Complete plastid genome of *Angelica*, *Ostericum* and related species (Apiaceae): genomic evolution and phylogenetic implications

Qiu-Ping Jiang
Sichuan University - Wangjiang Campus: Sichuan University https://orcid.org/0000-0001-9776-3781
Megan Price
Sichuan University - Wangjiang Campus: Sichuan University
Xian-Lin Guo
Sichuan University - Wangjiang Campus: Sichuan University
Wei Gou
Sichuan University - Wangjiang Campus: Sichuan University
Song-Dong Zhou
Sichuan University - Wangjiang Campus: Sichuan University
Xing-Jin He (xjhe@scu.edu.cn)
Sichuan University - Wangjiang Campus: Sichuan University https://orcid.org/0000-0003-2064-0112

Research article

Keywords: Angelica, Apiaceae, comparative analysis, Ostericum, phylogenetic relationships, plastid genome

Posted Date: March 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-304635/v1

Abstract

Background

Subtribe Angelicinae is a large and taxonomically complex group of Apiaceae, encompassing *Angelica*, *Archangelica*, *Coelopleurum*, *Conioselinum*, *Czernaevia*, *Glehnia*, *Levisticum* and *Ostericum* that are distributed in the Northern Hemisphere, and whether this taxa is natural is debatable, especially between *Angelica* and *Ostericum*. To determine genommic evolution and phylogenetic relationships between *Angelica*, *Ostericum*, and related species, we newly assembled the complete plastid genome sequences of eight subtribe Angelicinae species and *Melanosciadium pimpinelloideum* using next-generation sequencing technology.

Results

The nine plastid genomes we sequenced were conserved, and their size ranged from 146765 bp to 164329 bp, showing the typical quadripartite circular structure with an overall GC content of 37.5–37.8%. IR boundary analyses showed that the genes in the LSC region transfer into the IR regions and the SSC region was relatively stable. Codon usage patterns were similar among these species and we identified 66–86 SSRs, with the most abundant SSR being mononucleotide. The Pi analyses showed that *petA-psbJ*(0.02778), *atpI-atpH*(0.17333) and *petA-psbJ*(0.04726) intergenic regions had the highest Pi values in *Angelica*, *Ostericum*, and ten species, respectively.

Conclusions

Ostericum exhibited significant differences in size of genomes, content of genes and tRNAs, GC content, some type of SSRs, and IR boundaries to *Angelica*, and phylogenetic analyses found the relatedness between *Angelica* and *Ostericum* is more distant in protein-coding genes of the plastid genomes trees and nrITS trees.

Background

Subtribe Angelicinae is a large and taxonomically complex group of the Apiaceae-Apioideae-tribe Peucedaneae (tribe Selineae), which is comprised of eight genera; *Angelica*, *Archangelica*, *Coelopleurum*, *Conioselinum*, *Czernaevia*, *Glehnia*, *Levisticum* and *Ostericum*. Subtribe Angelicinae is a large group about 120 species in the world, widely distributed in north temperate regions, among which more than 60 species are found in China [1–5]. In this taxon, most species are perennial herbs and have important economic and medicinal value, such as *A. sinensis*, *O. citriodorum*, *L. officinale*, *G. littoralis*, etc [6–11]. There have been many studies of genera in subtribe Angelicinae, but most have focused on chemical composition and pharmacognosy [12–16] or simple sequence repeat (SSR) and DNA markers [17, 18]. Only a few studies have undertaken a phylogenetic analysis of the Angelica subtribe, and previous studies focusing on phylogenetic analysis usually used nrDNA (nrITS and nrETS) and several cpDNA genes, such as rpl16 and rps16 [19, 20]. Molecular phylogenetic studies of subtribe Angelicinae based on nrITS suggested that *Conioselinum*, *Ostericum* and *Levisticum* have rather distant phylogenetic relationships with *Angelica*, *Archangelica*, *Coelopleurum*, *Czernaevia*, and *Glehnia* and *Melanosciadium* (Apioideae-Smyrnieae, but previous studies showed this genus is closely related to *Angelica*) has a closer relationship with *Angelica* [20–24]. Nevertheless, these studies have indicated that phylogenetic relationships within subtribe Angelicinae are complex and disputed. Given the worldwide geographical distribution of the more than 127 species of subtribe Angelicinae [1, 2], it is very difficult to comprehensively investigate the phylogenetic relationships without extensive sampling or cooperation with international herbariums. Therefore, molecular information can be used to identify the phylogenetic relationships between genera more accurately without considerable material collection.

Plastids are plant organelles responsible for photosynthesis, producing organic matter and storing energy [25]. Plastid genomes are very conserved, mainly in their genome structure, gene sequence and gene type. A typical plastid genome can be divided into four regions: two inverted repeats regions (IRs), a large single-copy region (LSC) and a small single-copy region (SSC) [26, 27]. In higher plants, plastid DNA is a circular molecule of double-stranded DNA and in multi-copy form, which ranges from 120 kb to 170 kb in length and usually encodes 120 to 130 genes. Plastid DNA has many features, such as monolepsis, small subfractions, multiple replications, and slow molecular evolution. Because of these specific features, the plastid DNA sequence has been widely applied in molecular phylogenetic studies [28–35]. With the development of sequencing technology, especially the widespread use of next-generation sequencing, comparative genomics has become commonly used for the research of evolution, phylogeny, genome structure and population genetics [33, 36–38].

To date, there are 11 complete plastid genomes of *Angelica*, *Ostericum* and related species (Table S1). Although 11 species plastid genomes have been published, there are many questions unanswered. Such as, are there variations in plastid genome structure between species within genera of subtribe Angelicinae, especially between the *Angelica* and *Ostericum*? Based on plastid genome data, what are the phylogenetic relationships among species of *Angelica* and *Ostericum* and related species? Consequently, in this study we sequenced *de novo* eight species of four genera in the subtribe Angelicinae and 1 related species (3 of *Angelica*, 3 of *Ostericum*, 1 of *Coelopleurum*, 1 of *Conioselinum* and 1 of

Melanosciadium) (Table 1). We conducted structural and comparative analyses of ten plastid genome sequences (nine *de novo* and downloaded *Glehnia littoralis* of subtribe Angelicinae), including repetitive sequences, SSRs, codon usages, IR contraction and expansion, and nucleotide sequence diversity to define the relationship of these species. Additionally, we used 34 plastid genomes and 44 internal transcribed spacers (ITSs) to reconstruct phylogenetic relationships of subtribe Angelicinae based on maximum likelihood (ML) and Bayesian inference (BI) methods.

Table 1
Summary of assembly data for the nine species plastid genome

Species	Genome size (bp)	IR (bp)	LSC (bp)	SSC (bp)	Total number of genes	rRNA	tRNA	Protein–coding genes	A%	G%	C%	T%	G C%
Angelica amurensis	146931	18086(*2)	93201	17558	130	8	36	86	30.8	18.4	19.1	31.7	37.5
Angelica biserrata	146765	17982	93297	17504	130	8	36	86	30.8	18.4	19.1	31.6	37.5
Angelica tianmuensis	147308	18217	93238	17636	130	8	36	86	30.8	18.4	19.1	31.6	37.5
Coelopleurum saxatile	147032	17797	93901	17537	130	8	36	86	30.8	18.4	19.1	31.7	37.5
Conioselinum chinense	148653	18467	94111	17608	130	8	36	86	30.8	18.4	19.2	31.6	37.6
Melanosciadium pimpinelloideum	164329	35157	76450	17565	144	8	37	99	31.1	18.5	19.0	31.4	37.5
Ostericum grosseserratum	160489	26270	90517	17432	134	8	37	89	30.9	18.5	19.2	31.4	37.7
Ostericum huadongense	161095	26322	91015	17436	134	8	37	89	30.9	18.5	19.2	31.4	37.8
Ostericum sieboldii	156550	26033	86959	17525	134	8	37	89	31.0	18.5	19.1	31.5	37.6

Results

The plastid genome features of these species

The complete plastid genome of *Angelica*, *Ostericum* and related species was a single and typical quadripartite circular structure (Figure 1). The length of the nine species' plastid genomes ranged from 146765 bp in *Angelica biserrata* to 164329 bp in *Melanosciadium pimpinelloideum*. The typical complete plastid genome quadripartite structure of these species consisted of two identical IR regions ranged from 17797 in *Coelopleurum saxatile* to 35157 bp in *M. pimpinelloideum*, small single-copy regions (SSC) ranged from 17432 in *Ostericum grosseserratum* to 17636 bp in *A. tianmuensis*, and large single-copy regions (LSC) ranged from 76450 in *M. pimpinelloideum* to 93901 bp in *Coe. saxatile*. The genomes total A content was between 30.8–31.1%, the G content was between 18.4–18.5%, the C content was between 19.0–19.2%, the T content was between 31.6–31.7% and the total GC content was between 37.5%–37.8%. The plastid genomes contained 130–143 genes, including 86–98 protein-coding genes (PCGs), 36–37 transfer RNA genes (tRNAs) and 8 ribosomal RNA genes (rRNAs) (Table 1).

