Genotyping and Genetic Diversity Analysis of 47 Flowering Cherry (Cerasus) Germplasms

Yan Jiawen , Li Jianhui *, Bai Wenfu, Yu Lin, Nie Dongling, Xiong Ying, Li Bohai, Wu Sizheng

Hunan Forest Botanical Garden, Changsha, 410116, China

*These authors contributed equally to this work

Corresponding email: lioutu001@163.com

International Journal of Horticulture, 2020, Vol.10, No.6  doi: 10.5376/ijh.2020.10.0006

Received: 20 Oct., 2020
Accepted: 27 Oct., 2020
Published: 13 Nov., 2020

Copyright © 2020 Yan et al., This article was first published in Molecular Plant Breeding in Chinese, and here was authorized to translate and publish the paper in English under the terms of Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:
Yan J.W., Li J.H., Bai W.F., Yu L., Nie D.L., Xiong Y., Li B.H., and Wu S.Z., 2020, Genotyping and genetic diversity analysis of 47 flowering cherry (Cerasus) germplasms, International Journal of Horticulture, 10(6): 1-10 (doi: 10.5376/ijh.2020.10.0006).

Abstract In order to provide a theoretical basis for conservation and utilization of excellent flowering cherry germplasms, genetic diversity and genotype were analyzed based on single nucleotide polymorphism (SNP) molecular markers. A total of 47 germplasms were sequenced using restriction site-associated DNA sequencing (RAD-Seq) method, and highly consistent SNP sites have been identified for genetic diversity analysis and genotyping. The RAxML program was used for phylogenetic tree construction, based on the maximum likelihood (ML) method. The ADMIXTURE software was used to analyze genetic structure. The GCTA software was used for principal component analysis. Genetic diversity parameters, includes nucleotide diversity (π value), expected heterozygosity (exp He) and genetic differentiation index (FST) were calculated using the PopGenome and Arlequin 3.0 software. A total of 79,667 highly consistent SNPs were obtained, after genotype integrity filtering and closely linked sites screening. According to the phylogenetic, principal components and population structure analysis results, the 47 germplasms could be divided into eight genetic clusters. The π and exp He value of each cluster was as follows: I: 0.128 19, 0.119 94; II: 0.139 64, 0.130 55; III: 0.21 55, 0.164 73; IV: 0.086 82, 0.061 22; V: 0.095 49, 0.069 73; VI: 0.191 22, 0.10 61; VII: 0.16 85, 0.158 78; VIII: 0.290 98, 0.183 43. The cluster IV and V presented the largest genetic differentiation, with an FST value of 0.4612 89, followed by IV and VI, with an FST value of 0.456 958; the lowest genetic differentiation existed between the cluster VII and VIII, with an FST value of 0.0976 13. Except for cluster VII and VIII, the genetic differentiation among the other clusters was at or above the medium level. The results of genotyping showed that there were 29 individuals (61.70 %) with single genetic component, and 18 individuals (38.30 %) with two or three genetic components. Based on the results of genetic components analysis, there were four potential interspecific hybridizations. There was a high genetic diversity among 47 flowering cherry germplasms, obvious genetic differentiation among the eight clusters. The genetic components of natural hybrid individuals and artificial hybrid cultivars were revealed, and their parents were inferred, based on the genotyping results.

Keywords Flowering cherry; SNP; Genetic diversity; Genotyping

Flowering cherry is a general term for the Cerasus subgenus Cerasus in Rosaceae, as an ornamental woody plant cultivated worldwide (Jiang et al., 2018a). The germplasm resources of flowering cherry are distributed widely, frequent inter-species and intra-specific natural hybridization leads to its reticulated evolution, with a high degree of morphological variability, and the genetic background of some flowering cherry germplasms are not clear yet (Zhu et al., 2018a). In addition, the climatic conditions, cultivation techniques and other factors of different introduction sites lead to morphological differences in the same varieties (Jiang et al., 2018b). For these reasons, traditional methods such as morphological marker appears limited for classification and identification of flowering cherry germplasms, this is not conducive for the development and utilization of wild germplasms.

Due to many characteristics, such as abundant quantity, high polymorphism, not affected by tissue types, developmental periods and growth environment, molecular marker based on nucleic acid polymorphisms is a reliable strategy for identification of genetic relationship and diversity analysis (Zhu et al., 2018b). Currently, restriction fragment length polymorphism (RFLP), simple sequence repeats (SSR) and inter-simple sequence repeats (ISSR) markers have been widely used in the genetic diversity analysis of Cerasus (Zhu et al., 2018b;
Zhou et al., 2020). However, these traditional molecular markers have disadvantages such as high cost, technical difficulty and long experimental period (Zhou et al., 2018). Single nucleotide polymorphisms (SNP), as a third-generation molecular marker with many characteristics, such as large quantity, widely spread, strong representation and genetic stability, has been successfully applied in genetic diversity analysis, variety identification and genetic map construction of woody plants (Zhou et al., 2018).

