Development of Genomic Resources From *Crossostephium Chinense* (Asteraceae) Based on Genome Skimming Data

Luxian Liu
Henan University
Shook Ling Low
University of Bonn
Yu Feng
Zhejiang University
Binjie Ge
Shanghai Chenshan Botanical Garden: Shanghai Chen Shan Botanical Garden
Kamil Konowalik
Wroclaw University of Environmental and Life Sciences: Uniwersytet Przyrodniczy we Wroclawiu
Pan Li (panli_zju@126.com)
Zhejiang University https://orcid.org/0000-0002-9407-7740

Research Article

Keywords: Genome skimming, Simple sequence repeats (SSRs), Genetic resources, Molecular markers

Posted Date: March 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-280353/v1

Version of Record: A version of this preprint was published at Molecular Biology Reports on August 15th, 2021. See the published version at https://doi.org/10.1007/s11033-021-06590-9.

Abstract

Crossostephium chinense is a traditional Chinese medicinal herb and it is often cultivated as an ornamental plant. Previous studies on this species mainly focused on its chemical composition and it was represented rarely and marginally in genetic studies, which limited knowledge about its genetic background, and thus genomic resources remain scarce. To develop both chloroplast and nuclear polymorphic microsatellites for *C. chinense*, potential microsatellites were screened from genome skimming data of two individuals of *C. chinense*. Sixty-four and 63 cpSSR markers were identified from two chloroplast genomes of *C. chinense*. This study performed for the first-ever study on employment of genome skimming data and CandiSSR, consequently a total of 133 polymorphic nSSRs were developed. Ten nSSRs were randomly selected to test their transferability across 35 individuals from three populations of *C. chinense*, and 20 individuals each of *Artemisia stolonifera* and *A. argyi*. Cross-amplifications were successful done for *C. chinense*, and were partially successful amplified for both *Artemisia* species. The number of alleles varied from two to nine. The observed heterozygosity and expected heterozygosity per locus ranged from 0.000 to 0.286 and from 0.029 to 0.755, respectively. These genomic resources will be valuable for population genetics and conservation studies in *C. chinense* and *Artemisia*.

Introduction

Crossostephium chinense (L.) Makino (Asteraceae) forms a monotypic genus [1], or it is alternatively classified as *Artemisia chinensis* L. placed in the *Artemisia* subg. *Pacifica* C.R. Hobbs & B.G. Baldwin [2]. Whole plants of *C. chinense* are usually used for traditional Chinese medicinal treatments [3], and are widely cultivated for ornamental purposes [4, 5]. In the wild, the populations of *C. chinense* are restricted to the Southern region of China (Zhejiang, Fujian, Guangdong, and Taiwan), and the Ryukyus of Japan [6]. This pattern is congruent with the results of ecological niche modeling [2] and 215 occurrences available in the GBIF database (**GBIF**, https://doi.org/10.15468/39omei). The narrow distribution range may be a consequence of its coastal zone habitat limited to raised coral outcrops [6], as well as its physical adaptation to regional microclimates [2]. Personal observations found that each site forms an isolated population, especially in islands or the coast regions. Following WFO's red-listed *C. chinense* as a threatened species [7], this has raised public awareness on the importance and conservation of these rare populations. Previous studies of *C. chinense* mainly focused on its phytochemical composition [3, 8, 9]. The genetic studies by using fragment chloroplast and nuclear DNA of *C. chinense* have been partly carried out, nevertheless this species is always treated as a boundary species for other clades clarification in studies devoted to Asteraceae systematics [10]. Scarcity of the genetic information may hinder its effective utilization and protection, therefore there is a need for further studies.

Chloroplast genomes that possess an intermediate level of nucleotide substitution rate, are more conserved than nuclear and mitochondrial genomes [11]. Apart, its' non-recombinant nature and generally uniparental inheritance has led to the increasing utilization of the cp genome as a useful tool to understand the evolutionary history [12], as well as the genetic resources to develop abundant molecular markers such as chloroplast hotspot regions and chloroplast SSRs (cpSSRs) [13]. Unlike cpSSRs, nuclear SSRs (nSSRs) are highly polymorphic, codominant and biparentally inherited, making it widely applied for the evaluation of genetic variation [14], construction of genetic linkage maps [15], and conservation of the genetic resources [16]. The huge availability of genome database at reasonable costs, simultaneously coupled with a series of developed bioinformatics tools [17] had twisted the defect for large-scale investigations of molecular markers as compared to the traditional polymorphic SSR markers screening [18]. Including bioinformatics pipelines such as CandiSSR, is applied to detect candidate polymorphic SSRs from the next generation sequencing data [18].

The *de novo* or references-guided assembled chloroplast genomes of two *C. chinense* accessions are available on NCBI, nevertheless the microsatellites of *C. chinense* have never been studied. Thus, this paper is specifically aimed on utilizing the genome skimming data of *C. chinense* to develop cpSSRs and nSSRs markers. A set of randomly selected nSSRs markers was further used to validate the cross-amplification in 35 individuals collected from three populations of *C. chinense*, as well as 20 individuals each of species *Artemisia stolonifera* and *A. argyi*.

Materials And Methods

Plant material and DNA extraction

A total of 35 fresh young leaves were sampled from three populations of *C. chinense* for assessing and validation of genetic markers (Table S1). Genomic DNA was extracted using Plant DNAzol Reagent (LifeFeng, Shanghai) following the manufacturer's protocol. After isolation the material was frozen prior to the next downstream analyses.