Melanosciadium pimpinelloideum had the most genes (143 genes), with 14 more genes than *Angelica*, *Coe. saxatile* and *Conioselinum chinense* (*petD*, *rpoA*, *rps11*, *rpl36*, *infA*, *rps8*, *rpl14*, *rps3*, *rpl22*, *rps19*, *rpl2*, *rpl23*, *ycf2*, *trnI-CAU*), and contained 10 more genes than *Ostericum* (*petD*, *rpoA*, *rps11*, *rpl36*, *infA*, *rps8*, *rpl14*, *rps3*, *rpl22*, *rps19*). While *Ostericum* contained 4 more genes (*rpl2*, *rpl23*, *ycf2*, *trnI-CAU*) than *Angelica*, *Coe. saxatile* and *Con. chinense*. Moreover, *M. pimpinelloideum* had the most protein-coding genes (99 protein-coding genes), with 13 more protein-coding genes than *Angelica*, *Coe. saxatile* and *Con. chinense* (*infA*, *petD*, *rpl2*, *rpl14*, *rpl22*, *rpl23*, *rpl36*, *rpoA*, *rps3*, *rps8*, *rps11*, *rps19*, *ycf2*), and had 10 more protein-coding genes than *Ostericum* (*infA*, *petD*, *rpl14*, *rpl22*, *rpl36*, *rpoA*, *rps3*, *rps8*, *rps11*, *rps19*). Meanwhile, *Ostericum* contained 3 more protein-coding genes (*rpl2*, *rpl23*, *ycf2*, *trnI-CAU*) than *Angelica*, *Coe. saxatile* and *Con. chinense*. *M. pimpinelloideum* and *Ostericum* had the same number of tRNAs (37 tRNAs), and they had one *trnI-CAU* more than the other species. Referring to previous research, we considered *infA*, *ycf15*, *ycf68*, and the short copy of *ψycf1* and *ψrps19* as pseudogenes [34, 39, 40]. To distinguish pseudogene *ycf1* from the functional *ycf1* gene, we considered the pseudogene as *ψycf1* in this study, and the *ψrps19* is the same case.

Analysis of inverted repeat contraction and expansion

To assess the *Angelica*, *Ostericum* and related species' (including *Glehnia littoralis*) expansion and contraction of the IR regions, we identified and focused on the junctions of IR/LSC and IR/SSC (Figure 2). Gene *ycf2* of 277–766 bp in the junction of the LSC and the IRb region was located in the IRb region in *Angelica sp.*, *Coelopleurum saxatile*, *Conioselinum chinense*, and *G. littoralis*. In *Melanosciadium pimpinelloideum* the junction of the IRb and LSC coincided with the *petB* gene, with 118 bp in the LSC region and 1293 bp located in the IRb region. In *Ostericum sp.* the *rps19* gene extended 81 bp into the IRb region. The *ψycf1* gene was located in the junction of the SSC and IRb region in all ten species. Similarly, the *ndhF* gene was located in the SSC region 25–126 bp away from the IRb/SSC border except for *G. littoralis*. In *G. littoralis*, the *ndhF* gene extended 9 bp into the IRb region at the border of IRb/SSC. The *ycf1* gene crossed the SSC and IRa border in all ten species, with 3439–3592 bp in the SSC region and 1856–1975 bp located in IRa region. In all species, the *trnH* gene was closest to the IRa/LSC border and was located in the LSC region being 3–1155 bp away from the IRa region. However, there were differences in IRa region away from the IRa/LSC border. The *trnL* gene was in the IRa region in *Angelica sp.*, *Coe. saxatile*, *Con. chinense*, and *G. littoralis*, but was closest to the IRa/LSC border being 619–1329 bp away. Whereas the *petD* gene was closest to the IRa/LSC border and in the IRa region in *M. pimpinelloideum,* being 1472 bp away from the border. In *Ostericum sp.*, it was the *ψrps19* gene that was closest to the IRa/LSC border and located in the IRa region with 81 bp.

Analysis of codon usage and amino acids frequency

The 20 amino acids were encoded by 64 codons in the ten complete plastid genomes (i.e. *Angelica amurensis*, *Angelica biserrata*, *Angelica tianmuensis*, *Coelopleurum saxatile*, *Conioselinum chinense*, *Melanosciadium pimpinelloideum*, *Glehnia littoralis*, *Ostericum grosseserratum*, *Ostericum huadongense*, *Ostericum sieboldii*) (Figure 3). Methionine (Met) and tryptophan (Trp) only had the minimum type of codons with one codon, while leucine (Leu) and serine (Ser) had the maximum type of codons with six. The total number of codons of these species ranged from 22489 in *G. littoralis* to 26569 in *M. pimpinelloideum*. Among the amino acids, Leu had the maximum codon numbers ranging between 2399–2826, and cysteine (Cys) had the minimum codon numbers ranging from 240–285. The most used codon was AUU ranging from 927 to 10744, with the second being AAA ranging from 905 to 1105. The three least used were the termination codons UAA, UGA and UAG, with UGA using the least and ranging from 12 to14, while UAA was the most used and ranged from 33–43. Excluding the termination codons, the least used codon was UGC and ranged from 58–60. The codon with the highest RSCU value was UUA, then AGA, and GCU (deep red). The codon with the lowest RSCU value was AGC, followed by (in order) CGC, CUG, GAC, UAC, and CUC (deep blue) (Figure 3, Table S2). Furthermore, RSCU values of third position codons A or U were mostly greater than 1, whereas the third position codons C or G were mostly less than 1 (Table S2).

Analyses of repeats sequences and single sequence repeats (SSRs)

We detected forward, palindromic, reverse, and complementary repeats in these ten species. After filtering out duplicate data, we found that forward and palindromic repeats were typical, while reverse and complementary repeats were rare. Most repeat sequence lengths were 30–45 bp (Figure 4). Among these species, *Conioselinum chinense* had the smallest total number of repeats with 24, whereas *Melanosciadium pimpinelloideum* and *Ostericum sieboldii* shared the largest number of repeats 40.

We detected six types of SSRs (mononucleotide, dinucleotide, trinucleotide tetranucleotide, pentanucleotide and hexanucleotide) in these ten species. The SSR analysis showed *Ostericum huadongense* had the smallest number of SSRs with 66, and *Angelica biserrate* had the largest with 86. The number of mononucleotide SSRs was the largest, followed by dinucleotide, then tetranucleotide rather than trinucleotide. Pentanucleotide and hexanucleotide repeats were very rare, especially hexanucleotide. Only three species (*A. biserrate*, *Con. chinense* and *O. sieboldii*) had hexanucleotide SSRs (Figure 4). This is consistent with other Apiaceae species [34, 37, 40, 41] and *Allium* [35, 36, 42], but dinucleotide repeats are most numerous in *Forthysia* [43], and trinucleotide repeats are most abundant in *Nitotiana* [44].

Analysis of sequence diversity nucleotide

Nucleotide diversity (Pi) of the ten plastid genomes was calculated to estimate the sequence divergence level of different regions. Among all species, the SSC and LSC regions exhibited high divergence levels and IR regions exhibited low divergence (Table S3. Figure 5). The results indicated that the IR region was relatively conserved. The sequences with high Pi values were predominantly in intergenic spacers in all species. There were two exceptions, gene regions of *ycf2* and *ycf1*, which were in the boundaries of IR/SSC and IR/LSC.

In addition to the top ten Pi values, we found several more important peaks in Figure 5. Figure5A illustrates the *Angelica sp.* nucleotide diversity (Pi) level. We found that the regions with comparatively higher Pi values were the intergenic regions including *psbZ-trnG (GCC)-trnfM(CAU)-rps14*, *rps4-trnL(UAA)*, *atpB-rbcL*, *petA-psbJ-psbL-psbF-psbE*, *ψycf1-ndhF*, *ndhF-rpl32-trnL(UAG)* and additional gene region *ycf2*. The *petA-psbJ* showed the highest variability, with a Pi value of 0.02778. Figure5B, which shows the nucleotide diversity (Pi) level of *Ostericum*, demonstrates that the intergenic regions had higher values including *atpI-atpH*, *trnV (UAC)-ndhC*, *petA-psbJ-psbL-psbF*, *psbH-petB*, *infA-rpl36-rps11*, *ψycf1-ndhF* and gene region of *ycf1*. The *atpI-atpH* showed the highest variability, with a Pi value of 0.17333. Figure5C illustrates the nucleotide diversity (Pi) of the ten subtribe Angelicinae species. Of note in Figure 5C are the intergenic regions including *trnK(UUU)-rps16-trnQ(UUC)*, *atpI-atpH*, *trnC(CGA)-rpoB*, *trnE(UUC)-trnT(GGU)*, *petA-psbJ-psbL*, *rps4-trnT(UGU)-trnL(UAA)*, *ndhF-rpl32-trnL(UAG)-ccsA*, an gene regions of *ycf1* and *ycf2*. In these regions, *petA-psbJ* had the highest variability, with a Pi value of 0.04726, then *rpl32-trnL* with a similarly high Pi value of 0.04693. There were

several higher diversity regions that were different in *Angelica* and *Ostericum sp.*, such as the intergenic region *atpI-atpH* in *Ostericum*, and the intergenic region *atpB-rbcL* in *Angelica*.

