analysis was an important method to insight into genetic relationships among cultivars, and SSR molecular marker technology could be used to determine the degree of genetic variation and differentiation at the population level [26] [27] [28]. Genetic differentiation was an important parameter of population genetic structure. According to Wright [29], if $F_{ST} < 0.25$, genetic differentiation was low; in our study it was very low ($F_{ST} = 0.036$). The AMOVA indicated that only 4% of the genetic variation occurred among populations, whereas 39% occurred among individuals and 57% within individuals. Thus, genetic variation within the populations was the major variation. This was consistent with the results of Cai Yu Liang and Chen Jiao [26] [30]. Seed dispersal mechanisms had a significant effect on genetic differentiation among populations; because flowering cherries mainly relied on animal dispersal, the range of animal activities restricted dispersal [30].

Nm affected the genetic differentiation of populations. Generally speaking, when $Nm > 1$, genetic drift within populations was impeded, preventing differentiation [31]. The *Nm* value in this study was 1.577. So, genetic variation among the populations was restricted, and most genetic variation was found within populations.

4.3. Construction of Molecular ID Code System

Three methods were typically used to construct plant molecular ID code sys-

tems. The first was to use RFAP, RAPD, or AFLP molecular marker technologies, and labeled them 1 or 0 based on the presence or absence of the DNA bands and then the molecular ID code was added to the binary string [32] [33] [34] in decimal or by converting it to decimal, generating a decimal molecular ID code [35]. The second method, developed in recent years, used a combination of SSR technology and capillary electrophoresis to measure the sizes of amplified fragments. The amplified fragments were then sorted and encoded, and several primers' amplified fragment codes were combined to construct a molecular ID code [36] [37]. The third method was similar to the second method but produces a different type of ID code: the amplified fragments are encoded, and the resulting codes are combined in a sequence to form a molecular ID code [38].

We used the third method to construct a molecular ID code. To improve the distinguishability of SSR primers, we used both amplified fragments to encode in each locus. Given the increasing number of cherry cultivars and breeds, it was important to allow for increasing in future. Thus, we added one additional primer in the construction of our molecular ID code system. In this study, we encoded 99 fragments of 102 to 298 bp. If this was not sufficient for the number of codes required, two English letters (aa~zz) could also be appended, allowing for an additional 676 fragments. Therefore, this system could meet future needs and also ensure the ongoing uniformity of the 16-bit code.

Author Contributions

Experimental design: C.R.N; investigation: H.B.S., X.G.X.; data analysis: X.Q.Z. and C.R.N.; funding acquisition: C.R.N.; project administration, W..S.X., C.R.N., and X.Q.Z.; writing—original draft, C.R.N. writing—review & editing, C.R.N., X.Q.Z. and W.S.X. All the co-authors have read and approved the submitted version of the manuscript. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

We would like to acknowledge Hongling Kuang, Qinghua Wang who helped us with the sampling.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Kato, S., Matsumoto, A. and Yoshimura, K. (2014) Origins of Japanese Flowering Cherry (*Prunus subgenus Cerasus*) Cultivars Revealed Using Nuclear SSR Markers. *Tree Genetics & Genomes*, **10**, 477-487. https://doi.org/10.1007/s11295-014-0697-1

[2] Yu, D.J. and Li, C.L. (1986) Flora Reipublicae Popularis Sinicae. Vol. 38, Science

Press, Beijing.

[3] Wang, X.R. (2014) An Illustrated Monograph of Cherry Cultivars in China. Science Press, Beijing.

[4] Nie, C., Zhang, Y., Zhang, X., Xia, W., Sun, H., Zhang, S., Li, N., Ding, Z., Lv, Y. and Wang, N. (2023) Genome Assembly, Resequencing and Genome-Wide Association Analyses Provide Novel Insights into the Origin, Evolution and Flower Colour Variations of Flowering Cherry. *The Plant Journal*, **114**, 519-533. https://doi.org/10.1111/tpj.16151

[5] Kobayashi, Y. (1992) Flowering Cherry Cultivars. *Puranta*, **20**, 9-14.

