

Using genome-wide SSR analysis to understand the genetic diversity and structure of 174 *Akebia trifoliata* accessions

Wei Chen

Sichuan Agricultural University - Chengdu Campus

Huai Yang

Sichuan Agricultural University - Chengdu Campus

Hao Yang

Sichuan Agricultural University - Chengdu Campus

Shengfu Zhong

Sichuan Agricultural University - Chengdu Campus

Qing Dong

Sichuan Agricultural University - Chengdu Campus

Chen Chen

Sichuan Agricultural University - Chengdu Campus

Tianheng Ren

Sichuan Agricultural University - Chengdu Campus

Zhi Li

Sichuan Agricultural University - Chengdu Campus

Feiquan Tan

Sichuan Agricultural University - Chengdu Campus

Luo Peigao (✉ lpglab@sicau.edu.cn)

Sichuan Agricultural University - Chengdu Campus <https://orcid.org/0000-0002-4603-6960>

Research Article

Keywords: *Akebia trifoliata*, Genetic diversity, Heng Duan Mountains, SSR, Origin centre

Posted Date: December 9th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-2307089/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Genetic Resources and Crop Evolution on May 16th, 2023. See the published version at <https://doi.org/10.1007/s10722-023-01602-y>.

Abstract

Genetic diversity is usually regarded as an important role for understanding the origin, evolution and environmental adaptation of species. A total of 174 living collections of *Akebia trifoliata* and 72 high-polymorphism-potential (HPP) simple sequence repeat (SSR) markers were employed to systematically explore the diversity and structure of this species. The average number of alleles (N_a), Shannon's information index (I), number of effective alleles (N_e), observed heterozygosity (H_o), expected heterozygosity (H_e) and polymorphism information content (PIC) values were 9.01, 1.64, 4.98, 0.56, 0.72 and 0.69, respectively, and the linkage disequilibrium (LD) locus pairs percentage was on 1.17%. The PIC value of *AK78* was the highest (0.92), while that of *AK53* was the lowest (0.21). Sixty of 72 loci exhibited high polymorphism, while only one exhibited low polymorphism. In addition, the average PIC of chromosomes varied from 0.56 to 0.79. The genetic diversity of the SCB population was the highest, while that of NQH was the lowest. Further analysis of SCB suggested that JSJ-MJ, in the main part of the Heng Duan Mountain zone, had the highest genetic diversity and could be the origin centre of *A. trifoliata*. The main migration route of *A. trifoliata* could have been latitudinally from western to eastern China along the Yangtze River. Structure, cluster and principal coordinate analyses yielded similar results, revealing that the 6 populations could be divided into 2 groups or clusters. The geographic origin of several unknown accessions, such as Alc-166, Alc-171, Alc-172 and Alc-174, was revealed by cluster analysis. Overall, the results suggested that the living collections of *A. trifoliata* had high genetic diversity, good representativeness, and low duplication, which will benefit *A. trifoliata* germplasm collection, preservation, and breeding applications.

Introduction

The terrain of China is generally high in the west and low in the east, presenting a three-level descending trend (Wang et al., 2020). The Sichuan Basin, as the intermediate zone (the second level) between the Qinghai-Tibet Plateau (the first level) and the middle-lower reaches of the Yangtze River Plain (the third level), is composed of plateaus, mountains, valleys, hills and plains, with substantial altitude variation and complex climate types (Wang et al., 2020), suggesting the potential for high species diversity because multiple rich habitats usually accompany biological diversity.

The Heng Duan Mountain zone, west of the Sichuan Basin, is one of 34 global biodiversity hotspots (Hughes, 2017). Previous investigations revealed thousands of animals and plants in this area. For example, studies of vascular plants alone reported 8711 species of angiosperms, 89 species of gymnosperms and 708 species of ferns from 1600 genera of 230 families (López-Pujol et al., 2006; Lu et al., 2020). Moreover, many species, such as *Osmanthus serrulatus* (Chen et al., 2021), *Paeonia decomposita* (Chen et al., 2019) and *Impatiens chashanensis* (Bi et al., 2009), are endemic to this area. In fact, variation not only exists among species but also within the same species due to geographical barriers and different environmental conditions (Van Rossum et al., 2020). Therefore, studying the genetic diversity and population structure of a given species, especially endemic species, is valuable for understanding species origin, evolution, and distribution.

Akebia trifoliata ($2x = 32$), belonging to Lardizabalaceae, is a woody perennial liana (Du et al., 2012) and is mainly distributed in areas south of the Yellow River, such as Sichuan, Hunan, Hubei, Guizhou, Chongqing, Shanxi, and Yunnan, and is widespread in more than 10 provinces (Li et al., 2010). Economically, *A. trifoliata*

has major prospects for commercial exploitation with multiple purposes because it can be used as an edible fruit and oil crop (Zhao et al., 2014; Yang et al., 2021) as well as a medicinal and ornamental plant (Li et al., 2010; Wang et al., 2015). Evolutionarily, it could play an important role in elucidating angiosperm, especially eudicot, divergence because it is a basal eudicot and is often used as a representative of this group (Liu et al., 2010). Ecologically, this species is widely distributed in East Asia, mainly in China, Korea, and Japan (Wang et al., 2020), and especially in China (Li et al., 2010). Thus, systemically evaluating genetic diversity and completely understanding the population structure of *A. trifoliata* would help accelerate its practical exploitation, shed light on its evolutionary adaptation and aid in local ecosystem rehabilitation.

The genetic diversity and population structure of *A. trifoliata* could be evaluated based on various traits, such as phenotypic parameters and nutritional indices, but molecular markers are the most widely employed tools in this field (Guan et al., 2022). For example, DNA-based molecular markers such as sequence-related amplified polymorphism (SRAP) (Zhang et al., 2014, 2021), amplified fragment length polymorphism (AFLP) (Zhang et al., 2016), inter-simple sequence repeat (ISSR) (Zhang et al., 2021) and simple sequence repeat (SSR) markers (Li et al., 2019; Niu et al., 2019; Zhong et al., 2021; Chen et al., 2022) have been employed to estimate the genetic diversity of different *A. trifoliata* populations consisting of various individuals from different geographic regions of China, suggesting that the use of molecular markers to evaluate the population structure and genetic diversity of *A. trifoliata* has gradually become a research hotspot in theoretical as well as applied fields.

SSRs are the most widely employed markers in the population analysis of *A. trifoliata*. A few reports considered SSRs in *A. trifoliata*, but the numbers of materials were small (Li et al., 2009; Niu et al., 2019), or the markers lacked physical positions (Niu et al., 2019; Zhang et al., 2021). A total of 36160 high-polymorphism-potential (HPP) SSR markers were designed according to the high-quality assembled *A. trifoliata* genome, and they were also physically mapped to all 16 chromosomes of *A. trifoliata* (Zhong et al., 2022), which provided a set of powerful molecular tools for genetic improvement by molecular-assisted selection, elucidating environmental adaptations and protecting germplasm.

In addition, *A. trifoliata* is not only a typical woody perennial plant (Yu et al., 2021) but also an obligatory outcrossing species (Zou et al., 2019). Therefore, how to collect and preserve *A. trifoliata* should be given additional attention during the process of screening and establishing a valuable repository of plant genetic diversity because the biological characteristics, such as life history, seed biology and propagation methods, of *A. trifoliata* are inherently different from those of herbaceous annual species. Practice has shown that ex situ living conservation is an important strategy for applying the germplasm collections of woody perennial plants (Migicovsky et al., 2019).

In the past, various authors attempted to collect *A. trifoliata* and evaluate its genetic diversity, making great accomplishments in the field. A total of 62, 111, and 242 wild accessions from the Qinba Mountain area were reported in Zhang et al. (2014, 2016, 2021), respectively; 88 wild accessions were collected from the Yangtze River region (Niu et al., 2019); and more than 300 leaf samples of wild individuals in the Yangtze River region were analysed (Li et al., 2019). In addition, another study examined approximately 955 wild accessions, but their geographic origin was unclear (Zhong et al., 2021). Moreover, 43 cultivated *A. trifoliata* accessions were also collected for research (Zhang et al., 2021). Even though both genetic diversity and population structure

have been well evaluated and analysed in this species, the studies had shortcomings, such as small sample sizes (Zhang et al., 2014, 2021; Niu et al., 2019), sampling of narrow regions (Zhang et al., 2014, 2016, 2021), unclear geographic origins (Niu et al., 2019) and unavailability for breeding improvement (Li et al., 2019). Unfortunately, information about the distribution and density of *A. trifoliata* germplasm around the Heng Duan Mountain region is limited. Undoubtedly, collecting *A. trifoliata* germplasm from a broader geographic region, especially from biodiversity hotspots, and centralizing the diverse material at a defined location would be particularly valuable for theoretical studies and effective commercial exploitation of this species.