Phylogenetic analysis

We used 39 complete plastid genomes and 44 nrITS sequences to construct ML and BI trees to investigate the phylogenetic relationships in subtribe Angelicinae (Figure 6). The plastid genome and ITS trees produced incongruent topology trees. The plastid tree indicated that *Angelica* is monophyletic but subtribe Angelicinae is non-monophyletic. *Angelicasp*. were closely related to *Coelopleurum saxatile* and *Melanosciadium pimpinelloideum*. *Ostericum sp.* were closely related to *Pternopetalum*, and belonged to the *Acronrma* Clade [21]. *Conioselinum chinense* was in the *Sinodielsia* Clade [21]. In addition, *Angelica sinensis* was closely related to *Con. chinense* and also belonged to the *Sinodielsia* Clade [21]. In particular, *Glehnia littoralis* was embedded in *Angelica* (Figure6B, 6D, BS=69, PP=1).

However, these species and genera were in parallel branches in the nrITS tree, including *Melanosciadium*, *Coelopleurum*, *Glehnia*, *Peucedanum*, *Angelica acutiloba* and *Angelica*. The parallel branches and low support in BI analysis (Figure6A, 6C, BS=100, PP=0.5591) indicated that nrITS provided insufficient information to determine phylogenetic relationships and more nuclear gene and cpDNA sequences are needed to increase certainty in models. The ITS tree results from ML and BI analysis were a little different regarding the placement of *Peucedanum japonicum* where it was not one of the parallel branches of *Angelica* in BI analysis, but in ML analysis is one of the parallel branches of *Angelica*. Nevertheless, all trees indicated that the subtribe Angelicinae is non-monophyletic, and relatedness between *Angelica* and *Ostericum* is more distant.

Discussion

Plastid genome evolution

Plastid genomes are highly conserved in genome structure, GC content, gene numbers and gene order [29, 45]. Previous studies have confirmed that genome rearrangement, gene pseudogenization or deletion, expansion and contraction of IRs have occurred during the evolutionary process of other plants [46-49]. In this study, the highly conserved nature of the plastic genome does not mean that they are alike, there are important differences and diversity between plastid genomes. We found that the longest genome was 164329 bp in *Melanosciadium pimpinelloideum*, followed by *Ostericum sp.* with lengths around 160000 bp, while the other species (*Angelica amurensis*, *Angelica biserrata*, *Angelica tianmuensis*, *Coelopleurum saxatile* and *Conioselinum chinense*) had genomes with lengths about 148000 bp. The order of most to least number of genes and protein-coding genes was the same as length, with *M. pimpinelloideum* having the most, then *Ostericum sp.,* followed by all other species. The tRNA genes were divided into two groups, *M. pimpinelloideum* and *Ostericum sp.* had 37 tRNAs, the other species had 36 tRNAs. We compared our sequenced *O. grosseserratum* against downloaded *O. grosseserratum* (Genebank NO.: KT852844), where we compared their size, gene number, GC content, and tRNAs (Table S4). We conclude that *O. grosseserratum* (Genebank NO.: KT852844) is probably an identification error.

Codon usage bias is a crucial indicator for studying the evolution of genomes [50]. Codon usage bias may be caused by nucleotide mutation, random genetic drift, or the affection of translation efficiency [50–52]. In addition, gene sequence length, gene expression level, GC distribution position and tRNA abundance can affect the preference of synonymous codon usage [52–56]. We found that the genes of studied species preferred the codons ending with A or U, and this is a universal phenomenon in the plastid genome of higher plants [54, 57, 58].

Plastid genomes contain repeats (forward, palindromic, reverse, and complementary) and in most studies of angiosperm plastids, the authors reported repeats ≥ 30 bp [40, 59–61]. Repeat sequences are important in the identification of mutational hotspots and play an important role in genome rearrangement [62, 63]. Hence, these repeats are used extensively for population genetics and biogeography studies [37, 40, 41]. Overall, we found that forward and palindromic repeats were more common in our studied species and reverse and complementary repeats were rare, indicating these species tend to generate the forward and palindromic repeats rather than the reverse and complement repeats (Figure 4b).

Simple sequence repeats (SSRs) are valuable markers to detect variability within the same species and have been widely used in plant population genetics and evolutionary studies [40, 61, 64]. The most abundant SSRs were mononucleotide in our studied species, followed by (in decreasing abundance) dinucleotide, tetranucleotide, trinucleotide, pentanucleotide, and hexanucleotide repeats. This phenomenon is common in *Allium* [35, 36, 42] and Apiaceae [34, 37, 40, 41]. In our species, mononucleotide and dinucleotide compositions were similar, but trinucleotide, tetranucleotide and pentanucleotide were different across clades. For example, *Ostericum* was significantly different from *Angelica*, such as in the type of Trinucleotide, in Angelica amost is ATT and ATA, but in Ostericum is TAT, AAT and TTC (Table S3). All dinucleotide repeats were AT/TA, while other SSRs had motifs that were mostly A/T, causing overall AT richness in the studied plastid genomes [37, 65].

Contraction and expansion of IR regions are important for genome size variations and play a crucial role in evolution [66, 67]. The IR regions of our studied species can be divided into 3 groups by length. The first group has IR regions ranging from 17797–18467 bp and contains *Angelica*, *Co. saxatile*, *Con. chinense* and *G. littoralis*. The second group only contains *M. pimpinelloideum* because its IR regions were the largest with 35157 bp. The *Ostericum* species are grouped together with IR region lengths ranging from 26033–26322 bp. Through the length of LSC regions,

SSC regions and IR regions, we can conclude that size of IR regions is the key to affecting the size of genomes, the IR regions larger, the size of genomes larger. There are several models that have been proposed to explain the influence of IR regions on genome size. By examination of IR/LSC junctions in 13 *Nicotiana* species, Goulding *et al.* proposed a stepwise model involving a single-strand break, heteroduplex formation via a Holliday junction, and then small IR expansions via gene conversion [68]. In the study of *N. acuminata*, Wang *et al*. proposed a different model of double-strand break (DSB) followed by strand invasion and recombination that may cause the expansion and contraction of IR [69]. When studying *Pelargonium*, Zhu *et al.* proposed a model involving multiple inversions promoted by the dispersed repeats, expansions and contractions with several rounds of ebb-and-flow [46].

In this study there was no obvious change in SSC/IR borders, but we did detect an obvious shift of LSC/IR borders, indicating that these IR regions were undergoing a contraction and expansion. In Apiaceae the SSC region almost the same length, similar results were observed in the previously reported *Chamaesium*, *Ligusticum* and *Bupleurum* [34, 37, 40]. However, the size of the LSC region in our studied species was similar, ranging from 76450–94111 bp. We observed the transfer of LSC sequences and genes into the IR regions or the IR regions contracted in the IR/LSC border. We observed *Ostericum* had more had more *ycf2*, *trnI-CAU*, *rpl23*, *rpl2* in IRb and Ira regions than *Angelica*, and part *rps19* gene in IRb region in *Ostericum*. Compared to Angelica, *M. pimpinelloideum* had more *ycf2*, *trnI-CAU*, *rpl23*, *rpl2*, *rps19*, *rpl22*, *rps3*, *rpl16*, *rpl14*, *rps8*, *infA*, *rpl36*, *rps11*, *rpoA*, *petD*, *petB* in IRb region, and more *ycf2*, *trnI-CAU*, *rpl23*, *rpl2*, *rps19*, *rpl22*, *rps3*, *rpl14*, *rps8*, *infA*, *rpl36*, *rps11*, *rpoA*, *petD* in IRa region. This result is different from *Plantago* where genes are transferred from the SSC region into the IR region [60, 70]. During land plant evolution, the IR expansion or contraction that have transfered genes from the SC regions into the IR or vice versa is generally [44, 70, 71]. The IR contraction and expansion has been considered important evolutionary phenomena that resulted in the origin of pseudogenes, gene duplication, or the reduction of duplicate genes to single copy [61, 72–74].