Chloroplast genome marker (cpSSRs) development

To develop the chloroplast genome markers and to show the intraspecific variations in *C. chinense*, the two assembled chloroplast genome of *C. chinense* [19] were aligned and then adjusted manually in Geneious v8.1.7 [20]. The cpSSRs markers and single nucleotide substitutions (SNPs) for the *C. chinense* chloroplast genome were identified. Six types of repeats with minimum numbers of repetition, of which mononucleotide (10) > dinucleotide (5) > trinucleotide (5) > tetranucleotide (3) > pentanucleotide (3) > hexanucleotide (3) were implemented using Msatcommander v0.8.2 [21]. SNPs (nucleotide diversity [Pi]) were calculated using DNASP v5 software [22].

Polymorphic nuclear SSRs (nSSRs) development and validation

The genome skimming data of two *C. chinense* individuals we obtained previously were used to develop polymorphic nuclear SSRs markers [19]. The raw data were filtered and assembled into contigs using the CLC de novo assembler beta 4.06 (CLC Inc. Rarhus, Denmark). The chloroplast and mitochondria contigs from both *C. chinense* sequences were removed using the search engine on BLAST (NCBI BLAST v2.2.31). This was done by comparing to the chloroplast sequence of *C. chinense* (NCBI accession number: MH708561) and mitochondria sequence of *Helianthus annuus* (NCBI accession number: CM007908). Then,

software CandiSSR [18] was used to identify polymorphic nSSRs markers for *C. chinense*. The selected parameters in CandiSSR are performed by setting the flanking sequence length at 80, blast identity cutoff set at 95, blast e-value cutoff set at 1e-10, and blast coverage cutoff set at 95. For each target nSSRs, the primers are automatically designed in the pipeline developed for the Primer 3 package [23].

Ten developed polymorphic nSSRs markers were randomly selected to test the transferability to 35 individuals (three populations) of *C. chinense* collected from three different localities (Table S1). These ten nSSRs markers were also used for cross-amplification on *Artemisia stolonifera* and *A. argyi* (n= 20 respectively; Table S1). PCR amplifications were performed in a final volume of 10μL, which contained 1μL of genomic DNA, 5μL 2×Taq MasterMix (CWBIO, China), 0.1μM each of both forward and reverse fluorescently labeled universal primer (FAM, HEX, TAMRA; Table 1). The PCR conditions involved a single initial denaturation stage at 94°C for 1 min; followed by 28 cycles of denaturing, annealing and extending reactions respectively set at 94°C for 30s, 50-59°C for 30s, and 72°C for 30s. PCR reaction was completed with a final extension at 72°C for 5 min. Fragment lengths of PCR products were analyzed on an ABI PRISM 3720xl Genetic Analyzer (Applied Biosystems). Genotypes were scored by using the software GeneMarker v2.2.0 (SoftGenetics, LLC, State College, PA, USA). Deviations from Hardy-Weinberg equilibrium were tested through GENEPOP v4.2 [24]. We estimated genetic diversity parameters such as the number of alleles, observed and expected heterozygosity using CERVUS v3.0 [25].

Results

Chloroplast genome markers (cpSSRs) development

Sixty-four cpSSRs markers were identified from the *C. chinense*-ZJWZ cp genome. Among them, 51 markers were located in the LSC regions, whereas SSC and IR regions possess seven and six copies, respectively (Figure 1a-A). Of the genes and intergenic spaces of *C. chinense*-ZJWZ (NCBI accession number: MH708561), nine cpSSRs each were present in the protein-coding regions and in the introns, whereas 46 were identified from the intergenic spacer regions (Figure 1a-B). Among the lengths of repeated sequences, 48 cpSSRs are mononucleotides, 11 are dinucleotides, and five are tetranucleotides (Figure 1a-C). For the *C. chinense*-JPBB (NCBI accession number: MH708560), 63 cpSSRs markers were detected. Fifty cpSSRs were located in the LSC regions, whereas 7 and 6 were located in the SSC and IR regions, respectively. Among these markers, 47 are mononucleotides, 11 are dinucleotides and five are tetranucleotides. The distributions and types of cpSSRs in *C. chinense*-JPBB is shown in Figure 1b. All types of repeats were ATC-rich (Figure 1). Comparative analyses between the two *C. chinense* chloroplast genomes shown that nine cpSSRs loci are polymorphism, of which eight are located in the intergenic regions and one in the coding region (Table 1).

Nineteen SNPs, which also known as the mutational hotspots were detected from the pairwise alignment of both *C. chinense* chloroplast genomes. This include nine SNPs in the intergenic regions, two in the intron regions and eight SNPs in the coding sequences (Table 2). All the SNPs marker were located in the large and small single copy regions (LSC and SSC). All regions contained one substitution type, except the SNPs marker of *atp*A-*trn*R, *ndh*A, and *ycf*1 which contained two substitution types. Among the six substitution types, shifting from the T to G and A to C had the highest frequencies. Overall, the transition to transversions (Ts/Tv) marked ratio for 0.32. With the SNPs marker distinguish by LSC and SSC regions, the Ts/Tv values were 0.31 and 0.33 respectively. Meanwhile for category by gene types, the intron possesses Ts/Tv values of 0.5, while the spacer and exon respectively owned for 0.22 and 0.38. Further, narrow nucleotide diversity was examined for the sixteen SNPs marker, ranging from 0.0160 (*atp*A-*trn*R) to 0.0002 (*rpo*C2; Table 2).

