scientific reports

OPEN

Comparative and phylogenetic analysis of *Asparagus meioclados* Levl. and *Asparagus munitus* Wang et S. C. Chen plastomes and utility of plastomes mutational hotspots

Yulu Tian[1,2], Xue Liu[2✉], Yuanjiang Xu[2], Benxia Yu[2], Le Wang[3] & Xianyou Qu[2✉]

Tiandong is a vital traditional Chinese herbal medicine. It is derived from the tuber root of the *Asparagus cochinchinensis* according to the Pharmacopoeia of the people's republic of China (2020 Edition). On account of the similar morphology, *Asparagus meioclados* and *Asparagus munitus* were used as Tian-Dong in southwest China. Chloroplast (cp) genomes are highly active genetic components of plants and play an extremely important role in improving the efficiency of the identification of plant species. To differentiate the medicinal plants belonging to the genus *Asparagus*, we sequenced and analyzed the complete plastomes (plastid genomes) of *A. meioclados* and *A. munitus* and obtained two plastomes whose length changed to 156,515 bp and 156,381 bp, respectively. A total of 111 unique genes have been detected in plastome, which included 78 protein-coding genes, 29 tRNA genes and 4 rRNA genes. In plastomes of *A. meioclados* and *A. munitus*, 14,685 and 14,987 codons were detected, among which 9942 and 10,207 had the relative synonymous codon usage (RSCU) values higher than 1, respectively. *A. meioclados* and *A. munitus* have 26 SSRs patterns, among which *A. meioclados* was 25 and *A. munitus* 21. The average Ka/Ks value was 0.36, and positive selection was detected in genes of the photosynthetic system (*ndhF* and *rbcL*) in *Asparagus* species. To perform the comparative analysis of plastomes, the two newly sequenced plastomes of the *A. meioclados* and *A. munitus* species were compared with that of *A. cochinchinensis*, and 12 hotspots, including 5 coding regions and 7 inter-genomic regions, were identified. Based on the whole plastome of *Asparagus*, 2 divergent hotspots (*accD* and *rpl32-trnL-UAG*) and 1 international barcode fragment (*rbcL*) were screened, which may be used as particular molecular markers for the identification of *Asparagus* species. In addition, we determined the phylogenetic relationship between *A. meioclados* and *A. munitus* in the genus *Asparagus*. This study enriches our knowledge of the molecular evolutionary relationships of the *Asparagus* genus and provides treasured data records for species identification, molecular breeding, and evolutionary analysis of this genus.

Asparagus is a genus belonging to the family Liliaceae, which contains 300 species, and it spreads in temperate and tropical areas all over the world except America. There are 31 *Asparagus* species in China, including 29 wild species and 3 cultivated species[1]. Some have high medicinal value, such as *Asparagus cochinchinensis* (Lour.) Merr., *Asparagus meioclados* H. Lev., and *Asparagus munitus* F. T. Wang & S. C. Chen. Among *Asparagus* plant materials, only the dried root tuber of *A. cochinchinensis,* which is used as Tian-Dong has been indexed in the latest edition of the Chinese Pharmacopoeia in China[2]. Tiandong has been used in China for thousands of years with the prominent effects treating fever, cough and vomiting, sore throat, constipation, and other diseases. *A. meioclados*, a species closely related to *A. cochinchinensis*, has been used as the "Mo tonic" in Yi medicine to treat palpitation, anxiety, fatigue, whooping cough (pertussis), cough, chest pain, constipation, abdominal pain,

[1]School of Ethnic Medicine, Chengdu University of Traditional Chinese Medicine, No. 1166 Liutai Avenue, Wenjiang District, Chengdu 611137, China. [2]Chongqing Academy of Chinese Materia Medica, 34 Nanshan Road, Huangjueya, Nanan District, Chongqing 400065, China. [3]College of Life Science and Food Engineering, Chongqing Three Gorges University, 666 Tianxing Road, Wanzhou District, Chongqing 404100, China. ✉email: liu0906xue@163.com; quxy@sina.com

and rheumatism in particular provinces such as Sichuan, Guizhou, and Yunnan[3]. *A. munitus*, another species also closely related to *A. cochinchinensis*, has been mainly distributed in the southwest of Sichuan (Muli) and northern Yunnan (Yongning)[4]. Because of the medication habits and preferences in a few regions of China, the indiscriminate use of *A. meioclados* and *A. munitus* as *A. cochinchinensis* has posed hidden risks to the accuracy and safety of clinical trials of Tiandon[5], and been not conducive to the quality control of the traditional Chinese medicine Tiandong. Some phenotypic characteristics of *Asparagus* are overlap. Some phenotypic characteristics of *Asparagus* overlap. The medicinal plants of the genus *Asparagus* have branched leaves that are both diverse and variable in shape within species, and the base of the stem that produces scale-like leaves extends into stiff spines of different lengths. Moreover, the short flowering period and similar flower color make their identification difficult[6]. Therefore, the identification and classification of species in the genus *Asparagus* was controversial. To solve these problems, many scholars studied the phylogenetic relationships among a few species of the genus *Asparagus* based on morphology[7–9], ribosomal DNA (rDNA) sequence[10,11], and combination of a single or a few genes of chloroplast DNA (cpDNA)[12–18], identified a few species and clarified the relationships among them, however, *A. meioclados* and *A. munitus* were not among these species. The identity of and phylogenetic relationships among the related *Asparagus* species remain completely unresolved.

Chloroplast (cp) is a substantial plant organelle with a prokaryotic origin that performs vital functions in the transfer and expression of genetic information in the life cycle of plants and algae[19]. At present, the plastomes of numerous plant species sequenced rapidly and efficiently, have been used in phylogenetic studies and for the identification of related species and have a higher copy number and relatively small size compared to the nuclear genomes. The plastomes are maternally inherited in angiosperms and highly conserved and have been proven effective in phylogenetic analysis for clarifying complex phylogenetic relationships[20–22]. Furthermore, chloroplast genome data are a useful aid to the development of plastid genetic markers in phylogenetic studies[23–25].

To distinguish the medicinal plants in the genus *Asparagus* and determine the phylogenetic relationships among them, we sequenced the complete plastome of *A. meioclados* and *A. munitus* for the first time, conducted a comprehensive analysis of these plastomes to obtain useful super barcodes and specific DNA barcodes based on hypervariable fragment for the molecular identification of Tiandong, clarify the phylogenetic relationship of the genus and provided basic genetic information for further understanding the evolutionary relationship of *Asparagus*.

Results

Plastome sequencing and assembly. After DNA sequencing on the Illumina platform, 21,879,295 and 22,675,012 clean reads were obtained for A. meioclados and *A. munitus*, respectively. The average depth of *A. meioclados* and *A. munitus* were 1452 and 1952, respectively. The line plots (Figs. S1–2) based on the genome position information and coverage showed that all of the location of the genome had high coverage (no breakpoints), which proved that the assembly results of the plastids were reliable.

Plastome structure and characterization in *A. meioclados* and *A. munitus*. The plastomes of *A. meioclados* and *A. munitus* have the typical quadripartite structure consisting of two inverted repeat regions (IRA and IRB), a large single copy region (LSC) and a small single copy region (SSC). The difference in the size of the two plastomes was found to be 138 bp, which mostly occurred at the LSC (Fig. 1). In addition, 111 unique genes were encoded in each plastome, which included 78 protein-coding genes, 29 tRNA genes and 4 rRNA genes (Table 1). Moreover, there were a total of 19 genes located in the IR regions, including 7 protein-coding genes, 8 tRNAs, and 4 rRNA. Besides, the genes *ycf1* and *ndhF* were located at the junction of the SSC/IR borders, while *rpl22* was located at the LSC/IR border. A total of 16 genes, comprising 6 tRNAs (2 *TrnA-UGC*, 2 *trnI-GAU*, *trnK-UUU*, and *trnG-UCC*) and 10 protein-coding genes (*atpF*, *ndhA*, 2 *ndhB*, *petB*, *petD*, 2 *rpl2*, *rpl16*, and *rps16*), contained 1 intron. The genes *clpP*, *ycf3*, and *trnV-UAC*, however, contained 2 introns (Table S1). The guanine-cytosine (GC) content in IR regions was 42.91% and 42.92%, higher than that in other regions, in *A. meioclados* and *A. munitus*, respectively. Overall, the GC content in the IR, LSC, and SSC regions between the two species was similar.