With the continuous advancement of sequencing technology, the cost has been gradually decreased, this greatly facilitates the mining of candidate SNP sites (Yin et al., 2020). In order to reveal its genetic composition and diversity, and provide a reference for the utilization of flowering cherry germplasms, a total of 47 germplasms were sequenced using restriction site-associated DNA sequencing (RAD-Seq) method, and highly consistent SNP sites have been identified for genetic diversity analysis and genotyping.

1 Results
1.1 Sequence analysis and SNP calling
A total of 102.64 Gb clean data was generated by RAD sequencing, with average Q30 of 94.41%, average GC of 39.60%. Of the remaining trimmed reads, 93.45% to 96.82% was mapped to the reference genome (Table 1). A total of 7 550 159 SNPs were obtained, of which 79 667 SNPs with high consistency. Among them, there were 48 520 transitions (Ts) (A/G: 24 108, 30.26 %; C/T: 24 412, 30.64 %) and 31 147 transversions (Tv) (A/C: 7 998, 10.04 %; A/T: 9 467, 11.88 %; C/G: 5 872, 7.37 %; G/T: 7 810, 9.80 %). The observed transition/transversion (Ts/Tv) ratio was 1.56.

1.2 Genetic structure
1.2.1 Phylogenetic analysis
The phylogenetic analysis results (Figure 1; Table 2) show that 47 flowering cherry germplasms can be divided into eight clusters. There were 12 individuals in cluster I, of which nine (Yangmingshan No. 1, Yangmingshan No. 2, Yangmingshan No. 3, Meirensnahn No. 1, Alishan No. 1, Fujian No. 4, ‘Kanhizakur aplena’, Shanghai No. 1 and ‘Zhongguohong’) was Cerasus campanulata; Yunnan No. 1 and Yanling No. 12 belongs to C. yunnanensis and C. dielsiana, respectively; In addition, there is an unknown taxa in cluster I, was named Yongshun 46. Cluster II contains 11 individuals, including five C. pseudocerasus (Yongshun No. 1, Yongshun No. 5, Yongshun No. 50, Yongshun No. 51, and Tianmenshan No. 6), one C. conradiana (Yongshun No. 13), one C. discoidea (Yanling 13 No. 13), three unknown taxa (Zhangjiajie No. 1, Fujian No. 1 and Fujian No. 3), respectively. Cluster III Contains three individuals, including one cultivar (C. ‘Ryukyu-hizakura’) and two unknown taxa (Sangzhi No. 1 and Zhejiang No. 1). There were three C. clarofolia (Tianmenshan No. 1, Tianmenshan No. 4 and Tianmenshan No. 5) and two C. serrulata (Xinhua No. 1 and Sangzhi No. 5) in cluster IV and V, respectively. Cluster VI contains two cultivars or strains (C. ‘Accolade’ and C. subhirtella cv. subhirtella); Cluster VII Contains 11 cultivars or strains (C. × yedoensis ‘Yoshun’, C. × subhirtella ‘koshinensis’, C. spachiana ‘Jindai-akebono’, C. spachiana ‘Kamatsu-otome’, C. spachiana ‘Plena Rosea’, C. spachiana Pendula Rosea’, C. spachiana ‘Ujou-shidare’, C. × subhirtella ‘Autumnalis’, C. × subhirtella ‘Yaebeni-higan’, C. ‘Youkou’, and C. spachiana), and one C. subhirtella (Longshan No. 1). Cluster VIII contains one cultivar (C. × yedoensis ‘Somei-yoshino’) and one unknown taxa (Fujian No. 2).

1.2.2 Principal component analysis
The principal component analysis results (Figure 2; Table 2) also show that the 47 germplasms were divided into 8 clusters, the individuals in cluster II, III, IV, V and VI were completely consistent with the results of phylogenetic analysis; there were six individuals (Yangmingshan No. 2, Meirensnahn No. 1, Alishan No. 1, C. ‘Kanhizakur aplena’, Shanghai No. 1, and ‘Zhongguohong’) in cluster I; Cluster VII was included all individuals in cluster VII and VIII according the phylogenetic results; There were five individuals in cluster VIII, including three C. campanulata (Yangmingshan No. 1, Yangmingshan No. 3, and Fujian No. 4), one C. yunnanensis (Yunnan No. 1) and one C. dielsiana (Yanling No. 12). The was a high consistency between principal component analysis and phylogenetic analysis, which is 82.99%.