[6] Wang, Q.H., Liu, X.H. and Xu, L. (2015) An Illustrated Monograph of Main Cerasus Cultivars in China. Zhejiang Science and Technology Press, Hangzhou.

[7] Litt, M. and Luty, J.A. (1989) A Hypervariable Microsatellite Revealed by *in Vitro* Amplification of a Dinucleotide Repeat within the Cardiac Muscle Actin Gene. *American Journal Human Genetics*, **4**, 397-401.

[8] Tsuda, Y., Kimura, M., Kato, S., Katsuki, T., Mukai, Y. and Tsumura, Y. (2009) Genetic Structure of *Cerasus jamasakura*, a Japanese Flowering Cherry, Revealed by Nuclear SSRs: Implications for Conservation. *Journal of Plant Research*, **122**, 367-375. https://doi.org/10.1007/s10265-009-0224-x

[9] Lv, Y.L. (2006) Genetic Diversity and Breeding Strategy of *Cerasus campanulata* Maximum. Nanjing Forestry University, Nanjing.

[10] Zhang, Q. (2013) Study on the Cultivar Resources Investigation of Flowering Cherry (Cerasus) in China and Analysis of Particle Species and Cultivars Using SSR Markers. Nanjing Forestry University, Nanjing.

[11] Shuri, K., Asako, M., Kensuke, Y., Toshio, K., Kojiro, I., Takayuki, K. and Yuzuru, M. (2014) Origins of Japanese.

[12] Fu, T., Wang, Z.L. and Lin, L. (2015) Study on the Identification Method of a System of the Cerasus Plants Germplasm Resources. *Acta Horticulturae Sinica*, **42**, 2455-2468.

[13] Chen, J. (2016) Genetic Diversity Analysis of *Cerasus serrulata* Populations Based on SSR Markers. Nanjing Forestry University, Nanjing.

[14] Wu, P. (2016) Study on the Genetic Characteristics of Flowering Cherries Based on SSR Molecular Markers. South China Agricultural University, Guangzhou.

[15] Xu, K., Li, F. and Wu, J.F. (2014) Fingerprint Identification Platform of Capillary Electrophoresis Detection with Fluorescent SSR Markers on National Winter Rapeseed Cultivars (Lines) Field Trials. *Chinese Journal of Oil Crop Sciences*, **36**, 150-159.

[16] Oksanen, J., Blanchet, F., Kindt, R. and Legendre, P. (2010) Vegan: Community Ecology Package.

[17] Peakall, R. and Smouse, P.E. (2012) GenAlEx6.5: Genetic Analysis in Excel Population Genetic Software for Teaching and Research—An Update. *Bioinformatics*, **28**, 2537-2539. https://doi.org/10.1093/bioinformatics/bts460

[18] Peakall, R. and Smouse, P.E. (2006) GENALEX6: Genetic Analysis in Excel Population Genetic Software for Teaching and Research. *Molecular Ecology Notes*, **6**, 288-295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

[19] Jakobsson, M. and Rosenberg, N.A. (2007) CLUMPP: A Cluster Matching and Permutation Program for Dealing with Label Switching and Multimodality in Analysis of Population Structure. *Bioinformatics*, **23**, 1801-1806. https://doi.org/10.1093/bioinformatics/btm233

[20] Evanno, G., Regnaut, S. and Goudet, J. (2005) Detecting the Number of Clusters of Individuals Using the Software STRUCTURE: A Simulation Study. *Molecular Ecology*, **14**, 2611-2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

[21] Kruskal, J.B. (1964) Multidimensional-Scaling by Optimizing Goodness-of-Fit to Non-Metric Hypothesis. *Psychometrika*, **29**, 1-27. https://doi.org/10.1007/BF02289565

[22] Su, Q. (2007) Genetic Diversity among Populations of *Cerasus campanulata.* Nanjing Forestry University, Nanjing.

[23] Zhang, H.W. (2020) SSR Marker-Based Genetic Diversity and Population Structure Analysis of Potato Germplasm. *Molecular Plant Breeding*, 1-17.

