Comparative Analysis of Chloroplast Genome Structure and Phylogenetic Relationships Between a Typical Eucalypt Hybrid and Two Purebred Species

Gupeng Yi
Chinese Academy of Forestry
Guo Liu
liuguopz@163.com

Chinese Academy of Forestry
Jianzhong Luo
Chinese Academy of Forestry
Wanhong Lu
Chinese Academy of Forestry
Yan Lin
Chinese Academy of Forestry
Ying Cheng
Chinese Academy of Forestry
Zhijiao Song
Baoshan University

Research Article

Keywords: eucalypt, chloroplast genome, comparative analysis, codon usage, phylogenetic analysis

Posted Date: March 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9082114/v1

Additional Declarations: No competing interests reported.

Abstract

Background

The chloroplast genome serves a dual function as both a cytoplasmic marker and a functional contributor in hybrid species. A comparative study of chloroplast genomes in a eucalypt hybrid *Eucalyptus urophylla* × *Eucalyptus grandis* (*E. urograndis*) and pure species (*E. urophylla* and *E. grandis*) can provide a theoretical foundation for understanding forest evolution and genetic, it can offer certain technical support for advancements in forest ecological conservation, and biotechnological development.

Results

By employing next-generation sequencing technology, bioinformatics analysis, and other methods, we conducted whole-genome sequencing of the chloroplasts in the hybrids *E. urograndis* and the pure species *E. urophylla*. Our findings revealed that the *E. urograndis* (160,201 bp) had an intermediate chloroplast genome size between that of *E. urophylla* (160,283 bp) and *E. grandis* (160,137 bp). Significant differences was evident in gene composition, IR region expansion, and SSR site distribution. In particular, $trnK^{UUU}$, $trnT^{GGU}$, *psaB-psaA*, *ndhJ-ndhK*, and *rpl22-rps19-rpl2* were identified as significantly different regions. These regions can serve as potential barcode candidates in subsequent studies for species identification, allowing evaluation of their application potential in interspecific discrimination. A set of 14–16 optimal codons was identified in *Eucalyptus*, including GCA, CCA, UAA, GUU, ACA, UCA, CUU, GGU, CAA, UGU, AAU, AGU, GAA, ACU, UUG, and AAA, establishing a crucial foundation for the subsequent optimization of exogenous sequences based on codon usage patterns. This strategy provides essential technical support for advancing research in genetic engineering, trait improvement, and species conservation of *Eucalyptus* plants. Phylogenetic reconstruction confirmed that the eucalypt hybrid formed a monophyletic clade with the pure species *E. urophylla*.

Conclusions

The present study provides theoretical insights into the Structural variations in the chloroplast genome and evolutionary during eucalypt, while enhancing our understanding of interspecific diversity in chloroplast genomes.It also provides a theoretical foundation for sequencing the organelle genomes of eucalypt species, studying variations and evolution in organelle genomes, and developing molecular marker-assisted breeding strategies.

Highlights

1. The *E. urograndis* (160,201 bp) had an intermediate chloroplast genome size between that of *E. urophylla* (160,283 bp) and *E. grandis* (160,137 bp).

2. Analysis of SSRs in the chloroplast genomes of 12 eucalypt species revealed a variation ranging from 68 SSRs in *E. grandis* to 115 in *C. citriodora*.

3. The *psbL*, *psbN*, *infA*, and *pbf1* genes show differences among the three eucalypt species.

4. A set of 14-16 optimal codons was identified in eucalypt species, including GCA, CCA, UAA, GUU, ACA, UCA, CUU, GGU, CAA, UGU, AAU, AGU, GAA, ACU, UUG, and AAA, with a predominance of codons ending in A or U. This finding provides a key foundation for optimizing exogenous sequences based on codon usage patterns.

5. Consistent with established taxonomic frameworks, two *Angophora* species, three *Corymbia* species, and 29 eucalypt specie formed monophyletic clades with bootstrap support values exceeding 80%.

1. Introduction

Eucalypt, which includes the genera *Eucalyptus*, *Angophora*, and *Corymbia* in the Myrtaceae family, is a tree species known for its rapid growth, short rotation period, drought resistance, and broad adaptability. It is widely utilized in pulp and papermaking, plywood production, and solid wood processing, making it a key timber species in southern China's rapidly expanding plantations. According to the latest data, China's eucalypt plantation area has increased to 5.93 million hectares, with an annual wood production of about 40 million cubic meters[1]accounting for more than one-third of the country's total timber output. eucalypt plantations are critical for ensuring China's wood security, boosting economic growth, and promoting rural revitalization. *Eucalyptus urophylla* × *Eucalyptus grandis* (*E. urograndis*), a typical eucalypt hybrid, is produced by crossing *E. grandis* (father) with *E. urophylla* (mother). Initially created in the late 1980s at Guangxi Dongmen Forest Farm, it has since emerged as a key artificial forest species nationwide, owing to its robust adaptability, rapid growth, short growth cycle, ease of vegetative propagation, and favorable trunk morphology.