2
Table 1 The amounts of clean read, SNP, and match rate with the reference genome

| Sample No. | clean data (Gb) | Q30 (%) | GC content (%) | SNPs   | Mapped ratio (%) |
|------------|----------------|---------|----------------|--------|-----------------|
| 1          | 5.99           | 93.93   | 39.94          | 793 567 | 95.90           |
| 2          | 1.06           | 94.86   | 39.49          | 142 128 | 96.24           |
| 3          | 0.81           | 94.02   | 39.98          | 122 464 | 95.65           |
| 4          | 0.94           | 94.03   | 40.02          | 123 728 | 94.75           |
| 5          | 0.55           | 94.09   | 39.91          | 59 842  | 93.71           |
| 6          | 0.75           | 94.18   | 39.79          | 99 937  | 94.76           |
| 7          | 1.30           | 94.97   | 38.92          | 171 918 | 95.62           |
| 8          | 2.22           | 94.16   | 39.91          | 324 373 | 94.42           |
| 9          | 1.34           | 94.22   | 40.01          | 186 124 | 94.68           |
| 10         | 1.29           | 94.23   | 39.88          | 171 506 | 94.22           |
| 11         | 1.27           | 94.92   | 39.48          | 162 465 | 95.61           |
| 12         | 2.08           | 94.16   | 39.55          | 320 410 | 96.38           |
| 13         | 1.27           | 94.86   | 39.48          | 182 680 | 94.72           |
| 14         | 3.51           | 93.95   | 39.83          | 565 363 | 94.96           |
| 15         | 3.79           | 94.94   | 39.38          | 458 082 | 96.43           |
| 16         | 2.76           | 94.14   | 39.65          | 465 896 | 95.61           |
| 17         | 2.33           | 95.02   | 39.43          | 379 940 | 96.82           |
| 18         | 3.28           | 94.17   | 40.16          | 510 123 | 93.45           |
| 19         | 3.42           | 94.15   | 39.91          | 586 751 | 94.92           |
| 20         | 2.59           | 94.98   | 39.32          | 388 798 | 95.77           |
| 21         | 2.19           | 94.21   | 39.76          | 346 627 | 94.45           |
| 22         | 2.92           | 95.00   | 39.50          | 473 889 | 96.74           |
| 23         | 2.31           | 94.18   | 39.55          | 376 788 | 95.79           |
| 24         | 5.37           | 94.69   | 39.69          | 877 170 | 96.30           |
| 25         | 1.66           | 94.85   | 39.27          | 267 639 | 95.53           |
| 26         | 1.49           | 94.12   | 39.60          | 242 447 | 95.97           |
| 27         | 2.06           | 94.83   | 39.69          | 399 654 | 94.54           |
| 28         | 1.56           | 94.90   | 39.31          | 258 787 | 95.69           |
| 29         | 2.16           | 94.97   | 39.02          | 351 940 | 96.19           |
| 30         | 1.62           | 94.22   | 39.44          | 248 143 | 95.76           |
| 31         | 1.07           | 93.98   | 39.67          | 135 160 | 94.29           |
| 32         | 1.57           | 94.17   | 39.14          | 227 056 | 95.83           |
| 33         | 2.50           | 93.89   | 39.57          | 379 663 | 95.16           |
| 34         | 2.09           | 94.45   | 39.11          | 300 248 | 94.99           |
| 35         | 4.05           | 94.32   | 39.44          | 600 085 | 95.72           |
| 36         | 2.40           | 94.12   | 39.49          | 365 815 | 95.44           |
| 37         | 1.10           | 94.11   | 39.52          | 129 463 | 95.35           |
| 38         | 1.07           | 94.23   | 39.37          | 144 864 | 95.81           |
| 39         | 0.68           | 94.08   | 39.86          | 68 895  | 94.44           |
| 40         | 1.75           | 94.86   | 38.57          | 224 222 | 94.34           |
| 41         | 3.55           | 94.18   | 39.62          | 440 655 | 95.23           |
| 42         | 1.19           | 94.96   | 39.17          | 144 579 | 95.87           |
| 43         | 2.06           | 94.30   | 39.79          | 292 342 | 95.51           |
| 44         | 3.63           | 94.14   | 39.67          | 542 633 | 95.23           |
| 45         | 1.97           | 94.25   | 39.82          | 299 067 | 96.00           |
| 46         | 4.81           | 95.03   | 39.46          | 734 794 | 95.93           |
| 47         | 3.11           | 94.25   | 39.49          | 465 527 | 96.56           |
1.2.3 Genetic structure analysis

The results of genetic structure analysis showed that when the 47 germplasms were divided into 8 clusters (Figure 3; Table 2), the consistency with the results of the phylogenetic analysis was 87.23%. Among them, the individuals in cluster I, II, III, IV, V, and VI were completely consistent with the result of phylogenetic analysis. There were nine individuals in cluster VII, including seven cultivars (C. spachiana ‘Jindai-akebono’, C. spachiana ‘Komatsu-otome’, C. spachiana ‘Pendula Rosea’, C. spachiana ‘Ujou-shidare’, C. × subhirtella ‘koshinensis’, C. subhirtella cv. Subhirtella, and C. × subhirtella ‘Yaebeni-higan’), one C. subhirtella (Longshan No. 1), and one unknown taxa (Sangzhi No. 1). Cluster VIII was comprised of four cultivars or strains (C. spachiana, C. × yedoensis ‘Somei-yoshino’, C. × yedoensis ‘Yoshun’, and C. spachiana ‘Jindai-akebono’) and one unknown taxa (Fujian No. 2).

1.3 Genetic diversity

The results of genetic diversity analysis showed that there was a highest expected heterozygosity (Exp He) and Nucleotide diversity (\(\pi\)) in cluster VIII, with a value of 0.183 43 and 0.290 98, respectively, followed by cluster III (0.164 73 , 0.215 5), cluster VII (0.158 78, 0.1685), cluster II (0.130 55, 0.139 64), cluster I (0.119 94, 0.128 19), cluster VI (0.10 61, 0.191 22), cluster V (0.069 73, 0.09549), and cluster IV (0.061 22, 0.086 82) (Table 3).

