Original Article

Genetic diversity of food-medicinal *Lycium* spp. in China: Insights from chloroplast genome

Ruyu Yao [a,d,1], Bin Wang [a,b,1], Michael Heinrich [c], Qiuling Wang [a,*], Peigen Xiao [a,*]

[a] *Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China*
[b] *Hainan Provincial Key Laboratory of Resources Conservation and Development of Southern Medicine, Hainan Branch of the Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Haikou 570311, China*
[c] *Research Group 'Pharmacognosy and Phytotherapy', UCL School of Pharmacy, University of London, London WC1N1AX, United Kingdom*
[d] *Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China*

A R T I C L E   I N F O

A B S T R A C T

Objective: Goji (fruits of *Lycium* spp.) is commonly consumed as food and medicine. The increasing market demand for goji has led to its wide cultivation and broad breeding, which might cause loss of genetic diversity. This study aims to uncover the genetic diversity of the cultivated and wild *Lycium*.
Methods: The chloroplast genome (CPG) of 34 accessions of Chinese food-medicinal *Lycium* spp., including the popular cultivars and their wild relatives, was re-sequenced and assembled, based on which the genetic diversity was evaluated.
Results: Sequence structural comparison shows that CPG is comparatively conserved within species. Phylogenetic analysis indicates that CPG is sufficient for the discrimination of *Lycium* species; combined with nuclear ribosomal internal transcribed spacer (Nr ITS) sequences, materials with mixed genetic backgrounds can be identified. Nucleotide diversity analysis reveals that the modern cultivars are probably with a common maternal parent, while the wild accessions are with higher level of genetic diversity.
Conclusion: For the first time this study reveals the intraspecies genetic diversity of *Lycium* spp. using CPG, highlighting the urgent conservation demand of wild genetic resources of *Lycium*. Our study also demonstrates that CPG provides crucial evidence for identification of *Lycium* species with mixed genetic backgrounds and highlights the importance of the wild relatives in genetic diversity conservation. This CPG-based technology will contribute to the sustainable development of medicinal plants broadly.

1. Introduction

Many medicinal plants are consumed in diets for sustaining human health. Recently, such food-medicinal plants are increasingly accepted in global herbal market since their pivotal roles in daily health management (Heinrich, Yao, & Xiao, 2022; Yao, He, & Xiao, 2023). This increased demand has led to the supply of food-medicinal plants shifting from wild collection into cultivation (Kling, 2016; Xiao, Chen, Huang, & Xiao, 2009). While phylogenetically related plant species often have similar biochemical / medicinal properties, many landraces or cultivars have been the main source of the herbal products (Hao & Liu, 2023; Jiao et al., 2018). Together with other wild relatives, these landraces and cultivars

sustain the supply chain, as well, they contribute abundant genetic diversity which is important for the crop improvement and biodiversity conservation (Brozynska, Furtado, & Henry, 2016). On the other hand, blurry taxonomic relationships may bring in adulteration (Srirama et al., 2017). As a result, both authentication and genetic diversity assessment are of great importance in the sustainable development of food-medicinal plants (Chen et al., 2014; Chen et al., 2018; Chen, Yu, Luo, Wu, Li, & Steinmetz, 2016; Howes, et al., 2020).

Fruits of several *Lycium* species are used as both food and medicine in the worldwide (Yao, Heinrich, & Weckerle, 2018; Yao, Heinrich, Wei, & Xiao, 2021; Chen et al., 2023). Three of the seven native Chinese species (Editorial Committee of Flora of China, 1994), *i.e.*, *Lycium barbarum* L., *L. chinense* Mill. and *L. ruthenicum* Murr. are the common food-medicinal species (Yao, Heinrich, & Weckerle, 2018; Yao, Heinrich, Zhao, Wang, Wei, & Xiao, 2021). Known as goji (Gouqi in Chinese), fruits of *L. barbarum* and *L. chi-*

nense are increasingly accepted as a "super food" globally based in part on their nutritional value and health benefits (Heinrich, Kum, & Yao, 2022), but also since it can be produced in larger quantities agriculturally. The fruit of *L. ruthenicum* is an emerging health food which is with relative potent antioxidant activities (Yao, Heinrich, & Weckerle, 2018). Practically, goji may also mean fruits of other *Lycium* species, but those are less economically important because of their limited uses (Yao, Heinrich, & Weckerle, 2018). The increasing demand pushes some landraces and novel cultivars of goji to be cultivated widely in northern China, of which a landrace refers to a domesticated, locally adapted plant variety that has developed over time through natural selection and breeding practices in a particular geographical region, and a novel cultivar is a new variety of a plant that has been developed through selective breeding or genetic engineering (Yao, Heinrich, Wang, & Weckerle, 2018; Yao, Heinrich, Wei, & Xiao, 2021; Yao et al., 2018; Li & Liu, 2023). Therefore, the current commercial goji is sourced from diverse genotypes, and key questions have emerged which form a basis for the sustainable development of the goji industry. The blurry genetic backgrounds of several varieties may lead to the inappropriate use of species, such as *L. chinense* var. *potaninii* (Pojark) A. M. Lu (Editorial Committee of Flora of China, 1994; Yang, He, Wang, Zhang, Wang, & Liang, 2020). As the predominant source of goji, *L. barbarum* has been the focus of breeding since the 1960s (Yao, Heinrich, Wang, & Weckerle, 2018), however, the genetic diversity within cultivars of the food-medicinal *Lycium* is poorly known, which is crucial since the widespread cultivation may induce the loss of genetic diversity.

With a relatively small molecular size (120−160 kb) and moderate rates of nucleotide substitution, plant chloroplast genome (CPG) has been considered as a useful tool for studies on species identification, phylogeny and genetic diversity (Daniell, Jin, Zhu, Gitzendanner, Soltis, & Soltis, 2021). In recent years, chloroplast genomes (CPGs) were applied to investigate the genetic diversity of crops. For example, a study on melon (*Cucumis melo* L.) has demonstrated that the whole CPG is useful in population genetics, also in clarifying the intra-species relationship of plants (Cui et al., 2021). As of the genus *Lycium*, the CPGs of four species have been reported, *viz. L. barbarum*, *L. chinense*, *L. ruthenicum* and *L. ferocissimum* Miers, resulting in the phylogeny at species level (Cui et al., 2019; He, Tian, Yang, & Shi, 2020; Jia et al., 2018; Li, Zhang, Zhang, & Yisilam, 2020; Wang, Jia, Zhu, & Tian, 2019; Yang, Huang, An, Zheng, Huang, & Liang, 2019; Yisilam, Mamut, Li, Li, & Fu, 2018).

In this study, we re-sequenced and analyzed the CPG of 34 accessions including cultivars and wild relatives of the three Chinese medicinal *Lycium* species. The specific aims were: (1) to clarify the genetic background of the poorly understood varieties as well as key cultivars; (2) to evaluate the genetic diversity at the level of species.

2. Materials and methods

2.1. Plant materials

Young leaves and voucher specimens of the medicinal *Lycium* spp. were collected during fieldworks in several regions of China in 2018 and 2019. Specimens were identified morphologically referring to *Flora of China* (Editorial Committee of Flora of China, 1994), which were further confirmed by Prof. Anmin Lu from the Institute of Botany, Chinese Academy of Sciences, China. Voucher specimens were deposited at the Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, China. The young leaves (ca. 1.0 g of fresh weight) were sampled from the same individual and silica gel-dried. The geographic origins, sample IDs and vou-

cher IDs of the specimens and leaf samples were presented in Table S1.

2.2. DNA extraction and sequencing, CPG assembly and annotation

The silica gel-dried leaf samples were sent to Novogene (Beijing) for library construction and sequencing using Illumina sequencer as is described in Niu et al. (2018). On average, 8 248.72 M of clean data was obtained for each sample, and details of the sequencing and quality control information were shown in Table S2.