Nuclear microsatellite markers (nSSRs) development and validation

A total of 133 polymorphic nSSRs markers were generated for *C. chinense*, where the screening hit the criteria of similarity < 90% and no available markers were designed (Table S2). The standard deviation of these markers ranged from 0.5 to 2.5. Among them, di-, tri-, tetra-, penta- and hexanucleotides account for 57.10%, 39.90%, 1.50%, 0.75% and 0.75%, respectively (Figure 2). Ten selected primers for cross-amplification successfully amplified the nSSRs loci of 35 *C. chinense* individuals which were collected from three populations (Table 3). Observed alleles varied between 2 to 9 alleles per locus, while both H_O and H_E ranged from 0.000 to 0.286 and from 0.029 to 0.755, respectively (Table 4). For *Artemisia stolonifera* and *A. argyi*, the observed alleles ranged from 1 to 6, and H_O and H_E varied from 0.000 to 0.450 and from 0.000 to 0.565, respectively (Table 4). Besides, four loci (CC19, CC32, CC55 and CC66) showed significant deviation from expectations under Hardy-Weinberg equilibrium for *C. chinense* due to the presence of excess homozygotes.

Discussion

This study performs the first detailed study of the genome resources of *C. chinense*, although conservation strategies should have implemented due to its economically important value [4, 5]. Both *C. chinense*-ZJWZ and JPBB that possess chloroplast genome lengths of 151,024 bp and 151,097 bp have sharing identical gene contents information [19], which then causes the divergence hotspots could not be detected in *C. chinense*. Nevertheless, both sequences have harbored 19 highly variable loci (Table 2) that might serve the potential mutational hotspots. Comprehensive mutational hotspot markers screening from the whole chloroplast genome is useful in polymorphism sites identification to elucidate the evolutionary and resolving controversial in phylogenetic relationships, hybridization issues, and biogeography [26, 27]. For instance, 20 mutational hotspots were respectively recognized for *Artemisia scoparia* [28], *A. maritima* and *A. absinthium* [29]. Meanwhile, Kim et al. [30] compared 21 *Artemisia* (32 accessions) and suggested the markers *acc*D and *ycf*1 may represent the potential markers to be tested for the whole Asteraceae. Recognition of these two markers seem to be line with several other studies on the genera in Asteraceae, where either one of both markers were observed in the suggestion list [31, 32]. The marker *ycf*1 is included among the 19 divergent hotspots although it possesses a lower nucleotide diversity (Pi = 0.0004), imply the potential application of these markers over all species of Asteraceae. Overall, narrow nucleotide diversities for *C. chinense* are observed (Table 2), likely because of comparative analyses was made within species that sampling differently from two localities. The highly divergent hotspots usually are identified between closer species [30, 31, 32]. Furthermore, the chloroplast genomes of same species were relatively conserved, exhibited in less remarkable polymorphism. Thus, the listed mutational hotspots regions in this study, though with low nucleotide diversity, could still apply for inter-population genetic study and phylogeographic study to test the biogeography origin.

Nine polymorphism cpSSRs observed from both genome of *C. chinense*-ZJWZ and-JPBB are mononucleotide tandem repeats with intraspecific variation of polyA (polyadenine) represents the most repeated motif in six primer sets (Table 1). Overall, the repeat motif is varied between 10 and 12 nucleotides, with either polyA or polyT shown as the content. Among reported Asteraceae, the identified loci with abundance A/T content were also present for *A. scoparia* [28]. Moreover, mononucleotide SSRs were also the most frequent identified sequence in *Artemisia* species [28, 29], though multiple-nucleotide type SSRs may sometime present in least frequency. A distribution pattern of continual repeat sequences of polyT (polythymine) following by polyA are occurred in *atp*A-*trn*R (Table 1). Among the cpSSRs markers, three primer sets of which *rpo*C2-*rps*2 (cpSSR2), *atp*A-*trn*R (cpSSR4), and *ycf*1 (cpSSR9) were also suggested for mutational hotspots (Table 1). Repeat sequences have been proven crucial in chloroplast genome arrangement and sequence variation [27]. Further, the variable repeat sequences between lineages allow it significances used as microsatellites markers for genetic diversity, and population genetics studies of plant species [28, 33].

In this study, the employment of genome skimming data using CandiSSR represents the first-ever study in Asteraceae to identify the appropriate polymorphic nSSRs for *C. chinense*. Estimation on the expected heterozygosity that shown significant deviation on four loci (CC19, CC32, CC55 and CC66) may not only due to the presence of an excess homozygotes. Other factors including Wahlund effect, inbreeding, null alleles, and sampling effect are also the potential causes to the deviation [34, 35]. Attempt of 10 selected nSSRs tested for the transferability of loci among the populations of *C. chinense* is perfectly successful, whereas only four nSSRs is applicable for *Artemisia stolonifera* and *A. argyi* (Table 4). Verification of transferability loci onto other species would allow further understanding of phylogenetic relationship at both inter and intra level [36]. Thus, it is believed that more transferability markers could be select from the remaining 123 markers in this study. The approaches of applying nSSRs from various employment have been developed in Asteraceae for *Chresta* [37], *Solidago* [38], as well as to study the hybridization of two *Tithonia* species [39]. Application of nSSRs were also used for other plant species such as transferability test in *Sanguinaria* [40], and genetic structures studies in *Salix* [41], *Euptelea* [42], and *Engelhardia* [43]. The 133 successfully developed polymorphic nucleotide microsatellite markers can be further applied to reveal the genetic diversity, population structure, and to develop effective conservation as well as management strategies for *C. chinense*. This approach is applicable to other plants species.

