Genetic Diversity of a Natural Population of Akebia Trifoliata (Thunb.) Koidz and Extraction of a Core Collection Using Simple Sequence Repeat Markers

Yicheng Zhong
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Yue Wang
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Zhimin Sun
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Juan Niu
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Yaliang Shi
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Kunyong Huang
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Jing Chen
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Mingbao Luan (✉ luanmingbao@caas.cn)
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Jianhua Chen
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops

Research article

Keywords: Akebia trifoliata (Thunb.) Koidz, Genetic diversity, Simple sequence repeat markers, Core collection.

Posted Date: September 3rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-56041/v1

Abstract

Background: Developing a core collection can deepen our understanding of the genetic diversity of germplasm resources and lay the foundation for their rational utilization. *Akebia trifoliata* (Thunb.) Koidz is an important oil crop, particularly in China, but there is no relevant research report about the core collection of *A.trifoliate.*

Results: In this study, 28 simple sequence repeat (SSR) markers were used to assess the genetic diversity and genetic structure of a natural population of *A. trifoliata,* including 955 germplasms, and to extract a core collection. The genetic diversity of the natural population was moderately polymorphic. The average number of alleles (N_a), observed heterozygosity (H_O), expected heterozygosity (H_E), Shannon's information index (I^*), and polymorphic information content (PIC) were 3.71, 0.24, 0.46, 0.81, and 0.41, respectively. Two sub-populations were identified, indicating a weak genetic structure. The core collection was composed of 164 individuals (17.2% of 955 total germplasms in the population), and diversity parameters differed significantly from those of a random core germplasm collection.

Conclusions: The genetic diversity of the natural population of *A.trifoliata* was moderately polymorphic, and the genetic structure was weak. Moreover 164 individuals could represent the genetic diversity of the 995 individals. These results have implications for germplasm management and genomics studies in *A. trifoliata* as well as for the establishment of core collections of other perennial liana species.

Background

Akebia trifoliata (Thunb.) Koidz belongs to the Lardizabalaceae family and *Akebia* Decne. It is mainly distributed throughout China, Japan, North Korea, and Russia. In China, it is mainly found in the Yangtze River Basin, Yellow River Basin, and Shaanxi-Sichuan Area [1]. *A. trifoliata* has a long history of practical use in China as an important oil crop. Its seeds have a high oil content, with yields as high as 44%. Moreover, *A. trifoliata* fruits have many seeds (i.e., 57–200 seeds per fruit). The oil contains more than 10 kinds of fatty acids, and the saturated fatty acid:monounsaturated fatty acid:polyunsaturated fatty acid ratio is close to 1:1:1, fully conforming to the dietary nutrition standards recommended by the World Health Organization and the Chinese Nutrition Society, the oil is widely consumed in rural areas in southern China [2,3].

However, the species has been under considerable threat in China in the past few decades owing to changes in farming systems, economic development, urbanization, and other human disturbances. The natural wild species alone cannot meet the demand; therefore, it is important to accelerate germplasm preservation and to select and cultivate high-quality species with high oil yields. However, research has been focused on plant chemicals, and genetic analyses of *A. trifoliata,* particularly that of natural populations, are limited. Moreover, information on population genetic diversity is critical for crop breeding and production [4].

Molecular markers provide powerful tools for the identification of plant genetic diversity and the construction of core collections. Among different molecular markers, simple sequence repeats (SSRs) are widely used in plant science owing to their high polymorphism, reliability, rapid and simple detection, low cost, and easy operation [5]. However, to our knowledge, there had litter report on *A. trifoliata* genetic diversity based on SSR markers [6]. We have collected 955 accessions of *A. trifoliatan*. Redundant genetic resources present a challenge for the effective conservation, management, evaluation, and utilization of germplasms. To resolve this issue, it is necessary to construct a core collection. Frankel [7]first proposed the core collection, which provides preliminary information on diversity in a large collection. Core collections have since been developed for many oil crops, such as sesame (*Sesamum indicum* L.) [8], maize (*Zea mays* L.) [9], and soybean (*Glycine max*) [10]**.** A majority of the core collections include only 5–20% of the total germplasm but preserve most of the genetic diversity, thereby reducing the cost and increasing the speed of the work process. However, thus far, there is no core collection for *A. trifoliata*.

To evaluate the genetic diversity and genetic structure and to construct a core collection, 28 pairs of SSR markers were used to analyze 955 *A. trifoliata* germplasms collected from China. The genetic diversity and core collection provide a convenient resource for *A. trifoliata* breeding and protection as well as a foundation for follow-up research.

Methods

Plant materials and DNA isolation

The 955 *A. trifoliata* germplasms, for which the collection location was unclear, were cultivated in Taojiang experimental field of the Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences in 2012. Fresh tender leaves from each accession were placed in a liquid nitrogen tank, transported to the laboratory, and frozen at −80°C until genomic DNA extraction. Genomic DNA was extracted using a Rapid DNA Extraction Kit (Tiangen Biotech, Beijing, China). The purity and quality of extracted DNA were evaluated by 1% agarose gel electrophoresis and determined using a NanoDrop 2000 spectrophotometer.