GetOrganelle (Jin et al., 2020) was employed for CPG assembly. Default SPAdes kmers of 21, 55, 85, 115 were used and the superior result was selected; and then, target reads were selected using the default database of GetOrganelle, and CPGs of the samples were assembled referring to the published CPGs of *Lcyium* (NCBI accession numbers: MG729823, MG729824 and MG729825) (Cui et al., 2019). All the CPGs retrieved from NCBI for use in the current study had been evaluated for credibility based on the quality of publications as well as the reliability of the materials. The circular CPGs as ".fasta" files were then uploaded to CPGAVAS2 for annotation; in the meantime, the published CPGs of *Lycium* were retrieved from NCBI, and annotated using CPGAVAS2 as well (Shi et al., 2019). Accordingly, annotated files of diverse formats were generated. Table files as ".tbl" were generated using GB2sequin (Lehwark & Greiner, 2019) and the assembled CPGs were then uploaded to NCBI.

The nuclear ribosomal internal transcribed spacer (Nr ITS) sequences were extracted as follows: Sequences of Nr RNA were extracted using GetOrganelle (Jin et al., 2020) firstly, and then, ITS sequences were extracted from Nr RNA using the software ITSx (Bengtsson-Palme, et al., 2013).

2.3. Comparative analysis on structure of CPGs

The GenBank files generated from the annotation were then uploaded to CPGview-RSG (https://47.90.241.85:16100/cpgviewer/drawmap) to get the CPG gene maps. The length of each sub-region was detected by using IRscope (Amiryousefi, Hyvonen, & Poczai, 2018), as well, the junctions between them were visualized to show the IR expansion and contraction. The GC content of each sub-region was calculated using the package "seqinr" in R (Charif & Lobry, 2007; R Core Team, 2022); GC contents of a sliding window of 2 000 bp were calculated by species with "seqinr" (Charif & Lobry, 2007). The length of sub-regions as well as their GC contents were compared using Excel by F-test and *t*-test. REPuter was used to detect the dispersed repeats (forward, reverse, complement and palindromic) (Kurtz, Choudhuri, Ohlebusch, Schleiermacher, Stoye, & Giegerich, 2001).

2.4. Phylogenetic analysis based on CPGs and ITS

The newly assembled 34 CPGs as well as the 9 CPGs retrieved from NCBI were applied for phylogenetic analysis. Data were imported into MAFFT online server (Katoh, Rozewicki, & Yamada, 2019) for alignment with default parameters. Gblocks-0.91b was used to remove any ambiguous alignment regions (Castresana, 2000). Phylogenetic tree was inferred using maximum likelihood (ML) methods (Edler et al., 2021). An outgroup is not used since the present work is to study the phylogenetic relationship among the three species only.

2.5. Population genetics and nucleotide diversity (Pi) analysis

DnaSP (version. 6.12.03) (Rozas, et al., 2017) was used to calculate haplotype (h) and haplotype diversity (Hd), and PopArt (ver-

sion. 1.7) was used to construct a haplotype network graph based on TCS network (Leigh, Bryant, & Nakagawa, 2015).

As an indicator of genetic diversity, the nucleotide diversity (Nei, 1987) was calculated to measure the DNA polymorphism of the CPGs of the *Lycium* accessions. Firstly, the grouped CPGs were imported into PhyloSuite (Zhang et al., 2020); secondly, different functional sequences in CPGs were extracted using the function "extract" with default parameters; thirdly, the extracted coding sequences (CDSs) and intergenic regions were aligned using the module "MAFFT"; fourthly, the alignments were then loaded into DnaSP (Rozas et al., 2017) and the Pi values were then calculated using the function "DNA Polymorphic"; finally, the different Pi values were then complied manually and compared within Excel software.

3. Results

3.1. Structural comparison of CPGs

Thirty-four complete CPGs of *Lycium* spp. were obtained and assembled into circle DNA chains. These CPGs were then uploaded into NCBI, and their Accession IDs and structure details were shown in Table S1. Fig. 1 showed the gene map of the CPGs of these species.

The length of these CPGs ranges from 154 869 bp to 155 756 bp. With an average length of 155 735 bp, *L. chinense* has the longest CPG, followed by *L. barbarum* and *L. ruthenicum*. With a typical quadripartite structure, these CPGs are composed of four regions: one large single copy (LSC), one small single copy (SSC) and two inverted repeats (IR). As is shown in Fig. 1, a total of 131 genes were annotated in the CPGs of all samples, including 86 CDSs, 8 rRNAs and 37 tRNAs. Comparisons on the length as well as the GC content of the different regions among species are presented in Table 1 and Table 2. The length of SCC region is relatively stable within species, and the average length of the three species are 18 201, 18 207 and 18 208 bp; with regards to the IR region, the average length ranges from 25 396 bp to 25 460 bp among species; the length of LSC region differs greatly among species, ranging from 85 968 bp to 86 607 bp. *L. ruthenicum* has the shortest total CPG, LSC and IR, while *L. chinense* has the longest lengths of total CPG, LSC and IR. Moreover, SSC region has the highest GC content (40.83%−40.92%), followed by IR and LSC, and the average GC content of all species is 37.84%−37.91%. The high GC content of SSC region may be essential for its structural stability.

A sliding window analysis was performed to detect the GC content distribution along CPGs of the three species. There are two peaks at positions of 100 000−114 000 bp and 132 000−142 000 bp (Fig. 2), which correspond to the IR regions. With a relatively high GC content, IR regions are most conserved sections in CPGs. The GC content distribution patterns of *L. barbarum* and *L. chinense*, are highly similar with a Euclidean distance of 0.0268, while their Euclidean distances with *L. ruthenicum* are 0.1408 and 0.1557.

Overall, the IR junctions of *Lycium* CPGs (Fig. 3) do not differ significantly, but prone to be similar within closely related taxa, such as JLB (LSC/IRb) and JSA (SSC/IRa) in our case. However, JSB (IRb/SSC) and JLA (IRa/LSC) are only conserved within *L. barbarum*. For example, these is a 13 bp's overlap of trnH at JLA. Notably, the structural features of *L. chinense* var. *potaninii* (Y38) are similar to that of *L. barbarum* as featured by the 13 bp's overlap of trnH at JLA, rather than *L. chinense*, indicating its closer relationship with *L. barbarum*; meanwhile, the junction structure of Y35 is similar to that of *L. ruthenicum* especially at JSA and JLA, rather than *L. barbarum*. These features indicate that the maternal parents of the varieties are *L. barbarum* and *L. ruthenicum*, respectively.

Different types of repeat structure were detected in the CPG of these species (Fig. 4). It is worth noting that *L. barbarum* has fewer complement repeats than the others, and only three samples (Y20, Y33, Y34) have one complement repeats while other samples do not have; in the meanwhile, its forward repeats are relative higher. Differently, samples of *L. chinense* are with more complement repeats while those of *L. ruthenicum* are at an intermediate level. Moreover, compared with other two species, samples of *L. chinense* contain more reverse repeats averagely, and some of them are with less palindromic repeats.

3.2. Phylogeny and intraspecies genetic divergence

A phylogenetic tree (Fig. 5) was constructed using the aligned CPGs of the 43 *Lycium* accessions, including the 34 newly sequenced and the 9 sequences retrieved from NCBI. The samples are categorized into three clusters, including one cluster named as "barbarum", which consists of *L. barbarum* and *L. chinense* var. *potaninii*; one cluster named as "chinense", which consists of *L. chinense* and one cluster named as "ruthenicum", which mainly consists of *L. ruthenicum* individuals. The two sub-clades of "barbarum" are labeled as "b_I" and "b_II", respectively; the three sub-clades of "chinense" are labeled as "ch_I", "ch_II" and "ch_III", respectively; and the three sub-clades of "ruthenicum" are labeled as "ru_I", "ru_II" and "ru_III", respectively.

