Study on Population Variation and Phylogeography of Cherry in China(Cerasus conradinae)

Jing-Jing Dong[1], Xian-Gui Yi[1], Xian-Rong Wang[1], wu tong[1], Jie Chen[1], Yang hong[1], zhou huajing[1], Gao shucheng[1], chen xiangzhen[1], and LI meng[1]

[1]Nanjing Forestry University

January 21, 2023

Abstract

Cerasus conradinae is an important germplasm resource of wild cherry in China. In this work, sampling was expanded and genetic evidence was added for further study. The current and future potential populations were predicted by niche model. Based on three cpDNA sequences and one nrDNA sequence, and combined with the phylogeographic evolution of 12 populations of 244 individuals in C.conradinae, the temporal and spatial patterns of genetic variation in C.conradinae were investigated, and the effects of genetic drift and differentiation as well as natural environmental factors on the genetic variation and evolutionary distribution of C.conradinae were elucidated. Finally, morphological evidence combined with molecular evidence was used to discuss the species definition of population variation and differentiation. The results showed that Central China, East China and Southwest China were the core regions for the conservation and utilization of germplasm resources of C.conradinae. Support for variant Cerasus conradinae var. ruburm is established. The genetic diversity of C.conradinae was high (Hd = 0.830; Rd = 0.798). There was genetic variation among populations of C.conradinae, and genealogical geographic structure existed among the populations and three geographical groups, but the genetic differentiation coefficient at each level was low. The gene exchange was obvious in Southwest China, and the differentiation was obvious in Central China. The population and three geographic groups do not reject the expectation of expansion model. Two distinct lineages (three geographic groups) were identified from the population of C.conradinae: Central China+ East China lineage and East China+ (Central China+Southwest China) lineage, two lineages of 4.38Mya occurred in the early Pliocene based on geographical isolation. The southeastern part of Eastern China near Mount Wuyi was the most likely refuge for C.conradinae. The results provided a theoretical basis for the classification and identification of C.conradinae and the protection and utilization of germplasm resources.

1. Introduction

Cerasus conradinae (Koehne) Yu et Li belongs to *Cerasus*Mill. in the Rosaceae family(Yü et al.,1986; Shulaev et al., 2008; Phipps,2014). Deciduous trees (3-10m). This species of umbels, usually with 3 to 5 flowers, white or pink petals, single petals, flowers before the leaves open, large flowers. Leaves with blunt serrate, toothed end with small glands, fruit red (Wang., 2014). Trees of *C.conradinae* are tall and beautiful with long flowering time (Chen et al.,2015). At the same time, new varieties of cherry and cherry suitable for Chinese cultivation can be selected, bred and crossed (Bai et al.,2019), which is an important germplasm resource of wild cherry in China (Chen et al., 2016a). However, *C.conradinae* is in a semi-wild state at present, which needs to be studied and exploited.*C.conradinae* is distributed in Hunan, Hubei, Sichuan, Guizhou, Yunnan, Guangxi, Shaanxi, Henan and other middle and high altitude (500-2100 m) areas in China(Potter et al., 2007; Chen et al., 2020). Due to the long history of cultivation, interspecific hybridization,complex topography, climate and soil environment, China has a large range of external morphological variation, population variation and differentiation, and high genetic diversity(Zhang et al., 2017;Aranzana et al., 2019; Chen et al., 2019). Previous studies have focused on the resource survey, community structure and morphological characteristics, such as YuLin, BaiWenFu(Yu et al., 2007; Bai et al., 2020), wild cherry resources for investigating

the morphological characteristics were observed and the record shows that *C.conradinae* of inflorescence and flower color variation, there are some umbel floret number between 3 ˜ 10 flower, design and color is white or dark pink. However, specific studies on the phylogeography of *C.conradinae* are still blank, and its biogeography and character evolution are still unclear (Nybom et al., 2014) .

Nowadays, the combination of cpDNA and nrDNA sequences has become a common way to explore plant phylogeography. Maternally inherited chloroplast cpDNA markers can provide information on migratory routes and refuge locations of species without parental genetic confounding (Avise et al.,1987;Avise., 2000).The nrDNA inherited from both parents showed hybridization, multiploidy and gene introgression in the process of plant evolution(Vander Wall and Beck., 2012; Perdereau et al., 2014).With the addition of phenotypic characteristics, ecological geography and other theories to explore the unresolved issues of species definition, kinship, pedigree geography and historical evolution and change. In the era of information globalization, various software simulations to protect species germplasm resources, such as the addition of niche model to pedigree geography, The evolutionary history of population change was explored by setting different simulation parameters to assess the prediction of habitat and potential habitat. Niche simulation combined with studies on phylogeography can directly reflect the response of species to changes in climate factors, resulting in species isolation and differentiation (Chan et al., 2011). Among them, MaxEnt model was proposed by Phillips, which has the characteristics of high prediction accuracy, easy operation and strong practicality. It has been widely used to predict the potential habitat of species (Phillips et al., 2004). In order to further carry out*C.conradinae* group genetic and geographical research spectrum, we will make the sample and increase genetic evidence study of*C.conradinae* , finally proposed based on the three segments of chloroplast sequences and a sequence fragment nrDNA (Turkoglu et al., 2010), integrating the model and the morphological evidence of niche solve *C.conradinae* current and future potential ShiSheng District prediction, Species definition and phylogeography of population variation and differentiation, in order to reveal the historical causes of the current distribution pattern of *C.conradinae* from different perspectives, and provide theoretical basis for exploring its origin, migration routes and dispersal modes after the ice age, which is of great significance for the conservation and utilization of germplasm resources of *C.conradinae* (Bai., 2014; Yi et al., 2020).

2. Materials and methods

2.1Ecological Niche Modeling

Field investigation and species specimen database were used to collect longitude and latitude information of *C.conradinae* geographical distribution points, and 554 specimens from 24 herbarium were retrieved. The ambiguous geographical information, artificial planting or repeated specimen information were deleted, and the downloaded specimen information data were accurately compared one by one to ensure the accurate geographical location of the specimen. In each 2.5 'x2.5' grid, only one distribution point with the shortest distance from the center was selected, and the effective data of 201 geographical distribution points of *C.conradinae* specimens were finally collected (Qiu et al., 2018). Environmental factor data are mainly downloaded from World climate database WorldClim Version 2.1, January 2020 (http: //www.worldclim.org/), 19 climatic factors in the contemporary (1970-2000), 2050s (2041-2060), 2070s (2061-2080) and other periods, The spatial resolution is 2.5min, and the future climate data (2050s and 2070s) are based on the general atmospheric circulation model CCSM4. Through the method of obtaining climate information through DIVA-GIS version7.5, the environmental and climatic variables that play a key role in the influence of limiting climate factors on the geographical distribution of *C.conradinae* were screened (Zhu et al., 2017), and Pearson correlation analysis was conducted with the altitude, longitude and latitude of each geographical distribution point of*C.conradinae* (Garah et al., 2019). Based on the specimen records of *C.conradinae* in China combined with the geographical information positioning and climate data of the field distribution population, the MaxEnt version3.4.1 niche model was used to reconstruct the geographical distribution pattern of *C.conradinae* , and to simulate the suitable areas of potential distribution of*C.conradinae* in the contemporary, 2050s and 2070s in China.

2.2.Plant Materials

The sampling plan and strategy were formulated through the statistics and proofreading of the specimen data and the prediction of the contemporary suitable area of *C.conradinae.* From 2019 to 2022, the authors and members of the research group carried out field investigation and sample collection of *C.conradinae* , and collected 244 samples from 12 populations in 8 provinces (Figure 2, Table 1). During the sampling of *C.conradinae* in Hubei province, stable and continuous red-pink color was found in two populations of Phoenix Pool (FHC) in Yichang city and Gexian Mountain (GXS) in Xianning city, and stable and continuous capillaries were found on leaf peels and calyx tubes in one population of Wangcheng slope (WCP) in Enshi Autonomous, Hubei Province. Detailed morphological comparison and specimen examination were carried out among the three populations, and field observation was carried out one by one. The traditional phenotypic characteristics showed that there were stable and continuous variation in different populations in the same area (Table 1). Therefore,*Cerasus conradinae* var. *ruburm* and *Cerasus conradinae* var. *pubescens* were first established as variants, and samples from each population were collected and verified by molecular means.

2.3.DNA Extraction, Polymerase

Chain Reaction Amplification, Sequencing, and Sequence Alignment

Genomic DNA kit [Tiangen Biotechnology (Beijing) Co., LTD.] was used to extract DNA from 244 fresh leaves of *C.conradinae* according to the operating procedures. After extraction, 1% agarose gel electrophoresis was used to detect the concentration and purity of the extracted DNA. The DNA of qualified samples was placed in the refrigerator at -80 degC, and then the qualified DNA samples were sent to Shanghai Shenggong Co., LTD for sequencing to obtain haplotypes for phylogeographic analysis. By literature review and logging on NCBI website, universal primers for different sequences of cpDNA and nrDNA of*Cerasus* were screened and collected. Three pairs of cpDNA universal primers were encoded gene maturation enzyme gene fragment (*MatK*) (Heckenhauer et al., 2016), non-coding gene spacer fragment (*TrnL-F*)(Feng et al., 2017) and (*TrnD-E*)(Yi et al., 2018) and one pair of nrDNA sequence fragment (*ITS*) for genetic and lineage structure analysis of *C.conradinae* . The PCR amplification reaction system was 25 μl, which consisted of 2.0 μL DNA template, 0.5 μL (10 μmol/L) of upstream and downstream primers, 12.5 μL of 2×PCR Master Mix and 9.5 μL of ddH2O. Polymerase chain reaction (PCR) amplification was performed using the following procedures: initial denaturation at 94*C for 5 min, followed by denaturation at 94 ° C for 1 min for 30 cycles, annealing at 54-56 ° C for 1 min, and extension at 72*C for 1 min; Then 72 for a final 5 minutes.