Conclusions

In summary, 133 polymorphic nucleotide microsatellite markers were developed successfully and can be applied to reveal the genetic diversity, population structure and possible intra- and inter- population gene flow of *C. chinense*. It could also apply for effective conservation as well as management strategies for *C. chinense*. Moreover, our study confirms the suitability used of nSSRs across species and is applicable to *Artemisia*. This imply the potential use of these nSSRs for robust genetic studies.

Declarations

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant Nos. 31900188, 31970225), Natural Science Foundation of Zhejiang Province (Grant No. LY19C030007). The material was supported by National Wild Plant Germplasm Resource Center for Shanghai Chenshan Botanical Garden (ZWGX1902)

Author Contributions

PL designed the study. PL, YF and BJG collected the samples. LXL conducted the laboratory experiments. LXL, SLL and KK conducted bioinformatic and statistical analyses. All the authors drafted and revised the manuscript.

Compliance with ethical standards

Conflict of Interest

The authors declare no conflict of interest.

Ethical approval

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

References

1. Makino T (1906) Observations on the Flora of Japan. Shokubutsugaku Zasshi 20: en23-en35.
2. Hobbs CR, Baldwin BG (2013) Asian origin and upslope migration of Hawaiian *Artemisia* (Compositae-Anthemideae). J Biogeogr 40: 442-454.
3. Wu Q, Zou L, Yang XW, Fu DX (2009) Novel sesquiterpene and coumarin constituents from the whole herbs of *Crossostephium chinense*. J Asian Nat Prod Res 11: 85-90.
4. Tang F, Chen F, Chen S, Wang XE, Zhao H (2010) Molecular cytogenetic identification and relationship of the artificial intergeneric hybrid between *Dendranthema indica* and *Crossostephium chinense* by GISH. Plant Syst Evol 289: 91-99.
5. Yang H, Sun M, Lin S, Guo Y, Yang Y, Zhang T, Zhang J (2017) Transcriptome analysis of *Crossostephium chinensis* provides insight into the molecular basis of salinity stress responses. PLoS ONE 12: e0187124.

6. Zhu S, Yilin C, Yousheng C, Yourun L, Shangwu L, Xuejun G, ... & Gottschlich G (2011) Asteraceae (Compositae). Flora of China; Wu, ZY, Raven, PH, Hong, DY, Eds, 20-21.
7. WFO (2021) *Crossostephium chinense* (A.Gray ex L.) Makino. Published on the Internet; http://www.worldfloraonline.org/taxon/wfo-0000085439. Accessed on: 27 Jan 2021.
8. Yang XW, Zou L, Wu Q, Fu DX (2008) Studies on chemical constituents from whole plants of *Crossostephium chinense*. China journal of Chinese materia medica 33: 905-908. (In Chinese)
9. Uehara A, Kitajima J, Kokubugata G, Iwashina T (2013) Further characterization of foliar flavonoids in *Crossostephium chinense* and their geographic variation. Nat Prod Commun 9: 163-164.
10. Zhao HB, Chen FD, Chen SM, Wu GS, Guo WM (2010) Molecular phylogeny of *Chrysanthemum*, *Ajania* and its allies (Anthemideae, Asteraceae) as inferred from nuclear ribosomal ITS and chloroplast *trn*L-F IGS sequences. Plant Syst Evol 284: 153-169.
11. Drouin G, Daoud H, Xia J (2008) Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol Phylogenet Evol 49: 827-831.
12. Yang J, Yue M, Niu C, Ma XF, Li ZH (2017) Comparative analysis of the complete chloroplast genome of four endangered herbals of *Notopterygium*. Genes 8: E124.
13. Liu LX, Wang YW, He PZ, Li P, Lee J, Soltis DE, Fu CX (2018) Chloroplast genome analyses and genomic resource development for epilithic sister genera *Oresitrophe* and *Mukdenia* (Saxifragaceae), using genome skimming data. BMC Genomics 19: 235.
14. Kashi Y, King D, Soller M (1997) Simple sequence repeats as a source of quantitative genetic variation. Trends Genet 13: 74-78.
15. Jones E, Dupal M, Dumsday J, Hughes L, Forster J (2002) An SSR-based genetic linkage map for perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 105: 577-584.
16. Varshney RK, Chabane K, Hendre PS, Aggarwal RK, Graner A (2007) Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. Plant Sci 173: 638-649.
17. Kumar S, Dudley J (2007) Bioinformatics software for biologists in the genomics era. Bioinformatics 23: 1713-1717.
18. Xia EH, Yao QY, Zhang HB, Jiang JJ, Zhang LP, Gao LZ (2016) CandiSSR: an efficient pipeline used for identifying candidate polymorphic SSRs based on multiple assembled sequences. Front Plant Sci 6: 1171.
19. Chen TT, Du YX, Zhao HD, Dong MF, Liu LX (2019) The complete chloroplast genome of *Crossostephium chinense* (Asteraceae), using genome skimming data. Mitochondrial DNA B 4: 322-323.
20. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond, A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647-1649.
21. Faircloth BC (2008) Msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. Mol Ecol Resour 8: 92-94.
22. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452.
23. Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23: 1289-1291.
24. Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour 8: 103-106.
25. Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16: 1099-1106.
26. Wang W, Chen S, Zhang X (2018) Whole-genome comparison reveals divergent IR borders and mutation hotspots in chloroplast of herbaceous bamboos (Bambusoideae: Olyreae). Molecules 23: 1537.
27. Li J, Ye G, Liu H, Wang Z (2020). Complete chloroplast genomes of three important species, *Abelmoschus moschatus*, *A. manihot* and *A. sagittifolius*: Genome structures, mutational hotspots, comparative and phylogenetic analysis in Malvaceae. PLoS ONE 15: e0242591.
28. Iram S, Hayat MQ, Tahir M, Gul A, Ahmed I (2019) Chloroplast genome sequence of *Artemisia scoparia*: comparative analyses and screening of mutational hotspots. Plants 8: 476.
29. Shahzadi I, Mehmood F, Ali Z, Ahmed I, Mirza B (2020) Chloroplast genome sequences of *Artemisia maritima* and *Artemisia absinthium*: Comparative analyses, mutational hotspots in genus *Artemisia* and phylogeny in family Asteraceae. Genomics 112: 1454-1463.
30. Kim GB, Lim CE, Kim JS, Kim K, Lee JH, Yu HJ, Mun JH (2020) Comparative chloroplast genome analysis of *Artemisia* (Asteraceae) in East Asia: insights into evolutionary divergence and phylogenomic implications. BMC Genomics, 21: 1-17.
31. Cho MS, Kim SH, Yang J, Crawford DJ, Stuessy TF, López-Sepúlveda P, Kim SC (2020) Plastid phylogenomics of *Dendroseris* (Cichorieae; Asteraceae): Insights into structural organization and molecular evolution of an endemic lineage from the Juan Fernández Islands. Front Plant Sci 11: 594272.
32. Shen J, Zhang X, Landis JB, Zhang H, Deng T, Sun H, Wang H (2020) Plastome evolution in *Dolomiaea* (Asteraceae, Cardueae) using phylogenomic and comparative analyses. Front Plant Sci 11:376.
33. Jean Claude S, Park S (2020) *Aster spathulifolius* Maxim. a leaf transcriptome provides an overall functional characterization, discovery of SSR marker and phylogeny analysis. PLoS ONE 15: e0244132.
34. Geng QF, Liu J, Sun L, Liu H, Ou-Yang Y, Cai Y, Tang XS, Zhang HW, Wang ZS, An SQ (2015) Development and characterization of polymorphic microsatellite markers (SSRs) for an endemic plant, *Pseudolarix amabilis* (Nelson) Rehd. (Pinaceae). Molecules 20: 2685-2692.
35. Xu J, Hou FY, Wan DR, Wang S, Xu DM, Yang GZ (2015) Development and characterization of polymorphic microsatellite markers for *Sedum sarmentosum* (Crassulaceae) and their cross-species transferability. Molecules 20: 19929-19935.