Most interestingly, as is shown in Fig. 5, the variety *L. chinense* var. *potaninii* (LY21CY38) is not clustered with *L. chinense*, but is clustered with *L. barbarum*, indicating its close maternal relationship with *L. barbarum*. The phylogeny tree based on ITS (Fig. 6) indicates that Y38 has a close relationship with *L. chinense*. These together suggest that *L. chinense* var. *potaninii* (LY21CY38) is probably with a mixed genetic background, sourced from *L. barbarum* and *L. chinense*. However, this speculation will need to be confirmed with more evidence.

The sample Y35 with yellow fruits and narrow leaves could not be identified according to the *Flora of China*. The phylogenetic analysis shows that its CPG is clustered with *L. ruthenicum* (Fig. 5), while the ITS-based phylogeny indicates its close relationship with *L. barbarum* (Fig. 6). Morphologically, it is in-between of these two species. Taking the evidence both from molecular research and morphology, the genetic background of Y35 is probably a mixture of *L. ruthenicum* and *L. barbarum*.

Within the cluster "barbarum", clade "b_I" includes four samples of the non-cultivated or old forms from Ningxia, China. Specifically, Y20 is a procumbent wild variant, Y33 and Y34 are wild materials grow by rivers, and Y32 is a tree of over 110-year-old. The other clade "b_II", however, are modern cultivars from different provinces. Of these, Y30 and Y31 are the current widely planted cultivars "Mengqi" and "Ningqi No. 10". Y10 and Y39 are the Ningxia-oriented breeds which were introduced into Heilongjiang and Hebei in 1980s. Thus, materials of the clade "b_II" are modern cultivars, while "b_I" are the old / non-cultivated forms.

The cluster "chinense" consists of three clades. Clade "ch_I" is composed of materials from southwest of China except Y41, which is a material found in a botanical garden. Clade "ch_II" includes samples of diverse origins, including Guangdong, Guizhou, Hebei, Beijing and Heilongjiang, China. Clade "ch_III" comprises four samples from Yunnan Province, two samples from Beijing, and two samples from Shanxi and Ningxia, China, respectively.

The cluster "ruthenicum" includes three clades, but the distribution of samples does not show geographic patterns. Clade "ru_I" includes one sample from Xinjiang (Y27) and one sample from Ningxia (Y36); "ru_II" includes samples from Gansu, Ningxia and Qinghai; two samples of Ningxia and one sample from Qinghai; "ru_III" has samples of Qinghai, Xinjiang and Ningxia.
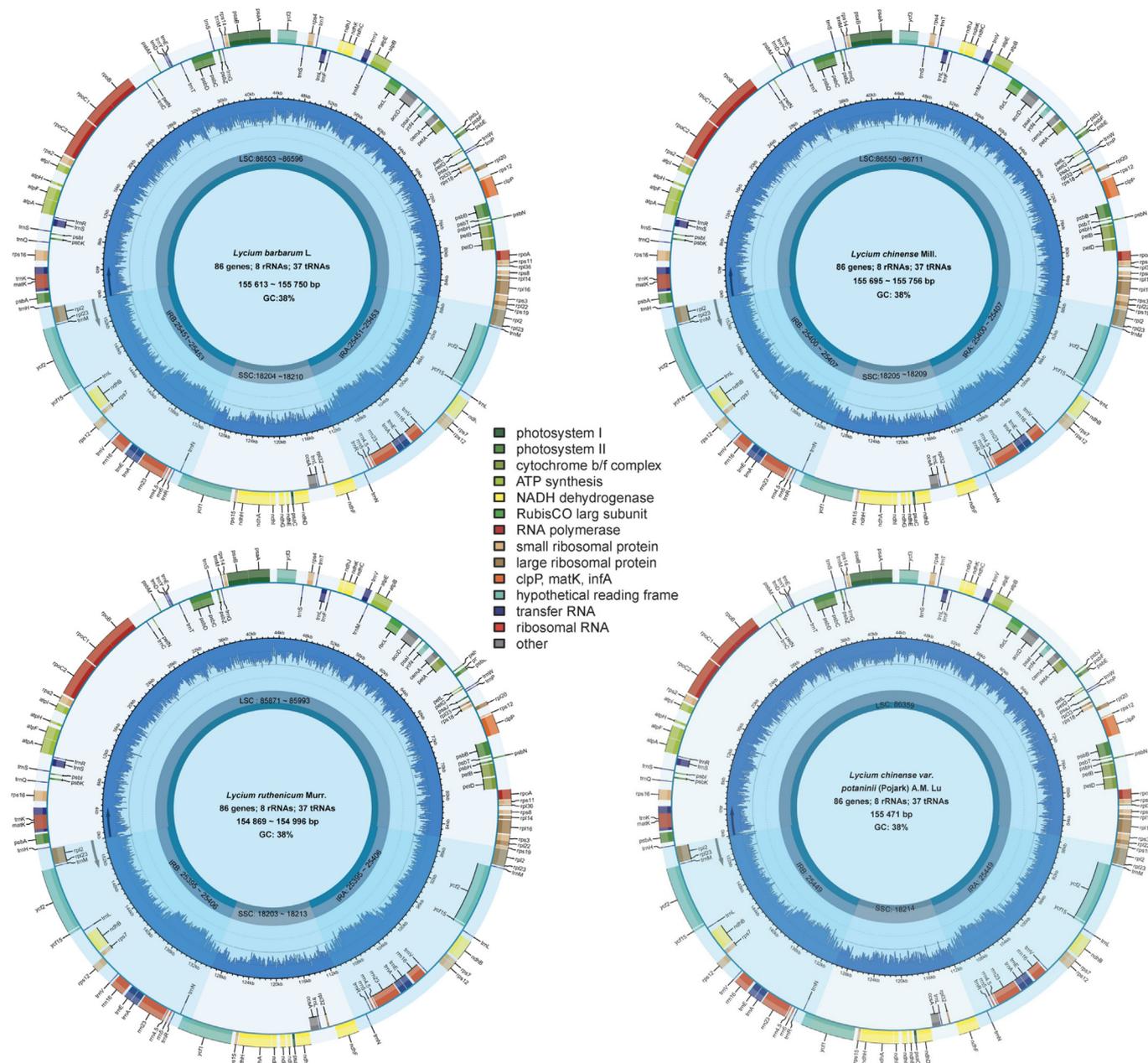
Fig. 1. CPG gene map of *L. barbarum*, *L. chinense*, *L. ruthenicum*, and *L. chinense* var. *potaninii* (Y38).

Table 1
Comparison of length of sub-regions of CPGs by species.

Species	Total (bp)	LSC (bp)	SSC (bp)	IR (bp)
L. barbarum	155 651 ± 22 B	86 548 ± 23 B	18 201 ± 4 B	25 451 ± 1 A
L. chinense	155 735 ± 17 A	86 607 ± 40 A	18 208 ± 2 A	25 460 ± 23 A
L. ruthenicum	154 968 ± 38 C	85 968 ± 38 C	18 207 ± 4 A	25 396 ± 4 B

Note: values are average ± SD; different capital letters showing significant differences ($P < 0.01$) by t-test.

Table 2
Comparison of GC content of sub-regions of CPGs by species.