Phylogenetic relationship analysis

Until now molecular phylogenetic studies based on nrITS and a few cpDNA sequences did not support the monophyly of subtribe Angelicinae [21, 23, 24]. In this study, we performed phylogenetic analyses for subtribe Angelicinae using protein-coding genes of the plastid genomes and ITS sequences, and we conclude that subtribe Angelicinae is non-monophyletic. *Angelica* is closely related to *Coelopleurum saxatile*, *Melanosciadium pimpinelloideum*, and *Angelica acutiloba* belongs to the tribe Peucedaneae (tribe Selineae) [21]. Regardless of which tree, ITS or cpDNA, *A. acutiloba* is one of the parallel branches of *Angelica* and *A. acutiloba* may not belong to *Angelica*. The specific phylogenetic position of *A. acutiloba* needs more research to ascertain its correct position. *Ostericum* is closely related to *Pternopetalum* and it belongs to the *Acronrma* Clade [21]. We found that *Pterygopleurum neurophyllum* was nested within *Ostericum* and there are three possible reasons for this unexpected result. The first is that *P. neurophyllum* should belong to *Ostericum* and that the plastid tree suggests that *P. neurophyllum* clusters with *Ostericum* forming a monophyletic taxon. The second is that the downloaded ITS (GeneBank No.: AY509127) data is from a species misidentified as *P. neurophyllum* and since there is only one sequence in GeneBank this cannot be compared against other sequences of the species. The third is that it was caused by the conflict between nuclear and plastid, like *Glehnia littoralis* and *Angelica sp.*. Nevertheless, more data and sequences of field collected materials is needed to solve this unexpected result.

The plastid genome and ITS trees produced incongruent tree topologies. The main difference is in *Angelica* and related species, especially the phylogenetic position of *G. littoralis*. In the cpDNA tree, *G. littoralis* was embedded in *Angelica*, but in the nrITS tree *G. littoralis* forms a parallel branch of *Angelica*. The organellar relationships and nuclear relationships discord is common in angiosperms, such as in Rosidae [75], Temperate Bamboos [76], and *Heuchera* [77]. Given that nuclear and plastid genomes have distinct evolutionary histories, organellar genomes have experienced unique population genetic pressures. Several factors can drive such conflict, including incomplete lineage sorting (ILS), hybridization and gene flow [76–78]. Together, our results suggested that subtribe Angelicinae species, and possibly Apiaceae, may have experienced a complex evolutionary history and speciation process.

Conclusion

In this study, we used the next-generation sequencing to sequence and annotate the complete plastid genomes of *Angelica amurensis*, *Angelica biserrata*, *Angelica tianmuensis*, *Coelopleurum saxatile*, *Conioselinum chinense*, *Melanosciadium pimpinelloideum*, *Ostericum grosseserratum*, *Ostericum huadongense*, and *Ostericum sieboldii*. The structure and organization of these plastid genomes are similar to each other and previously reported genomes from Apiaceae. We compared these nine genomes to another Angelicinae species, *Glehnia littoralis*, for repetitive sequences, SSRs, codon usages, IR contraction and expansion, and nucleotide sequence diversity. We found that *Ostericum* exhibited striking differences in size of genomes, content of genes and tRNAs, GC content, some type of SSRs, and IR boundaries to *Angelica*, and confirmed that the relatedness between *Angelica* and *Ostericum* is more distant in protein-coding genes of the plastid genomes trees. And the cpDNA tree was inconsistent with the ITS tree and this may have been caused by incomplete lineage sorting (ILS), hybridization and gene flow. The current plastid genomic dataset and the detailed analysis of subtribe Angelicinae species provide abundant genetic resources for the future molecular phylogeny, evolution and population genetic studies of subtribe Angelicinae and Apiaceae.

Methods

DNA sequencing, assembly, and annotation

Fresh green leaves were collected from adult plants of nine species from the field, and then they were desiccated and stored in silica gel. Qiu-Ping Jiang underwork the formal identification of the plant materials in this study. The herbarium specimens of these species were stored in the Herbarium, College of Life Sciences, Sichuan University (SZ). Specimen voucher details can be found in the table S1. Permission is not required to sample these plants because they are not key protected plants.

We extracted the total genomic DNA from the stored dry leaves, using the modified CTAB method [79]. The nine species in this study were *Angelica amurensis*, *A. biserrata*, *A. tianmuensis*, *Coelopleurum saxatile*, *Conioselinum chinense*, *Melanosciadium pimpinelloideum*, *Ostericum grosseserratum*, *O. huadongense*, *O. sieboldii*. Voucher specimens of these nine species were deposited in the herbarium of Sichuan University (SZ) (Table S1). The Illumina Novaseq 6000 platform (Illumina, San Diego, CA, USA) at Novogene (Beijing, China) was used to sequence the resultant DNA with Novaseq 150 sequencing strategy. The remaining clean data were assembled using NOVOPlasty 2.7.1 [80] with the default K-mer value 39, and *rbcL* of *A. biserrata* (GenBank accession No.: JN704956.1) was used as seed input for *A. biserrata*, *rbcL* of *A. sylvestris* (GenBank accession No.: DQ133798.1) was used as seed input for *A. amurensis* and *A. tianmuensis*, *rbcL* of *Coe. maritimum* (GenBank accession No.: KX527043.1) was used as seed input for *Coe. saxatile*, *rbcL* of *M. pimpinelloideum* (GenBank accession No.: KX527530.1) was used as seed input for *M. pimpinelloideum*, *rbcL* of *Con. chinense* (GenBank accession No.: MG224187.1) was used as seed input for *Con. chinense*, *rbcL* of *O. sieboldii* (GenBank accession No.: D44579.1) was used as seed input for *O. grosseserratum*, *O. huadongense* and *O. sieboldii*. Preliminary genome annotation was conducted using PGA [81], manual modifications for uncertain genes, and uncertain start and stop codons were corrected based on comparison with other affinis' plastid genomes using GENEIOUS R11 [82]. The nine species' annotated genome sequences were submitted to GenBank, and their corresponding accession numbers are listed in (Table S1). We also downloaded *G. littoralis* from NCBI, GenBank accession No.: MH142518, to enhance genome comparative analyses.

Genome comparative analyses of codon usage

Circular gene maps of the annotated genomes were drawn by the online program OGDRAW[83]. Genome features were compared between the nine species of subtribe Angelicinae and one related species based on the program GENEIOUS R11. We removed protein-coding genes (PCGs) less than 300 bp from the ten species' genomes, leaving protein-coding genes (PCGs) (>300 bp) that were analysed for codon usage bias in the program codon W [84]. We also calculated the relative synonymous codon usage (RSCU) [85] of the ten species.

Characteristics of cpSSRs and repetitive sequences

The plastid simple sequence repeats (cpSSRs) of the 10 species were generated using Perl script MISA[86] with the same settings: 10 repeats for mononucleotide, 5 repeats for dinucleotide, 4 repeats for trinucleotide and 3 repeats for tetranucleotide, pentanucleotide and hexanucleotide. Furthermore, the repetitive sequences (i.e. complementary, forward, palindromic, and reverse repeats) of plastid genomes were identified using REPuter [87]. The parameters were set as follows: (1) minimum repeat size of sequence was 30 bp; (2) sequence identity was more than 90% between two repeats; and (3) Hamming distance was equal to 3. Manual modifications of the data included removal of all overlapping repeat sequences.

Sequence diversity analysis

The alignment of the ten species' plastid genomes were visualized using MAFFT v7.402 [88] and calibrated manually. The single nucleotide polymorphism (SNP) analysis was generated using DnaSP v5 [89]. The parameters were set as follows: the windows length was 600 bp, and step size was 200 bp.

Phylogenetic analysis

To infer phylogenetic relationships within subtribe Angelicinae, the nine plastid genomes (nine *do novo* sequenced including one non-subtribe) were compared to another 25 Apiaceae Apioideae plastid genomes, with *Bupleurum* and *Chamaesium* as the outgroup. All plastid genome sequences were aligned using MAFFT v7.402, and we then extracted protein-coding genes. Maximum likelihood (ML) analyses was undertaken using RAxML v7.2.8 [90] and the best-fit model with GTR+G model and 1000 bootstrap replicates. Bayesian Inference (BI) analyses were performed in MrBayes version 3.2 [91], the best-fit model for the sequence (GTR+G+I) was conducted using Modeltest 3.7 [92] with optimized parameters. Four simultaneous runs were performed using Markov chain Monte Carlo (MCMC) simulations for 10 million generations, starting from a random tree and sampling one tree every 1000 generations. The first 20% of obtained trees were discarded as burn-in and the remaining were used to calculate a 50% majority-rule consensus topology and posterior probability (PP) values.

Declarations

Author Contributions

Q-PJ and X-JH conceived and designed the work. Q-PJ, X-LG, and WG analyzed the sequence data. Q-PJ wrote the manuscript. MP, S-DZ and X-JH revised the manuscript. All authors gave final approval of the paper

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 32070221, 31872647), National Specimen Information Infrastructure, Educational Specimen Sub-Platform (Grant No. 2005DKA21403-JK), the fourth national survey of traditional Chinese medicine resources (Grant No. 2019PC002). The funding agencies had no role in the design of the experiment, analysis, and interpretation of data and in writing the manuscript.