In this study, 174 *A. trifoliata* accessions selected from the centralized preservation site and 100 pairs of SSR markers of the dinucleotide repeat (DNR) type covering the 16 chromosomes of *A. trifoliata* were employed to understand the genetic diversity and population structure of the accessions. These results are helpful for identifying the geographic origin centre of this species and screening breeding germplasm with economic traits of interest. Moreover, the data and conclusions will benefit *A. trifoliata* germplasm collection, preservation, and breeding applications.

Materials And Methods

Germplasm accession selection

A total of 174 accessions of *A. trifoliata* were collected from more than 3,000 ex situ living germplasm collections in the Germplasm Nursery at Sichuan Agricultural University Chongzhou Research Station (30°43'N, 103°65'E; Zhong et al., 2022a, 2022b); the selection factors included plant growth or health status, representativeness of different ecological regions, distribution density of the corresponding wild resource and a large geographic distance to avoid duplication due to clonal propagation. The details of all the accessions are shown in Table S1, and these accessions could be divided into six populations: Sichuan Basin (SCB), Yunnan-Guizhou Plateau (YG), south of the Qinglin Huaihe line (SQH), north of the Qinglin Huaibei line (NQH), Qinba Mountains (QB) and unknown geographic origin (UN), consisting of Alc-1 to Alc-78, Alc-79 to Alc-97, Alc-98 to Alc-145, Alc-146 to Alc-151, Alc-152 to Alc-163 and Alc-164 to Alc-174, respectively. For example, a total of 78 accessions were selected from SCB because it is one of 34 biodiversity hotspots in the world, and 15 accessions (19.2%) were from a narrow region—Shimian County of Sichuan Province—with large altitudinal variation covering approximately 180 degrees around the Dadu River (Fig. 1).

Genomic DNA extraction

Total genomic DNA was extracted from 10 ~ 15 young, fresh and healthy (Guan et al., 2022). The quality and concentration of the DNA were determined on 1% (w/v) agarose gels by comparison with known concentrations of uncut lambda DNA standard. The isolated genomic DNA was diluted to 10 ng/μL with 0.1x TE buffer (1 mmol/L Tris-HCl, 0.1 mmol/L EDTA, pH = 8.0) and stored at -80°C for SSR amplification according to previously described methods (Zhong et al., 2022).

Primer selection and PCR amplification

The previously reported 100 pairs of SSR markers physically mapped to all 16 chromosomes of *A. trifoliata* (Zhong et al., 2022a, 2022b) were used in the present study. DNA amplifications were performed in a PTC-200

thermocycler. Each PCR mixture contained each SSR primer at a concentration of 200 nmol/L, 0.2 mmol/L dNTPs, 1.5 mmol/L MgCl₂, 1 unit of Taq polymerase, and 60 ng of template DNA. PCR was performed as follows: 94°C for 1 min, followed by 30 cycles of 94°C for 45 seconds, 55–60°C for 30 seconds, and 72°C for 30 seconds and a final amplification step of 10 min at 72°C. Then, 4 µl of each PCR product was mixed with 2 µl of loading buffer and separated by capillary electrophoresis.

SSR data analysis

Polymorphic bands were scored as “1” for present or “0” for absent. The total number of different alleles (N_a), number of effective alleles (N_e), observed and expected heterozygosity (H_o and H_e , respectively), Shannon’s information index (I), Inter-individual fixation index (F_{is}), Total population fixation index (F_{it}), coefficient of genetic differentiation (F_{st}) and percentage of polymorphic loci (PPL) were calculated and principal coordinate analysis (PCoA) was performed by GenAlEx 6.5 (Peakall and Smouse, 2006). The polymorphism information content (PIC) was calculated by Power Marker 3.25 (Liu and Muse, 2005) and a phylogenetic tree was constructed using the website “iTOL” (<https://itol.embl.de/>). Pairwise genetic distance and Nei’s genetic identity were determined by POPGENE32 (Yeh et al., 2000), and the model-based software program STRUCTURE v2.3.4 was used to infer population structure by a Bayesian approach. The threshold of $Q = 0.8$ was adopted to judge whether an individual was a migrant accession, where we calculated gene penetration as the number of migrant accessions divided by the size of the given group. The optimal K value was deduced by evaluating $K = 1 \sim 7$. The genetic ancestry of each individual was estimated based on the admixture model; estimates were obtained with the Markov chain Monte Carlo (MCMC) method with 100,000 iterations followed by a burn-in period of 100,000 iterations. The best K values were identified based on a previously described method (Evanno et al., 2005) in STRUCTURE HARVESTER. TASSEL2.1 software was used to calculate the level of linkage disequilibrium (LD) and support probability among different SSR loci combinations. The trend line of the LD value decay with genetic distance was simulated by nonlinear algebraic regression (Bradbury et al., 2007).

Results

SSR amplification and polymorphism

Among the previously published 100 HPP SSR markers designed according to DNRs (Zhong et al., 2022), 96 and 83 produced clear bands and polymorphic bands in a small population consisting of 25 accessions separated by large geographic distances, respectively (Guan et al., 2022). According to the brightness, distinctness and distribution of polymorphic bands, 72 polymorphic SSR markers were further selected for population analysis.

All 72 polymorphic SSR markers physically assigned to all chromosomes of *A. trifoliata* produced at least one band and two bands in a given accession and in the population, respectively. A total of 649 alleles at 72 SSR loci were obtained in the population consisting of 174 accessions. The number of alleles per SSR locus was 9.01 on average and varied from 2 at AK4 and AK99 to 23 at AK66 (Table S3). The number of different alleles among 21 (29.2%) loci was less than or equal to five, while that among 24 (33.3%) loci was more than ten. The average I , N_e , H_o , H_e and PIC values were 1.64, 4.98, 0.56, 0.72 and 0.69, ranging from 0.48 to 2.64, 1.28 to 12.67, 0.07 to 0.95, 0.22 to 0.92 and 0.21 to 0.92, respectively (Table S3), which indicated high genetic diversity. The genetic diversity of AK53 was the lowest, for which the I , N_e , H_o , H_e and PIC values were 0.48, 1.28, 0.21,

0.22 and 0.21, respectively. In contrast, that of AK78 was the highest, with I , N_e , H_o , H_e and PIC values of 2.62, 12.67, 0.79, 0.92 and 0.92, respectively. We further artificially classified all 72 markers into three groups based on the number of different alleles produced by them: class I, class II and class III, including 21 (fewer than or equal to 5), 27 (more than 5 and fewer than or equal to 10) and 24 (more than 10) markers, respectively. Statistical analysis revealed that almost all differences in I , N_e , H_o , H_e and PIC among them were significant at the $P = 0.05$ level, and only that in H_o between class I and class II was not significant at that level (Table S4), which showed that generally, the larger the number of alleles produced by the marker was, the higher the genetic diversity of the marker.

In addition, according to the recommended standard (Botstein et al., 1980), 60 loci (83.3%) exhibited high polymorphism, while only one (AK53, 1.4%) locus exhibited low polymorphism. There were 11 (15.3%) moderately polymorphic loci, including both AK4 and AK99, which produced only two bands in the population. We also found that the average PIC exhibited a large degree of variation from 0.56 on chromosome 16 to 0.79 on chromosome 12 (Fig. 2). There were six chromosomes (chromosomes 1, 3, 4, 11, 12 and 13) with an average PIC greater than 0.70 and three chromosomes (chromosomes 2, 14 and 16) with an average PIC less than 0.60. AK53, with the lowest PIC (0.21), was assigned to chromosome 9, and AK78, with the highest PIC (0.92), was assigned to chromosome 13. These results showed that genetic diversity was different among chromosomes or chromosomal regions (Fig. 2).