Codon usage bias analysis. To keep away from sampling errors in the analysis of codon usage bias, we estimated the relative synonymous codon usage frequency (RSCU) for 36 common eligible protein-coding sequences of *A. meioclados* and *A. munitus* (Fig. 2). The consequences showed that the codons of all 20 amino acids in the plastome of *A. meioclados* and *A. munitus* were identical, with a total of 36 protein-coding genes containing 14,685 and 14,987 codons, respectively, among which 9942, 10,207 had RSCU > 1 (Tables S2, S3). Leucine (Leu) and cysteine (Cys) were the most and least frequently used amino acids. All amino acids except Met and Trp were encoded by multiple synonymous codons, and Arg, Leu and Ser synonymous codons more than others. There were 14 amino acids with at least one codon which RSCU > 1, indicating that the codon is strongly preferred. In addition to the termination codon, the UUA codon of Leucine (Leu) had a RSCU value of 1.95, indicating its high frequency of use in the genome. Except for Met (Methionine-AUG) and Trp (Tryptophan-UGG), all of the other amino acid codons showed a strong bias towards A or U at the third position in codon. In addition, the EMBOSS software package was used to calculate the GC content at the first, second and third base in codon (GC1, GC2 and GC3, respectively) of the genome. Then, the GC3 and GC12 (mean values GC content of GC1 and GC2), were divided into horizontal and vertical coordinates to draw neutral plots to analyze the relative contribution of mutation pressure and natural selection to the formation of codon usage patterns (Fig. 3). The neutral map showed a weak correlation between GC3 and GC12, and the correlation coefficients for *A. meioclados* and *A. munitus* were 0.01376 and 0.03533, respectively. According to the neutrality plots, the slope of the regression line was close to 0, and all the genes represented by scatter points were on the
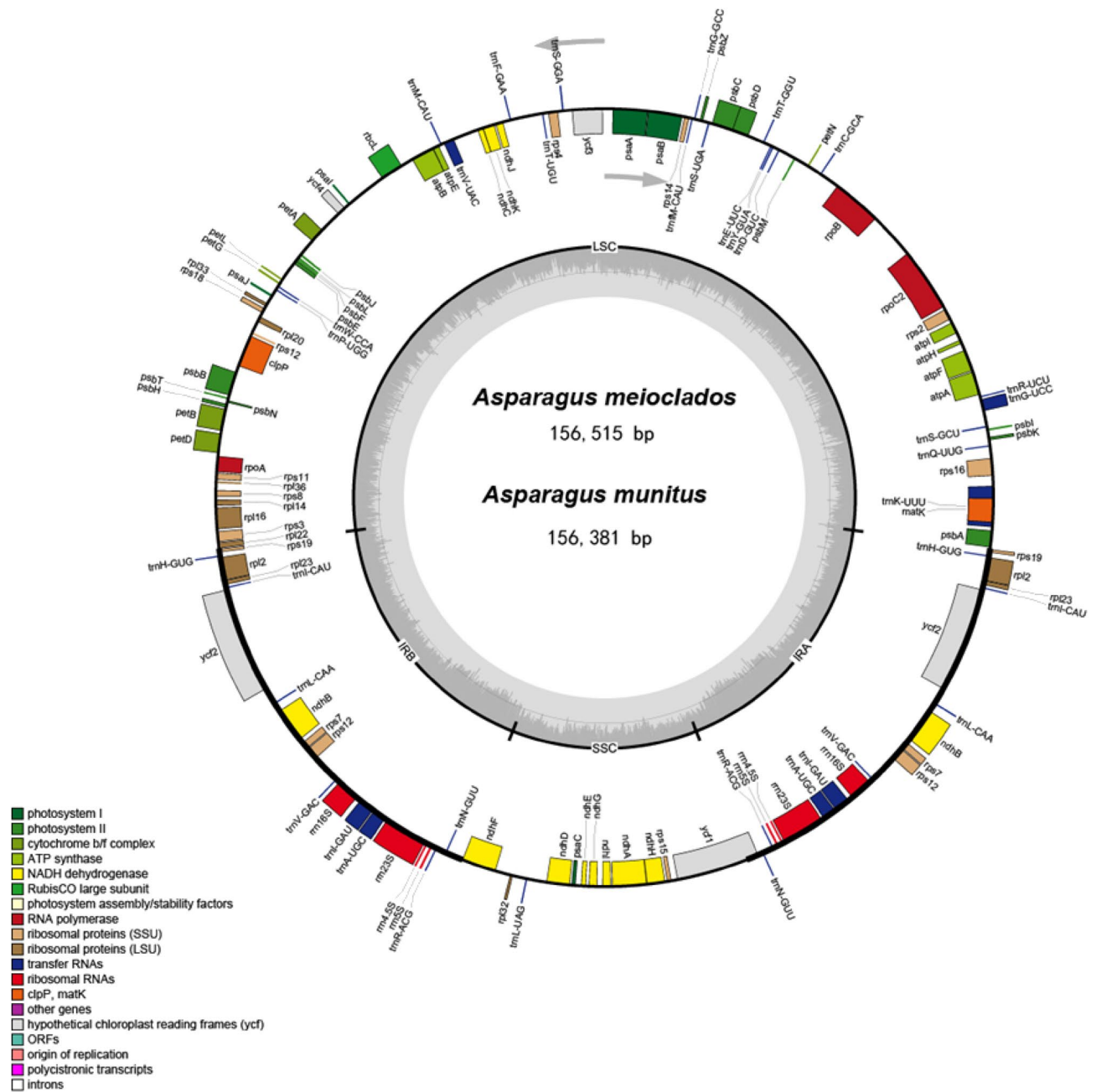
Figure 1. Plastome map of *A. meioclados* and *A. munitus*. The circle diagram showed all the genes in the small single copy (SSC) region, the large single copy (LSC) region, and the reverse repeat (IRa and IRb) region, with different functional groups of genes represented in different colors. Genes in the circle were transcribed clockwise and genes outside the circle were transcribed anticlockwise. The AT/GC contents correspond to the light and dark gray of the inner circle, respectively.

diagonal line. The slope of the regression line for *A. meioclados* was lower than that for *A. munitus*, and the points fell almost on a horizontal line.

Characterization of repeat sequences and simple sequence repeats (SSRs). After scanning SSRs among the plastomes, a total of 79 and 83 SSRs in *A. meioclados* and *A. munitus* were detected, respectively, which comprised mono-, di-, tri-, and tetra-nucleotide repeats, but pentanucleotide repeats were only detected in *A. munitus* (Table 2). We observed a total of 26 SSRs patterns in two plastomes among which 25 in *A. meioclados* and 21 in *A. munitus* (Fig. 4). Among the mononucleotide SSRs, polyadenine (polyA) or polythymidine (polyT) repeat patterns occurred most frequently, with only 1 cytosine (C) repeat and no guanine (G) repeat. Tandem repeats were the most frequently occurring repeats with 39 and 28 repeats detected in *A. meioclados* and *A. munitus*, respectively. Nineteen palindromic repeats were detected in both *A. meioclados* and *A. munitus*, 17 and 13 forward repeats were found, respectively. Only 1 reverse repeat was detected in *A. munitus*. The length of

Species	*A. meioclados*	*A. munitus*
Genome size (bp)	156,515	156,381
LSC (bp)	84,775	84,635
IRs (bp)	53,092	53,118
SSC (bp)	18,648	18,628
Total number of genes (unique)	111	111
Protein-coding gene (unique)	78	78
rRNA (unique)	4	4
tRNA (unique)	29	29
GC content (%)	37.55	37.57
LSC (%)	35.55	35.56
IR (%)	42.91	42.92
SSC (%)	31.40	31.43

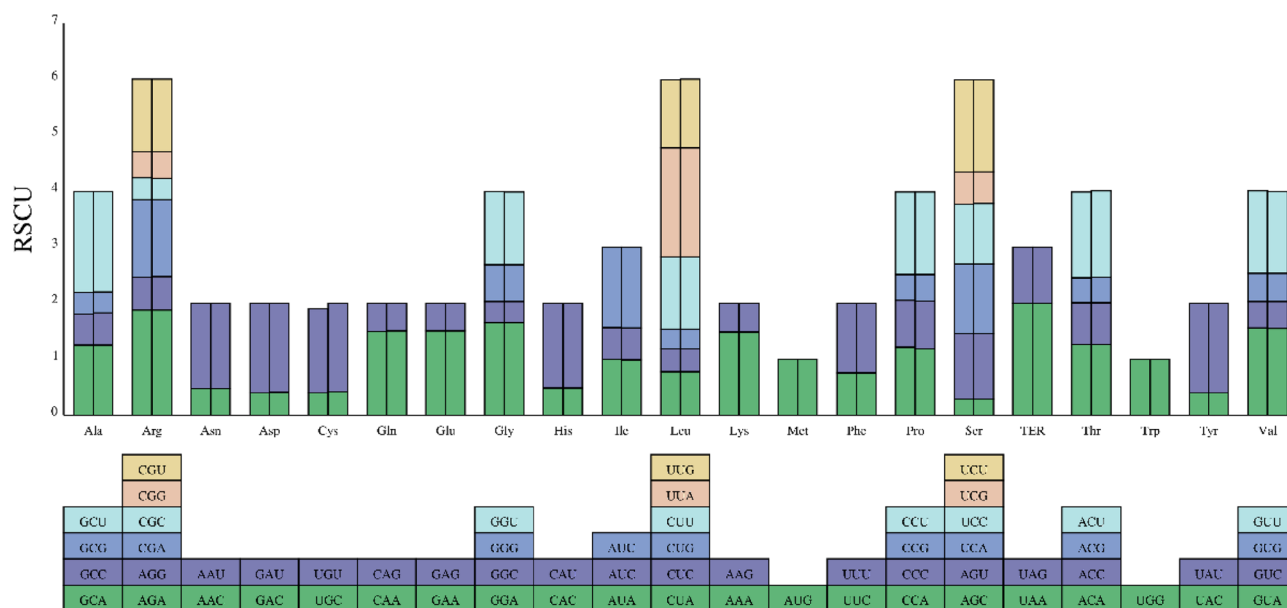**Table 1.** Features of two *Asparagus* plastomes.

