

Phylogeography and population genetic structure of flowering cherry *Cerasus serrulata* (Lindley) Loudon in subtropical and temperate China

Xian-Gui Yi¹, Jie Chen¹, Hong Zhu¹, Yong-Fu Li¹, Xue-Xia Li¹, Meng Li¹, Yi-Fan Duan¹, Lin Chen¹, and Xian-Rong Wang¹

¹Nanjing Forestry University

May 5, 2020

Abstract

Cerasus serrulata (Rosaceae) is an important flowering cherry resource. It is almost the most widely distributed species in the genus, mainly included in the subtropical and temperate China, which enables the geographic evolutionary pattern to be a representative. Besides, the morphological traits are greatly varied especially in ornamental characters. All of these makes *Cerasus serrulata* a valuable research object. Thus, phylogeographic analysis was conducted to apprehend the spatial pattern and evolutionary history, which can also add insights into the phylogeography of the genus *Cerasus* and plants in subtropical and temperate China, as well as to deeper understand the genetic diversity and structure of the germplasm to make better and more effective utilization. A total of 327 individuals of 18 populations were collected. Three cpDNA fragments (matK, trnD-E, trnS-G) and the nuclear internal transcribed spacer (ITS) were utilized. The result showed a high genetic diversity both in species level and population level of *Cerasus serrulata*. The high genetic differentiation among populations and the existence of phylogeographic structure in whole were detected. In addition, no bottleneck was identified. The the distribution pattern and center were formed before the LGM. Two geographical lineages were inferred. One was confined to Qingling Mountain and Taihang Mt. The other was from the Wuling Mt to Lu-Huang Mt, and then went northeast to the coast of Asia. Besides, taxonomic treatments of the *Cerasus serrulata* complex were reconsidered.

Concise cover letter

Cerasus serrulata (Rosaceae) is an important flowering cherry resource. It is almost the most widely distributed cherry in China. The study of phylogeography on *C. serrulata* can bring insights into the spatio-temporal evolutionary pattern of the genus *Cerasus*. Furthermore, its spreading pattern covers the main subtropical and temperate China in eastern Asia, and also occupies the majority of Sino-Japanese forest subkingdom, making *C. serrulata* valuable in providing clues to the evolutionary history of these regions. As for *C. serrulata* itself, it harbors a great deal of morphological variations, especially in ornamental traits. Phylogeographic study can reveal the spatio-temporal evolutionary pattern, help us understand the current differentiation level of different geographic populations and how the diversification was formed. Besides, it can also recognize the genetic diversity and structure of *C. serrulata* to help us better comprehend the condition of this important resource so as to make much more effective utilization such as to breed new flowering cherry cultivars.

Introduction

Phylogeography focuses on the gene genealogy spatio-temporal pattern of related species or intraspecies as well as how the pattern was formed (Hickerson *et al.*, 2010; Bai & Zhang, 2014). Taking advantages of genetic diversity, genetic structure analysis and molecular time dating, temporal and spatial population

divergence and expansion, evolution history can be inferred; the glacial refugia can also be identified combining the climate changes and glacial events in late Quaternary (Hickerson *et al.*, 2010). It is widely acknowledged that the glacial-interglacial oscillations in Quaternary glaciation, especially the LGM, had eminent effects on shaping the contemporary distribution of species and their genetic diversity (Chung *et al.*, 2017).

The Sino-Japanese forest subkingdom, involved in the Sino-Japanese Floristic Region (SJFR) except the Himalayan forest region (Wu & Wu, 1996), covers the south and east of China, as well as the Korean Peninsula and the Japanese Archipelago. This region is embroiled in the subtropical and temperate climate zones in the northern hemisphere, mainly consisted of the subtropical and temperate China. There are remarkably rich flora in this region due to the high rainfall brought by the south-eastern monsoon, and they may have experienced dramatic distribution changes during the Quaternary climatic oscillations (Mitsui *et al.*, 2007). Mountains are believed to play important roles as refugia and dispersal corridors for species during the Quaternary global climate changes. Nanling Mountain and Qinling Mountain acted dual roles as a dispersal corridor in east-west direction and as a glacial refugium in subtropical China during the late Quaternary (Tian *et al.*, 2018; Guo *et al.*, 2014; Wulufu, 1964; Zhou, 1997). The main Korean mountain range [the so-called “Baekdudaegan” (BDDG)] is also thought to be a glacial refugium, mainly for the boreal and temperate flora of northeastern Asia (Chung *et al.*, 2017). In addition to large mountains, microrefugia such as smaller massifs or lowland sites are also of dominant importance for sustaining species. It was hypothesized that multiple microrefugia locate in the northern part of the southern refuge (24°N-33°N) (Wang & Ge, 2006).

Cerasus serrulata (Lindley) Loudon is one of the most widely spread species in genus *Cerasus*, distributing from the north in Heilongjiang Province to the south in Guangzhou Province (Li & Bartholomew, 2003), covering a widest latitude range (~ 24°N- 45°N) (Li *et al.*, 2014); and from west in Guizhou Province to Japan (Li & Bartholomew, 2003), also across a broad longitude range. The distribution pattern is involved in the Sino-Japanese forest subkingdom, and is also within the subtropical and temperate monsoon climate zones. All the extensive coverage, various and special habitats as well as complex climate, indicating a strong adaptation, have bred high variants in *C. serrulata*, and highlighted a high research value of phylogeography to bring new clues to the history of geological events, as well as the development and utilization of cherries such as breeding new or improved ornamental cultivars since *C. serrulata* itself possesses good ornamental characteristics and discrepancies of flower color, flower type, leaf color, tree shape, high resistance, etc. Nevertheless, the high variant level has also well caused some controversies on taxonomic relationships of related species, which needs further studies.

In this study, *C. serrulata* was selected as a typical model of cherry in subtropical and temperate China along with Korean peninsula. Three maternal-inherited chloroplast DNA and the bi-parentally inherited nuclear DNA ITS sequences were used together for phylogeographic analysis of *C. serrulata* for the first time to better understand the genetic diversity and genetic structure, as well as to clarify the temporal and spatial population divergence and expansion history of *C. serrulata*, which can also provide new clues for cherry evolutionary history and identifying refugia of subtropical and temperate China in SJFR. In addition, controversial taxonomic groups within the species and of related species, seemingly caused by the different processes of speciation in *Cerasus serrulata*, were also discussed to add more proofs for further clarifying their relationships.