Species	Total (%)	LSC (%)	SSC (%)	IRa/IRb (%)
L. barbarum	37.83 C	35.86 C	40.87 B	38.64 A
L. chinense	37.84 B	35.88 B	40.83 C	38.63 B
L. ruthenicum	37.91 A	35.97 A	40.92 A	38.65 A

Note: values are average ± SD; different capital letters showing significant differences ($P < 0.01$) by t-test.
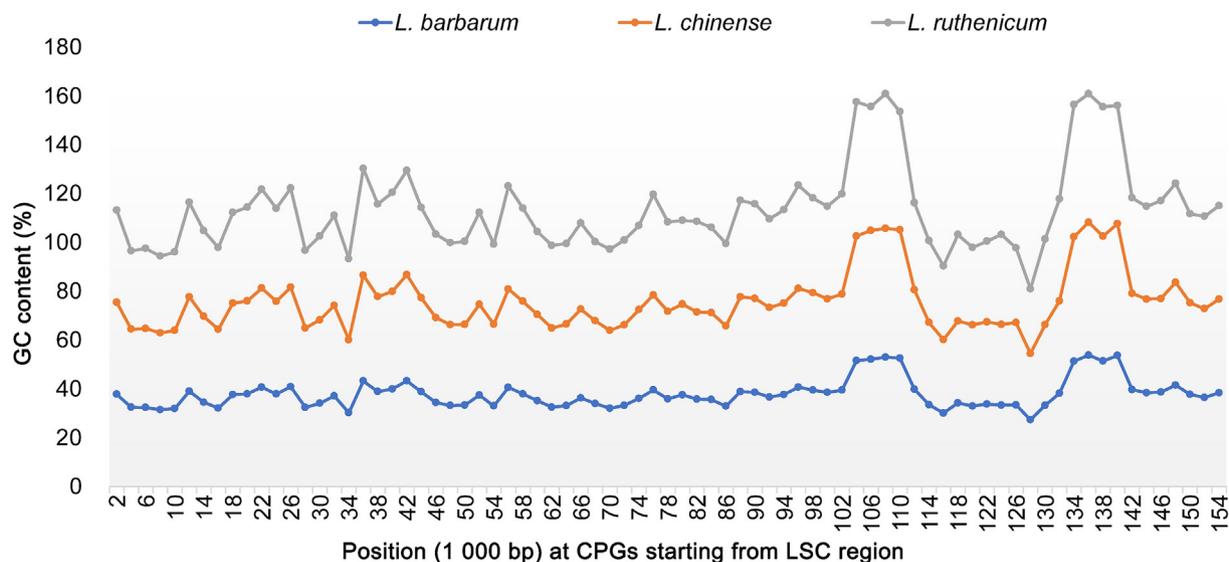
Fig. 2. Sliding window analysis of average GC content along CPGs of Chinese medicinal *Lycium*.

Twenty-seven haplotypes (Hap 1−Hap 27) were identified from the *Lycium* accessions as indicated by the haplotype network based on cpDNA at species level (Fig. 7). Haplotypes are found to be abundant within species of *L. barbarum*, *L. chinense* and *L. ruthenicum*. It can be seen that the South Africa sourced accession Hap 27 is the branch of *L. chinense*. Moreover, those of Y35 and Y18 are the branch of *L. ruthenicum*, and Y38 (Hap 17) is located in the branch of *L. barbarum*, suggesting the mixed genetic backgrounds of these accessions.

3.3. Nucleotide diversity of medicinal Lycium species

The nucleotide diversity was calculated to measure the DNA polymorphism of the CPGs of the *Lycium* accessions. Table 3 shows the Pi value of the CPGs within the three phylogenetic clusters, as well as that within the sub-clades.

The high genetic diversity within the three phylogenetic clusters is reflected by their high Pi values. Among them, the cluster "ruthenicum" has the highest Pi of 0.000 20. While Pi of the sub-clade "b_I" is 0.000 14, which is relatively high, Pi of "b_II" is 0. Materials of the "b_II" are the diverse modern cultivars, and the low Pi indicates their low genetic diversity in CPGs. This implies that these cultivars are sourced from maternal parents with close phylogenetic relationships. While these popular cultivars are already under large-scale cultivation, this also is a risk for loss of diversity in the goji farmland. On the contrary, the sub-clade "b_I" is composed of the old cultivars or non-cultivated materials, of which the genetic diversity is relatively higher. Pi of the sub-clades of "chinense" is low in general, especially the "ch_I", which is mainly composed of wild materials from southwestern China, has a low Pi of 0.000 01. The sub-clades of "ruthenicum" have high nucleotide diversity as are proofed by their high Pi values, especially for "ru_III", which with Pi of 0.000 70.

The nucleotide diversity within *L. barbarum* is 0.00015, but that of its two sub-clades indicates that all the diversity comes from the old cultivars and non-cultivated materials. *L. chinense* from southwestern China also has a low level of nucleotide diversity. Notably, *L. ruthenicum* has a relatively high nucleotide diversity. *L. ruthenicum* is widely distributed in Qinghai-Tibetan Plateau, and has been transplanted in farmland since around 2015 as an emerging health food. It is believed that artificial selection has rare impacts on its

genetic diversity during the short time of cultivation, and this may lead to its high level of nucleotide diversity.

The nucleotide diversity of CPGs of old cultivars or non-cultivated materials (the clade "b_I") and the modern cultivars (the clade "b_II") of *L. barbarum* was compared, and the different Pi values of their intergenic region and CDS are shown in Fig. 8.

Interestingly, Pi values of sequences in the intergenic region of "b_II" are all 0, indicating their conserved structure as well as low level of nucleotide diversity; conversely, that of the clade "b_I" has a relatively high nucleotide diversity with a mean Pi of 0.000 49, and high Pi are observed in several sequences (Fig. 8A). In terms of the CDS, the two clades manifest variations in different genes, except for "ycf1" whose Pi is not zero in both clades (Fig. 8B). Importantly, the modern cultivars have a relatively low nucleotide diversity with a mean Pi of 0.000 09, while that of its counterpart is 0.002 12. Overall, the old cultivars and non-cultivated materials have a higher Pi, and the modern cultivars perform diversity in some sequences of CDS only.

4. Discussion

4.1. CPGs reveal intraspecies genetic diversity of Chinese medicinal Lycium spp.

Former studies on CPGs of *Lycium* have focused on the structural feature of the species as well as the interspecies difference (Cui et al., 2019; He, Tian, Yang, & Shi, 2020; Jia et al., 2018; Wang, Jia, Zhu, & Tian, 2019; Yang, Huang, An, Zheng, Huang, & Liang, 2019; Yisilam, Mamut, Li, Li, & Fu, 2018). The current work uses CPGs to elaborate the intraspecies genetic varieties creatively. Genetic diversity is a primary indicator for biodiversity conservation and crop molecular breeding (Castaneda-Alvarez et al., 2016; Roa, Hamilton, Wenzl, & Powell, 2016). Goji is an important economic crop, and a variety of genetic markers have been applied to study the genetic diversity of *Lycium*. Our study has provided crucial adjuvant information for the genetic diversity of *Lycium* based on the nucleotide diversity of their CPGs.