The chloroplast genomes of higher plants exhibit a highly conserved quadripartite structure, in which two inverted repeat sequences (IR) partition the entire circular chloroplast genome into a large single-copy region (LSC) and a small single-copy region (SSC). The chloroplast genome (cpDNA) [2–4], characterized by its structural conservation, low recombination rate, and typical maternal inheritance pattern, serves as an ideal model for investigating species evolution and hybridization. Although the NCBI database currently includes cpDNA data for 73 eucalypt species (https://ngdc.cncb.ac.cn/cgir/genome?term=eucalyptus, accessed on March 28 2025) and phylogenetic studies have confirmed a deep divergence between *Eucalyptus* and *Corymbia* lineages[5], research in this area remains limited compared to other plant systems. Notably, Research on specific nuclear genes or functional traits of important economic eucalypt species such as *Eucalyptus grandis* and *Eucalyptus regia* has been extensively conducted. However, systematic comparative genomics and evolutionary analysis data at the complete chloroplast genome level remain significantly lacking. Therefore, this work presents the first comparative framework for analyzing the chloroplast genomes of the hybrid species *E. urograndis* and the pure species *E. urophylla, E. grandis*. By conducting a comprehensive analysis of genome architecture, simple sequence repeats (SSRs),synonymous codon usage preference, optimal codon, inverted repeat (IR) boundary variations, nucleotide diversity, and phylogenetic relationships among *E. urophylla*, *E. grandis* and *E. urograndis*, this research aims to clarify the distinctive characteristics and evolutionary patterns of their chloroplast genomes. The results of this work will serve as a scientific basis for the conservation and sustainable utilization of these genetic resources. Furthermore, it provides a theoretical foundation for future studies, including organelle genome sequencing of other eucalypt species, in-depth exploration of organelle genome variations and evolutionary relationships, and the development of molecular marker-assisted breeding strategies.

2. Methods

2.1 Plant materials

We obtained improved varieties of *E. urograndis* (No. Gui S-SC-EUG-001-2009) from the Guangxi State-owned Dongmen Forest Farm (107°51'52"E, 22°19'53"N) and *E. urophylla* from the Jijia Town, Leizhou City, Guangdong Province (109°47'44"E, 20°56'15"N). For both species, healthy mature leaves were collected from the southern and northern sides of three trees exhibiting uniform growth and favorable conditions. These leaves were combined into a

single composite sample for analysis. The same sampling protocol was applied to *E. urograndis* (3-year-old trees) and *E. urophylla* (6-year-old trees). The collected leaf samples were submerged in liquid nitrogen immediately for rapid freezing and stored at -80°C for future analysis.

2.2 DNA extraction, library construction, and sequencing

Total DNA from leaf samples of the two eucalypt species was extracted using the DNAsecure Plant kit (TIANGEN, Beijing, China). Following a quality assessment of the samples, we constructed the library using the MGIEasy Fast Enzyme-Cutting Library Preparation procedures. The NanoDrop 2000 (Thermo Scientific, Wilmington, DE, USA) was used to determine DNA concentrations and quality. Trimmomatic software[6] was used to control the raw reads received from sequencing using the default parameter setting. This was followed by the filtration of the adaptor and low-quality sequences to obtain high-quality clean reads.

2.3 Chloroplast genome splicing, annotation, and physical mapping

The splicing process was divided into three stages: initial splicing, contig screening, and extension, filling, and re-splicing. The limited data (2 × 150 bp, 10 G) prompted the use of GetOrganelle v1.7.7.1 software, which demonstrated superior performance for circular genome splicing. The chloroplast genomes were annotated using CPGAVAS2 software[7], and the physical maps of the chloroplast genomes of *E. urograndis* and *E. urophylla* were created by Organellar Genome DRAW online software[8].

2.4 Repeat sequence analysis

The SSRs of the chloroplast genome were analyzed by MISA software[9] using the following parameters: mononucleotides ≥ 10, dinucleotides ≥ 5, trinucleotides ≥ 4, tetranucleotides ≥ 3, pentanucleotides ≥ 3, and hexanucleotides ≥ 3. The minimum spacing between two SSRs was fixed at 100 bp. The dispersed repeats in the chloroplast genome were detected using REPuter software (http://bibiserv.techfak.uni-bielefeld.de/reputer)[10] Forward (F), palindrome (P), reverse (R), and complementary long repeats (C) were detected with a repeat length of more than 30 bp and a Hamming distance of 3.