Materials and methods

Population sampling

We have conducted field investigations of *Cerasus serrulata* according to specimen and literature records since 2010, and a total of 18 populations of *Cerasus serrulata* with 327 accessions have been collected (Table 1, Figure 1). Among them, seventeen populations were from China and one was from Korea. Green, young and healthy leaves were gathered from individuals at least 30 m apart, and were quickly stored in silica gel. All accessions were used for cpDNA analysis, while 174 were used for ITS analysis, with a maximum accession number of 10 for each population (Table 1). The whole collections should have covered the most of the

distributions of *Cerasus serrulata* .

DNA extraction, PCR amplification, and sequencing

The genomic DNA was extracted by CTAB-SiO₂ method (Nan, 2012; Zhang et al, 2004). 3 chloroplast DNA (cpDNA) fragments, selected after screening: *matK*, *trnD-E*, *trnS-G*, and the nuclear internal transcribed spacer (ITS) were amplified. Polymerase chain reaction (PCR) amplifications (25 μ L) contained: 12.5 μ L 2 \times Taq PCR MasterMix (Tiangen, Beijing, China), 1 μ L each primer, 2 μ L DNA template and 8.5 μ L ddH₂O. The PCR procedure was as follows: initial denaturation (95 $^{\circ}$ C, 5 min), 35 cycles of denaturation (94 $^{\circ}$ C, 45 s), annealing (52/56 $^{\circ}$ C, 45 s) and extension (72 $^{\circ}$ C, 45 s), and a final extension (72 $^{\circ}$ C, 8 min). The PCR products were checked by a 1.0% agarose gel and were sent for sequencing on ABI Prism 3730 Genetic Analyzer (Sangon, Shanghai, China).

Genetic diversity and phylogeographic analysis

The sequences were checked visually and edited manually by BioEditor ver. 7.09 (Hall, 1999), and then aligned by Clustal W in MEGA ver. 5.0 (Tamura et al., 2011). The three cpDNA sequences were concatenated for subsequent analysis. Identification of cpDNA haplotypes (H) and ITS ribotypes (R), and calculation of their diversities (H_d/R_d) as well as nucleotide diversity (P_i) were performed by DnaSP v. 5.0 (Librado & Rozas, 2009). Analysis of molecular variance (AMOVA) was conducted to estimate the genetic variances within and between populations and groups by Arlequin ver. 3.1 (Laurent *et al.*, 2005). Network of cpDNA haplotypes and ITS ribotypes were drew by TCS1.2.1 (Clement *et al.*, 2000). Phylogenetic analysis, taking *Prunus salicina*, *Cerasus tomentosa* and *C. yedoensis* as outgroups, was constructed based on maximum parsimony (MP), maximum likelihood (ML) and Bayesian (BI) to make the best use of the CIPRES Science Gateway V. 3.3. (Miller & Gross, 2011). The robustness of each internal branch of the genealogical trees was evaluated with 1,000 bootstrap replicates. All the above analysis involved both cpDNA and ITS sequences.

G_{ST} , representing haplotype frequency, and N_{ST} , representing haplotype difference, were calculated to examine the existence of population structure by PERMUT v. 2.0 with 10,000 permutations (Pons & Petit, 1996). $N_{ST} > G_{ST}$, $P < 0.05$ indicates the presence of a phylogeographic structure. Mismatch distribution analyses and neutrality tests were conducted to test population expansion using ARLEQUIN v. 3.5 with 10,000 permutations. These analysis were only conducted with cpDNA sequences.

Divergence time dating

Best-fit models of nucleotide substitution of chloroplast haplotypes were chosen by jModelTest2 v2.1.6. The divergence times were estimated by BEAST v1.8.4 (Drummond et al., 2016). Lacking fossil records, secondary calibration points according to the diversification of Rosaceae were applied to calibrate node ages. Divergence time dating was calculated in two steps with two trees respectively. Five secondary calibration points were applied to the first tree: Rosaceae Crown, 90.18 Mya (Zhang et al., 2017; Li et al., 2017); Amygdaleae, 84.84 Mya (Zhang et al., 2017); Spiraceae, 84.30 Mya (Zhang et al., 2017); Roseae+Rubeae, 59.51 Mya (Zhang et al., 2017); Maleae Crown, 50.06 Mya (Zhang et al., 2017; Li et al., 2017) to evaluate the divergence time of *Prunus* (node 1) and the time of *Cerasus* Crown (node 2). These two nodes were subsequently used to calibrate the node ages in the second tree of haplotypes of *Cerasus serrulata*. The GTR substitution model and the uncorrelated log-normal relaxed clock and Yule process as for tree prior were selected. MCMC was run for 1×10^7 generations with sampling every 5,000 generations. The parameters were checked in Tracer ver. 1.7.1 (Rambaut et al., 2018). Final phylogenetic trees were edited in FigTree v. 1.4.

Results

Genetic diversity

The total length of aligned sequences for the three chloroplast fragments (*matK*, *trnD-E*, *trnS-G*) was 2161 bp with 25 variable sites, including 21 single nucleotide variant sites, 2 6-bp insertions and 5 deletions in 323

individuals, and 661 bp for ITS with 10 variable sites in 174 individuals from 18 *Cerasus serrulata* populations. Results of diversity analyses were shown in Table 1. At species level, the haplotype diversity (H_d) was 0.828 and the ribotype diversity (R_d) was 0.673; the nucleotide diversity (P_i) of cpDNA and ITS were 1.490 and 2.521, respectively. All of the data suggested a high genetic diversity of *Cerasus serrulata*. At population level, H_d varied from 0.000 to 0.700 while R_d varied from 0.000 to 0.733; $P_i \times 10^{-3}$ of cpDNA varied from 0.000 to 1.090 while that of ITS varied from 0.000 to 2.820. Population CPL showed the highest haplotype diversity both in cpDNA and ITS regions. Combing the geographic distribution (Figure 1) along with haplotype distribution (Figure 1, 4), networks (Figure 2, 5) and cladograms (Figure 3, 6), four regions of groups were divided (Table 1). At region level, H_d varied from 0.266 to 0.747 and R_d varied from 0.246 to 0.752, while P_i varied from 0.280 to 1.940 for cpDNA, and from 0.370 to 1.670 for ITS. Both kinds of sequences revealed a lowest diversity level in Northwest China (NW). But differently, region in Southeast China (SE) showed a highest H_d while region in Southwest China (SW) showed a highest R_d , indicating a highest haplotype/ribotype diversity level in South China.

