Genetic Diversity Analysis of Chinese Plums Based on Whole-genome Resequencing

Xiao Wei[1,#], Fei Shen[2,#], Qiuping Zhang[1,*], Ning Liu, Yuping Zhang, Ming Xu, Yujun Zhang, Xiaoxue Ma, Weisheng Liu[1,*]

1 *Liaoning Institute of Pomology, Yingkou 115009*

2 *Beijing Agro-biotechnology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100193*

[#] Authors that contributed equally to this work.

[*] Corresponding author: Qiuping Zhang (e-mail: lbzhangqiuping2@163.com) and Weisheng Liu (e-mail: wsliulaas@163.com)

Abstract

Plum is gaining importance due to their extensive genetic diversity and nutritional attributes beneficial for human health. Single-nucleotide polymorphisms (SNPs) are the most abundant form of genomic polymorphisms and are widely used in population genetics research. Here, we used the high-quality SNPs generated from high-depth whole-genome resequencing of 67 *Prunus* accessions to evaluate the genome-level diversity and population structure. The 67 plum accessions were classified into four groups, the southern cultivar group (SCG), the northern cultivar group (NCG), the foreign cultivar group (FG), and the mixed cultivar group (MG), basing on the results of the phylogenetic analysis, the principal component analysis, and the population structure analysis. Several cultivars from South China were observed distributing in the other three groups, and gene flow was detected between the SCG and the FG. For the two groups mainly consisted of plum landraces, the genetic diversity indices including the private allele number, the observed heterozygosity, the expected heterozygosity, and the nucleotide diversity of the SCG were higher than those of the NCG. We speculated that, for SCG, the genetic background was relatively broader, the genetic exchange was more frequent with other groups, and the possible origin center of Chinese plum was at the Yangtze River Basin in South China. This study provided detailed genetic variation features of plum cultivars, and preliminarily dissected their genetic diversity and population structure, laying a foundation for breeders to use diverse germplasm and allelic variants towards developing improved plum varieties.

Keywords: *Prunus salicina* Lindl., whole-genome resequencing, population structure, genetic introgression

1. Background

Plum is one of the most important fruit crops in the world, possessing extensive genetic diversity and high economic value. The largest plum producer is China, with an annual production of 6,801,187 metric tons in 2018, accounting for 53.9% of the world's total (FAOSTAT, 2018). Within the genus, *Prunus*, Chinese plum (*Prunus salicina* L., 2n = 2x = 16), also known as Japanese plum, is widely grown for fresh market consumption and the canning industry, which includes both pure Chinese plum and its hybrids with other diploid plum species, such as *Prunus simonii* Carr., *Prunus cerasifera* Ehrh., *Prunus americana* Marsh., and others (Liu et al., 2007).

Chinese plum originates from the Yangtze River Basin, and now there are abundant plum germplasm resources in China (Hartmann and Neumulle, 2009). With its long growing history and extensive geographical distribution, there are more than 800 indigenous plum cultivars in China derived from *Prunus salicina* L. alone. Over 600 of them are now preserved at National Germplasm Repository for Plums and Apricots (NGRPA) located in Xiongyue, Liaoning Province, China (Zhang, 1990; Liu et al., 2006). Yu et al. (2011) carried out comprehensive phenotyping of 405 Chinese plum cultivars and its hybrids from NGRPA, totally 32 morphological and agronomic characters were investigated, of which their coefficients of variations (CVs) were calculated ranging from 14.85% to 47.09%, suggesting that the genetic variability of Chinese plums distributed widely.

Genetic variability is the prerequisite for any plant breeding program. Learning the extent and structure of genetic variation in germplasm collections is a crucial step for

61 the efficient conservation and utilization of biodiversity in cultivated crops, and using
62 diverse plum resources to broaden the genetic base of worldwide plum cultivars is the
63 critical objective for plum breeders (Liu et al., 2007; Urrestarazu et al., 2018).
64 Previous efforts have been made to understand the genetic diversity of plum better.
65 Zhang and Zhang (1998) took the lead in collecting plum germplasm resources and
66 evaluating their genetic diversity on the basis of morphological traits and isozyme
67 polymorphisms, whereas, morphological traits were highly susceptible to
68 environmental factors, so the estimates of genetic diversity were not precise; while
69 isozymes had a low degree of polymorphism and hence were not efficient enough for
70 the characterization of germplasm genetic diversity (Khush, 2002). Overcoming the
71 above-mentioned problems, DNA-based markers became widely applied in plum
72 genetic diversity analysis, including random amplified polymorphic DNAs (RAPDs)
73 (Liu et al., 2006; Ben et al., 2015), simple sequence repeats (SSRs) (Pop et al., 2018;
74 Zhang et al., 2018; Acuña et al., 2019), inter-simple sequence repeats (ISSRs) (Liu et
75 al., 2007; Wu et al., 2019). In addition, benefiting from the advances in
76 next-generation sequencing (NGS) technologies, genotyping by sequencing (GBS),
77 provides a great wealth of information that makes it possible to identify thousands of
78 single nucleotide polymorphisms (SNPs), which, after adequate filtering, allow us to
79 carry out detailed genetic diversity studies (Elshire et al., 2011; Urrestarazu et al.,
80 2018; Salazar et al., 2019; Zhebentyayeva et al., 2019).

81 By conducting DNA-based marker assessments, Chinese plums were generally
82 classified into two major groups, the southern cultivar group (SCG) and the northern

cultivar group (NCG) (Qiao, 2003; Liu et al., 2005; Wei et al., 2019). The *P. salicina* system in South China was rather primitive than that in North China basing on a palynology study, and the spread of *P. salicina* was from the south to the north (Guo, 2006). However, simply cultivar clustering works are inadequate for the integrated study of genetic diversity of Chinese plum. As described in peach, large scale SNP data-based diversity analysis boosted the deciphering of the evolution and domestication of peach germplasm resources (Cao et al., 2014; Yu et al., 2018; Li et al., 2019).

In the present study, we used for the first time SNP markers generated from high depth whole-genome resequencing data to elucidate the pattern of genetic diversity, structuration, and domestication of a diverse *P. salicina* collection. An in-depth understanding of their genetic relationships could benefit plum germplasm conservation and utilization, and lead to new cultivar improvement.