Competing Interests

The authors declare no conflict of interest.

Availability of data and materials

The nine annotated plastid sequences and nrITS sequences have been submitted to NCBI (https://www.ncbi.nlm.nih.gov) with accession numbers can be found in Table S1. Information for phylogenetic analysis download from Genbank can be found in Table S1.

Ethics approval and consent to participate

Not applicable

Acknowledgements

Not applicable

References

1. Pan ZH, Waston MF: **Flora of China**, vol. 14. Beijing, China: Science Press; 2005.
2. Shan RH, Yuan CQ: **Flora Reipublicae Popularis Sinicae**, vol. 55. Beijing, China: Science Press; 1992.
3. Shan RH: **New taxa of *Angelica* Linn. from China**. *Acta Phytotaxonomica Sinica* 1995, **33**(5):476–483.
4. Pimenov MG, Kljuykov EV: **Notes on some Sino-Himalayan species of *Angelica* and *Ostericum* (Umbelliferae)**. *Willdenowia* 2003, **33**(1):121–137.
5. Pan ZH, Li XH: **The study on one new spcies of *Ostericum* Hoffman in China**. *Journal of Plant Resources and Environment* 1996, **5**(2):45–49.
6. Xie L, Yang LH, Li XH: **Advances in pharmacological action of *Angelica sinensis***. *Research on Traditional Chinese Medicine* 2000, **16**(6):56–58.
7. Li XB, Qi YD, Zhang J, Zhu YX, Zhang Z, Zhang BG: **Advances in the Research of Chemical constituents and Bioactivities of *Conioselinum***. In: *The 10th Nabonal Symposium on natural Medicinal Resources Proceedings and Abstracts: 2012* 2012: 552–555.
8. Qiao YJ, Fang Y **The precious medicinal plant of "Bei Sha Shen"**. *Life World* 2003:21.
9. Tian WY, Lan F, Li SP, Luo JY: **The preliminary study on the pharmacology of *Ostericum citriodorum***. *Chinese Pharmacological Bulletin* 1989, **5**(4):249.
10. Liu K, Wang NG, Chen LS: **Research Brief of *Levisticum officinale***. *Chinese Pharmaceutical Journal* 1979, **14**(8):373–375.
11. Li XJ, Yuan GX: **Resources and Identification of medicinal products "DuHuo"**. *Primary Journal of Chinese Materia Medica* 1998, **12**(2):5–6.
12. Chu SS, Liu QZ, Du SS, Liu ZL: **Chemical Composition and Insecticidal Activity of the Essential Oil of the Aerial Parts of *Ostericum grosseserratum* (Maxim) Kitag (Umbelliferae)**. *Trop J Pharm Res* 2013, **12**(1):99–103.
13. Forycka A, Buchwald W: **Variability of composition of essential oil and coumarin compounds of *Angelica archangelica* L.**. *Herba Pol* 2019, **65**(4):62–75.
14. Kerrola K, Galambosi B, Kallio H: **Characterization of volatile composition and odor of *Angelica* (*Angelica archangelica* subsp. *archangelica* L.) root extracts**. 1994, **42**(9):1979–1988.
15. Wedge D, Gao Z, Tabanca N, Demirci B, Baser K, Pridgeon J, Becnel J, Sampson B, Werle C: **Chemical Composition and Biological Activities of Two Angelica Essential Oils from China**. In: *8th Annual Oxford International Conference on the Science of Botanicals* 2009, **75**(04):422.

16. Yang XB, Zhao Y, Zhou SY, Liu L, Mei QB: **Analysis of monosaccharide composition in *Angelica polysaccharides* by precolumn derivatization high performance liquid chromatography**. *ANALYTICAL SCIENCES* 2005, **33**(9):1287–1290.
17. Gil J, Um Y, Kim S, Kim OT, Koo SC, Reddy CS, Kim SC, Hong CP, Park SG, Kim HB *et al*: **Development of Genome-Wide SSR Markers from *Angelica gigas* Nakai Using Next Generation Sequencing**. *Genes* 2017, **8**(10):238.
18. Park I, Yang S, Kim WJ, Song JH, Lee HS, Lee HO, Lee JH, Ahn SN, Moon BC: **Sequencing and Comparative Analysis of the Chloroplast Genome of *Angelica polymorpha* and the Development of a Novel Indel Marker for Species Identification**. *Molecules* 2019, **24**(6):1038.
19. Liao C, Downie SR, Li Q, Yu Y, He X, Zhou B: **New Insights into the Phylogeny of *Angelica* and its Allies (Apiaceae) with Emphasis on East Asian Species, Inferred from nrDNA, cpDNA, and Morphological Evidence**. *Systematic Botany* 2013, **38**(1):266–281.
20. Zhou J, Gong X, Downie SR, Peng H: **Towards a more robust molecular phylogeny of Chinese Apiaceae subfamily Apioideae: additional evidence from nrDNA ITS and cpDNA intron (rpl16 and rps16) sequences**. *Mol Phylogenet Evol* 2009, **53**(1):56–68.
21. Downie SR, Spalik K, Katz-Downie DS, Reduron J-P: **Major clades within Apiaceae subfamily Apioideae as inferred by phylogenetic analysis of nrDNA ITS sequences**. *Plant Diversity and Evolution* 2010, **128**(1):111–136.
22. Zhou J, Downie SR, Liu WZ, Gong X: **A molecular phylogeny of Chinese Apiaceae subfamily Apioideae inferred from nuclear ribosomal DNA internal transcribed spacer sequences**. *TAXON* 2008, **57**(2):402–416.
23. Feng T, Downie SR, Yu Y, Zhang X, Chen W, He X, Liu S: **Molecular systematics of *Angelica* and allied genera (Apiaceae) from the Hengduan Mountains of China based on nrDNA ITS sequences: phylogenetic affinities and biogeographic implications**. *J Plant Res* 2009, **122**(4):403–414.
24. Liao C-Y, He X-J: **Angelica dabashanensis(Apiaceae), a new species from Shaanxi, China**. *Ann. Bot. Fennici* 2012, **49**:1–9.
25. Douglas SE: **Plastid evolution: origins, diversity, trends**. *Current Opinion in Genetics and Development* 1998, **8**:655–661.
26. Raubeson LA, Jansen RK: **Chloroplast genomes of plants, Plant diversity and evolution: genotypic and phenotypic variation in higher plants**. UK:CABI publishing; 2005.
27. Jansen RK, Ruhlman TA: **Plastid Genomes of Seed Plants**. Springer Netherlands; 2012.
28. Haberle RC, Fourcade HM, Boore JL, Jansen RK: **Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes**. *J Mol Evol* 2008, **66**(4):350–361.
29. Wicke S, Schneeweiss GM, dePamphilis CW, Muller KF, Quandt D: **The evolution of the plastid chromosome in land plants: gene content, gene order, gene function**. *Plant Mol Biol* 2011, **76**(3-5):273–297.
30. Dong W, Liu J, Yu J, Wang L, Zhou S: **Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding**. *PLoS One* 2012, **7**(4):e35071.
31. Magallon S, Hilu KW, Quandt D: **Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates**. *Am J Bot* 2013, **100**(3):556–573.
32. Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA *et al*: **Phylotranscriptomic analysis of the origin and early diversification of land plants**. *Proc Natl Acad Sci U S A* 2014, **111**(45):E4859–4868.
33. Daniell H, Lin CS, Yu M, Chang WJ: **Chloroplast genomes: diversity, evolution, and applications in genetic engineering**. *Genome Biol* 2016, **17**(1):134.
34. Guo XL, Zheng HY, Price M, Zhou SD, He XJ: **Phylogeny and Comparative Analysis of Chinese *Chamaesium* Species Revealed by the Complete Plastid Genome**. *Plants (Basel)* 2020, **9**(8).
35. Xie FM, Xie DF, Xie C, Yu Y, Zhou SD, He XJ: **Adaptation Evolution and Phylogenetic Analyses of Species in Chinese *Allium* Section Pallasia and Related Species Based on Complete Chloroplast Genome Sequences**. *Biomed Res Int* 2020, **2020**:8542797.
36. Xie DF, Huan-Xi YU, Price M, Xie C, He XJ: **Phylogeny of Chinese *Allium* Species in Section Daghestanica and Adaptive Evolution of *Allium* (Amaryllidaceae, Allioideae) Species Revealed by the Chloroplast Complete Genome**. 2019, **10**:460
37. Ren T, Li ZX, Xie DF, Gui LJ, Peng C, Wen J, He XJ: **Plastomes of eight *Ligusticum* species: characterization, genome evolution, and phylogenetic relationships**. *BMC Plant Biol* 2020, **20**(1):519.
38. Carbonell-Caballero J, Alonso R, Ibanez V, Terol J, Talon M, Dopazo J: **A Phylogenetic Analysis of 34 Chloroplast Genomes Elucidates the Relationships between Wild and Domestic Species within the Genus Citrus**. *Mol Biol Evol* 2015, **32**(8):2015–2035.
39. Logacheva MD, Shipunov AB: **Phylogenomic analysis of *Picramnia*, *Alvaradoa*, and *Leitneria* supports the independent Picramniales**. *J. Syst. Evol.* 2017, **55**:171–176.
40. Li J, Xie DF, Guo XL, Zheng ZY, He XJ, Zhou SD: **Comparative Analysis of the Complete Plastid Genome of Five *Bupleurum* Species and New Insights into DNA Barcoding and Phylogenetic Relationship**. *Plants (Basel)* 2020, **9**:543.
41. Gou W, Jia SB, Price M, Guo XL, Zhou SD, He XJ: **Complete Plastid Genome Sequencing of Eight Species from *Hansenia*, *Haplosphaera* and *Sinodielsia* (Apiaceae): Comparative Analyses and Phylogenetic Implications**. *Plants (Basel)* 2020, **9**:1523.

