Posted on Authorea 14 Oct 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.166573612.23507910/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Phylogeographic analysis of Siraitia grosvenorii in subtropical China provides insight into the origin of cultivated monk fruit and conservation of genetic resources

Bingbin Xie[1], Bowen Lai[1], Liping Chen[1], Sujuan Wei[1], and Shaoqing Tang[1]

[1]Guangxi Normal University

October 14, 2022

Abstract

Siraitia grosvenorii, an economically important plant species with high medicinal value, is endemic to subtropical China. To determine the population structure and origin of cultivated S. grosvenorii, we examined the variation in three chloroplast DNA regions (trnR-atpA, trnH-psbA, trnL-trnF) and two nuclear gene orthologs (CHS and EDL2) of S. grosvenorii in 130 individuals, selected from 13 wild populations across its natural distribution range, and 21 cultivated accessions using a phylogeographic approach. The results showed non-overlapping distribution of chlorotypes, three distinct chloroplast genetic groups restricted to different mountain ranges, and comparable nuclear diversity among the distinct geographical groups, suggesting the existence of at least three separate refugia. The current phylogeographic patterns of S. grosvenorii probably resulted from long-term survival in multiple refugia and limited expansion. Our results also demonstrated that wild populations in northeastern Guangxi share the same gene pool as cultivated S. grosvenorii accessions, suggesting that the current cultivars originated from wild populations distributed in northeastern Guangxi. The results of this study provide insight into improving the efficiency of S. grosvenorii breeding using a genetic approach, and outline measures for the conservation of its genetic resources.

Phylogeographic analysis of *Siraitia grosvenorii* in subtropical China provides insight into the origin of cultivated monk fruit and conservation of genetic resources

Bingbin Xie[a, b#], Bowen Lai[a, b#], Liping Chen[a, b], Sujuan Wei[a, b], Shaoqing Tang[a, b*]

[a] Key Laboratory of Ecology of Rare and Endangered Species and Environmental Protection, Ministry of Education, Guangxi Normal University, Guilin 541004, China

[b] Guangxi Key Laboratory of Rare and Endangered Animal Ecology, College of Life Science, Guangxi Normal University, Guilin 541004, China

*Author for correspondence. College of Life Science, Guangxi Normal University, Guilin 541006, China.

Email: shaoqing@mailbox.gxnu.edu.cn

Bingbin Xie and Bowen Lai should be considered joint first author

Abstract

Siraitia grosvenorii , an economically important plant species with high medicinal value, is endemic to subtropical China. To determine the population structure and origin of cultivated *S. grosvenorii* , we examined the variation in three chloroplast DNA regions (*trn* R-*atp* A,*trn* H-*psb* A, *trn* L-*trn* F) and two nuclear gene orthologs (*CHS* and *EDL2*) of *S. grosvenorii* in 130 individuals, selected from 13 wild populations across its natural distribution range, and 21 cultivated accessions using a phylogeographic approach. The results showed non- overlapping distribution of chlorotypes, three distinct chloroplast genetic groups restricted to

Posted on Authorea 14 Oct 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.166573612.23507910/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

different mountain ranges, and comparable nuclear diversity among the distinct geographical groups, suggesting the existence of at least three separate refugia. The current phylogeographic patterns of *S. grosvenorii* probably resulted from long- term survival in multiple refugia and limited expansion. Our results also demonstrated that wild populations in northeastern Guangxi share the same gene pool as cultivated *S. grosvenorii* accessions, suggesting that the current cultivars originated from wild populations distributed in northeastern Guangxi. The results of this study provide insight into improving the efficiency of *S. grosvenorii* breeding using a genetic approach, and outline measures for the conservation of its genetic resources.

Keywords:*Siraitia grosvenorii* , phylogeography, population structure, chloroplast DNA, nuclear gene ortholog

Dear Editor,

We are pleased to submit a manuscript entitled "**Phylogeographic analysis of *Siraitia grosvenorii* in subtropical China provides insight into the origin of cultivated monk fruit and conservation of genetic resources** " for publication as a reserach article in*Ecology and Evolution* .

Siraitia grosvenorii (Cucurbitaceae) is an important economic crop in Guangxi with high value of medicinal. The population genetic structure of this species at different geographical scales and its evolutionary history, also the genetic origin of cultivated *S. grosvenorii* is still unknown. In this new manuscript, we reported the results of the phylogeographical study on *S. grosvenorii* in subtropical China. This study provides a genetic basis for the utilization and conservation of *S. grosvenorii* germplasm.

All authors contributed significantly to the work. All authors have read and approved the final manuscript and agree with its submission to*Ecology and Evolution* . We confirm that this manuscript has not been published elsewhere and is not under consideration by any other journal. We have no conflicts of interest to declare.

Thank you for your consideration our manuscript. We look forward to hearing from you.

Sincerely,

Shaoqing Tang, Ph.D.

College of Life Science,

Guangxi Normal University,

Guilin,

China.

1. Introduction

Siraitia grosvenorii (Cucurbitaceae), also known as monk fruit or Luohanguo, is an economically important crop in Guangxi, China. *S. grosvenorii* has been cultivated for more than 100 years and is prevalent in the Yongfu, Longsheng, and Lingui counties of Guilin (Guangxi, China). *S. grosvenorii* was introduced from Guangxi to other provinces in southern China, which was followed by an attempt to introduce the species into India (Shivani et al., 2021). In addition to its use as Chinese traditional medicine (Chinese Pharmacopoeia Commission, 2015), *S. grosvenorii* is highly valued in the food industry. The*S. grosvenorii* fruit contains mogrosides, which are approximately 300 times sweeter than sucrose (Kasai et al., 1989). Because mogrosides are natural and low- calorie compounds, the mogroside extract from*S. grosvenorii* is considered to be an ideal sugar substitute (Yan et al., 2008). Given the increasing demand for non- nutritive sweeteners from natural sources, *S. grosvenorii* sweeteners have been widely used in the food and beverage industries for the development of low- calorie products (Pandey & Chauhan, 2019). In 2020,*S. grosvenorii* was planted on an area of over 10,000 hm^2 in Guilin, accounting for more than 85% of the global production (Lu et al., 2022). However, because of the continuous vegetative propagation of *S. grosvenorii* , its cultivated accessions have become more vulnerable to pests and diseases (Li et al., 2004). Wild genetic resources allow plant breeders

Posted on Authorea 14 Oct 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.166573612.23507910/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

to select and breed for the desired characteristics, which facilitates not only the maintenance but also the improvement of agricultural productivity (Day-Rubenstein et al., 2005; Prescott-Allen & Prescott-Allen, 2013). To fully utilize the genetic resources of*S. grosvenorii* , a proper understanding of the population structure of its wild accessions and the origin of its cultivated accessions is greatly needed.