36. Andrés-Sánchez S, Temsch EM, Rico E, Martínez-Ortega MM (2013) Genome size in *Filago* L. (Asteraceae, Gnaphalieae) and related genera: phylogenetic, evolutionary and ecological implications. Plant Syst Evol 299: 331-345.
37. Siniscalchi CM, Loeuille B, De Siqueira Filho JA, Pirani JR (2019) *Chresta artemisiifolia* (Vernonieae, Asteraceae), a new endangered species from a recently created protected area in the Brazilian Caatinga. Phytotaxa 399: 119-126.
38. Sakata Y, Kaneko S, Hayano A, Inoue-Murayama M, Ohgushi T, Isagi Y (2013) Isolation and characterization of microsatellite loci in the invasive herb *Solidago altissima* (Asteraceae). Appl Plant Sci 1: 1200313.
39. López-Caamal A, Reyes-Chilpa R, Tovar-Sánchez E (2018) Hybridization between *Tithonia tubaeformis* and *T. rotundifolia* (Asteraceae) evidenced by nSSR and secondary metabolites. Plant Syst Evol 304: 313-326.
40. Liao R, Luo Y, Yisilam G, Lu R, Wang Y, Li P (2019) Development and characterization of SSR markers for *Sanguinaria canadensis* based on genome skimming. Appl Plant Sci 7: e11289.
41. Kosiński P, Sękiewicz K, Walas Ł, Boratyński A, Dering M (2019) Spatial genetic structure of the endemic alpine plant *Salix serpillifolia*: genetic swamping on nunataks due to secondary colonization?. Alpine Bot 129: 107-121.
42. Cao YN, Comes HP, Sakaguchi S, Chen LY, Qiu YX (2016) Evolution of East Asia's Arcto-tertiary relict *Euptelea* (Eupteleaceae) shaped by late Neogene vicariance and quaternary climate change. BMC Evol Biol 16: 1-17.
43. Zhang CY, Low S, Song YG, Kozlowski G, Van Do T, Li L, Zhou SS, Tan YH, Cao GL, Zhou Z, Meng HH, Li, J (2020) Shining a light on species delimitation in the tree genus *EngelhardiaLeschenault* ex *Blume* (Juglandaceae). Mol Phylogenet Evol 152: 106918.

Tables

Table 1
The location of nine polymorphism cp SSRs in the *Crossostephium chinense* chloroplast genome.

