BMC Genomics

RESEARCH Open Access

The population chloroplast genomes of *Populus* reveal the phylogenetic relationship between three new taxa of sect. *Leucoides* and their parents

Yujie Shi[1†], Jingliang Huang[2†], Xueqin Wan[2*], Jinglian Shi[3], Zhen Chen[1] and Wei Zeng[1*]

Abstract

Background Poplars are important woody plants, which are widely distributed in the forests from the subtropics to the north of the Northern Hemisphere. Poplars have high ecological and economic value. However, there are frequent interspecific and intraspecific hybrids in *Populus*, resulting in a large number of intermediate taxa, which makes the morphological identification of *Populus* very challenging. Plastid genome is an important tool to study the evolutionary relationship of plants. Therefore, comparison and phylogenetic analysis were carried out based on the population chloroplast genomes of 34 individuals from 7 taxa.

Results In this study, seven newly assembled and annotated chloroplast genomes of *Populus* were reported. They all had typical quadripartite structures with the same GC content (37.6%), but there were differences within the population, and the genome size ranged from 155,736 bp to 156,812 bp. In all *Populus* species, 134 genes were identified, including 88 protein coding genes (PCGs), 37 tRNA and 8 rRNA genes. The gene sequences alignment of different taxa showed that the gene sequences and content were relatively conservative, there was no gene rearrangement, and only 3 highly variable regions (*psbZ-trnG*, *ndhC-trnV* and *trnN-trnR*) were identified, which can be used as molecular markers. Most PCGs had high codon usage bias and 3 positive selection genes (*rps7*, *rps12* and *rpl16*) have been identified. The analysis of population genetic structure and phylogeny showed that the chloroplast genomes supported that *Populus* was a monophyletic taxon, which could be divided into four sections (*Abaso*, *Turanga*, *Populus* and ATL (*Aigeiros*, *Tacamahaca* and *Leucoides*)). Among them, *P. dafengensis*, *P. butuoensis* and *P. szechuanica* had the closest genetic relationship, *P. gonggaensis* and *P. cathayana* had the closest genetic relationship, it was speculated that the taxa of Sect. *Tacamahaca* may be the main female parent of the three new taxa from Sect. *Leucoides*.

†Yujie Shi and Jingliang Huang contributed equally to this work.

*Correspondence:
Xueqin Wan
w-xue@163.com
Wei Zeng
zengw@tzc.edu.cn

Full list of author information is available at the end of the article

Conclusion In general, this study provides valuable insights for new species identification, phylogenetic relationships, breeding and resource development, and genetic diversity of *Populus*.

Keywords *Populus*, Population chloroplast genome, *Leucoides*, Comparative analysis, Phylogeny

Introduction

Poplar is an important woody plant belonging to the genus *Populus*, which is widely distributed in the forests from subtropics to the north of the Northern Hemisphere, mainly in temperate forests [1]. Poplars are one of the most important afforestation tree species in the world, which play an important role in global wood production and ecological environment construction [2]. Furthermore, poplars have the advantages of fast growth, easy asexual reproduction, relatively small genome, moderate number of species, easy to obtain several high-quality genomes, mature genetic operation system, and are also recognized as a model forest tree species for molecular biology research [3, 4].

Although the genus has high ecological and economic value and has been used as a model for the study of tree diversity for a long time, the phylogenetic relationship within the genus is still unclear. Frequent interspecific hybridization and clonal reproduction haunt taxonomists of the genus, with recognized species, varieties and hybrids ranging from 22 to 85 [5]. According to the morphological characteristics, the international classification system divides the genus *Populus* into 6 sections (Sect. *Abaso*, Sect. *Turanga*, Sect. *Populus*, Sect. *Leucoides*, Sect. *Tacamahaca*, Sect. *Aigeiros*) and 29 species [5], while the Chinese classification system divides the genus into 5 sections and 71 species [6]. Therefore, there are great differences in species and intraspecific in the classification system of *Populus* (genus-sections-species), especially in the taxa from Sect. *Tacamahaca* and Sect. *Leucoides*.

There are 57 *Populus* species in China, of which 25 endemic species are distributed in southwest China, which is the area with the most abundant and diverse poplar resources, and is also considered to be one of the centers of natural distribution and variation of poplars in China [7–9]. This area is the main natural distribution area of the taxa of Sect. *Tacamahaca* and Sect. *Leucoides*, which is dominated by mountain and plateau landforms and plays an important role in speciation [10]. Moreover, the unique and complex topographic conditions of the region have also become a refuge for plants to cope with climate change, providing opportunities for plant hybridization [11]. There are many species in the genus *Populus*, with few flower and fruit traits that can be used for classification, and the morphological variation of vegetative organs is very complex. Moreover, interspecific hybridization and even intraspecific hybridization are quite common in the genus, and there are a large number of taxa with hybrid origin. Therefore, the genus *Populus* is recognized as one of the few most taxonomically difficult groups among plants [9, 12], and many problems still need to be solved.

Chloroplast genome is one of the plant plastid genetic systems, which is a circular double-stranded tetrad composed of large single copy region (LSC), small single copy region (SSC) and a pair of inverted repeats (IRs) interspersed between them. It usually encodes about 130 genes mainly related to photosynthesis and chloroplast replication [13]. Because of its small genome, relatively conservative and slow nucleotide replacement rate, it is used as a widespread tool for plant identification, evolutionary biology and genetic diversity [14]. So far, a large number of poplar chloroplast genomes have been published one after another, such as *Populus trinervis* [15], *P. maximowiczii* [16], *P. wilsonii* [17] and *P. szechuanica* [18], etc. Moreover, comparative chloroplast genomes have become an effective method to solve many biological and evolutionary problems at different levels, especially for deeper phylogenetic analysis [19, 20]. Zong et al. conducted a comparative analysis based on the chloroplast genomes of five *Populus* species from the western Sichuan Plateau of China and found that their chloroplasts were highly similar, and *P. schneideri* was a sister group of *P. kangdingensis* and *P. pseudoglauca*, the *P. xiangchengensis* was a sister to *P. cathayan*a [21]. Then, Phylogenetic studies based on the chloroplast genomes of 24 poplar species were carried out. It was found that *Populus* was a monophyletic group, and the 24 species were divided into 5 clades, but the species of Sect. *Tacamahaca* and Sect. *Leucoides* were difficult to distinguish [4]. Similar results were obtained by Zhou et al. phylogenetic tree constructed based on chloroplast genomes of 39 *Populus* species [22].

In recent years, with the rapid development of sequencing technology, the chloroplast genomes of the population have been widely studied to explore the matrilineal genetic evolution and phylogeny of plants at the population level. In this study, *Populus dafengensis*, *P. gonggaensis* and *P. butuoensis* were new taxa from Sect. *Leucoides* of *Populus* which were identified by the comprehensive information of morphology, geography and whole genome in the early stage of our research group [10]. They all belong to hybrid origin, in which the parents of *P. dafengensis* are *P. lasiocarpa* and *P. szechuanica*, the parents of *P. gonggaensis* are *P. lasiocarpa* and *P. cathayana*, and the parents of *P. butuoensis* are *P. wilsonii* and *P. szechuanica*. However, there is still a lack of

chloroplast genome research between them and their parents. Therefore, based on the genome resequencing data from 34 individuals of 7 taxa, we assembled complete chloroplast genomes, and compared their sequence differences, codon bias, simple repeat sequences, nucleotide polymorphisms, genetic structure, and constructed phylogenetic trees based on 58 other chloroplast genome sequences of Salicaceae. This study provides an important reference for molecular evolution and phylogenetic reconstruction of *Populus*.

Results

Characteristics of chloroplast genomes from the seven *Populus* species

The complete chloroplast genomes from 34 individuals of 7 poplars ranged from 155,736 bp (*P. wilsonii*) to 156,812 bp (*P. cathayana*). They were all divided into typical quadripartite structures (Fig. 1), the LSC length was distributed between 84,733 bp (*P. szechuanica*) and 85,058 bp (*P. wilsonii*), the SSC length was ranged from 16,447 bp (*P. butuoensis*) to 16,628 bp (*P. wilsonii*), the IR length was between 27,034 bp (*P. wilsonii*) and 27,657 bp (*P. cathayana*), and the overall GC content was relatively consistent, which was 36.7% (Table S3). Furthermore, there were also differences among different individuals of the same species, in which *P. lasiocarpa* and *P. szechuanica* had rich intra-individual variations, and the length of their chloroplast genomes had 137 bp and 135 bp differences, respectively. The intra-individual variations of *P. wilsonii*, *P. gonggaensis* and *P. dafengensis* were small, and there was almost no variation in *P. butuoensis* and *P. cathayana* (Table S3).

Although the length of chloroplast genome regions of 7 poplars were different between species and within species, the number of genes encoded by them was the same. The all individuals encoded 134 functional genes, of which 114 were single copy genes, including 79 protein coding genes (PCGs), 30 tRNAs, 4 rRNAs and 1 pseudogene (*infA*). They were divided into four categories (photosynthesis, self-replication, functionally unknown genes and other genes). Most of these genes existed in the form of single copy, but there were also 20 multi-copy genes, including 9 PCGs, 7 tRNA and 4 rRNA genes. Moreover, 18 genes had 1 intron, 4 genes (*rps12* (2), *ycf3* and *clpP*) had 2 introns, and *rps12* was trans-spliced with the duplicated 3′ end in the IRs and the 5′ end located in the LSC region (Fig. 1 and Table S4).