Figure 2. Codon content of CDS in the plastomes of *A. meioclados* (left) and *A. munitus* (right). The abscissa represented 20 amino acids and terminators, and the ordinate represented the RSCU value.
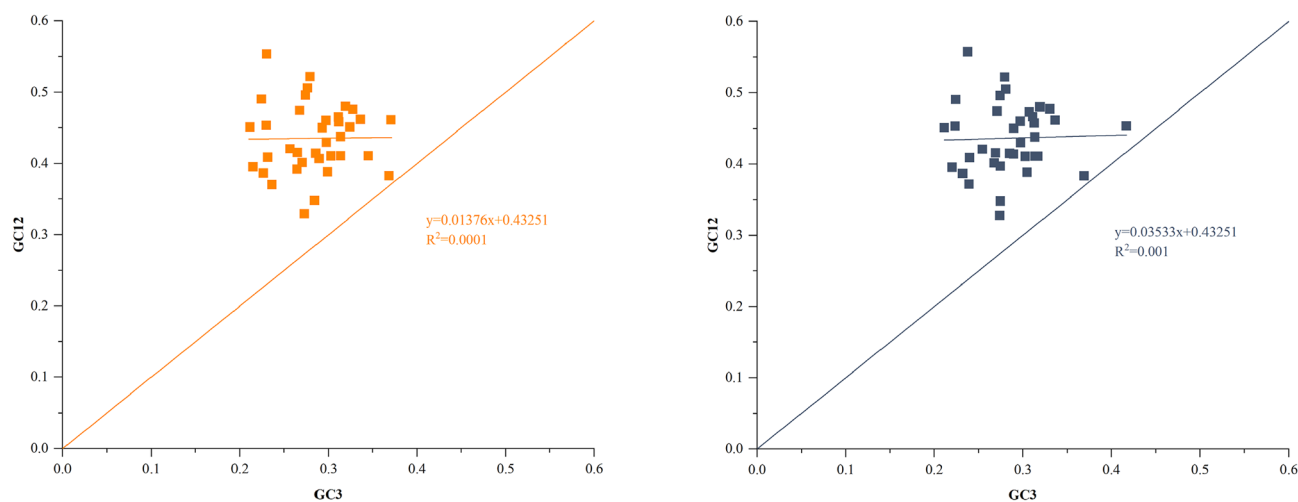

Figure 3. Neutrality analysis performed by plotting GC12 values against GC3 values for the plastomes of *A. meioclados* (orange) and *A. munitus* (blue). The diagonal line on the neutrality plot showed that the value of GC12 was equal to GC3.

Type	A. meioclados	A. munitus
Simple sequence repeat (SSR)		
LSC	51	47
IR	12	12
SSC	20	20
Intron	7	4
Intergenic	47	52
Exon	29	23
Mono-nucleotide	50	49
Di-nucleotide	13	13
Tri-nucleotide	4	5
Tetra-nucleotide	12	11
Penta-nucleotide	4	1
Total	83	79
Tandem repeat		
LSC	26	17
IR	7	7
SSC	6	4
Intron	1	2
Intergenic	30	16
Exon	8	10
Total	39	28
Reverse repeat		
Forward repeat	17	13
Palindromic repeat	19	19
Reverse repeat	0	1
Total	36	33

Table 2. Distribution of repeat number in plastomes in *A. meioclados* and *A. munitus*.
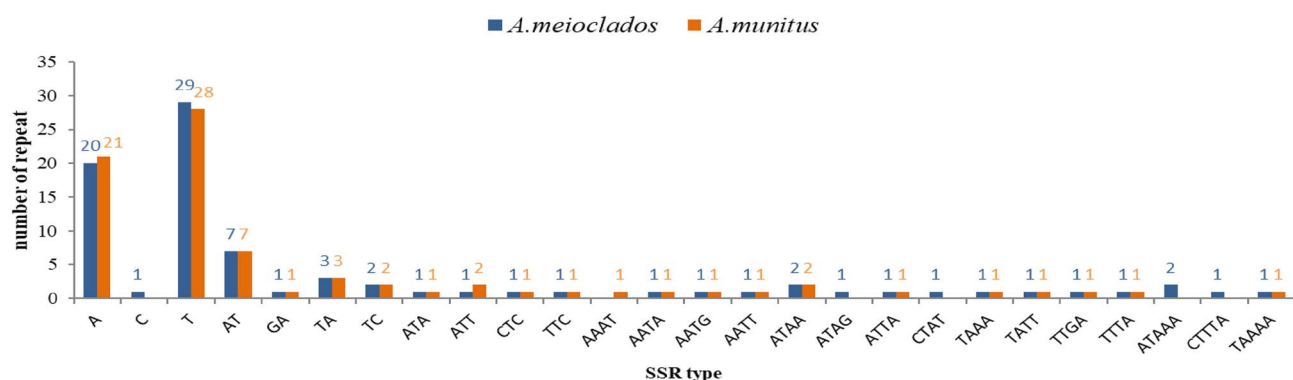

Figure 4. Repeat types, number and length of plastomes in *A. meioclados* and *A. munitus*.

LSR in *A. meioclados* ranged from 30 to 52 bp, while for *A. munitus*, it was between 30 and 56 bp. The plastome of *A. meioclados* contained 24 repeats of 30–35 bp, 5 repeats of 36–40 bp, and 7 repeats longer than 40 bp. The plastome of *A. munitus*, however, contained 23 repeats of 30–35 bp, 5 repeats of 36–40 bp, and 6 repeats longer than 40 bp. Except for *accD* and *atpF* genes, most long sequence repeats (LSR) were located in non-coding regions (Fig. 5).

Divergent hotspots identification of three medicinal species. To detect the sequence polymorphisms that could be used as candidate sites in molecular markers, we performed mVISTA, IRscope, and sliding window analyses of the plastomes of *A. meioclados* and *A. munitus*, with *A. cochinchinensis* used as the reference. The results showed that the three plastomes were similar. Divergent hotspots mainly concentrated in the non-coding region and a few exonic regions, while the coding region was found to be more conserved. The sequence variations detected in LSC and SSC regions were greater than those in IR regions. There were 15 intergene spacer regions (*rps16-trnQ*, *trnS-trnG*, *atpF-atpH*, *rpoB-trnC*, *petN-psbM*, *trnE-trnT*, *psaA-ycf3*,
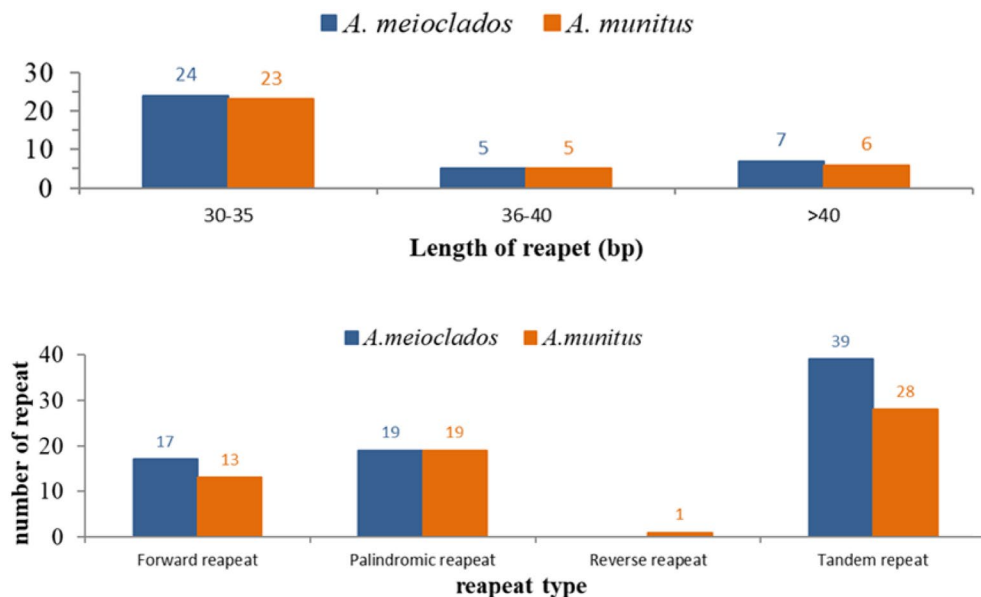
Figure 5. Frequency of repeat sequences of the *A. meioclados* and *A. munitus* plastomes determined by REPuter.

trnT-trnL, ndhC-trnV, petA-psbJ, rps18-rpI20), rpl16-rps3, rpl32-trnL, ccsA-ndhD, and *rps15-ycf1* and 2 gene regions (*accD* and *ycf1*) had the highest difference (Fig. 6).

The expansion and contraction of IR boundaries are the main reasons for differences in the size among cp genomes of plants during evolution[25]. The IRscope analysis showed that the gene types of LSC, SSC, IRa, and IRb joining areas were the same in the genome of the three species (Fig. 7), which had the characteristics of conservation. The genes of *rpl22, rps19, trnN,* and *ndhF*, exhibited differences in the location of the boundaries; the *rpl22* genes in *A. meioclados* and *A. munitus* were 2 bp away from the IRb-LSC border; while the *A. cochinchinensis* was 58 bp away from the IRb-LSC border. The *rps19* genes in *A. cochinchinensis* and *A. munitus* were located 53 bp away from the IRb-LSC border, whereas, in *A. meioclados,* they were 55 bp away from the IRb-LSC border. It is noteworthy that in these three plastomes, the *ycf1* gene existed in both the IRa and SSC regions, including the IRa-SSC junction. The length of IRa in the *ycf1* in *A. munitus* was 802 bp, while in *A. meioclados*, it was 790 bp.

We compared the nucleotide diversity (Pi) values of the three plastomes, which were found to be within the range of 0–0.02333. We mapped 12 mutation hotspots (Fig. 8), which Pi > 0.01, including 5 coding regions (*accD, ndhF, ndhD, rps3,* and *ycf1*) and 7 intergenic regions (*ndhG-ndhI, rpl32-trnL-UAG, trnE-UUC-trnT-GGU, trnG-UCC-trnR-UCU, atpB-rbcl, trnR-UCU-atpA,* and t*rnS-UGA-psbZ*).