2. Materials and Methods

2.1 Plant materials

A diverse collection consisting of 67 *Prunus* spp. accessions, including 65 *P. salicina* accessions and two *P. simonii* accessions, were selected for the whole-genome resequencing study (Table S1). All these trees were held at NGRPA located in Xiongyue county, Liaoning Province, China (40°18′ N, 122°16′ E) at a density of 3.0 × 4.0 m, trained to the open vase system with three or four main branches, and were maintained under conventional management and pest control operations.

105 *2.2 DNA extraction, library preparation, and sequencing*

106 Genomic DNAs were isolated from the young and healthy leaf samples of the 67
107 accessions using a modified CTAB protocol (Doyle and Doyle, 1990). The quality
108 and integrity of the DNAs were examined using a NanoDrop® spectrophotometer
109 (ND-1000, Thermo Fisher Scientific Inc., USA) followed by electrophoresis on 1%
110 agarose gels. Quantification of the DNA samples was assessed with Qubit™ (2.0
111 Fluorometer, Invitrogen, Carlsbad, CA, USA).

112 High molecular weight DNA aliquots with 230/260 and 260/280 ratios ranging
113 between 1.8 – 2.0 and 1.8 – 2.2, respectively, were then sent to BGI (Shenzhen, China)
114 for library construction and sequencing. The insert-size of the libraries was 500 bp,
115 and the length of pair-end reads was 150 bp. All the libraries were sequenced using
116 the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA).

117 *2.3 Reads mapping, SNP calling, and SNP annotation*

118 The qualified paired-end reads of each accession were aligned against the peach
119 reference genome v2.0 (Verde et al., 2013) using BWA v0.7.12-r1039 with the
120 parameters mem -t 4 -k 32 –M (Li and Durbin, 2010), and SNPs were identified using
121 SAMtools v1.4 (Li et al., 2009). Low-quality SNPs were filtered out by minimum
122 minor allele frequency (mnMAF < 0.01) and missing data per site (MDpS > 10%),
123 and finally converted into Variant Call Format file (VCF). Gene-based SNP
124 annotation was performed using the ANNOVAR package v2018-04-16 (Wang et al.,
125 2010). Based on the reference genome annotation, SNPs were categorized as
126 occurring in exonic regions (overlapping with a coding exon), intronic regions

127 (overlapping with an intron), upstream and downstream regions (within a 1 kb region
128 upstream or downstream from the transcription start site), or intergenic regions.
129 Additionally, the heterozygosity was calculated using VCFtools v0.1.14 (Danecek et
130 al., 2011).

131 *2.4 Population structure analysis*

132 TreeBeST v1.9.2 software was used to calculate the distance matrix (Vilella et al.,
133 2009). RAxML v8.2.12 was used to construct the ML phylogenetic tree, and 1000
134 bootstrap replicates were used (Stamatakis et al., 2014). The obtained phylogenetic
135 tree was visualized using MEGA v5.05 (Tamura et al., 2011).

136 The principal component analysis was performed using the PLINK v1.07 software
137 with default parameters (Purcell et al., 2007).

138 We used ADMIXTURE v1.23 to infer the population structure (Alexander et al.,
139 2009). To identify the best genetic clusters *K*, cross-validation error was tested for
140 each *K* value from 2 to 10. The termination criterion was 1e-6 (stopping when the
141 log-likelihood increased by less than 1e-6 between iterations).

142 *2.5 Linkage disequilibrium (LD) analysis*

143 Linkage disequilibrium (LD) was calculated using SNPs with MAF greater than 0.05
144 through PLINK v1.07 software with the following setting: --file --r2 --ld-window
145 99999 --ld-window-kb 200 --out. LD decay was calculated based on the squared
146 correlation coefficient (r^2) values between two SNPs and the physical distance
147 between the two SNPs (Purcell et al., 2007).

148

2.6 Gene flow analysis

The TreeMix v1.13 software was used to evaluate the gene flow among different groups with the parameters -se-bootstrap-k 1000 -m, where the number (-m) varied from one to three (Pickrell and Pritchard, 2012).

2.7 Genetic diversity and differentiation analysis

To evaluate the genetic diversity and differentiation, we used a 100-kb sliding window with a step size of 10 kb to calculated the number of private alleles (A_P), the observed heterozygosity (Ho), the expected heterozygosity (He), the nucleotide diversity (π), the Tajima's D value, Wright's F-statistic (F_{IS}), and the population-differentiation statistics (Fixation index, F_{ST}) using VCFtools v0.1.14 (Danecek et al., 2011).

3. Results

3.1 Markers development based on whole-genome resequencing

We collected 67 plum accessions representing different geographic and morphological characteristics (Table S1). The 67 plum accessions were sequenced using the Illumina HiSeq 2500 platform with a sequence depth higher than 20×; a totally 462.6 Gb clean data were retained with an average of 6.9 Gb for each accession after filtering of the low-quality reads. The quality of the sequencing data was high with a Q30 > 85%, and a GC content slightly fluctuated around 38.0%, and the average mapping rate was 87.49% (Table S1). We mapped the reads to the peach genome due to the relative phylogenetic relationship, and nearly 87.49% of reads could be mapped. We constructed SNPs calling based on the unique mapped reads, and a total of 16,600,033 SNPs and 2,107,664 INDELs were identified across the 67 accessions using

SAMtools v1.4 software, and 14,549,234 highly informative and well well-distributed SNPs were obtained after filtering, which ensured the accuracy and reliability of subsequent genetic diversity and population structure analyses (Figure 1). Approximately 43.81% of the SNPs were located in intergenic regions, and 11.88% were located in coding regions. The non-synonymous to synonymous substitution ratio (dN/dS) for the SNPs in the coding regions was 15.92. Transitions were found in a proportion of 59.58% (10315764/16600033), with a transition/transversion ratio (Ts/Tv ratio) of 1.47 (Figure 1, Table 1).