2.5 Relative Usage of Synonymous Codons

To improve the accuracy of subsequent codon preference analysis, the codon usage of three chloroplast protein-coding genes from eucalypt species was screened using MEGA 7.0.26 software based on the following criteria: length > 300 bp, start codon ATG, stop codons TGA, TAA, or TAG, and nucleotide count as a multiple of three. Ultimately, 45, 48, and 44 CDS sequences were selected for subsequent data analysis, respectively. Using the online platform EMBOSS Explorer (https://www.bioinformatics.nl/emboss-explorer/), the nucleotide composition percentages of the three eucalypt chloroplast genomes were calculated[11], along with the GC content at the first, second, and third codon positions (denoted as GC1, GC2, and GC3, respectively). Meanwhile, using Rstudio software, genes were ranked from high to low based on their ENC values, and the top and bottom 10% of genes were selected as the high-expression and low-expression libraries, respectively. Codons with ΔRSCU > 0.08 within each library were then identified as high-expression codons based on the genes' RSCU values. Codons meeting both ΔRSCU > 0.08 and RSCU > 1 were defined as optimal codons[12]. Finally, the corresponding results were visualized through plotting.

2.6 Genome comparison and analysis, IR border and divergence analyses

Geneious Prime v2021.1.1[13] was used to determine the chloroplast genome length, GC level, and LSC, SSC, and IRs. It was also used to estimate the number and classification of genes. Chloroplast genomes from the 10 eucalypt species were examined using the IRscope online software[14] to identify shrinkage or expansion at the boundaries.

MAFFT v.7.450 was used to carry out multiple sequence alignment of *E. urograndis* and its parental chloroplast genome sequences [15]. DNAPARS v.6.0 [16]was used to calculate nucleotide polymorphisms (π values) and identify mutation genes and hotspots. The step size and search window length were set at 200 bp and 600 bp, respectively, and the results were plotted in Excel.

2.7 Phylogenetic tree analysis

Table S1 lists species and accessions used in phylogenetic analysis. Thirty-four species were selected to represent major clades, including main sections and series with *Eucalyptus* and related *taxa*. The samples contained 18 *Eucalyptus* species belonging to the subgenus *Symphyomyrtus*, representing nine sections within that subgenus, five species of subgenus *Eucalyptus* that represented four sections, one species of subgenus *Alveolata* (*E. microcorys*), one species of subgenus Cuboidea (*E. tenuipes*), one species of subgenus *Eudesmia* (*E. erythrocorys*), and one species of subgenus *Idiogenes* (*E. cloeziana*). In addition, we also included two *Angophora* species, three *Corymbia* species, and three outgroup species, including *Arabidopsis thaliana* and *Populus trichocarpa*. We utilized MAFFT v7.313 to align the complete chloroplast genome sequences, followed by trimAl to remove indels[17] Maximum likelihood (ML) was used to assess the phylogenetic relationships. Meanwhile, ML reconstruction was performed with IQ-TREE[18] using ModelFinder's best-fit nucleotide substitution model[19]. Ten thousand ultrafast bootstrap replications were used to evaluate branch support for the ML tree [20].

3. Results

3.1 Structural and functional characteristics of three Eucalyptus chloroplast genomes

The chloroplast genomes of *E. urophylla* and *E. urograndis* were sequenced and assembled, whereas the chloroplast genome of *E. grandis* was retrieved from NCBI (Reference sequence: NC_014570.1). The assembled genomes of the three species were similar in size, ranging from 160,137 bp for *E. grandis* to 161,283 bp for *E. urophylla* (Fig. 1).

All three chloroplast genomes were AT-rich (GC level = 36.86–36.89%, Table 1) and had a quadripartite structure similar to many angiosperms, with two inverted repeats (IR) region copies (52780–52798 bp), one large single-copy region (LSC, 88876–88987 bp), and one small single-copy region (SSC, 18481–18498 bp) (Table 1). The GC level in each region differed between the three chloroplast genomes with 42.7%, 30.4%-30.5%, and 34.7%-34.8% variance in IR, SSC, and LSC regions, respectively (Table 1).

Table 1
The complete chloroplast genome features for three *Eucalyptus* species

Species	Size/bp	Number of genes	LSC(bp) /GC	SSC(bp) /GC	IRs(bp) /GC	Total GC%	PCGs/(bp)	rRNA /(bp)	tRNA /(bp)
E. urograndis	160201	127	34.7%	30.4%	42.7%	36.86%	83/78016	8/9056	36/2736
E. grandis	160137	127	34.8%	30.5%	42.7%	36.89%	83/91066	8/9050	36/2745
E. urophylla	160283	129	34.7%	30.4%	42.7%	36.86%	85/85394	8/9050	36/2736

Notes:LSC, large single-copy region; SSC, small single-copy region; IRs, two inverted repeats; PCGs, Protein-coding genes.