[24] Jordano, P. and Godoy, J.A. (2000) RAPD Variation and Population Genetic Structure in *Prunus mahaleb* (Rosaceae), an Animal-Dispersed Tree. *Molecular Ecology*, **9**, 1293-1305. https://doi.org/10.1046/j.1365-294x.2000.01009.x

[25] Yin, X.G. (2018) Analysis of Genetic Diversity of *Cerasus serrulata* Population by SSR. *Journal of Nanjing Forestry University* (*Natural Science Edition*), **42**, 25-31.

[26] Chen, F. (2013) Analysis of Population Structure in 82 Medicago spp. Germplasm. Yangzhou University, Yangzhou.

[27] Wang, J., Du, J.C. and Wang, Z.L. (2016) Genetic Diversity Analysis of 36 Alfalfa Accessions Based on SSR Markers. *Chinese Journal of Grassland*, **38**, 20-25, 33.

[28] Shi, X.H. and Guo, P.Z. (2000) Molecular Marker and Analysis of Population Genetic Structure in Wheat. *Journal of Capital Normal University* (*Natural Science Edition*), **1**, 59-65.

[29] Wright, S. (1965) The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. *Evolution*, **19**, 395-420. https://doi.org/10.1111/j.1558-5646.1965.tb01731.x

[30] Cai, Y.L. (2006) Genetic Analysis of the Wild Cherry Germplasm and Identification of Cultivated Cherry Cultivars Using DNA Fingerprints. Northwest University, Xi'an.

[31] Slatkin, M. (1985) Gene Flow in Natural Populations. *Annual Review of Ecology and Systematics*, **16**, 393-430. https://doi.org/10.1146/annurev.es.16.110185.002141

[32] Ohtsubo, K. and Nakamura, S. (2007) Cultivar Identification of Rice (*Oryza sativa* L.) by Polymerase Chain Reaction Method and Its Application to Processed Rice Products. *Journal of Agricultural and Food Chemistry*, **55**, 1501-1509. https://doi.org/10.1021/jf062737z

[33] Liu, X.L., Ma, L. and Chen, X.K. (2010) Establishment of DNA Fingerprint ID in Sugarcane Cultivars in Yunnan, China. *Acta Agronomica Sinica*, **36**, 202-210. https://doi.org/10.3724/SP.J.1006.2010.00202

[34] Zhao, X.Y., Huang, L. and Ren, X.P. (2010) Establishment of DNA Fingerprint Identity of Arachis Species with High Oil Content. *Acta Agriculturae Boreali-Sinica*, **25**, 64-70.

[35] Hao, D.M., Qiu, C.S. and Yu, W.J. (2011) Construction of Flax Molecular ID card System by RAPD Marker and Genetic Diversity Analysis. *Chinese Agricultural Science Bulletin*, **27**, 168-174.

[36] Xu, L.F., Ge, L. and Yuan, S.X. (2014) Using the Fluorescent Labeled SSR Markers to Establish Molecular Identity of Lily Germplasms. *Acta Horticulturae Sinica*, **41**, 2055-2064.

[37] Wan, Y.L., Zhang, J. and Liu, A.Q. (2018) Molecular Identification of Herbaceous Peony Cultivars Based on SSR Markers. *Journal of Northwest A & F University*

(*Natural Science Edition*), **46**, 90-97.

[38] Yin, Y. and Zhao, J.H. (2018) Establishment of Molecular Identity for Wolfberry Cultivars Based on SSR Markers. *Biotechnology Bulletin*, **34**, 195-201.

Supplementary

Table S1. Details of the 13 SSR primers used for flowering cherry analysis.