Locus	Sample ID	Start BP	Repeat motif	End BP	Sample ID	Start BP	Repeat motif	End BP	Position
cpSSR1	*C. chinense*-ZJWZ	9640	$(A)_{12}$	9652	*C. chinense*-JPBB	9640	$(A)_{13}$	9653	*trn*C-*pet*N
cpSSR2	*C. chinense*-ZJWZ	23059	$(A)_{11}$	23070	*C. chinense*-JPBB	23060	$(A)_{12}$	23072	*rpo*C2-*rps*2
cpSSR3	*C. chinense*-ZJWZ	25928	$(T)_{11}$	25939	*C. chinense*-JPBB	25930	$(T)_{12}$	25942	*atp*I-*atp*H
cpSSR4	*C. chinense*-ZJWZ	29439	$(T)_{11}$	29450	*C. chinense*-JPBB	29442	$(T)_{10}$	29452	*atp*A-*trn*R
cpSSR5	*C. chinense*-ZJWZ	29451	$(A)_{10}$	29461	*C. chinense*-JPBB	29453	$(A)_{12}$	29465	*atp*A-*trn*R
cpSSR6	*C. chinense*-ZJWZ	34538	$(T)_{11}$	34549	*C. chinense*-JPBB	34541	$(T)_{10}$	34551	*psb*C-*trn*S
cpSSR7	*C. chinense*-ZJWZ	72932	$(A)_{12}$	72944	*C. chinense*-JPBB	72966	$(A)_{11}$	72977	*psb*B-*psb*T
cpSSR8	*C. chinense*-ZJWZ	115011	$(A)_{10}$	115021	*C. chinense*-JPBB	115044	$(A)_{11}$	115055	*ndh*D-*psa*C
cpSSR9	*C. chinense*-ZJWZ	123986	$(A)_{12}$	123998	*C. chinense*-JPBB	124060	$(A)_{11}$	124071	*ycf*1
Note: BP = base pair									

Table 2
The patterns of SNP marker in the *Crossostephium chinense* chloroplast genome.

Position	Gene Type	Gene Region	Substitution Type	*C. chinense*-ZJWZ	*C. chinense*-JPBB	No. of mutations	Region length (bp)	Nucleotide diversity
*trn*K	Intron	LSC	T/C	C	T	1	2,616	0.0004
*psb*M-*trn*D	Spacer	LSC	T/G	G	T	1	667	0.0015
*rpo*C2	Exon	LSC	A/C	A	C	1	4,152	0.0002
*rps*2	Exon	LSC	A/C	A	C	1	711	0.0014
*atp*A-*trn*R	Spacer	LSC	A/T	T	A	2	126	0.0160
	Spacer	LSC	A/C	C	A			
*trn*T	Exon	LSC	A/G	G	A	1	68	0.0147
*psa*A	Exon	LSC	T/G	T	G	1	2,253	0.0004
*ycf*3-*trn*S	Spacer	LSC	T/G	T	G	1	841	0.0012
*trn*F-*ndh*J	Spacer	LSC	A/C	A	C	1	707	0.0014
*psa*I-*ycf*4	Spacer	LSC	T/C	T	C	1	373	0.0027
*trn*P-*psa*J	Spacer	LSC	C/G	C	G	1	308	0.0033
*rpl*20-*rps*12	Spacer	LSC	A/G	A	G	1	726	0.0005
*ndh*F	Exon	SSC	A/G	A	G	1	2,226	
*ndh*A	Exon	SSC	T/G	T	G	2	2,176	0.0009
	Intron	SSC	A/C	C	A			
*ndh*H-*rps*15	Spacer	SSC	T/G	G	T	1	91	0.0110
*ycf*1	Exon	SSC	T/C	C	T	2	5,037	0.0004
	Exon	SSC	A/T	A	T			

Table 3
Characteristics of the ten selected polymorphic nucleotide microsatellite markers for *Crossostephium chinense*.

Locus	Primer sequences (5'–3')	Repeat motif		Allele size range (bp)	*Ta* (°C)	Fluorescent dye
		C. chinense-ZJWZ	*C. chinense*-JPBB			
CC6-8	F: TCGAGCCAAACATGCTGAGA R: TCCAATAAGTGTGTGCTAGCT	$(AC)_{10}$	$(AC)_8$	152–156	56.6	FAM
CC8-11	F: ACCCGGACCTCTAACCTCTC R: TGACGGGGTATGTGAGTCAA	$(AC)_{11}$	$(AC)_{10}$	109–111	55.0	TAMRA
CC16-24	F: CCAATCTTCACCATCCGAGCT R: TCCTTACGATCCTGCAAGCC	$(AC)_7$	$(AC)_8$	139–144	55.0	TAMRA
CC19-30	F: TGGACGTTTGGGAGAGAACA R: GTGTCTCACCCAAGTAGCGA	$(AC)_9$	$(AC)_8$	143–153	56.5	TAMRA
CC32-34	F: AGGCCAGTTTCACGAACCAA R: TTGGCGACACAAACCTAGCT	$(AG)_7$	$(AG)_6$	155–159	55.2	TAMRA
CC33-50	F: CCTCGAAAGTTAACGGTGGGT R: CACCGATCACCACACCTCAA	$(AG)_8$	$(AG)_6$	189–207	55.4	FAM
CC54-95	F: CCGATCCGATCCCAAGCTTT R: GCTGCTGGAAGTTGACTGTC	$(CAA)_{11}$	$(CAA)_{13}$	167–190	54.9	FAM
CC55-96	F: CACGTCAATATCCAACGGCAA R: GGTCCAGGGTCCATTTGGTT	$(CAA)_{14}$	$(CAA)_{12}$	173–181	55.8	HEX
CC62-110	F: CGAAGAGGATCCGAAGCCAA R: ATTCTGGTGCTGGTGGGATG	$(CAC)_5$	$(CAC)_7$	145–152	57.1	TARMA
CC66-118	F: TCTCATCACCCCACTAACCAC R: AGTGGTGGTGGTCACGAATC	$(CAT)_8$	$(CAT)_6$	181–193	56.9	HEX

Table 4

Characteristics of the selected ten polymorphic nuclear microsatellite markers in three populations of *C. chinense* and two species of *Artemisia*.