The genetic diversity of domesticated crops and their wild relatives provides a useful genetic resource for breeding and is of crucial ecological importance in the global ecosystem. Crop cultivation and domestication may lead to purifying selection, which have

Fig. 3. Comparison of junctions between LSC, IRb, SSC and IRa of selected CPGs.

been found in CPGs, e.g., ginseng and tea (Srirama et al., 2017; Peng et al., 2021; Zhao et al., 2014). Notably, our study shows that the modern cultivars of *L. barbarum* have a low level of genetic diversity as indicated by their CPGs, suggesting that the widely cultivated material may be bred from a single maternal parent. The wide cultivation of *Lycium* dated back to 1960s in China, and the

push for large-scale breeding began in the end of 1980s (Yao et al., 2018). Seedlings of *L. barbarum* were introduced into various regions from Ningxia, and the natural mutations were utilized in the new lines or cultivars; therefore, the cultivated *Lycium* has a narrow gene pool. This may lead to genetic swamping because of the large volume of agricultural cultivation compared with the
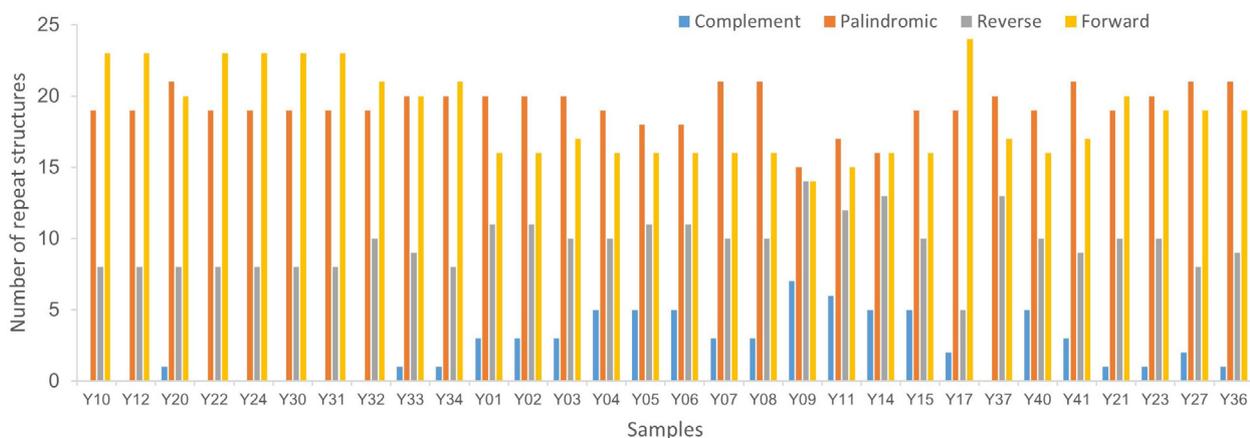
Fig. 4. Number of different types of repeat structure. Note: from left to right, samples Y10 to Y34 are *L.barbarum*; samples Y01-Y41 are *L. chinense*; samples Y21-Y36 are *L. ruthenicum*.
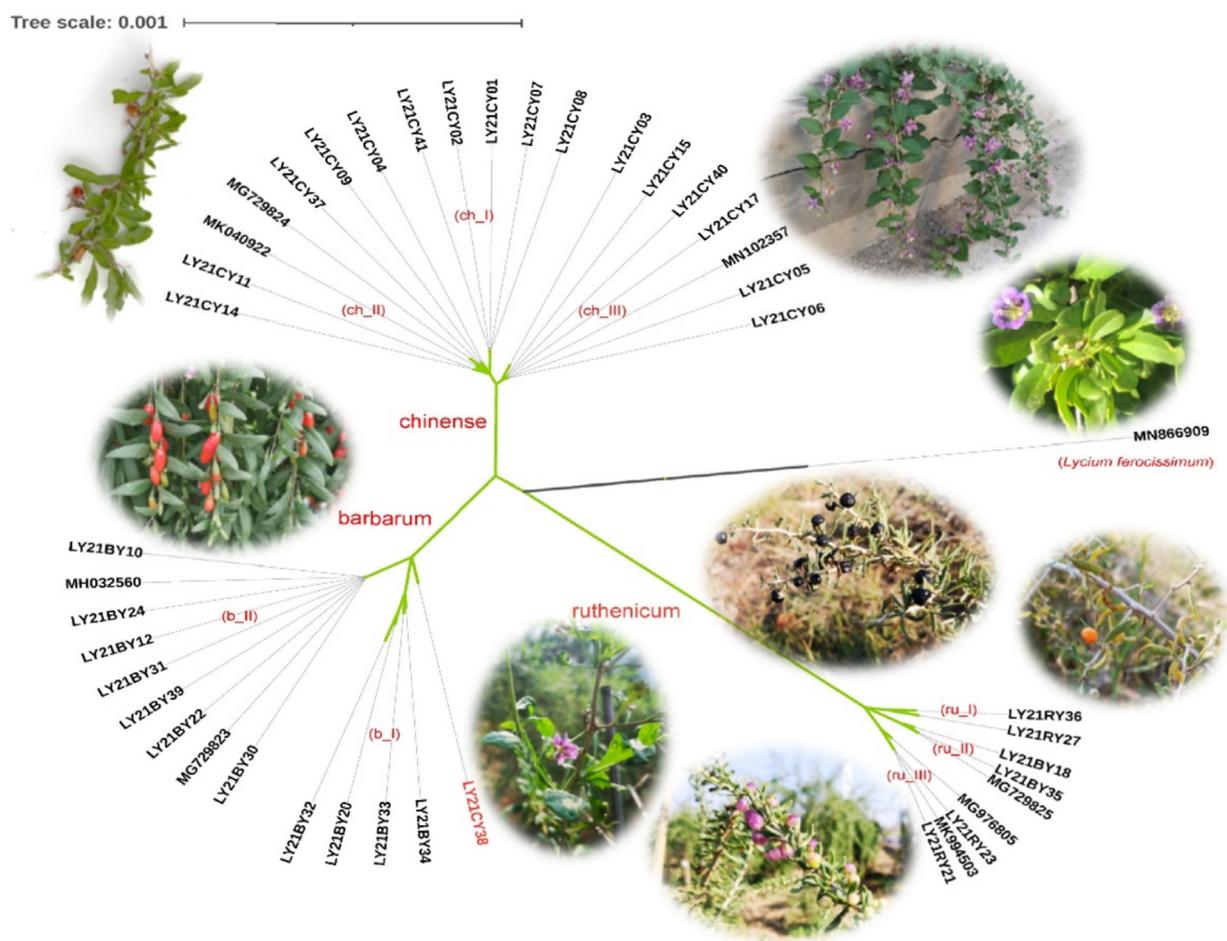

Fig. 5. An unrooted maximum likelihood tree of *Lycium* accessions based on CPGs.

non-cultivated materials, and will potentially result in genetic diversity loss and might trigger agricultural or ecological disaster (Castaneda-Alvarez et al., 2016). Therefore, we strongly claim to emphasize the application and conservation of the wild genetic resources in the future cultivation and breeding.