3.2 Dispersed repeat and SSR analysis

The *E. urophylla* chloroplast genome included 102 qualified SSRs, comprising 79 mononucleotide SSRs, 4 dinucleotide SSRs, 4 trinucleotide SSRs, and 15 tetranucleotide SSRs (Table S2). The chloroplast genome of *E. grandis* exhibited fewer SSRs (68), which included 48 mononucleotide SSRs, 2 dinucleotide SSRs, 4 trinucleotide SSRs, and 14 tetranucleotide SSRs (Fig. 2). The hybrid *E. urograndis* exhibited identical SSR types and numbers as *E. urophylla*. SSRs were predominantly distributed in the LSC region of all three species, with fewer in the IRs and SSC regions. Mononucleotide SSRs accounted for the highest proportion, while dinucleotide, trinucleotide, and tetranucleotide SSRs were relatively less common. Furthermore, SSRs in the chloroplast genomes of all three *Eucalyptus* species included predominantly adenine (A) and thymine (T). *E. urophylla* possessed 4 dinucleotide SSRs with A/T and T/C motifs, 4 trinucleotide SSRs with ATT/TAA motifs, and 15 tetranucleotide SSRs rich in A and T. *E. grandis* contained 2 dinucleotide SSRs (A/T and T/C motifs), 5 trinucleotide SSRs (ATT/TAA motifs), and 13 A/T-rich tetranucleotide SSRs. The hybrid *E. urograndis* shared the same nucleotide repeat composition as *E. urophylla*.

Dispersed repeats are a type of repeat sequence interspersed throughout the genome. They are classified into four types: forward, palindromic, reverse, and complementary. *E. urophylla* comprised 31 dispersed repeats, including 16 forward and 15 palindromic (Fig. 3). The longest forward repeat spanned 518 bp (0.32% of the genome), while the longest palindromic repeat measured 26,923 bp (16.81%). *E. grandis* shared the same number of dispersed repeats (31) as *E. urophylla*, with the longest forward repeat at 524 bp (0.32%) and the longest palindromic repeat at 26,838 bp (16.80%). The hybrid *E. urograndis* included 32 dispersed repeats, comprising 17 forward and 15 palindromic repeats. The longest palindromic repeat was 26,923 bp long, accounting for 16.80% of the genome. However, no reverse or complement repeats were observed in the three *Eucalyptus* species (Fig. 3).

3.3 Comparative analysis of SSR among eucalypt species

Analysis of SSRs in the chloroplast genomes of 12 eucalypt species revealed a variation ranging from 68 SSRs in *E. grandis* to 115 in *C. citriodora* (Fig. 4A). The SSR composition included 48 to 87 mononucleotide repeats, 1 to 4 dinucleotide repeats, 3 to 8 trinucleotide repeats, 4 to 17 tetranucleotide repeats, 0 to 1 pentanucleotide repeats, and 0 to 1 hexanucleotide repeats. These SSRs were predominantly located in the LSC region of the chloroplast genomes. Tetranucleotide repeats were generally more abundant than trinucleotide repeats across the 12 species, whereas pentanucleotide and hexanucleotide repeats were relatively rare. *E. pauciflora* exhibited the highest proportion of mononucleotide repeats (81.74%). Four species—*E. urograndis*, *E. urophylla*, *E. nitens*, and *E. robusta*—shared an identical SSR composition, each containing 79 mononucleotide repeats, along with 4 dinucleotide and 4 trinucleotide repeats. *C. citriodora* was the most distinctive species, uniquely possessing both pentanucleotide and hexanucleotide repeats among the 12 examined species, whereas all other species contained only the other four repeat types.

Dispersed repeats represent a distinct category of repetitive sequences, distinguished by their scattered distribution throughout the genome. These repeats are classified into four types: Forward (F), Palindromic (P), Reverse (R), and Complement (C). An analysis of dispersed repeats within the chloroplast genomes of 12 eucalypt species (Fig. 4B) demonstrated that forward and palindromic repeats were predominat, whereas complement repeats were the least prevalent. Only *C. citriodora* and *E. nitens* exhibited all four repeat types (F, P, R, C) in their chloroplast genomes; the remaining ten species lacked complement repeats. *E. globulus* displayed a notable disparity in repeat counts, with seven more forward repeats than reverse repeats. In other species, the numbers of forward and palindromic repeats were generally comparable. Complement repeats remained consistently rare across all 12 species, occurring exclusively in *C. citriodora* and *E. nitens*.

3.4 Analysis of Relative Synonymous Codon Usage (RSCU)

The total average GC1, GC2, and GC3 values of the chloroplast genome CDS codons of *E. urograndis*, *E. urophylla*, and *E. grandis* are 38.4%, with individual averages of 47.14%, 38.67%, and 29.387%, respectively, showing a pattern of GC1 > GC2 > GC3. Using an ENC value of 45 as the threshold for bias classification, the overall average ENC values for 45, 48, and 44 CDS codons in the chloroplast genomes of the three eucalypt species range from 48.83 to 49.41, all exceeding 45(Table 2). This indicates a relatively weak codon usage bias in the chloroplast genomes of these three species.

Table 2
Genomic features of chloroplast genomes of three *Eucalyptus* plant species.

Parameters	*E. urograndis*	*E. urophylla*	*E. grandis*
CDS number (before processing)	83	85	83
CDS number (after processing)	45	48	44
L_aa	51687	58938	44406
GC1(%)	47.54	46.02	47.86
GC2(%)	38.9	38.03	39.08
GC3(%)	29.56	28.74	29.86
ENc	49.41	48.83	49.36
Average GC(%)	38.67	37.6	38.93

Note: L_aa means the total number of amino acids; GC1, GC2 and GC3 indicate the GC content at the first, second and third codon positions.