Locus	*Crossostephium chinense*												*Artemisia stolonifera*			*Artemisia argyi*		
	ZJWZ (*N* = 11)			ZJPY (*N* = 17)			JPSD (*N* = 7)			All (*N* = 35)			AHJH (*N* = 20)			ZJHZ (*N* = 20)		
	A	H_O	H_E[b]	*A*	H_O	H_E[b]	*A*	H_O	H_E[b]	*A*	H_O	H_E[b]	*A*	H_O	H_E[b]	*A*	H_O	H_E[b]
CC6	2	0.091	0.087	1	0.000	0.000	1	0.000	0.000	2	0.029	0.029	6	0.450	0.565***	1	0.000	0.000
CC8	2	0.000	0.298***	1	0.000	0.000	1	0.000	0.000	2	0.000	0.388	1	0.000	0.000	1	0.000	0.000
CC16	2	0.091	0.087	2	0.059	0.057	1	0.000	0.000	3	0.057	0.057	3	0.400	0.436*	3	0.150	0.526***
CC19	3	0.455	0.368	2	0.059	0.057	1	0.000	0.000	4	0.171	0.585***	/	/	/	/	/	/
CC32	1	0.000	0.000	2	0.059	0.057	1	0.000	0.000	2	0.029	0.506***	/	/	/	/	/	/
CC33	6	0.364	0.512**	7	0.294	0.519*	1	0.000	0.000	9	0.257	0.755	/	/	/	/	/	/
CC54	5	0.545	0.450	3	0.235	0.213	2	0.000	0.490**	7	0.286	0.537	/	/	/	/	/	/
CC55	3	0.182	0.169	1	0.000	0.000	2	0.143	0.133	4	0.086	0.535***	/	/	/	/	/	/
CC62	1	0.000	0.000	2	0.059	0.057	1	0.000	0.000	2	0.029	0.029	/	/	/	/	/	/
CC66	4	0.545	0.665*	3	0.235	0.403	2	0.000	0.490**	5	0.286	0.646***	3	0.400	0.329	2	0.050	0.049

Note: *A* = number of alleles per locus; H_E = expected heterozygosity; H_O = observed heterozygosity; *N* = number of individuals sampled.

[a] Locality and voucher information are available in Appendix 1.

[b] Significant deviations from Hardy-Weinberg equilibrium at *P < 0.05, **P < 0.01, and ***P < 0.001, respectively.