Genetic structure

Analysis of molecular variance (AMOVA) revealed a high differentiation level of both cpDNA ($F_{st}=0.69325$) and ITS ($F_{st}=0.78692$) in *C. serrulata* (Table 2). It was estimated that 69.32% of the whole variations of cpDNA sequences and 78.69% of the whole variations of ITS sequences existed among populations, signifying a phylogeographic structure in *C. serrulata* ($N_{ST}=0.388 > G_{ST}=0.328$, $P < 0.05$) (Table 3). But in detail, groups in northwest (NW) ($N_{ST}=0.305 < G_{ST}=0.461$, $P > 0.05$) and northeast (NE) ($N_{ST}=0.496 > G_{ST}=0.473$, $P < 0.05$) showed no or weak phylogeographic structure.

Distribution pattern and phylogenetic relationship of cpDNA haplotypes

A total of 19 haplotypes were detected (Figure 1, Table 1), among which 8 haplotypes were unique only to one population, while haplotype H13 was the most common one, followed by H5. Number of haplotypes in each populations ranged from 1 to 4, whereas 5 populations owned only one kind of halotype.

The distribution showed a geographically relevant pattern with 3 geographic groups, also conferred by the haplotype network (Figure 2) and phylogenetic tree (Figure 3). The first group distributed in the northwest (NW), clearly involving population MP, LiS and BTM, contained 3 haplotypes (H1-H3) with H3 being the most frequent one. The second group scattered in the southwest (SW), west from Dabie Mt. to the east of Hengduan Mt., comprising 6 haplotypes (H4-H9) in which H5 was the most dominant one. The third group dispersed from the northeast of Dabie Mt. to Korea, regarded as the east region (E), possessed the most abundant haplotypes (H10-H19), in which H13 took up the most common part.

In a higher resolution, Group E can be further separated into two groups: the northeast group (NE) involving haplotypes of H14-H19, and the southeast group (SE) comprising haplotypes of H10-H12, as four regional groups described above (Table 1). Nonetheless, group SE was not monophyletic in this partition. In a lower resolution, all the haplotypes could be allocated into 2 groups: group NW, and all the other groups, amalgamating group SW and E (NE+SE).

Distribution pattern and phylogenetic relationship of ITS ribotypes

A total of 9 ITS ribotypes were identified in the 174 individuals (Figure 4, Table 1). Of all the ribotypes, 5 were unique only to one population and ribotype R4 was the most standard ribotype which covered 12 populations, and was the unique ribotype for 7 populations.

Taken together, the distribution pattern (Figure 4), ribotype network (Figure 5) and phylogenetic tree (Figure 6) exhibited a structure with 2 groups of these ribotypes which was consistent with the cpDNA haplotype relationship in a lower resolution. Group one, consisting of 5 ribotypes (R1-R5) with R4 being the most extraordinary ribotype, and involving 13 populations as the utmost group, located in Southeast China and the southwest region (SW+SE). Group two incorporated 4 ribotypes (R6-R9), taking R7 as the main ribotype, distributed discontinuously in the northwest region and around Korea (including population FHS in China), covering the northwest and northeast groups (NW+NE).

Divergence time dating

The divergence time to the most recent common ancestor (TMRCA) of *C. serrulata* was estimated to be around 10.6 Mya (PP=1; 95% HPD: 8.83-12.46 Mya), near the middle of Miocene (Figure 7, Table 4). During 10.6 Mya, the northwest group began its divergence from the other groups. When it came to 8.02 Mya, the southwest group and the east group involving the northeast and southeast groups, started to separate away.

Population dynamics

In neutrality test, the value of Tajima's D and Fu's F_S of all populations and geographical groups was mainly positive, or unremarkable negative but $P > 0.05$ (Table 5), denoting that the hypothesis of neutral evolution was not to be excluded. Mismatch distribution analyses implied that all geographical groups did not refuse to have experienced sudden expansion ($P(SSD) > 0.05$, $P(Rag) > 0.05$), except the northeast group whose $P(Rag) = 0.04$, slightly lower than 0.05, but also indicating a potential expansion (Table 6). Unimodal distribution curves detected in all and in each region further proved a recent demographic expansions (Figure 8). The expansion time of each groups were accessed according to the formula $T = \tau/2u$ (Rogers & Harpending, 1992; Rogers, 1995). It was estimated that the southwest group (SW) expanded first in 0.719 Mya (95% CI: 0.022-1.452), followed by the southeast group (SE) in 0.665 Mya (95% CI: 0.000-1.249), and the last northwest group (NW) in 0.586 Mya (95% CI: 0.000-0.684).

Discussion

Genetic diversity

Cerasus serrulata is a widely distributed species in China of the genus *Cerasus* and is involved in the subtropical and temperate monsoon climate zones. Both cpDNA ($H_d = 0.828, P_i = 1.490$) and ITS ($R_d = 0.673, P_i = 2.521$) analysis proved a high genetic diversity of *C. serrulata*, higher than wild *C. pseudocerasus* (cpDNA: $h = 0.557, \pi = 2.09$) (Chen et al., 2015). The diversity based on ITS sequences was lower than *C. dielsiana* (ITS: $H_d = 0.879, \pi = 3.56$) (Zhu et al., 2019), which was ought to be a result of less sampling, might not really mean a lower diversity level. The broad spreading pattern, allowed by its high adaptation ability, must be a significant reason for the high diversity, since diverse topographies such as alps, mountains and hills, as well as plains, additionally along with different climate zones have provided variant habitats for *C. serrulata* during adaptation in diffusion which impelled variations and high diversity. Obviously, large sampling, covering the most of its natural distribution records, was also a fundamental factor. Besides, another reason may be the long evolutionary history of *C. serrulata*, which was dated back to 10.60 Mya, even earlier than *C. dielsiana* (6.28 Mya) (Zhu et al., 2019), because the long existence allowed a more multiple accumulation of variants, especially strengthened by a wide and strong adaptation.