Differences in genetic diversity among geographic populations

The percentage of polymorphic loci (PPL) and the average N_a , I , N_e , H_o , H_e and PIC values as well as the detailed geographic data of *A. trifoliata* living collections from different regions are shown in Table 1. Six parameters (except N_e and H_o) were the highest in SCB, while all of the parameters were lowest in NQH; moreover, the differences in both N_e and H_o were very small between SCB and SQH. Generally, all geographic populations had high genetic diversity, possibly resulting from the underexploited status of the species, and the order of genetic diversity from high to low among the different populations was SCB, SQH, YG, QB, UN and NQH, which indicated that SCB, with the largest N_a (7.76), I (1.54), H_e (0.70), PIC (0.66) and PPL (1), could be the origin centre of *A. trifoliata*. In addition, both H_o and H_e were larger than 0.50 in all geographic regions except NQH (Table 1), indicating high heterozygosity. Nei's genetic identity and pairwise genetic distance were the largest (0.81) and the smallest (0.21) between SCB and SQH, respectively, while they were 0.57 between SCB and NQH, which was the smallest Nei's genetic identity and the largest pairwise genetic distance (Table 2).

Table 1
Genetic diversity indices of 6 *A. trifoliata* geographic populations

Population	Size	Longitude	Latitude	Area	Na	Ne	I	Ho	He	PIC	PPL
SCB	78	99.99 ~ 109.27	26.49 ~ 32.57	56.42	7.76	4.50	1.54	0.56	0.70	0.66	1.00
YG	19	100.83 ~ 109.14	24.42 ~ 28.55	34.32	5.24	3.57	1.25	0.54	0.62	0.60	0.94
SQH	48	109.17 ~ 118.98	25.28 ~ 31.85	64.45	7.28	4.52	1.51	0.57	0.69	0.65	0.99
NQH	6	111.22 ~ 113.64	34.72 ~ 35.14	1.016	2.93	2.42	0.84	0.45	0.46	0.56	0.74
QB	12	104.96 ~ 109.50	32.34 ~ 35.64	13.5	4.63	3.38	1.17	0.50	0.58	0.61	0.89
UN	11	/	/	/	4.13	3.02	1.05	0.51	0.52	0.57	0.83

Note. Na: No. different alleles, Ne: No. effective alleles, I: Shannon's information index, Ho: Observed heterozygosity, He: Expected heterozygosity, PIC: Polymorphism information content, PPL: Percentage of polymorphic loci, SCB: Sichuan Basin, YG: Yunnan-Guizhou Plateau, SQH: South of the Qinglin Huaihe line, NQH: South of the Qinglin Huaihe line, QB: Qinba Mountains, UN: Unknown origin.

Table 2
Pairwise genetic distances among 6 *A. trifoliata* subpopulations

Population	SCB	YG	SQH	NQH	QB	UN
SCB	****	0.68	0.81	0.57	0.66	0.69
YG	0.38	****	0.77	0.62	0.74	0.73
SQH	0.21	0.27	****	0.62	0.74	0.75
NQH	0.57	0.48	0.48	****	0.68	0.65
QB	0.42	0.30	0.30	0.39	****	0.76
UN	0.37	0.31	0.29	0.43	0.27	****

Note: Nei's genetic identity appears above the diagonal, and pairwise genetic distance appears below the diagonal; the other information is the same as in Table 1.

To further narrow the geographic region of the origin centre of *A. trifoliata*, we divided SCB into four geographic subregions according to the main branches of the Yangtze River and the river itself: the south of the Yangtze River (YZR-S); the region between Jinshajiang and the Mingjiang River (JSJ-MJ), which is the core zone of the Heng Duan Mountains region; the zone between Mingjing and the Jialingjiang River (MJ-JLJ); and the zone between Jialingjiang and Dabashan (JLJ-DBS). Comparative analysis revealed that germplasms from JSJ-MJ had the highest Na (6.68), Ne (4.27), I (1.46), He (0.67), PIC (0.64) and PPL (0.99) (Table S5), and the diversity index of the germplasms from MJ-JLJ was the second highest, while that from YZR-S was the smallest (Table

S5). Among the four subpopulations, JSJ-MJ and MJ-JLJ showed the largest (0.86) and the smallest (0.15) Nei's genetic identity and pairwise genetic distance, respectively, while those between JLJ-DBS and YZR-S were the smallest (0.58) and the largest (0.54), respectively (Table S6). The results indicated that JSJ-MJ could indeed be the origin centre of *A. trifoliata*.

Structure analysis

The ΔK values calculated by the previously described methods showed a maximum at $K = 2$, and the second highest peak was at $K = 4$ (Fig. 3).

When K equalled 2, the 174 accessions were divided into two groups (Group A and Group B) (Table 3). Group A consisted of 91 samples (75 accessions, 15 accessions and 1 accession from SCB, SQH and UN, respectively), and Group B consisted of 83 samples (33, 19, 12, 10, 6, and 3 accessions from SQH, YG, QB, UN, NQH, and SCB, respectively). The Q values of four genotypes (Alc-4, Alc-10, Alc-9 and Alc-18) in Group A and two genotypes (Alc-79 and Alc-139) in Group B were lower than 0.8; therefore, gene penetration (4.40%) in Group A was slightly stronger than that in Group B (2.41%). Likewise, when K equalled 4, the accessions could be divided into 4 groups (Group 1, Group 2, Group 3 and Group 4). Group 1 was the smallest group, consisting of 32 accessions from SCB, and Group 2 was the largest group, comprising 59 accessions (43 accessions, 15 accessions and 1 accession from SCB, SQH and UN, respectively). Group 3 consisted of 38 accessions (12, 10, 6, 4, 4 and 2 accessions from YG, UN, QB, SQH, NQH and SCB, respectively), and Group 4 consisted of 45 accessions (29 accessions, 7 accessions, 6 accessions, 2 accessions and 1 accession from SQH, YG, QB, NQH and SCB, respectively). Further comparison showed that Group A was a combination of Group 1 and Group 2, and Group B was a combination of Group 3 and Group 4. Gene penetration among the 4 groups was ranked as follows: Group 2 (16.95%) > Group 4 (11.11%) > Group 3 (10.34%) > Group 1 (6.25%) (Table 3).

Table 3
Genetic diversity indices for *A. trifoliata* accessions with $K = 2$ and 4

Population	K value	Size	Na	Ne	I	Ho	He	PPL	PIC
Group A	K = 2	91	7.93	4.49	1.54	0.56	0.7	1	0.66
Group B		83	7.58	4.38	1.48	0.55	0.67	0.97	0.63
Group 1	K = 4	32	6.15	3.90	1.37	0.59	0.65	0.97	0.61
Group 2		59	6.60	3.85	1.34	0.53	0.62	0.94	0.60
Group 3		38	5.72	3.50	1.21	0.49	0.57	0.96	0.61
Group 4		45	6.69	4.12	1.39	0.56	0.64	0.88	0.61

Note: the information is the same as in Table 1

Cluster analysis

The phylogenetic tree (Fig. 4) constructed by PowerMarker 3.25 software showed that when the clade length was 0, all 174 accessions could also be divided into two clusters: Cluster 1 and Cluster 2, containing 90 and 84 accessions, respectively. Cluster 1 had a simple structure and consisted of 90 accessions (71 accessions, 17 accessions, 1 accession and 1 accession from SCB, SQH, YG and UN, respectively), while Cluster 2 had a

relatively complex structure and consisted of 84 accessions (31, 18, 10, 12, 7 and 6 accessions from SQH, YG, UN, QB, SCB and NQH, respectively). Similarly, when the branch length was 0.0044, both Cluster 1 and Cluster 2 could be further subdivided into two subclusters. Cluster 1–1 consisted of only two accessions: one from UN and one from SQH; however, Cluster 1–2 was composed of 88 accessions (71 and 17 accessions from SCB and SQH, respectively). Cluster 2 – 1 was composed of 39 accessions (26, 7, 4 and 2 accessions from SQH, YG, QB and SCB, respectively). Cluster 2–2 was composed of 45 accessions (11, 10, 8, 6, 5 and 5 accessions from YG, UN, QB, NQH, SQH and SCB, respectively) (Fig. 3).