Codon usage bias

The relative synonymous codon usage (RSCU) values of PCGs in seven poplar chloroplast genomes were analyzed. The results showed that all samples had the same codon usage bias for each amino acid. There were 26,460 – 26,879 codons in all PCGs of the seven poplars,

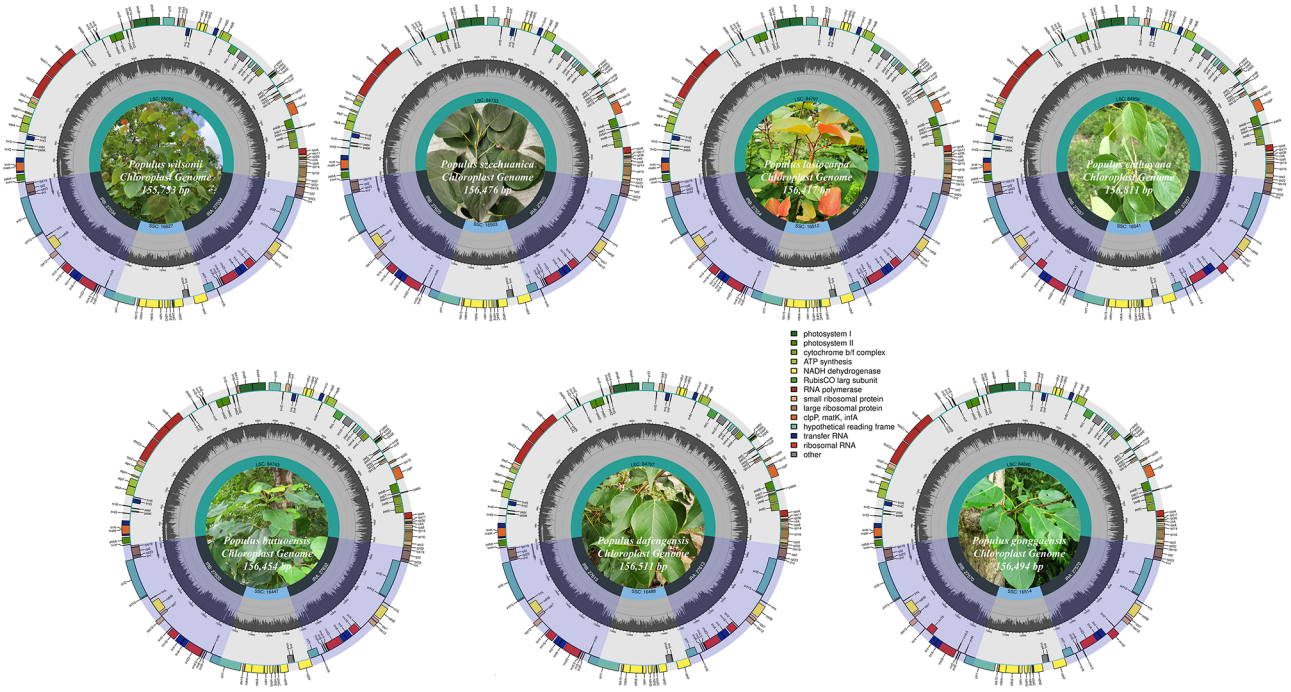

Fig. 1 Chloroplast genome maps of seven *Populus* species. The forward coding genes are located on the outside of the ring, while the inverted coding genes are located on the inside of the ring. The gray circle inside represents the GC content and shows the corresponding regions and corresponding lengths of the four blocks (LSC/SSC/IRs). Genes with different functions are given different colors. The center of the circle is the typical morphological traits of each poplar

of which the number of codons encoding leucine was the largest, 2858–2881 (10.71–10.80%), and the number of codons encoding cysteine was the least, only 304–309 (1.13–1.15%) (Fig. 2B). Among the 61 codons, RSCU values of 29 codons were greater than 1, among which the RSCU values of AGA were the largest (1.90–1.93), 30 codons were less than 1, and 2 codons (AUG and UGG) were equal to 1. Furthermore, 97% of the codons with a RSCU value greater than 1 end in A/U, indicating that the end of A /U in their chloroplast genomes tends to play a dominant role (Fig. 2A).

SSRs and long sequence repeats analysis

Simple sequence repeats (SSRs) were widely distributed in chloroplast genomes. A total of 129–146 SSRs loci were detected in 7 poplar chloroplast genomes, mainly located in the LSC region, and 81 SSR loci were detected in the PCGs (*ycf1*, *rpoC2*, *rpoB*, *ndhD*, *ndhF* and *rps8*), the rest were located in intergenic spacer regions (IGS) and introns (Table S5 and Fig. 3A). The types of SSRs ranged from mononucleotide (Mono-) sequence repeats to hexanucleotide (Hexa-) sequence repeats, although not all six types existed in each species. Among them, mono-sequence repeats were the most common (99–116 loci), mainly A/T type repeats, and only 0–2 G/C type repeats. The dinucleotide (Di-) sequence repeats were composed of AG/GT and AT/AT, and each species had only one AG/CT. There was only one type of trinucleotide (Tri-) sequence repeat (AAT/AAT), and the number was 4–10. There were 10–11 tetranucleotide (Tetra-) sequence repeats, including 6 types. The repeats of pentanucleotide (Penta-) and hexa-sequence repeats were specific, for example, AATAT/ATATT type only existed in *P. cathayana*, AAAAAT/ATTTTT type only existed in *P. dafengensis*, AAATT/AATTT and AATATT/AATATT types only existed in *P. gonggaensis* (Fig. 3B).

In addition, we also detected a total of 362 long sequence repeats (39–70 loci) in 7 poplar chloroplast genomes, mainly located in the LSC region (223 loci). There were 69 repeats located in the PCGs (*ccsA*, *psaA*, *psaB*, *ycf1* and *ycf2*), and the rest were located in IGS and intron region, including four repeat types: forward repeats (F-), reverse repeats (R-), palindromic repeats (P-) and complementary repeats (C-) (Table S6). In all species, the repetition frequency of F- and P- repeats were higher, being detected 153 and 122 times, respectively; while the repetition frequency of R- and C- repeats were lower, being detected 65 and 22 times, respectively (Fig. 3C). There were great differences in the number of long sequence repeats among different species, in which the total number of *P. wilsonii* and *P. cathayana* were much more than that of other species, but their distribution trend was the same. The length of most long sequence repeats varied from 30 bp to 40 bp (Fig. 3D).

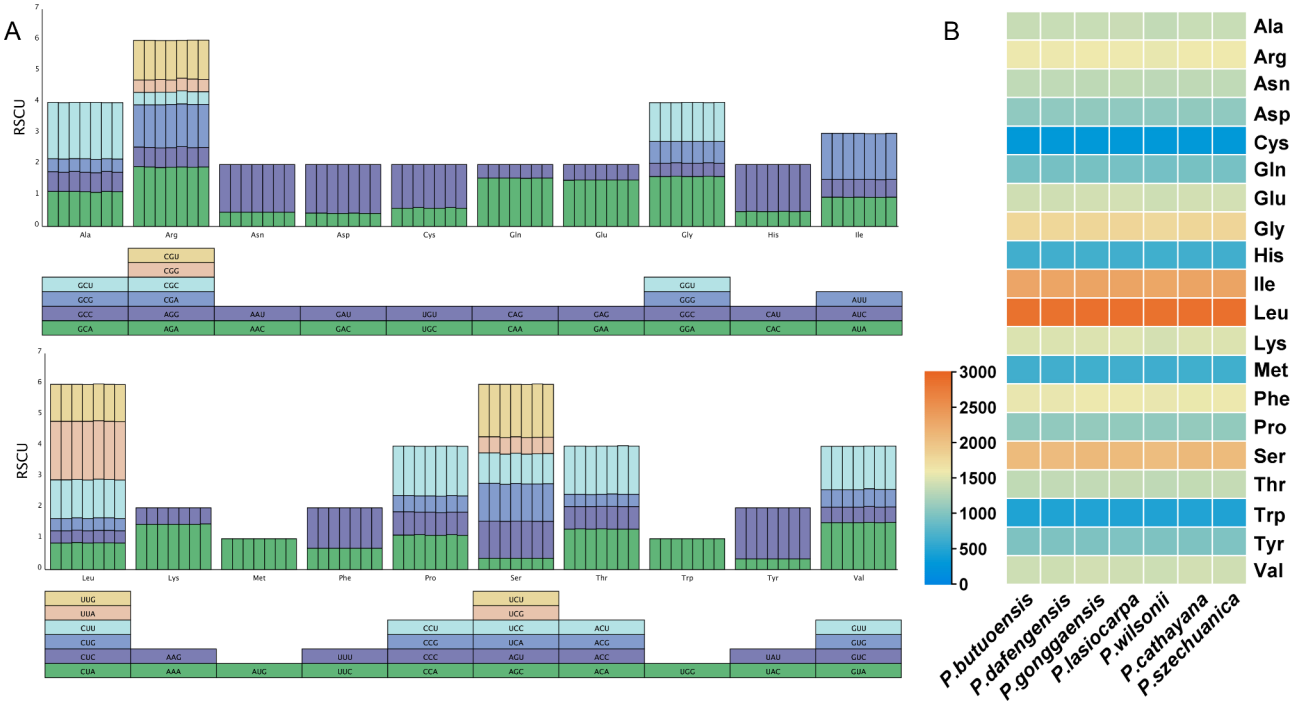

Fig. 2 Codon bias analysis of 7 *Populus* species. (**A**) RSCU analysis from 61 codons of 20 amino acids, each column represents a kind of poplar, from left to right are *P. butuoensis*, *P. dafengensis*, *P. gonggaensis*, *P. lasiocarpa*, *P. wilsonii*, *P. cathayana* and *P. szechuanica*, respectively. (**B**) The total number of codons of the PCGs. The value of each species is the average of all repeating individuals of the population