Selective pressure analysis. The non-synonymous to synonymous substitution ratio (Ka/Ks) of 72 protein-encoding genes in 9 plastomes were compared with those of *A. cochinchinensis* to evaluate the selection pressure between species. The average Ka/Ks ratio of these genes was 0.36. Since the Ks value of 33 genes was 0, the rest (only 39 genes) were used in the Ka/Ks ratio test (Fig. 9). Using the KaKs calculator software, the Ka/Ks ratio of less than 0.5 was obtained for most genes, indicating that there were obvious patterns of purifying selection for these protein-encoding genes. A few photosynthesis-related genes showed positive selection in partially paired species (Ka/Ks ratio > 1), e.g., the *ndhF* gene was positively selected in 2 pairs (*A. cochinchinensis* & *A. meioclados* and *A. cochinchinensis* & *A. munitus*), *rbcL* gene was positively selected in 4 two-pair species (*A. cochinchinensis* & *A. meioclados, A. cochinchinensis* & *A. racemosus, A. cochinchinensis* & *A. setaceus* and *A. cochinchinensis* & *A. schoberioides*). This suggested that the need for adequate light in some species of *Asparagus* might have exerted a strong selective force on these genes during evolution.

The *accD* genes of all paired species except *A. cochinchinensis* & *A. aethiopicus* and *A. cochinchinensis* & *A. densiflorus* were positive selected, and the expression of *accD* genes might indirectly affect the efficiency of photosynthesis, possibly as a result of the genus's adaptation to its poor living environment.

Phylogenetic analysis. To avoid the heterogeneity of evolutionary rate between genes, misjudgment of orthologous lineages, and incomplete lineage sorting, which may affect the phylogenetic reconstruction, we used concatenation method (maximum likelihood and Bayesian inference) and coalescence method (wASTRAL) to reconstruct the phylogenetic relationship of *Asparagu*s and evaluated the phylogenetic positions of *A. meioclados* and *A. munitus*. Based on the maximum likelihood (ML), Bayesian inference (BI) analysis and wASTRAL of the entire plastomes, a consistent topology was obtained. The phylogenetic positions detected in this study were strongly supported by bootstrap values and posterior probabilities from ML, BI and wASTRAL analyses,
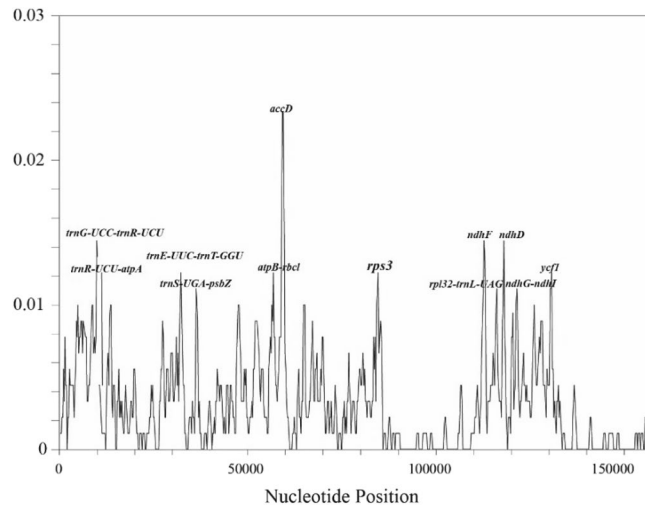
Figure 6. Comparison of three plastomes using mVISTA program with *A. cochinchinensis* plastome as a reference. The top gray arrows above the alignment indicate genes and their orientation. Genome regions are color coded. Blue and red areas indicate protein coding regions and the conserved non-coding sequences (CNS) regions, respectively. The green areas represent untranslated region (UTR) that do not encode proteins but have translational regulatory functions.

Figure 7. Boundary difference analysis of the quadripartite structure of three plastomes of the genus *Asparagus*.

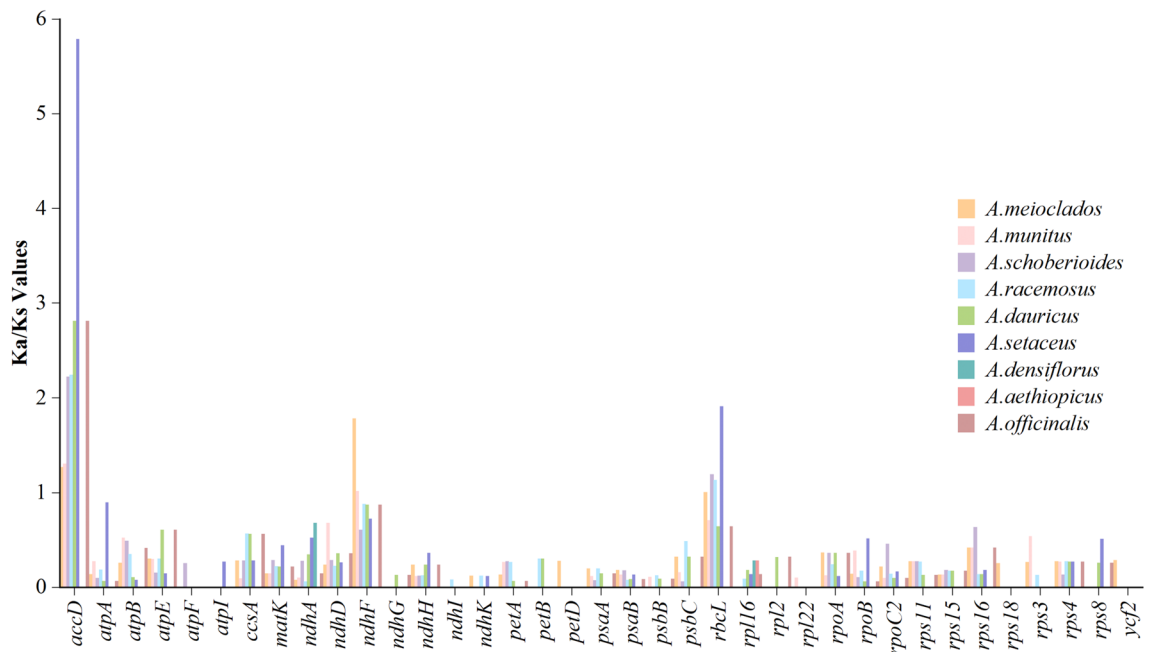**Figure 8.** Nucleotide diversity (Pi) values among the three plastomes.

Figure 9. The Ka/Ks ratios of 37 protein-coding genes of *A. cochinchinensis* plastome when compared to other nine species of *Asparagus*.

respectively, which were annotated at corresponding nodes (Fig. 10, Fig S3). *Asparagus* species and the outgroup species were divided into different clades, e.g., 10 *Asparagus* species were grouped into four clades, which were sister clades, and *A. meioclados*, *A. munitus*, and *A. racemosus* were clustered into a single branch, while *A. dauricus*, *A. officinalis*, and *A. schoberioides* were gathered into a single branch. Furthermore, *A. cochinchinensis*, *A. densiflorus*, and *A. aethiopicus* were clustered together, but *A. setaceus* alone was in a separate branch.

Evaluation of DNA barcoding. To find the genomic fragments that could identify *Asparagus* species, we evaluated 12 divergent hotspots and 3 international barcode fragments (*matK*, *rbcL*, and *trnH-psbA*) in the plastomes in terms of the degree of the variation in fragments and the applicability of DNA barcoding for species identification. The length of 14 fragments ranged from 130 to 5694 bp, with *trnR-UCU-atpA* and *ycf1* being the shortest and the longest fragments, respectively (Table 3).The proportion of variable sites ranged from 0.77% (*trnR-UCU-atpA*) to 41.33% (*trnH-psbA*), parsimony informative sites ranged from 0.32% (*matK*) to 3.08% (*ndhG-ndhI*), Singleton sites accounted for 0.00% (*trnR-UCU-atpA*)–40.00% (*trnH-psbA*), in which the variable sites with proportion > 2% were *rbcL*, *accD*, *rnH-psbA*, *ndhG-ndhI*, *rpl32-trnL-UAG*, *trnE-UCC-trnT-GGU*, *trnS-UGA-psbZ*. The NJ tree was constructed using 12 divergent hotspots and 3 international common
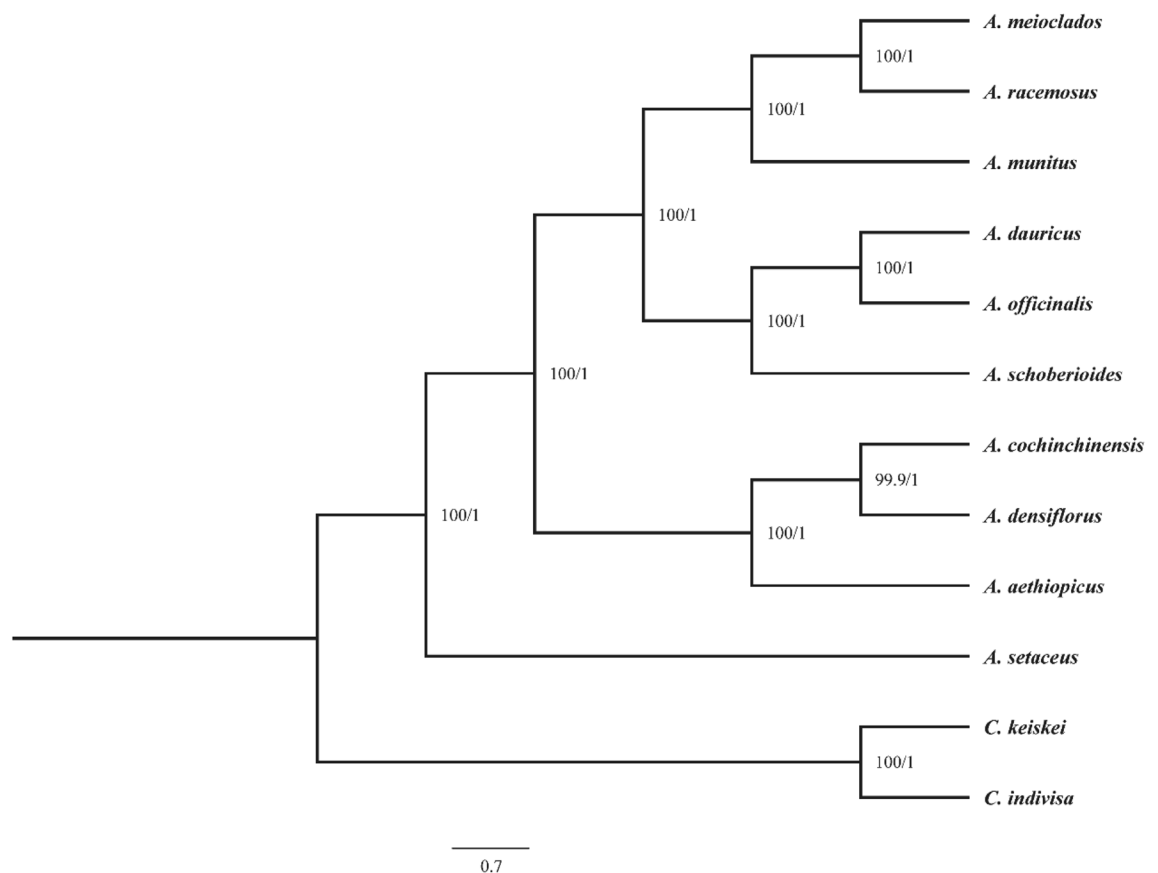
Figure 10. Phylogenetic tree reconstruction of the 10 species inferred from Bayesian inference (BI) and maximum likelihood (ML) based on cp DNA. Numbers represent ML/ BI posterior probabilities.