PCoA analysis

The results of the PCoA showed that the first three principal coordinates together were responsible for 35.5% of the variation, accounting for 16.06%, 12.46% and 6.93%, respectively. According to the PCoA, the SQH accessions had the widest distribution of all the *A. trifoliata* accessions from the 6 geographical regions. Moreover, these accessions overlapped with those of the other 5 populations. The accessions from SCB largely overlapped with those from SQH but were clearly distinct from the other four geographical subpopulations and densely distributed in the first quartiles. The YG, NQH, QB and UN accessions were distributed in the second and third quartiles (Fig. 5a). When information from clustering based on STRUCTURE analysis was integrated into the PCoA scatter plot, Group 1 and Group 2 were distributed in the first, third and fourth quartiles, while Group 3 and Group 4 were distributed in the second and third quartiles. Groups 3 and 4 and Groups 1 and 2 overlapped, but Groups 1 and 2 were clearly distinct from Groups 3 and 4. Furthermore, adding Group 1 to Group 2 recreated Group A, and similarly, Group B consisted of Group 3 and Group 4 (Fig. 5b).

Linkage disequilibrium analysis

Seventy-two SSR markers produced 151 linked and 2405 non-linked locus pairs in the population, and 21 (13.91%) linked and 220 (9.15%) non-linked pairs had a significant LD at $p < 0.01$ level (Table S7). In addition, the number of locus pairs in LD with both $r^2 > 0.1$ and $p < 0.01$ was only four (2.65%) and 26 (1.08%) in the linked and non-linked locus pairs, respectively. This suggested that the level of LD across all accessions was low.

Discussion

Elucidating the genetic diversity of a given species would provide a theoretical reference for both the protection and utilization of that resource (Chan and Sun., 1997), and DNA-based molecular markers such as SSRs and AFLPs are powerful tools for analysing genetic diversity. Compared with AFLPs, SSRs are simpler, with a lower cost and higher polymorphism (Julier et al., 2003); therefore, SSR technology is widely used for crop gene mapping (Huang et al., 2014), population genetics (Emanuelli et al., 2013), DNA fingerprinting (Ercisli et al., 2011), varietal identification (Harisha et al., 2021) and molecular marker-assisted breeding (Prasanna et al., 2010). Although SSRs have been employed to estimate the genetic diversity of different *A. trifoliata* populations by various authors (Li et al., 2019; Niu et al., 2019; Zhong et al., 2021), there are two main shortcomings in these studies: fewer or uncertain chromosomal positions of SSR markers and the use of fragmented populations or populations with narrow geographic origins.

Variation in the richness of polymorphic information among different markers

In a recent report, 56 (70.8%) and 16 (20.2%) of 79 SSR markers successfully amplified products and produced clear polymorphic bands in a winged bean (*Psophocarpus tetragonolobus* L.) population consisting of 457 accessions, respectively (Laosatit et al., 2022). However, in the present study, 96 and 83 of 100 SSR markers produced clear bands and polymorphic bands in a small population consisting of 25 accessions, respectively (Zhong et al., 2022). On the one hand, SSR polymorphisms possibly have a large degree of variation among different species (Fan et al., 2013). On the other hand, we prefer to accept the view that SSR polymorphism could be related to the repeat type (Moxon and Wills, 1999; Everett and Wood, 2004).

The wide application of SSRs in crop gene mapping is mainly because of their evolutionarily neutral characteristics and high polymorphism (Luo et al., 2005). Compared with trinucleotide repeats (TNRs), DNRs experience relative little evolutionary selection pressure or functional optimization and therefore exhibit greater neutrality and more polymorphism (Everett and Wood, 2004). In addition, SSR polymorphism is affected by SSR length, where the longer the SSR is, the higher the SSR polymorphism (Singh et al., 2010). Thus, SSRs longer than 50 bp commonly show high polymorphism potential. In the present study, 72 DNR and HPP SSR markers that were physically assigned to 16 chromosomes were ultimately used to evaluate the genetic diversity of an *A. trifoliata* population, and their average PIC was up to 0.69 (Table S3), obviously higher than the previously reported PICs of 0.43 (Niu et al., 2019) and 0.41 (Zhong et al., 2021). Moreover, there was abundant variation in PIC not only among different markers (from 0.21 for AK53 to 0.92 for AK78) but also among different chromosomes (from 0.56 for chromosome 16 to 0.79 for chromosome 12) (Table S3 and Fig. 2). The number of HPP SSRs with a PIC larger than 0.50 was 60 (83.3%), while the PIC of the low-polymorphism markers was less than 0.25. The PIC was only one among 72 SSRs (Table S3). Comprehensively, the observed N_a as well as all calculated I , N_e , H_o , H_e and PIC values suggested that the 72 SSR markers generally exhibited high polymorphism and could be reliable and ideal markers for population analysis.

Large differences in genetic diversity among *A. trifoliata* germplasms from different geographic regions

According to traditional Chinese ecological regions, in the present study, the 174 accessions were divided into six populations: SCB, YG, SQH, NQH, QB and UN (Table S1). The results of the comparative analysis showed that SCB had the largest I (1.54) and PIC (0.66), which suggested that SCB had the richest genetic diversity and could be the centre of origin. Two recent studies showed that in *A. trifoliata* germplasm, there was a significant positive correlation between genetic distance and geographic distance (Li et al., 2019; Zhang et al., 2021). In our study, compared with SQH, both QB and YG had a smaller geographic interval but exhibited a larger genetic distance from SCB (Fig. 4). In addition, the genetic characteristics of NQH were similar to those of QB (Table 1 and Fig. 4). The climate factors such as rainfall and sunlight of SQH were similar to those of SCB, and both NQH and QB also had similar climate factors. Therefore, we assume that SCB could be the centre of *A. trifoliata* origin, which could be explained by both multiple geomorphological and climate types in SCB (Wang et al., 2020), and climate similarity rather than geographic distance could be an important factor influencing *A.*

trifoliata migration. In addition, the other indices, especially He, also provided some persuasive evidence supporting this view.

To further narrow the scope of the potential *A. trifoliata* origin centre, four geographic subregions, namely, YZR-S, JSJ-MJ, MJ-JLJ and JLJ-DBS, of SCB were divided according to the main branches of the Yangtze River. We found that JSJ-MJ had the largest I (1.46) and PIC (0.64) (Table 2), exhibiting the highest genetic diversity. Thus, it could be the origin centre of *A. trifoliata*, which was well supported by the fact that the Heng Duan Mountain zone mainly including JSJ-MJ is a global biodiversity hotspot (Hughes, 2017). Finally, based on the change in genetic diversity among different geographic regions and subregions, there was evidence that the latitudinal migration of *A. trifoliata* from West to East China along the Yangtze River could have been the major route.

Structure and cluster analyses could be useful methods for identifying the geographic origin of unknown accessions

Previous studies suggested that directional selection had reduced the genetic diversity (Santiago and Caballero., 1998) and exhibited a positive correlation with the level of LD (Mallet et al., 1990) in many crops such as wheat (Hao et al., 2010), tomato (Van et al., 2008) and maize (Ching et al., 2002). In this study, there only were 30 (1.17%) LD locus pairs among 2556 pairs across 174 accessions, which could be explained by a very short cultivating history of *A. trifoliata* (Guan et al., 2022; Yu et al., 2021) and also indicated that employing the accession to evaluate the genetic diversity of wild *A. trifoliata* germplasms was reasonable.

Although there were some slight differences in numbers and members, the division results of structure and cluster analyses were generally very similar (Figs. 3 and 4) and in good agreement with the pairwise genetic distance and PCoA results (Fig. 5 and Table 2). Thus, these methods were capable of providing information on a given accession of unknown origin.

The collection, preservation, and utilization of accession resources is a systematic, lasting and challenging task (Jackson, 1997), in which the loss of geographical origin data or other information is unavoidable (Singh et al., 2019). However, this information is very important for applications in plant breeding and germplasm preservation. Therefore, determining the possible geographical origin is of great significance for plant breeders. In our study, there were 11 accessions with unknown geographic origins provided by the Chinese Academy of Forestry Sciences (Alc-164, Alc-165, Alc-167, Alc-168, and Alc-169), Jiangxi Academy of Forestry (Alc-166) and various farmers (Alc-170, Alc-171, Alc-172, Alc-173, and Alc-174).