3.2 Phylogenetic analysis

We constructed a phylogenetic analysis based on the high-quality and high-density markers across the genome. A phylogenetic tree of the 67 accessions was been built for better visualization of sample distribution and relationships (Figure 2a). The accessions could be divided into four main groups revealed by the phylogenetic tree. The plum cultivars from Sichuan, Guizhou, Yunnan, Guangdong, Guangxi, Zhejiang, and Fujian Province clustered as the Southern cultivar group (SCG). The Northern cultivar group (NCG) comprised plum cultivars mainly from Hebei, Henan, Shandong, Shaanxi Province, and the south part of Liaoning Province. The foreign cultivar group (FG) included plum cultivars from the USA and Japan. The mixed cultivar group (MG) consisted of newly selected cultivars from the NGRPA and several cultivars from SCG and FG. Notably, we could find that there were several cultivars from the south part of China distributing in the other three groups.

3.3 Population structure analysis

Population genetic structure analysis was performed based on high-quality SNPs. We used different population *K* values in this study to calculate the population structure (Figure 2b). When the *K* value was equal to 2~4, the NCG exhibited a consistent genetic constitution with SCG, suggesting the NCG was derived from SCG and also underwent an environment/human selection to some extent. We could also find that the population structure of SCG was complicated, mostly when the *K* value was higher, suggesting the higher genetic diversity of SCG, which was consistent with the π value analysis (Figure 3b). For FG, we could find that the population structure was relatively independent comparing with SCG and others. Thus, we proposed that the existence of intra-species hybrid in the group and led to a high π value (Figure 3b).

3.4 Principal component analysis

We performed principal component analysis (PCA) on the 67 plum accessions to give a better understanding of the relationships between different cultivar groups (Figure 2c). Based on the two-component, the SCG exhibited a relatively close relationship with NCG, which was consistent with the phylogenetic analysis and population structure analysis. NCG could be considered as a more independent subgroup of SCG and may be derived from a specific ecotype. Samples from SCG showed some degree of correlation, suggesting the close relationship between SCG and others.

3.5 Genetic diversity analysis

As shown in Table 2, the expected heterozygosity (He) of the *Prunus* populations

varied between 0.234 and 0.304; the observed heterozygosity (Ho) of the *Prunus* populations ranged between 0.328 and 0.429; Wright's F-statistic (F_{IS}) of the *Prunus* populations varied between –0.241 and –0.146; the number of private alleles (A_P) in the *Prunus* populations varied between 546 and 31,285; the nucleotide diversity (π) ranged between 0.00358 and 0.00467. The mean nucleotide variation of *P. salicina* was higher than that of other perennial crops, such as peach (π = 0.0015) (Verde et al., 2013), cassava (π = 0.0026) (Kawuki et al., 2009), and apricot (π = 0.0027) (Li et al., 2020), but lower than that of date palm (π = 0.0092) (Hazzouri et al., 2015). The Tajima's D values of the four groups were all tested positive (1.002 – 1.497) and significantly different from zero. Thus, the null hypothesis of neutral evolution could be rejected. As shown in Figure 3a, the LD decay rate was fastest in the SCG and was lowest in the FG, suggesting a lower frequency of genetic recombination in the FG, as they included artificially hybridized cultivars.

3.6 Genetic differentiation analysis

The F_{ST} between the four groups varied between 0.0485 and 0.1303 (Fig. 3). The F_{ST} value of MG-FG, MG-SCG, and SCG-NCG pairs was relatively lower among all the group-pairs analyzed, indicating that the genetic differences within populations were higher than those between populations and the possible genetic exchange between populations. The F_{ST} value between the NCG and FG groups were the largest among all the groups analyzed, which may have been due to geographical isolation, low gene flow between populations, and significant genetic differences.

3.7 Gene flow analysis and inferred evolutionary path

We analyzed the gene flow between the four geographic groups (Figure 4b, 4c). Allowing one or two migration events (m = 1 or 2), we observed that gene flow occurred between the SCG and FG accessions, which likely reflected their many shared genomic components due to hybridization in their domestication and breeding histories. Considering the geographical locations of SCG and NCG, though these two ecological groups were close to each other, we found no gene flow between them, indicating a relatively independent domestication process.

4. Discussion

SNPs have become tremendously important as markers for genetics research in plants because they have been found to be in high frequency, display a lower mutation rate compared to SSR-based markers, and they are uniformly distributed across the genome (Dirlewanger et al., 2002). NGS is a kind of low-cost technologies to discover large numbers of SNPs for extensive genetic studies, including within-species diversity analysis, linkage map construction, and genome-wide association studies (GWAS), which is allowing significant advances in plant genetics and breeding (Cao et al., 2016; Yu et al., 2018; Li et al., 2019; Pinosio et al., 2020). The genus *Prunus* has shown conserved intraspecific and intragenic collinearity in the *Rosaceae* family, and peach has been considered as a model species for the genus *Prunus* regarding multiple genetic types of research (Arús et al., 2012; Carrasco et al., 2018; Marti et al., 2018; Zhang et al., 2020). On these bases, for the first time, we conducted high-depth (~20×) whole-genome resequencing of 67 plum accessions using the peach genome sequence as a reference, obtaining an average of 6.9 Gb

259 high-quality data for each acquisition, which ensured the accurate and comprehensive
260 population genetic analysis.

261 Based on the high-quality SNPs, the 67 plum accessions in this study were divided
262 into four groups, the SCG, NCG, FG, and MG, supported by the phylogenetic tree, the
263 population structure analysis, and the PCA results, which were partly in accordance
264 with the previously obtained results through RAPD, ISSR, and SSR marker assessing
265 (Qiao, 2003; Liu, 2005; Wei et al., 2019). We noticed that several cultivars from the
266 south part of China distributed in other three groups (Figure 2a), combining with the
267 fact that the *P. salicina* system in South China was rather primitive than that in North
268 China basing on a palynology study, we speculated that, for SCG, the genetic
269 background was relatively broader, the gene exchange was more frequent with other
270 groups, and the possible origin center of Chinese plum was at the Yangtze River Basin
271 in South China (Guo, 2006; Hartmann and Neumulle, 2009; Topp et al., 2012; Wei et
272 al., 2019). For the five newly selected cultivars from the NGRPA in MG, one of their
273 parents was from the USA (‘Blackamber’, ‘Friar’) or Japan (‘Akihime’) and the other
274 parent was native to China, and the gene flow between SCG and FG was detected
275 (Figure 4). Thus, we assumed that those cultivars in MG were genetically related. As
276 mentioned in other studies, most of the Chinese plum cultivars from Japan and the
277 improved Chinese plum hybrids from the USA were found distributed across Chinese
278 indigenous plum group through dendrogram analysis (Liu et al., 2006; Liu et al., 2007;
279 Wei et al., 2019), which was according with the *P. salicina*-predominant genetic
280 components of improved Chinese plum hybrids (Faust and Suranyi, 1999), and in

281 agreement with the fact that China-originated plum cultivars were initially introduced
282 to Japan, and then imported to the USA in 1870 from Japan by Luther Burbank
283 (Hartmann and Neumulle, 2009).