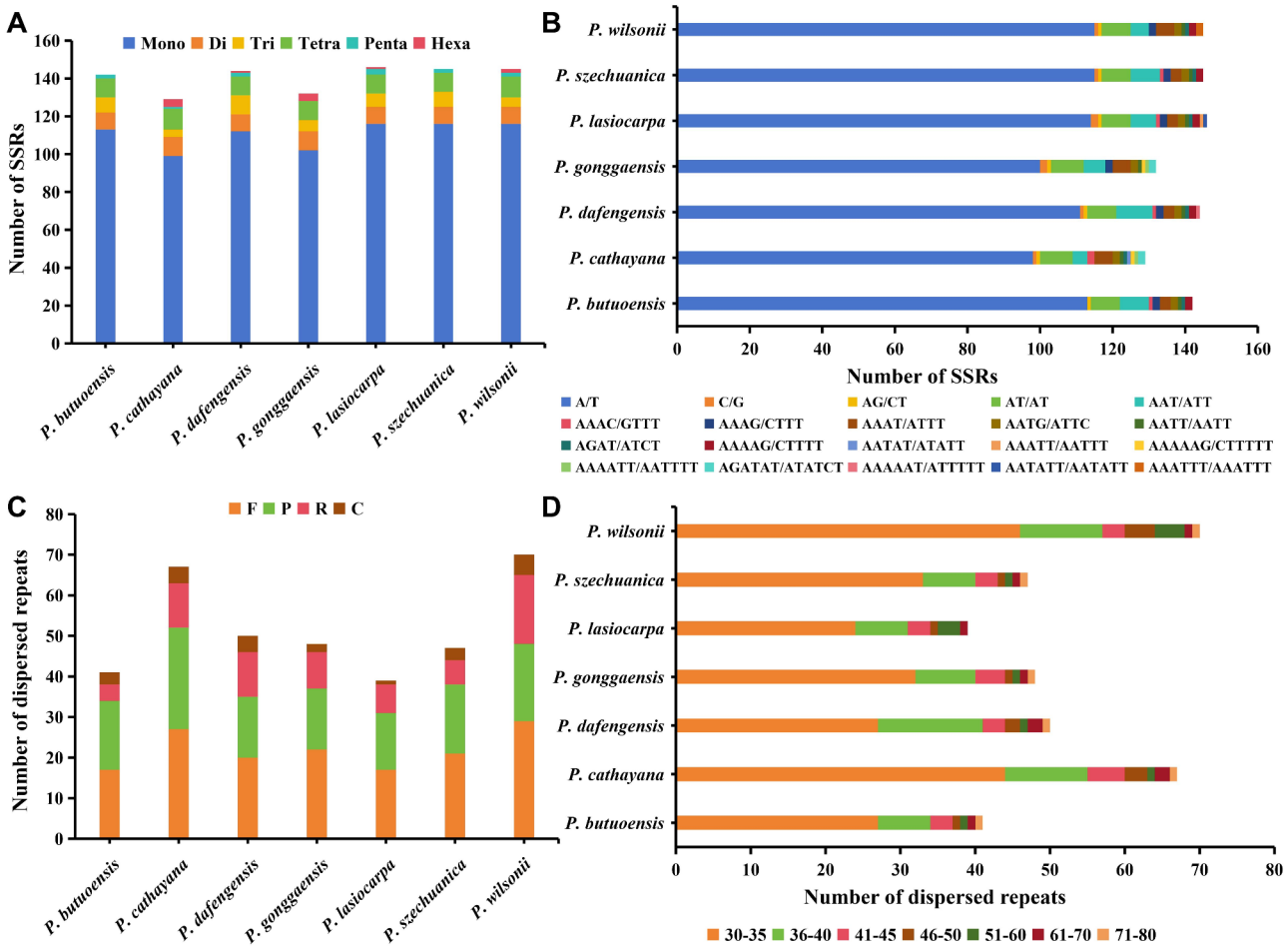

Fig. 3 SSRs and long sequence repeats analysis of 7 poplars. (**A**) The number of the different SSRs types. (**B**) The number of SSR motifs in different repeat types. (**C**) The number of different long sequence repeat type. (**D**) The distribution of the length of long sequence repeats

IRs expansion and contraction, sequence difference and nucleotide diversity

The junction boundaries of LSC, IRs and SSC regions from 7 poplars were compared comprehensively. The junction between IRs and SC region showed a high conservatism. Several key genes located at the junctions of LSC/IRb (JLB), IRb/SSC (JSB), SSC/IRa (JSA) and IRa/LSC (JLA) were found in the chloroplast genomes of the seven poplars, namely *rps22*, *rps19*, *ycf1*, *ndhF*, *trnN* and *trnH* (Fig. 4). One copy of *ycf1* gene straddled the boundary of IRs/SSC, mainly located in the IRb region, only 33–158 bp located in the SSC region, while the other copy spanned the IRa/SSC region, mostly in the SSC region (3758–3774 bp), and only 1/3 fragment located in the IRa region. In contrast, JLB and JLA region were relatively stable, *rps22* spanned LSC/IRb region; *rps19* located in IRb and IRa regions, with 218 bp from boundary; the trnH located in the LSC region, which was away from the boundary 3 bp, and had the same distribution pattern in 7 poplars.

Taking the chloroplast genome of *P. trichocarpa* as the reference sequence, the similarity of the chloroplast genome from 7 poplars was compared (Fig. 5). The results showed that the interspecific differences were small and the exons were highly conserved. The main differences came from *psbZ-trnG*, *trnL-ndhB*, *rps7*, *trnN-trnR* and *ndhB-trnL* regions, most of which were located in IGS. We also detected the rearrangement and reversal among these sequences, and the results showed that all sequences were highly conservative and there was no gene rearrangement (Fig. 6A).

In addition, the nucleotide polymorphisms (Pi) values of 7 poplar chloroplast genomes were calculated, and a total of 767 polymorphic sites were detected. The Pi ranged from 0 to 0.0254, with an average of 0.00193. Due to the high sequence similarity of these species, only three highly variable sites (Pi > 0.011) were found, which were 2 IGS (*psbZ-trnG*, *ndhC-trnV*) from LSC region and *trnN-trnR* from IRa region (Fig. 6B). To sum up, the non-coding region showed more variation than the coding region, and these highly variable region sequences can be