SSR analysis

Twenty-eight SSR primer pairs (Table S1) were synthesized according to Niu et al [6]. SSR-primed polymerase chain reactions (PCRs) were carried out in a 10 uL reaction volume with 1× PCR buffer, 0.2 mmol/L dNTP, 1 U of Taq DNA polymerase (Tiangen), 0.5 uL of forward primer (10 nmol/L), 0.5 uL of reverse primer (10 nmol/L), and 0.5 uL of DNA from each accession. PCR was performed under the following conditions: 94°C for 5 min, followed by 33 cycles each of 30 s at 95°C, 30 s at the primer-specific annealing temperature, 30 s at 72°C, and a final extension of 10 min at 72°C. The PCR products were separated on 8% polyacrylamide gels, and silver dyeing was conducted according to the methods of Zhang et al [11]. Molecular weights were estimated using a DNA marker. The allele with the maximal molecular weight was recorded as “A,” followed by B, C, D, etc. If only one band was obtained for a set of primers, the germplasm was recorded as homozygous.

Data analysis

PowerGene version 1.3.2 [12] was used to analyze the effective number of alleles (N_e), Shannon–Weaver diversity index (I^*), genetic distance (Nei's genetic distance), observed heterozygosity (H_O), and expected heterozygosity (H_E); PowerMarker version 3.2.5 [13] was used to estimate the polymorphic information content (PIC) and number of alleles (N_a). Based on Nei's genetic distances, a clustering tree was constructed using PowerMarker, and visualized using MEGA version 7.0 and iTol [14]. Population genetic structure was assessed using the mixed model and the correlated allele frequency model in STRUCTURE version 2.3.4 [15] and Structure Harvester version 6.0 [16]. A principal component analysis (PCA) was performed using NTSYS10.2 [17]. The variance analysis was implemented in SAS version 9.0 [18].

Extraction of a core germplasm collection

The stepwise clustering (SC) method can effectively preserve the genetic diversity of the original germplasm [19]. Accordingly, in this study, SC was used to extract a core collection based on SSR markers. First, genetic distances were calculated for the original collection, and a cluster analysis was then performed according to the genetic distances. Next, a tree diagram was obtained. According to the principle of clustering, the differences within groups are smallest at the lowest level; therefore, one of the two genetic materials in each group were randomly selected to enter the next round of the cluster analysis. If only one genetic material was available, it was used in the next round of the cluster analysis. All retained genetic material was re-entered into the next round of the cluster analysis. The method was repeated until the material met the set requirements to obtain the core collection.

Results

Genetic diversity of the natural population

A total of 104 alleles were detected at 28 SSR markers. As summarized in Table 1, N_a per locus ranged from two to five (mean, 3.71). Seventeen primer pairs amplified four alleles and five primer pairs amplified two alleles, but only one amplified five alleles. N_E ranged from 1.2018 to 2.8556 (mean, 1.9873), H_O ranged from 0 to 0.83 (mean, 0.2382), H_E ranged from 0.1681 to 0.6521 (mean, 0.4604), Nei's distance ranged from 0.1679 to 0.6535 (mean, 0.4600), I^* ranged from 0.3083 to 1.1866 (mean, 0.8086), and PIC ranged from 0.1538 to 0.5936 (mean, 0.4085). The PIC indicates that the 28 SSR markers were moderately polymorphic ($0.25 < PIC < 0.5$); the most highly polymorphic SSR marker had 3.86 times higher variance than that of the least polymorphic marker. Seven microsatellites exhibited high polymorphism ($PIC > 0.5$) and four microsatellites exhibited low polymorphism ($PIC < 0.25$). The heterozygosity of *A. trifoliata* is relatively low based on H_O and H_E (i.e., 0.238 and 0.460, on average).

Genetic structure of the natural population

A cluster analysis was performed to analyze the genetic relationships among the 955 *A. trifoliata* accessions, and a dendrogram based on genetic distances is shown in Fig. 1. The cluster analysis divided 955 germplasms into two main groups, accounting for 56.44% and 43.56% of the natural population.

We evaluated K-values (population number) of 2–9 for a STRUCTURE analysis. The most significant change in likelihood occurred when the K-value increased from 2 to 3 and the highest ΔK value was observed between K = 2 and K = 3. Therefore, according to Evanno et al [20] (Fig. 2a), the optimal K value in this study was 2. The division of the natural population into two subgroups (Fig. 2b) was consistent with the results of the cluster analysis.

The PCA also divided most of the 955 *A. trifoliata* accessions into two populations, excluding only a few germplasms (Fig. 3). These results were similar to those of the cluster analysis and STRUCTURE analysis, supporting the division of the 955 *A. trifoliata* germplasms into two groups.

Extraction of a core collection

SC was used to extract a core collection using $N_{a,}$ Nei’s distance, H_O, and PIC over the 28 SSR markers as indicators. The core collection consisted of 164 genetic individuals, representing only 17.2% of the original genetic population. In comparison with the total natural population, the final core collection showed 94.2%, 98.7%, 116.4%, and 116.73% of the variation based on N_a, H_O, Nei’s distance, and PIC (Table 2). These results indicated that the core germplasm is representative of the entire genetic population.