284 Southwest China was speculated as a primitive domestication center of plum as
285 there were wild plum populations discovered in Yunnan, Sichuan, and Guizhou
286 Province (Wei et al., 2019). Suggesting by the high Shannon index of diversity, the
287 high value of average sufficient allele, and the high expected heterozygosity, the
288 Chinese plums cultivated in southwest China were tested with high genetic diversity
289 using ISSR and SSR markers (Chen et al., 2014; Wei et al., 2019). Additionally, by
290 analyzing the phenotyping variation of 405 plum accessions distributed in different
291 geographical regions in China, the fruit weight of cultivars in South China exhibited
292 the most significant coefficient of variation (Yu et al., 2011). The π, Ho, and He of the
293 SCG calculated with SNPs in our study were higher than those of the NCG,
294 corresponding with the higher genetic diversity of the SCG; these genetic diversity
295 indices of the SCG were lower than that of the FG and MG, presumably due to the
296 artificial hybridized cultivars in the FG and MG.

297 Generally, the genetic differentiation between populations is considered moderate
298 when the F_{ST} value is higher than 0.05 and is considered highly differentiated when
299 the F_{ST} value is higher than 0.15 (Xu, 2009). In our study, the NCG-FG pair possessed
300 the highest F_{ST} value, following by the SCG-FG couple. In contrast, the F_{ST} value of
301 the MG-FG was significantly lower as there were five newly selected plum cultivars
302 by NGRPA in the MG with foreign lineage from the FG. Human intervention via the

artificial selection of favorable phenotypic traits to enhance production and improve desirable agronomic traits can both reduce the levels of genetic variability and skew allele frequencies (Tajima, 1989).

5. Conclusions

In the current study, using high-quality SNPs generated from high-depth whole-genome resequencing of 67 *Prunus* accessions, we provided detailed genetic variation features of plum cultivars, and preliminarily dissect their genetic diversity and population structure, laying a foundation for breeders to use diverse germplasm and allelic variants towards developing improved plum varieties.

Acknowledgments

This work was supported by National Natural Science Foundation of China (31401826), the Program of Conservation and Utilization of Crop Germplasm Resources (2018NWB003) and National Crop Germplasm Resources Platform of China (NICGR2018-056).

Data availability

All of the raw reads of the plum accessions generated in this study have been deposited in the public database of National Center of Biotechnology Information under PRJNA659814.

Authors' contributions

ZQP and LWS conceived and designed the experiments. SF and WX performed the experiments. WX, SF and ZQP analyzed the data and wrote the paper. LN, ZYP, XM, MXX and ZYJ collected samples and managed the materials. LWS revised the manuscript. All authors have read and approved the final manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

Acuña C V., Rivas JG, Brambilla SM, et al (2019) Characterization of Genetic Diversity in Accessions of *Prunus salicina* Lindl: Keeping fruit flesh color ideotype while adapting to water stressed environments. Agronomy 9:1–12. https://doi.org/10.3390/agronomy9090487

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664. https://doi.org/10.1101/gr.094052.109

Arús P, Verde I, Sosinski B, et al (2012) The peach genome. Tree Genet Genomes 8:531–547. https://doi.org/10.1007/s11295-012-0493-8

Ben Tamarzizt H, Ben Mustapha S, Baraket G, et al (2015) Assessment of genetic diversity and relationships among wild and cultivated Tunisian plums (*Prunus* spp.) using random amplified microsatellite polymorphism markers. Genet Mol Res 14:1942–1956. https://doi.org/10.4238/2015.March.20.4

Cao K, Zheng Z, Wang L, et al (2014) Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops. Genome Biol 15:1–15. https://doi.org/10.1186/s13059-014-0415-1

Carrasco B, González M, Gebauer M, et al (2018) Construction of a highly saturated linkage map in Japanese plum (*Prunus salicina* L.) using GBS for SNP marker calling. PLoS One 13:1–14. https://doi.org/10.1371/journal.pone.0208032

Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools.

352 Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

353 Dirlewanger E, Cosson P, Tavaud M, et al (2002) Development of microsatellite
354 markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity
355 analysis in peach and sweet cherry (*Prunus avium* L.). Theor Appl Genet
356 105:127–138. https://doi.org/10.1007/s00122-002-0867-7

357 Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15.

358 Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A robust, simple
359 genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One
360 6:1–10. https://doi.org/10.1371/journal.pone.0019379

361 Faust M, Suranyi D (1999) Origin and dissemination of plum. Horticultural Reviews,
362 23, 179–231.

363 Food and Agriculture Organization of the United Nations. 2018. World production of
364 plums and sloes. 23 April. 2020. http://www.fao.org/faostat/zh/#data/QC

365 Guo Z (2006) Collection and utilization of Prunus salicina germplasm resource in
366 South China. Nanjing: Nanjing Agricultural University. *In Chinese*

367 Hartmann W, Neumulle M (2009) Plum breeding. In *Breeding Plantation Tree Crops:*
368 *Temperate Species*; Jain, S.M., Priyadarshan, P.M., Eds.; Springer Science +
369 Business Media: New York, NY, USA, 2009; pp. 161–231.