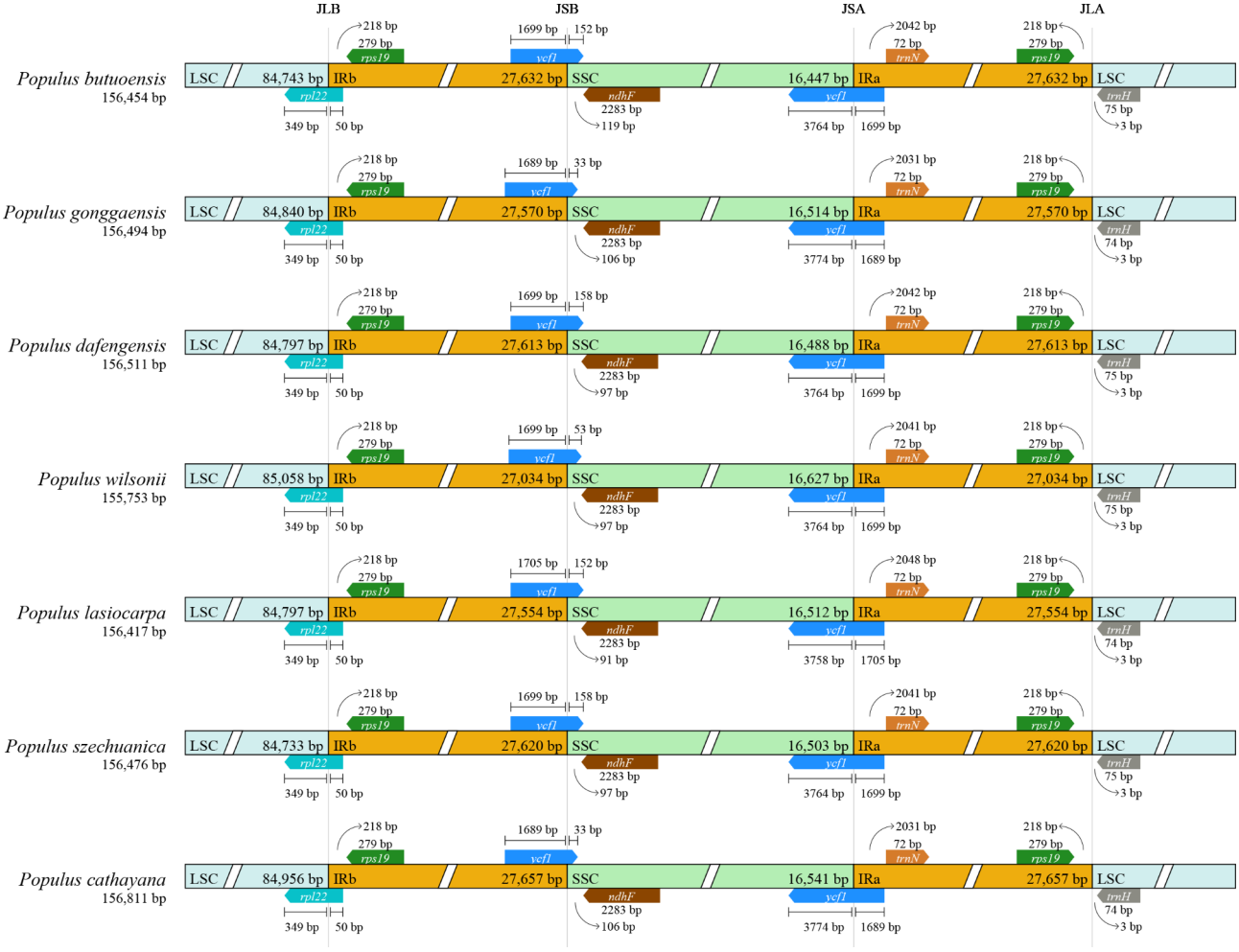

Fig. 4 Comparison of LSC, IRs and SSC junction positions of chloroplast genomes from 7 poplars. JLB: the junction of LSC and IRb; JSB: the junction of SSC and IRb; JSA: the junction of SSC and IRa; JLA: the junction of LSC and IRa. The boxes above and below the main line represent adjacent boundary genes. The gap between the gene and the boundary is represented by base length (bp)

used as marker codes for species identification and phylogenetic study of *Populus*.

Synonymous (Ks) and nonsynonymous (Ka) substitution rate analyses

Taking the chloroplast genome of *P. trichocarpa* as the reference, we calculated the Ka/Ks ratios of 79 PCGs in 7 poplars (Fig. 7). The results showed that the Ka/Ks values of 29 PCGs in 7 poplars were less than 1, or could not be determined because the Ka or Ks values were zero, which also indicated that they were conservative. Moreover, the Ka/Ks values of some genes were different among different species, such as the Ka/Ks values of *rpl16* gene from *P. gonggaensis*, *P. wilsonii*, *P. szechuanica* and *P. cathayana* were greater than 1; the Ka/Ks values of *rps12* and *rps7* genes were greater than 1 in all species, while the Ka/Ks values of rest genes were less than 1, which were purifying selections.

Population genetic evolution analysis of chloroplast genomes

The matrilineal genetic evolution was analyzed based on the chloroplast genomes of 34 individuals from 7 populations. First of all, through the analysis of genetic structure, it was found that when K = 4 or 5 (the lowest error rate of cross-validation, CV = 0.18), the 7 populations were clearly divided into 4 groups (Fig. 8E). Among them, 5 individuals of *P. cathayana* were a group, and 5 individuals of *P. gonggaensis* were a group, and there was a close relationship between them. The 5 individuals of *P. wilsonii* were in a group, while the genetic components of *P. dafengensis*, *P. lasiocarpa* and *P. szechuanica* were mixed and clustered into a group, and they were closely related to *P. butuoensis*. The results of phylogenetic tree and principal component analysis (PCA) based on the whole chloroplast genome were consistent with the genetic structure analysis (Fig. 8A-B). Furthermore, on the basis of complete chloroplast genomes alignment, we

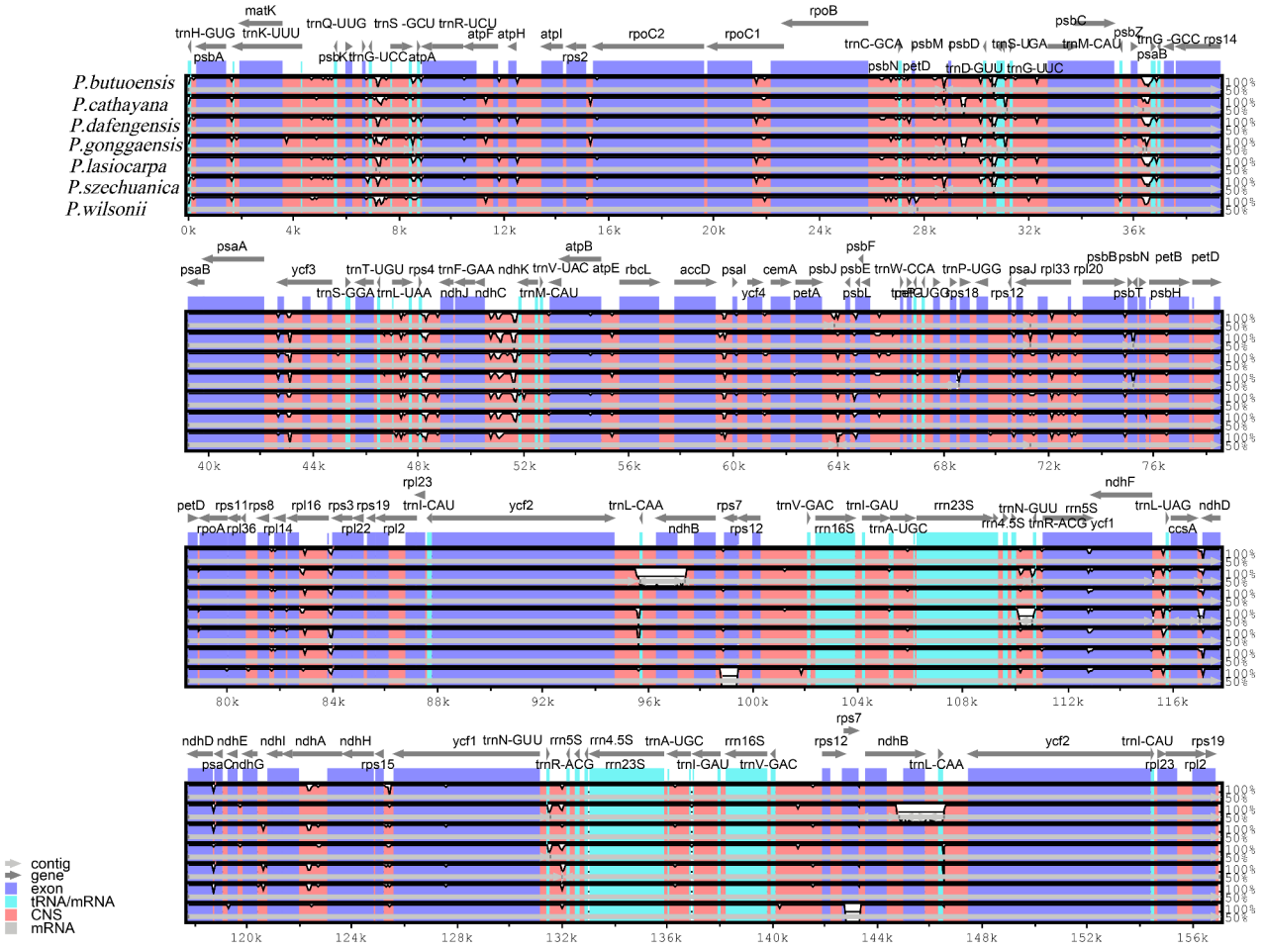

Fig. 5 The similarity maps of chloroplast genomes from 7 poplars were compared by MVISTA. Taking the chloroplast genome of *P. trichocarpa* as the reference sequence, the X axis represents the coordinates of the chloroplast gene, and the Y axis represents the average percentage similarity (50–100%). The pink region was a conservative non-coding region (CNS), purple was the gene exon region, and the blue region was tRNA or rRNA

constructed a median-joining network (Fig. 8C). The 34 individuals were divided into 16 haplotypes, which were consistent with the results of phylogenetic tree, PCA and structure analysis mentioned above. The 7 populations could be divided into 4 groups, among which *P. dafengensis*, *P. lasiocarpa*, *P. szechuanica* and *P. butuoensis* were closely related, and some individuals of *P. szechuanica* and *P. dafengensis* even shared the same haplotype. In addition, we analyzed the nucleotide polymorphisms (Pi) and genetic differentiation coefficient (Fst) of 7 populations (Fig. 8D), among which *P. dafengensis* had the highest Pi (4.10×10^{-2}), *P. gonggaensis* had the lowest Pi (4.22×10^{-3}), *P. gonggaensis-P. butuoensis* had the highest Fst (0.976), *P. lasiocarpa-P. dafengensis* had the smallest differentiation (Fst = 0.171), and the genetic diverseness among populations of *P. lasiocarpa*, *P. dafengensis*, *P. szechuanica* and *P. butuoensis* were all low (Fst = 0.371 – 0.171). To sum up, these results were in good agreement.