S. grosvenorii is naturally distributed in subtropical China, and mainly grows in the Nanling, Dayaoshan, Yunkai, Jiulianshan, and Wugongshan Mountains at an altitude of 400–1400 meters. Global climatic fluctuations, particularly Quaternary climate oscillations, might have profoundly shaped the distribution range and genetic structure of plants (Abbott et al., 2000; Avise, 2000). Although glaciers are not found at low elevations (< 2500 m) in subtropical China (Ni et al., 2010), a cold and dry climate has developed in this region because of the intensification of winter monsoons (Shi et al., 2006; Wang et al., 2012). *S. grosvenorii* diverged from members of the Cucurbitaceae family approximately 40–60 million years ago (Xia et al., 2018; Guo et al., 2020), and its current genetic structure and diversity across its distributional range might be influenced by the Quaternary climate changes. Previously, the genetic diversity of wild accessions and cultivars of *S. grosvenorii* was analyzed using random amplified polymorphic DNA (RAPD), inter-simple sequence repeat (ISSR), and amplified fragment length polymorphism (AFLP) markers, which revealed a high level of genetic differentiation among wild populations (Tang et al., 2007a). Additionally, the genetic diversity of cultivars was discovered to be much lower than that of wild accessions (Peng et al., 2005; Zhou et al., 2005; Zhou & Tang, 2006; Tang et al., 2007a). Based on RAPD markers, Tang et al. (2007b) speculated that clonal growth plays a role in shaping the spatial genetic structure of *S. grosvenorii* . However, the population genetic structure of*S. grosvenorii* at different geographical scales and its evolutionary history, such as potential refugia and population expansion, remain unknown, mainly because the DNA markers used in previous studies offer limited utility in inferring population structure and dynamics.

In this study, we report the results of the phylogeographical analysis of *S. grosvenorii* in subtropical China using three chloroplast DNA (cpDNA) fragments and two nuclear gene orthologs. We aimed to (1) investigate the genetic structure of *S. grosvenorii* , and infer its main potential refugial locations and demographical history, and (2) determine the geographical origin and ancestral populations of cultivated *S. grosvenorii* . Information on the population structure and demographical history of *S. grosvenorii*is important for the genetic improvement of its cultivars and the conservation of its wild germplasm.

2. Materials and methods

2.1 Sampling

A total of 151 individuals of*S. grosvenorii* were collected in this study, including 130 individuals selected from 13 wild populations and 21 individuals (16 female and 5 male) of 7 cultivated accessions (hereafter referred to as cultivars). Details of sampling sites are shown in Table 1. To avoid duplicate sampling, plants were sampled from wild populations located approximately 10 m apart, and 10 individuals were samples from each population. Fresh leaves were dried in silica gel for subsequent DNA extraction.

2.2 DNA extraction, PCR amplification, and DNA sequencing

Genomic DNA was extracted from the dried leaves using the modified cetyltrimethylammonium bromide (CTAB) method (Doyle & Doyle, 1987). Three highly polymorphic cpDNA fragments, including*trn* R-*atp* A (Dane & Lang, 2004), *trn* H-*psb* A (Sang et al., 1997), and *trn* L-*trn* F (Taberlet et al., 1991), and two nuclear genes, including *chalcone synthase*(*CHS*) and *EID1-like F-box protein 2* (*EDL2*), were analyzed in this study. Nucleotide sequences of the *S. grosvenorii CHS* gene (GenBank accession no. GU980155.1) and*Momordica charantia EDL2* gene (GenBank accession no. XR_002601804.1) were downloaded from the National Center for Biotechnology Information (NCBI), and both these genes were identified as orthologs using the OrthoMCL database (*http://orthomcl.org/orthomcl/*). The nuclear genes were PCR amplified with sequence-specific primers designed using Primer3Web (*http://primer3.wi.mit.edu/*). Primer details are given in Table 2. The amplification was carried out in 50-μL reactions, each containing 0.5 μL of genomic DNA, 5 μL of 10× PCR buffer (Mg^{2+}plus), 4 μL of dNTP mix (2.5 μM), 0.5 μL of each primer (50 mM), and 2.5 U of ExTaq DNA polymerase (TaKaRa, Dalian, China). The amplification program was as follows:

Posted on Authorea 14 Oct 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.166573612.23507910/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

pre-denaturation at 94 °C for 5 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at various temperatures (Table 2) for 30 s, and elongation at 72 °C for 1 min, and a final extension at 72 °C for 10 min. PCR products were purified and sequenced by Sangon Biotech (Shanghai, China).

Genes with double peaks at polymorphic sites was regarded as heterozygous. Haplotypes were determined using the method described by Clark (1990) if the gene sequence contained a single heterozygous site, and through the cloning and sequencing of PCR products if the sequence contained two or more peaks. To clone the gene, DNA was amplified using high-fidelity PrimeSTAR HS DNA polymerase (TaKaRa, Dalian, China), as described above. The cloned DNA was purified and sequenced by Sangon Biotech (Shanghai, China); 10 clones were sequenced per PCR product.