370 Hazzouri KM, Flowers JM, Visser HJ, et al (2015) Whole genome resequencing of
371 date palms yields insights into diversification of a fruit tree crop. Nat Commun 6:.
372 https://doi.org/10.1038/ncomms9824

373 Kawuki RS, Ferguson M, Labuschagne M, et al (2009) Identification, characterisation

374 and application of single nucleotide polymorphisms for diversity assessment in
375 cassava (*Manihot esculenta* Crantz). Mol Breed 23:669–684.
376 https://doi.org/10.1007/s11032-009-9264-0
377 Khush G (2002) Molecular genetics-plant breeder's perspective, p. 1–8. In: S.M. Jain,
378 D.S. Brar, and B.S. Ahloowalia (eds.). Molecular techniques in crop
379 improvement. Kluwer Academic Publishers, Dordrecht, The Netherlands.
380 Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler
381 transform. Bioinformatics 26:589–595.
382 https://doi.org/10.1093/bioinformatics/btp698
383 Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format
384 and SAMtools. Bioinformatics 25:2078–2079.
385 https://doi.org/10.1093/bioinformatics/btp352
386 Li R, Li Y, Fang X, et al (2009) SNP detection for massively parallel whole-genome
387 resequencing. Genome Res 19:1124–1132. https://doi.org/10.1101/gr.088013.108
388 Li W, Liu L, Wang Y, et al (2020) Genetic diversity, population structure, and
389 relationships of apricot (*Prunus*) based on restriction site-associated DNA
390 sequencing. Hortic Res 7:. https://doi.org/10.1038/s41438-020-0284-6
391 Li Y, Cao K, Zhu G, et al (2019) Genomic analyses of an extensive collection of wild
392 and cultivated accessions provide new insights into peach breeding history 06
393 Biological Sciences 0604 Genetics. Genome Biol 20:1–18.
394 https://doi.org/10.1186/s13059-019-1648-9
395 Liu W (2005) Studies on the genetic diversity among plum germplasm resources and

396 the phylogenetic relationships of main plum species. Beijing: China Agricultural
397 University. *In Chinese*
398 Liu W, Liu D, Feng C, et al (2006) Genetic diversity and phylogenetic relationships in
399 plum germplasm resources revealed by RAPD markers. J Hortic Sci Biotechnol
400 81:242–250. https://doi.org/10.1080/14620316.2006.11512057
401 Liu W, Liu D, Zhang A, et al (2007) Genetic diversity and phylogenetic relationships
402 among plum germplasm resources in China assessed with inter-simple sequence
403 repeat markers. J Am Soc Hortic Sci 132:619–628.
404 https://doi.org/10.21273/jashs.132.5.619
405 Marti AFI, Saski CA, Manganaris GA, et al (2018) Genomic sequencing of Japanese
406 plum (*Prunus salicina* Lindl.) mutants provides a new model for rosaceae fruit
407 ripening studies. Front Plant Sci 9:. https://doi.org/10.3389/fpls.2018.00021
408 Pickrell JK, Pritchard JK (2012) Inference of Population Splits and Mixtures from
409 Genome-Wide Allele Frequency Data. PLoS Genet 8:.
410 https://doi.org/10.1371/journal.pgen.1002967
411 Pinosio S, Marroni F, Zuccolo A, et al (2020) A draft genome of sweet cherry (*Prunus*
412 *avium* L.) reveals genome‐wide and local effects of domestication. Plant J 1–13.
413 https://doi.org/10.1111/tpj.14809
414 Pop R, Hârţa M, Szabo K, et al (2018) Genetic diversity and population structure of
415 plum accessions from a Romanian germplasm collection assessed by simple
416 sequence repeat (SSR) markers. Not Bot Horti Agrobot Cluj-Napoca 46:90–96.
417 https://doi.org/10.15835/nbha46110884

418 Purcell S, Neale B, Todd-Brown K, et al (2007) PLINK: A tool set for whole-genome
419 association and population-based linkage analyses. Am J Hum Genet 81:559–575.
420 https://doi.org/10.1086/519795
421 Qiao Y (2003) Establishment of RAPD, ISSR and SSR reactions system and analysis
422 of genetic diversity of Japanese plum cultivars. Nanjing: Nanjing Agricultural
423 University. *In Chinese*
424 Salazar JA, Pacheco I, Silva C, et al (2019) Development and applicability of GBS
425 approach for genomic studies in Japanese plum (*Prunus salicina* Lindl.). J Hortic
426 Sci Biotechnol 94:284–294. https://doi.org/10.1080/14620316.2018.1543559
427 Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and
428 post-analysis of large phylogenies. Bioinformatics 30:1312–1313.
429 https://doi.org/10.1093/bioinformatics/btu033
430 Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by
431 DNA polymorphism. Genetics 123:585–595.
432 Tamura K, Peterson D, Peterson N, et al (2011) MEGA5: Molecular evolutionary
433 genetics analysis using maximum likelihood, evolutionary distance, and
434 maximum parsimony methods. Mol Biol Evol 28:2731–2739.
435 https://doi.org/10.1093/molbev/msr121
436 Topp B, Russell D, Neumuller M, et al (2012) Plum//Badenes M, Byrne D. Fruit
437 Breeding. Media: Springer Science Business, pp571-621.
438 Urrestarazu J, Errea P, Miranda C, et al (2018) Genetic diversity of Spanish *Prunus*
439 *domestica* L. germplasm reveals a complex genetic structure underlying. PLoS

440 One 13:1–21. https://doi.org/10.1371/journal.pone.0195591

441 Verde I, Abbott AG, Scalabrin S, et al (2013) The high-quality draft genome of peach
442 (*Prunus persica*) identifies unique patterns of genetic diversity, domestication
443 and genome evolution. Nat Genet 45:487–494. https://doi.org/10.1038/ng.2586

444 Vilella AJ, Severin J, Ureta-Vidal A, et al (2009) EnsemblCompara GeneTrees:
445 Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res
446 19:327–335. https://doi.org/10.1101/gr.073585.107

447 Wang K, Li M, Hakonarson H (2010) ANNOVAR: Functional annotation of genetic
448 variants from high-throughput sequencing data. Nucleic Acids Res 38:1–7.
449 https://doi.org/10.1093/nar/gkq603

450 Wei X, Zhang Q, Liu N, et al (2019) Genetic diversity of the *Prunus salicina* L. from
451 different Sources and their related species. Sci. Agric. Sin. 52: 568–578
452 https://doi:10.3864/j.issn.0578-1752.2019.03.017 *In Chinese*

453 Wu W, Chen F, Yeh K, Chen J (2019) ISSR analysis of genetic diversity and structure
454 of plum varieties cultivated in Southern China. Biology (Basel) 8:1–13.
455 https://doi.org/10.3390/biology8010002