In order to further explore the phylogenetic relationship between these seven poplars and other sections (Sect. *Aigeiros*, *Turanga*, *Populus*, *Abaso*, *Leucoides* and *Tacamahaca*) of *Populus*, we constructed a phylogenetic tree based on the common PCGs in the chloroplast genomes from 62 species of *Populus*. The results showed that the phylogenetic trees constructed by Maximum likelihood (ML) and Bayesian inference (BI) had good support value, and the tree topology was similar (Fig. 9A and B). Among them, the taxa of Sect. *Populus*, Sect. *Turangaand* and Sect. *Abaso* were grouped into one group respectively, while the genetic relationships of taxa from Sect. *Aigeiros*, Sect. *Tacamahaca*, and Sect. *Leucoides* were clustered into a large group (ATL). Moreover, *P. dafengensis*, *P. butuoensis*, *P. lasiocarpa* and *P. szechuanica* were closely related, while *P. gonggaensis*, *P. wilsonii* and *P. cathayana* were close, and the intraspecific repeatability was stable.

Fig. 6 Chloroplast genomes alignment and hotspots detection of 7 poplars. (**A**) MAUVE alignment of seven chloroplast genomes from *Populus*. Within each of the alignments, local collinear blocks are represented by blocks of the same color connected by lines. (**B**) Comparative analysis of nucleotide polymorphisms (Pi) of chloroplast genomes in 7 poplars. X-axis represents positions of the midpoints of each window, Y-axis represents Pi value in each 600 bp window

Discussion

In this study, we compared the chloroplast genomes from 3 new taxa of *Populus* and 4 parents. They all showed typical circular quadripartite structures, ranging in length from 155,736 bp to 156,812 bp, similar to the chloroplast genome size of most poplars [8, 23–25]. The GC content was consistent among all species, but there were some differences in the length of different regions among different individuals from the same taxa. There were 137 bp and 135 bp differences in chloroplast sequence length from population of *P. lasiocarpa* and *P. szechuanica*, respectively, and the differences were mainly in IRs region (66 bp) of *P. lasiocarpa* and LSCs region (130 bp) of *P. szechuanica*. However, there was almost no diversity in the respective population of *P. wilsonii* and *P. cathayana*, which were very conservative. The chloroplast genomes of 7 *Populus* species were annotated a total of 134 genes, including 88 PCGs, 37 tRNAs and 8 rRNAs genes. The IR region of *P. wilsonii* was much smaller than that of other groups, which may be caused by the contraction of IR region and the main reason for the difference in chloroplast genome size [26, 27].

The sequence differences of chloroplast genomes from 7 poplars were compared. The results showed that they

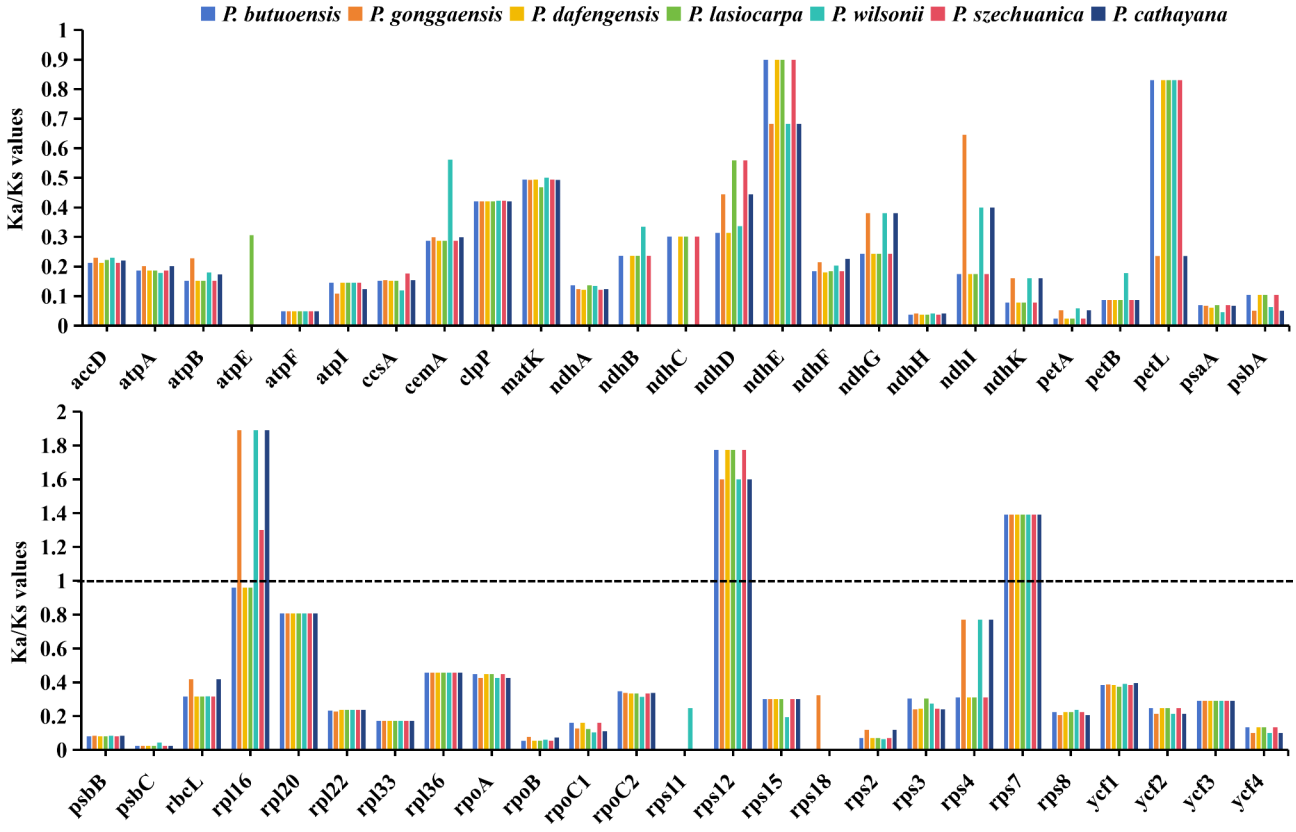

Fig. 7 The Ka/Ks ratios of 50 PCGs from 7 *Populus* species taking *P. trichocarpa* as a reference

were highly conservative, there was no gene rearrangement and reversal, and only three regions (*psbZ-trnG*, *ndhC-trnV* and *trnN-trnR*) showed high variation. These three intergenic regions can be used for phylogenetic analysis of related taxa of *Populus*, or as potential molecular markers for population genetics and phylogeny.

The sequence repeats affected gene transcriptional regulation, protein translation, chromosome formation and metabolism, reflecting the differences in mutation frequency and evolution rate of species [28]. There were 39–70 long sequence repeats in the chloroplast genome of *Populus* species, and there were four repeat types (F/P/R/C) in 7 taxa, but the number of F- and P- repeats were much more than that of R- and C- repeats. Moreover, SSRs loci were widely distributed in the genome of species, which would lead to diversity due to the different number of repeat loci [29]. At the same time, it had the characteristics of genetic stability, high abundance and wide distribution, which was often used to study genetic structure, diversity, differentiation and related taxa [30]. In this study, a total of 129–146 SSRs loci were detected in *Populus* species, including 6 repeat types. Among them, the mono-sequence repeats were the most common, accounting for 76.74–80.00% of the total SSRs, and the basic repetitive unit was the A/T. The individuals of *P. lasiocarpa*, *P. szechuanica* and *P. wilsonii* had high genetic diversity (3.94×10^{-2}, 3.23×10^{-2}, 3.18×10^{-2}) might be due to the large number of SSRs loci (146, 145, 145).

Ka and Ks nucleotide substitution patterns were very important indicators in gene evolution [31]. The selection pressure of genes was reflected by the ratio of Ka/Ks. When the value of Ka/ Ks < 1, = 1 and > 1, respectively, it indicated the purified selection, neutral evolution and positive selection of genes, and most of the PCGs were in purified selection [8, 32, 33]. When we compared 79 PCGs in *Populus* species, only 3 genes (*rps7*, *rps12* and *rpl16*) had Ka/Ks greater than 1, and the others were less than 1, indicating that these 3 genes were affected by positive selection and were undergoing a period of rapid evolution. Furthermore, codon usage bias can reflect the evolutionary characteristics and influencing factors of chloroplast genome, and plays an important role in genome expression [34]. The RSCU value reflects the codon usage patterns of different genes, and the higher the value, the higher the codon usage frequency [35]. In the *Populus* species, the RSCU values of 29 codons were greater than 1, and 97% of the high-frequency codons end with A/U, indicating that the third base of the chloroplast codon was asymmetrical and there was an A/U bias, which was consistent with the chloroplast genome codon bias of most plants [36].