2.3 Sequence analysis

DNA sequences were aligned using MAGA 7.0.14 (Kumar et al., 2016), and edited manually in BioEdit 7.0.1 (Hall, 1999) where necessary. A haplotype network map was constructed using the median-joining method in Network 5.0 (Bandelt et al., 1999), and the geographic distribution of haplotypes was drawn on the map using ArcMap GIS (ESRI, Redlands, CA, USA). Haplotype diversity (h) and nucleotide diversity (π) were calculated using DNAsp 5.0 (Librado & Rozas, 2009). A phylogenetic tree of cpDNA haplotypes was constructed using the Maximum Likelihood (ML) method, and implemented with the IQ-TREE program (Nguyen et al., 2015).*Siraitia siamensis* (GenBank accession no. MK75585) was selected as the outgroup. Total genetic diversity (H_T) and the genetic diversity within a population (H_S) were calculated using PERMUT version 2.0 (Pons & Petit, 1996). To infer the phylogeographic structure of *S. grosvenorii* , significant differences between G_{ST} andN_{ST} were calculated using PERMUT version 2.0 (Pons & Petit, 1996). Analyses of molecular variance (AMOVA) were performed using Arlequin v3.11 (Excoffier et al., 2005). In addition, AMOVA analyses with three chlorotype-based genetic groups identified in this study were also conducted. Genetic variation was quantified at three hierarchical levels: among populations, within population, and among groups of populations identified by three chlorotype-based genetic groups found in this study.

The historical population dynamics of *S. grosvenorii* was determined by performing neutrality tests and estimating mismatch distribution. Neutrality tests were performed by calculating Tajima'sD (Tajima, 1989) and Fu and Li's D (Fu & Li, 1993) using DNAsp 5.0 (Librado & Rozas, 2009), and Fu's F s (Fu, 1997) using Arlequin v3.11 (Excoffier et al., 2005). Mismatch distribution analyses were performed using a sudden (stepwise) expansion model (Rogers & Harpending, 1992) with DNAsp 5.0 (Librado & Rozas, 2009). The goodness-of-fit between observed and expected mismatch distributions was tested by calculating the sum of squared deviations (SSD) and raggedness index (H_{Rag}) in Arlequin v3.11 (Excoffier et al., 2005).

3. Results

3.1 Population structure based on chlorotypes

Three cpDNA regions (*trn* R-*atp* A, *trn* H-*psb* A, and *trn* L-*trn* F) were amplified from 151 individuals and aligned. The alignment of the three cpDNA sequences was 1225 bp in length, and contained 21 polymorphic sites, resulting in six chlorotypes (C1–C6). The chlorotype sequences of *trn* R-*atp* A,*trn* H-*psb* A, and *trn* L-*trn* F were deposited in GenBank under the accession numbers MK357026–MK357031, MK356960–MK356965, and MK357032–MK357037, respectively.

The different chlorotypes and their network and geographical distribution are presented in Table 3 and Figure 1a and c. Most populations contained only one chlorotype, and only the JX population contained two chorotypes (C1, C4). Thus, within-population diversity was observed only in the JX population (Table 1). Values of total $h$and π were 0.735 and 0.00614, respectively (Table 1). No ancestral haplotype was found in the chlorotype network. C1 was the most abundant chlorotype, and was shared by five populations (MES, SJ, BL, JT, JX) and cultivars. Two chlorotypes were shared by sets of three different populations; chlorotype C2 was shared by HZ, DX, and ZQ, and chlorotype C3 by SG, JLS, and WGS. Chlorotypes C4, C5, and C6 were unique to populations JX, PB, and RY, respectively. Chlorotype C4 was separated from

Posted on Authorea 14 Oct 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.166573612.23507910/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

C1 by one mutational step, and chlorotype C1 was separated from C5 by 10 mutational steps. The other part of the network contained three chlorotypes (C2, C3, and C6), which were found in seven populations (HZ, DX, ZQ, SG, JLS, WGS, RY). Chlorotype C6 was separated from C3 by two mutational steps.

The results of network analysis were corroborated by the phylogenetic analysis of chlorotypes. The phylogenetic tree constructed using the ML method divided the six chlorotypes into three genetic groups (Figure 1d). Chlorotype C5 formed a separate clade (YK group); chlorotypes C1 and C4 clustered together in the YD group; and chlorotypes C2, C3, and C6 formed the MD group. Total genetic diversity across all wild populations ($H_T = 0.797$) was higher than the average within-population genetic diversity ($H_S = 0.062$), resulting in high population differentiation ($G_{ST} = 0.922$, $N_{ST} = 0.964$). A permutation test revealed no significant phylogeographic structure in *S. grosvenorii*($N_{ST} > G_{ST}$; $P > 0.05$). The overall AMOVA revealed that 100% of the molecular variation in cpDNA was found among the 13 populations, whereas 88.80% was found among the three cpDNA-based phylogenetic groups (YK, YD, and MD) (Figure 1, Table 6).

3.2 Population structure based on *CHS* and *EDL2* haplotypes

The *CHS* sequence alignment was 1014 bp in length and contained 25 polymorphic sites. Forty-five *CHS* haplotypes (H1–H45, GenBank accession nos. MK356966–MK357010) were identified among all samples surveyed. Because 30 of the 45 haplotypes were unique to a single population, the inference of the ancient haplotype was difficult. Haplotype H12 was shared by seven populations (MES, SJ, HZ, DX, SG, WGS, PB), making it the most common and the most widely distributed haplotype (Figure 2). Additionally, multiple haplotypes such as H1 and H5 were derived from haplotype H12 (Figure 2). Six haplotypes (H2, H4, H6, H8, H10, H31) were found in two wild populations (MES and JT) and in cultivars. Populations BL and JX contained only H5 and H3 haplotypes, respectively (Figure 2, Table 4). The total values of h and π were 0.914 (range: 0–0.932) and 0.00417 (range: 0–0.00423), respectively (Table 1). The highest values of h (0.932) and π (0.00421) were found in the JLS population (Table 1).