456 Xu G (2009) Plant Population Genetics. Beijing: Science Press. *In Chinese*

457 Yu X, Zhang Q, Liu W, et al (2011) Genetic diversity analysis of morphological and
458 agronomic characters of Chinese plum (*Prunus salicina* Lindl.) germplasm. J
459 Plant Genet Resour, 12(3): 402-40.
460 https://doi.org/10.13430/j.cnki.jpgr.2011.03.024 *in Chinese*

461 Yu Y, Fu J, Xu Y, et al (2018) Genome resequencing reveals the evolutionary history

of peach fruit edibility. Nat Commun 9:1–13. https://doi.org/10.1038/s41467-018-07744-3

Zhang J (1990). Report on the investigation for national plum and apricot resources. China Fruits, 4, 29–34. *In Chinese*

Zhang J, Zhou E (1998) China fruit-plant monographs, plum flora. China Forestry Press, Beijing. *In Chinese*

Zhang Q, Wei X, Liu N, et al (2020) Construction of an SNP-based high-density genetic map for Japanese plum in a Chinese population using specific length fragment sequencing. Tree Genet Genomes 16:. https://doi.org/10.1007/s11295-019-1385-y

Zhang Q ping, Wei X, Liu W sheng, et al (2018) The genetic relationship and structure of some natural interspecific hybrids in *Prunus* subgenus *Prunophora*, based on nuclear and chloroplast simple sequence repeats. Genet Resour Crop Evol 65:625–636. https://doi.org/10.1007/s10722-017-0559-4

Zhebentyayeva T, Shankar V, Scorza R, et al (2019) Genetic characterization of worldwide *Prunus domestica* (plum) germplasm using sequence-based genotyping. Hortic Res 6:12. https://doi.org/10.1038/s41438-018-0090-6

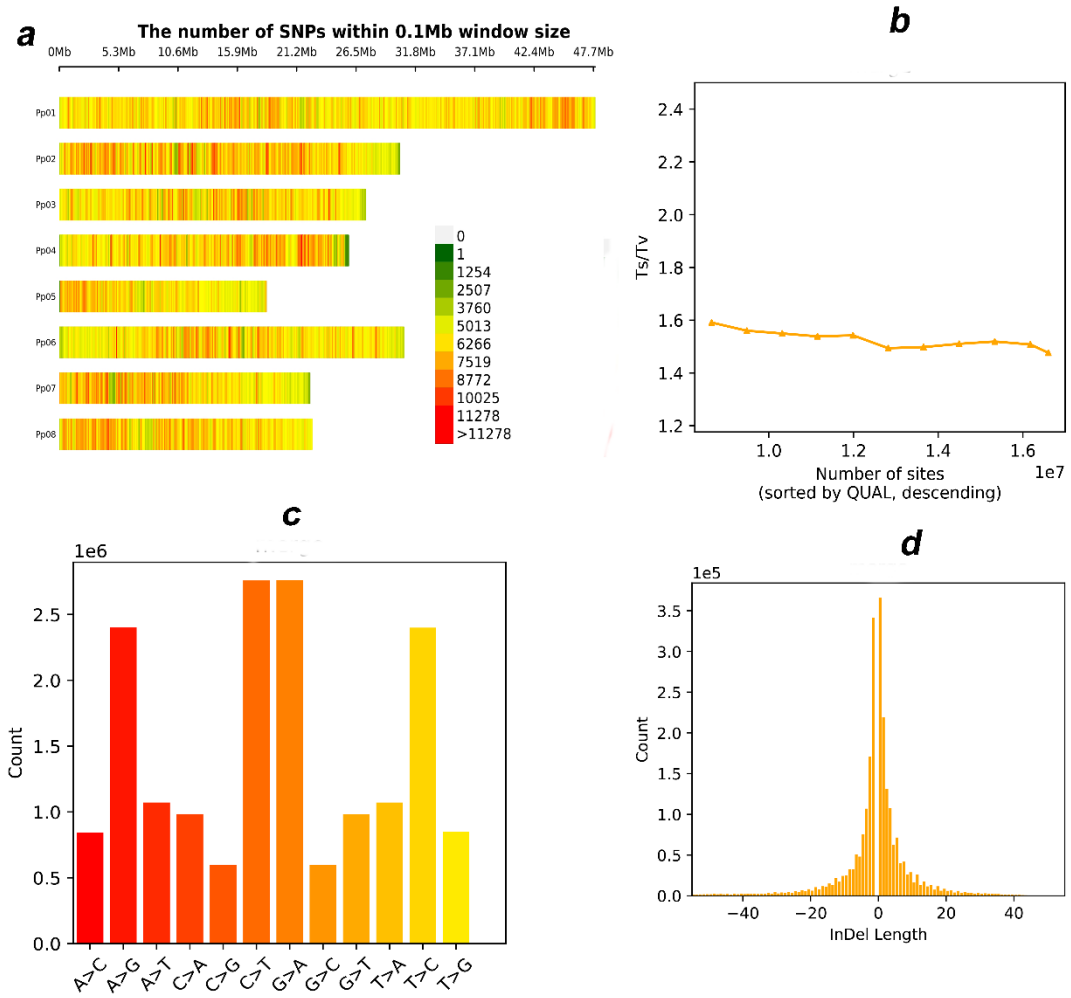

Figure 1. The statistics of markers generated from the whole-genome resequencing of the 67 plum accessions. a) The chromosome-scale SNP distribution; b) The distribution of transition/transversion ratio (Ts/Tv ratio); c) The counts of different types of transitions and transversions; d) The counts of genome-wide insertions and deletions.