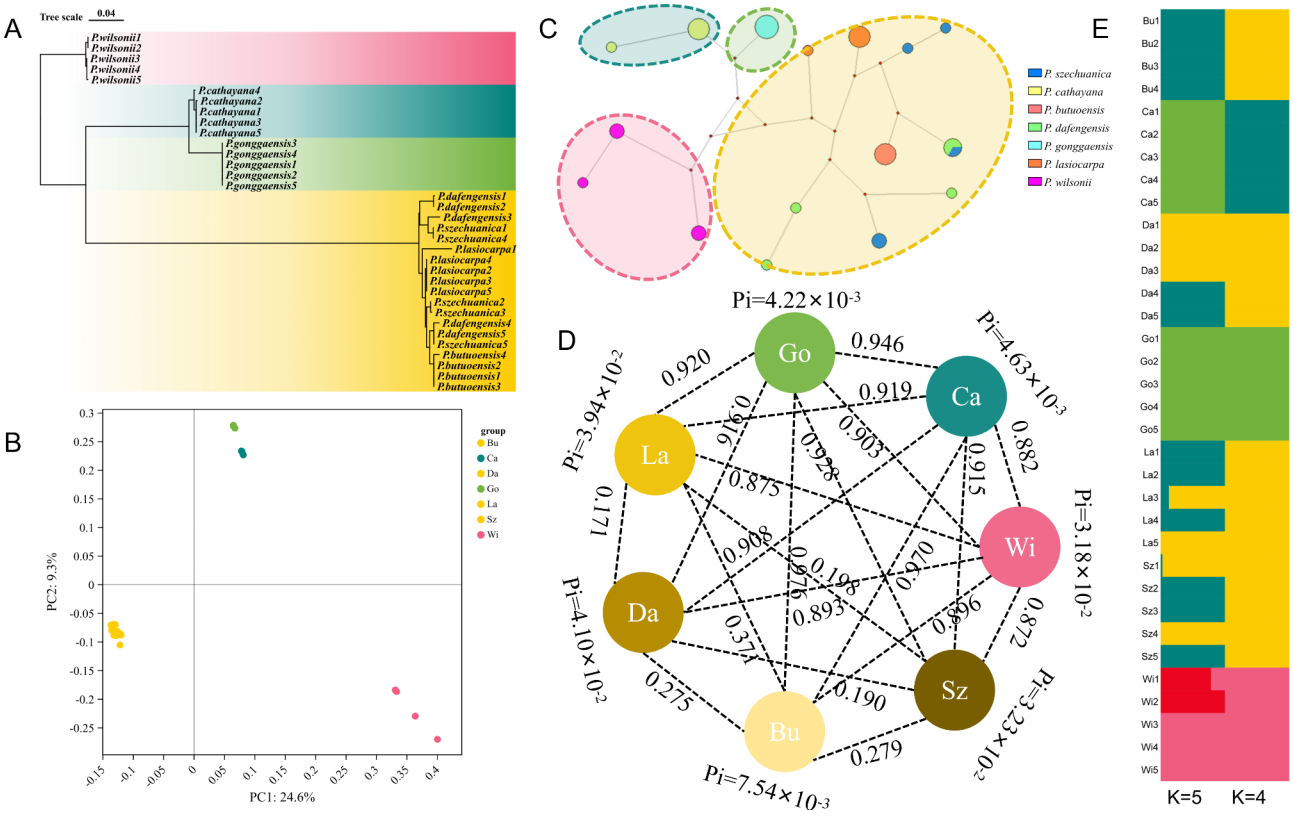

Fig. 8 Population genetic evolution analysis of 34 individuals in 7 *Populus* species. (**A**) Phylogenetic tree based on the complete chloroplast genomes. (**B**) Principal component analysis. (**C**) Haplotype network analysis. (**D**) Intraspecific nucleotide polymorphism and interspecific genetic differentiation coefficient. (**E**) Analysis of genetic components from chloroplast genomes

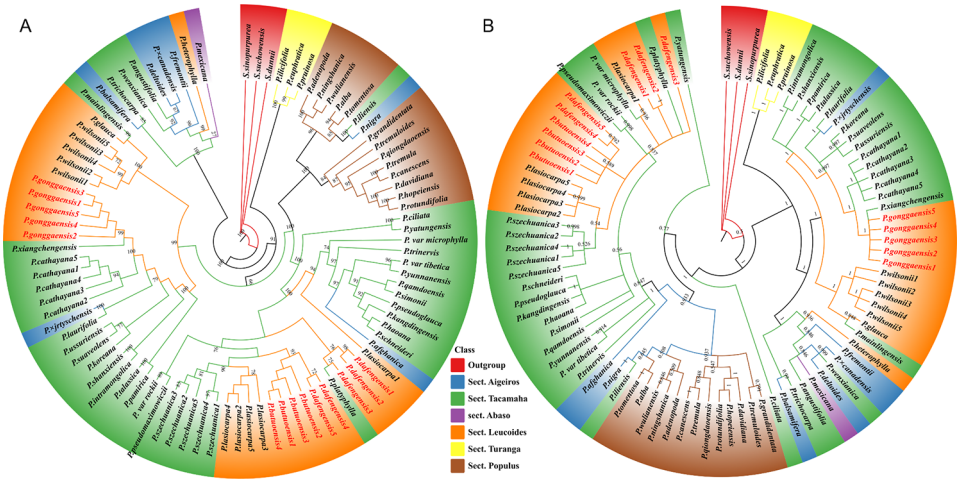

Fig. 9 Phylogenetic analysis of 62 species from 6 sections of *Populus*. (**A**) Phylogenetic tree based on ML method. (**B**) Phylogenetic tree based on BI method. Different sections are given different colors, and three new species are highlighted in red fonts. The support values are displayed on the branches of the tree

Due to the widespread interspecific and intraspecific hybridization and gene introgression of *Populus* species [37, 38], its taxonomic status was controversial in academic circles. Some characters used for morphological classification were affected by environmental and geographical factors, so it was difficult to establish an accurate phylogenetic relationship. Chloroplast genome was a highly conservative genome, which was widely used in phylogenetic reconstruction and population genetic evolution at low classification level [39]. The main purpose of this study was to use the chloroplast genomes of *Populus* species to explore the phylogenetic relationship

between 3 new taxa and known parents and to identify possible matrilineal genetic parent. The results showed that all taxa of *Populus* were divided into 4 sections (Sect. *Turanga*, *Populus*, *Abaso* and ATL). Among them, 3 new taxa had close genetic relationship with 4 parents, *P. gonggaensis* and *P. cathayana* had the closest genetic relationship, *P. butuoensis*, *P. dafengensis* and *P. szechuanica* had the closest genetic relationship, and similar results were also shown in the analysis of genetic structure. Because the chloroplast genome was matrilineal inheritance, it was inferred that *P. cathayana* may be the female parent of *P. gonggaensis*, and *P. szechuanica* may be the female parent of *P. butuoensis* and *P. dafengensis*. Moreover, the genetic differentiation among *P. butuoensis*, *P. dafengensis* and *P. szechuanica* was small, which may be caused by strong gene flows, which was consistent with the inference of the whole genome [10].

Conclusions

In this study, we assembled, annotated, and compared chloroplast genomes of three new *Populus* taxa from Sect. *Leucoides* and four parents. The plastids of the seven *Populus* species were highly conserved, only the genome size, the number of repeats and the IR boundary were slightly different. In addition, we identified three highly variable regions that can provide genetic markers for the identification and phylogenetic reconstruction of related *Populus* species. Phylogenetic analysis showed that *Populus* was a monophyletic group, which could be divided into four sections. In summary, the taxa from Sect. *Tacamahaca* might be the female parent of 3 new species, while the taxa of Sect. *Leucoides* might be the male parent. This study is not only helpful to the development and utilization of *Populus* resources, but also provides favorable data support for subsequent phylogenetic network reconstruction and population genetics research.