To verify the reliability of the results, three false core collections composed of 164 individuals were randomly selected four genetic diversity indexes (N_a, H_O, Nei’s distance, and PIC) were estimated. As shown in Table 3, all four indicators were significantly different from those in the newly established core collection. These results support the validity of the method for extracting the core germplasm and further suggest that the core germplasm effectively represents the entire genetic population.

Discussion

A. trifoliata is an important oil crop. Most studies of the species have focused on active components, such as quinatic acid [21], triterpene saponins [22], and akebiaoside K [23]**.** Only a few studies have evaluated the biology of *A. trifoliata*. For example, Zou et al. [24] studied recurrent somatic embryogenesis and the development of somatic embryos. Niu et al [6] developed SSR markers via de novo transcriptome assembly and Zou et al [25] showed the effectiveness of recurrent selection in *A. trifoliata*breeding. However, this approach is not conducive to the development of *A. trifoliata* as an oil crop.

In this study, the collected 955 *A. trifoliata* germplasms were not registered, and so the geographical origin of each germplasm resource was unclear, which limits our understanding of the genetic diversity of germplasm resources and the development of a core collection. However, SSR molecular marker technology is not affected by geographical origin and complex factors, such as collection organs, development period, and external environment, and results in high polymorphism, stable results, and good repeatability.

Progress in *A. trifoliata* breeding has been slow, in part because it is a perennial plant and new plants do not bear fruit for 4 years [26]. Therefore, the generation of new *A. trifoliata* varieties is time-consuming. Furthermore, little is known about the biological characteristics of *A. trifoliata*, making it difficult to choose good parents. *A. trifoliata* has many uses that may guide breeding. For example, the consumption of *A. trifoliata* fruits is limited by the thick skin and abundant seeds [27], suggesting that breeding for thin skin and fewer seeds will improve market value. Similarly, *A. trifoliata* can be cultivated for use as an oil crop by focusing on seed properties. Molecular genetic markers are widely used in plant breeding, and genetic diversity must be considered when identifying trait populations and choosing parental strains to ensure the success of breeding. The results obtained in this study deepen our understanding of the genetic diversity of germplasm resources and facilitate the rational utilization of germplasm resources.

In this study, an SSR analysis of 955 *A. trifoliata* germplasms was performed to evaluate genetic diversity. In a previous study, 49 pairs of SSR markers were used to analyze 88 *A. trifoliata* germplasms [6]; PIC and H_O values were 0.43 and 0.2210, respectively, similar to those in our study (PIC = 0.41; H_O = 0.2382), thereby verifying that the species is moderately polymorphic (0.25 < PIC < 0.5). Additionally, 14 pairs of EST–SSR markers have been used to evaluate polymorphisms in 106 individuals from four natural populations of *Dysosma versipellis* (Berberidaceae) [28], with average N_a, H_O, H_E, and PIC values at 6.286, 0.296, 0.534, and 0.467, which were higher than the corresponding values in this study, but still demonstrated a moderate level of polymorphism. With respect to other oil plants, *A. trifoliata* polymorphism was similar to that in sesame (*Sesamum indicum* L.) [29] and peanut (*Arachis hypogaea* L.) [30], but lower than that estimated in maize (*Zea mays* L.) [31], soybean (*G. max*) [32]**,** and sunflower (*Helianthus annuus* L.) [33]**.**

Abundant crop germplasm resources are the basis of crop breeding. However, excessive germplasms have various limitations. For example, it is difficult to precisely and rapidly identify useful resources for plant breeders. The management and preservation of germplasm resources is expensive and time-consuming; a core collection can effectively resolve these issues [34]**.** This study demonstrates the feasibility of establishing a core germplasm collection in perennial oil crops and is the first core collection established in *A. trifoliata*. Although core collections have been reported for some oil crops, most are not perennial crops. The core germplasm represented 17.1% of all accessions, which is higher than the range of 5–10% recommended by Brown [35] as well as the values reported in other plants, e.g., sesame (*Sesamum indicum* L.) (28/277) [8], maize (*Zea mays* L.) (951/13521) [9], and soybean (*G. max*) [10], whereas they are slightly less than those for the rubber tree (*Hevea brasiliensis*) (128/505) [36], ramie (*Boehmeria nivea* L.) (22/105) [37], and Gympie messmate (*Eucalyptus cloeziana* F. Muell., family Myrtaceae Juss.) (247/707) [38]. However, if we apply one additional filter, the N_a and H_O are reduced to 82.1% and 84.2% of those for the full population, and the core collection is reduced to eight genetic individuals. The maintenance of the vast majority of germplasm diversity should be a priority for guiding the determination of an optimal fraction; accordingly, we did not aim for a low rate of germplasm retention.