Gene	Aligned length	Variable sites	Parsimony informative sites	Singleton sites	% of variable sites	% of parsimony informative sites	% of Singleton sites
matK	1563	18	5	13	1.15	0.32	0.83
rbcL	1443	33	20	13	2.29	1.39	0.90
trnH-psbA	75	31	1	30	41.33	1.33	40.00
accD	1220	54	25	29	4.43	2.05	2.38
ndhD	1245	18	9	9	1.45	0.72	0.72
ndhF	2223	40	19	21	1.80	0.85	0.94
rps3	687	8	3	5	1.16	0.44	0.73
ycf1	5694	104	75	29	1.83	1.32	0.51
atpB-rbcL	812	14	4	10	1.72	0.49	1.23
ndhG-ndhI	390	20	12	8	5.13	3.08	2.05
rpl32-trnL-UAG	1020	29	13	16	2.84	1.27	1.57
trnE-UCC-trnT-GGU	508	19	12	7	3.74	2.36	1.38
trnG-UCC-trnR-UCU	141	8	3	5	5.67	2.13	3.55
trnR-UCU-atpA	130	1	1	0	0.77	0.77	0.00
trnS-UGA-psbZ	385	19	9	10	4.94	2.34	2.60

Table 3. Sequence diversity comparison of 12 divergent hotspots and 2 international barcode fragments in 10 *Asparagus* species.

barcode fragments, and the identification rate of 10 *Asparagus* species was 30% (*trnR-UCU-atpA*)–100% (*atpB-rbcL*, *ndhF*, and *accD*), among them, the identification rate > 80% were *accD*, *ndhF*, *rbcL*, *ycf1*, *atpB-rbcL*, and *rpl32-trnL-UAG* (Table S4). Finally, 2 divergent hotspots (*accD* and *rpl32-trnL-UAG*) and 1 international barcode fragment (*rbcL*) were selected, which identification rate greater than 80% and the proportion of variable sites greater than 2%. These divergent hotspots could be used as specific DNA barcodes the identification of

Asparagus species. The Neighbor Joining (NJ) tree (Figs. S4–S6) based on divergent hotspots showed highly similar topological relationships to the plastome ML tree, with only a few differences in *A. cochinchinensis*, *A. densiflorus* and *A. aethiopicus*. *A. cochinchinensis* was sister to *A. densiflorus* and *A. aethiopicus* in the ML tree of the plastome. However, the neighbor-joining (NJ) tree results based on *accD* showed that *A. cochinchinensis*, *A. densiflorus* and *A. aethiopicus* were one clade. The phylogenetic relationship obtained by Neighbor Joining (NJ) tree of *Asparagus* based on *rpl32-trnL-UAG* was consistent with that obtained by *accD*. The phylogenetic relationship obtained by Neighbor Joining (NJ) tree of *Asparagus* based on *rbcL* was completely consistent with that obtained by plastome ML tree. Except for the *A. aethiopicus* and *A. cochinchinensis* branches (Bootstrap value = 0.31), the other branches were given high support (Bootstrap value > 0.5).

Discussion

Similar to most angiosperm plastid genomes[25], the plastomes of *A. meioclados* and *A. munitus* exhibited a typical quadripartite structure consisting of 2 reverse repeat regions (IRA and IRB), a large single copy region (LSC) and a small single copy region (SSC). The plastomes of *A. meioclados* and *A. munitus* were slightly smaller than those of other *Asparagus* species[26,27]. The length of LSC, IR and SSC regions, gene content (Table 1), gene order, and number of introns for *A. meioclados* and *A. munitus* showed high similarity, similar with plants of the genus *Camellia*[28], suggesting that the chloroplast genome was relatively conserved. The variation in the frequency of synonymous codons used in protein synthesis is of great significance for studying the origin and evolution of species, gene functions, and protein expression[29,30,73]. The codons encoding all amino acids, except for Met and Trp, showed a strong bias for A or U in the third codon positions. It was in agreement with *Cicer echinospermum*[31] and *Carpesium abrotanoides*[32], but not with *Glechoma longituba*[33] and *Aconitum coreanum*[34]. The difference might be related to the different number of tRNAs in different species, and the use of codons with high abundance of paired tRNAs could guarantee efficient protein expression. The neutral map showed a weak correlation between GC3 and GC12, indicating that the GC content of the genome displayed a conservative pattern, and the codon usage bias was mainly affected by selection. In addition, the slope of the regression curve of *A. meioclados* and *A. munitus* was 0.01376 and 0.03533, respectively, indicating that natural selection had an important effect on the formation of codon bias in *A. meioclados* and *A. munitus*.

The intra-species variations in simple sequence repeats (SSRs) could provide a valuable resource for population genetics of polymorphism[35,36]. Similar to other studies, the nucleotide repeats consisting of A or T had the highest number in *A. meioclados* and *A. munitus*, and the SSRs were mainly found in the non-coding region and large and small single-copy regions (LSC and SSC) in our study[18].

Mononucleotide repeats were found to contribute more to genetic variations than other SSRs. Notably, most of the dinucleotide SSRs of *A. meioclados* and *A. munitus* were AT/TA/TC, which was consistented with previous reports of other angiosperms[37]. This confirmed that the simple repetitive sequence of chloroplast genome mainly consists of short poly real A and poly real T, rather than C or G[38]. Tandem repeated sequences have played a crucial role in genome size changes, genome rearrangement, gene duplication, evolution, and inheritance, and genetic variation[39,40]. Three types of tandem repeats were detected in *A. meioclados* and *A. munitus*, with the palindromic repeats being the most abundant repeats, followed by forward repeats. Only the plastome of *A. munitus* had inverted repeats, similar to the case observed for Chinese Polyspora[41]. These differences suggest significant sequence variation and genome rearrangements that have occurred during evolution.

Although the cp genomes of plants are conserved in size and structure, IR amplification and contraction are common evolutionary phenomena[42]. Comparative analysis of IR boundaries clearly showed differences in chloroplast genomes among *Asparagus* species. The results of mVISTA, IRscope, and sliding window analyses of the plastomes of *A. meioclados* and *A. munitus* confirmed that the differences in plastomes in the LSC region were greater than those in the SSC region and also greater in the non-coding region than in the coding region, possibly as a result of gene copy number differences in the inverted repeats sequences caused by gene conversion[43]. The results of the present study demonstrated that the differences in the nucleotide diversity greater than 1.5% in hotspots could be used for phylogenetic analysis, genetic analysis, and identification of plant seed species[44,45]. The identified hotspots included 7 intergenic regions (*ndhG-ndhI*, *rpl32-trnL-UAG*, *trnE-UUC-trnT-GGU*, *trnG-UCC-trnR-UCU*, *atpB-rbcl*, trnR-UCU-atpA, and *trnS-UGA-psbZ*), which could serve as potential molecular markers for identifying the three *Asparagus* species.

Adaptive evolution is defined as the improvement in the fitness of a plant species during evolution and has always been driven by evolutionary processes, like natural selection, that act to increase genetic variation sourced from mutation, genetic recombination, and gene flow and also contribute to biodiversity at all levels of biological organization[46]. In *Asparagus*, the *ndhF* and *rbcL* genes, which belong to the photosynthetic system, were found to be under positive selection. The *ndh* gene encodes cp NADH dehydrogenase[47]. The NADH dehydrogenase complex in higher plants is not only involved in photosynthetic electron transport but also resistant to photooxidative stress[48,49]. The *rbcL* gene encodes ribulose 1,5-bisphosphate carboxylase/oxygenase, and *rbcL* plays an important role as a regulator of photosynthetic electron transport and is required for photosynthesis[50]. This suggests that the need for enough light that some *Asparagus* species have might have exerted strong selective forces on these genes during evolution. The *accD* gene encodes the β-carboxyltransferase subunit of the acetyl-Coa carboxylase complex[51] and plays an important role in leaf development and affects leaf lifespan and seed yield[52]. Additionally, the positive expression of the *accD* gene might indirectly affect the efficiency of photosynthesis, which could be due to the adaptation of the species of this genus to poor environments in which they grow. In summary, positive selection would possibly contribute to *Asparagus* species diversification and adaptation.