Genetic differentiation

AMOVA pointed out a high genetic differentiation level among *Cerasus serrulata* populations (cpDNA: $F_{st} = 0.69325$, ITS: $F_{st} = 0.78692$), signifying the existence of a phylogeographic structure ($N_{ST} = 0.388 > G_{ST} = 0.328, P < 0.05$). A similar high diversity level was also detected in *C. dielsiana* (ITS: $F_{ST} = 0.7827$) (Zhu et al., 2019), but contrarily low levels were unraveled in *C. pseudocerasus* landraces (ISSR: $F_{ST} = 0.308$) (Chen et al., 2018), *C. pseudocerasus* containing both wild populations and landraces (cpDNA: $F_{ST} = 0.24922$, ITS: $F_{ST} = 0.31571$) (Chen et al., 2015), *C. tomentosa* (SSR: $F_{ST} = 0.181$) (He et al., 2015), *C. lannesiana* var. *speciosa* (cpDNA: $F_{ST} = 0.1655$) (Kato et al., 2011), *C. mahaleb* (RSPD: $F_{ST} = 0.1646$) (Jordano and Godoy, 2000), *C. campanulata* (SSR: $F_{ST} = 0.1477$) (Lv, 2006; Su, 2007), *C. jamasakura* (SSR: $F_{ST} = 0.043$) (Tsuda et al., 2009). This might because both *C. serrulata* and *C. dielsiana* are much more widely distributed than most of species in *Cerasus*. Additionally, adequate seed flow were also believed to weaken the interspecific divergence of *C. pseudocerasus* and other similar *Cerasus* species (Chen et al., 2015; Kato et al., 2011; Tsuda et al., 2009; Zhu et al., 2019).

Together of the cladograms and networks of both cpDNA and ITS sequences, finally two groups of *C. serrulata* within all the regions were defined: one mainly consisted of the northwest groups (NW), and the

other was primarily composed of the left groups (SE+SW). Nevertheless, the northeast group (NE), involving population FHS in Liaoning Province and a foreign population K-NS in Korea, appeared a confusing phylogeographic pattern. It was assigned to group SE+SW in cpDNA analysis, but was related to group NW in ITS analysis. An earlier study based on nuclear SSRs, involving the main northeast and southeast populations of *C. serrulata*, indicated a consistent connection within the eastern regions involving population FHS and K-NS (Yi et al., 2018), which in accordance with the result of cpDNA analysis, making the distribution more puzzled. N_{ST}/G_{ST} analysis manifested no or weak phylogeographic structure within groups in northwest (NW) ($N_{ST}=0.305 < G_{ST}=0.461$, $P > 0.05$) and northeast (NE) ($N_{ST}=0.496 > G_{ST}=0.473$, $P < 0.05$). Actually, *C. serrulata* distributions were also recorded in Hebei Province and Heilongjiang Province in China, as well as in Japan, but we failed to obtain samples in these regions. Thus, the lack of the comprehensive sampling in Northeast China and Japan was mainly ought to be the reason of this confusion.

Demographic dynamic and potential glacial refugia

Collectively, two genealogies of *C. serrulata* were detected: the northwest (NW) genealogy and the other genealogy consisted of the left groups involving the northeast (NE), southeast (SE) and southwest (SW) groups in our sampling areas, located mainly in the south and east of China within the subtropical and temperate monsoon climate zones. But the phylogeographic pattern did not compare the geographic pattern or climate zones.

The most recent common ancestor of *C. serrulata* appeared around 10.6 Mya in the mid-late of Miocene, which was also the divergence time of the two genealogies, calculated in the molecular dating analysis. To be detailed, the northwest group (NW) which formed one lineage began to divided out firstly in 10.6 Mya; afterwards, the southwest (SW) and the east (E, containing NE+SE) groups in the other lineages started to differentiate away from each other during 8.02 Mya (Fig. 7, Table 4). In other species, adjacent time was also detected of *C. dielsiana* ancestor (6.28 Mya) (Zhu et al., 2019), *Polystichum glaciale* ancestor (6.89 Mya) (Dong et al., 2018) and the divergence of *Ixiolirion tataricum* and *I. songaricum* (~7 Mya) (Li et al., 2019). It seems that the divergence event of many species was concentrated during such a period of time, and this time overlapped with that of the Himalayan movement and that of the Arctic ice-sheet to begin developing while the global climate was cooling down (Liu et al., 1998) and drier (Zhao et al., 2018; Ma & Tian, 2015). The strong landform and climate changes might lead to such a concentration of species divergence. In this, the uplifting of Qinling Mountains might took place accompanied by the Himalayan movement (Xue et al., 2004), and gradually isolated the north groups from the other, finally resulting in the divergence of the two geographical lineages.

Demographic expansion was recognized both in neutrality test and mismatch distribution analysis, except the northeast group (NE). The expansion time of group southwest (SW), southeast (SE) and northwest (NW) were dated back to 0.719 Mya, 0.665 Mya and 0.586 Mya, respectively (Table 3). All the time was before the the Last Glaciation Maximum (LGM, from 0.0265 Mya to 0.019-0.02 Mya) during the Last Glaciation of Quaternary glacial period, indicating that the distribution center and pattern of *C. serrulata* was formed before the LGM, which was accordant to that of genus *Cerasus* (Cao, 2006; Li, 2009). Comparing these time with the global temperature change curve of the Earth since 5 Mya (Cao, 2006; Li, 2009; Wang, 2014), it was found that these time was all sited in the uprising temperature periods in the frequent cooling-warming climate oscillations (Lisiecki & Raymo, 2005). The warmed-up temperature which also brought humidity should have enabled the expansions. Besides, within all these populations, CPL possesses the highest diversity which was located in Guizhou Province in the northwest group (NW). Regarding the divergence time and genetic diversity level together, two lineage colonization events were inferred. One was confined to Qingling Mountain and Taihang Mt. The other comprised two separate steps: firstly from the Wuling Mt. to Jiangnan Hilly; and then northeast to the coast of Asia. Glacial refugia which protected species to survive are usually considered to be large mountains. Nevertheless, importance of microrefugia such as smaller massifs or lowland sites is rising more and more attentions (Tomasz et al., 2019; Zhu et al., 2019). The broad land of the southeast of Asia, linked together by Bohai Sea and the coast of Asia today as