The cluster IV and V presented the largest genetic differentiation, with an \(F_{ST}\) value of 0.4612 89, followed by IV and VI, with an \(F_{ST}\) value of 0.456 958; the lowest genetic differentiation existed between the cluster VII and VIII, with an \(F_{ST}\) value of 0.0976 13 (Table 4).
Table 2 The information of 47 flowering cherry (Cerasus) germplasms

| No. | Germplasm name         | Genetic background                     | Origin       |
|-----|------------------------|----------------------------------------|--------------|
| 1   | Yangmingshan No.1      | C. campanulata                         | Taipei, Taiwan |
| 2   | Yangmingshan No.2      | C. campanulata                         | Taipei, Taiwan |
| 3   | Yangmingshan No.3      | C. campanulata                         | Taipei, Taiwan |
| 4   | Meirenshan No.1        | C. campanulata                         | Taidong, Taiwan |
| 5   | Alishan No.1           | C. campanulata                         | Jiayi, Taiwan |
| 6   | Fujian No.4            | C. campanulata                         | Fuzhou, Fujian |
| 7   | C. ‘Kanhzakur aplena’  | C. campanulata                         | Fuzhou, Fujian |
| 8   | Shanghai No.1          | C. campanulata                         | Shanghai     |
| 9   | C. ‘Zhongguohong’      | C. campanulata                         | Fujian       |
| 10  | Yunnan No.1            | C. yunnanensis                         | Kunming, Yunnan |
| 11  | Yanling No.12          | C. dielsiana                           | Yanling, Hunan |
| 12  | Yongshun No.1          | C. pseudocerasus                       | Yongshun, Hunan |
| 13  | Yongshun No.5          | C. pseudocerasus                       | Yongshun, Hunan |
| 14  | Yongshun No.50         | C. pseudocerasus                       | Yongshun, Hunan |
| 15  | Yongshun No.51         | C. pseudocerasus                       | Yongshun, Hunan |
| 16  | Tianmenshan No.6       | C. pseudocerasus                       | Zhangjiajie, Hunan |
| 17  | Yongshun No.13         | C. conradiinae                         | Yongshun, Hunan |
| 18  | Yanling No.13          | C. discoides                           | Yanling, Hunan |
| 19  | Tianmenshan No.1       | C. clarofolia                          | Zhangjiajie, Hunan |
| 20  | Tianmenshan No.4       | C. clarofolia                          | Zhangjiajie, Hunan |
| 21  | Tianmenshan No.5       | C. clarofolia                          | Zhangjiajie, Hunan |
| 22  | Xinhua No.1            | C. serrulata                           | Xinhua, Hunan |
| 23  | Sangzhi No.5           | C. serrulata                           | Sangzhi, Hunan |
| 24  | Longshan No.1          | C. subhirtella                         | Longshan, Hunan |
| 25  | C. spachiana           | C. spachiana                           | Japan        |
| 26  | C. × yedoensis ‘Somei-yoshino’ | C. spachiana × C. speciosa | Japan     |
| 27  | C. × yedoensis ‘Yoshun’  | C. spachiana × C. sargentii             | Japan       |
| 28  | C. ‘Accolade’          | C. spachiana × C. sargentii             | Japan       |
| 29  | C. spachiana ‘Jindai-akebono’ | C. spachiana × ?                     | Japan       |
| 30  | C. spachiana ‘Komatsu-otome’ | C. spachiana complex                | Japan       |
| 31  | C. spachiana ‘Plena Rosea’ | C. spachiana complex                | Japan       |
| 32  | C. spachiana ‘Pendula Rosea’ | C. spachiana complex                | Japan       |
| 33  | C. spachiana ‘Ujou-shidare’ | C. spachiana complex                | Japan       |
| 34  | C. × subhirtella ‘koshinensis’ | C. spachiana × C. sargentii           | Japan       |
| 35  | C. × subhirtella ‘Autunnalis’ | (C. spachiana × C. incisa) × C. subhirtella | Japan  |
| 36  | C. subhirtella cv. subhirtella | Unknown                                  | Japan       |
| 37  | C. × subhirtella ‘Yaebeni-higan’ | C. incisa × C. spachiana               | Japan       |
| 38  | C. ‘Youkou’            | C. × yedoensis × C. campanulata        | Japan       |
| 39  | C. ‘Ryukyu-hizakura’   | Unknown                                 | Japan       |
| 40  | Sangzhi No.1           | Unknown                                 | Sangzhi, Hunan |
| 41  | Yongshun No.46         | Unknown                                 | Yongshun, Hunan |
| 42  | Xinhua No.13           | Unknown                                 | Xinhua, Hunan |
| 43  | Zhejiang No.1          | Unknown                                 | Hangzhou, Zhejiang |
| 44  | Zhangjiajie No.1       | Unknown                                 | Zhangjiajie, Hunan |
| 45  | Fujian No.1            | Unknown                                 | Fuzhou, Fujian |
| 46  | Fujian No.2            | Unknown                                 | Fuzhou, Fujian |
| 47  | Fujian No.3            | Unknown                                 | Fuzhou, Fujian |
Figure 2 Principal component analysis (PCA) of 47 germplasms of flowering cherry