To the best of our knowledge, this study is the first to apply SSR markers to a large number of *A. trifoliata* germplasms. Estimates of genetic diversity and genetic structure can provide a foundation for future *A. trifoliata* breeding. The core collection can reduce the management cost and improve the protection of germplasm resources. However, the establishment of a core germplasm collection is a dynamic process and subsequent studies are needed to continuously improve the core collection of *A. trifoliata.*

Conclusions

This study showed moderate genetic diversity and weak genetic structure in the natural population of *A. trifoliata,* based on 28 SSR markers. A core germplasm collection consisting of 164 germplasm was generated, accounting for 17.2% of the original germplasm. Further, these findings confirmed the feasibility of using SSR markers to establish a core collection for perennial vines and lay a foundation for further breeding and genomics studies of *A. trifoliata.*

Abbreviations

SSR: simple sequence repeat

$N_{a:}$ number of alleles

N_e: effective number of alleles

H_o: observed heterozygosity

H_e: expected heterozygosity

Nei’ s: genetic distance

I^*: Shannon’ s information index

PIC: polymorphic information content

Declarations

Acknowledgments

We would like to thank Jun Liu, Han Deng for technical assistance in DNA extraction and SSR markers experiments. We also thank Xiyang Chen, Xinkang Feng, Qiaoyun He, Zhoumei Zhang for the help in paper writing and data analysis.

Authors' contributions

ML and JC conceived and designed the project, YZ, YW, ZS, JN, YS, KH, and JC collected the plant materials. YZ, YW, ZS, JN, YS, and JC performed molecular labwork and scored the markers. YZ, YW, JN, and YS analyzed the data and wrote the manuscript with assistance from all other authors. All authors read and approved final manuscript

Finding

This research was financially supported by the Central Public- interest Scientific Institution Basal Research Fund (1610242020010) and Agricultural Science, Technology Innovation Program (ASTIP) of CAAS (Grant No.2017IBFC). The funding bodies had no role in the design, collection, and analysis, interpretation of data or in writing the manuscrip

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Availability of data and materials

No specific permits were required for the described field studies and the field studies did not involve endangered or protected species. The datasets supporting the conclusions of this article are included within the article and its additional files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors argeed to publich .

Competing interests

The authors declare that they have no competing interests

References

1. Xie J, Li XH, Zhang CJ, Ouyang HN. Distribution of Akebia trifoliata (Thunb.) Koidz wild resources. Journal of Shaanxi Normal University: Natural Science Edition. 2006;34(3): 272-274.
2. Wan ML, Liu XW, Ban XT, Luo KM, Shi LJ, Zhang CJ, Yang RY, Li ZJ, Hao RC. The Fruit Character and Nutrition Composition of *Akebia trifoliata*(Thunb.)Koidz under the Cultivation Condition. Guizhou Agricultural ences.2008; 36(3) 121-122.

3. Zhong WM, Ma YH. Analysis and evaluation of nutritional components in *Akebia trifoliate* Southwest China Journal of Agricultural Sciences. 2016;29(1):160-173.
4. Glaszmann JC, Kilian B, Upadhyaya HD, Varshney RK. Accessing genetic diversity for crop improvement. Curr. Opin. Plant Biol.2010; 13 (2):167–173.
5. Powell W, Machray GC, Provan J. Polymorphisms revealed by simple sequence repeats. Trends Plant Sci. 1996;1(7): 215-222.
6. Niu J, Wang YJ, Shi YL, Wang XF, Sun ZM, Huang KR, Gong C, Luan MB, Chen JH. Development of SSR markers via de novo transcriptome assembly in *Akebia trifoliata* (Thunb.) Koidz. Genome.2019; 62(12): 817-831
7. Frankel H. Genetic perspectives of germplasm conservation. In: Arber, W.K.,Llimensee, K., Peacock, W.J., Starlinger, P. (Eds.), Genetic Manipulation: Impact on Man and Society. Cambridge University Press, Cambridge, UK, 1984 pp. 161–170.
8. Park JH, Suresh S, Cho GT, Choi NG, Baek HJ, Lee CW, Chung JK. Assessment of molecular genetic diversity and population structure of sesame (sesamum indicum l.) core collection accessions using simple sequence repeat markers. Plant Genetic Resources.2014; 12(1):112-119.
9. Li Y, Shi Y, Cao Y, Wang T. Establishment of a core collection for maize germplasm preserved in Chinese National Genebank using geographic distribution and characterization data. Genetic Resources and Crop Evolution. 2005; 51(8):845-852.
10. Oliveira MF, Nelson RL, Geraldi IO, Cruz CD, Toledo José Francisco F. Establishing a soybean germplasm core collection. Field Crops Research.2010; 119(23): 277-289.
11. Zhang J, Wu YT, Guo WZ. Fast screening of microsatellite markers in cotton with page/silver staining. Acta Gossypii Sin.2000; 12: 267-269.
12. Yeh FC, Boyle TJB. Population genetic analysis of codominant and dominant markers and quantitative traits. Belg.j.bot. 1997;129.
13. Liu KJ, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 2015; 21: 2128–2129.
14. Tamura K, Stecher G, Peterson D, Filipsk A, Kumar S. Mega6: molecular evolutionary genetics analysis version 6. 0. Mol. Biol. Evol. 2013; 30: 2725–2729.
15. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155(4): 9197-9201.
16. Earl DA, Vonholdt BM. Structure harvester: a website and program for visualizing structure output and implementing the evanno method, Conservation Genetics Resources. 2012; 4(2): 359–361.
17. Rohlf FJ. NTSYSpc: numerical taxonomy system, Version 2.20. Exeter Publishing, Ltd, Seatauker, NY.2008
18. Park H. SAS Institute, Inc Cary North Carolina. Sign.2002
19. Wang JC, Hu J, Xu HM, Zhang S. A strategy on constructing core collections by least distance stepwise sampling. Theoretical & Applied Genetics.theoretische Und Angewandte Genetik. 2007;

115(1): 1-8.

20. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology. 2005; 14: 2611-2620.
21. Liu GY, Ma SC, Zheng J, Yu ZX, Lin RC. Determination of components quinatic acid and akeboside stc commonly contained in mutong of akebia. Chinese Journal of Analytical Chemistry. 2008; 36(5): 683-686.
22. Iwanaga S, Warashina T, Miyase T. Triterpene saponins from the pericarps of akebia trifoliata. Chemical and Pharmaceutical Bulletin. 2012; 60(10): 1264-1274.
23. Xu QL, Wang J, Dong LM, Zhang Q, Luo B, Jia YX, Wang HF, Tan JW. Two new pentacyclic triterpene saponins from the leaves of *Akebia trifoliata*. Molecules. 2016; 21(8): 962.
24. Zou S, Yao X, Zhong C, Li D, Huang H. Recurrent somatic embryogenesis and development of somatic embryos in *Akebia trifoliata* (thunb.) koidz (lardizabalaceae). Plant Cell, Tissue and Organ Culture (PCTOC). 2019; 139(3): 493-504.
25. Zou S, Yao X, Zhong C, Zhao T, Huang H, Effectiveness of recurrent selection in *Akebia trifoliata* (lardizabalaceae) breeding. Scientia Horticulturae. 2019; 246: 79-85.
26. Zhang YJ, Dang HS, Yang LL, Wei GY, Wang Y. Geographical distribution and resource survey of wild medicinal plant *Akebia trifoliata* trifoliata. Chinese Wild Plant Resources. 2013; 32(3): 58-62.
27. Zhou X, Zhang LB, Peng YH, Jiang LJ, Chen JZ, Yu PY, Feng XC, Li PW, Xiang M. Prospect of Breeding of Improved Varieties and Propagation Technology of Akebia trifoliata. Molecular Plant Breeding. http://kns.cnki.net/kcms/detail/46.1068.S.20200213.1629.007.html
28. Guo R, Mao YR, Cai JR, Wang JY, Wu J, Qiu YX. Characterization and cross-species transferability of EST–SSR markers developed from the transcriptome of *dysosma versipellis* (berberidaceae) and their application to population genetic studies. Molecular Breeding. 2014; 34(4): 1733-1746.
29. Hernán EL, Karlovsky P. Genetic relationship and diversity in a sesame (*Sesamum indicum*) germplasm collection using amplified fragment length polymorphism (AFLP). BMC Genetics. 2006; 7(10).
30. Zhang XR, Liu FZ, Zhang K, Wan YS. Population structure and genetic diversity analysis of peanut (*Arachis hypogaea*) using molecular markers. International Conference on Applied Biotechnology**.** 2018.
31. Inghelandt DV, Melchinger AE, Lebreton C, Stich B. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theoretical and Applied Genetics. 2010; 20(7): 1289-1299.
32. Lin H, Ke W, Kai Y. Genetic diversity of semi-wild soybean using SSR markers. Acta Botanica Boreali-occidentalia Sinica. 2002; 4: 751-757.
33. Carla VF, Natalia AJGR, Jeremias ZAP, Diego CMYM, Corina MF, Daniel ARAH, Horacio EH, Norma BP, Veronica VL. Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers, BMC Plant Biology. 2015; 52(15): 1-12.

34. Frankel OH, Brown AHD. Plant genetic resources today: a critical appraisal. Crop Genetic Resources : Conservation and Evaluation.1984
35. Brown AHD. The case for core collections. In: Brown, A.H.D., Frankel, O.H.,Marshall, D.R., Williams, J.T. (Eds.), The Use of Plant Genetic Resources. Cambridge University Press, Cambridge, UK, 1989, pp. 136–156
36. Fauzi AF, Rafii MY, Maiden AN, Roslinda S, Sulaiman Z. Core collection of Hevea brasiliensis from the 1995 RRIM Hevea germplasm for effective utilisation in the rubber breeding programme. Journal of Rubber Research. 2019; 23(1): 33-40.
37. Luan MB, Zou ZZ, Zhu JJ, Wang XF, Xu Y, Ma QH, Sun ZM, Chen JH. Development of a core collection for ramie by heuristic search based on ssr markers. Biotechnology and Biotechnological Equipment. 2014; 28(5): 798-804.
38. Lv JB, Li CR, Zhou CP, Chen JB, Li FG, Weng QJ, Li M, Wang YQ, Chen SK, Chen JC, Gan SM. Genetic diversity analysis of a breeding population of Eucalyptus cloeziana, F. Muell. (Myrtaceae) and extraction of a core germplasm collection using microsatellite markers. Industrial Crops and Products. 2020; 145: 112157.