42. Yang X, Xie DF, Chen JP, Zhou SD, Yu Y, He XJ: **Comparative Analysis of the Complete Chloroplast Genomes in *Allium* Subgenus Cyathophora (Amaryllidaceae): Phylogenetic Relationship and Adaptive Evolution**. *Biomed Res Int* 2020, **2020**:1732586.
43. Wang WB, Yu H, Wang JH, Lei WJ, Gao JH, Qiu XP, Wang JS: **The Complete Chloroplast Genome Sequences of the Medicinal Plant *Forsythia suspensa* (Oleaceae)**. *Int J Mol Sci* 2017, **18**:2288.
44. Asaf S, Khan AL, Khan AR, Waqas M, Kang SM, Khan MA, Lee SM, Lee IJ: **Complete Chloroplast Genome of *Nicotiana otophora* and its Comparison with Related Species**. *Front Plant Sci* 2016, **7**:843.
45. Ravi V, Khurana JP, Tyagi AK, Khurana P: **An update on chloroplast genome**. *Pl Syst Evol* 2008, **271**: 101–122.
46. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK: **The complete chloroplast genome sequence of *Pelargonium* x *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants**. *Mol Biol Evol* 2006, **23**(11):2175–2190.
47. Guisinger MM, Kuehl JV, Boore JL, Jansen RK: **Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage**. *Mol Biol Evol* 2011, **28**(1):583–600.
48. Martin M, Sabater B: **Plastid ndh genes in plant evolution**. *Plant Physiol Biochem* 2010, **48**(8):636–645.
49. Wicke S, Muller KF, de Pamphilis CW, Quandt D, Wickett NJ, Zhang Y, Renner SS, Schneeweiss GM: **Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family**. *Plant Cell* 2013, **25**(10):3711–3725.
50. Hanson G, Coller J: **Codon optimality, bias and usage in translation and mRNA decay**. *Nat Rev Mol Cell Biol* 2018, **19**(1):20–30.
51. Morton BR: **The role of context-dependent mutations in generating compositional and codon usage bias in grass chloroplast DNA**. *J Mol Evol* 2003, **56**(5):616–629.
52. Rao Y, Wu G, Wang Z, Chai X, Nie Q, Zhang X: **Mutation bias is the driving force of codon usage in the Gallus gallus genome**. *DNA Res* 2011, **18**(6):499–512.
53. Behura SK, Severson DW: **Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes**. *Biol Rev Camb Philos Soc* 2013, **88**(1):49–61.
54. Harrison RJ, Charlesworth B: **Biased gene conversion affects patterns of codon usage and amino acid usage in the Saccharomyces sensu stricto group of yeasts**. *Mol Biol Evol* 2011, **28**(1):117–129.
55. Novoa EM, Ribas de Pouplana L: **Speeding with control: codon usage, tRNAs, and ribosomes**. *Trends Genet* 2012, **28**(11):574–581.
56. Rensing SA, Fritzowsky D, Lang D, Reski R: **Protein encoding genes in an ancient plant: analysis of codon usage, retained genes and splice sites in a moss, *Physcomitrella patens***. *BMC Genomics* 2005, **6**:43.
57. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, *et al*: **Evolution of genes and genomes on the Drosophila phylogeny**. *Nature* 2007, **450**(7167):203–218.
58. Zuo LH, Shang AQ, Zhang S, Yu XY, Ren YC, Yang MS, Wang JM: **The first complete chloroplast genome sequences of *Ulmus* species by de novo sequencing: Genome comparative and taxonomic position analysis**. *PLoS One* 2017, **12**(2):e0171264.
59. Goro T, Ryoma K, Moore CE, Tyler M, Onodera NT, Yuichiro K, Archibald JM, Yuji I, Tetsuo H: **Comparative plastid genomics of *Cryptomonas* species reveals fine-scale genomic responses to loss of photosynthesis**. *Genome Biology and Evolution* 2020, **2**:2–46.
60. Asaf S, Khan AL, Lubna, Khan A, Khan A, Khan G, Lee IJ, Al-Harrasi A: **Expanded inverted repeat region with large scale inversion in the first complete plastid genome sequence of *Plantago ovata***. *Sci Rep* 2020, **10**(1):3881.
61. Henriquez CL, Abdullah, Ahmed I, Carlsen MM, Zuluaga A, Croat TB, McKain MR: **Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae)**. *Genomics* 2020, **112**(3):2349–2360.
62. Ogihara Y, Terachi T, Sasakuma T: **Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species**. *PNAS* 1988, **85**(22):8573–8577.
63. Weng ML, Blazier JC, Madhumita G, Jansen RK: **Reconstruction of the Ancestral Plastid Genome in Geraniaceae Reveals a Correlation between Genome Rearrangements, Repeats, and Nucleotide Substitution Rates**. *Mol Biol Evol* 2013,(3):645–659.
64. Yang X, Zhou T, Su X, Wang G, Zhang X, Guo Q, Cao F: **Structural characterization and comparative analysis of the chloroplast genome of *Ginkgo biloba* and other gymnosperms**. *Journal of Forestry Research* 2021, **32**:765–778.
65. Wang X, Zhou T, Bai G, Zhao Y: **Complete chloroplast genome sequence of *Fagopyrum dibotrys*: genome features, comparative analysis and phylogenetic relationships**. *Scientific Reports* 2018, **8**:12379.
66. Kode V, Mudd EA, Iamtham S, Day A: **The tobacco plastid accD gene is essential and is required for leaf development**. *Plant J* 2005, **44**(2):237–244.
67. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, Jansen RK: **Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus***. *BMC Genomics* 2007, **8**:174.
68. Goulding SE, Olmstead RG, Morden CW, Wolfe KH: **Ebb and flow of the chloroplast inverted repeat**. *Molecular and General Genetics* 1996, **252**:195–206.

69. Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM: **Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots**. *BMC Evol Biol* 2008, **8**:36.
70. Zhu A, Guo W, Gupta S, Fan W, Mower JP: **Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates**. *New Phytol* 2016, **209**(4):1747–1756.
71. Lin CP, Wu CS, Huang YY, Chaw SM: **The complete chloroplast genome of *Ginkgo biloba* reveals the mechanism of inverted repeat contraction**. *Genome Biol Evol* 2012, **4**(3):374–381.
72. Li B, Zheng Y: **Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae**. *Sci Rep* 2018, **8**(1):9285.
73. Mehmood F, Abdullah, Shahzadi I, Ahmed I, Waheed MT, Mirza B: **Characterization of *Withania somnifera* chloroplast genome and its comparison with other selected species of Solanaceae**. *Genomics* 2020, **112**(2):1522–1530.
74. Menezes APA, Resende-Moreira LC, Buzatti RSO, Nazareno AG, Carlsen M, Lobo FP, Kalapothakis E, Lovato MB: **Chloroplast genomes of *Byrsonima* species (Malpighiaceae): comparative analysis and screening of high divergence sequences**. *Sci Rep* 2018, **8**(1):2210.
75. Zhao L, Li X, Zhang N, Zhang SD, Yi TS, Ma H, Guo ZH, Li DZ: **Phylogenomic analyses of large-scale nuclear genes provide new insights into the evolutionary relationships within the rosids**. *Mol Phylogenet Evol* 2016, **105**:166–176.
76. Guo C, Ma PF, Yang GQ, Ye XY, Guo Y, Liu JX, Liu YL, Eaton DAR, Guo ZH, Li DZ: **Parallel ddRAD and Genome Skimming Analyses Reveal a Radiative and Reticulate Evolutionary History of the Temperate Bamboos**. *Syst Biol* 2020,syaa076.
77. Folk RA, Mandel JR, Freudenstein JV: **Ancestral Gene Flow and Parallel Organellar Genome Capture Result in Extreme Phylogenomic Discord in a Lineage of Angiosperms**. *Syst Biol* 2017, **66**(3):320–337.
78. Yang J, Yue M, Niu C, Ma XF, Li ZH: **Comparative Analysis of the Complete Chloroplast Genome of Four Endangered Herbals of *Notopterygium***. *Genes (Basel)* 2017, **8**(4):124.
79. Doyle J: **A rapid DNA isolation procedure for small quantities of fresh leaf tissue.** *Phytochem. Bull.* 1987, **19**:11–15.
80. Dierckxsens N, Mardulyn P, Smits G: **NOVOPlasty: de novo assembly of organelle genomes from whole genome data**. *Nucleic Acids Res* 2017, **45**(4):e18.
81. Qu XJ, Moore MJ, Li DZ, Yi TS: **PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes**. *Plant Methods* 2019, **15**:50.
82. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al*: **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data**. *Bioinformatics* 2012, **28**(12):1647–1649.
83. Lohse M, Drechsel O, Kahlau S, Bock R: **OrganellarGenomeDRAW–a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets**. *Nucleic Acids Res* 2013, **41**:W575–581.
84. Peden, John F: **Analysis of codon usage**. *University of Nottingham* 2000, **90**(1):73–74.
85. Sharp PM, Li WH: **An Evolutionary Perspective on Synonymous Codon Usage in Unicellular Organisms**. *Journal of Molecular Evolution* 1986, **24**:28–38.
86. Thiel T, Michalek W, Varshney RK, Graner A: **Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.)**. *Theor Appl Genet* 2003, **106**(3):411–422.
87. Kurtz S, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R: **REPuter: the manifold applications of repeat analysis on a genomic scale**. *Nucleic Acids Research* 2001, **29**(22):4633–4642.
88. Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: improvements in performance and usability**. *Mol Biol Evol* 2013, **30**(4):772–780.
89. Librado P, Rozas J: **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data**. *Bioinformatics* 2009, **25**(11):1451–1452.
90. Stamatakis A: **RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models**. *Bioinformatics* 2006, **22**(21):2688–2690.
91. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP: **MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space**. *Syst Biol* 2012, **61**(3):539–542.
92. Posada D, Crandall KA: **MODELTEST: testing the model of DNA substitution**. *Bioinformatics* 1998, **14**(9):817–818.

Figures

Figure 1

Plastid genome map of nine sequenced species. The genes shown outside of the circle are transcribed clockwise and inside of the circle are transcribed anticlockwise. The genes belonging to different functional groups are separated by color. The darker grey area in the inner circle shows the GC content, while the lighter grey corresponds to the AT content. LSC, large single-copy; SSC, small single-copy; IR, inverted repeat.

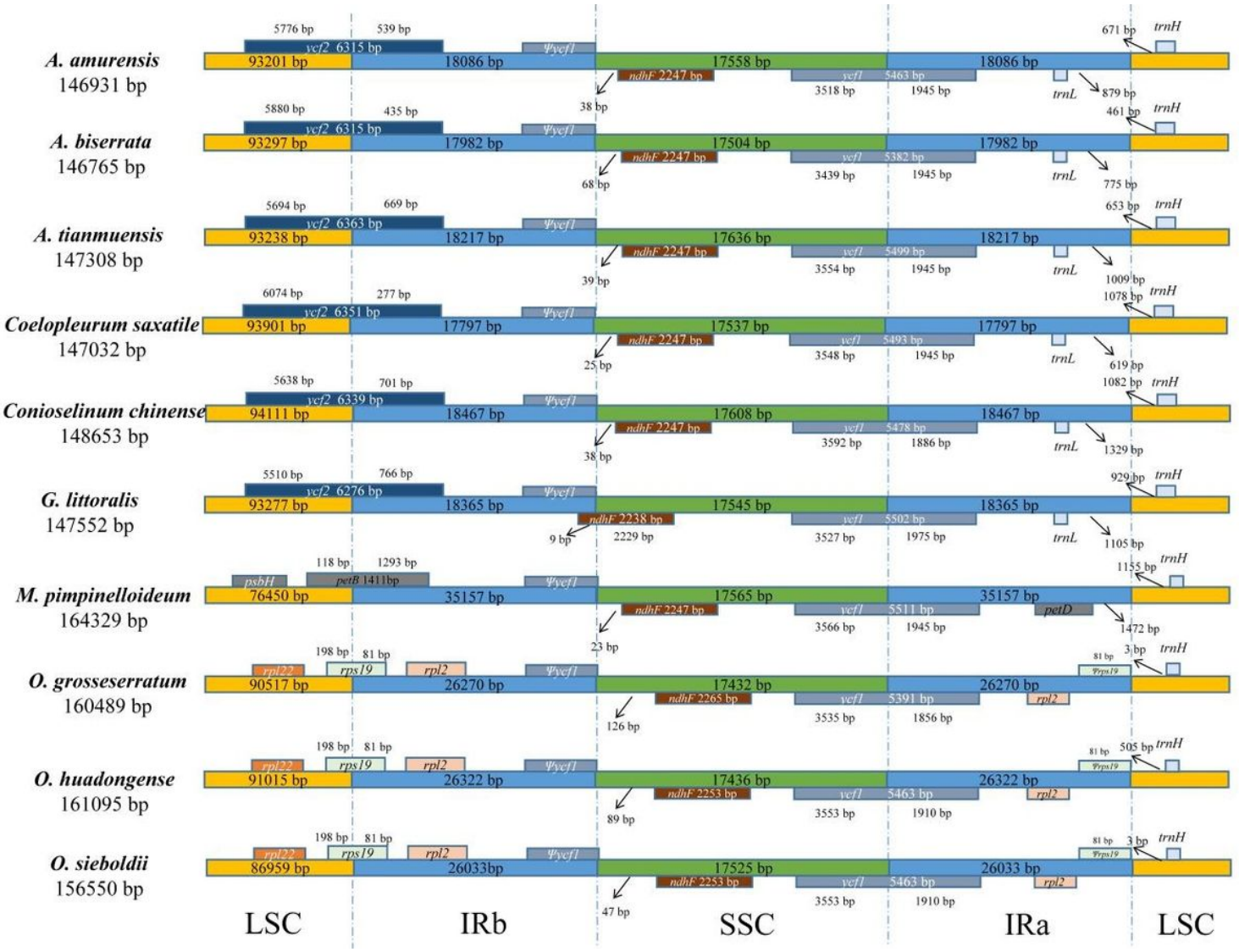

Figure 2

Comparison of the border regions of the ten species' plastid genomes. LSC (large single-copy), SSC (small single-copy), and IR (inverted repeat) regions.

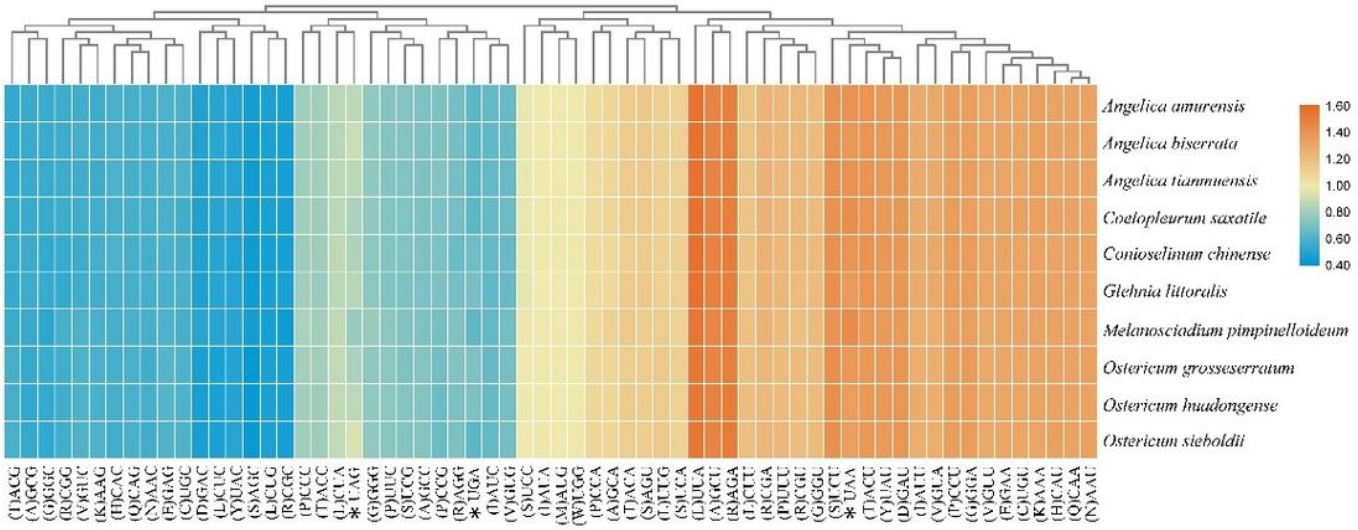

Figure 3

The RSCU values of ten related species. The orange shows higher RSCU values and the blue shows lower RSCU values. * indicated the termination codon.