The *EDL2* sequence alignment was 607 bp in length, and contained 13 polymorphic sites, resulting in 15 haplotypes (E1–E15, GenBank accession nos. MK357011–MK357025). The *EDL2* haplotype network showed a 'star-like' topology, with haplotype E3 at the center (Figure 3). Eight haplotypes (E2, E4, E6, E7, E9, E10, E12, E14) were separated from haplotype E3 by one mutational step. In addition, E3 was the most common and the most widely distributed haplotype, because it was found in all populations, except HZ and WGS. This suggests that E3 is likely the ancient haplotype. Six haplotypes (E4, E10, E11, E13, E14, E15) were exclusive to a single population (Figure 3, Table 5). The WGS population contained only the E9 haplotype. The total values of h and π inferred from the *EDL2* data were 0.834 (range: 0–0.774) and 0.00255 (range: 0–0.00315), respectively (Table 1). The highest values of h (0.774) and π (0.00263) were found in population JLS (Table 1).

Similar to the results of cpDNA data analysis, the H_T of nuclear genes (0.965 for *CHS* , 0.559 for *EDL2*) was higher than the H_S of these genes (0.904 for *CHS* , 0.589 for *EDL2*). Although the N_{ST} values of both genes (0.464 for *CHS* , 0.349 for *EDL2*) were higher than the corresponding G_{ST} values (0.420 for *CHS* , 0.344 for *EDL2*), the difference was not significant ($P > 0.05$), suggesting that the phylogeographic structure, based on the nuclear genes, was also not significant. The results of AMOVA showed that the variation in *CHS* sequence among populations (50.95%) was only slightly higher than that within populations (49.05%) (Table 6). In contrast to the results of cpDNA and *CHS* data analyses, the among-population variation in *EDL2* (37.86%) was lower than the within-population variation (62.14%) (Table 6). When the populations were divided into three chlorotype-based genetic groups, the sequence data of both *CHS* and *EDL2* showed a similar level of within-population variation (47.40% for *CHS* , 59.71% for *EDL2*), but the among–population variation in both *CHS* and *EDL2* was much lower (8.45% for *CHS* , 1.19% for *EDL2*) (Table 6).

3.3 Demographic history of *S. grosvenorii*

The values of Tajima's D , Fu and Li's D , and Fu's F s for cpDNA haplotypes of *S. grosvenorii* were positive (Table 7). In addition, the mismatch distribution was multimodal (Figure 4a), and both the SSD and H_{Rag}

Posted on Authorea 14 Oct 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.166573612.23507910/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

statistics were significant (Table 7), indicating that *S. grosvenorii* did not undergo population expansion. The nonsignificant Tajima' s D and Fu and Li's D values of the two nuclear genes and the significant SSD and H_{Rag} statistics of *EDL2* also rejected the hypothesis of rapid expansion (Table 7). However, the resulting unimodal mismatch distribution detected by *EDL2* data in*S. grosvenorii* may indicated a model of sudden expansion. This pattern was also indicated by the significant Fu's *Fs* value (-16.58458, $P < 0.001$) as well as the nonsignificant SSD and H_{Rag}statistics ($P < 0.05$) of *CHS* .

4. Discussion

4.1 Potential refugia and phylogeographic structure of *S. grosvenorii*

The chlorotypes of *S. grosvenorii* were confined to different mountain ranges, showing a clear-cut geographical distribution pattern of *S. grosvenorii* (Figure 1). In addition, the clear chlorotype-based structuring of *S. grosvenorii* into three genetically differentiated groups was consistent with three distinct geographical groups (Figure 1). Additionally, the sequences of nuclear genes showed no evidence of a reduction in the genetic diversity of*S. grosvenorii* throughout its natural distribution range but revealed comparable levels of genetic diversity in each of the three geographical groups. These results suggest that *S. grosvenorii*occupied at least three separate refugia during the last glaciation in southern China. The first major refugium, including the MES, SJ, BL, JT, and JX populations, might have been located in Yuechengling-Dayaoshan Mountains (in the western range of Nanling Mountains), which was represented by chlorotype C1 (the ancestral haplotype in its clade) and C4. The populations of HZ, DX, ZQ, JLS, RY, SG, and WGS harbored chlorotypes C2, C3, and C6, located in Mengzhuling-Dayuling Mountains (in the central and eastern Nanling Mountains), is speculated to be the second refugium for *S. grosvenorii.* The third refugium probably occurred in Yunkai Mountains, which houses only the PB population and the C5 chlorotype. The Nanling and Yunkai Mountains, which are characterized by a rugged topography, provided stability by buffering the Pleistocene fluctuations (Qiu et al., 2011), and are well recognized as refugia of many other subtropical species, such as *Eurycorymbus cavaleriei* (Wang et al., 2009), *Pinus kwangtungensis*(Tian et al., 2010), *Loropetalum chinense* (Gong et al., 2016),*Bretschneidera sinensis* (Wang et al., 2018), and *Eomecon chionantha* (Tian et al., 2018).

Chlorotypes of *S. grosvenorii* were private to and fixed in single local populations, indicating harbor stable long-term refugia (Petit et al., 2003; Médail & Diadema, 2009; Qiu et al., 2011). In addition, the long genetic distances among three chlorotype clades (10 or 12 mutations) imply that populations in each refugia region have been isolated for a certain period of time, given the low mutation rate of cpDNA (Petit & Vendramin, 2007). Conversely, neutrality tests and mismatch distribution analysis suggest that limited population expansion in *S. grosvenorii* . If any population expansion occurred, it occurred on a limited geographical scale (Figure 1), as is apparent for refugia regions in Yuechengling-Dayaoshan Mountains and Mengzhuling-Dayuling Mountains (the origin of private chlorotypes C4 and C6). The results of AMOVA based on cpDNA also indicated that the variation among groups (88.8%) was higher than that among populations within a group (11.2%) (Table 6). Such a phylogeographic pattern has been observed in most plant species in subtropical China, which suggests long-tern survival of species in multiple Quaternary refugia and limited expansion (Lopez-Pujol et al., 2011; Qiu et al., 2011; Lei et al., 2012; Zhang et al., 2013).