Tables

Table 1 Genetic diversity parameters for original genetic population at the 28 SSR markers.

Marker	Na	Ne	Ho	He	Nei’s	I*	PIC
s3	4.0000	2.1867	0.8331	0.5430	0.5427	0.8772	0.4518
s4	3.0000	1.2635	0.0000	0.2087	0.2085	0.3772	0.1881
s5	2.0000	1.2018	0.0000	0.1681	0.1679	0.3083	0.1538
s13	4.0000	2.4260	0.3389	0.5881	0.5878	0.9884	0.5184
s19	3.0000	1.2717	0.0597	0.2137	0.2136	0.4423	0.2029
s22	4.0000	1.9081	0.3488	0.4762	0.4759	0.8338	0.4180
s24	4.0000	1.4441	0.2630	0.3077	0.3075	0.5870	0.2857
s25	5.0000	2.8864	0.4837	0.6539	0.6535	1.1603	0.5917
s27	4.0000	1.5906	0.1328	0.3715	0.3713	0.6942	0.3412
s28	4.0000	2.8708	0.4882	0.6521	0.6517	1.1866	0.5936
s30	4.0000	2.6522	0.3899	0.6233	0.6230	1.1080	0.5553
s32	3.0000	1.3067	0.1883	0.2349	0.2347	0.4760	0.2217
s34	4.0000	1.6833	0.2664	0.4062	0.4059	0.7404	0.3665
s40	3.0000	2.0286	0.2000	0.5079	0.5071	0.8584	0.4434
s46	4.0000	1.6737	0.2225	0.4028	0.4025	0.7544	0.3686
s50	3.0000	1.5539	0.0881	0.3567	0.3565	0.5908	0.3030
s52	4.0000	1.9707	0.2440	0.4928	0.4926	0.8646	0.4415
s57	4.0000	1.6181	0.3394	0.3822	0.3820	0.6975	0.3494
s59	4.0000	1.5182	0.1046	0.3416	0.3423	0.6152	0.3037
s67	4.0000	2.0343	0.0986	0.5087	0.5084	0.8766	0.4509
s68	4.0000	2.6354	0.2705	0.6210	0.6206	1.0916	0.5578
s72	3.0000	2.2823	0.2358	0.5624	0.5619	0.9275	0.4834
s74	4.0000	2.8566	0.2374	0.6504	0.6499	1.1549	0.5879
s77	3.0000	2.3400	0.0400	0.5746	0.5726	0.9219	0.4787
s84	4.0000	1.6493	0.1595	0.3940	0.3937	0.7671	0.3654
s89	4.0000	2.6996	0.2158	0.6307	0.6296	1.0835	0.5599

s92	4.0000	1.9091	0.0346	0.4764	0.4762	0.7087	0.3712
s100	4.0000	2.1819	0.3846	0.5420	0.5417	0.9489	0.4854
Mean	3.7143	1.9873	0.2382	0.4604	0.4600	0.8086	0.4085

Marker: the name of SSR marker

Na: number of alleles

Ne: effective number of alleles

Ho: observed heterozygosity

He: expected heterozygosity

Nei’s: genetic distance

I*: Shannon’s information index

PIC: polymorphic information content

Table 2 Comparisons of the genetic diversity among core collection and original genetic population

	Original genetic population	Core collection	Retention
Number	955	164	17.2%
Na	3.7143	3.5000	94.2%
Ho	0.2382	0.2351	98.7%
Nei’s	0.4600	0.5356	116.4%
PIC	0.4085	0.4752	116.3%

Table 3 Differences between genetic diversity of core collection and pseudo-core collection

	Na	Ho	Nei’s	PIC
Core collection	3.5000	0.2351	0.5356	0.4753
First random	3.3929	0.2333	0.4436	0.3932
Second random	3.3214	0.2374	0.4672	0.4154
Third random	3.4286	0.2399	0.4524	0.4018
P value	<0.001	<0.001	<0.001	<0.001