Note: 1–47: Table 2

Figure 3 Genetic structure of 47 germplasms of flowering cherry

Note: 1–47: Table 2

Table 3 Genetic diversity of flowering cherry germplasms

| Clusters | Observed heterozygosity | Observed homozygosity | Expected heterozygosity | Expected homozygosity | Nucleotide diversity (π) |
|----------|-------------------------|-----------------------|-------------------------|-----------------------|--------------------------|
| I        | 0.058 15                | 0.941 85              | 0.119 94                | 0.880 06              | 0.128 19                 |
| II       | 0.141 35                | 0.858 65              | 0.130 55                | 0.869 45              | 0.139 64                 |
| III      | 0.231 16                | 0.768 84              | 0.164 73                | 0.835 27              | 0.215 5                  |
| IV       | 0.093 01                | 0.906 99              | 0.061 22                | 0.938 78              | 0.086 82                 |
| V        | 0.040 42                | 0.959 58              | 0.069 73                | 0.930 27              | 0.095 49                 |
| VI       | 0.198 23                | 0.801 77              | 0.106 1                | 0.893 9               | 0.191 22                 |
| VII      | 0.183 16                | 0.816 84              | 0.158 78                | 0.841 22              | 0.168 5                  |
| VIII     | 0.358 63                | 0.641 37              | 0.183 43                | 0.816 57              | 0.290 98                 |

Table 4 Pairwise genetic differentiation index ($F_{ST}$) values among eight clusters of flowering cherry germplasms

| Clusters | II     | III    | IV     | V      | VI     | VII    | VIII   |
|----------|--------|--------|--------|--------|--------|--------|--------|
| I        | 0.172 865 | 0.165 023 | 0.332 268 | 0.327 347 | 0.348 987 | 0.379 151 | 0.377 077 |
| II       | -      | 0.182 35 | 0.203 843 | 0.237 985 | 0.266 357 | 0.323 268 | 0.307 898 |
| III      | -      | -      | 0.315 755 | 0.252 891 | 0.285 669 | 0.297 074 | 0.293 864 |
| IV       | -      | -      | -      | 0.461 289 | 0.456 958 | 0.320 849 | 0.425 051 |
| V        | -      | -      | -      | -      | 0.382 665 | 0.295 951 | 0.368 801 |
| VI       | -      | -      | -      | -      | -      | 0.116 995 | 0.244 266 |
| VII      | -      | -      | -      | -      | -      | -      | 0.097 613 |
2 Discussion
2.1 Genetic diversity and differentiation

Previous studies have shown that there are obvious differences in the nucleotide diversity of woody plants. The π value of *Pinus bungeana* is only 0.001 59 on average, which is extremely low (Yang et al., 2019), while the π value of *Litsea populifolia* is generally higher than 0.3, with a highly diversity (Wang et al., 2019). In this study, the nucleotide diversity of cluster I, II, III, VI, VII, and VIII was higher than the single species cluster IV and V, due to the former contains multiple species or interspecies hybrids. In addition, the π values of cluster IV and V were higher than *C. dielsiana*, with π value of 0.003 4 ~ 0.003 72 (Zhu et al., 2019). For the same reason, the expected heterozygosity of cluster IV (0.061 22) and V (0.069 73) was lower than that of other clusters, and also much lower than that of *Prunus mira* (0.3~0.63) (Bao et al., 2018), *Cerasus jamasakura* (0.665~0.817) (Tsuda et al., 2009) and *C. serrulata* (0.488) (Yi et al., 2018) and other related species. In actual research, when the $F_{ST} \leq 0.05$, no obvious genetic differentiation among populations; when the $F_{ST} \geq 0.1$, genetic differentiation was moderate or large (Willing et al., 2012). In this study, the genetic differentiation between cluster IV and V was the largest, with a $F_{ST}$ value of 0.461 289; and the genetic differentiation between cluster VII and VIII was the smallest, with a $F_{ST}$ value of 0.097 613. The genetic differentiation among most clusters was at or above the medium level, except for cluster VII and VIII.