Based on the analysis criteria of RSCU values (where RSCU > 1 and > 1.5 represent relatively high and high-frequency codons, respectively), the chloroplast genomes of the three *Eucalyptus* species exhibit similar codon usage preferences, with their RSCU values for all encoded amino acids primarily distributed in the range of 0–3 (Fig. 5). *E. urograndis*, *E. urophylla*, and *E. grandis* each have 30 high-frequency codons with RSCU values greater than 1, accounting for 46.87% of the total codons. Among these high-frequency codons, only one ends with G, while the rest end with A or U. The results indicate that the codon usage in the chloroplast genomes of the three *Eucalyptus* species tends to favor A or U endings. Additionally, in the chloroplast genomes of the three ucalypt species, the codon UUA for leucine (Leu) acts as an overlapping codon and is used more frequently than other codons. Specifically, the RSCU values are 1.91 for *E. urograndis*, 1.98 for *E. urophylla*, and 1.91 for *E. grandis.*

3.5 Analysis of Optimal Codons

In *E.urograndis*, 14 optimal codons were identified with the highest ΔRSCU value of 1.243 and an average ΔRSCU value of 0.535 (Fig. 6). For *E. urophylla*, the highest number of optimal codons was observed, totaling 16, along with a maximum ΔRSCU value of 1.343 and an average ΔRSCU value of 0.417. In *E. grandis*, 14 optimal codons were identified, matching the highest ΔRSCU value observed in *E. urograndis*, and an average ΔRSCU value of 0.446. Across the three *Eucalyptus* species, a total of 14–16 optimal codons were identified, including GCA, CCA, UAA, GUU, ACA, UCA, CUU, GGU, CAA, UGU, AAU, AGU, GAA, ACU, UUG, and AAA(Fig. 7).

3.6 IR border and divergence analyses

In the chloroplast genomes of the ten eucalypt species (Fig. 8), variations were observed in the positions of the *rpl22*, *rps19*, and *rpl2* genes at the IR boundaries, as well as in their gene lengths. The *rpl22* and *rps19* genes were located at the IR/LSC boundary (JLA) in *E. urophylla*, positioned to the right of the junction, with *rps19* located 1 bp from the boundary. In *E. grandis*, these genes were found at the LSC/IRb boundary (JLB), positioned to the left of the junction, with *rps19* located 2 bp from the boundary. *E. urograndis*, both genes were located at the SSC/IRA boundary (JSA), positioned to the left of the junction, with *rps19* located 2 bp from the boundary. The *rpl2* gene was located to the right of the JSA junction in *E. urograndis* but was not detected in *E. grandis*. In the remaining eight species, *rpl2* consistently appeared to the right of the IR/LSC boundary (JLA). The *ndhF* gene was located 187 bp to the right of the JLA junction in *E. urograndis*. It was located 187 bp and 94 bp to the right of the JSB junction in *E. urophylla* and *C. citriodora*, respectively, but was absent in the other seven species.

To identify sequence variation sites within the three *Eucalyptus* chloroplast genomes, we calculated nucleotide diversity (Pi) values (range, 0-0.34889; average, 0.21225) (Table S3, Fig. 9). We detected five highly divergent regions (*trnK*^{UUU}, *trnT*^{GGU}, *psaB-psaA*, *ndhJ-ndhK*, and *rpl22-rps19-rpl2*), with nucleotide diversity exceeding 0.34. The *trnT*^{GGU} gene exhibited the most significant divergence, the highest Pi value (0.34889), and is located in the LSC region. These highly variable regions could be candidate molecular biomarkers for phylogenetic reconstruction in eucalypt species.

3.7 Phylogenomic analysis

We constructed a ML phylogenetic tree using chloroplast genomes of *E. urophylla*, *E. grandis*, *E. urograndis*, and 29 published eucalypt species, with *A. thaliana* and *P. trichocarpa* as outgroups (Fig. 10). The results revealed that the outgroups (*A. thaliana* and *P. trichocarpa*) differed significantly from the 32 eucalypt species. Consistent with established taxonomic frameworks, two *Angophora* species, three *Corymbia* species, and 29 eucalypt specie formed monophyletic clades with bootstrap support values exceeding 80%. The 29 eucalypt species were classified into six subgenera, with chloroplast genome sequences effectively separating most except *Eucalyptus queenslandica*. This is consistent with previous studies by Liu et al[21] and Steane et al[22], which found a close genetic relationship between the subgenus *Eucalyptus* and *Monocalyptus*. this work found that the two eucalypt hybrids (*E. sideroxylon* × *E. melliodora* and *E. urograndis*) were genetically similar to their maternal parents, which is consistent with the maternal inheritance pattern outlined in eucalypt chloroplast genomes. Within the subgenus *Symphyomyrtus*, *E. camaldulensis* from *Exsertaria* exhibited closer genetic proximity to species from *Transversaria* and section *Globulus*. This finding provides a theoretical foundation for using interspecific hybridization among these three divisions to develop superior hybrids in breeding programs. However, the genetic similarity among species from various sections within the same subgenus differs from the most recent published classification of the *Eucalyptus* genus [23]. For example, within the subgenus *Eucalyptus*, *E. regnans*, *E. elata*, and *E. baxteri* are classified under the section *Eucalyptus*, yet in this work, *E. regnans* exhibited a closer genetic affinity to *E. marginata* (section *Longistylus*) than to its section members. This discrepancy may reflect the limitations of chloroplast genomes-inherited maternally and highly conserved-in resolving fine-scale taxonomic relationships [24].