Figures

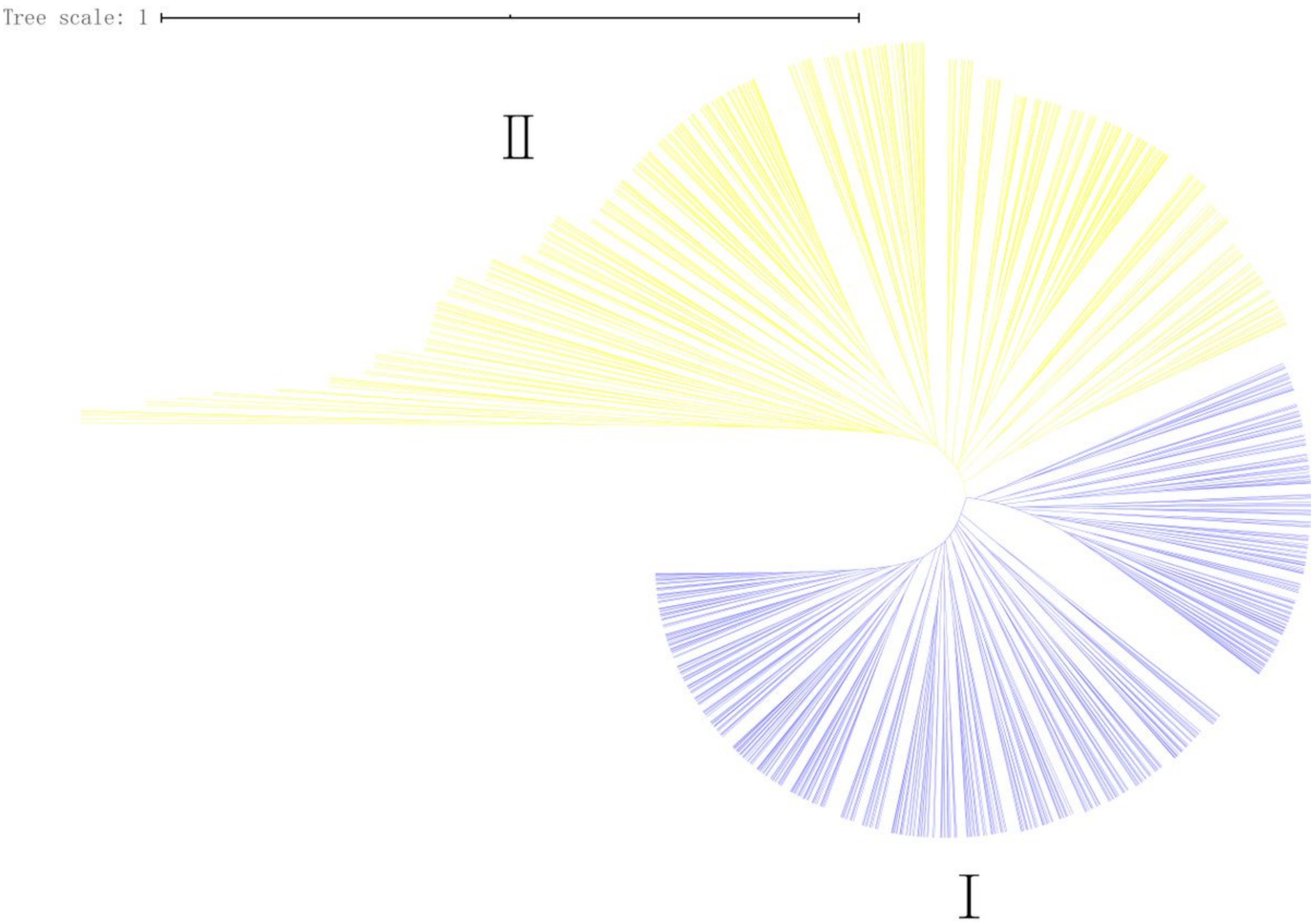

Figure 1

Phylogenetic tree of 955 A.trifoliate accessions based on genetic distance. Blue and orange indicate different clusters.