2.2 Identification of unknown taxa

The intermediate types in the offspring of *Cerasus* interspecific hybrids are hard to accurately classified by using morphological markers, SNP marker-based genotyping technology is a reliable method to solve this problem. The morphological characteristics of Sangzhi No. 1 were similar to those of *C. subhirtella*, and its genetic composition was exactly the same as that of Longshan No. 1 (*C. subhirtella*), *C. spachiana* ‘Komatsu-otome’, *C. spachiana* ‘Plena Rosea’, and *C. spachiana* ‘Pendula Rosea’. In addition, the phylogenetic results show that it was closely related to Longshan No. 1. Based on the above results, it is speculated that the main parent of Sangzhi No. 1 may be *C. subhirtella*. The morphological characteristics of Xinhua No. 13 were similar to those of *C. serrulata*, and it has the genetic components of *C. serrulata* and *C. campanulata*, with the proportion of 56.2% and 43.8%, respectively. The phylogenetic results showed that Xinhua No. 13 was most closely related to *C. ‘Youkou’, which is the hybrid offspring of *C. × yedoensis* ‘Amagi Yoshino’ and *C. campanulata* (Oba et al., 2007). It is speculated that the main parents of Xinhua No. 13 may be *C. serrulata* and *C. campanulata*. The morphological characteristics of Zhejiang No. 1 and Zhangjiajie No. 1 were similar to those of *C. discoidea*, and their genetic components were exactly the same as those of *C. pseudocerasus*, *C. conradinae*, and *C. discoidea*. The phylogenetic result showed that they are clustered with *C. discoidea*. It is speculated that the main parent of Zhejiang No. 1 and Zhangjiajie No. 1 may be *C. discoidea*. The morphological characteristics of Fujian No. 1 were similar to those of *C. campanulata*, and it has the genetic components of *C. campanulata*, *C. pseudocerasus*, and *C. clarofolia*, with the proportion of 51.0%, 37.4%, and 11.6%, respectively. The phylogenetic result showed that it was clustered with *C. campanulata*, *C. yunnanensis*, and *C. dielsiana* together. It is speculated that the main parent of Fujian No. 1 may be *C. campanulata*. The morphological characteristics of Fujian No. 2 were similar to the cultivar *C. × yedoensis* ‘America’, and its genetic components are exactly the same as those of *C. × yedoensis* ‘Somei-yoshino’, and they were clustered together. *C. × yedoensis* ‘America’ was bred from the seed progeny of *C. × yedoensis* ‘Somei-yoshino’ (Oba et al., 2007). Therefore, we speculate that Fujian No. 2 may be the cultivar *C. × yedoensis* ‘America’. The morphological characteristics of Fujian No. 3 were similar to those of *C. conradinae*, and it has the genetic components of *C. conradinae* and *C. campanulata*, with the proportion of 82.5% and 17.5%, respectively. The phylogenetic result showed that it was clustered into a single line with *C. conradinae*, *C. pseudocerasus*, and *C. discoidea*. It is speculated that the main parent of Fujian No. 3 may be *C. conradinae*. Understanding the genetic background of these unknown taxa can provide a theoretical basis for the classification and breeding of flowering cherry.

2.3 Genetic components of cultivars

There are more than 200 cultivars of flowering cherry are commonly grown in the world (Kato et al., 2012), most cultivars belong to a species complex with ten basic diploid founders, *C. apetala*, *C. campanulata*, *C. incisa*, *C.
jamasakura, C. leveilleana, C. maximowiczii, C. nipponica, C. sargentii, C. spachiana and C. speciosa (Shirasawa et al., 2019). Among the 16 cultivars of flowering cherry in this study, the genetic background of seven cultivars is clear, as follows: C. × yedoensis ‘Somei-yoshino’ (C. spachiana × C. speciosa), C. ‘Accolade’ (C. spachiana × C. sargentii), C. × subhirtella ‘Yaebeni-higan’ (C. incisa × C. spachiana), C. × subhirtella ‘koshinensis’ (C. spachiana × C. sargentii), C. × yedoensis ‘Yoshun’ (C. spachiana × C. sargentii), C. ‘Youkou’ (C. × yedoensis × C. campanulata) and C. × subhirtella ‘Autumnalis’ [(C. spachiana × C. incisa) × C. subhirtella] (Oba et al., 2007; Katsuki and Iketani, 2016); The female parent of C. spachiana ‘Jindai-akebono’ is C. spachiana, its male parent is not clear; C. spachiana ‘Komatsu-otome’, C. spachiana ‘Plena Rosea’, C. spachiana ‘Pendula Rosea’ and C. spachiana ‘Ujou-shidare’ are C. spachiana complex (Oba et al., 2007). In this study, the genetic components of each cultivars were revealed by genetic structure analysis. C. ‘Accolade’ and C. × subhirtella ‘Autumnalis’ contained the components of C. spachiana, C. serrulata and C. × yedoensis ‘Somei-yoshino’, with the proportion of 36.6%, 43.5%, 19.9% and 53.0%, 36.6%, 10.4%, respectively; C. × subhirtella ‘koshinensis’, C. spachiana ‘Ujou-shidare’ and C. × subhirtella ‘Yaebeni-higan’ all contained the components of C. spachiana and C. serrulata, with the proportion of 75.4%, 24.6%; 84.5%, 15.5% and 70.6%, 29.4%; C. spachiana ‘Jindai-akebono’ contained the components of C. spachiana and C. × yedoensis ‘Somei-yoshino’, with the proportion of 34.4% and 65.5%; C. ‘Youkou’ contained the components of C. serrulata and C. campanulata, with the proportion of 47.0% and 53.0%. The genetic background of the above cultivars was consistent with the previous reports (Oba et al., 2007). The components of C. spachiana ‘Komatsu-otome’, C. spachiana ‘Plena Rosea’ and C. spachiana ‘Pendula Rosea’ were exactly the same as those of C. spachiana, this is consistent with the fact that they all belong to the C. spachiana complex (Oba et al., 2007). Understanding the genetic components and proportion of these flowering cherry cultivars or germplasms can provide reference for identification of cultivars. The plant materials in this study did not contain all the original species, it leads to the exact parents could not be clearly determined.