Primer Name	Sequence (5′ → 3′)	Length	Species	Origin
	R: GTTTGTTGCAATAGTCCCATCACTGC		EST-SSRs	
CPSCT012	F: ACGGGAGACTTTCCCAGAAG	156	*Prunus salicina*	Mnejja M *et al.* 2004
	R: CTTCTCGTTTCCTCCCTCCT		EST-SSRs	
DY640364	F: ACAACTCTTTCTGGGTTCATTGCT	228 - 238	*Cerasus.jamasakura*	Yoshiaki *et al.* 2009
	R: GTTTAAAACTCGTATCGTTCCCAAGGGT		EST-SSRs	
DY652293	F: ATACTTCGCGAAAATCACAAATCG	297 - 300	*Cerasus.jamasakura*	Yoshiaki *et al.* 2009
	R: GTTTCCACGAGAAGAAGACCGTGAGAAT		EST-SSRs	
EMPA022	F: CGATCTCTCTTCTCTTCGCTTC	161	*Cerasus avium*	Clarke J B *et al.* 2009
	R: CCACCCAAACCTCTCAAACC		genomic cDNA	
EMPA026	F: ATTGAAAAAGCCAAAGAGCG	219	*Cerasus avium*	Clarke J B *et al.* 2009
	R: TTCACGGTTTGAAGCAAGTG		genomic cDNA	
EMPA027	F: GCCAACACCCAAATGGTTAG	210	*Cerasus avium*	Clarke J B *et al.* 2009
	R: CTCTCCACGGTCTTGCTTTC		genomic cDNA	
EMPAS02B	F: CTACTTCCATGATTGCCTCAC	131 - 145	*Cerasus avium*	Vaughan S P *et al.* 2004
	R: AACATCCAGAACATCAACACAC		genomic cDNA	
M13B	F: AAGTGTGGGAGTCGGTGTCG	172 - 195	*Prunus.persica* (peach)	Yamamoto *et al.* 2002
	R: GCTCAATTTCGCTGCTTCCT		cDNA	Ohta *et al.* 2005
PBBCT34	F: CTACCTGAAATAAGCAGAGCCAT	228	*Prunus persica*	Dirlewanger E *et al.* 2002
	R: CAATGGAGAATGGGGTGC		genomic cDNA	
PCEGA25	F: GCAATTCGAGCTGTATTTCAGATG	145 - 198	Sour cherry	Cantini *et al.* 2001
	R: CAGTTGGCGGCTATCATGTCTTAC		genomic cDNA	
PCHGMS1	F: GGGTAAATATGCCCATTGTGCAATC	194	*Prunus persica*	Sosinski B *et al.* 2000
	R: GGATCATTGAACTACGTCAATCCTC		genomic cDNA	
PCHGMS3	F: ACGCTATGTCCGTACCATTCCCATG	170 - 230	*Prunus.persica* (peach)	Sosinski *et al.* 2000
	R: CAACCTGTGATTGCTCCTATTAAAC		genomic cDNA	Downey L. *et al.* 2000
UDP96-018	F: TTCTAATCTGGGCTATGGCG	232 - 271	*Prunus.persica* (peach)	Cipriani G. *et al.* 1999
	R: GAAGTTCACATTTACGACAGGG		genomic cDNA	Tesolin R. *et al.* 2000

Table S2. The amplified fragment codes.

code	Fragment (bp)	code	Fragment (bp)	code	Fragmen (bp)
0	-	34	168	68	236
1	102	35	170	69	238
2	104	36	172	70	240
3	106	37	174	71	242

Continued

4	108	38	176	72	244
5	110	39	178	73	246
6	112	40	180	74	248
7	114	41	182	75	250
8	116	42	184	76	252
9	118	43	186	77	254
10	120	44	188	78	256
11	122	45	190	79	258
12	124	46	192	80	260
13	126	47	194	81	262
14	128	48	196	82	264
15	130	49	198	83	266
16	132	50	200	84	268
17	134	51	202	85	270
18	136	52	204	86	272
19	138	53	206	87	274
20	140	54	208	88	276
21	142	55	210	89	278
22	144	56	212	90	280
23	146	57	214	91	282
24	148	58	216	92	284
25	150	59	218	93	286
26	152	60	220	94	288
27	154	61	222	95	290
28	156	62	224	96	292
29	158	63	226	97	294
30	160	64	228	98	296
31	162	65	230	99	298
32	164	66	232		
33	166	67	234		