2.4.Data Analysis

MAFFT version 7 (Katoh and Standley, 2013) was used for sequence alignment of all cpDNA sequences, and then PhyloSuite version 1.2.2 was used for manual inspection and revision. After cutting, use Concatenate Sequence in the sequence of *MatK* , *TrnL-F* , and*TrnD-E* . Genetic diversity parameters of each population were obtained from DnaSP version 6 (Librado and Rozas,2009), which mainly include: number of haplotypes,h 、haplotype diversity,Hd 、nucleotide diversity,$\Pi\imath/\pi$ 、number of polymorphic sites,S (Nei, 1987). Arlequin version 3.1 (Excoffier and Lischer, 2010) was used for AMOVA analysis of molecular variance, and 1000 non-parametric permutations were used for significance test. The degree of freedom (*d.f* .), total variance, variation component, variation variance distribution, and genetic differentiation index (F_{ST}) among populations were estimated within and among populations of *C.conradinae* . The numerical values of *Nst* and *Gst* and the significance of *P*were obtained by setting parameters, and the reasons affecting genetic differentiation of *C.conradinae* population and the existence of obvious pedigree geographical structure were discussed. Using PopArt version 1.7 and Notepad++ version 7.7 and other software. TCS Network (Clement et al., 2000) was used to construct haplotype network diagram and explore haplotype kinship. According to the results of TCS haplotype relationship, the geographical distribution map of haplotypes was drawn by Arcmap version 10.2 combined with haplotype species and distribution proportion. Arlequin version 3.1 analysis of Neutrality test (NT) and Pairwise Mismatch distribution (MDA), In the neutral test analysis, Tajima's test and Fu'sFs test are mainly used (Tajima, 1989). Combined with DnaSP version 6.1, if the Expected value curve obtained by clicking Population Size Changes and Observed value curve coincide, the combination of the two is used to test the hypothesis of sudden population expansion. And whether the

C.conradinae population experienced bottleneck or expansion events (Harpending, 1994).

2.5.Molecular Dating and DemographicAnalyses

BEAST version 6(Drummond and Rambaut, 2007) was used to construct the time tree at the level of Rosaceae family and calculate the divergence time of *C.conradinae* lineage. According to the latest phylogenetic tree of Rosaceae (Zhang et.al, 2017), four representative groups of Rosaceae subfamilies and their chloroplast genome sequences were selected by NCBI. All chloroplast genomes were aligned by MAFFT version7(Katoh and Standley, 2013). Maximum likelihood method (ML) was used to construct the optimal nucleotide substitution model by phylogeny. IQ-Tree version1.6.12 (Nguyen et al., 2015) software was used to obtain the model. Bayesian Information Criterion (BIC) was used to calculate the best nucleotide substitution model (Darriba et al., 2012). Random start tree was used for 10,000 generations, sampling frequency was once every 1000 generations, and the best nucleotide replacement model was selected. An intergenus (four subfamilies) phylogenetic tree of Rosaceae was constructed using five molecular clocks or fossil correction points. The whole chloroplast genome was used to determine the differentiation time of *C.conradinae* and other *Cerasus*plants in the same clade. Then, 10 ribotypes divergence time phylogenetic trees of nrDNA(ITS) of *C.conradinae* were constructed from two fossil correction points, and the differentiation time and migration route of*C.conradinae* were predicted.Software parameters were finally set as the GTR substitution model and exponential uncorrelated relaxed model with Yule process. Two independent MCMC runs were conducted, each with 300,000,000 generations and sampling every 1,000 generations. The first 12,000,000 generations in each run were removed as burn-in. According to fossil divergence time, the Rosaceae Crown (N1) was constrained by lognormal distribution, with its mean value set at 90.18Mya and standard deviation at 0.05(Zhang et al, 2017). (Tribe Maleae + Tribe Spiraeeae) + Tribe Amygdaleae (N2) was constrained by lognormal distribution, with the mean value set at 72.62Mya and the standard deviation at 0.05(Zhang et al, 2017; Zhang et al., 2021). The Tribe Amygdaleae (N3) was constrained by lognormal distribution, with a mean of 68.58Mya and a standard deviation of 0.01 (Wehr and Hopkins, 1994; Xiang et al, 2017;Zhang et al., 2021). According to fossil divergence time, the differentiation time of*Prunus* in a narrow sense is 60.7 ˜ 62.4Mya, and the mean and standard deviation of *Prunus* divergence time (N4) is 55 Mya and 0.09131(Li et al., 2011; Chin et al., 2014). Node*Cerasus* (N5) is constrained by lognormal distribution, with the mean value set at 28.21Mya and the standard deviation at 0.05(Zhang et al, 2021).The tree and log files of two separate runs are combined with LogCombiner version 2.6.6 (part of the BEAST package). All effective sample size (ESS) values were well above 200. TreeAnnotator version 1.7.3 (part of the BEAST version 1.7.3 package) has 25% wear and is used to estimate the mean divergence time and 95% highest posterior density (HPD) interval. Finally, FigTree version 1.3.1 (Rambaut, 2009) is used. The age of each node and its 95% HPD interval are displayed (Li et al., 2011; Chin et al., 2014).

3. Results

3.1.Ecological Niche Modeling

According to the model evaluation criteria of receiver operating curve (ROC), the detection results of *C.conradinae* (training setting 0.945, test setting 0.949) reached the excellent standard (Alvarado-Serrano and Knowles, 2014). In the current potential suitable areas, the distribution of *C.conradinae* in Hubei, Zhejiang, Fujian, Jiangxi, Anhui, Henan, Yunnan, Chongqing, Guizhou, Hunan, Sichuan is potentially suitable for growth. There is also a distribution trend in Jiangsu and Shandong. The core distribution area is mainly concentrated in Central China and East China. In the future, with the gradual warming of the climate, the total suitable area in the 2050s decreases compared with the contemporary area, but the total suitable area in the 2070s increases, and the extremely suitable area has a trend of spreading to high latitude and northeast direction, and the high suitable area has a trend of spreading to low latitude direction. However, the core suitable areas remain in Central China and East China (Figure 2). Based on DIVIA-GIS, 10 environmental factors were extracted for the habitats of the potential distribution areas. Finally, it was determined that the wettest seasonal precipitation (bio16) and annual precipitation (bio12) were the primary factors limiting the contemporary potential geographical distribution of *C.conradinae* . Annual low temperature (bio7) was an important climate factor. Water, the dominant climate limiting factor for *C.conradinae* , had a greater im-

pact than temperature, which was speculated to be related to the climatic zone of *C.conradinae* distribution area being affected by north subtropical and subtropical humid monsoon regions. In 2070s China, especially in Central China and Southwest China, the rainfall increased, and the temperature increase changed significantly. The predicted increase of the total suitable area of *C.conradinae* in 2070s was related to the change trend of the above comprehensive climate factors (Garah et al., 2019).

3.2.Sequence Variation Haplotype Frequency and Distribution

The total length of the 3 cp DNA fragments was 2238bp, *MatK*fragment was 740bp, *TrnL-F* fragment was 868bp, and *TrnD-E*fragment was 630 bp. A total of 22 mutation sites were detected, including 2 mutation sites in *MatK* fragment, 12 mutation sites in*TrnL-F* fragment and 8 mutation sites in *TrnD-E* fragment. A total of 18 chloroplast haplotypes (H1-H18) were identified from 12 populations. H1, H7 and H15 were the most common, with frequencies of 30.33% (74), 24.59% (60) and 9.43% (23), respectively (Table 3). H1 had the highest frequency of haplotype, which appeared in Central China (GXS, DWS) and East China (MYS, WYS, DBS, QLF, ZXTC). H7 appeared in Southwest China (EMS, HFG), Central China (WCP, FHC, GXS) and East China (MYS, WYS, DYH). H15 was found in Southwest China (EMS, HFG) and Central China (WCP, FHC). The lowest frequency was H10, H11, H18, and the number of distribution was 2. The common haplotypes are H1-H9, H11-H13 and H15. Five haplotypes, H10, H14, H16, H17 and H18, were endemic and occurred in East China (MYS), Central China (GXS, FHC) and Southwest China (WCP, HFG). The populations with the highest and lowest number of haplotypes were GXS (7) and DBS (2), respectively (Table 3; Table 4).

The length of 1 nrDNA fragment sequencing sequence was 578bp, and a total of 19 mutation sites were detected. A total of 11 Ribotypes (R1-R11) were identified from 12 populations. R4, R1 and R3 were the most common, with frequencies of 30.3% (85), 24.5% (47) and 9.4% (28), respectively (Table 3). R4 had the highest frequency of Ribotypes, which appeared in Central China (GXS, DWS) and East China (MYS, WYS, QLF, ZXTC, DYH), R1 appeared in Southwest China (EMS, HFG) and Central China (WCP, FHC), R3 appeared in Central China (WCP, FHC), and R1 appeared in Central China (WCP, FHC). R6 is the unique ribotype of GXS, the number of ribotypes is 4, the number of ribotypes is rare ribotypes, and the most number of ribotypes (R) is the population GXS (6) (Table 3; Table 4).