Studies on sweet potato and tea have also indicated that the cultivated populations were poor in genetic diversity (Peng et al., 2021; Xiao et al., 2021; Yao et al., 2018). Interestingly, cultivated *L. ruthenicum* has a relatively higher level of genetic diversity. As an emerging crop in recent years, the cultivation of *L. ruthenicum*
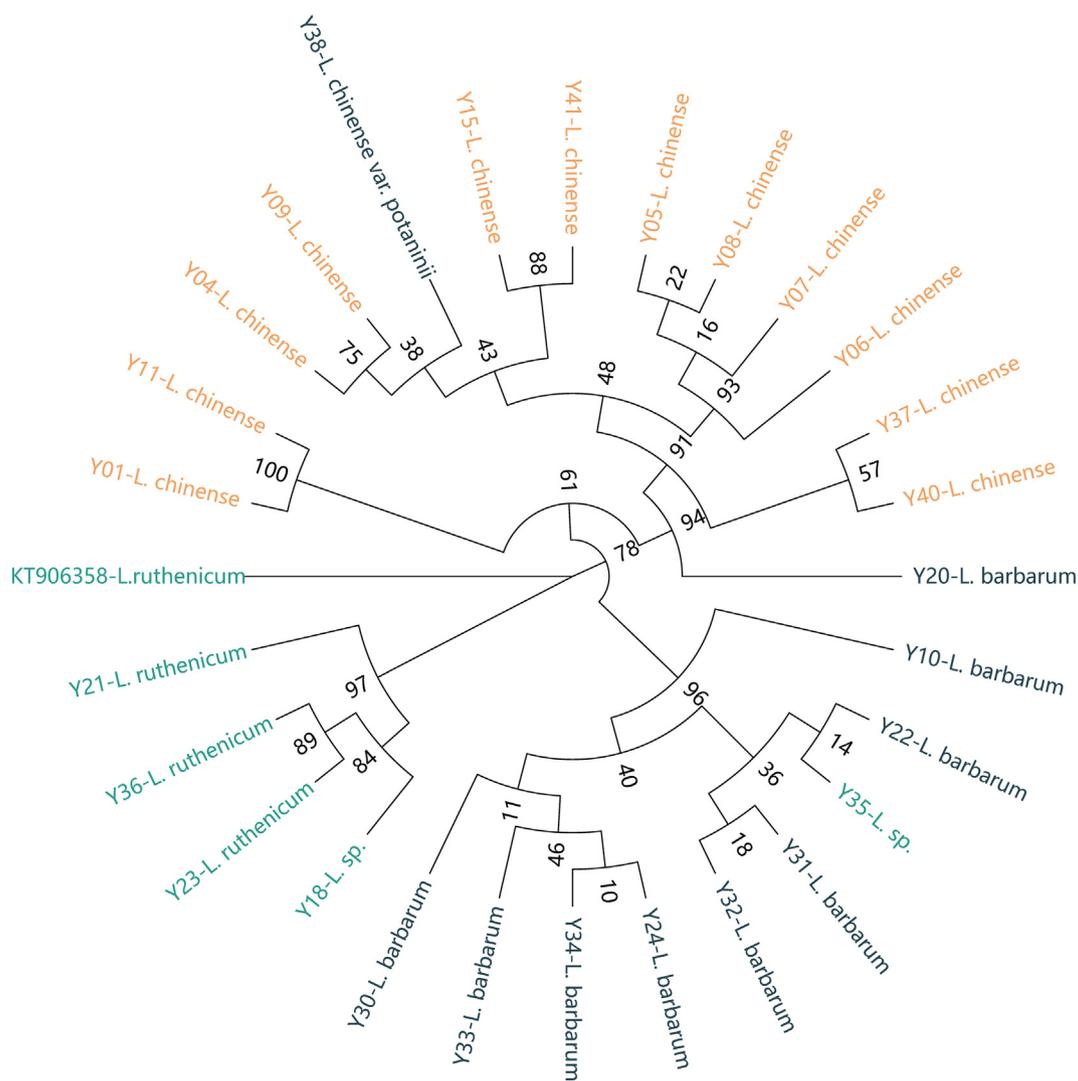
Fig. 6. Maximum likelihood tree based on ITS sequences of selected samples.
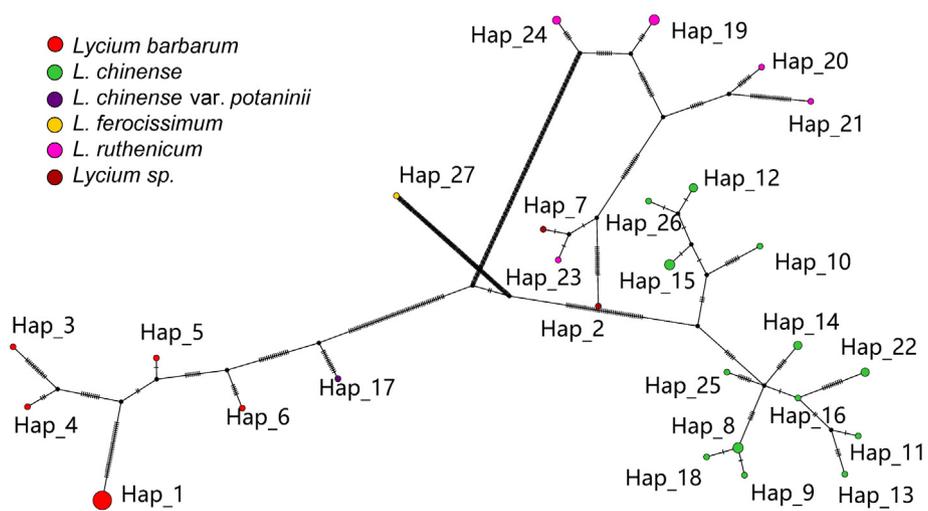

Fig. 7. Haplotype network of *Lycium* accessions based on cpDNA.

Table 3
Nucleotide diversity (Pi) of CPGs within three *Lycium* clusters.

Clusters	Number of CPGs	Number of segregating sites	Number of haplotypes	Haplotype diversity	Nucleotide diversity
Barbarum	15	79	6	0.571 43	0.000 15
Chinense	20	62	14	0.957 89	0.000 09
Ruthenicum	10	84	7	0.911 11	0.000 20
b_I	4	36	4	1.000 00	0.000 14
b_II	10	0	1	0.000 00	0.000 00
ch_I	5	2	3	0.700 00	0.000 01
ch_II	8	33	7	0.964 29	0.000 08
ch_III	7	13	4	0.809 52	0.000 03
ru_I	2	29	2	1.000 00	0.000 19
ru_II	3	18	3	1.000 00	0.000 08
ru_III	5	17	2	0.600 00	0.000 70

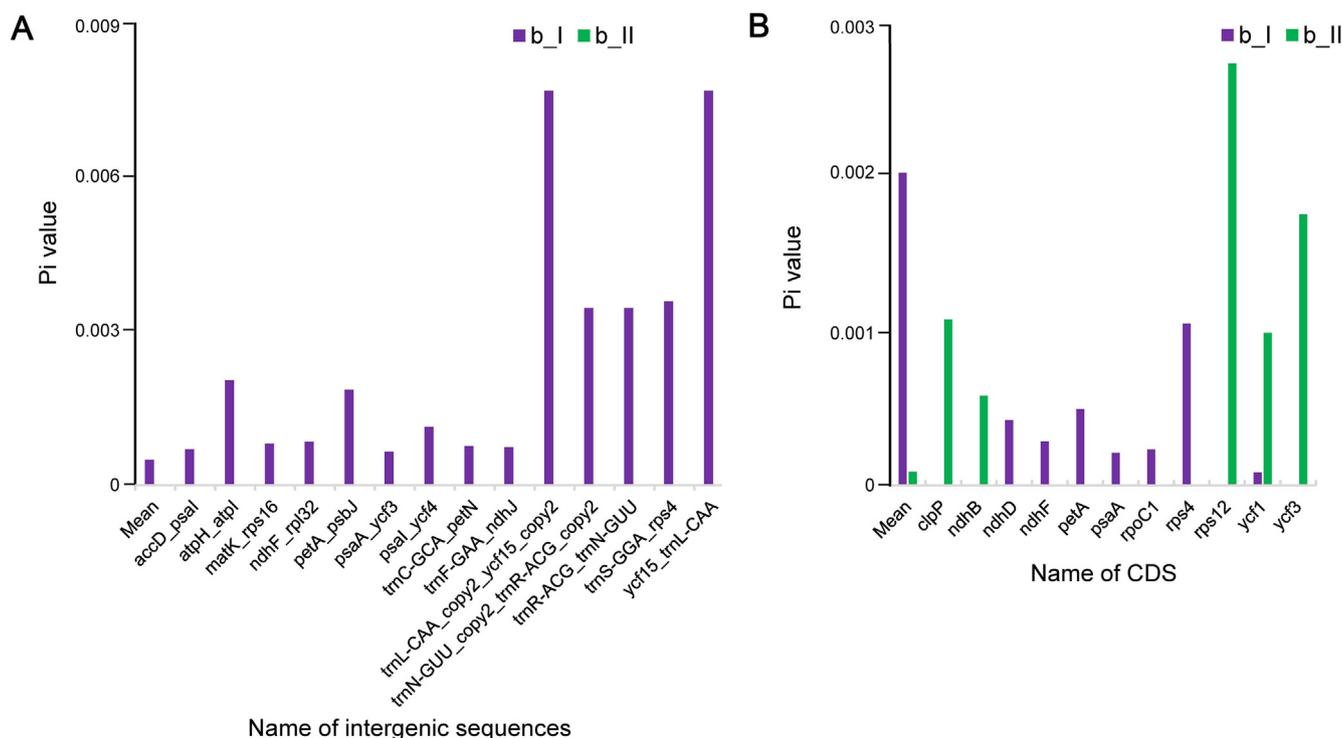Note: as a variety, Y38 is not included in sub-clade.