We found that accession Alc-166 was always clustered with Group A and Group 2, while the other 10 UN accessions were also always clustered into Group B and Group 3 in the structure analysis with K = 2 and 4, respectively (Fig. 4). Likewise, Alc-166 was classified into Cluster 1–1, while the other 10 UN accessions were classified into Cluster 2–2 by cluster analysis (Fig. 4). Interestingly, the accessions of unknown geographic origin Alc-166 and Alc-106 from SQH always belonged to the same group (Group A and Group 2 at K = 2 and 4, respectively) (Fig. 4) according to structure analysis and were classified into the same cluster, 1–1, by cluster analysis (Fig. 4). Importantly, Cluster 1–1 consisted of only Alc-166 and Alc-106 (Fig. 4). Therefore, the results provided strong evidence that Alc-166 possibly originated from SQH, where the Jiangxi Academy of Forestry is located.

Similarly, we inferred that Alc-171, Alc-172 and Alc-174 may have originated from YG because they were assigned to a small cluster that included only some accessions from YG with a branch length of 0.0124 (0.0025) (Fig. 4). Evidently, structure and cluster analyses are two helpful methods for inferring the geographic origin of unknown accessions in the germplasm preservation and utilization processes.

Ex situ living collections provide a helpful resource for preserving and exploiting woody perennial plant resources

A. trifoliata is a perennial vine with major prospects for commercial exploitation, having edible fruits, also called augmelon, and value as an oil crop (Zhao et al., 2014; Yang et al., 2021), as well as medicinal and ornamental value (Wang et al., 2015). In situ conservation is a traditional method used to protect wild precious woody perennial plant resources, but it is also usually time-consuming, expensive, inconvenient and less likely to be systemically studied and applied than other methods.

In contrast, ex situ living collections are an effective and valuable tool for conserving and utilizing woody perennial species such as *A. trifoliata* because they could reduce costs, enhance efficiency, avoid trafficking dangers and increase germplasm use in plant breeding programs (Migicovsky et al., 2019). Over approximately ten years, we have constructed an ex situ living collection consisting of more than 3,000 accessions of *A. trifoliata* from more than ten provinces of China (Zhong et al., 2022). In the present study, molecular data analysis of 174 accessions revealed rich genetic density within regions and among different regions (Table 1). The molecular data as well as putative origin and migration information will be valuable for scientifically and commercially evaluating the living collections of this species and for effectively using these resources in breeding programs with specific goals. Therefore, systemically investigating phenotypic characteristics and finally screening some accessions with objective traits of interest for use as parents are priorities for the future.

Conclusions

Seventy-two pairs of polymorphic SSR markers physically mapped to all 16 chromosomes of *A. trifoliata* and 174 living preserved accessions collected from five geographical populations (SCB, SQH, YG, QB and NQH) and provided by various scientists and farmers (UN) were employed to evaluate the genetic density of this species. The genetic density of every population was generally high. Comparative analysis revealed that JSJ-MJ of SCB had the richest genetic density and could be the origin centre of *A. trifoliata*. According to the change in genetic diversity among different regions, *A. trifoliata* may have migrated mainly latitudinally from the west to the east of China along the Yangtze River rather than longitudinally. In addition, we inferred that Alc-171, Alc-172 and Alc-174 may have originated from YG and that Alc-166 originated from SQH by cluster analysis. This research provides global insights into the origin, distribution and migration of *A. trifoliata*.

Declarations

Acknowledgments The authors thank Dr. Bin Li of the Chinese Academy of Forestry Sciences for affording the accessions Alc-75 to Alc-77, Alc-79, Alc-80, Alc-102, Alc-139 to Alc-141, Alc-146 to Alc-148, Alc-154 to Alc-165 and Alc-167 to Alc-169 and Ms. Yujuan Wang of the Jiangxi Academy of Forestry for affording the accessions Alc-101, Alc-131 to Alc-138 and Alc-166.

Author contribution Data curation, Huai Yang and Feiquan Tan; Formal analysis, Hao Yang; Methodology, Zhi Li; Resources, Qing Dong, Chen Chen and Tianhen Ren; Software, Shengfu Zhong; Writing – original draft, Wei Chen; Writing – review & editing, Peigao Luo. All authors reviewed and approved the final submission.

Funding This research was supported by grants from the Science and Technology Department of Sichuan Province, China (2020YFN0091 and 2020YJ0331). This work was also supported by the Science and Technology Department of Sichuan Province and the Bureau of Yaan.

Data availability All data generated or analysed during this study are included as Supplementary Data file.

Conflict of interest The authors declare that there is no conflict.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Bi HY, Hou YT, Zhou XR, Yu SX (2009) *Impatiens chashanensis* (*Balsaminaceae*) sp. nov. from Sichuan, China. *Nord J Bot*, 27: 372-375
2. Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*, 32: 314-331
3. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: Software for Association
4. Mapping of Complex Traits in Diverse Samples. *Bioinformatics*, 23: 2633-2635
5. Chan KF, Sun M (1997) Genetic diversity and relationships detected by isozyme and RAPD analysis of crop and wild species of *Amaranthus*. *Theor Appl Genet*, 95: 865-873
6. Chen W, Yang H, Zhong SF, Z Jun, Zhang QY, Li Z, Ren TH, Tan FQ, Shen JL, Li Q, Luo PG (2022) Expression Pro-files of Microsatellites in Fruit Tissues of *Akebia trifoliata* and Development of Efficient EST-SSR Markers. *Genes*, 13(8):1451
7. Chen L, Pan T, Qian H, Zhang M, Yang G, Wang X (2021) Genetic diversity and population structure revealed by SSR markers on endemic species *Osmanthus serrulatus* rehder from Southwestern Sichuan basin, China. *Forests*, 12: 1365
8. Chen Y, Zhou Q, Sun L, Ng WL, Zhou R, Wu W (2019) The chloroplast genome of *Paeonia decomposita* (*Paeoniaceae*), an endangered wild tree peony from Sichuan, China. *Conserv Genet Resour*, 11: 59-61
9. Ching A, Caldwell KS, Jung M, Dolan M, Smith O, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC genetics*, 3: 1-14
10. Du Y, Jiang Y, Zhu X, Xiong H, Shi S, Hu J, Peng H, Zhou Q, Sun W (2012) Physicochemical and functional properties of the protein isolate and major fractions prepared from *Akebia trifoliata* var. *australis* seed. *Food Chem*, 133: 923-929
11. Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troggio M, Myles S, Martinez-Zapater JM, Zyprian

12. E, Moreira FM, Grando MS (2013) Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grapes. *BMC Plant Biol*, 13: 39
13. Ercisli S, Ipek A, Barut E (2011) SSR marker-based DNA fingerprinting and cultivar identification of olives (*Olea europaea*). *Biochem Genet*, 49: 555-561
14. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol*, 14: 2611-2620
15. Everett CM, Wood NW (2004) Trinucleotide repeats and neurodegenerative disease. *Brain*, 127: 2385-2405
16. Fan L, Zhang MY, Liu QZ, Li LT, Song Y, Wang LF, Zhang SL, Wu J (2013) Transferability of newly developed pear SSR markers to other *Rosaceae* species. *Plant Mol Biol Report*, 31: 1271-1282
17. Guan J, Fu P, Wang X, Yu X, Zhong S, Chen W, Yang H, Chen C, Yang H, Luo PG (2022) Assessment of the breeding potential of a set of genotypes selected from a natural population of *Akebia trifoliata* (three-leaf *Akebia*). *Horticulturae*, 8: 116
18. Hao CY, Perretant MR, Choulet F, Wang LF, Paux E, Sourdille P, Zhang XY, Feuillet C, Balfourier F (2010) Genetic diversity and linkage disequilibrium studies on a 3.1-Mb genomic region of chromosome 3B in European and Asian bread wheat (*Triticum aestivum* L.) populations. *Theor Appl Genet*, 121: 1209-1225
19. Harisha R, Bhadru D, Vanisri S, Gourishanakar V, Satish L (2021) SSR and morphological traits based fingerprints and DNA barcodes for varietal identification in rice. *Biotechnol Biotechnol Equip*, 35: 1461-1473
20. Huang Q, Li X, Chen WQ, Xiang ZP, Zhong SF, Chang ZJ, Zhang M, Zhang HY, Tan FQ, Ren ZL, Luo PG (2014).
21. Genetic mapping of a putative *Thinopyrum intermedium*-derived stripe rust resistance gene on wheat chromosome 1B. *Theor Appl Genet*, 127: 843-853
22. Hughes CE (2017) Are there many different routes to becoming a global biodiversity hotspot? *Proc Natl Acad Sci U S A*, 114: 4275-4277
23. Jackson MT (1997) Conservation of rice genetic resources: the role of the international Rice Genebank at IRRI. *Plant Mol Biol*, 35: 61-67
24. Julier B, Flajoulot S, Barre P, Cardinet G, Santoni S, Huguet T, Huyghe C (2003) Construction of two genetic linkage maps in cultivated tetraploid alfalfa (*Medicago sativa*) using microsatellite and AFLP markers. *BMC Plant Biol*, 3: 9
25. Laosatit K, Amkul K, Chankaew S, Somta P (2022) Molecular genetic diversity of winged bean gene pool in Thailand assessed by SSR markers. *Hortic Plant J*, 8: 81-88
26. Li L, Xu Q, Yao X (2019) Microsatellite analysis reveals the resilience of genetic diversity within extant populations of three *Akebia* species to chronic forest fragmentation in China. *Plant Ecol*, 220: 69-81
27. Li L, Yao X, Chen X, Huang H (2009) Development and characterization of microsatellite loci in Chinese medicinal plant *Akebia trifoliata ssp. australis* and cross-species amplification in closely related taxa. *Conserv Genet*, 10: 959-962
28. Li L, Yao X, Zhong C, Chen X, Huang H (2010) *Akebia*: a potential new fruit crop in China. *HortScience*, 45: 4-10