Numerous practical studies have demonstrated that organelle phylogenies are an important tool for clarifying the genetic relationships between species[46,53]. In the past two decades, there are a large variety of researches

on biological networks emerging, which have attracted great attention from scientists and scholars[54]. In order to get more accurate phylogenetic relationship of the genus asparagus, we compared the concatenate method and coalescent method, obtained a consistent topology. The 10 species of *Asparagus* were divided into four evolutionary clades, *A. meioclados*, *A. munitus* and *A. racemosus* formed a single branch, while *A. dauricus*, *A. officinalis* and *A. schoberioides* gathered into another branch. *A. cochinchinensis*, *A. densiflorus*, and *A. aethiopicus* were grouped in a third branch, and *A. setaceus* was located in the most basic branch, which was only including its own. These results are consistent with the phylogenetic relationships reported in previous studies[18,55], and further clarify the phylogenetic relationships within *Asparagus* species. According to the phylogenetic position of the species within the genus *Asparagus* determined in this study, *A. meioclados* was most closely related to *A. racemosus*, and *A. cochinchinensis* was closest to *A. densiflorus*. According to the traditional classification of *Asparagus*[56], dioecious and monoecious species were divided into two major factions, but *A. setaceus* and *A. densiflorus* (both dioecious species) were in different branches with high bootstrap values in our study. The results showed that at the molecular level, monoecious species of the genus *Asparagus* were not necessarily completely divergent, more data might be needed to confirm this speculation.

In recent years, DNA barcoding has played an extremely important role in improving the efficiency of plant species identification. From single-locus DNA barcodes to combined markers and even to the genome level, DNA barcoding has provided more and more genetic information. The plastome has the potential to serve as a highly accurate "super DNA barcode" used for taxonomic revision, phylogeny inference, and species identification[57,58]. But the complete plastome sequencing is cost and time-consuming. A short, easily amplified, and variable sequence as an ideal DNA barcoding is becoming increasingly necessary for the study of Traditional Chinese Medicine[59], and it has been successfully used for many taxa. The development of specific barcodes has been shown to make the rapid and accurate identification of specific taxa feasible[60–63]. However, due to the complex evolutionary history of plants, some groups cannot be identified using the existing international common barcodes (*matK*, *rbcL*, and *trnH-psbA*)[64,65]. Therefore, the development of new DNA barcodes with high identification rates and group specificity is currently being researched. In this study, 2 divergent hotspots (*accD*, and *rpl32-trnL-UAG*) and 1 international barcode fragment (*rbcL*) were identified from the whole plastomes of *Asparagus* species, with the discrimination rates higher than 80% and proportions of variable sites greater than 2%. These potential molecular markers could be used for identifying medicinal species of the genus *Asparagus*. In addition, other analysis and application of the chloroplast genome of *Asparagus*, such as SNP loci and digital PCR analysis techniques, could be used to further accurately identify *Asparagus* species.

Conclusion

In this study, for the first time, the plastomes of *A. meioclados* and *A. munitus* were obtained. The results of genome sequencing, assembly, and annotation, as well as comparative analyses showed that both plastomes had similar gene and GC contents but differed in tandem repeats and SSRs. Twelve mutation hotspots, including 5 coding regions and 7 intergenic regions, were detected in the two plastomes compared with that of *A. cochinchinensis*. Phylogenetic analysis showed that *A. meioclados* and *A. munitus* were in different clades from *A. cochinchinensis*, and a more accurate assessment of phylogenetic relationships in *Asparagus* was carried out in the present work compared to previous studies. Two divergent hotspots and 1 international barcode fragment were selected as specific DNA barcodes for the identification of *Asparagus* species with a discrimination rate higher than 80% and a proportion of variable sites greater than 2%. With the development of plant science, plastid transformation is becoming an important tool for plant species identification. This study used chloroplast genome fragments for the molecular identification of *Asparagus* species and it can provide valuable information about the genetic diversity and evolutionary patterns of *Asparagus* species.

Materials and methods

Plant material, DNA extraction and sequencing.
Fresh and healthy leaves of *A. meioclados* and *A. munitus* used in this study were collected from Miyi and Muli counties, Sichuan Province, respectively, and then they were frozen at − 80 °C. The tissue of all plant samples was used for the extraction of DNA with the modified cetyltrimethylammonium bromide (CTAB) method[66] and the concentration and quality of DNA were determined by the Qubit 3.0 Fluorometer (Invitrogen) and agarose gel electrophoresis (1%), respectively. High-quality DNA samples (concentration ≥ 5 ng/μL, total yield higher than or equal to 0.5 μg, integrity of the main dispersion band above 5 kb, and no obvious impurities below the qualification threshold) were used for library construction and then sequenced by the Illumina NovaSeq 6000 (San Diego, CA, United States). Eight *Asparagus* plastomes for the comparison with the main 3 plastomes used in this study, including *A. cochinchinensis* (MW970105), *A. schoberioides* (KX790361), *A. officinalis* (NC_034777), *A. dauricus* (MT712151), *A. racemosus* (MN736960), *A. setaceus* (MT712152), *A. densiflorus* (MT740250), and *A. aethiopicus* (MZ337394), were downloaded from the NCBI GenBank.

Plastome assemblage and annotation.
After DNA sequencing on the Illumina platform, clean reads were obtained. Raw reads with a low average quality (the Phred quality score (Q) ≤ 20) were removed using Trimmomatic v 0.3[67] with the parameters set as follows: sliding window 4–15, trailing 3, leading 3, and the minimum length of reads 50. The coverage depth of each site was counted using SAMTOOLS[68], and the assembly quality of plastids was judged by drawing line plots based on genomic position information and coverage[69]. The plastomes were assembled de novo by GetOrganelle[70]. The online tool CPGAVAS2 (http://47.96.249.172:16019/analyzer/home)[71], an integrated plastome sequence annotator and analyzer, was used to annotate the plastid genomes of *A. meioclados* and *A. munitus*, with *A. officinalis* (MT712156) and *A. schoberioides* (KX790361) used as references, respectively. The OGDRAW tool (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html)

with the default settings was used to draw the circular genome maps and then manually check them[72]. Finally, the annotated sequences for *A. munitus* and *A. meioclados* were submitted to GenBank (accession numbers OQ628361 and OQ628362, respectively).

Codon usage bias.   The variation in the frequency of synonymous codons used in protein synthesis is of great significance for the studies on the origin and evolution of species, gene functions, and protein expression[73]. To estimate the relative synonymous codon usage (RSCU) more accurately, all coding sequences (CDSs) were screened according to the following conditions: (1) deletion of repeated sequences; (2) removing CDSs shorter than 300 bp; (3) using ATG as the initiation codon; and (4) sequencing with termination codons (TAA, TAG, and ATG)[74,75]. CodonW v.1.4.2[76] was used to calculate the RSCU value for qualified sequences to quantify the degree of codon use bias. The GC content was calculated by the EMBOSS software suite[77]. The neutrality plot analysis was performed to determine the relative contributions of mutation pressure and natural selection to the formation of codon usage patterns.

Characterization of repeat sequences and SSRs.   SSRs and Tandem repeat sequences of cp genome are important for genome structure, phylogenetic relationship and population genetic analysis[78]. Repeats in the newly sequenced plastomes of the two species, including palindromic, reverse, and direct repeats were identified using REPuter[79] with the following settings: (1) a hamming distance of 3; (2) a minimum repeat size of 30 bp; and (3) maximum computed repeats of 5000. The simple sequence repeats (SSRs) of the plastomes of *A. meioclados* and *A. munitus* were identified using the MISA Perl script[80] with the search parameters set at 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats, respectively. Tandem repeats were detected by the online tool Tandem Repeats Finder (https://tandem.bu.edu/trf/trf.html) with default parameter settings[81].

Genome comparison and divergent hotspots identification.   To further develop molecular markers for the identification of medicinal plants of the genus *Asparagus*, we compared the newly sequenced plastomes of the two species with that of *A. cochinchinensis*. We used the mVISTA v.2.0[82] program in the Shuffle-LAGAN mode to compare the plastomes. On line program Irscope[83] (https://irscope.shinyapps.io/irapp/) was conducted to compare the large single copy (LSC), small single copy (SSC) and reverse repeat (IR) regions in plastomes. Finally, we identified the polymorphic regions. MAFFT[84] was used to perform the multiple sequence alignment (MSA) analysis of plastomes, and the DnaSP v6[85] software was used for performing the sliding window analysis. The nucleotide variability (Pi) of the coding and non-coding regions was calculated, and the window length and the step size were set to 600 and 200 bp, respectively.

Selective pressure analysis.   The ratio of the nonsynonymous substitution rate (Ka) to the synonymous substitution rate (Ks) (Ka/Ks) of protein-coding genes was used to explore the relationship between the growing environment and the evolutionary rates of protein-coding genes in the genus *Asparagus*. The Ka/Ks ratios of 10 species of the *Asparagus* genus were calculated using the YN model in KaKs_Calculator 2.0[86], Since the YN model considers the characteristics of DNA sequence evolution, it has been increasingly applied in the study of molecular evolution[87,88].