3 Materials and Methods

3.1 Plant materials

The 47 flowering cherry germplasms in this study were introduced and collected by Hunan Forest Botanical Garden (Table 2). The fresh leaves were collected and frozen in liquid nitrogen, then stored in - 80°C ultra-low temperature refrigerator for further study.

3.2 Extraction of genomic DNA

Total DNA was extracted by plant DNA Extraction Kit (Takara, Dalian, China) according to the instructions. The qualified samples were used for construction of library.

3.3 Library construction and sequencing

Library construction protocol was as follows: (i) Genomic DNA was digested with a restriction enzyme Taq I and the P1 adapter was ligated to the fragments, the P1 adapter contains a forward amplification primer site, an Illumina sequencing primer site, and a barcode; (ii) Adapter-ligated fragments were combined, sheared and (iii) ligated to a second adapter P2, a divergent ‘Y’ adapter, containing the reverse complement of the reverse amplification primer site preventing amplification of genomic fragments lacking a P1 adapter; (iv) RAD tags, which have a P1 adapter, will be selectively and robustly enriched. The library for Illumina sequencing was constructed from 200 ng of each DNA sample. All libraries were sequenced using Illumina HiSeq™ at Shanghai Major Biological Medicine Technology Co., Ltd. Oryza sativa was used as the control in the experiment to evaluate the accuracy of library construction. The filtering of low-quality sequences was as based on three criteria: (i) remove the adapter-contained sequences; (ii) remove the reads with 10% N content; (iii) remove the adapter and the fragments with length less than 25 bp.

3.4 Single nucleotide polymorphism (SNP) calling

The Burrows-Wheeler Aligner (http://manpages.ubuntu.com/manpages/bionic/man1/bwa.1.html) was applied for sequence alignment between the individual reads and the C. avium genome sequence (https://www.rosaceae.org/
species/prunus_avium/genome_v1.0.a1), the Stacks software package (http://catchenlab.life.illinois.edu/stacks/) was used to detect SNP loci, and the vcftools tool of SAMtools (https://github.com/samtools/samtools/releases/) was used to filter out SNP loci. The filtering of SNP loci was based on four criteria: (i) an average sequence depth of ≥ ten-fold; (ii) select the non-repeated reads that can be mapped to the reference genome; (iii) the allele with frequency of 40%~60%; (iv) distance between adjacent SNPs is more than 50 kb.

3.5 Genetic structure analysis
We reconstructed a phylogeny employing the GTRGAMMA model and 1000 bootstrap (BS) replicates under the maximum-likelihood (ML) inference in RAxML (https://cme.h-its.org/exelixis/web/software/raxml/index.html). The number and proportion of genetic components of each sample were calculated by Bayesian algorithm of admixture software (https://wikis.utexas.edu/display/bioteam/Admixture). The criteria of maximum value of ΔK was used for detecting the objective number of clusters of individuals (Evanno et al., 2005). GCTA software (https://cnsgenomics.com/software/gcta/#PCA) was used for principal component analysis.

3.6 Genetic diversity analysis
The nucleotide diversity (π value) was estimated by PopGenome software (https://www.biostars.org/p/319671/). Expected heterozygosity, observed heterozygosity, and genetic differentiation among clusters (FST) were calculated using Arlequin 3.0 (http://cmpg.unibe.ch/software/arlequin3/).

Authors’ contributions
Yan Jiawen and Li jianhui are the experimental design and executors; Yan Jiawen completed the data analysis and the writing of the first draft of the paper; Bai Wenhui, Yu Lin, Nie Dongling, Xiong Ying and Li Bohai participated in the experimental design and analysis of the experimental results; Wu Sizheng is the conceiver and person in charge of the experiment design, data analysis, thesis writing and revision. All authors read and approved the final manuscript.

Acknowledgements
This work was supported by the Key Research and Development Project of Hunan Province, China (2017NK251), and the Forestry Science and Technology Development Project of China National Forestry and Grassland Administration (KJZXSA2019037).