3.3.Haplotype Network

Based on the TCS haplotype network diagram, Hap1 and Hap7 were found in the inner part of the network diagram, while H1, H7 and H15 contained more individuals. Therefore, Hap1 and Hap7 are presumed to be ancient haplotypes, and other haplotypes are derived haplotypes. Among the 18 haplotypes, two groups were formed: Central China+ East China and Central China+Southwest China. Different haplotypes under each branch communicated in the same area, and Central China contained a total of 17 haplotypes (except H18). H10, H14, H16, H17 and H18 were unique haplotypes of MYS (Jiangxi Province, Yichun City), GXS (Hubei Province, Xianning City), WCP (Hubei Province, Enshi Autonomous), FHC (Hubei Province, Yichang City) and HFG (Chongqing City), respectively (Figure 2, Table 3).

According to the TCS Ribotypes network diagram (Figure 3), R4 is in the inner part of the network relationship diagram, and R1, R4 and R5 individuals contain a large number. Therefore, it is speculated that R4 is an ancient ribotype. The 11 Ribotypes in central China were divided into two groups: East China + Central China (GXS, DWS) and Southwest China + Central China (WCP, FHC). Central China contained a total of 10 ribosomes (except R11). R6 was the endemic Ribotypes type of GXS (Figure 3, Table 3).

3.4.Genetic Diversity and Population Genetic Structure

The haplotype genetic diversity parameter(H_d)was the highest in population WYS (Fujian). Population DBS(Anhui) had the lowest genetic diversity index. The overall haplotype diversity was $H_d = 0.830$, the nucleotide diversity was$P_i \times 10^{-3}$=0.878, and the average nucleotide difference number $K = 1.916$. The haplotype diversity (H_d) of different populations varied from 0.257 to 0.836, the nucleotide diversity ($P_i \times 10^{-3}$) varied from 0.230 to 1.113, and the mean nucleotide difference number (K) varied from 0.524 to 2.191

(Table 4). The variation range of haplotype diversity (H_d) in the Southwest China (Chongqing, Sichuan), Central China (Hubei, Hunan) and East China (Jiangxi, Anhui, Fujian, Zhejiang) was 0.627- 0.774, and the variation range of nucleotide diversity ($P_i \times 10^{-3}$) was 1.050-1.240. The mean nucleotide difference (K) ranged from 1.895-2.771. The genetic diversity parameters in Central China were higher than those in other regions. The population differentiation index (F_{st}) of cpDNA level in *C.conradinae* was 0.48886, indicating a high degree of differentiation. The genetic variation among populations was 48.89%, and the genetic variation within populations was 51.11%. The genetic variation within populations was slightly higher than that between populations, and the values tended to be similar(Heckenhauer et al., 2016). The AMOVA results showed that the genetic variation among regional groups was 3.06%, the genetic variation among populations within regional groups was 46.32%, and the genetic variation within populations within regional groups was 50.62%. The genetic variation within populations within regional groups was slightly higher than that between populations within regional groups. Genetic differentiation parameters of *C.conradinae* population (N_{st} = 0.29843, N_{st} = 0.28176, P < 0.05) showed pedigree geographical structure. Genetic differentiation parameters in Southwest China:N_{st} = 0.081, Gst = 0.072, P < 0.05; Genetic differentiation parameters in Central China: Nst = 0.22810, G_{st} = 0.18051, P < 0.05; Genetic differentiation parameters in East China: Nst =0.33970, Gst =0.30473, P<0.05, pedigree geographic structure was detected in all three groups of geographic regions (Table 4; Table 5).

The total Ribotypes diversity(R_d)was 0.798, the nucleotide diversity wasP_i x10^{-3} = 0.886, and the average nucleotide difference number (K) =3.799. The Ribotypes diversity (R_d)of different populations varied from 0.000 to 0.798, the nucleotide diversity (P_i x 10^{-2}) varied from 0.000 to 0.886, and the average nucleotide difference number (K) varied from 0.000 to 3.799 (Table 4). The variation range of Ribotypes diversity (R_d), nucleotide diversity (P_i x 10^{-2}) and mean nucleotide difference number (K) in Southwest, Central and East China was 0.486-0.745, 0.169-0.756 and 0.972-4.341, respectively. The genetic diversity parameters in Central China were higher than those in other regions. The population differentiation index(F_{ST} = 0.82511)at the nuclear gene level, indicating that the horizontal differentiation degree of *C.conradinae* species was high. The genetic variation among populations was 82.51%, and the genetic variation within populations was 17.49%. The genetic variation within populations was higher than that between populations. The results of AMOVA showed that the genetic variation among regional groups was 16.51%, and the genetic variation among populations within regional groups was 56.25%, with the lowest coefficient of differentiation among populations in Southwest China. In the three geographic subgroups, the genetic variation among populations within regional subgroups was higher than that within populations within regional subgroups (Table 4; Table 5).

3.5.Molecular Dating, and Demographic Analyses

N1: The differentiation time of Rosaceae (Subfamily Amygdaloideae + Subfamily Rosoideae) was estimated to be 92.17Mya (95% HPD: 92.07-92.29Mya). N2: (Tribe Maleae + Tribe Spiraeeae) + Tribe Amygdaleae had an estimated differentiation time of 75.61 Mya (95% HPD: 75.39-75.76 Mya). N3:*Prinsepia*+((*Laurocerasus* +(*Padus* + *Maddenia*))+ (*Amygdalus* +(*Prunus* + *Armeniaca*))+*Cerasus*) has an estimated differentiation time of 68.53 Mya (95%) HPD: 68.27-68.72 Mya). N4: (*Laurocerasus* +(*Padus* +*Maddenia*))+ (*Amygdalus* +(*Armeniaca* +*Prunus*))+*Cerasus* has an estimated differentiation time of 49.81 Mya (95% HPD: 49.47-50.0 Mya). N5: (*Amygdalus* + (*Armeniaca* + *Prunus*))+*Cerasus* has an estimated differentiation time of 28.13Mya (95% HPD: 28.0-28.26Mya). In the AI subbranch (Subg.*Cerasus*), *Cerasus mahaleb* is in 15.25 Mya (95% HPD: Mya Mya) diverged from the common ancestor of Cerasus, The estimated differentiation time of *Cerasus clarofolia* + (*Cerasus pseudocerasus* + *Cerasus scopulorum*) + *C. conradinae* was 11.71 Mya (95% HPD: 8.09-15.31Mya). (Figure 4 (A); Table 6)

C.conradinae 10 ribotypes coancestor time is 4.38Mya (95% HPD: 2.38-6.51Mya). The results of divergence analysis showed that the differentiation time of Central China+ East China lineage and East China+ (Central China+Southwest China) lineage of *C.conradinae*could be traced to 4.38Mya(95% HPD: 2.38-6.51Mya)in Pliocene. R10 was the first ribotype differentiated from the Central China + East China lineage, belonging to the Central China specific ribotype, indicating that Central China and East China differentiated earliest.

The differentiation time should be 3.32Mya (95% HPD: 1.12-5.17Mya) in Pliocene. R11 is the first haplotype differentiated from East China+ (Central China+Southwest China) lineage, belonging to the East China specific ribotype, and the differentiation time is about 2.17Mya (95% HPD: 0.77-4.54mya). It indicated that there was obvious gene exchange between East China and Central China. R3 is the first ribotype differentiated from the lineage of Central China + Southwest China. It is unique to Central China and differentiated at about 1.10Mya (95%HPD: 0.11-2.85Mya) during Pleistocene. (Figure 4 (B); Table 6)

If the curve of mismatch distribution analysis is bimodal or multi-modal, and the result values of SSD and Hrag are low (p-value < 0.05 for the transient expansion model), it indicates that due to the passing of time, the population of *C.conradinae* is in a relatively stable or gradually degenerate level. If it shows a single peak and the result values of SSD and Hrag are high (p-value >0.05), it indicates that the population has undergone a recent expansion event(Figure5; Table 7). Or based on the degree to which the curves of the expected value and the observed value coincide. The greater the coincidence similarity is, the expansion has occurred in the past. otherwise, the opposite is true(Figure5). Based on the analysis of the mismatch distribution at the species level between populations, the results showed that the double p-value values of 12 populations from 8 provinces were all greater than 0.05, indicating that there was expansion between populations (Table 7). From the level of population and geographical grouping, the results showed that population, Central China and East China showed an insignificant double-peak phenomenon, while Southwest China showed an obvious double-peak phenomenon, and all values of p-value were greater than 0.05. The results show that the population and three geographical groups of *C.conradinae* do not reject the expectation of the expansion model.

3.6.Population Variation and Taxonomic Treatment

Based on Bayesian information criteria (BIC), the optimal nucleotide replacement model was constructed using IQ-Tree version1.6.12. The best nucleotide replacement model was GTR+F+I+G4. Among the 18 haplotypes, H10, H14, H16, H17, and H18 were specific haplotypes in MYS (Yichun, Jiangxi), GXS (Xianning, Hubei), WCP (Enshi, Hubei), FHC (Yichang, Hubei), and HFG (Chongqing). The phylogenetic relationship of nrDNA (ITS) molecular markers showed that R6 was GXS (Xianning, Hubei Province) specific ribotype among 11 ribotype (Table 3;Figure6). In conclusion, the cpDNA sequence-binding phenotype supported the formation of the two varieties. However, the combination of nrDNA(ITS) sequence and phenotype supported *Cerasus conradinae* var. *ruburm* as a variety. The molecular evidence of *C.conradinae* var.*ruburm* was insufficient, and there was no obvious variation between *C.conradinae* and other populations. Therefore, the results of morphological markers and sequence markers support *C.conradinae* and *C.conradinae* var. *pubescens* (Hodel et al., 2021)(Table 3;Figure 6).