4.2 Origin of cultivated *S. grosvenorii* and implications forimprovement and conservation of genetic resources

Population genetics-based insights will enable a better understanding the origin of the crop and help to identify raw material for breeding and crop improvement (Ramanatha & Hodgkin, 2002; Turner-Hissong et al., 2020). *S. grosvenorii* is documented to have been cultivated in Longsheng, Yongfu, and Lingui counties for at least 140 years (Zhou et al., 1981; Yang, 2004), and then was introduced from Guangxi to Guangdong, Hunan, Jiangxi, and Guizhou Provinces (Li et al., 2004). Our data revealed the presence of only one chlorotype (C1) in cultivated*S. grosvenorii* (Figure 1), which suggests that *S. grosvenorii* was domesticated probably from a single center of origin. In addition, populations from Yuechengling-Dayaoshan Mountains shared the same chlorotype (C1) with cultivars, as well as large numbers of nuclear haplotypes, indicating that currently main cultivars in Yongfu, Lingui, and Longsheng counties have originated from wild popula-

Posted on Authorea 14 Oct 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.166573612.23507910/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

tions distributed in northeastern Guangxi, consistent with the principals of "nearby domestication". Indeed, the initial cultivation of *S. grosvenorii* followed traditional practices, whereby plants were directly collected from the local wild resources and then cultivated in the field (Li et al., 2004).

Although wild populations of *S. grosvenorii* are believed to possess abundant variation (H_T= 0.797 for cpDNA; $H_T = 0.965$ for *CHS* ; and$H_T = 0.559$ for*EDL* 2), it is obvious that gene-rich pools of wild relatives have not been exploited for the improvement of cultivars. The three genetic groups identified in this study contained a large number of unique haplotypes (Figures 2 and 3, Tables 4 and 5). This unique composition of the genetic groups shows a great potential for improving the cultivated varieties. However, during the field work, we noted that wild populations of *S. grosvenorii* had suffered a rapid decline, and were even extirpated at some distribution points, because of habitat deterioration. Notably, the YK group contained only one population (PB). Establishing a germplasm bank for *S. grosvenorii* is critical for conservation purposes, and germplasm collections should be exhaustive to compensate for the unique genetic compositions. The YK group (population PB) deserves particular attention as a prime ex situ conservation target. Additionally, we propose that populations from the MD group and from the geographical origin of cultivars (YD group) should be utilized as the plant material or core gene pool for breeding and crop improvement, and both the in situ and ex situ methods should be adopted.

5. Conclusions

The molecular data collected in the current study reveal the population genetic structure of *S. grosvenorii* , with at least three separate refugia during glacial periods in subtropical China. Our results suggest that *S. grosvenorii* was exposed to long-term geographic isolation in these three refugia, and experienced limited inter/postglacial population expansion, which might be responsible for its current phylogeographic patterns.*S. grosvenorii* cultivars have originated from wild populations in northeastern Guangxi. This phylogeographical study also provides a genetic basis for the utilization and conservation of *S. grosvenorii* germplasm.

Author contributions

Bingbin Xie: Conceptualization (Equal); Data curation (Lead); Formal analysis (Lead); Investigation (Equal); Methodology (Lead); Software (Lead). **Bowen Lai** : Conceptualization (Equal); Data curation (Supporting); Formal analysis (Supporting); Writing – original draft (Lead); Writing – review & editing (Supporting). **Liping Chen** : Data curation (Supporting); Formal analysis (Supporting); Investigation (Equal); Methodology (Supporting); Software (Supporting). **Sujuan Wei** : Conceptualization (Equal); Methodology (Supporting); Supervision (Supporting); Validation (Supporting); Writing – review & editing (Equal). **Shaoqing Tang** : Conceptualization (Equal); Funding acquisition (Lead); Project administration (Lead); Supervision (Lead); Validation (Lead); Visualization (Lead); Writing – review & editing (Equal).

Acknowledgments

All authors contributed equally to this study. This work was supported by the 2018 Chinese Medicine Public Health Service Subsidy Special "National Chinese Medicine Resources Census Project" (No. GXZYZYPC18-7-2).

Conflict of Interest

The author declares that they have no conflict of interest.

Data Availability Statement

The data that support the findings of this study are openly available in the the NCBI Nucleotide database at https://www.ncbi.nlm.nih.gov/nuccore/, under accession number: MK356960–MK357037.

References

Abbott, R. J., Smith, L. C., Milne, R. I., Crawford, R. MM., Wolff, K., & Balfour, J. (2000). Molecular analysis of plant migration and refugia in the Arctic. *Science* , *289* (5483), 1343-1346.

Posted on Authorea 14 Oct 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.166573612.23507910/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Avise, J. C. (2000). *Phylogeography: the history and formation of species* . Harvard University Press.

Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* , *16* (1), 37-48.

Chinese Pharmacopoeia Commission. (2015). *Pharmacopoeia of the People's Republic of China, Volume I* . China Medical Science Press.

Clark, A. G. (1990). Inference of haplotypes from PCR-amplified samples of diploid populations. *Molecular Biology and Evolution* ,*7* (2), 111-122.

Dane, F., & Lang, P. (2004). Sequence variation at cpDNA regions of watermelon and related wild species: implications for the evolution of*Citrullus* haplotypes. *American Journal of Botany* ,*91* (11), 1922-1929.

Day-Rubenstein, K. A., Heisey, P. W., Shoemaker, R. A., Sullivan, J., & Frisvold, G. B. (2005). Crop genetic resources: an economic appraisal.

Economic Information Bulletin Number 2.

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phyotochemical Bulletin* , *19* , 11-15.

Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary bioinformatics* , *1* (4), 47-50.

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection.*Genetics* , *147* (2), 915-925.

Fu, Y. X., & Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* , *133* (3), 693-709.

Gong, W., Liu, W. Z., Gu, L., Kaneko, S. G., Koch, M. A., & Zhang, D. X. (2016). From glacial refugia to wide distribution range: demographic expansion of *Loropetalum chinense* (Hamamelidaceae) in Chinese subtropical evergreen broadleaved forest. *Organisms Diversity and Evolution* , *16* (1), 23-38.

Guo, J., Xu, W. B., Hu, Y., Huang, J., Zhao, Y. Y., Zhang, L., Huang, C. H., & Ma, H. (2020). Phylotranscriptomics in Cucurbitaceae reveal multiple whole-genome duplications and key morphological and molecular innovations. *Molecular Plant* , *13* (8), 1117-1133.