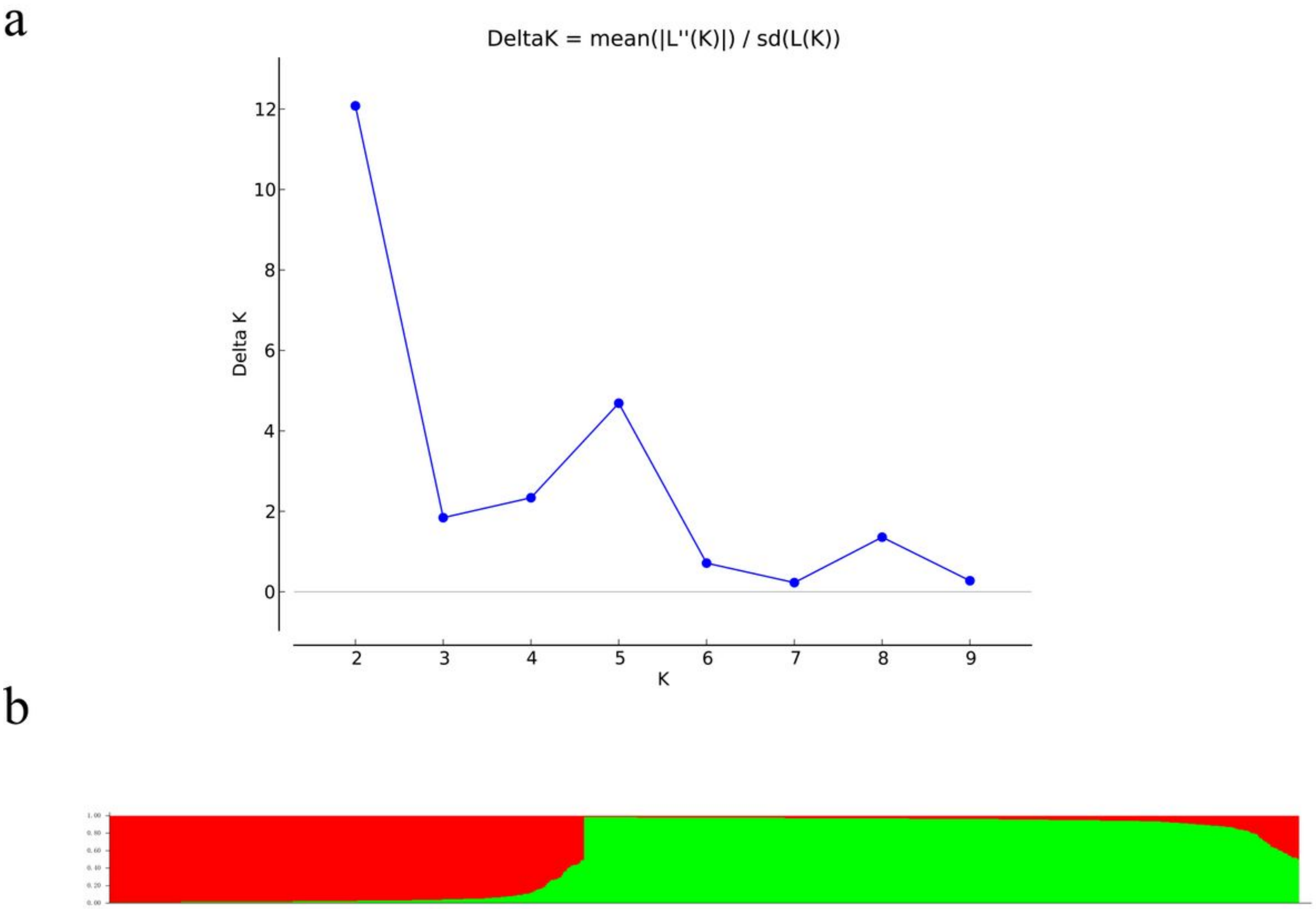

Figure 2

Population structure analysis of 955 A.trifoliate accessions. (a) Delta K based on the rate of change of L (K) between successive K values. (b) Population structure based on K = 2, Red: group 1, Green: group 2.

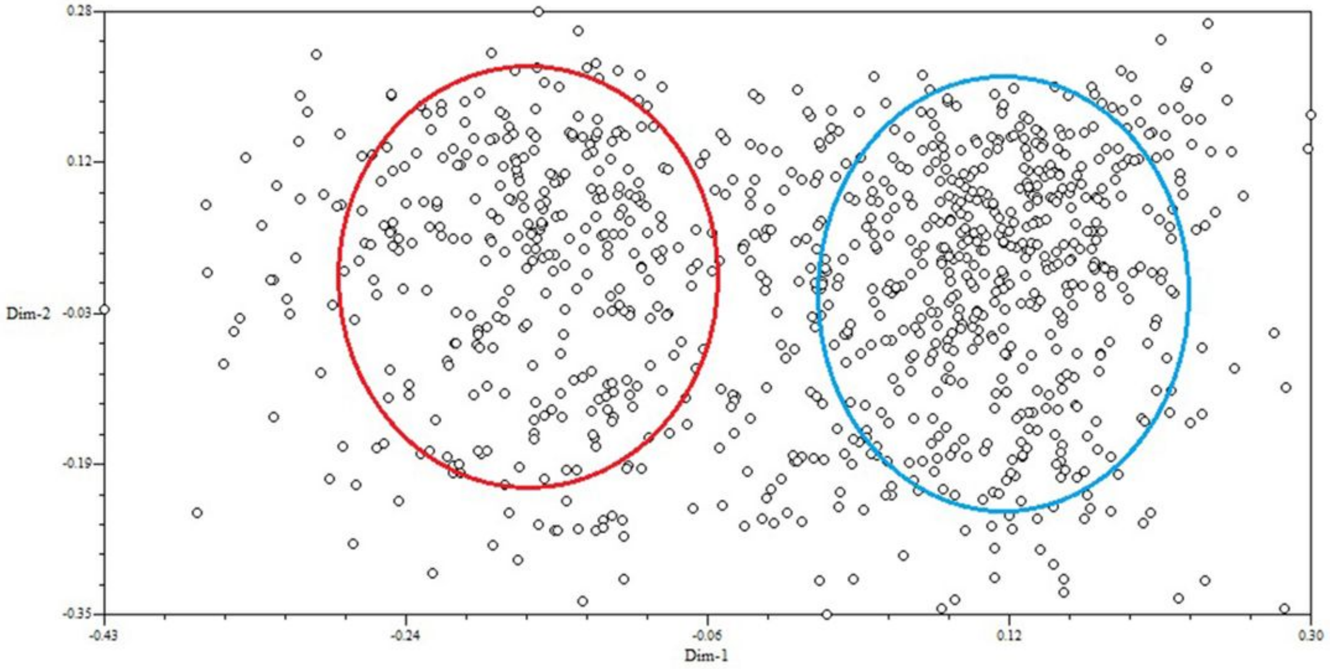

Figure 3

Principal coordinate analysis of 955 A.trifoliate accessions. Two circles represent two groups,Red:group 1,Blue: gupup two. The germplasm outside the circle is not in these two groups.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable.docx
- SupplementaryTableS5.doc
- SupplementaryTableS4.doc
- SupplementaryTableS3.doc
- SupplementaryTableS2.doc
- SupplementaryTableS1.doc