4. Discussion

4.1.Genetic Diversity and Population Genetic Structure

C.conradinae was widely distributed with high morphological diversity and obvious phenotypic variation. cpDNA and nrDNA markers showed that there was genetic differentiation among populations, and the population level genetic diversity of *C.conradinae* was high (Hd = 0.830; Rd = 0.798), higher than the mean value of 170 seed plants (mean value :h_T = 0.67; Petit et al., 2005). Generally, the high cpDNA haplotype diversity is considered to reflect a long evolutionary history, which also indicates restricted gene flow between populations (Varvioet al., 1986; Qiu et al., 2009). The population variation of *C.conradinae* was mainly between populations.The genealogical structure existed among the population and three geographical groups, but the genetic differentiation coefficient at each level was low. The southwest of China is the lowest and the gene exchange is obvious, which is closely related to the abundance of wild cerulea resources in southwest China. The differentiation was obvious in Central China, which was speculated to be related to the strong adaptability of *C.conradinae* and seed propagation of birds, beasts, water and wind (Zhang et al., 2010).

4.2. Geographical structure of pedigree

Based on the cpDNA and nrDNA haplotype phylogenetic tree and the TCS network map, the population of *C.conradinae* divided into two distinct lineages (three geographic groups) : lineage of Central China+ East China and lineage of East China+ (Central China+Southwest China) (Li et al., 2011; Chin et al., 2014). The lineages structure is basically the same as the clustering results of haplotype TCS network diagram. The phylogeographic group differentiation time was estimated to be 4.38Mya (95%HPD: 2.38-6.51Mya) in the early Pliocene, and two lineages differentiated based on geographic isolation (Yi et al., 2020). Changes to plants brought about by climate changes in the Pliocene largely contributed to the expansion of the population of *C.conradinae* (Chen et al.,2015). One lineage is the Central China + East China geographic group, and R10 is the first ribotype differentiated from the Central China + East China lineage, belonging to the unique ribotype of Central China(Turkoglu et al., 2010). It is inferred that *C.conradinae*originated from Central China and differentiated from East China on 3.32Mya (95% HPD: 1.12-5.17Mya) in Pliocene. The other lineage is East China+ (Central China+Southwest China), R11 is the first ribotype differentiated from East China+ (Central China+Southwest China) lineage and belongs to the East China specific ribotype. The differentiation time is about 2.17Mya (95% HPD: 0.47-4.54Mya), indicating significant gene exchange between East China and Central China, showing significant morphological variation and intraspecific genetic differentiation. Most of the other subclades diverged during the Pliocene. R3 is the first ribotype differentiated from the lineage of Central China + Southwest China. It is unique to Central China and differentiated at about 1.10Mya (95%HPD: 0.11-2.85Mya) during Pleistocene. The population of*C.conradinae* differentiated from Central China to Southwest China about 1.10Mya (95%HPD: 0.11-2.85Mya) in the early Pleistocene of the Quaternary. Therefore, the distribution center and overall pattern of*C.conradinae* have been formed during the Pliocene and Pleistocene transition, which is consistent with the prediction that the distribution center of *Cerasus* was formed before the onset of the Quaternary glacial period. The mountains of the northern Hemisphere are in the middle latitudes and the lower latitudes. Glacial activities in the Quaternary led to multiple alternations of cold glacial periods and warm interglacial periods, resulting in a significant rise and fall of sea level (Hewitt.,1999). It is speculated that *C.conradinae*differentiated into Southwest China. The results of mismatch distribution analysis (MDA) showed that the population of*C.conradinae* and the three geographical groups did not reject the expectation of expansion model (Wang et al., 2010). Haplotype distribution and diversity analysis and TCS results showed that Central China had the highest genetic diversity. The haplotype species were most abundant in the southeast region near Wuyi Mountain in Eastern China and were the most likely sanctuary of *C.conradinae* . Combined with the haplotype distribution range and diversity index analysis, it can be inferred that the southwest of Central China is another sanctuary of*C.conradinae* (Qiu et al., 2010).

During 2.82mya (95% HPD: 0.77-4.98Mya), *C.conradinae* in Central China (Dawei Mountain, Gexian Mountain, etc.) expanded to the southeast (Wuyi Mountain, Qingliang Feng, etc.). Then it spread to Dabie Mountain (R11) in Anhui Province in the northwest. *C.conradinae* expansion event occurred mainly in the middle Pleistocene(Arenas et al., 2012). The geological period of this stage is the third glacial age (Quaternary glaciation)(Mona et al., 2014). It is speculated that the climate change during the Quaternary interglacial period caused the expansion of the population of *C.conradinae* . The natural distribution area of*C.conradinae* was mainly in cool and humid regions in central China, East China and Southwest China. The habitat and climate characteristics of the distribution area could also explain the expansion event at this stage. Thus, *C.conradinae* formed a distribution center and refuge before the onset of the Quaternary glaciation (southeast of eastern China near Wuyi Mountain). During the Pleistocene 1.10Mya (95% HPD: 0.11-2.85Mya), about the interglacial period of the Quaternary glacial stage, *C.conradinae* spread from central and eastern China to the southwest (Heifeng Valley of Chongqing city and Mount Emei of Sichuan province)(Zhang et al, 2017; Zhang et al., 2021). This study reveals the distribution prediction, phenotypic variation, classification and phylogeography of potential suitable areas of *C.conradinae* (Alvarado-Serrano et al., 2014). The results provided a theoretical basis for the classification and identification of *C.conradinae* and the protection and utilization of germplasm resources.

5. Conclusion

Taking *C.conradinae* as the research object, a systematic combination of morphological markers and molecu-

lar markers was carried out. The population variation, differentiation and taxonomic location of *C.conradinae* were clarified, and the genetic diversity, genetic structure and evolutionary distribution of *C.conradinae* were revealed(Yi et al., 2018). Support for variant *C.conradinae* var.*ruburm* is established(Bai., 2014; Yi et al., 2020). Central China, East China and Southwest China were the core regions for the conservation and utilization of germplasm resources of *C.conradinae* . The genetic diversity of *C.conradinae* was high ($H_d = 0.830$; $R_d = 0.798$). There was genetic variation among populations of *C.conradinae* , and genealogical geographic structure existed among the populations and three geographical groups, but the genetic differentiation coefficient at each level was low. The gene exchange was obvious in Southwest China, and the differentiation was obvious in Central China. Two distinct lineages (three geographic groups) were identified from the population of *C.conradinae* : Central China+ East China lineage and East China+ (Central China+Southwest China) lineage, two lineages of 4.38Mya occurred in the early Pliocene based on geographical isolation. The southeastern part of Eastern China near Mount Wuyi was the most likely refuge for *C.conradinae* (Zhang et al, 2017); Southwestern of Central China is another refuge for *C.conradinae* .

Acknowledgments

We would like to thank all members of Cerasus Research Center of Nanjing Forestry University and Yongmei Yi from Hubei Minzu University, who participated in the sampling work of *C.conradinae* . 2022 is the 120th anniversary of my alma mater, Nanjing Forestry University. I would like to congratulate she on her happy birthdays.

Funding

This research was funded by Jiangsu Province Modern Agriculture Key Project [BE2020343] : Breeding and promotion of new varieties of red cherry in Yangtze River Basin. Xuzhou Science and Technology Project [KC21336]: Breeding, demonstration and promotion of excellent cherry variety resources in Xuzhou area.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability Statement

Data analysed in this study is publicly available

at Figshare (https://figshare.com/articles/dataset/PS-DE/21904659).

References

Yu, D. J., Lu, L. T., Ku, T. C., Li, C. L., and Chen, S. X. (1986). Flora of China, Vol.

38. Beijing: Science Press.

Shulaev, V., Korban, S. S., Sosinski, B., Abbott, A. G., Aldwinckle, H. S., Folta, K. M., et al. (2008). Multiple models for Rosaceae genomics. Plant Physiol. 147, 985–1003. doi: 10.1104/pp.107.115618

Phipps, J. B. (2014). Flora of North America North of Mexico, Vol. 9. Oxford: Oxford

University Press.

Wang, R.X. (2014).An illustrated monograph of cherry cultivars in China. Science Press 12: 24–28.

Chen, N.L. (2015) Evaluation of ornamental traits of *Cerasus*species and study on suitable areas of Chinese wild species in Zhejiang. Zhejiang: Zhejiang Science and Technology University.

Bai, F.W., Yu, L., Li, H. J,. et al. (2019). Investigation of *Cerasus* resources on Dawei Mountain. Hunan Forestry Science Technology. 46(6):85-90. doi: 10.3969 /j. issn.1003-5710.2019.06.013

Chen, T., Li, L., Zhang, J., Huang, Z. L., Zhang, H. W., Liu, Y., et al. (2016a). Investigation, collection and preliminary evaluation of genetic resources of Chinese cherry [*Cerasus pseudocerasus* (Lindl.) G. Don]. J. Fruit Sci. 33, 917– 933.doi: 10.13925/j.cnki.gsxb.20150549

Potter, D., Eriksson, T., Evans, R. C., Oh, S., Smedmark, J. E. E., Morgan, D. R., et al.