29. Liu C, Zhang J, Zhang N, Shan H, Su K, Zhang J, Meng Z, Kong H, Chen Z (2010) Interactions among proteins of floral MADS-box genes in basal eudicots: implications for evolution of the regulatory network for flower development. *Mol Biol Evol*, 27: 1598-1611
30. Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, 21: 2128-2129
31. López-Pujol J, Zhang FM, Ge S (2006) Plant biodiversity in China: richly varied, endangered, and in need of conservation. *Biodivers Conserv*, 15: 3983-4026
32. Lu Y, Xu P, Li Q, Wang Y, Wu C (2020) Planning priority conservation areas for biodiversity under climate change in topographically complex areas: a case study in Sichuan province, China. *PLoS One*, 15: e0243425
33. Luo PG, Ren ZL, Zhang HQ, Zhang HY (2005) Identification, chromosome location, and diagnostic markers for a new gene (YrCN19) for resistance to wheat stripe rust. *Phytopathology*, 95: 1266-1270
34. Mallet J, Barton N, Lamas G, Santisteban J, M Muedas, H Eeley (1990) Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics*, 124: 921-936
35. Migicovsky Z, Warschefsky E, Klein LL, Miller AJ (2019) Using living germplasm collections to characterize, improve, and conserve woody perennials. *Crop Sci*, 59: 2365-2380
36. Moxon ER, Wills C (1999) DNA microsatellites: agents of evolution? *Sci Am*, 280: 94-99
37. Niu J, Wang Y, Shi Y, Wang X, Sun Z, Huang K, Gong C, Luan M, Chen J (2019) Development of SSR markers via de novo transcriptome assembly in *Akebia trifoliata* (Thunb.) Koidz. *Genome*, 62: 817-831
38. Peakall ROD, Smouse PE (2006) genalex 6: genetic analysis in excel. Population genetic software for teaching and research. *Mol Ecol Notes*, 6: 288-295
39. Prasanna BM, Pixley K, Warburton ML, Xie CX (2010) Molecular marker-assisted breeding options for maize improvement in Asia. *Mol Breed*, 26: 339-356
41. Santiago E, Caballero A (1998) Effective size and polymorphism of linked neutral loci in populations under directional selection. *Genetics*, 149: 2105-2117
42. Singh H, Deshmukh RK, Singh A, Singh AK, Gaikwad K, Sharma TR, Mohapatra T, Singh NK (2010) Highly variable
43. SSR markers suitable for rice genotyping using agarose gels. *Mol Breed*, 25: 359-364
44. Singh N, Wu S, Raupp WJ, Sehgal S, Arora S, Tiwari V, Vikram P, Singh S, Chhuneja P, Gill BS, Poland J (2019)
45. Efficient curation of genebanks using next generation sequencing reveals substantial duplication of germplasm accessions. *Sci Rep*, 9: 650
46. Van BR, Zhu A, Ursem R, Verbakel H, Gort G, Eeuwijk FAV (2008) Diversity and linkage disequilibrium analysis within a selected set of cultivated tomatoes. *Theor Appl Genet*, 117: 89-101
47. Van Rossum T, Ferretti P, Maistrenko OM, Bork P (2020) Diversity within species: interpreting strains in microbiomes. *Nat Rev Microbiol*, 18: 491-506

48. Wang J, Ren H, Xu QL, Zhou ZY, Wu P, Wei XY, Cao Y, Chen XX, Tan JW (2015) Antibacterial oleanane-type triterpenoids from pericarps of *Akebia trifoliata*. Food Chem, 168: 623-629
49. Wang L, Deng H, Qiu X, Wang P, Yang F (2020) Determining the impact of key climatic factors on geographic distribution of wild *Akebia trifoliata*. Ecol Indic, 112: 106093
50. Wang N, Cheng W, Wang B, Liu Q, Zhou C (2020) Geomorphological regionalization theory system and division methodology of China. J Geogr Sci, 30: 212-232
51. Yang H, Chen W, Fu P, Zhong S, Guan J, Luo PG (2021) Developmental stages of *Akebia trifoliata* fruit based on volume. Hortic Sci Technol, 39: 823-831
52. Yeh F, Yang R, Boyle T, Ye Z, Xiyang J, Yang R, Boyle T (2000) PopGene32, Microsoft Window-Based Freeware for Population Genetic Analysis, Version 1.32. Molecular Biology and Biotechnology Centre, University of Alberta Edmonton, Alberta, Canada.
53. Yu X, Zhong S, Yang H, Chen C, Chen W, Yang H, Guan J, Fu P, Tan F, Ren T, Shen J, Zhang M, Luo PG (2021)
54. Identification and characterization of NBS resistance genes in *Akebia trifoliata*. Front Plant Sci, 12: 758559
55. Zhang Y, Zhang Z, Dong S (2014) Study on optimization of SRAP-PCR reaction system and genetic diversity in *Akebia trifoliata*. J Shaanxi Norm Univ, 42: 1672-4291
56. Zhang Z, Jian JY, Wang Z (2016) AFLP analysis on genetic diversity of *Akebia trifoliata* in Qinling mountains of Shaanxi. Chin Tradit Herb Drugs, 2016: 3890-3895
57. Zhang Z, Yang Q, Niu Y, Zhang Y, Dong S, Zhang W, Wang Z (2021) Diversity analysis and establishment of core collection among *Akebia trifoliata* (Thunb.) Koidz. in Qinba mountain area of China using ISSR and SRAP markers. Genet Resour Crop Evol, 68: 1085-1102
58. Zhao SQ, Hu JN, Zhu XM, Bai CQ, Peng HL, Xiong H, Hu JW, Zhao Q (2014) Characteristics and feasibility of trans- free plastic fats through lipozyme TL IM-catalyzed interesterification of palm stearin and *Akebia trifoliata* variety australis seed oil. J Agric Food Chem, 62: 3293-3300
59. Zhong SF, Chen W, Yang H, Shen J, Ren T, Li Z, Tan F, Luo PG (2022a) Characterization of microsatellites in the *Akebia trifoliata* genome and their transferability and the development of a whole set of effective, polymorphic and physically mapped SSR markers. Front Plant Sci, <https://doi.org/10.3389/fpls.2022.860101>
60. Zhong SF, Li B, Chen W, Wang LL, Guan J, Wang Q, Yang Z, Yang H, Wang XS, Yu XJ, Fu P, Liu HC, Chen C, Tan
61. FQ, Ren TH, Shen JL, Luo PG (2022b) The chromosome-level genome of *Akebia trifoliata* as an important resource to study plant evolution and environmental adaptation in the Cretaceous. Plant J. doi: 10.1111/tbj.16011
62. Zhong Y, Wang Y, Sun Z, Niu J, Shi Y, Huang K, Chen J, Chen J, Luan M (2021) Genetic diversity of a natural population of *Akebia trifoliata* (Thunb.) Koidz and extraction of a core collection using simple sequence repeat markers. Front Genet, 12: 716498
63. Zou S, Yao X, Zhong C, Zhao T, Huang H (2019) Effectiveness of recurrent selection in *Akebia trifoliata* (Lardizabalaceae) breeding. Sci Hortic, 246: 79-85