Phylogenetic analysis.   The whole chloroplast genomes of the above mentioned 8 species were used for phylogenetic analysis to determine the phylogenetic position of *A. meioclados* and *A. munitus* in the genus *Asparagus*, with *Convallaria keiskei* (accession number, NC_042228.1) and *Cordyline indivisa* (accession number, NC_035998) used as outgroups, respectively. The plastomes of these species were aligned using MAFFT[84], and the best-fitting model of nucleotide substitution (GTR + I + G) was determined using the Akaike Information Criterion (AIC) in jModelTest V2.1.10[89], MrBayes v3.2.6[90,91], and IQ-TREE v2.1.4[92] were use to perform phylogenetic analysis using Bayesian inference (BI) and maximum likelihood (ML) methods, respectively. The former used the Markov chain Monte Carlo (MCMC) algorithm and ran four chains for 1,000,000 generations at the same time, sampling every 1,000,000 generations. The first 25% of trees were discarded as burn-in. The remaining trees were used to generate a consensus tree, which was considered to reach a plateau, as the mean standard deviation (SD) of the split frequencies remained below 0.001. The latter [maximum likelihood (ML)] was calculated using ModelFinder in the IQ-TREE package and the Akaike Information Criterion (AIC) with 1000 bootstrap copies[93]. In addition, we reconstructed the species tree using two coalescence-aware methods: CASTER and wASTRAL. The aligned genomes were split into 340 segments of 500 base-pairs, and we estimated a gene tree for each segment by IQ-Tree with approximate Bayesian supports. The best nucleotide substitution model for each gene tree were find by IQ-Tree. CASTER-site v1.15.0.0[94] was used to infer the species tree directly from the aligned genomes. Local bootstrap supports were assessed based on blocks of 100 base-pairs. We also inferred the species tree from estimated gene trees using wASTRAL-hybrid v1.15.2.3[95] which utilized both branch lengths and branch supports. The built-in local posterior probabilities were displayed as branch supports. The phylogenetic tree visualized using FigTree v1.4 (http://tree.bio.ed.ac.uk/software/fgtree/).

Evaluation of DNA barcoding.   The divergent hotspots (pi > 0.015) and the international common barcode fragments (*matK*, *rbcL,* and *trnH-psbA*) in the plastomes of species were evaluated in terms of the degree of fragment variation and the efficiency of species identification. The MEGA v.11[96] software was used to measure the proportion of variable sites, parsimony informative sites and singleton sites. Based on the K2P model, MEGA v.11 was used to construct the neighbor-joining (NJ) tree, and the bootstrap method was used to repeat the operation 1000 times to test the reliability (support value) of each branch. The success of sequence

analysis was judged by the support value of species clustering (higher than 0.5). The number of fragments was counted to screen for suitable fragments as ideal DNA barcodes for the genus *Asparagus*.

Ethical approval.   The collection of plant materials (*A. meioclados* and *A. munitus*) used in this study complied with relevant institutional, national and international guidelines and legislation.

Data availability

References

1. Chen, S. C. & Tamanian, K. G. ASPARAGUS Linnaeus. *Flora China* **24**, 208–215 (2000).
2. State Pharmacopoeia Commission. *Pharmacopoeia Of The People's Republic of China, Vol ***1* 56–57 (China Medical Science and Technology Press, 2020).
3. Li, H. R. & Huang, Y. A brief introduction of Yi medicines for treatment of "Sise" and their types of "Sise" treated. *Chin. J. Ethn. Med.* **27**, 47–52 (2021).
4. Editorial Committee of Chinese Flora. *Flora of China, Vol ***15* 108 (Science Press, 1978).
5. Lv, X. Y. & Chen, A. M. Research status and prospects on cultivation and processing technique of *Asparagus cochinchinensis* in Neijiang Area. *Mod. Agric. Sci. Technol.* **14**, 77–79 (2019).
6. Altntas, S., Pakyürek, M., Sensoy, S., Erez, M. E. & Inal, B. Genetic diversity among some Asparagus species using rDNA ITS, cp DNA trnL intron sequence and screening for antioxidant activity. *Pol. J. Environ. Stud.* **28**(4), 1–7 (2019).
7. Ni, J. M. & Zhao, R. N. Comparative observation on pollen morphology of Asparagus in Gansu Province. *J. Lanzhou Univ. (Med. Sci.)*. **1**, 17–19 (1990).
8. Lu, L. Y. Comparative identification of *Asparagus cochinchinensis* and its counterfeit of *Asparagus filicinus*. *Lishizhen Med. Mater. Med. Res.* **3**, 662 (2007).
9. Fukuda, T. *et al.* Molecular phylogeny of the genus Asparagus (Asparagaceae) inferred from plastid petB intron and petD–rpoA intergenic spacer sequences. *Plant Species Biol.* **20**(2), 121–132 (2005).
10. Ou, L. J. *et al.* Comparison of rDNA internal transcribed spacer sequences in Asparagus. *J. Chin. Med. Mater.* **33**(10), 1542–1545 (2010).
11. Moreno, R., Espejo, J. A., Cabrera, A. & Gil, J. Origin of tetraploid cultivated asparagus landraces inferred from nuclear ribosomal DNA internal transcribed spacers' polymorphisms. *Ann. Appl. Biol.* **153**, 233–241 (2008).
12. Castro, P., Gil, J., Cabrera, A. & Moreno, R. Assessment of genetic diversity and phylogenetic relationships in Asparagus species related to *Asparagus officinali*. *Genet. Resour. Crop Evol.* **60**(2), 1275–1288 (2013).
13. Boonsom, T., Waranuch, N., Ingkaninan, K., Denduangboripant, J. & Sukrong, S. Molecular analysis of the genus Asparagus based on matK sequences and its application to identify *A. racemosus*, a medicinally phytoestrogenic species. *Fitoterapia* **83**(5), 947–953 (2012).
14. Ou, L. J. *et al.* Comparison of trnH-psbA sequences of some species in Asparagus. *Lishizhen Med. Mater. Med. Res.* **24**(11), 2678–2679 (2013).
15. Lee, Y., Kanno, A. & Kameya, T. Phylogenetic relationships in the genus Asparagus based on the restriction enzyme analysis of the chloroplast DNA. *Breed. Sci.* **47**, 375–378 (1997).
16. Chen, S. C., Kim, D. K., Chase, M. W. & Kim, J. H. Molecular analysis of the genus Asparagus based on matK sequences and its application to identify *A. racemosus*, a medicinally phytoestrogenic species. *PLoS One* **8**(3), e59472 (2013).
17. Wong, K. H. *et al.* Complete plastomes of *Asparagus aethiopicus* L., *A. densiflorus* (Kunth) Jessop 'Myers', and *A. cochinchinensis* (Lour.) Merr.: Comparative and phylogenetic analysis with congenerics. *PLoS One* **17**(4), e0266376 (2022).
18. Sheng, W. The entire plastome sequence of *Asparagus cochinchinensis* and genetic comparison to Asparagus species. *Open Life Sci.* **17**(1), 893–906 (2022).
19. Li, J. *et al.* Comparative plastid genomics of four Pilea (Urticaceae) species: Insight into interspecific plastid genome diversity in Pilea. *BMC Plant Biol.* **21**, 1–13 (2021).
20. Mcneal, J. *et al.* Using partial genomic fosmid libraries for sequencing complete organellar genomes. *Bio Techniques* **41**, 69–73 (2006).
21. Wicke, S. *et al.* The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **51**, 273–297 (2011).
22. Shaw, J. *et al.* The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* **92**, 142–166 (2005).
23. Shi, H. W. *et al.* Complete plastomes of two Siraitia Merrill species: Comparative analysis, positive selection and novel molecular marker development. *PLoS One* **14**, 12 (2019).
24. Zhang, G. *et al.* Identification of the original plants of cultivated Bupleuri Radix based on DNA barcoding and plastome analysis. *PeerJ* **10**, e13208 (2022).
25. Wang, R. J. *et al.* Dynamics and evolution of the inverted repeat-large single copy junctions in the plastomes of monocots. *BMC Evol. Biol.* **8**(1), 1–14 (2008).
26. Li, J. R. *et al.* Characterization of the complete plastome of *Asparagus setaceus*. *Mitochondrial DNA Part B*. **4**(2), 2639–2640 (2019).
27. Zhang, R. & Sheng, W. The complete plastome of *Asparagus densiflorus* (Kunth) Jessop L.. *Mitochondrial DNA Part B*. **7**(5), 722–724 (2022).
28. Wu, Q. *et al.* Comparative transcriptomic analysis unveils the deep phylogeny and secondary metabolite evolution of 116 Camellia plants. *Plant J.* **111**(2), 406–421 (2022).
29. Ikemura, T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* **2**(1), 13–34 (1985).
30. Tuller, T., Waldman, Y. Y., Kupiec, M. & Ruppin, E. Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci. USA* **107**, 3645–3650 (2010).
31. Mehmetoglu, E. *et al.* The complete chloroplast genome sequence of *Cicer echinospermum*, genome organization and comparison with related species. *Sci. Hortic.* **296**, 110912 (2022).
32. He, X. *et al.* The complete chloroplast genome of *Carpesium abrotanoides* L. (Asteraceae): Structural organization, comparative analysis, mutational hotspots and phylogenetic implications within the tribe Inuleae. *Biologia* **77**(7), 1861–1876 (2022).

33. Ding, R. *et al.* Analysis of codon usage in the chloroplast genome of *Cypripedium calceolus*, Linye Kexue Yanjiu. *For. Res.* **34**(5), 177–185 (2021).

34. Park, I. *et al.* The complete chloroplast genome sequence of *Aconitum coreanum* and *Aconitum carmichaelii* and comparative analysis with other Aconitum species. *PLoS One* **12**(9), e0184257 (2017).

35. Asaf, S. *et al.* Complete plastome of nicotiana otophora and its comparison with related species. *Front. Plant Sci.* **7**, 843 (2016).