(2007). Phylogeny and classification of Rosaceae. Plant Syst. Evol. 266, 5–43.doi: 10.1007/s00606-007-0539-9

Chen, T., Hu, G. P., Wang, Y., Chen, Q., Zhang, J., Wang, L., et al. (2020). Survey, collection and conservation of wild *Cerasus* Mill. germplasm resources in China. J. Plant Genet. Resour. 21, 532–541. doi: 10.13430/j.cnki.jpgr.20190716001

Zhang, S. D., Jin, J. J., Chen, S. Y., Chase, M. W., Soltis, D. E., Li, H. T., et al. (2017). Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. New Phytol. 214, 1355–1367.doi: 10.1111/nph.14461

Aranzana, M. J., Decroocq, V., Dirlewanger, E., Eduardo, I., Gao, Z. S., Gasic, K.,et al. (2019). *Prunus* genetics and applications after de novo genome sequencing: achievements and prospects. Hortic. Res. 6:58. doi: 10.1038/s41438-019-0140-8

Chen, F., Song, Y. F., Li, X. J., Chen, J. H., Mo, L., Zhang, X. T., et al. (2019). Genome sequences of horticultural plants: past, present, and future. Hortic. Res.6:112. doi: 10.1038/s41438-019-0195-6

Yu, L., Li, H. J., Bai, F, W,. et al. (2019). Investigation and Utilization Evaluation of Wild *Cerasus* Resources in Daxiong Mountain of Hunan Province. Forestry and Environmental Science.35(6):38-43.

Nybom, H., Weising, K., and Rotter, B. (2014). DNA fingerprinting in botany: past, present, future. Invest. Genet. 5:1. doi: 10.1186/2041-2223-5-1

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., et al.(1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. 18, 489–522.doi: 10.1146/annurev.es.18.110187.002421

Avise, J. C., (2000)Phylogeography: The History and Formation of Species[M]. Phylogeo ography: The History and Formation of Species.

Vander Wall, S. B., and Beck, M. J. (2012). A comparison of frugivory and scatter-hoardin seed- dispersal syndromes. Bot Rev. 78, 10–31. doi: 10.1007/s12229-011-9093-9

Perdereau AC, Kelleher CT, Douglas GC, Hodkinson TR.(2014). High levels of gene flow and genetic diversity in Irish populations of *Salix caprea* L. inferred from chloroplast and nuclear SSR markers. BMC Plant Biol.14:202. doi: 10.1186/s12870-014-0202-x.

Chan, L. M., Brown, J. L., and Yoder, A. D. (2011). Integrating statistical genetic and geospatial methods brings new power to phylogeography. Mol. Phylogenet. Evol. 59, 523–537. doi: 10.1016/j.ympev.2011.01.020

Phillips, S., Dudik, M., Schapire, R. (2004). A maximum entropy approach to species distribution modeling. ACM, 2004.

Turkoglu, Z., Bilgener, S., Ercisli, S., *et al* . (2010). Simple sequence repeat-based assessment of genetic relationships among*Prunus* rootstocks. Genetics & Molecular Research. 9(4): 2156-2165. doi: 10.4238/vol9-4gmr957

Bai, N. W., Zhang, Y. D. (2014). Current status and future directions in plant phylogeography. 26(2):125-137. doi: 10.13376/j.cbls/2014020

Yi, G. X., Yu, X., Chen, J., *et al* . (2020). The genome of Chinese flowering cherry (*Cerasus serrulata*) provides new insights into*Cerasus* species. Horticulture Research, 7(1):14. doi: 10.1038/s41438-020-00382-1

Qiu, J., Zhu. H., Chen, X. (2018) Modeling the suitable areas and ecological characteristics of *Sorbus alnifolia* using DIVA-GIS software. Journal of Beijing Forestry University. 40(9):25-32.

Zhu, H., You, L. X., Li, F.Y.(2017) Modeling the Geographical Distribution Pattern and Climatic Limited Factors of *Cerasus schneideriana.* Journal of Tropical and Subtropical Botany. 25(4):315-322. doi: 10.11926/jtsb.3702

Garah, K., Bentouati, A. (2019). Using the MaxEnt model for assessing the impact of climate change on the Aurasian *Aleppo pine*distribution in Algeria[J]. African Journal of Ecology, 57(4):500-511. doi:10.1111/aje.12630

Heckenhauer, J., Barfuss, M., Samuel, R. (2016). Universal multiplexable*matK* primers for DNA barcoding of angiosperms. Applications in Plant Sciences. 4(1),14. doi: 10.3732/apps.1500137

Feng, Y., Liu, T., Wang, X.Y., *et al* . (2017) Characterization of the complete chloroplast genome of the Chinese cherry *Prunus pseudocerasus* , (Rosaceae). Conservation Genetics Resources, 2 , 1-4. doi: 10.1007/s12686-017-0770-9

Yi, G. X.(2018). Study on population variation and phylogeography of*Cerasus serrulata* [D]. Nanjing: Nanjing Forestry University.

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol.30, 772–780. doi: 10.1093/molbev/mst010

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.doi: 10.1093/bioinformatics/btp187

Nei, M. (1987). Molecular Evolutionary Genetics. New York, NY: Columbia University Press.

Excoffier, L., and Lischer, H. E. (2010). Arlequin suite version 3.5: a new series of

programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Clement, M. Podada, D., and Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657–1659.doi: 10.1046/j.1365-294x.2000.01020.x

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by

DNA polymorphism. Genetics 123, 585–595. doi: 10.1093/genetics/123.3.585

Harpending, H. C. (1994). Signature of ancient population growth in a low- resolution mitochondrial DNA mismatch distribution. Hum. Biol. 66, 591–600.

Drummond, A. J., and Rambaut, A. (2007). BEAST: bayesian evolutionary analysis

by sampling trees. BMC Evol. Biol. 7:214.doi: 10.1186/1471-2148-7-214

Zhang, S. D., Jin, J. J., Chen, S. Y., Chase, M. W., Soltis, D. E., Li, H. T., et al. (2017). Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. New Phytol. 214, 1355–1367. doi: 10.1111/nph.14461

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. doi: 10.1093/molbev/msu300

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9:772. doi:10.1038/nmeth.2109

Zhang, J., Wang Y., Chen T., et al. (2021) Evolution of Rosaceae Plastomes Highlights Unique *Cerasus* Diversification and Independent Origins of Fruiting Cherry. Frontiers in Plant Science, 12. doi: 10.3389/fpls.2021.736053

Wehr, W. C., and Hopkins, D. Q. (1994). The Eocene orchards and gardens of Republic, Washington. Washington Geol. 22, 27–34.

Xiang, Y. Z., Huang, C. H., Hu, Y., Wen, J., Li, S. S., Yi, T. S., et al. (2017). Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. Mol. Biol. Evol. 34, 262–281. doi:10.1093/molbev/msw242

Li, Y., Smith, T., Liu, C. J., Awasthi, N., Y ang, J., Wang, Y. F., et al. (2011). Endocarps of *Prunus* (Rosaceae: Prunoideae) from the early Eocene of Wutu, Shandong Province, China. Taxon 60, 555–564. doi: 10.1002/tax.602021

Chin, S. W., Shaw, J., Haberle, R., Wen, J., and Potter, D. (2014). Diversification of almonds, peaches, plums and cherries - molecular systematics and biogeographic history of *Prunus* (Rosaceae). Mol. Phylogenet. Evol. 76, 34–48. doi: 10.1016/j.ympev.2014.02.024

Rambaut, A. (2009). FigTree 1.3.1. Available online at: http://tree.bio.ed.ac.uk/software/figtree

Chin, S. W., Shaw, J., Haberle, R., Wen, J., and Potter, D. (2014). Diversification of almonds, peaches, plums and cherries - molecular systematics and biogeographic history of Prunus (Rosaceae). Mol. Phylogenet. Evol. 76, 34–48.doi: 10.1016/j.ympev.2014.02.024

Hodel, R., Zimmer, E., Wen, J. (2021) A phylogenomic approach resolves the backbone of *Prunus* (Rosaceae) and identifies signals of hybridization and allopolyploidy. Molecular Phylogenetics and Evolution. 107118. doi: 10.1016/j.ympev.2021.107118

Petit, R. J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D., and Vendramin, G. G. (2005). Invited review: comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol. Ecol. 14, 689–701. doi: 10.1111/j.1365-294X.2004.02410.x

Varvio, S. L., Chakraborty, R., and Nei, M. (1986). Genetic variation in subdivided populations and conservation genetics. Heredity 57, 189–198. doi: 10.1038/hdy.1986.109

Qiu, Y. X., Guan, B. C., Fu, C. X., and Comes, H. P. (2009). Did glacials and/or interglacials promote allopatric incipient speciation in East Asian temperate plants? Phylogeographic and coalescent analyses on refugial isolation and divergence in Dysosma versipellis. Mol. Phylogenet. Evol. 51, 281–293. doi: 10.1016/j.ympev.2009.01.016

Zhang, Q., Chiang, T., George, M., *et al* . (2010). Phylogeography of the Qinghai-Tibetan Plateau endemic *Juniperus przewalskii*(Cupressaceae) inferred from chloroplast DNA sequence variation. Molecular Ecology,14(11):3513-3524.doi: 10.1111/j.1365-294X.2005.02677.x

Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. Biol. J. Linn. Soc. 68, 87–112. doi: 10.1111/j.1095-8312.1999.tb01160.x