Materials and methods

Sample collection, DNA extraction and sequencing

The branches or seedlings of 34 individuals from 7 taxa (including 5 *Populus wilsonii* C. K. Schneider, 5 *Populus lasiocarpa* Olivier, 5 *Populus gonggaensis* N. Chao & J. R. He, 5 *Populus dafengensis* X. Q. Wan & Y. J. Shi, and 4 *Populus butuoensis* X. Q. Wan & Y. J. Shi) were collected in the field (Table S1) and planted in the seeding base of Sichuan Agricultural University (103°38′43″ E, 30°33′24″ N). Among them, the morphological identification of *P. szechuanica*, *P. cathayana*, *P. lasiocarpa* and *P. wilsonii* were based on the description from Flora of China [6], and the morphological identification of *P. gonggaensis* was based on the literature published by Chao. *P. butuoensis* and *P. dafengensis* were new species from Sect. *Leucoides* of *Populus* discovered and identified by X. Q. Wan & Y. J. Shi [10]. The specimens of these 7 taxa were kept in the herbarium of Sichuan Agricultural University, Chengdu, China. After the fresh leaves of the plants were collected, the leaves were frozen with liquid nitrogen, and the whole genomic DNA were extracted by slightly modified CTAB method [40]. The purity, integrity and concentration of DNA were analyzed by gel electrophoresis and Nanodrop. The qualified DNA were selected and sent to Novogene Biotechnology Co., Ltd. (Beijing, China), and the library was constructed and sequenced through the Illumina HiSeq 4000 platform, and the paired end reads of 2×150 bp were produced. The raw reads were preprocessed by Trimmomatic v0.39 software [41] to obtain clean reads, including filtering out reads with adapters, with a proportion of N more than 10% and of low-quality. Clean reads were used for follow-up analysis after quality evaluation by FastQC v0.11.7 software [42], and the basic indicator statistics of the clean reads were recorded in Table S7.

Chloroplast genomes assembly and annotation

First of all, taking the complete chloroplast genome of *P. trichocarpa* (NC_009143.1) as a reference sequence, the chloroplast genome from 34 individuals of 7 *Populus* taxa assembled by Getorganelle v1.7.7 software [43]. Clean reads mapping to chloroplast genome sequences via Samtool v.0.1.19 tool to detect its integrity [44]. Then, the chloroplast genomes were annotated by online tools CPGAVAS2 [45] and Geseq [46], and the annotations information were further corrected manually by Geneious Prime v11.0.18 software [47] according to the reference genome. Finally, the annotated chloroplast genome maps of poplars were visualized by CHLOROPLOT online software and submitted to NCBI database to obtain the accession numbers.

Chloroplast genomes characteristics analysis

The basic characteristic information of seven poplar chloroplast genomes were obtained by CPGView online tool, including the length and GC content of total genome sequences, the length of LSCs, SSCs and IRs sequences, and the number of protein coding genes (PCGs), tRNA genes and rRNA genes. The detailed information was recorded in Table S3.

Codon usage bias analysis

As an important indicator of codon usage bias, the relative synonymous codon usage (RSCU) values of PCGs were calculated by CodonW v1.4.2 software. A RSCU value greater than 1 indicates usage preference, a RSCU value equal to 1 indicates no codon preference, and a RSCU value less than 1 indicates lower codon usage frequency. Three stop codons (TAA, TAG and TGA) were deleted due to lack of degeneracy. This codon usage bias

was also considered to be the result of a combination of natural selection, species mutation and genetic drift [35].

Simple sequence and long sequence repeats identification

The simple sequence repeats (SSRs) in 7 chloroplast genomes of poplar were detected by microsatellite online identification tool MISA-web [48]. The minimum repeat numbers of mononucleotide (Mono-), dinucleotide (Di-), trinucleotide (Tri-), tetranucleotide (Tetra-), pentanucleotide (Penta-) and hexanucleotide (Hexa-) nucleotide sequences were 10, 5, 4, 3, 3 and 3, respectively. Furthermore, we used REPuter online software to identify and locate forward repeats (F-), reverse repeats (R-), complementary repeats (C-) and palindromic repeats (P-) [49]. The following settings for repeat identification were used: Hamming distance equal to 3, minimal repeat size was set to 30 bp, maximum computed repeat was set to 300 bp.

Chloroplast genomes comparison

In order to study the differences of chloroplast genomes from seven species of poplar, taking the chloroplast genome of *P. trichocarpa* as a reference, using mVISTA online tool [50] with shuffle-LAGAN model to visualize their identities.

The PhyloSuit v1.2.3 software [51] was used to extract the PCGs from all chloroplast genomes of 7 poplars. The extracted PCGs first carried on the multi-sequence alignment through the MAFFT v7.427 software [52], then used the KaKs_Calculator 2.0 tool to calculate the nonsynonymous substitution (Ka), the synonymous substitution (Ks), and their ratio (Ka/Ks) with MLWL model [53].

In order to detect the expansion and contraction of the boundary regions of IRs, SSCs and LSCs of chloroplast genomes, we used the IRScope online tool to visualize the junction sites in each region [54].

The structural variation from chloroplast genomes of *Populus* species could be explored by rearranging and comparing the structure of chloroplast genome using Mauve v2.4.0 program in Geneious Prime software and analyzing missing genes, repetition, rearrangement or translocation. Moreover, in order to clarify the level of sequence variation, we used the sliding window method to evaluate chloroplast nucleotide polymorphism (Pi) in DnaSP v6.12.03 software [55]. The window size and step size were set to 600 bp and 200 bp, respectively.

Population evolution and phylogenetic analysis

Based on the complete chloroplast genomes from 34 individuals of 7 poplars, the genetic structure of all samples was inferred by using ADMIXTURE v.1.3.0 software with default parameters [56], and K values were setting between 1 and 7. The K values corresponding to the lowest error rate of cross verification (CV) in the result of structure was interpreted as the best K value. The principal component analysis (PCA) of the population was carried out by Plink v2.0 software [57], and the genetic differentiation coefficient (Fst) and nucleotide polymorphism (Pi) of the population were calculated by pixy program [58].

In order to explore the phylogenetic relationship of 3 new species in *Populus* and their 4 parents, the complete chloroplast genomes of 55 poplars were obtained from NCBI database, and 3 *Salix* were taken as outgroups (Table S2). Firstly, the common PCGs of all species were extracted by PhyloSuit v1.2.3 software, and multiple sequences were compared and trimmed by MAFFT module and trimAl v1.2 software [59]. The sequences after trimming were concatenated into a super sequence, then the phylogenetic tree was constructed by the maximum likelihood (ML) method and Bayesian inference (BI) method, respectively. The ML tree was constructed via IQ-TREE v2.2.0.3 software with 5000 ultra-fast bootstraps [60], and the optimal model (GTR + F + I + G4) was obtained by the ModelFinder program [61]. The MrBayes v3.2.7a software was used to construct BI tree with GTR + I model. The number of chains and runs was set to 4 and 3, respectively. Then, the MCMC was run for 5,000,000 generations, with sampling each 1,000 generations, and the top 25% of the trees aged to discard, then the rest constructed as a 50% consistent tree; convergence was determined by keeping the average standard deviation of the crossover frequencies < 0.01. Finally, the online tool Chipplot was used to visualize the evolutionary trees [62].

We further constructed a haplotype network to explore the genetic relationship among seven poplars. Firstly, the chloroplast genomes of 34 individuals from 7 species were aligned by MAFFT software. Then their haplotype sequences were obtained by DnaSP software, and the haplotype network map was visualized via Network v10.2 [63].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-11099-z.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

S.Y.J. and H.J.L designed the study. Z.W. and W.X.Q. directed the study. Z.W. revised the paper. S.Y.J. analyzed the data and wrote the original draft. S.J.L. and C.Z. participated in the guidance of the analysis. All authors have read and agreed to the published version of the manuscript.

Funding

This work was funded by Basic Public Welfare Research Project of Zhejiang Province (LGN22C020001 to C.Z.), Startup Funding of Taizhou University for

the Biomass Polysaccharide Metabolism Institute (T20231801002 to Z.W.), the National Natural Science Foundation of China (31870645 to W.X.Q.)

Data availability
The newly assembled 7 complete chloroplast genomes of Populus can be obtained on GenBank. The submission IDs are provided temporarily and accession number will be provided after the article is published. The resequencing data have been deposited in the Genome Sequence Archive (GSA) at the NGDC, BIG, CAS / CNCB (GSA: CRA009307), which are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, College of Life Sciences, Taizhou University, Taizhou 318000, China
[2]Sichuan Province Key Laboratory of Ecological Forestry Engineering on the Upper Reaches of the Yangtze River, College of Forestry, Sichuan Agricultural University, Chengdu 611130, China
[3]School of electronics and information engineering, Taizhou University, Taizhou 318000, China

Received: 29 March 2024 / Accepted: 27 November 2024
Published online: 17 February 2025