Fig. 8. Comparison of nucleotide diversity in intergenic region (A) and CDS (B) between CPGs of modern cultivars (b_II) and old cultivars or non-cultivated materials (b_I) of *L. barbarum*.

has a short history, and the cultivated materials are mainly *trans*-planted from the wild.

4.2. CPG of Lycium is an effective tool for genetic background identification

The phylogenetic tree based on CPGs of *Lycium* discriminates the three medicinal species effectively, indicating CPG's function as super DNA barcoding. The morphologic traits of *Lycium*, especially for the diverse cultivars, vary greatly among different environments (Yao, Heinrich, Zhao, Wang, Wei, & Xiao, 2021; Yao et al., 2018). Consequently, it is risky to identify the species relying on morphologic traits solely. For example, in the present work, some materials of the "ruthenicum" cluster are with morphologic traits of *L. barbarum*. Therefore, such a reliable super-DNA barcoding provides an important complement for species identification. CPG is superior to the traditional short sequence-based DNA markers since the Insertions and Deletions (InDels) included are also operative (Xiao et al., 2021).

Moreover, hybridization within the genus *Lycium* may make the identification even more difficult (Wu, Wei, Yang, & Zhang, 2011; Yang, He, Wang, Zhang, Wang, & Liang, 2020). As CPG of *Lycium* is matrilineal inheritance, CPG-based barcoding is able to deduce maternal parent; which combined with nuclear genes, like ITS as we used in this study, could be used for the identification of interspecific hybrids. *L. chinense* var. *potaninii* was supposed to be a hybrid from *L. chinense* and *L. barbarum* by the evidence of population genetics (Yang, He, Wang, Zhang, Wang, & Liang, 2020), which was supported by our CPG plus ITS strategy, and its parents were identified additionally. Additionally, the haplotype network analysis also suggested the mixed genetic backgrounds of these accessions

5. Conclusion

The present study for the first time reveals the intraspecies genetic diversity of *Lycium* spp. using CPG, and demonstrates that CPG is an efficient tool for the identification of *Lycium* accessions

with mixed genetic backgrounds. The CPG- and ITS-based phylogenetic analyses demonstrate that the variety *L. chinense* var. *potaninii* (Y38) and the sample Y35 have mixed genetic backgrounds. The genetic diversity loss risk of the cultivated *L. barbarum* is implied by their low diversity within CPG, highlighting the urgent conservation demand of wild genetic resources of *Lycium*.

The increasing demand for food-medicinal plants requires a sustainable supply system. As the most important healthcare materials, the sustainable development of food-medicinal plants is in the spotlight. However, the wild to cultivation herbs are facing with problems such as blurry biological origins and loss of genetic diversity. Our study demonstrates that CPG is a useful tool for the species identification and genetic diversity evaluation, and is expected to solve such general problems in the production of food-medicinal plants. Hopefully, the current technology will contribute to the sustainable development of medicinal plants more generally.

CRediT authorship contribution statement

Ruyu Yao: Conceptualization, Methodology, Validation, Investigation, Resources, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition. **Bin Wang:** Methodology, Validation, Investigation, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition. **Michael Heinrich:** Conceptualization, Supervision, Writing – review & editing. **Qiuling Wang:** Methodology, Resources, Data curation, Writing – review & editing, Project administration, Funding acquisition. **Peigen Xiao:** Conceptualization, Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

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

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chmed.2024.02.003.

References

Amiryousefi, A., Hyvonen, J., & Poczai, P. (2018). IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformation, 34*, 3030–3031.

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., ... Bunce, M. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology and Evolution, 4*, 914–919.

Brozynska, M., Furtado, A., & Henry, R. J. (2016). Genomics of crop wild relatives: Expanding the gene pool for crop improvement. *Plant Biotechnology Journal, 14*, 1070–1085.

Castaneda-Alvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., ... Toll, J. (2016). Global conservation priorities for crop wild relatives. *Nature Plants, 2*, 16022.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution, 17*, 540–552.

Charif, D., & Lobry, J. (2007). SeqinR 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In U. Bastolla, M. Porto, H. Roman, & M. Vendruscolo (Eds.), *Structural approaches to sequence evolution: Molecules, networks, populations, series Biological and Medical Physics, Biomedical Engineering* (pp. 207–232). New York: Springer Verlag.

Chen, C., Zheng, L., Ma, Q., Zhou, W. B., Lu, Y., Zhao, Y. P., & Fu, C. X. (2018). Impacts of domestication on population genetics of a traditional Chinese medicinal herb, *Atractylodes macrocephala* (Asteraceae). *Journal of Systematics and Evolution, 57*, 222–233.

Chen, D. J., Guo, S., Yi, Y. L., Zhang, F., Duan, R., Xu, P., & Duan, J. A. (2023). Research progress on amides from *Lycium* and their biological activities. *Chinese Traditional and Herbal Drugs, 54*(1), 317–333.

Chen, S., Pang, X., Song, J., Shi, L., Yao, H., Han, J., & Leon, C. (2014). A renaissance in herbal medicine identification: From morphology to DNA. *Biotechnology Advances, 32*, 1237–1244.

Chen, S., Yu, H., Luo, H., Wu, Q., Li, C., & Steinmetz, A. (2016). Conservation and sustainable use of medicinal plants: Problems, progress, and prospects. *Chinese Medicine, 11*, 37.

Cui, H., Ding, Z., Zhu, Q., Wu, Y., Qiu, B., & Gao, P. (2021). Population structure, genetic diversity and fruit-related traits of wild and cultivated melons based on chloroplast genome. *Genetic Resources and Crop Evolution, 68*(3), 1011–1021.

Cui, Y., Zhou, J., Chen, X., Xu, Z., Wang, Y., Sun, W., ... Yao, H. (2019). Complete chloroplast genome and comparative analysis of three *Lycium* (Solanaceae) species with medicinal and edible properties. *Gene Reports, 17*, 100464.

Daniell, H., Jin, S., Zhu, X. G., Gitzendanner, M. A., Soltis, D. E., & Soltis, P. S. (2021). Green giant-a tiny chloroplast genome with mighty power to produce high-value proteins: History and phylogeny. *Plant Biotechnology Journal, 19*, 430–447.

Editorial Committee of Flora of China (1994). Verbenaceae through Solanaceae: *Lycium*. In Z. Y. Wu & P. H. Raven (Eds.), *Flora of China (Volume 17)* (pp. 301–304). Science Press & Missouri Botanical Garden Press.

Edler, D., Klein, J., Antonelli, A., & Silvestro, D. (2021). raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods in Ecology and Evolution, 12*(2), 373–377.

Hao, D., & Liu, C. (2023). Deepening insights into food and medicine continuum within the context of pharmacophylogeny. *Chinese Herbal Medicines, 15*, 1–2.

He, S. L., Tian, Y., Yang, Y., & Shi, C. Y. (2020). Phylogenetic analyses of *Lycium chinense* (Solanaceae) and its coordinal species applying chloroplast genome. *Mitochondrial DNA Part B, Resources, 5*, 2201–2202.