Wang, Y. H., Jiang, W. M., *et al* . (2015) Molecular phylogeography and ecological niche modelling of a widespread herbaceous climber,*Tetrastigma hemsleyanum* (Vitaceae): insights into Plio-Pleistocene range dynamics of evergreen forest in subtropical China. The New Phytologist. 206(2):852-867. doi: 10.1111/nph.13261

Qiu, X.Y., Lu, X. Q., Zhang, H, Y., Cao, N. Y. (2017). Phylogeography of East Asia's Tertiary relict plants: current progress and future prospects. Biodiversity Science. 25(2):136-146. doi:10.17520/biods.2016292

Arenas, M., Ray, N., Currat, M., and Excoffier, L. (2012). Consequences of range contractions and range shifts on molecular diversity. Mol. Biol. Evol. 29, 207–218.doi: 10.1093/molbev/msr187

Mona, S., Ray, N., Arenas, M., and Excoffier, L. (2014). Genetic consequences of habitat fragmentation during a range expansion. Heredity 112, 291–299. doi: 10.1038/hdy.2013.105

Alvarado-Serrano, D. F., and Knowles, L. L. (2014). Ecological niche models in phylogeographic studies: applications, advances and precautions. Mol. Ecol.Resour. 14, 233–248.doi: 10.1111/1755-0998.12184

FIGURES

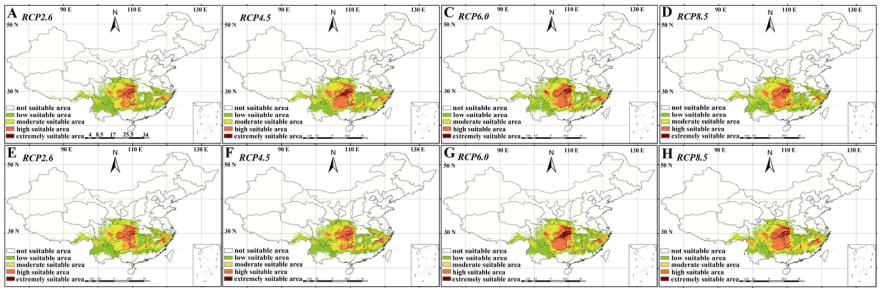

FIGURE 1 | (A-D) Potential adaptive area of *C.conradinae* under the scenarios of RCP2.6, RCP4.5, RCP6.0 and RCP8.5 in 2050s based on MaxEnt model; (E-H) Potential adaptive area of *C.conradinae* under the scenarios of RCP2.6, RCP4.5, RCP6.0 and RCP8.5 in 2070s based on MaxEnt model.

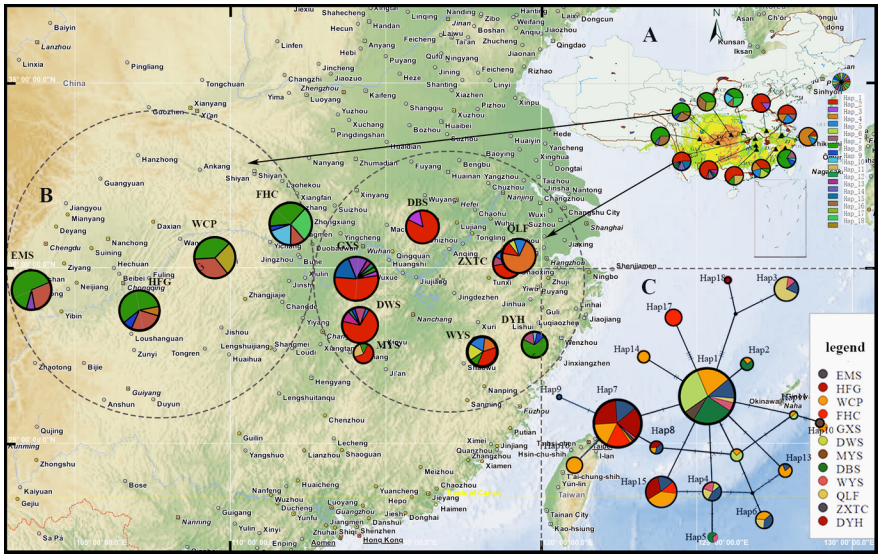

FIGURE 2 | (A) The distribution range of *C.conradinae in China* and Potential habitat for *C.conradinae* in China under current climate. (B) Location of the 12 natural populations and geographic distribution of 18 cpDNA haplotypes (H1-H18). The radius of the pie charts is proportional to the number of individuals sampled. (C) TCS network showing the relationship of haplotypes. The haplotypes are indicated by circles and the colors correspond with the color of the haplotypes in all populations. The size of each pie chart is proportional to the frequency of corresponding haplotype.

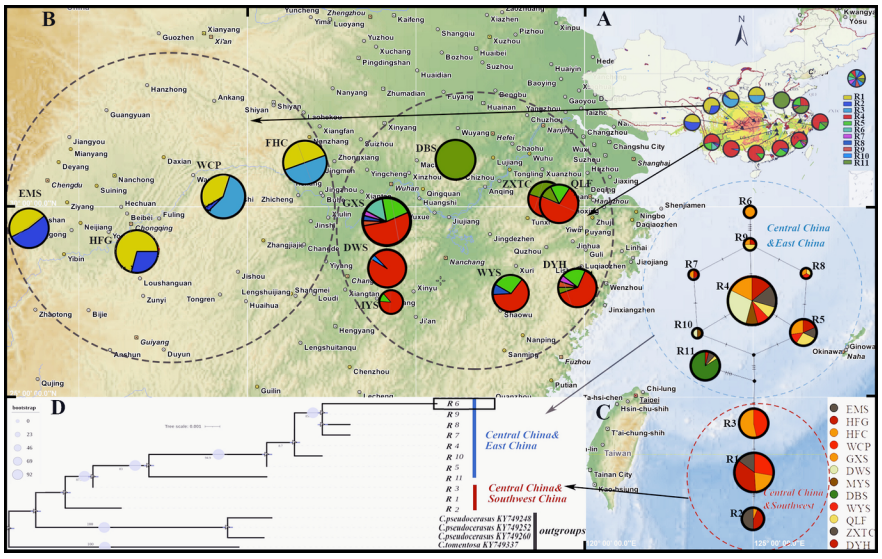

FIGURE 3 | (A) The distribution range of*C.conradinae in China* and Potential habitat for*C.conradinae* in China under current climate. (B) Location of the 12 natural populations and geographic distribution of 11 nrDNA Ribotypes (R1-R11). The radius of the pie charts is proportional to the number of individuals sampled. (C) TCS network showing the relationship of Ribotypes. The Ribotypes are indicated by circles and the colors correspond with the color of the Ribotypes in all populations. The size of each pie chart is proportional to the frequency of corresponding Ribotypes. (D) Bayesian phylogenetic tree of Ribotypes size based on nrDNA(ITS) sequences.

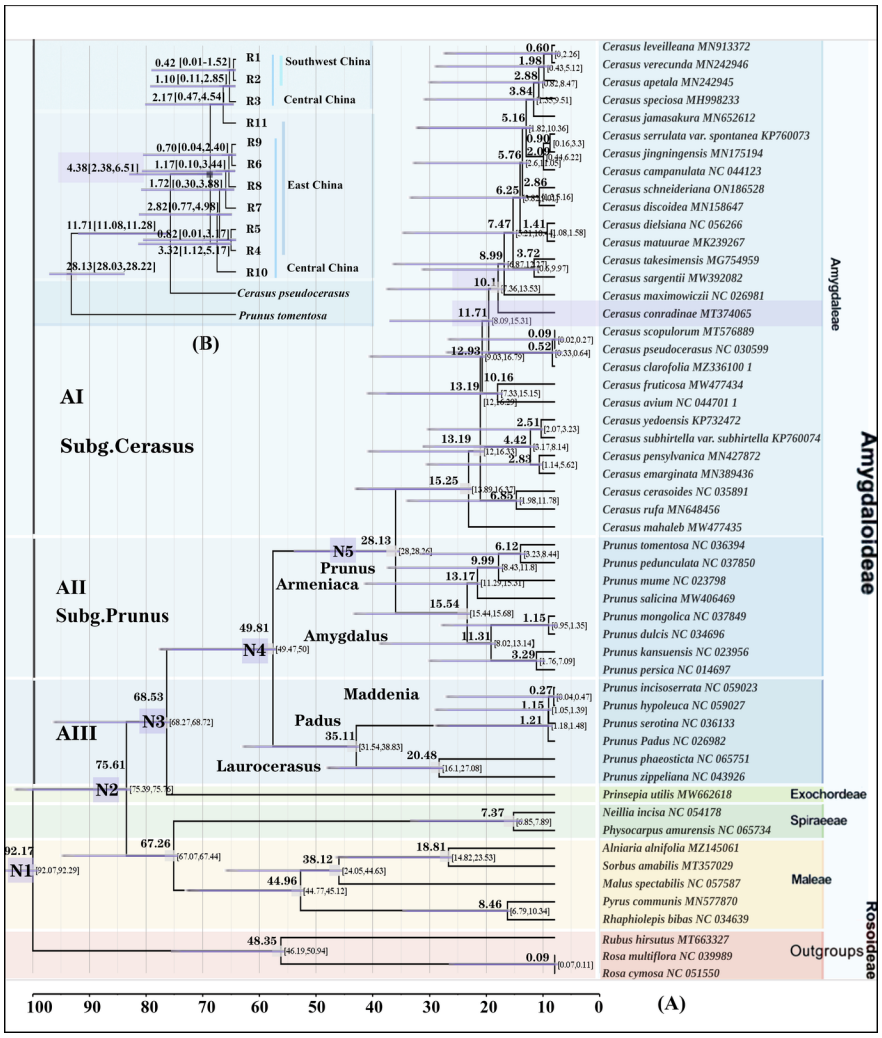

FIGURE 4 | (A) The whole chloroplast genome was used to calculate the differentiation time of different Rosaceae based on 5 differentiation time nodes(B)Construction of nrdna (its) ribosomal divergence time of *C.conradinae* based on two differentiation time nodes