sea level declined resulted from the glaciation (Guo et al., 2014), as well as valleys in east-west mountains such as Qinling Mt, provided refuges and expansion accesses for many species (Wulufu, 1964; Zhou, 1997). In this, the large mountain of Qinling, Wuling and Taibai Mts in Korea, as well as some microrefugia in Jiangnan Hilly especially the Huang Mt., Lu Mt. and Tianmu Mt. were inferred to be glacial refugia for *C. serrulata*, also having regard to their higher genetic diversity. In addition, the consistent distribution pattern in cpDNA and SSR analysis (Chen, 2016; Yi et al., 2018) of the northeast (NE) and southeast (NW) groups which are isolated by the Bohai Sea is also elucidated by the conjunction of Korea and eastern China during glacial period. All the lineage colonizations and potential refugia are generally coincident with the result of phylogeographic analysis of *Cerasus* (Li, 2009), indicating that the wide spread *C. serrulata* can be a representative species of the genus *Cerasus* in the evolutionary process.

Taxonomic reconsideration

Cerasus serrulata, a complex with ample variations and transition phenotypes some of which are morphologically specialized in the process of diversification by specific environments, has brought about taxonomic controversies. Here, *C. xueluoensis*, published in 2013 (Nan et al., 2013), treated as an outgroup in this study, was totally mixed up with haplotypes or ribotypes of *C. serrulata* in cladograms built by both cpDNA and ITS sequences (Figure 3, Figure 6). To further discuss the relationship between *C. xueluoensis* and *C. serrulata*, more wild related species were added to rebuild the cladograms based on ITS sequences. This time, *C. xueluoensis* was separated from the clade formed by all ribotypes of *C. serrulata* (Figure 9). However, more adequate and accurate information is needed to define the interrelationship between *C. xueluoensis* and *C. serrulata*.

Interestingly, *C. serrulata* var. *lannesiana* was significantly detached from *C. serrulata* in this tree (Figure 9), supporting its treatment as *C. lannesiana*. *C. laoshanensis* published in 2017 was described to resemble *C. serrulata* (Zang, 2017). Nevertheless, population LaoS in this study representing *C. laoshanensis*, showed no specificity from *C. serrulata* individuals in any relationship analysis. Thus, it was endorsed to merge *C. laoshanensis* into *C. serrulata*. Besides, *C. huangangensis*, a record in 2007 (Yi, 2007), represented by population HGS which showed distinctiveness as exclusively possessed only one kind of haplotype or ribotype both in cpDNA and ITS (Figure 1, Figure 4), also within the big clade in both trees of *C. serrulata*, is ought to be treated as *C. serrulata* var. *huangangensis*.

Conclusion

We found dominant genetic diversity and existence of phylogeographic structure in *Cerasus serrulata*. The estimated divergence time and expansion events further indicated that the distribution pattern and center of *C. serrulata* was formed before LGM. Two geographical lineages were inferred. One was confined to Qinling Mountain and Taihang Mt. The other was from the Wuling Mt to Lu-Huang Mt, and then went northeast to the coast of Asia, totally with two separate steps. Taxonomic statements of several controversial related species were reconsidered. *C. xueluoensis* was close to *C. serrulata* but also exhibited differentiation at species level, which needs further study. *C. lannesiana* is confirmed rather than *C. serrulata* var. *lannesiana*. And finally a new variation, *C. serrulata* var. *huangangensis* is identified.

References

- Bai W. N. & Zhang D. Y. (2014). Current status and future directions in plant phylogeography. *Chinese Bulletin of Life Sciences*, 26(02), 125-137.
<https://doi.org/10.13376/j.cbcls/2014020>
- Cao D. W. (2006). Molecular phylogeography of *Prunus* subgenus *Cerasus*. Xi'an: Northwest University.
<https://doi.org/10.7666/d.y893947>
- Chen J. (2016). Genetic diversity analysis of *Cerasus serrulata* populations based on SSR markers. Nanjing: Nanjing Forestry University.

<http://cdmd.cnki.com.cn/Article/CDMD-10298-1017800064.htm>

Chen T., Chen Q., Luo Y., Huang Z. L., Zhang J., Tang H. R., Pan D. M., Wang X. R. (2015). Phylogeography of Chinese cherry (*Prunus pseudocerasus* Lindl.) inferred from chloroplast and nuclear DNA: insights into evolutionary patterns and demographic history. *Plant Biology* , 17(4), 787-97.

<https://doi.org/10.1111/plb.12294>

Chen T., Hu Y. Y., Chen Q., Wang Y., Zhang J., Tang H. R., Wang X. R. (2018). Molecular and morphological data reveals new insights into genetic diversity and population structure of Chinese cherry (*Prunus pseudocerasus* Lindl.) landraces. *Genetic Resources and Crop Evolution volume* , 65, 2169-2187.

<https://doi.org/10.1007/s10722-018-0683-9>

Chung M. Y., López-Pujol J. & Chung, M. G. (2017). The role of the Baekdudaegan (Korean Peninsula) as a major glacial refugium for plant species: A priority for conservation. *Biological Conservation* , 206, 236-248.

<https://doi.org/10.1016/j.biocon.2016.11.040>

Clement M., Posada D., Crandall K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular ecology* , 9(10), 1657-1659.

<https://doi.org/10.1046/j.1365-294x.2000.01020.x>

Dong L., Xu B., Rana S. K., Li Z. M., Sun H. (2018). Phylogeography of rare fern *Polystichum glaciale* endemic to the subnival zone of the Sino-Himalaya. *Plant Systematics and Evolution* , 304(4), 485-499.

<https://doi.org/10.1007/s00606-018-1495-2>

Drummond A. J., Rambaut A., Suchard M. BEAST v1.8.4.