Figures

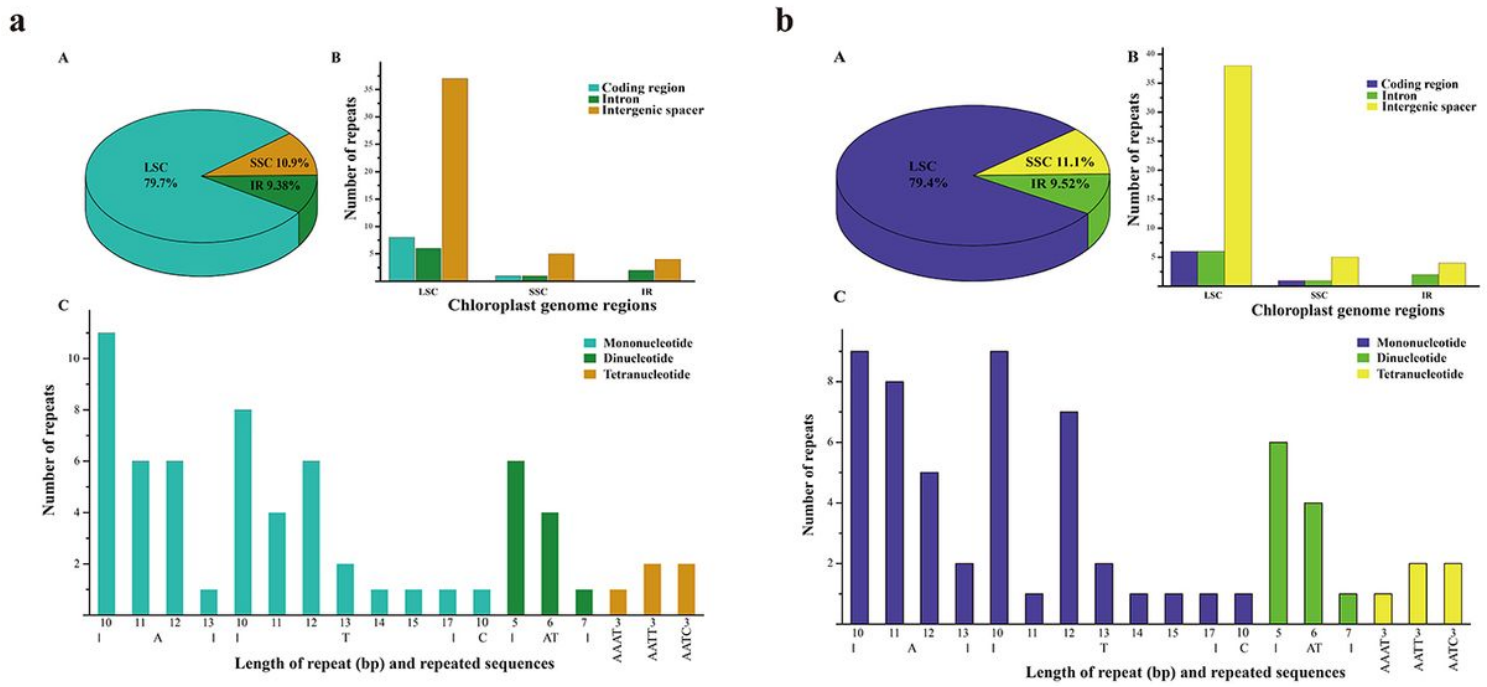

Figure 1

The distribution, type, and presence of chloroplast simple sequence repeats (cpSSRs) in Crossostephium chinense-ZJWZ (a) and -JPBB (b). (A) Presence of cpSSRs in the LSC, SSC, and IR regions. (B) Presence of cpSSRs in the protein-coding regions, intergenic spacers and introns of LSC, SSC, and IR regions. (C) Presence of polymers in the cp genome.

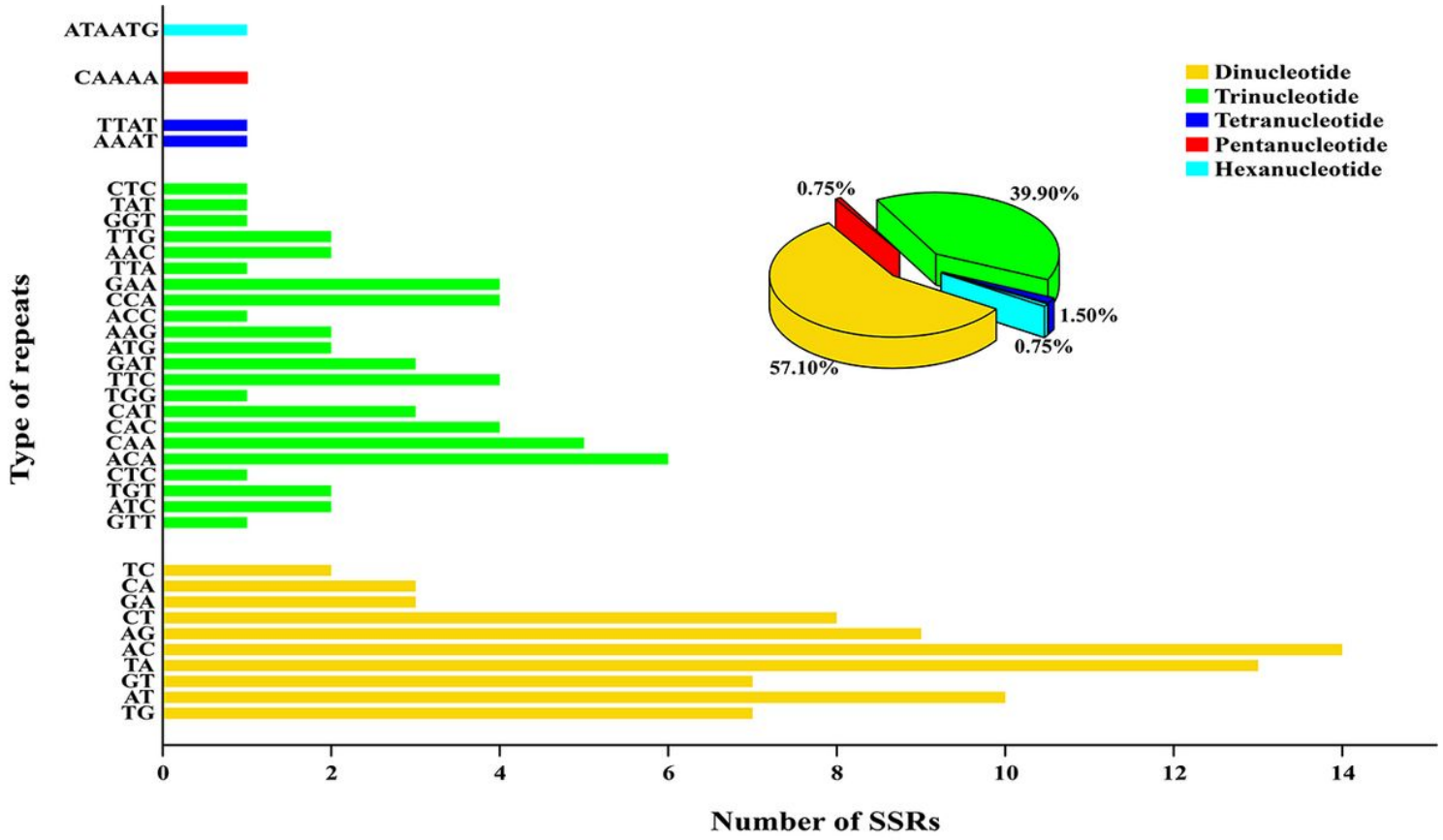

Figure 2

The distribution of polymorphic nuclear simple sequence repeats (nSSRs) for Crossostephium chinense.

Supplementary Files

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

- TableS1.docx
- TableS2.xlsx