Figures

● >4 samples ● =4 samples ● =3 samples ● =2 samples

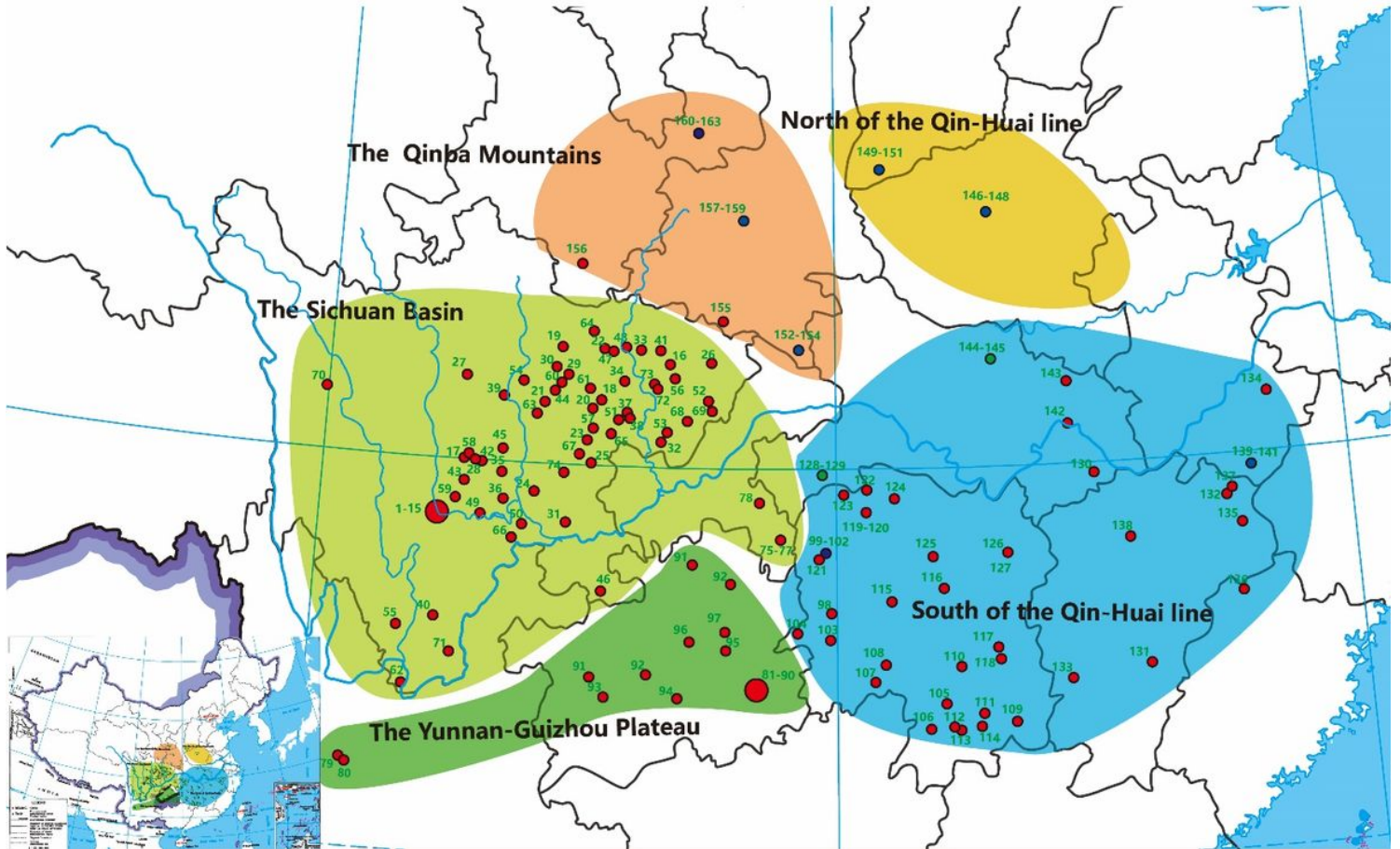


Figure 1

The distribution of the 174 *A. trifoliata* accessions

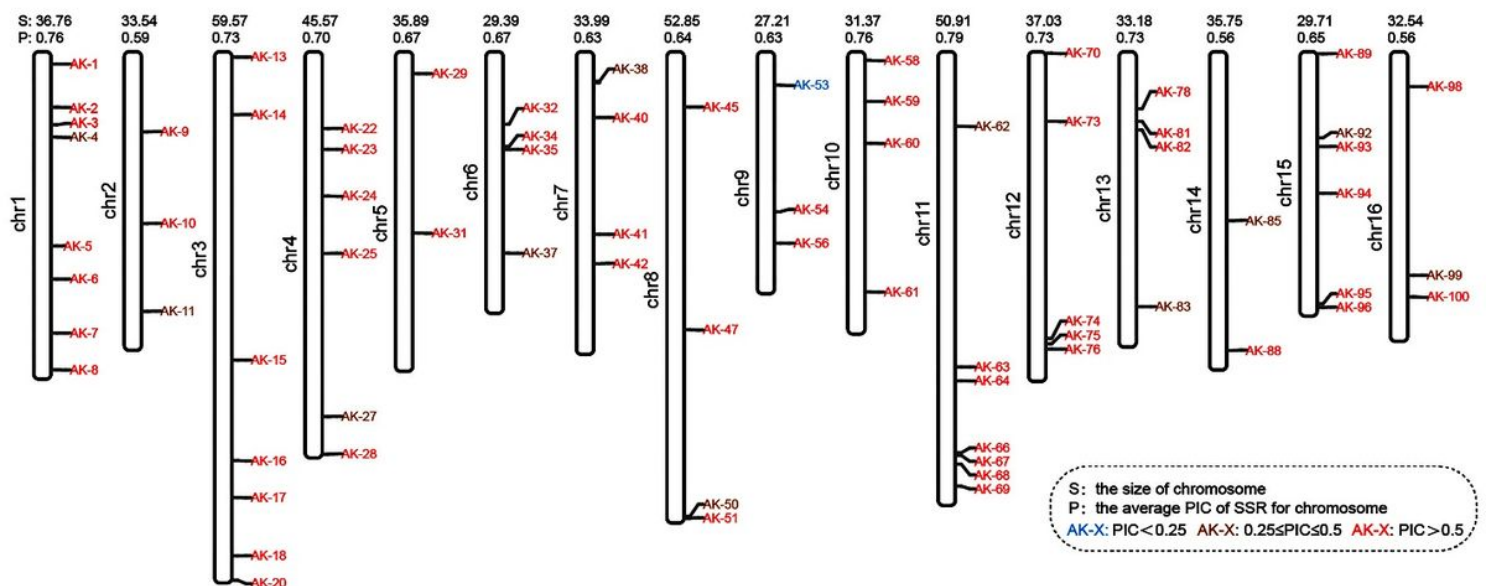


Figure 2

The position of SSRs on the chromosome and their average PIC value for the chromosome