4. Discussion

The elucidation of chloroplast genomes lays the groundwork for in-depth research into chloroplast functions, intracellular gene transfer, species diversity, plant resource conservation, and genetic breeding [25]. Our study sequenced, assembled, and annotated the chloroplast genomes of *E. urograndis* (a widely used hybrid in Chinese eucalypt plantations) and the pure species *E. urophylla*, the complete chloroplast genome data were collected and compared to the published *E. grandis* chloroplast genome(NC_014570.1). *E. urophylla*, *E. grandis*, and *E. urograndis* exhibited the typical angiosperm quadripartite structure in their chloroplast genomes, with sizes ranging from 160,137 bp to 160,283 bp. This finding is consistent with the results of Bayly et al., which reported minimal size variation

among chloroplast genomes in 39 eucalypt species [5]. *E. urograndis*, *E. grandis*, and *E. urophylla* chloroplast genomes comprised 127, 127, and 129 genes (Table 1), respectively, including 83, 83, and 85 protein-coding genes, 36 tRNA genes, and 8 rRNA genes. These findings are consistent with previously reported eucalypt chloroplast genomes[26–28], demonstrating a conserved chloroplast genome organization across the genus. The hybrid *E. urograndis* was more closely related to *E. urophylla*.

The *psbL* gene was detected in the chloroplast genomes of *E. urograndis* and *E. grandis* but not in *E. urophylla*. The *psbN* gene is retained in *E. urophylla* and *E. grandis*, but lost in *E. urograndis*, most likely due to evolutionary gene loss. This is consistent with the findings of Zhang et al.[29] who demonstrated a high conservation of *psbN* in angiosperms. Due to their distinct life-history strategies, parasitic plants may lose or modify certain chloroplast genes (including *psbN*) [30],potentially reflecting ecological adaptations. This work found variations in the *psbL*, *psbN*, *infA*, and *pbf1* genes between *E. urograndis*, *E. urophylla* and *E. grandis*. These discrepancies may arise from structural variations in the chloroplast genome or from adaptive evolutionary processes. Alternatively, chloroplast genome of the hybrid could have been inherited from an as-yet uncharacterized parental lineage.These factors may operate independently or in combination, leading to subtle. Future research will focus on expanding the sample set and validating these genomic regions.

SSRs, which are highly variable, drive chloroplast genome rearrangement and expansion. this work revealed 68 to 102 SSRs in the three *Eucalyptus* chloroplast genomes (Fig. 2). Notably, the hybrid *E. urograndis* harbored one additional forward repeat compared to other two eucalypt, with the longest forward repeats varying by 48 and 42 bp (Fig. 3), perhaps due to IR contraction during evolution. Mononucleotide SSRs dominated the repeat composition, providing a foundation for interspecies genomic variation studies in eucalypt. Under environmental stress[31], SSR dynamics may alter gene expression and influence adaptive capacity. Significant differences in SSR abundance and distribution patterns between *E. urograndis*, *E. urophylla*, and *E. grandis* demonstrate their potential as molecular biomarkers for species identification and phylogenetic reconstruction in eucalypt.

The chloroplast genomes of the three *Eucalyptus* species exhibit weak codon usage bias, and their preference pattern aligns with that of most plant chloroplasts, showing a significant tendency to use codons ending with A or U. This finding is consistent with previous studies on Cupressaceae [32], Theaceae [33], and *Davidia involucrata* [34]. Moreover, a total of 14–16 optimal codons were identified across the three *Eucalyptus* species, including GCA, CCA, UAA, GUU, ACA, UCA, CUU, GGU, CAA, UGU, AAU, AGU, GAA, ACU, UUG, and AAA, with a predominance of codons ending in A or U (Fig. 7). This pattern shows a high degree of similarity to the codon preferences observed in *Euphorbia esula* [35] and the *Hevea genus* [36]. Studies have demonstrated that codon usage bias can serve as an effective molecular marker for distinguishing closely related species, such as bamboos[37]. The results of this study provide a strong reference for the accurate classification and identification of eucalypt species, aiding in the recognition of their genetic diversity and characteristic differences. Additionally, leveraging their codon usage patterns to optimize exogenous sequences can effectively enhance gene expression efficiency, offering technical support for subsequent research on genetic engineering improvements and species conservation in the genus *Eucalyptus*.