36. Pauwels, M. *et al.* Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, *Arabidopsis helleri* (Brassicaceae). *New Phytol.* **193**, 916–928 (2012).

37. Kuang, D. Y., Wu, H., Wang, Y. L., Gao, L. M. & Zhang, S. Z. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): Implication for DNA barcoding and population genetics. *Genome* **54**(8), 663–673 (2011).

38. Julian, T. F. *et al.* What can we do with 1000 plastid genomes?. *Plant J.* **90**(4), 808–818 (2017).

39. Nie, X. *et al.* Complete plastome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PLoS One* **7**, e36869 (2012).

40. Li, R., Ma, P. F., Wen, J. & Yi, T. S. Complete sequencing of five *Araliaceae plastomes* and the phylogenetic implications. *PLoS One* **8**(10), e78568 (2013).

41. Fan, Z. F. & Ma, C. L. Comparative plastome and phylogenetic analyses of Chinese Polyspora. *Sci. Rep.* **12**, 15984 (2022).

42. Huang, H., Shi, C., Liu, Y., Mao, S. Y. & Gao, L. Z. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: Genome structure and phylogenetic relationships. *BMC Evol. Biol.* **14**, 151 (2014).

43. Khakhlova, O. & Bock, R. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* **46**(1), 85–94 (2006).

44. Sarkinen, T. & George, M. Predicting plastid marker variation: Can complete plastid genomes from closely related species help?. *PLoS One* **8**, e82266 (2013).

45. Korotkova, N., Nauheimer, L., Hasmik, T. V., Allgaier, M. & Borsch, T. Variability among the most rapidly evolving plastid genomic regions is lineage-specific: Implications of pairwise genome comparisons in Pyrus (Rosaceae) and other angiosperms for marker choice. *PLoS One* **9**, e112998 (2014).

46. Ge, Q. Y., Cai, Y., Wang, J. F. & Zhao, S. G. Complete genome analysis of mt DNA in Carp and Crucian. *Jiyinzuxue yu Yingyong Shengwuxue (Genom. Appl. Biol)* **39**(1), 37–43 (2020).

47. Wang, B., Gao, L., Su, Y. J. & Wang, T. Adaptive evolutionary analysis of chloroplast genes in euphyllophytes based on complete plastome sequences. *Acta Sci. Nat. Univ. Sunyatseni* **51**, 108–113 (2012).

48. Joët, T., Cournac, L., Horvath, E. M., Medgyesy, P. & Peltier, G. Increased sensitivity of photosynthesis to antimycin A induced by inactivation of the chloroplast ndhB gene. Evidence for a participation of the NADH-dehydrogenase complex to cyclic electron flow around photosystem I. *Plant Physiol.* **125**, 1919–1929 (2001).

49. Horváth, E. M. *et al.* Targeted inactivation of the plastid ndhB gene intobacco results in an enhanced sensitvity of photosynthesis to moderate stomatal closure. *Physiol. Plant.* **123**, 1337–1349 (2000).

50. Allahverdiyeva, Y., Mamedov, F., Mäenpää, P., Vass, I. & Aro, E. M. Modulation of photosynthetic electron transport in the absence of terminal electron acceptors: Characterization of the rbcL deletion mutant of tobacco. *Biochim. Biophys. Acta* **1709**, 69–83 (2005).

51. Tseng, C. C. *et al.* Editing of accD and ndhF chloroplast transcripts is partially affected in the Arabidopsis vanilla cream1 mutant. *Plant Mol. Biol.* **73**, 309–323 (2010).

52. Madoka, Y. *et al.* Chloroplast transformation with modified accD operon increases acetyl-CoA carboxylase and causes extension of leaf longevity and increase in seed yield in tobacco. *Plant Cell Physiol.* **43**, 1518–1525 (2002).

53. Lei, H., Zhu, S. J., Zhao, Y. & Duan, X. X. Sequencing and characteristic analysis of the complete plastome of *Trollius chinensis*. *Jiyinzuxue yu Yingyong Shengwuxue (Genom. Appl. Biol.)* **38**(8), 3595–3604 (2019).

54. Liu, J. B., Bao, Y. & Zheng, W. T. Analyses of some structural properties on a class of hierarchical scale-free networks. *Fractals* **30**(07), 2250136 (2022).

55. Sheng, W. The complete plastome of two traditional medical plants: *Asparagus cochinchinensis* (Lour.) Merr. and *Asparagus dauricus* Fisch. ex Link. *Mitochondrial DNA Part B* **7**(5), 725–726 (2022).

56. The Angiosperm Phylogeny Group. An update of the giosperm phylogeny group classification for the orders and families of flowering plants 2016: APGIV. *Bot. J. Linn. Soc.* **181**(1), 1–20 (2016).

57. CBOL Plant Working Group. A DNA barcode for land plants. *Proc. Natl. Acad. Sci. USA* **106**(31), 12794–12797 (2009).

58. Kane, N. C. & Cronk, Q. Botany without borders: Barcoding in focus. *Mol. Ecol.* **17**, 5175–5176 (2008).

59. Han, J. *et al.* An authenticity survey of herbal medicines from markets in China using DNA barcoding. *Sci. Rep.* **6**, 18723 (2016).

60. Li, Y. *et al.* Complete plastome of seven fritillaria species, variable dna markers identification and phylogenetic relationships within the genus. *PLoS One* **13**, e0194613 (2018).

61. Menezes, A. P. A. *et al.* Plastomes of byrsonima species (malpighiaceae): Comparative analysis and screening of high divergence sequences. *Sci. Rep.* **8**, 1 (2018).

62. Zhang, N. *et al.* An analysis of *Echinacea plastomes*: Implications for future botanical identification. *Sci. Rep.* **7**(1), 216 (2017).

63. Yang, J. *et al.* Development of chloroplast and nuclear dna markers for chinese oaks (quercus subgenus quercus) and assessment of their utility as dna barcodes. *Front. Plant Sci.* **8**, 816 (2017).

64. Piredda, R. *et al.* Prospects of barcoding the Italian wild dendroflora: Oaks reveal severe limitations to tracking species identity. *Mol. Ecol. Resour.* **11**(1), 72–83 (2011).

65. Ngai, H. L., Kong, B. L. H., Lau, D. T. W. & Shaw, P. C. Differentiation of Lingxiaohua and Yangjinhua by chloroplast genome sequencing and DNA barcoding markers. *Genome* **66**(2), 21–33 (2022).

66. Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**, 11–15 (1987).

67. Bolger, A. M. *et al.* Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **3**, 2114–2120 (2014).

68. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**(16), 2078–2079 (2009).

69. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**(2), giab008 (2021).

70. Jin, J. J. *et al.* GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**, 241 (2020).

71. Shi, L. C. *et al.* CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* **1**, 1–9 (2019).

72. Lohse, M., Drechsel, O. & Bock, R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **52**, 267–274 (2007).

73. Quax, T. E. *et al.* Codon bias as a means to fine-tune gene expression. *Mol. Cell.* **59**, 149–161 (2015).

74. Liu, H. *et al.* Analysis of codon usage in the plastome of broomcorn millet (*Panicum miliaceum* L.). *Plant Sci. J.* **35**, 362–371 (2017).

75. Wang, Z. *et al.* Comparative analysis of codon usage patterns in plastomes of six Euphorbiaceae species. *PeerJ* **8**, e8251 (2019).

76. Sharp, P. M. & Li, W. H. Codon usage in regulatory genes in Escherichia coli does not reflect selection for "rare" codons. *Nucl. Acid Res.* **14**, 7737–7749 (1986).

77. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European molecular biology open software suite. *Trends Genet.* **16**, 276–277 (2000).

78. Inkyu, P. *et al.* The complete chloroplast genomes of six ipomoea species and Indel Marker development for the discrimination of authentic pharbitidis semen. *Front. Plant Sci.* **9**, 1664–2462 (2018).

79. Kurtz, S. *et al.* REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **29**, 4633–4642 (2001).

80. Thiel, T., Michalek, W., Varshney, R. & Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **106**, 411–422 (2003).

81. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**(2), 573–580 (1999).

82. Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **32**, W273–W279 (2004).

83. Amiryousefi, A., Hyvönen, J. & Poczai, P. IRscope: An online program tovisualize the junction sites of plastomes. *Bioinformatics* **34**, 3030–3031 (2018).

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

85. Rozas, J. *et al.* DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).

86. Wang, D., Zhang, Y., Zhang, Z., Zhu, J. & Yu, J. KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom. Bioinform.* **8**(1), 77–80 (2010).

87. Zhang, R. *et al.* Exploration of plastid phylogenomic conflict yields new insights intothe deep relationships of Leguminosae. *Syst. Biol.* **69**, 613–622 (2020).

88. Zhang, Z. *et al.* ParaAT: A parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* **419**(4), 779–781 (2012).

89. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest2: More models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).

90. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**(8), 754–755 (2001).

91. Ronquist, F. & Huelsenbeck, J. P. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**(12), 1572–1574 (2003).

92. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

93. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

94. Zhang, C. Scalable Algorithms for Inference and Simulation under Complex Phylogenetic Models. https://escholarship.org/uc/item/59x9m4bw (2022).

95. Zhang, C. & Siavash, M. Weighting by gene tree uncertainty improves accuracy of quartet-based species trees. *Mol. Biol. Evol.* **39**(12), 2 (2022).

96. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).

Author contributions

Y.L.T., X.L., X.Y.Q. designed the experiments and drafted and revised the manuscript. Y.J.X., B.X.Y., analyzed the data. L.W., preparation of plant material and collection of samples. All the authors reviewed the manuscript.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-42945-x.

Correspondence and requests for materials should be addressed to X.L. or X.Q.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.