<https://github.com/beast-dev/beast-mcmc/releases/tag/v1.8.4>

Guo X. D., Wang H. F., Bao L., Bai W. N., Ge J. P. (2014). Asymmetric pattern of climate oscillations and the origin of biodiversity. *Journal of Beijing Normal University (Natural Science)* , 50(06), 622-628.

<http://d.wanfangdata.com.cn/periodical/bjsfdxxb201406012>

Hall, T.A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* , 41, 95-98.

He H. L., Cai Y. L., Gao T. X., Li B., Wan T., Wang Y. (2015). SSR evaluation of genetic diversity of seven wild *Prunus tomentosapopulations* in Shaanxi. *Journal of Northwest A&F University (Natural Science Edition)* , 43(06), 193-198,204.

<https://doi.org/10.13207/j.cnki.jnwafu.2015.06.021>

Hickerson M. J., Carstens B. C., Cavender-Bares J., Crandall K. A., Graham C. H., et al. (2010) Phylogeography's past, present, and future: 10 years after *Avisé*. *Molecular Phylogenetics and Evolution* , 54, 291-301.

<https://doi.org/10.1016/j.ympev.2009.09.016>

Hu Y., Wang X., Zhang X. X., Zhou W., Chen X. Y., Hu X. S. (2019). Advancing phylogeography with chloroplast DNA markers. *Biodiversity Science* , 27(2), 219-234.

<https://doi.org/10.17520/biods.2018319>

Jordano P. & Godoy J. A. (2000). RAPD variation and population genetic structure in *Prunus mahaleb* (Rosaceae), an animal-dispersed tree. *Molecular Ecology* , 9(9), 1293-1305.

<https://doi.org/10.1046/j.1365-294x.2000.01009.x>

Kato S., Iwata H., Tsumura Y., Mukai Y. (2011). Genetic structure of island populations of *Prunus lan-nesiana* var. *speciosa* revealed by chloroplast DNA, AFLP and nuclear SSR loci analyses. *Journal of Plant Research* , 124(1), 11-23.

<https://doi.org/10.1007/s10265-010-0352-3>

Laurent E., Guillaume L., Stefan S. (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* , 1, 47-50.

Li C. L., Bartholomew B. (2003). Flora of China. Vol. 9. Beijing, St. Louis: Science Press; Missouri Botanical Garden Press.

<http://www.iplant.cn/info/Cerasus%20serrulata?t=foc>

Li M. M. (2009). Phylogeography of *Prunus* subgenus *Cerasus* and genetic diversity of *P. pseudocerasus* . Xi'an: Northwest University.

<https://doi.org/10.7666/d.y1454041>

Li M., Ohi-Toma T., Gao Y. D., Xu B., Zhu Z. M., & Ju W. B., Gao X. F. (2017). Molecular phylogenetics and historical biogeography of *Sorbus sensu stricto* (rosaceae). *Molecular Phylogenetics and Evolution* , 111, 76-86.

<https://doi.org/10.1016/j.ympev.2017.03.018>

Li M., Yi X. G., Wang H. C., Shang T., Gu Y., Wang X. R. (2014). Studies on the relationship between *Cerasus serrulata* distribution region and the environmental factor S. *Journal of Nanjing Forestry University (Natural Science Edition)* , 38(S1), 74-80.

<https://doi.org/10.3969/j.issn.1000-2006.2014.S1.016>

Li Y., Song F., Zhang X. N., Lv G. H. (2019). Phylogeography suggest the Yili Valley being the glacial refuge of the genus *Ixiolirion* (Amaryllidaceae) in China. *Systematics and Biodiversity*, 17(4), 385-401.

<https://doi.org/10.1080/14772000.2019.1612966>

Librado P. & J. Rozas J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* , 25(11), 1451-1452.

<https://doi.org/10.1093/bioinformatics/btp187>

Lisiecki L. E. & Raymo M. E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic δ^{18} records. *Paleoceanography* , 20, PAa1003.

<https://doi.org/10.1029/2004PA001071>

Liu T. S., Zheng M. P., Guo Z. T. (1998). Nitiation and evolution of the Asian monsoon system timely coupled with the ice-sheet growth and the tectonic movements in Asia[J]. *Quaternary Sciences* , 18(3), 194-204.

<http://www.dsyyj.com.cn/CN/Y1998/V18/I3/194>

Lv Y. L. (2006). Genetic diversity and breeding strategy of *Cerasus campanulata* Maxim. Naning: Nanjing Forestry University.

<https://doi.org/10.7666/d.y918223>

Ma X. L & Tian J. (2015). East Asian monsoon evolution and aridification of Northwest China viewed from land and sea on the tectonic-orbital time scale since 15Ma. *Quaternary Sciences* , 6(35), 1320-1330.

<https://doi.org/10.11928/j.issn.1001-7410.2015.06.02>

Miller, A. J. & Gross B.L. (2011). From forest to field: Perennial fruit crop domestication. *American Journal of Botany* , 98(9), 1389-1414.

<https://doi.org/10.3732/ajb.1000522>

Mitsui, Y. , Chen, S. T. , Zhou, Z. K. , Peng, C. I. , Deng, Y. F. , & Setoguchi, H. (2007). Phylogeny and biogeography of the genus *Ainsliaea* (asteraceae) in the Sino-Japanese region based on nuclear rDNA and plastid DNA sequence data. *Annals of Botany* , 101(1), 111-124.

<https://doi.org/10.1093/aob/mcm267>

Nan C. H. (2012). Study on population variation and reproductive biology of *Cerasus discoidea* Yü et Li. Nanjing: Nanjing Forestry University.

<http://cdmd.cnki.com.cn/Article/CDMD-10298-1012033442.htm>

Nan C. H., Wang X. R., Tang G. G., Yi X. G. (2013) *Cerasus xueluoensis* (Rosaceae), a new species from China. *Annales Botanici Fennici*, 50(1-2), 79-82.

<https://doi.org/10.5735/085.050.0114>

Pons O. & Petit R. J. (1996). Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* , 144(3), 1237-1245.