Figure 4

Analysis of repeat sequences and simple sequence repeats (SSRs) in ten species. (A) Numbers of four repeat types, (B) numbers of repeats divided by the length in species, (C) Numbers of six SSR types

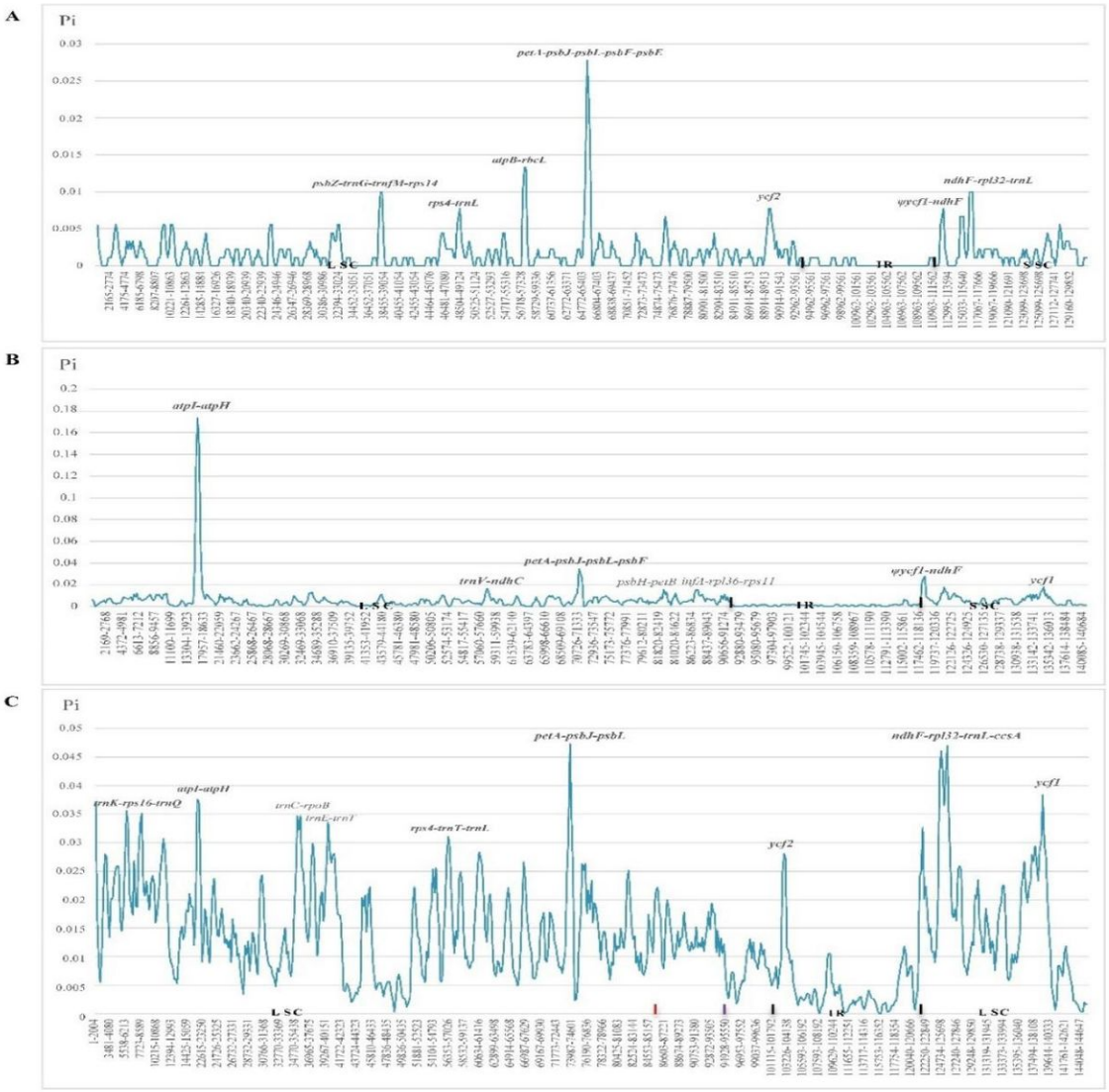

Figure 5

The nucleotide diversity of (A) 3 Angelica species, (B)3 Ostericum species and (C) ten related species. The regions with the higher Pi values were marked out. LSC (large single-copy), SSC (small single-copy), and IR (inverted repeat) regions. In (C) the red line indicates the LSC/IR border of Melanosciadium pimpinelloideum, the purple line indicates the LSC/IR border of Ostericum.

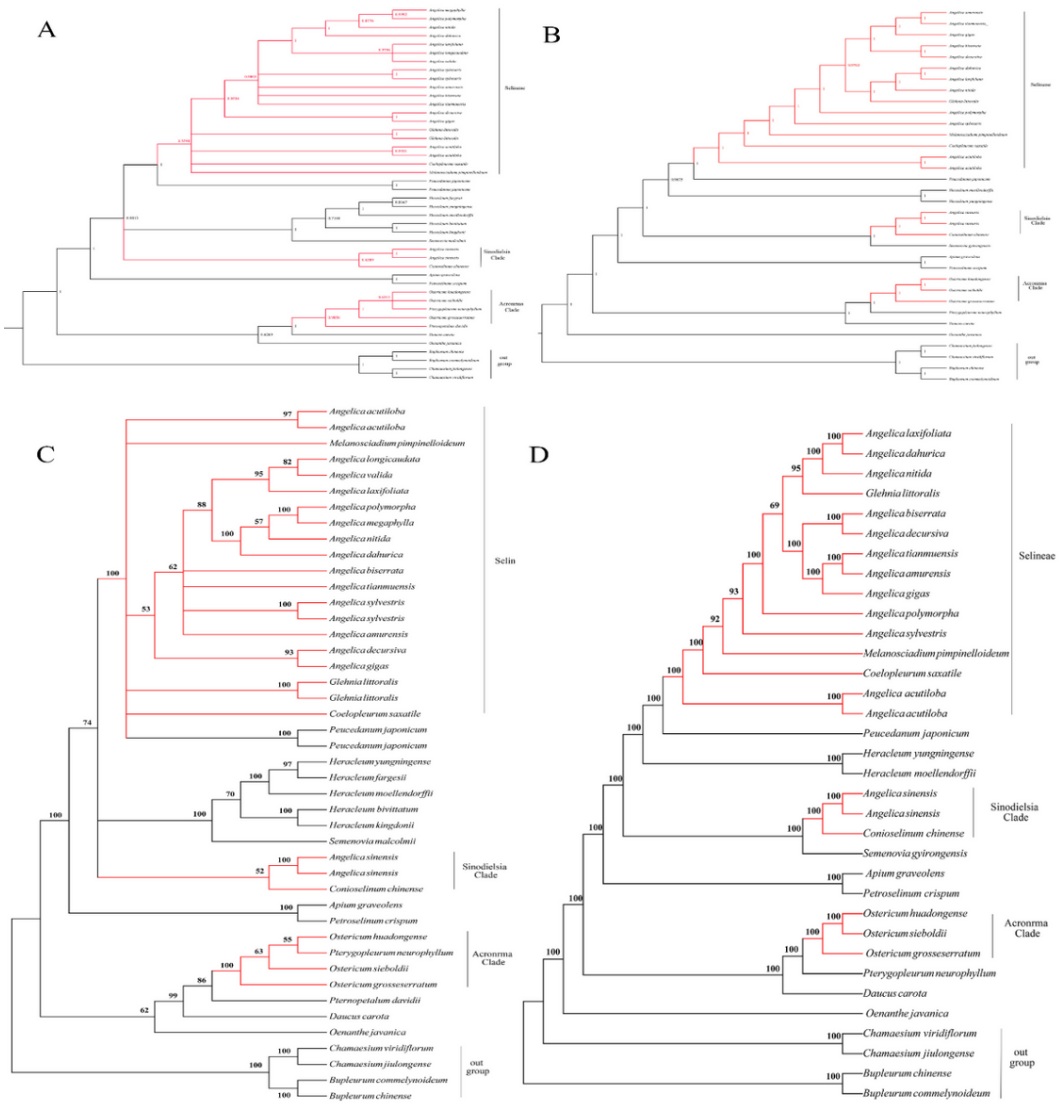

Figure 6

Phylogenetic relationships among Angelica, Ostericum and related groups are inferred from maximum likelihood (ML) and Bayesian inference (BI) based on protein-coding genes and nrITS. (A) Bayesian inference of nrITS; (B) Bayesian inference of protein-coding genes; (C) maximum likelihood of nrITS; (D) maximum likelihood of protein-coding genes. Maximum likelihood bootstrap support (ML BS) and Bayesian posterior probabilities (BI PP) are presented at the nodes.

Supplementary Files

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

- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS5.docx