References

1. Hamzeh M, Dayanandan S. Phylogeny of *Populus* (Salicaceae) based on nucleotide sequences of chloroplast *TRNT-TRNF* region and nuclear rDNA. Amer J Bot. 2004;91:1398–408.
2. Dickmann DI, Kuzovkina J. Poplars and willows of the world, with emphasis on silviculturally important species. Poplars and willows: trees for society and the environment. Wallingford, UK: CABI; 2014. pp. 8–91.
3. Jansson S, Douglas CJ. *Popalus*: A model system for plant biology. Annu Rev Plant Biol. 2007;58:435–58.
4. Zong D, Gan P, Zhou A, Zhang Y, Zou X, Duan A, Song Y, He C. Plastome sequences help to resolve deep-level relationships of Populus in the family Salicaceae. Front Plant Sci. 2019;10:5.
5. Eckenwalder JE. Systematics and evolution of Populus. Ottawa: NRC Research; 1996.
6. Fang ZF, Zhao SD, Skvortsov AK. Populus. In: Wu ZY, Raven PH, editors. Flora of China. Vol. 4. Beijing: Missouri Botanical Garden; 1999.
7. Shi Y-J, Mi J-X, Huang J-L, He F, Chen L-H, Wan X-Q. Study on the origin and classification of two Poplar species on the Qinghai–Tibet Plateau. Forests. 2023;14(5):1003.
8. Zong D, Zhou A, Zhang Y, Zou X, Li D, Duan A, He C. Characterization of the complete chloroplast genomes of five *Populus* species from the western Sichuan plateau, southwest China: comparative and phylogenetic analyses. PeerJ. 2019;7:e6386.
9. Wan XQ, Zhang F. An overview of Populus genetic resources in southwest China. Forestry Chron. 2013;89(1):79–87.
10. Shi Y-J, Huang J-L, Mi J-X, Li J, Meng F-Y, Zhong Y, He F, Tian F-F, Zhang F, Chen L-H, et al. A model of hybrid speciation process drawn from three new poplar species originating from distant hybridization between sections. Mol Phylogenet Evol. 2024;190:107966.
11. Wu S, Wang Y, Wang Z, Shrestha N, Liu J. Species divergence with gene flow and hybrid speciation on the Qinghai–Tibet Plateau. New Phytol. 2022;234:392–404.
12. Wan XQ, Zhang F, Zhong Y, Ding YH, Wang CL, Hu TX. Study of genetic relationships and phylogeny of the native *Populus* in Southwest China based on nucleotide sequences of chloroplast *trnt–trnf* and nuclear DNA. Plant Syst Evol. 2013;299:57–65.
13. Daniell H, Lin C-S, Yu M, Chang W-J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17:1–29.
14. Qiao J, Cai M, Yan G, Wang N, Li F, Chen B, Gao G, Xu K, Li J, Wu X. High-throughput multiplex cpDNA resequencing clarifies the genetic diversity and genetic relationships among Brassica napus, Brassica rapa and Brassica oleracea. Plant Biotechnol J. 2016;14(1):409–18.
15. Liu X, Sun C, Chen Z, Gong X. The complete chloroplast genome sequence of Populus Trinervis, and a comparative analysis with other Populus species. Mitochondrial DNA B Resour. 2020;5(3):2466–8.
16. Zhang Y, Shang H, Liu W, Shan X, Wang N, Peng R. Complete chloroplast genome sequence and phylogenetic analysis of Populus maximowiczii Henry 1913 (Salicaceae Mirb). Mitochondrial DNA B Resour. 2024;9(8):1137–41.
17. Du S. The complete chloroplast genome sequence of Populus Wilsonii based on landscape design, and a comparative analysis with other Populus species. Mitochondrial DNA Part B. 2020;5(3):2716–8.
18. Chen X, Ou CH. The complete chloroplast genome sequence of Populus Szechuanica, a medicinal plant for child anesthesia application. Mitochondrial DNA B Resour. 2021;6(2):696–8.
19. Yang J-B, Tang M, Li H-T, Zhang Z-R, Li D-Z. Complete chloroplast genome of the genus Cymbidium: lights into the species identification, phylogenetic implications and population genetic analyses. BMC Evol Biol. 2013;13:1–12.
20. Song Y, Dong W, Liu B, Xu C, Yao X, Gao J, Corlett RT. Comparative analysis of complete chloroplast genome sequences of two tropical trees Machilus yunnanensis and Machilus balansae in the family Lauraceae. Front Plant Sci. 2015;6:662.
21. Zong D, Zhou A, Zhang Y, Zou X, Li D, Duan A, He C. Characterization of the complete chloroplast genomes of five Populus species from the western Sichuan plateau, southwest China: comparative and phylogenetic analyses. PeerJ. 2019;7:e6386.
22. Zhou J, Zhang S, Wang J, Shen H, Ai B, Gao W, Zhang C, Fei Q, Yuan D, Wu Z, et al. Chloroplast genomes in Populus (Salicaceae): comparisons from an intensively sampled genus reveal dynamic patterns of evolution. Sci Rep. 2021;11(1):9471.
23. Zong D, Gan P, Zhou A, Li J, Xie Z, Duan A, He C. Comparative analysis of the complete chloroplast genomes of seven Populus species: insights into alternative female parents of Populus tomentosa. PLoS ONE. 2019;14(6):e0218455.
24. Wang T, Fan L, Guo X, Luo X, Wang K. Characterization of the complete chloroplast genome of Populus Qiongdaoensis T. Hong Et P. Luo. Conserv Genet Resour. 2016;8(4):435–7.
25. Kersten B, Faivre Rampant P, Mader M, Le Paslier M-C, Bounon R, Berard A, Vettori C, Schroeder H, Leplé J-C, Fladung M. Genome sequences of Populus tremula Chloroplast and Mitochondrion: implications for holistic Poplar breeding. PLoS ONE. 2016;11(1):e0147209.
26. Wu M-l, Li Q, Xu J, Li X-w. Complete chloroplast genome of the medicinal plant Amomum compactum: gene organization, comparative analysis, and phylogenetic relationships within Zingiberales. Chin Med. 2018;13(1):10.
27. Xie F-M, Xie D-F, Xie C, Yu Y, Zhou S-D, He X-J. Adaptation Evolution and Phylogenetic Analyses of Species in Chinese *Allium* Section *Pallasia* and Related Species Based on Complete Chloroplast Genome Sequences. *BioMed Research International 2020*, *2020*: 8542797
28. Hu J, Gui S, Zhu Z, Wang X, Ke W, Ding Y. Genome-wide identification of SSR and SNP markers based on whole-genome re-sequencing of a Thailand wild sacred lotus (Nelumbo nucifera). PLoS ONE. 2015;10(11):e0143765.
29. Kaur S, Panesar PS, Bera MB, Kaur V. Simple sequence repeat markers in genetic divergence and marker-assisted selection of rice cultivars: a review. Crit Rev Food Sci Nutr. 2015;55(1):41–9.
30. Gao X, Zhang X, Meng H, Li J, Zhang D, Liu C. Comparative chloroplast genomes of Paris Sect. Marmorata: insights into repeat regions and evolutionary implications. BMC Genomics. 2018;19(10):133–44.
31. Kimura M. The neutral theory of molecular evolution. Sci Am. 1979;241(5):98–129.
32. Lawrie DS, Messer PW, Hershberg R, Petrov DA. Strong purifying selection at synonymous sites in D. Melanogaster. PLoS Genet. 2013;9(5):e1003527.
33. Wang D, Zhang S, He F, Zhu J, Hu S, Yu J. How do variable substitution rates influence Ka and Ks calculations? Genomics Proteom Bioinf. 2009;7(3):116–27.
34. Boël G, Letso R, Neely H, Price WN, Wong K-H, Su M, Luff JD, Valecha M, Everett JK, Acton TB. Codon influence on protein expression in E. Coli correlates with mRNA levels. Nature. 2016;529(7586):358–63.

35. Sharp PM, Li W-H. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 1987;15(3):1281–95.
36. Duan H, Zhang Q, Wang C, Li F, Tian F, Lu Y, Hu Y, Yang H, Cui G. Analysis of codon usage patterns of the chloroplast genome in Delphinium grandiflorum L. reveals a preference for AT-ending codons as a result of major selection constraints. PeerJ. 2021;9:e10787.
37. Liu S, Zhang L, Sang Y, Lai Q, Zhang X, Jia C, Long Z, Wu J, Ma T, Mao K. Demographic history and natural selection shape patterns of deleterious mutation load and barriers to introgression across *Populus* genome. Mol Biol Evol. 2022;39:msac008.
38. Wang M, Zhang L, Zhang Z, Li M, Wang D, Zhang X, Xi Z, Keefover-Ring K, Smart LB, DiFazio SP. Phylogenomics of the genus Populus reveals extensive interspecific gene flow and balancing selection. New Phytol. 2020;225(3):1370–82.
39. Zhou J, Zhang S, Wang J, Shen H, Ai B, Gao W, Zhang C, Fei Q, Yuan D, Wu Z. Chloroplast genomes in Populus (Salicaceae): comparisons from an intensively sampled genus reveal dynamic patterns of evolution. Sci Rep. 2021;11(1):1–14.
40. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19(1):11–5.
41. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
42. Wingett S, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control [version 2; peer review: 4 approved]. *F1000Research* 2018, 7(1338).
43. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, Li D-Z. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:1–31.
44. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
45. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 2019;47(W1):W65–73.
46. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–11.
47. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.
48. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5.
49. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29(22):4633–42.
50. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004;32(suppl2):W273–9.
51. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20(1):348–55.
52. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
53. 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. 2010;8(1):77–80.
54. Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018;34(17):3030–1.
55. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 2003;19(18):2496–7.
56. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655–64.
57. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. PLINK: a Tool Set for whole-genome Association and Population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.
58. Korunes KL, Samuk K. Pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. Mol Ecol Resour. 2021;21:1359–68.
59. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3.
60. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.
61. Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.
62. Xie J, Chen Y, Cai G, Cai R, Hu Z, Wang H. Tree visualization by one table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. Nucleic Acids Res. 2023;51(W1):W587–92.
63. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16(1):37–48.

Publisher's note