<https://www.genetics.org/content/144/3/1237>

Rambaut A., Drummond A., Xie D., Baele G., Suchard M. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic biology*, 67(5), 901-904.

<https://doi.org/10.1093/sysbio/syy032>.

Rogers A. R. (1995). Genetic evidence for a Pleistocene population explosion. *Evolution* , 49, 608-615

Rogers A. R. & Harpending H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* , 9, 552-569.

Su Q. (2007). Genetic diversity among populations of *Cerasus campanulata* . Nanjing: Nanjing Forestry University.

<https://doi.org/10.7666/d.y1111861>

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* , 28(10), 2731–2739.

<https://doi.org/10.1093/molbev/msr121>

Tian S., Kou Y., Zhang Z., Yuan L., & Zhang Z. (2018). Phylogeography of *Eomecon chionantha* in subtropical China: the dual roles of the Nanling Mountains as a glacial refugium and a dispersal corridor. *BMC Evolutionary Biology* , 18(20).

<https://doi.org/10.1186/s12862-017-1093-x>

Tomasz S., Marek M. & Michal R. (2019). Relict populations and Central European glacial refugia: The case of *Rhododendron ferrugineum* (Ericaceae). *Journal of Biogeography* , 46(2), 392-404.

<https://doi.org/10.1111/jbi.13512>.

Tsuda Y., Kimura M., Kato S., Katsuki T., Mukai Y., Tsumura Y. (2009). Genetic structure of *Cerasus jamasakura* , a Japanese flowering cherry, revealed by nuclear SSRs: implications for conservation. *Journal of Plant Research* , 122(4), 367-375.

<https://doi.org/10.1007/s10265-009-0224-x>

Wang H. W. & Ge S. (2006). Phylogeography of the endangered *Cathaya argyrophylla* (Pinaceae) inferred from sequence variation of mitochondrial and nuclear DNA. *Molecular Ecology* , 15, 4109–4122.

<https://doi.org/10.1111/j.1365-294X.2006.03086.x>

Wang X. R. (2014). An Illustrated Monograph of Cherry Cultivars in China. Beijing: Science Press.

Wu Z.Y. & Wu S.G. (1996). A proposal for a new floristic kingdom (realm): the E. Asiatic Kingdom, its delineation and characteristics. In: Zhang AL, Wu SG, eds. Floristic characteristics and diversity of East Asian plants. Beijing/Berlin: China Higher Education Press/Springer Verlag, 3-42.

Wulufu E. B. Translated by Zhong C. X., et al. (1964). Historical Phytogeography. Beijing: Science Press.

Xue X. X., Li H. H., Li Y. X., Liu H. J. (2004). The new data of the uplifting of Qinling Mountains since the Middle Pleistocene. *Quaternary Sciences* , 24(1), 82-87.

<http://www.dsyyj.com.cn/CN/Y2004/V24/I1/82>

Yi X. G. (2007). Studies on investigation and exploitation of the resources of *Cerasus* Mill. in Wuyi Mountain. Nanjing: Nanjing Forestry University.

<https://doi.org/10.7666/d.y1111964>

Yi X. G., Chen J., You L. X., Cong R., Wang H. C., Daun Y. F., Wang X. R. (2018). Genetic diversity of *Cerasus serrulata* populations assessed by SSR markers. *Journal of Nanjing Forestry University(Natural Science Edition)* , 42(05), 25-31.

<http://d.wanfangdata.com.cn/periodical/njlydxxb201805004>

Zang D. K. (2017). *Cerasus laoshanensis* (Rosaceae), a new species from Shandong, China. *Annales Botanici Fennici* , 54(1-3), 135-137.

<https://doi.org/10.5735/085.054.0321>

Zhang B., Zhang L., Zhu G. Q., Wang M. X., Huang M. R. (2004). A rapid and simple method of total DNA extraction from tree. *Journal Of Nanjing Forestry University (Natural Science Edition)* , 28(1), 13-16.

<https://doi.org/10.3969/j.issn.1000-2006.2004.01.003>

Zhang S. D., Jin J. J., Chen S. Y., Chase M. C., Soltis D. E., Li H. T., Yang J. B., Li D. Z., Yi T. S. (2017). Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. *New Phytologist* , 214(3), 1355-1367.

<https://doi.org/10.1111/nph.14461>

Zhao L., Lu H. Y., Tang L. Y. (2018). Cenozoic palynological records and vegetation evolution in the Weihe Basin, Central China. *Quaternary Sciences* , 38(5), 1083-1093.

<http://www.dsyyj.com.cn/CN/Y2018/V38/I5/1083>

Zhou Y. L. (1997). Phytogeography of Northeast China. Beijing: Science Press.

Zhu H., Yi X. G., Li Y. F., Zhu S. X., Li M., Duan Y. F., Wang X. R. (2019). Phylogeography and population genetic structure of flowering cherry species *Cerasus dielsiana* in subtropical China. *Systematics And Biodiversity* .

<https://doi.org/10.1080/14772000.2019.1670750>

Tables

Table 1. Location, haplotypes and genetic diversity of *Cerasus serrulata* populations
Symbol

Table 1. Location, haplotypes and
Geographic coordinate

MP	106.680°E, 33.382° N
LiS	111.999°E, 35.340° N
BTM	111.930°E, 33.374° N
CPL	106.671°E, 27.658° N
HeS	112.694°E, 27.291° N
TTZ	115.761°E, 31.164° N
LS	116.982°E, 29.545° N
HGS	117.767°E, 27.841° N
HS	118.168°E, 30.130° N
TMS	119.430°E, 30.343° N
NS	119.514°E, 31.175° N
BHS	119.078°E, 32.138° N
YTS	119.443°E, 34.719° N
TS	117.111°E, 36.232° N
MS	117.958°E, 35.551° N
LaoS	120.602°E, 36.196° N
FHS	124.092°E, 40.429° N
K-NS	126.987°E, 37.548° N
Northwest China, NW (MP, LiS, BTM)	Northwest China, NW (MP, LiS, B
Southwest China, SW (CPL, HeS)	Southwest China, SW (CPL, HeS)
Southeast China, SE (TTZ, LS, HGS, HS, TMS, NS, BHS, YTS, TS ,MS, LaoS)	Southeast China, SE (TTZ, LS, HC
Northeast China & Korea, NE (FHS, K-NS)	Northeast China & Korea, NE (FH
Mean	Mean
All	All