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

Kasai, R., Nie, R. L., Nashi, K., Ohtani, K., Zhou, J., Tao, G. D., & Tanaka, O. (1989). Sweet cucurbitane glycosides from fruits of*Siraitia siamensis* (chi-zi luo-han-guo), a Chinese folk medicine.*Agricultural and Biological Chemistry* , *53* (12), 3347-3349.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets.*Molecular Biology and Evolution* , *33* (7), 1870-1874.

Lei, M., Wang, Q., Wu, Z. J., López-Pujol, J., Li, D. Z., & Zhang, Z. Y. (2012). Molecular phylogeography of *Fagus engleriana*(Fagaceae) in subtropical China: limited admixture among multiple refugia. *Tree Genentic and Genomes* , *8* (6), 1203-1212.

Li, F., Li, D. P., Jiang, S. Y., & Zhang, H. R. (2004).*Cultivation, development and utilization of Luo Han Guo* . China Forestry Publishing House.

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

Lopez-Pujol, J., Zhang, F. M., Sun, H. Q., Ying, T. S., & Ge, S. (2011). Mountains of southern China as "Plant Museums" and "Plant Cradles": evolutionary and conservation insights. *Mountain Research and Development* , *31* (3), 261-269.

Posted on Authorea 14 Oct 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.166573612.23507910/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Lu, Y., Qin, S. Y., Hu, T. T., & Wei, Z. X. (2022). Create a ten billion "sweet industry" in Guangxi. *Sci-Tech and Development of Enterprise* ,(2), 25-27.

Médail, F., & Diadema, K. (2009). Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography* , *36* (7), 1333-1345.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* ,*32* (1), 268-274.

Ni, J., Yu, G., Harrison, S. P., & Prentice, I. C. (2010). Palaeovegetation in China during the late Quaternary: Biome reconstructions based on a global scheme of plant functional types.*Palaeogeography, Palaeoclimatology, Palaeoecology* , *289* (1), 44-61.

Pandey, A. K., & Chauhan, O. P. (2019). Monk fruit (*Siraitia grosvenorii*)-health aspects and food applications. *Pantnagar Journal of Research* , *17* (3), 191-198.

Peng, Y. T., Tang, S. Q., Li, B. L., & Liu, Y. H. (2005). Genetic diversity of *Siraitia grosvenorii* detected by ISSR markers.*Biodiversity Science* , *13* (1), 36-42.

Petit, R. J., & Vendramin, G. G. (2007). Plant phylogeography based on organelle genes: an introduction, In: S. Weiss, & N. Ferrand (Eds.),*Phylogeography of southern European refugia: evolutionary perspectives on the origins and conservation of European biodiversity*(pp. 23-97). Springer Netherlands.

Petit, R.J., Aguinagalde, I., Beaulieu de, J. L., Bittkau, C., & Brewer, S. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* , *300* (5625), 1563-1565.

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

Prescott-Allen, R., & Prescott-Allen, C. (2013). *Genes from the wild: using wild genetic resources for food and raw materials* . Routledge.

Qiu, Y. X., Fu, C. X., & Comes, H. P. (2011). Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. *Molecular Phylogenetics and Evolution* , *59* (1), 225-244.

Ramanatha, R. V., & Hodgkin, T. (2002). Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell, Tissue and Organ Culture* , *68* (1), 1-19.

Rogers, A. R., & Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* , *9* (3), 552-569.

Sang, T., Crawford, D. J., & Stuessy, T. F. (1997). Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia*(Paeoniaceae). *American Journal of Botany* , *84* (8), 1120-1136.

Shi, Y. F., Cui, Z. J., & Su, Z. (2006). *The Quaternary glaciations and environmental variations in China* . Hebei Science and Technology Press.

Shivani, Thakur, B. K., Mallikarjun, C. P., Mahajan, M., Kapoor, P., Malhotra, J., Dhiman, R., Kumar, D., Pal, P. K., & Kumar, S. (2021). Introduction, adaptation and characterization of monk fruit (*Siraitia grosvenorii*): a non-caloric new natural sweetener.*Scientific Reports* , *11* (1), 1-13.

Taberlet, P., Gielly, L., Pautou, G., & Bouvet, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* , *17* (5), 1105-1109.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* , *123* (3), 585-595.

Posted on Authorea 14 Oct 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.166573612.23507910/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Tang, S. Q., Bin, X. Y., Peng, Y. T., Zhou, J. Y., Wang, L., & Zhong, Y. (2007a). Assessment of genetic diversity in cultivars and wild accessions of Luohanguo (*Siraitia grosvenorii* [Swingle] AM Lu et ZY Zhang), a species with edible and medicinal sweet fruits endemic to southern China, using RAPD and AFLP markers. *Genetic Resources and Crop Evolution* , *54* (5), 1053-1061.

Tang, S. Q., Li, Y., Geng, Y., Zhang, G. R., Wang, L., & Zhong, Y. (2007b). Clonal and spatial genetic structure in natural populations of Luohanguo (*Siraitia grosvenorii*), an economic species endemic to South China, as revealed by RAPD markers. *Biochemical Systematics and Ecology* , *35* (9), 557-565.

Tian, S., Kou, Y. X., Zhang, Z. R., Yuan, L., Li, D.R., López-Pujol, J., Fan, D. M., & Zhang, Z. Y. (2018). Phylogeography of *Eomecon chionantha* in subtropical China: the dual roles of the Nanling Mountains as a glacial refugium and a dispersal corridor. *BMC Evolutionary Biology* , *18* (1), 1-12.

Tian, S., López-Pujol, J., Wang, H. W., Ge, S., & Zhang, Z. Y. (2010). Molecular evidence for glacial expansion and interglacial retreat during Quaternary climatic changes in a montane temperate pine (*Pinus kwangtungensis* Chun ex Tsiang) in southern China. *Plant Systematics and Evolution* , *284* (3), 219-229.

Turner-Hissong, S. D., Mabry, M. E., Beissinger, T. M., Ross-Ibarra, J., & Pires, J. C. (2020). Evolutionary insights into plant breeding.*Current Opinion in Plant Biology* , *54* , 93-100.