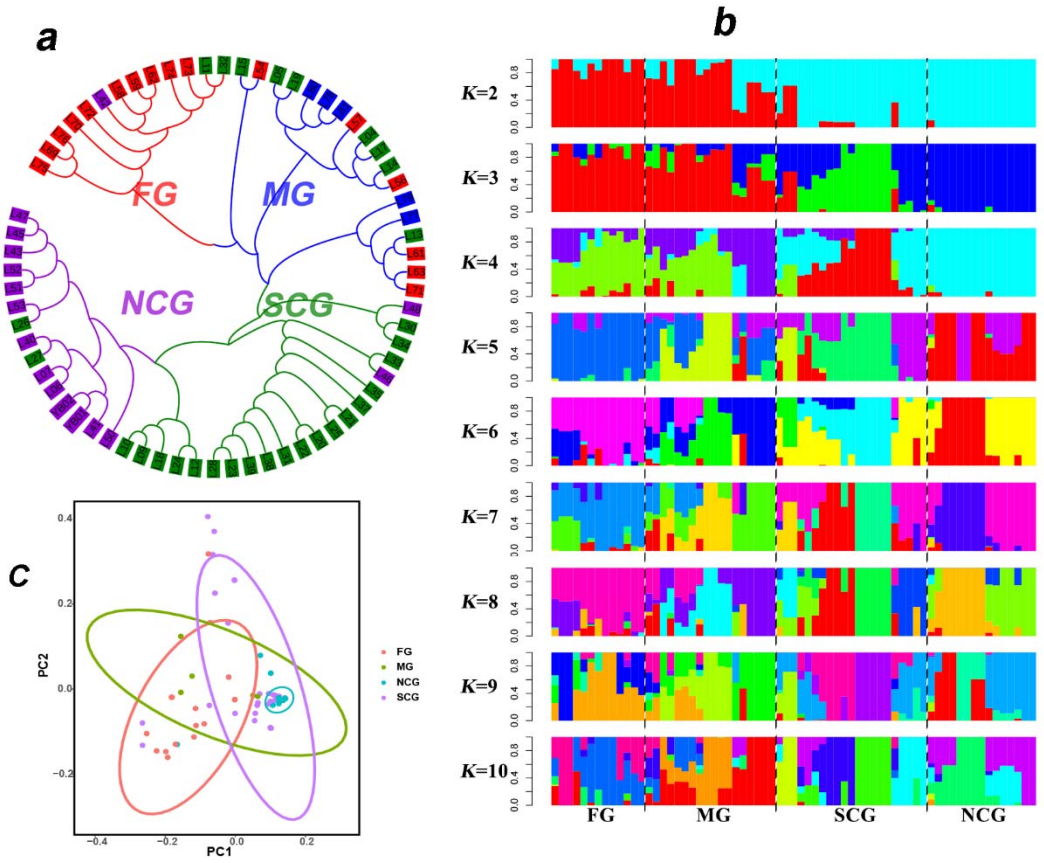

Figure 2. The population structure of the 67 plum accessions. a) Neighboring-joining phylogenetic tree constructed using SNPs at fourfold degenerate site. Each group was color coded; b) Bayesian model-based clustering of the 67 plum accessions with the number of ancestry kinship (*K*) from 2 to 10; c) Principle Component Analysis (PCA) of the 67 plum accessions. FG, the foreign cultivar group; MG, the mixed cultivar group; SCG, the southern cultivar group; NCG, the northern cultivar group.

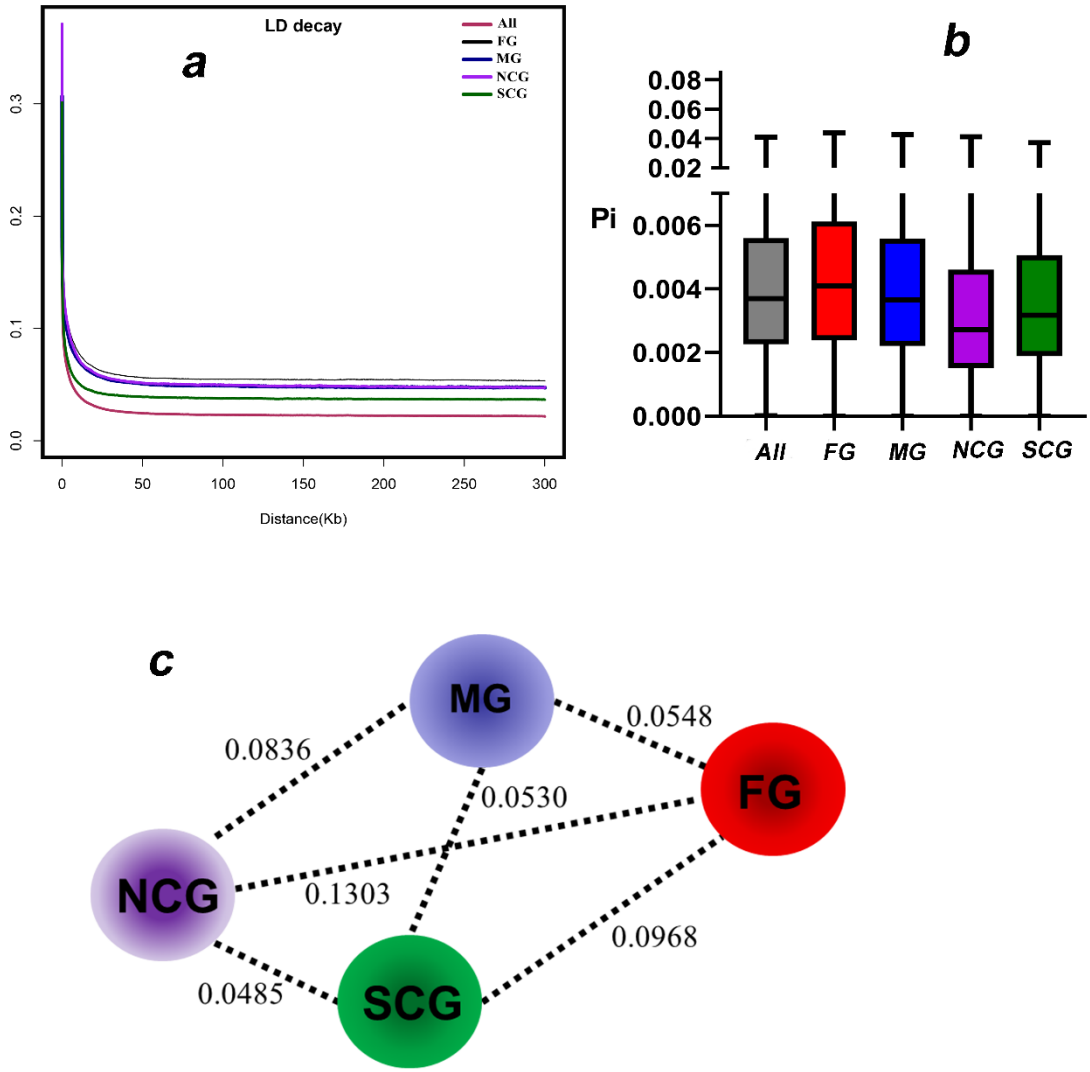

Figure 3. The genetic diversity of different groups. a) The decay of linkage disequilibrium (LD) measured as the squared correlation coefficient (r^2) by pairwise physical distance; b) The nucleotide diversity (π) of different groups; c) The genetic differentiation analysis between groups. The values between pairs indicate population divergence (F_{ST}). FG, the foreign cultivar group; MG, the mixed cultivar group; SCG, the southern cultivar group; NCG, the northern cultivar group.