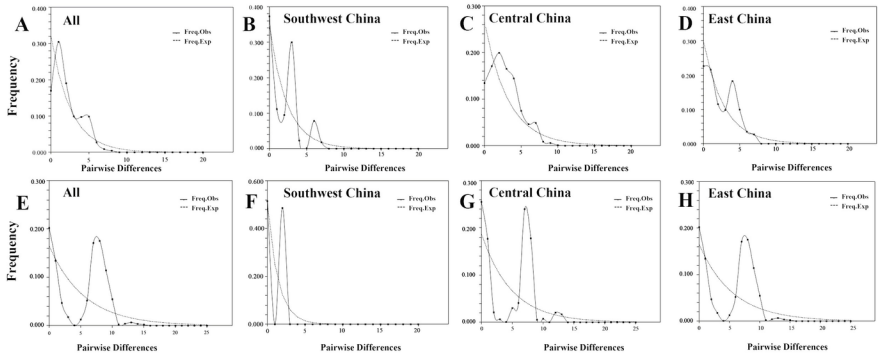

FIGURE 5 | (A-D) Mismatch distribution map of different geographical groups (lineages) of *C.conradinae*

based on nrDNA fragment; (E-H) Mismatch distribution map of different geographical groups (lineages) of *C.conradinae* based on cpDNA fragment.

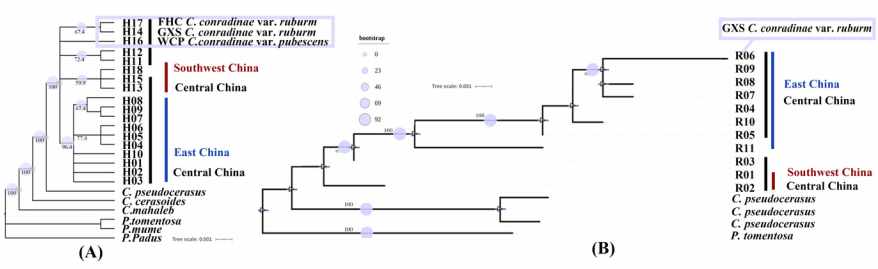

FIGURE 6 | (A) Cladograms of *Cerasus conradinae*haplotypes based on cpDNA(*MatK,TrnD-E,TrnL-F*) sequence; (B) Cladograms of *C.conradinae* ribotypes based on cpDNA(*ITS*) sequence.

TABLES

TABLE 1 | Voucher information and geographic characteristics of 12 *C.conradinae* populations.

Taxa	Code	Country	Location	GPS	Altitude/m
C.conradinae	EMS	CHINA	Mount Emei, Leshan City, Sichuan Province	103.4680 E 29.5770 N	680
	HFG	CHINA	Heifeng Valley, Heishan Town, Chongqing City	106.9800 E 28.8690 N	580
	DWS	CHINA	Dawei Mountain, Changsha City, Hunan Province	114.1640 E 28.3620 N	692
	MYS	CHINA	Mingyue Mountain, Yichun City, Jiangxi Provin	114.2970 E 27.5870 N	1100
	DBS	CHINA	Dabie Mountain, Lu 'an City, Anhui Province	116.2010 E 31.1300 N	890
	WYS	CHINA	Wuyi Mountain, Nanping City, Fujian Province	117.9630 E 27.6680 N	980
	QLF	CHINA	Qingliangfeng Peak, Lin 'an City, Zhejiang Province	118.9140 E 30.1150 N	1050
	ZXTC	CHINA	Tianchi Lake, West Zhejiang, Lin 'an City, Zhejiang Province	119.1270 E 30.3000 N	1200
	DYH	CHINA	Dayang Lake in Lishui City, Zhejiang Province	119.7480 E 27.8740 N	1230

C.conradinae var. *ruburm*	FHC	CHINA	Phoenix Pool, Yichang City, Hubei Province	111.8470 E 31.1440 N	780
	GXS	CHINA	Gexian Mountain , Xianning City, Hubei Province	114.0710 E 29.6380 N	680
C.conradinae var. *pubescens*	WCP	CHINA	Wangcheng slope, Enshi Autonomous Prefecture, Hubei Province	109.4520 E 30.3480 N	850

TABLE 2 | Fossils and molecular estimation used as calibration points for molecular dating.

Node	Mean values/standard deviation used at calibration points	References
N1# Rosaceae Crown	90.18/0.05	Zhang et al., 2017
N2# (Tribe Maleae + Tribe Spiraeeae) + Tribe Amygdaleae	75.62/0.05	Zhang et al., 2017; Zhang et al., 2021
N3# Tribe Amygdaleae	68.58 /0.01	Wehr and Hopkins, 1994; Xiang et al., 2017; Zhang et al., 2021
N4# Node *Prunus*	55.0/0.09	Li et al., 2011; Chin et al., 2014
N5# Node *Cerasus*	28.21/0.05	Zhang et al., 2021

TABLE 3 | Distribution of haplotypes (Ribotypes) in *C.conradinae* among individuals, populations, and glaciated/unglaciated regions of China.

Haplotypes	Number of individuals	Frequencies in individuals(%)	Number of populations	Frequencies in populations(%)	Geographical distributions
Hap1	74	30.33%	7	58.33%	GXS DWS MYS DBS WYS QLF ZXTC
Hap2	4	1.64%	1	8.33%	GXS
Hap3	17	6.97%	3	25.00%	WYS QLF ZXTC
Hap4	9	3.69%	4	33.33%	DWS WYS QLF ZXTC
Hap5	3	1.23%	2	16.67%	WYS QLF
Hap6	8	3.28%	3	25.00%	DWS ZXTC DYH
Hap7	60	24.59%	8	66.67%	EMS HFG WCP FHC GXS MYS WYS DYH
Hap8	6	2.46%	4	33.33%	HFG WCP FHC GXS
Hap9	6	2.46%	2	16.67%	FHC DYH

Haplotypes	Number of individuals	Frequencies in individuals(%)	Number of populations	Frequencies in populations(%)	Geographical distributions
Hap10	2	0.82%	1	8.33%	MYS
Hap11	2	0.82%	2	16.67%	GXS DWS
Hap12	4	1.64%	2	16.67%	GXS DWS
Hap13	5	2.05%	1	8.33%	GXS
Hap14	5	2.05%	1	8.33%	GXS
Hap15	23	9.43%	4	33.33%	EMS HFG WCP FHC
Hap16	7	2.87%	1	8.33%	WCP
Hap17	7	2.87%	1	8.33%	FHC
Hap18	2	0.82%	1	8.33%	HFG
Ribotypes	Number of individuals	Frequencies in individuals(%)	Number of populations	Frequencies in populations(%)	Geographical distributions
Rib 1	47	19.75%	4	33.33%	EMS HFG WCP FHC
Rib 2	16	6.72%	3	25.00%	EMS HFG WCP
Rib 3	28	11.76%	2	16.67%	WCP FHC
Rib 4	85	35.71%	7	58.33%	GXS DWS MYS WYS QLF ZXTC DYH
Rib 5	22	9.24%	6	50.00%	GXS MYS WYS QLF ZXTC DYH
Rib 6	4	1.68%	1	8.33%	GXS
Rib 7	2	0.84%	2	16.67%	GXS DYH
Rib 8	3	1.26%	3	25.00%	GXS WYS QLF
Rib 9	4	1.68%	3	25.00%	GXS WYS DYH
Rib10	2	0.84%	1	8.33%	MYS
Rib11	25	10.50%	4	33.33%	DBS QLF ZXTC DYH

TABLE 4 | Genetic characteristics of 12 *C.conradinae* populations studied.