Wang, J., Gao, P. X., Kang, M., Lowe, A. J., & Huang, H. W. (2009). Refugia within refugia: the case study of a canopy tree (*Eurycorymbus cavaleriei*) in subtropical China. *Journal of Biogeography* , *36* (11), 2156-2164.

Wang, M. N., Duan, L., Qiao, Q., Wang, Z. F., Zimmer, E. A., Li, Z. C., & Chen, H. F. (2018). Phylogeography and conservation genetics of the rare and relict *Bretschneidera sinensis* (Akaniaceae). *PLoS One* , *13* (1), e0189034.

Wang, S. Y., Lu, H. Y., Han, J. T., Chu, G. Q., Liu, J. Q., & Negendank, J. F. W. (2012). Palaeovegetation and palaeoclimate in low-latitude southern China during the Last Glacial Maximum.*Quaternary International* , *248* , 79-85.

Xia, M., Han, X., He, H., Yu, R. B., Zhen, G., Jia, X. P., Cheng, B. J., & Deng, X. W. (2018). Improved de novo genome assembly and analysis of the Chinese cucurbit *Siraitia grosvenorii* , also known as monk fruit or luo-han-guo. *GigaScience* , *7* (6), giy067.

Yan, X., Rivero-Huguet, M. E., Hughes, B. H., & Marshall, W. D. (2008). Isolation of the sweet components from *Siraitia grosvenorii* .*Food Chemistry* , *107* (3), 1022-1028.

Yang, M. Q. (2004). Germplasm resources and production advice of Luohanguo in Longsheng County. *Southern Horticulture* , *5* , 19-20.

Zhang, Z. Y., Wu, R., Wang, Q., Zhang, Z. R., López-Pujol, J., Fan, D. M., & Li, D. Z. (2013). Comparative phylogeography of two sympatric beeches in subtropical China: Species-specific geographic mosaic of lineages. *Ecology and Evolution* , *3* (13), 4461-4472.

Zhou, J. Y., & Tang, S. Q. (2006). Genetic diversity of cultivated Luohanguo (*Siraitia grosvenorii*) revealed by RAPD markers.*Molecular Plant Breeding* , *1* , 71-78.

Zhou, J. Y., Tang, S. Q., Xiang, W. S., & Bin, X. Y. (2005). Genetic diversity of cultivated Luohanguo (*Siraitia grosvenorii*) based on ISSR marker. *Guihaia* , *5* , 431-436+503.

Zhou, L. C., Zhang, B. Y., Qin, L., & Li, F. (1981). Luohanguo variety resources survey research and utilization opinions. *Guihaia* ,*3* , 29-33.

Table 1. Details of *S. grosvenorii* sampling sites and genetic variation.

Wild population/cultivar	Latitude (N)/ Longitude (E)	Altitude (masl)	Sample size	No. of haplotypes (cpDNA/ *CHS*
MES	25°52'/110deg24'	700	10	1/7/3

Posted on Authorea 14 Oct 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.166573612.23507910/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

BL	24°47'/110deg05'	550	10	1/1/3
SJ	25°40'/109deg41'	400	10	1/2/3
JT	25°26'/109deg50'	500	10	1/3/2
JX	24°06'/110deg15'	850	10	2/2/5
YD group	YD group	YD group	YD group	0/12/8
DX	25°14'/111deg47'	719	10	1/6/5
HZ	24°37'/111deg33'	600	10	1/4/3
ZQ	23°36'/112deg30'	347	10	1/3/4
RY	24°52'/113deg05'	900	10	1/8/2
SG	24°42'/113deg49'	400	10	1/7/4
JLS	24°37'/114deg29'	443	10	1/12/4
WGS	27°29'/114deg15'	428	10	1/5/1
MD group	MD group	MD group	MD group	3/30/11
PB	22°27'/109deg49'	850	10	1/4/3
YK group	YK group	YK group	YK group	1/4/3
Cultivar	–	–	21	1/10/3
Total	–	–	151	6/45/15

Note: MES, BL, SJ, JT, JX, DX, HZ, ZQ, RY, SG, JLS, WGS, and PB represent wild populations, which were divided into three genetic groups—YD (MES, BL, SJ, JT, JX), MD (DX, HZ, ZQ, RY, SG, JLS, WGS), and YK (PB)—based on the phylogenetic analysis of chlorotypes.

Table 2. Primers used in this study for DNA amplification and sequencing.

Target DNA	Primer sequence (5'–3')	Reference	Tm ()
cpDNA	cpDNA	cpDNA	cpDNA
*trn*R-*atp*A	F: AGGTTCAAATCCTATTGGACGCA	(Dane & Lang, 2004)	55
	R: TTTTGAAAGAAGCTATTCARGAAC		
*trn*H-*psb*A	F: GTTATGCATGAACGTAATGCTC	(Sang et al., 1997)	55
	R: CGCGCATGGTGGATTCACAATCC		
*trn*L-*trn*F	F: GGTTCAAGTCCCTCTATCCC	(Taberlet et al., 1991)	55
	R: ATTTGAACTGGTGACACGAG		
Nuclear genes	Nuclear genes	Nuclear genes	Nuclear genes
CHS	F: GCCACCCGTCTTATTAGCCA	This study	58
	R: TGACGCGCTGTGTGTGTGCACACACACCC		
EDL2	F: AAAGGGRCAYCTYAGTGAG	This study	55
	R: AACTCRGAYGTCTCYTCAG		

Note: cpDNA, chloroplast DNA; Tm, melting temperature of primer pair.

Table 3. Distribution of chlorotypes, based on three cpDNA regions (*trn* R-*atp* A, *trn* H-*psb* A, and*trn* L-*trn* F), in the wild populations and cultivars of*S. grosvenorii* .

Chlorotype	Wild population MES	Wild population BL	Wild population SJ	Wild population JT	Wild population JX	Wild populatio HZ
C1	10	10	10	10	3	
C2						10
C3						
C4					7	
C5						

Posted on Authorea 14 Oct 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.166573612.23507910/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Table 4. Distribution of *CHS* haplotypes in the wild populations and cultivars of *S. grosvenorii* .