References
Bao W.Q., Wayun T., Du H.Y., Li T.Z., Liu H.M., Wang L., and Bai Y.E., 2018, Genetic diversity and population structure of Amygdalus mira in the Tibet plateau in China based on SSR markers, Scientia Silvae Sinicae, 54(2): 30-41 (in Chinese with English abstract)
Chang K.S., Chang C.S., Park Y.T., and Roh M., 2007, Reconsideration of the Prunus serrulata complex (Rosaceae) and related taxa in eastern Asia, Botanical Journal of the Linnean Society, 154(1): 35-54 [https://doi.org/10.1111/j.1095-8339.2007.00631.x]
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(2): 2611-2620
[https://doi.org/10.1111/j.1365-294X.2005.02553.x]
PMid:15969739
Jiang D.Y., Shen F.Q., Xu L.Y., Li Y.G., Shen X., and Liu X.H., 2018a, Genetic diversity analysis of Cerasus cultivars based on ISSR markers, Journal of Southwest Forestry university, 38(3): 21-30 (in Chinese with English abstract)
Jiang D.Y., Shen X., Wu F., Li Y.G., Shi C.G., Yang S.Z., and Liu X.H., 2018b, Genetic diversities of Cerasus spp. from Zhejiang Province based on ISSR markers, Journal of Zhejiang Forest Science Technology, 38(4): 1-7 (in Chinese with English abstract)
Kato S., Matsunotomo A., Yoshimura K., Katsuki T., Iwamoto K., Tsuda Y., Ishio S., Nakamura K., Moriwaki K., Shiroishi T., Gojobori T., and Yoshimaru H., 2012, Clone identification in Japanese flowering cherry (Prunus subgenus Cerasus) cultivars using nuclear SSR markers, Breeding Science, 62(3): 248-255 [https://doi.org/10.1270/jsbbs.62.248]
PMid:23226085 PMCID:PMC3501942
Katsuki T. and Ikatani H., 2016, Nomenclature of Tokyo cherry (Cerasus × yedoensis ‘Somei-yoshino’, Rosaceae) and allied interspecific hybrids based on recent advances in population genetics, TAXON, 65(6): 1415-1419 [https://doi.org/10.1270/taxon.65.6.1]
Oba H., Kawasaki T., and Tanaka H., eds., 2007, The cherry blossoms in Japan, Yama-kei Publishers Co., Ltd., Tokyo, Japan, pp.199-247
Shirasawa K., Esumi T., Hirakawa H., Tanaka H., Itai A., Ghelfi A., Nagasaki H., and Isobe S., 2019, Phased genome sequence of an interspecific hybrid flowering cherry, ‘Somei-Yoshino’ (Cerasus × yedoensis), DNA Research, 26(5): 1-11 [https://doi.org/10.1093/dnares/dsz016]
Tsuada 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
PMid:19340524

Wang X., Gao M., Wu L.W., Wang Y.D., Liu H.C., and Chen Y.C., 2019, A study on population genetic diversity among *Litsea populifolia* (Hems1.) gamble in mount Emei area, Journal of Plant Genetic Resources, 20(2): 359-369 (in Chinese with English abstract)

Willing E.M., Dreyer C., and van Oosterhout C., 2012, Estimates of genetic differentiation measured by *F*<sub>ST</sub> do not necessarily require large sample sizes when using many SNP markers, PLoS One, 7(8): e42649
https://doi.org/10.1371/journal.pone.0042649
PMid:22905157 PMCID:PMC3419229

Yang Y.X., Zhi L.Q., Jia Y., Zhong Q.Y., Liu Z.L., Yue M., and Li Z.H., 2019, Nucleotide diversity and demographic history of *Pinus bungeana*, an endangered conifer species endemic in China, Journal of Systematics and Evolution, 58(3): 282-294
https://doi.org/10.1111/jse.12546

Yi X.G., Chen J., You L.X., Cong R., Wang H.C., Duan Y.F., and Wang X.R., 2018, Genetic diversity of *Cerasus serrulata* populations assessed by SSR markers, Journal of Nanjing Forestry University (Natural Sciences Edition), 42(5): 25-31 (in Chinese with English abstract)

Yin M.H., Wang Q., Zhang H.L., Cai X.H., Xu C.Q., Chen F.L., Liu S.Y., Zhang Q.W., Cai H., and Chen R.H., 2020, Whole genome Re-sequencing analysis of alpine potato and local farm potato in Huaiyu Mountain under high altitude habitats, Genomics and Applied Biology, 39(3): 1198-1207 (in Chinese with English abstract)

Zhou C.C., Yang D.M., Li S.K., Wan J.Y., Rong J.D., and Zheng Y.S., 2020, Genetic relationship analysis of 18 *Cerasus* materials based on inter-simple sequence repeat molecular marker, Journal of Forest and Environment, 40(1): 46-53 (in Chinese with English abstract)

Zhou L., Duan Y., Wen B., Ma Y.C., Zhu X.J., Wang Y.H., and Fang W.P., 2018, Research and application of SNP molecular markers in ligneous plants genetics and breeding, Subtropical Plant Science, 47(2): 187-193 (in Chinese with English abstract)

Zhu H., Yi X.G., Zhu S.X., Wang H.C., Duan Y.F., and Wang X.R., 2018a, Analysis on relationship and taxonomic status of some species in Subg. *Cerasus* Koehne with chloroplast DNA *atpB-rbcL* fragment, Bulletin of Botanical Research, 38(6): 820-827 (in Chinese with English abstract)

Zhu H., Yi X.G., Zhu S.X., Wang H.C., and Wang X.R., 2018b, Application of the molecular marker technology to *Cerasus* Mill. (Rosaceae), World Forestry Research, 31(6): 16-24 (in Chinese with English abstract)

Zhu H., Yi X.G., Li Y.F., Zhu S.X., Li M., Duan Y.F., and Wang X.R., 2019, Phylogeography and population genetic structure of flowering cherry species *Cerasus sieboldiana* in subtropical China, Systematics and Biodiversity, 17(6): 622-633
https://doi.org/10.1080/14772200.2019.1670750