Population code	Population code	Pop. size	H_d	$P_i \times 10^{-3}$	Haplotypes/ Ribo-types(no.of individuals)
Genetic diversity parameters of sampled populations and their chloroplast genes	**Genetic diversity parameters of sampled populations and their chloroplast genes**	**Genetic diversity parameters of sampled populations and their chloroplast genes**	**Genetic diversity parameters of sampled populations and their chloroplast genes**	**Genetic diversity parameters of sampled populations and their chloroplast genes**	**Genetic diversity parameters of sampled populations and their chloroplast genes**

1	EMS	1 4	0.538	0.680	H7(9)H13(1)H15(4)
2	HFG	26	0.683	0.960	H7(15)H8(2)H15(7)H18(
3	WCP	26	0.686	0.810	H7(10)H15(9)H16(7)
4	FHC	26	0.785	1.780	H7(10)H8(1)H9(5)H15(3
5	GXS	28	0.680	0.540	H1(15)H2(1)H7(1)H11(1
6	DWS	30	0.687	0.950	H1(21)H4(1)H6(3)H8(1)
7	MYS	8	0.607	0.690	H1(5)H7(1)H10(2)
8	DBS	21	0.257	0.230	H1(18)H2(3)
9	WYS	11	0.836	1.113	H1(4)H3(2)H4(2)H5(2)H
10	QLF	17	0.500	1.070	H1(2)H3(12)H4(2)H5(1)
11	ZXTC	18	0.699	1.010	H1(9)H3(3)H4(4)H6(2)
12	DYH	19	0.526	0.700	H6(3)H7(13)H8(2)H9(1)
Southwest China	Southwest China	40	0.627	1.050	H7(24) H8(2)H13(1)H15(11) H18(2)
Central China	Central China	110	0.866	1.240	H1(41)H2(1)H7(21)H8(2 H13(4)H14(5)H15(12) H16(7)H17(7)
East China	East China	94	0.774	1.150	H1(36)H2(1)H4(1)H6(3)
Mean	Mean		0.623	0.878	
All	All	244	0.830	1.04	
Genetic diversity parameters of sampled populations and their nuclear genes (ITS)	**Genetic diversity parameters of sampled populations and their nuclear genes (ITS)**	**Genetic diversity parameters of sampled populations and their nuclear genes (ITS)**	**Genetic diversity parameters of sampled populations and their nuclear genes (ITS)**	**Genetic diversity parameters of sampled populations and their nuclear genes (ITS)**	**Genetic diversity parameters of sampled populations and their nuclear genes (ITS)**
1	EMS	15	0.533	0.186	R1(7)、R2(8)
2	HFG	24	0.431	0.150	R1(17)、R2(7)
3	WCP	26	0.538	0.115	R1(10)、R2(1)、R3(15)
4	FHC	26	0.520	0.091	R1(13)、R3(13)
5	GXS	28	0.667	0.325	R4(15)、R5(6)、R6(4)、
6	DWS	25	0.080	0.014	R4(24)、R10(1)
7	MYS	8	0.250	0.044	R4(7)、R5(1)
8	DBS	21	0.000	0.000	R11(21)
9	WYS	11	0.564	0.120	R4(7)、R5(3)、R8(1)
10	QLF	17	0.728	0.311	R4(7)、R5(6)、R8(1)、
11	ZXTC	18	0.464	0.368	R4(13)、R5(3)、R11(2)
12	DYH	19	0.579	0.269	R4(12)、R5(4)、R7(1)、
Southwest China	Southwest China	39	0.486	0.169	R1(24)、R2(15)
Central China	Central China	105	0.745	0.756	R1(23)、R2(1)、R3(28) R6(4)、R7(1)、R8(1)、
East China	East China	94	0.666	0.669	R4(46)、R5(17)、R7(1)
All	All	238	0.798	0.886	R1(24)、R2(15)

TABLE 5 | Analyses of molecular variance (AMOVAs) based on cpDNA and nrDNA data for populations of *C.conradinae*

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	Fixation Indices	Fixation Indices	Gst/Nst	Gst/Ns
chloroplast DNA fragments	**chloroplast DNA fragments**	**chloroplast DNA fragments**	**chloroplast DNA fragments**	**chloroplast DNA fragments**	**chloroplast DNA fragments**	**chloroplast DNA fragments**	**chloroplast DNA fragments**	**chloro DNA frag ments**
All groups	**All groups**							
Among populations	11	1560.531	6.69964 Va	48.89	Fst=0.48886	Fst=0.48886	0.28176 /0.29843 (P<0.05)	0.28176 /0.2984 (P<0.0
Within populations	232	1625.161	7.00501 Vb	51.11				
Southwest China	**Southwest China**							
Among populations	1	0.402	0.1090 Va	11.01	Fst=0.08126	Fst=0.08126	0.072/0.081 (p<0.05)	0.072/0 (p<0.0
Within populations	38	36.651	0.96451 Vb	88.99				
Central China	**Central China**							
Among populations	3	27.065	0.29098 Va	22.04	Fst=0.22038	Fst=0.22038	0.18051/0.22810 (P<0.05)	0.18051 (P<0.0
Within populations	106	109.116	1.02940 Vb	77.96				
East China								
Among populations	5	1173.349	14.16449Va	46.05	Fst=0.46046	Fst=0.46046	0.30473/0.33970 (P<0.05)	0.30473 (P<0.0
Within populations	88	1460.534	16.59698Vb	53.95				
Southwest & Central & East	**Southwest & Central & East**	**Southwest & Central & East**						
Among regions	2	359.203	0.42291 Va	3.06	FSC= 0.47782	FSC= 0.47782	FSC= 0.47782	FSC= 0.47782
Within regions	9	1201.328	6.41002 Vb	46.32	FST=0.49378	FST=0.49378	FST=0.49378	FST=0
Within populations	232	1625.161	7.00501 Vc	50.62	FCT=0.03056	FCT=0.03056	FCT=0.03056	FCT=0
nuclear DNA fragments	**nuclear DNA fragments**	**nuclear DNA fragments**	**nuclear DNA fragments**	**nuclear DNA fragments**	**nuclear DNA fragments**	**nuclear DNA fragments**	**nuclear DNA fragments**	**nuclea DNA frag ments**
All groups								
Among populations	11	499.251	2.28341 Va	82.51	Fst= 0.74918	0.37339 /0.74924 (P<0.05)	0.37339 /0.74924 (P<0.05)	0.37339 /0.7492 (P<0.0
Within populations	226	109.380	0.48398 Vb	17.49				

Southwest China								
Among populations	1	1.213	0.03707 Va	6.55	Fst=0.06157	0.03221 /0.06157 (p<0.05)	0.03221 /0.06157 (p<0.05)	0.03221 /0.0615 (p<0.05
Within populations	37	19.556	0.52853 Vb	93.45				
Central China								
Among populations	3	255.125	3.23420 Va	94.38	Fst=0.85663	0.38164/0.8574	0.38164/0.8574	0.38164
Within populations	101	19.451	0.19258 Vb	5.62		(P<0.05)	(P<0.05)	(P<0.05
East China								
Among populations	5	126.356	1.60432 Va	73.73	Fst=0.73734	0.29616/0.5598	0.29616/0.5598	0.29616
Within populations	88	50.293	0.57151 Vb	26.27		(P<0.05)	(P<0.05)	(P<0.05
Southwest & Central & East	**Southwest & Central & East**							
Among regions	2	197.775	0.46342 Va	16.51	F_{SC}=0.67382			
Within regions	12	577.426	1.57873 Vb	56.25	F_{ST}=0.72768			
Within populations	461	352.317	0.76424 Vc	27.23	F_{CT}=0.16513			

TABLE 6 | Summary of cpDNA and nrDNA-based divergence time estimation by Bayesian

Node	**Median v**
N1: # Rosaceae(Subfamily Amygdaloideae + Subfamily Rosoideae)	92.17
N2: # (Tribe Maleae + Tribe Spiraeeae) + Tribe Amygdaleae	75.61
Tribe Maleae + Tribe Spiraeeae	67.26
Tribe Maleae	44.96
N3: #*Prinsepia* + ((*Laurocerasus*+(*Padus* + *Maddenia*))+ (*Amygdalus* + (*Prunus* + *Armeniaca*))+*Cerasus*)	68.53
N4: # (*Laurocerasus*+(*Padus* + *Maddenia*))+ (*Amygdalus* + (*Armeniaca* + *Prunus*))+*Cerasus*	49.81
Laurocerasus+(*Padus* + *Maddenia*)	35.11
N5: # (*Amygdalus* + (*Armeniaca* + *Prunus*))+*Cerasus*	28.13
1.*Cerasus mahaleb*+*Cerasus*	15.25
2.(*C. clarofolia* + (*C. pseudocerasus* + *C. scopulorum*)) + *C. conradinae*	11.71
3. *C. conradinae*	4.38
4. *C. conradinae* (Central China + East China)	3.32
5. *C. conradinae* (East China +(Central China+Southwest China))	2.17
6. *C. conradinae* (Central China + Southwest China)	1.1

TABLE 7 | Neutrality and population expansion tests for

C.conradinae

Groups	Tajima's D (P-value)	Fu's Fs (P-value)	**Demographic expansion** (SSD) (P-value)	**Demographic expansion** Raggedness index (P-value)	**Spatial expansion** (SSD) (P-value)	**Spatial expansion** Raggedness index (P-value)
Based on the detection results of chloroplast DNA fragments	**Based on the detection results of chloroplast DNA fragments**	**Based on the detection results of chloroplast DNA fragments**	**Based on the detection results of chloroplast DNA fragments**	**Based on the detection results of chloroplast DNA fragments**	**Based on the detection results of chloroplast DNA fragments**	**Based on the detection results of chloroplast DNA fragments**
12 populations 244 individuals	0.09900	0.10000	0.38800	0.44200	0.26900	0.43600
Southwest China	0.26400	0.64200	0.14900	0.10000	0.19300	0.27900
Central China	0.24400	0.08000	0.12400	0.39100	0.25400	0.43400
East China	0.40800	0.50900	0.68500	0.84200	0.55700	0.87500
Based on the detection results of nuclear DNA fragments	**Based on the detection results of nuclear DNA fragments**	**Based on the detection results of nuclear DNA fragments**	**Based on the detection results of nuclear DNA fragments**	**Based on the detection results of nuclear DNA fragments**	**Based on the detection results of nuclear DNA fragments**	**Based on the detection results of nuclear DNA fragments**
12 populations 238 individuals	0.35291	N.A.	0.17772	0.03300	0.08500	0.12093
Southwest China	1.00000	0.95300	0.32324	0.00000	0.12400	0.32879
Central China	0.98600	0.79100	0.12764	0.02900	0.21900	0.17021
East China	0.62164	N.A.	0.28450	0.11800	0.56100	0.39557