CHS haplotype	Wild population MES	Wild population BL	Wild population SJ	Wild population JT	Wild population JX	Wild popu HZ
H1				2		
H2				16		
H3					20	3
H4						
H5	2	20	18			
H6				1		
H7						
H8						
H9						
H10						
H11						
H12	11		2			11
H13						
H14						
H15						5
H16						1
H17						
H18						
H19						
H20						
H21						
H22						
H23						
H24	2					
H25						
H26						
H27						
H28						
H29				1		
H30	1					
H31	2					
H32	1					
H33	1					
H34						
H35						
H36						
H37						
H38						
H39						
H40						
H41						
H42						
H43						
H44						
H45						

Posted on Authorea 14 Oct 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.166573612.23507910/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Table 5. Distribution of *EDL2* haplotypes in the wild populations and cultivars of *S. grosvenorii* .

EDL2 haplotype	Wild population MES	Wild population BL	Wild population SJ	Wild population SJ	Wild population JT	Wild pop JT
E1		9	9	10	10	
E2	1					10
E3	7	10	10	5	5	10
E4		1	1			
E5						
E6						
E7						
E8						
E9						
E10						
E11						
E12	12			5	5	
E13						
E14						
E15						

Table 6. Analysis of molecular variance (AMOVA) of *S. grosvenorii* populations, based on the cpDNA regions (*trn* R-*atp* A, *trn* H-*psb* A, and*trn* L-*trn* F) and nuclear genes (*CHS* and*EDL2*).

Grouping	Source of variation	Sequence variation (%) cpDNA	Sequence variation (%) *CHS*	Sequence variati *EDL2*
All populations	Among populations	100.00**	50.95**	37.86**
	Within population	0.00**	49.05**	62.14**
Genetic groups	Among groups	88.80**	8.45**	1.19**
	Among populations within a group	11.20**	44.15**	39.10**
	Within population	0.00**	47.40**	59.71**

Note: Asterisks indicate statistical significance (**$P< 0.001$). Sequence variation in cpDNA and nuclear genes was analyzed in all 13 wild populations of *S. grosvenorii* (MES, BL, SJ, JT, JX, DX, HZ, ZQ, RY, SG, JLS, WGS, PB) as one group and as three genetic groups (YD, MD, YK; identified by the phylogenetic analysis of chlorotypes).

Table 7. Demographic expansion probabilities computed on the combined cpDNA and nuclear gene data of *S. grosvenorii.*

DNA	Tajima's *D*	Fu and Li's *D*	Fu' *Fs*	SSD	Raggedness index
cpDNA	2.81958 ns	1.50170 ns	17.56927 ns	0.09105 **	0.14300 **
CHS	0.02792 ns	-0.42074 ns	-16.58458 **	0.00844 ns	0.01332 ns
EDL2	-0.57491 ns	0.77656 ns	-3.99761 ns	0.00970*	0.08172 **

Note: SSD, sum of squared deviations; H_{Rag} , raggedness index; *$P < 0.05$; **$P < 0.001$; ns, not significant.

Figure legends

Figure 1. Geographical distribution, median-joining network and Maximum Likelihood (ML) phylogenetic

Posted on Authorea 14 Oct 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.166573612.23507910/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

tree for 6 chlorotypes of *S. grosvenorii.* **a** Chlorotypes of 13 wild populations of*S. grosvenorii.*Each color represents a chlorotype. **b** Chlorotypes of 21 individuals of cultivated *S. grosvenorii* accessions. **c**Median-joining network of six chlorotypes resolved in *S. grosvenorii.* **d**Phylogenetic tree constructed using the Maximum Likelihood method. Numbers at nodes represent the result of the ML bootstrap (BP) analysis. Nodes without numbers correspond to supports weaker than 80% BP.

Figure 2. Geographical distribution, median-joining network for 45 *CHS* haplotypes of*S. grosvenorii.* **a***CHS*haplotypes of 13 wild populations of *S. grosvenorii.* The color of H1–H10 was based on the *CHS* haplotypes of cultivars. Haplotypes that unique to a single population were in white color. For the rest of *CHS* haplotypes, each color represents a haplotype.**b***CHS*haplotypes of 21 individuals of cultivated *S. grosvenorii*accessions.Colors denote the origin of cultivated *S. grosvenorii* as inferred by cpDNA data. **c** Median-joining network of 45 *CHS*haplotypes resolved in *S. grosvenorii.* The colors correspond to those in Fig. 2a.

Figure 3. Geographical distribution, median-joining network for 15 *CHS* haplotypes of*S. grosvenorii.* **a** *EDL2* haplotypes of 13 wild populations of *S. grosvenorii.* The color of E1–E3 was based on the *EDL2* haplotypes of cultivars. For the rest of *EDL2*haplotypes, each color represents a haplotype. **b** *EDL2*haplotypes of 21 individuals of cultivated *S. grosvenorii*accessions. Colors denote the origin of cultivated *S. grosvenorii* as inferred by cpDNA data. **c** Median-joining network of 15 *EDL2* haplotypes resolved in *S. grosvenorii.*

Figure 4. Mismatch distribution analysis of *S. grosvenorii* based on cpDNA and nuclear gene data. **a** cpDNA; **b** *CHS* ; **c***EDL2* .

Hosted file

`Figure 1.eps` available at https://authorea.com/users/514724/articles/590360-phylogeographic-analysis-of-siraitia-grosvenorii-in-subtropical-china-provides-insight-into-the-origin-of-cultivated-monk-fruit-and-conservation-of-genetic-resources

Hosted file

`Figure 2.eps` available at https://authorea.com/users/514724/articles/590360-phylogeographic-analysis-of-siraitia-grosvenorii-in-subtropical-china-provides-insight-into-the-origin-of-cultivated-monk-fruit-and-conservation-of-genetic-resources

Hosted file

`Figure 3.eps` available at https://authorea.com/users/514724/articles/590360-phylogeographic-analysis-of-siraitia-grosvenorii-in-subtropical-china-provides-insight-into-the-origin-of-cultivated-monk-fruit-and-conservation-of-genetic-resources