IR region contraction and expansion in plant chloroplast genomes are common phenomena that determine genome size variation across species[38]. We observed that *E. urograndis* had a shorter LSC region than *E. urophylla* and *E. grandis*, with LSC variances accounting for the majority of the differences in genome size (Fig. 5). It is hypothesized that variations in the LSC region length are primarily responsible for the observed differences in the chloroplast genome size of *E. urograndis*. Studies indicate that longer IR regions offer additional repetitive sequences[39], improve genome stability, and lower the chances of structural rearrangements[40]. Our results demonstrate that *E. urograndis* has a modest IR expansion compared to other two species of eucalypt, which may confer increased

genomic stability and adaptability under environmental stress. The contraction-expansion divergence at IR boundaries in three species of *Eucalyptus* likely contributes to chloroplast genome size variance, which may affect overall genome architecture and function. These findings are consistent with observations in *Paeonia* species, where alterations in the LSC/IRb boundary positions correlate with evolutionary divergence[41]. Minor boundary variations may thus serve as evolutionary markers for chloroplast genome differentiation[42].

We discovered five highly divergent regions between the chloroplast genomes of eucalypt hybrid and pure species, predominantly located in the LSC region. The nucleotide diversity of IR regions was significantly lower than the SSC regions (Fig. 7), consistent with prior findings[43]. Zhang et al.[44] conducted a similar study, comparing entire chloroplast genomes in 24 Myrtales species using MVISTA and variable locus proportions across 114 non-coding regions and 74 coding genes with DnaSP. The study demonstrated that IR regions are less divergent than SC regions in Asteraceae and Myrtales. Variations were also observed in genes near the IR/SSC boundaries, such as *rpl22*, *rps19*, and *rpl2*, consistent with the results of the IRscope boundary analysis.

Previous studies have shown that the chloroplast genome can be used for species identification and phylogenetic reconstruction [45–47]. Our findings reveal significant divergence in the chloroplast genomes of *Angophora*, *Corymbia*, and *Eucalyptus*, allowing for the unambiguous distinction between the three genera. However, chloroplast genomes cannot discriminate between species in the subgenera *Eucalyptus* and *Monocalyptus*, which share closer genetic proximity. However, chloroplast genomes continue to be effective in elucidating phylogenetic relationships for other subgenera. As the largest *Eucalyptus* subgenus, *Symphyomyrtus* exhibits close genetic distances across its sections, allowing for interspecific hybridization between section members to produce high-performance hybrids and clonal variants. For instance, hybrids derived from crosses between sections Latoangulata (*E. urophylla*, *E. grandis*, *E. pellita*), Exsertaria (*E. camaldulensis*, *E. tereticornis*), and Maidenaria (*E. dunnii*, *E. globulus*) have demonstrated superior heterosis in plantation applications [6][48–51]. this study also revealed close genetic relationships among the sections *Globulus, Exsertaria*, and *Transversaria*. The tree species within these three sections may exhibit higher hybrid compatibility, which could facilitate the successful production of hybrid offspring. However, this inference, based on genomic similarity, requires further functional validation through direct hybridization experiments.

5. Conclusions

As previously mentioned, eucalypt species have highly conserved boundaries and similar flanking genes but with slightly different expansion degrees. The only variations detected were in genome size, GC level, repetitive sequences, and IR boundary. In conclusion, we propose five hypervariable regions, $trnK^{UUU}$, $trnT^{GGU}$,*psaB-psaA*, *ndhJ-ndhK*, and *rpl22-rps19-rpl2*, as candidate molecular biomarkers for identifying *E. urograndis*. A set of 14–16 optimal codons was identified in eucalypt species, establishing a crucial foundation for the subsequent optimization of exogenous sequences based on codon usage patterns. This strategy provides essential technical support for advancing research in genetic engineering, trait improvement, and species conservation of eucalypt plants. These findings enrich the chloroplast genomic database for eucalypt species and provide a theoretical foundation for reconstructing their phylogenetic relationships. this study increases our understanding of interspecific diversity in chloroplast genomes between eucalypt species .

Abbreviations

A
Adenine
C

Complementary
F
Forward
IR
Inverted repeats
JLA
IRA/LSC
JLB
LSC/IRb
JSA
SSC/IRA
JSB
IRB/LSC
LSC
Large single-copy region
ML
Maximum likelihood
P
Palindromic
PCGs
Protein-coding genes
Pi
Nucleotide diversity
R
Reverse
SSC
Small single-copy region
SSRs
Simple sequence repeats
T
thymine.

Declarations

Acknowledgments

Not applicable.

Funding

This research was funded by the Fundamental Research Funds of CAF (Grant No. CAFYBB2023MB034) and the National Key R&D Program of China (Grant No. 2022YFD2200203 and 2023YFD2201001).

Availability of data and materials

The datasets supporting the results of this article are available on NCBI, https://www.ncbi.nlm.nih.gov/ (accession number: PV464070 and OL804288).