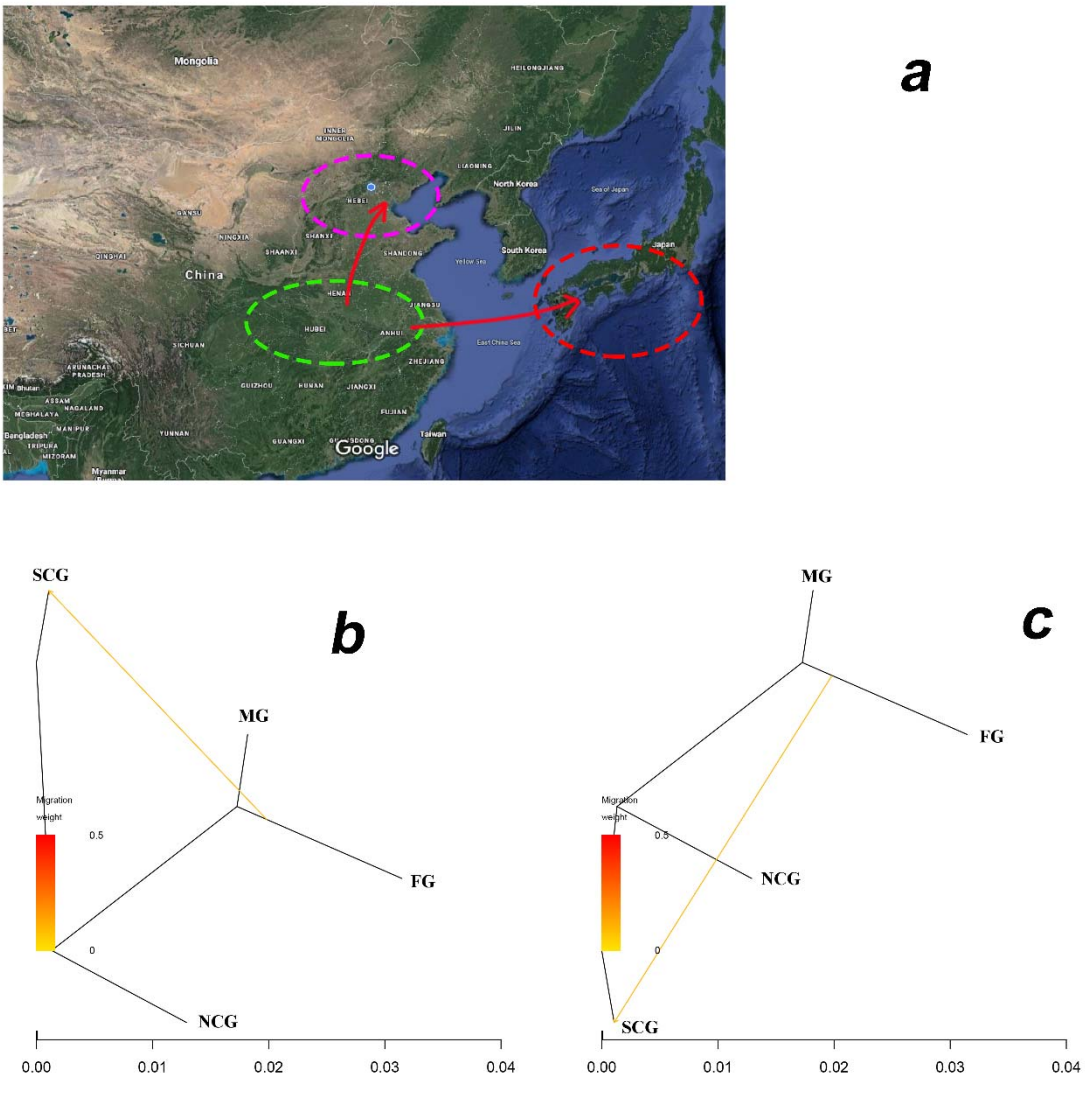

Figure 4. The inferred evolutionary path (a) and gene flow analysis between populations as inferred by TreeMix using a model with one (b) and two (c) admixture events. Admixtures are colored according to their weight. FG, the foreign cultivar group; MG, the mixed cultivar group; SCG, the southern cultivar group; NCG, the northern cultivar group.

Table 1. Statistics of variations across different chromosomes

Categories	Numbers	Ratio
Raw SNPs	16,600,033	88.73%
InDels	2,107,664	11.27%
Filtered SNPs	14,549,234	87.65%
Intergenic regions SNPs	7,272,007	43.81%
Coding regions SNPs	1,971,748	11.88%
Transition	10,315,764	59.58%
Transversion	6,999,126	40.42%
Transition/Transversion	--	147.00%
Non-synonymous SNPs	1,855,201	94.09%
Synonymous SNPs	116,547	5.91%

Table 2.The statistical values of genetic diversity within different populations.

Groups	Private allele number (Ap)	Observed heterozygosity (Ho)	Expected heterozygosity (He)	Nucleotide diversity (π)	Wright's F-statistic (F_{IS})	Tajima's D
FG	31,285	0.4290	0.3036	0.00467	-0.2411	1.4974
MG	30,112	0.3658	0.2846	0.00433	-0.1560	1.0094
SCG	27,287	0.3365	0.2606	0.00395	-0.1460	1.0022
NCG	546	0.3280	0.2336	0.00358	-0.1829	1.3211