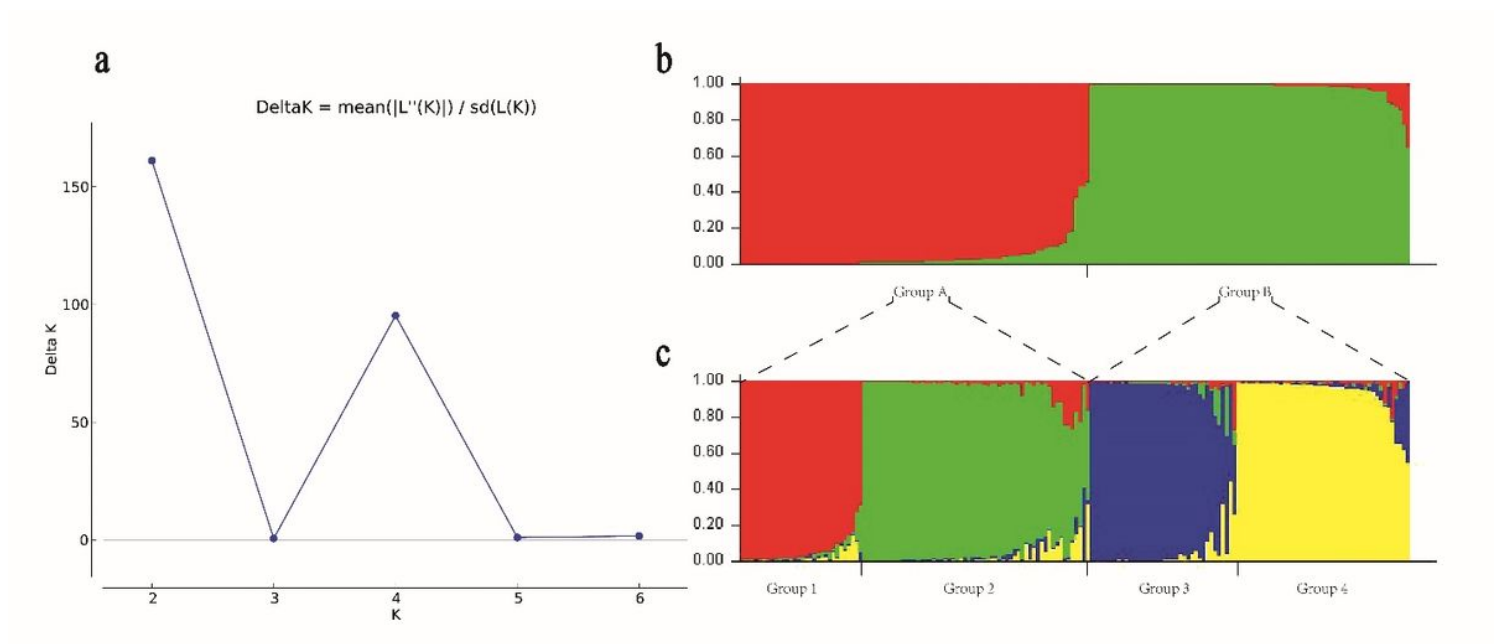


Figure 3

Population structure analysis. (a) Delta K based on the rate of change in L (K) between successive K-values. (b) Population structure based on K = 2. (c) Population structure based on K = 4.

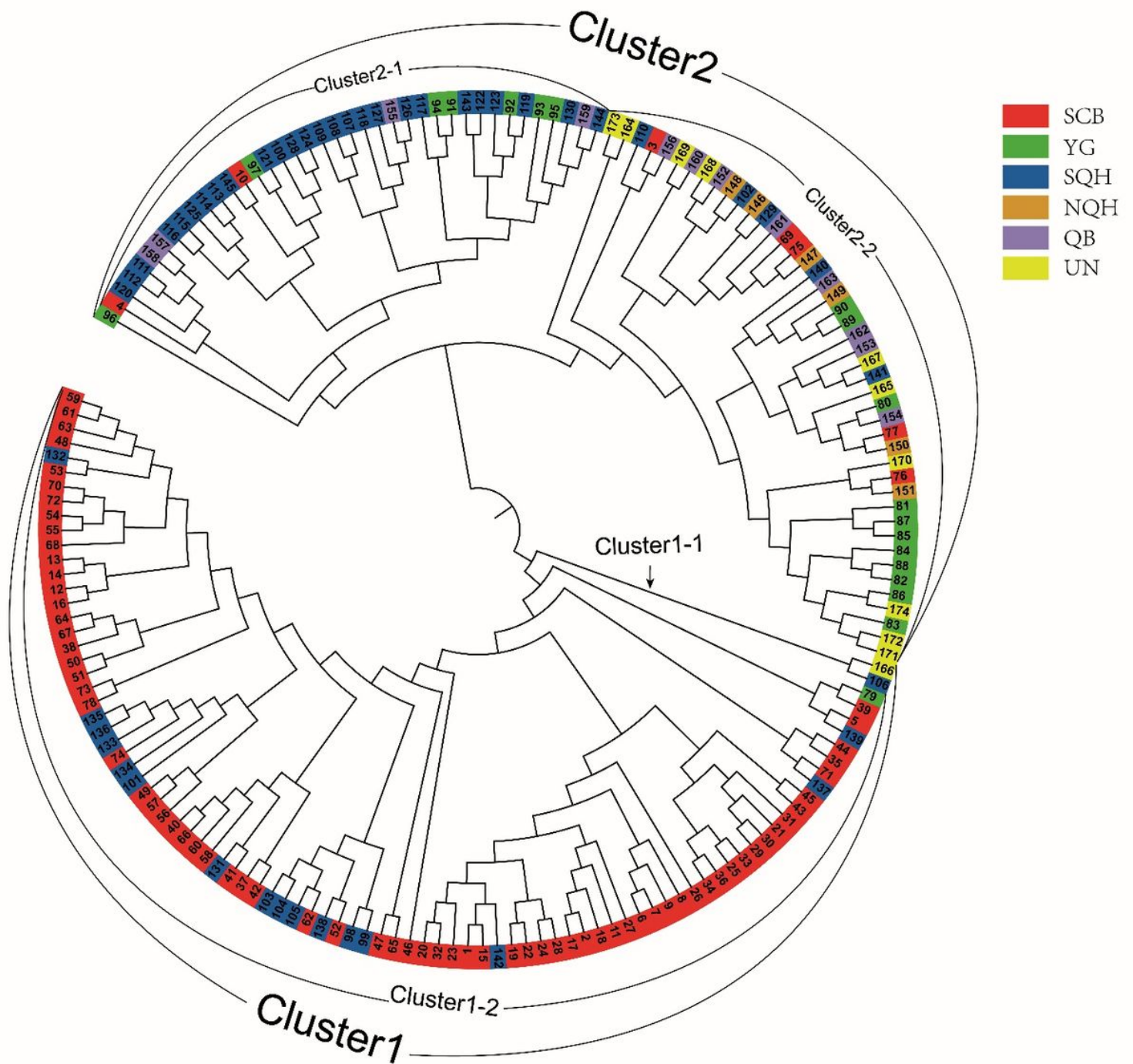


Figure 4

Phylogenetic tree of the 174 accessions of *A. trifoliata* based on 72 simple sequence repeat loci

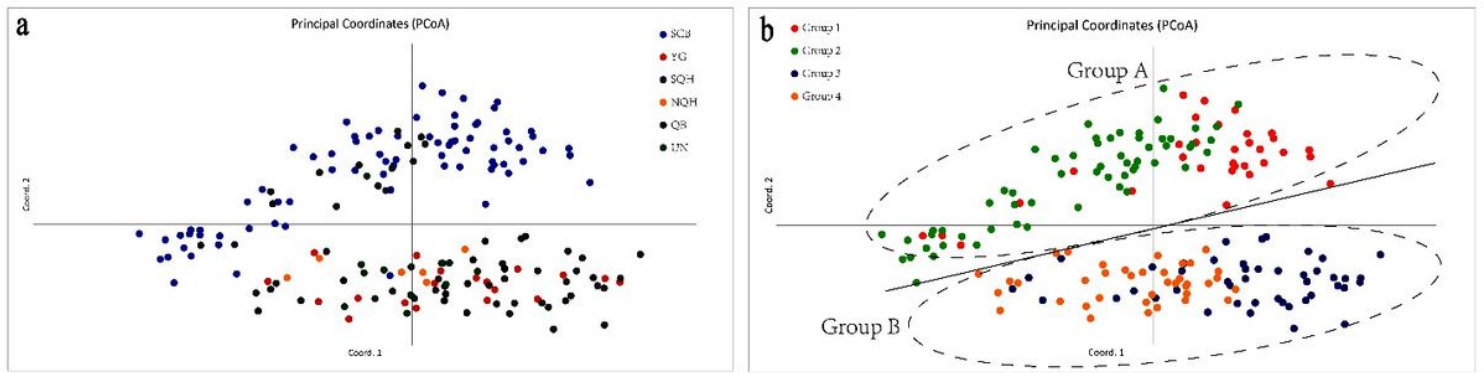


Figure 5

Scatter plots (a and b) from principal coordinate analysis. (a) PCoA based on geographical regions, (b) PCoA based on $K = 4$.