Table 6. Neutrality test for four geographical groups of <i>Cerasus serrulata</i>	Table 6. Neutrality test for four geographical g
Statistics	Northwest group (NW)
Tajima's <i>D</i>	-0.289
<i>p</i> -value	0.406
Fu's <i>F_S</i>	0.67
<i>p</i> -value	0.645

Table 3. G_{st}/N_{st} and mismatch distribution analysis(MDA)of <i>Cerasus serrulata</i>	Table 3. G_{st}/N_{st} and mismatch distrib
Group	G_{st}/N_{st}
Northwest group (NW)	0.461/0.305 (p>0.05)
Southwest group (SW)	0.186/0.204 (P<0.01)
Southeast group (SE)	0.180/0.326 (P<0.01)
Northeast group (NE)	0.473/0.496 (p>0.05)
All	0.328/0.388 (P<0.05)

Table 4. Summary of cpDNA-based divergence time estimation by Bayesian approach	Table 4. Summary of cpDNA-ba
Node	Anchor age (Mya)
A# Rosaceae Crown	90.18
B# Amygdaleae+ (Maleae,Spiraceae)	84.84
C# Maleae+Spiraceae	84.3
D# Roseae+Rubeae	59.51

E# Maleae Crown	50.06
1# Prunus +Cerasus	24.79
2# Cerasus	10.31
3# NW +(SW, East Group)	
4# SW+(SE, NE)	
5# H10	
6# H12 + [NE, SE(H11+H13)]	
7# NE	
8# NE(H14+H15)+SE(H11+H13)	
9# H19	

Table 6. Neutrality test for four geographical groups of *Cerasus serrulata*

Statistics

Tajima's D

p -value

Fu's F_S

p -value

Table 6. Neutrality test for four geographical groups of *Cerasus serrulata*

Northwest group (NW)

-0.289

0.406

0.67

0.645

Table 5. Neutrality test for populations of *Cerasus serrulata*

Statistics

Tajima's D

p -value

Fu's F_S

p -value

Table 5. Neutrality test for populations of *Cerasus serrulata*

MP

0.801

0.795

1.422

0.804

Figure legends

Figure 1 Geographic distribution of the cpDNA haplotypes in 18 natural populations of *Cerasus serrulata* .

Figure 2 TCS network of cpDNA haplotypes in 18 natural populations of *Cerasus serrulata* . Each numbered circle (H1-H19) represents a unique haplotype, and the size of the circle is proportional to the overall frequency of each haplotype in the entire sample of the species

Figure 3 Cladograms of the cpDNA haplotypes of *Cerasus serrulata* . Numbers at nodes indicate values of Bayesian posterior probability (>0.5)/MP bootstrap percentage (>50%)/ML bootstrap percentage (>50%).

Figure 4 Geographic distribution of the ribotypes based on ITS sequences in 18 natural populations of *Cerasus serrulata* .

Figure 5 TCS network of the ribotypes based on ITS sequences in 18 natural populations of *Cerasus serrulata* . Each numbered circle (R1-R9) represents a unique ribotype, and the size of the circle is proportional to the overall frequency of each ribotype in the entire sample of the species

Figure 6 Cladograms of the ribotypes based on ITS sequences of *Cerasus serrulata* . Numbers at nodes indicate values of Bayesian posterior probability (>0.5)/MP bootstrap percentage (>50%)/ML bootstrap percentage (>50%).

Figure 7 BEAST-derived chronograms of *Cerasus serrulata* based on cpDNA sequences. Numbers at each node indicate the node age (million years ago, Mya). Critical nodes are marked in red circles and their node ages are enlarged in red squares. The blue bars illustrate the extent of the 95% highest posterior density (HPD) credibility intervals for each divergence time.

Figure 8 Mismatch distribution for the four geographic groups and all populations of *Cerasus serrulata* . The blue line shows observed values (Obs.) while the red line represents the expected values (Exp.) under a model of sudden expansion/spatial expansion.

Figure 9 Cladograms of *Cerasus serrulata* complex based on ITS sequence, taking *C. sargentii* , *C. speciosa* , *C. jamasakura* , *C. yedoensis* , *C. subhirtella* , *C. tomentosa* and *Prunus salicina* as outgroups. Numbers at nodes indicate values of Bayesian posterior probability (>0.5)/MP bootstrap percentage (>50%)/ML bootstrap percentage (>50%).

DATA ACCESSIBILITY

DNA sequences: GenBank accessions MT007600-MT007789

Competing Interests Statement

None declared.

Author Contributions section

Yi Xian-Gui substantially contributed to the conception and design. Yi Xian-Gui, Li Meng, Zhu Hong, Chen Lin and Duan Yi-Fan conducted the field research and sample collection. Chen Jie and Li Yong-Fu did the laboratory work. Yi Xian-Gui, Zhu Hong, Li Yong-Fu, Li Xue-Xia and Li Meng analyzed data. Yi Xian-Gui and Chen Jie drafted the article. All authors participated in ameliorating the research and discussions of the results, as well as polishing up the manuscript. Wang Xian-Rong finally approved the version to be published.

ACKNOWLEDGMENTS

This research was supported by the Science Foundation of Jiangsu Province, China 116 [BE2017037], and the Forestry Science and Technology Innovation and Popularization Project in 117 Jiangsu Province, China[LYKJ[2017]14,LYKJ[2018]29, LYKJ[2018]43]. We are grateful to Dr. Sun Zhong-Shuai for his valuable suggestions on the discussion. We are also grateful to Xu Zi-Kun, Shi Cai-Xia, Ke Xian-Shi and Qiu Jing for their help with the field work, and Ren Qing-Hong for sample collection in Korea. Besides, we are thankful to Shang Tao and Cong Rui for help in laboratory work.

Hosted file

Figures.rar available at <https://authorea.com/users/295883/articles/424755-phylogeography-and-population-genetic-structure-of-flowering-cherry-cerasus-serrulata-lindley-loudon-in-subtropical-and-temperate-china>