Heinrich, M., Kum, K. Y., & Yao, R. (2022). Decontextualized Chinese medicines: Their uses as health foods and medicines in the 'global North'. *Routledge Handbook of Chinese Medicine*, eds. V. Lo & M. Stanley-Baker. 1st ed: Routledge, 721–741.

Heinrich, M., Yao, R., & Xiao, P. (2022). 'Food and medicine continuum' – Why we should promote cross-cultural communication between the global East and West. *Chinese Herbal Medicines, 14*, 3–4.

Howes, M. J. R., Quave, C. L., Collemare, J., Tatsis, E. C., Twilley, D., Lulekal, E., ... Nic Lughadha, E. (2020). Molecules from nature: Reconciling biodiversity conservation and global healthcare imperatives for sustainable use of medicinal plants and fungi. *Plants, People, Planet, 2*, 463–481.

Jia, G., Xin, G., Ren, X., Du, X., Liu, H., Hao, N., ... Liu, W. (2018). Characterization of the complete chloroplast genome of *Lycium barbarum* (Solanales: Solanaceae), a unique economic plant to China. *Mitochondrial DNA Part B, Resources, 3*, 1062–1063.

Jiao, J., Bai, Z., Huang, W., Xia, P., Liu, F., Ma, C., & Liang, Z. (2018). Evaluation of genetic diversity and population structure of *Polygonati rhizoma* germplasms: Implications for better crop development and conservation of a traditional Chinese medicine. *Acta Physiologiae Plantarum, 40*, 184.

Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., Depamphilis, C. W., Yi, T. S., & Li, D. Z. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology, 21*, 241.

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics, 20*, 1160–1166.

Kling, J. (2016). Protecting medicine's wild pharmacy. *Nature Plants, 2*, 16064.

Kurtz, S., Choudhuri, J., Ohlebusch, E., Schleiermacher, C., Stoye, J., & Giegerich, R. (2001). REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research, 29*, 4633–4642.

Lehwark, P., & Greiner, S. (2019). GB2sequin - A file converter preparing custom GenBank files for database submission. *Genomics, 111*, 759–761.

Leigh, J. W., Bryant, D., & Nakagawa, S. (2015). POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution, 6*, 1110–1116.

Li, M. Z., & Liu, Z. G. (2023). Application of omics technology in *Lycium ruthenicum* resources. *Chinese Traditional and Herbal Drugs, 54*(1), 272–282.

Li, Z., Zhang, X., Zhang, Q., & Yisilam, G. (2020). Complete chloroplast genome of *Lycium ferocissimum* (Solanaceae), a species native to South Africa. *Mitochondrial DNA Part B, Resources, 5*, 756–757.

Nei, M. (1987). DNA polymorphism within and between populations. *Molecular Evolutionary Genetics*. Columbia University Press, 254–286.

Niu, Y. T., Jabbour, F., Barrett, R. L., Ye, J. F., Zhang, Z. Z., Lu, K. Q., ... Chen, Z. D. (2018). Combining complete chloroplast genome sequences with target loci data and morphology to resolve species limits in *Triplostegia* (Caprifoliaceae). *Molecular Phylogenetics and Evolution, 129*, 15–26.

Peng, J., Zhao, Y., Dong, M., Liu, S., Hu, Z., Zhong, X., & Xu, Z. (2021). Exploring the evolutionary characteristics between cultivated tea and its wild relatives using complete chloroplast genomes. *BMC Ecology and Evolution, 21*, 71.

R Core Team (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Roa, C., Hamilton, R. S., Wenzl, P., & Powell, W. (2016). Plant genetic resources: Needs, rights, and opportunities. *Trends in Plant Science, 21*, 633–636.

Rozas, J., Ferrer-Mata, A., Sanchez-Delbarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sanchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution, 34*, 3299–3302.

Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., & Liu, C. (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Research, 47*, W65–W73.

Srirama, R., Santhosh Kumar, J. U., Seethapathy, G. S., Newmaster, S. G., Ragupathy, S., Ganeshaiah, K. N., ... Ravikanth, G. (2017). Species adulteration in the herbal trade: Causes, consequences and mitigation. *Drug Safety, 40*, 651–661.

Wang, C., Jia, G., Zhu, L., & Tian, F. (2019). The complete chloroplast genome sequence of *Lycium ruthenicum* (Solanaceae), a traditional medicinal plant in China. *Mitochondrial DNA Part B, Resources, 4*, 2495–2496.

Wu, L. L., Wei, R. X., Yang, Q. W., & Zhang, Z. Y. (2011). A preliminary study on the hybrid origin of new taxa in *Lycium* (Solanaceae). *Guihaia, 31*, 304–311.

Xiao, S., Xu, P., Deng, Y., Dai, X., Zhao, L., Heider, B., ... Cao, Q. (2021). Comparative analysis of chloroplast genomes of cultivars and wild species of sweetpotato (*Ipomoea batatas* [L.] Lam). *BMC Genomics, 22*, 262.

Xiao, X. H., Chen, S. L., Huang, L. Q., & Xiao, P. G. (2009). Survey of investigations on daodi Chinese medicinal materials in China since 1980s. *China Journal of Chinese Materia Medica, 34*(5), 519–523.

Yang, L., He, Y., Wang, S., Zhang, Y., Wang, Y., & Liang, Q. (2020). Population genetic divergence in *Lycium chinense* Mill. (Solanaceae) is driven by both hybridization and geo-environmental factors. *Genetic Resources and Crop Evolution, 67*, 1505–1520.

Yang, Z., Huang, Y., An, W., Zheng, X., Huang, S., & Liang, L. (2019). Sequencing and structural analysis of the complete chloroplast genome of the medicinal plant *Lycium chinense* Mill. *Plants, 8*, 87.

Yao, R., He, C., & Xiao, P. (2023). 'Food and medicine continuum' in the East and West: Old tradition and current regulation. *Chinese Herbal Medicines, 15*, 6–14.

Yao, R., Heinrich, M., Wang, Z., & Weckerle, C. S. (2018). Quality control of goji (fruits of *Lycium barbarum* Mill): A value chain analysis perspective. *Journal of Ethnopharmacology, 224*, 349–358.

Yao, R., Heinrich, M., & Weckerle, C. S. (2018). The genus *Lycium* as food and medicine: A botanical, ethnobotanical and historical review. *Journal of Ethnopharmacology, 212*, 50–66.

Yao, R., Heinrich, M., & Xiao, P. (2021). Cross-cultural ethnobotanical assembly as a new tool for understanding medicinal and culinary values - the genus *Lycium* as a case study. *Frontiers in Pharmacology, 12*, 708518.

Yao, R., Heinrich, M., Zhao, X., Wang, Q., Wei, J., & Xiao, P. (2021). What's the choice for goji: *Lycium barbarum* L. or *L. chinense* Mill.? *Journal of Ethnopharmacology, 276*, 114185.

Yao, R., Heinrich, M., Zou, Y., Reich, E., Zhang, X., Chen, Y., & Weckerle, C. S. (2018). Quality variation of goji (fruits of *Lycium spp.*) in China: A comparative morphological and metabolomic analysis. *Frontiers in Pharmacology, 9*, 151.

Yisilam, G., Mamut, R., Li, J., Li, P., & Fu, C. (2018). Characterization of the complete chloroplast genome of *Lycium ruthenicum* (Solanaceae). *Mitochondrial DNA Part B, Resources, 3*, 361–362.

Zhang, D., Gao, F., Jakovlic, I., Zou, H., Zhang, J., Li, W. X., & Wang, G. T. (2020). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular Ecology Resources, 20*, 348–355.

Zhao, Y., Yin, J., Guo, H., Zhang, Y., Xiao, W., Sun, C., ... Xiao, J. (2014). The complete chloroplast genome provides insight into the evolution and polymorphism of *Panax ginseng. Frontiers in Plant Science, 5*, 696.