Authors' contributions

GupengYi: Writing-original draft, Methodology, Investigation, Formal analysis, Data curation. **Wanhong Lu** and **Yan Lin:** Software, Methodology, Investigation. **Jianzhong Luo** and **Guo Liu:** Writing-review and editing, Supervision, Funding acquisition, Conceptualization. **Ying Cheng** and **Zhijiao Song:** Visualization, Validation, Methodology, Investigation, Conceptualization.

Supplementary information

Table S1. 34 species and accessions used for phylogenetic analysis; Table S2 .SSR information in the chloroplast genomes of tewlve eucalypt species; Table S3. Sliding window test of Pi in the Hybrid tail *E.urograndis*, *E.urophylla* and *E.grandis* chloroplast genomes.

Ethics approval and consent to participate

The appropriate permissions were obtained for all materials used in this work. We complied with all relevant institutional, national, and international guidelines and legislation. The Guangxi State-owned Dongmen Forest Farm (Chongzuo, Guangxi Zhuang Autonomous Region, China) and Leizhou Forestry Bureau (Zhanjiang, Guangdong Province, China) gave permission for the leaves of *E. urograndis* and *E. urophylla* to be collected for the study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Liu T, Xie YJ. Analysis and prospects of the rapid development of Eucalyptus plantations in China. Eucalypt Sci Technol. 2020;37(4):38–47.
2. Maheswari P, Kunhikannan C, Yasodha R. Chloroplast genome analysis of Angiosperms and phylogenetic relationships among Lamiaceae members with particular reference to teak (*Tectona grandis* L.f). J Biosci. 2021;46:43. https://doi.org/10.1007/s12038-021-00166-2.
3. Zhang J, Yuan Q, Wei Y, Liu H, Zheng X, Wang Y, Liu H. Evolutionary analysis of chloroplast genomes of *Sect.Isika* (Lonicera)species, Nematotinus plants. Chin Tradit Herb Drugs. 2024;55(9):3085–97.
4. Yan A, Zhu D. Application of chloroplast genome in phylogenetics and genetic engineering. Chin J Cell Biology. 2004;26:153–6.
5. Bayly MJ, Rigault P, Spokevicius A, Ladiges PY, Ades PK, Anderson C, Bossinger G, Merchant A, Udovicic F, Woodrow IE, Tibbits J. Chloroplast genome analysis of Australian eucalypts - *Eucalyptus*, *Corymbia*, *Angophora*, *Allosyncarpia* and *Stockwellia* (Myrtaceae). Mol Phylogenet Evol. 2013;69(3):704–16. https://doi.org/10.1016/j.ympev.2013.07.006.
6. Guo Y, Yao B, Yuan M, Li J. The complete chloroplast genome and phylogenetic analysis of Astragalus scaberrimus Bunge 1833. Mitochondrial DNA Part B. 2021;6(12):3364–6. https://doi.org/10.1080/23802359.2021.1997108.

7. 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. https://doi.org/10.1093/nar/gkz345.
8. Zhu B, Gao Z, Qian F, Yang X, Lv X, Cai M. The complete chloroplast genome of a purple Ethiopian rape (Brassica carinata: Brassicaceae) from Guizhou Province, China and its phylogenetic analysis. Mitochondrial DNA Part B. 2021;6:1821–3. https://doi.org/10.1080/23802359.2021.1926365.
9. Wang W, Yu H, Wang J, Lei W, Gao J, Qiu X, Wang J. The complete chloroplast genome sequences of the medicinal plant Forsythia suspensa (Oleaceae). Int J Mol Sci. 2017;18(11):2288. https://doi.org/10.3390/ijms18112288.
10. 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. https://doi.org/10.1093/nar/29.22.4633.
11. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet. 2000;16(6):276–7.
12. Wu P, Xiao W, Luo Y, Xiong Z, Chen X, He J, Sha A, Gui M, Li Q. Comprehensive analysis of codon bias in 13 Ganoderm a mitochondrial genomes. Fron-tiersin Microbiol. 2023;14:1170790. https://doi.org/10.3389/fmicb.2023.1170790.
13. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9. https://doi.org/10.1093/bioinformatics/bts199.
14. Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018;34(17):3030–1. https://doi.org/10.1093/bioinformatics/bty220.
15. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2017;20(4):1160–6. https://doi.org/10.1093/bib/bbx108.
16. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34:3299–302. https://doi.org/10.1093/molbev/msx248.
17. 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. https://doi.org/10.1093/bioinformatics/btp348.
18. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74. https://doi.org/10.1093/molbev/msu300.
19. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9. https://doi.org/10.1038/nmeth.4285.
20. Minh BQ, Nguyen MAT, Von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013;30(5):1188–95. https://doi.org/10.1093/molbev/mst024.
21. Liu G, Arnold RJ, Xie Y, Wu Z. Genetic relationships among 40 species of Eucalyptus based on simple sequence repeat markers. J Trop For Sci. 2018;30(3):402–14. https://www.jstor.org/stable/26512525.
22. Steane DA, Nicolle D, McKinnon GE, Vaillancourt RE, Potts BM. Higher-level relationships among the eucalypts are resolved by ITS-sequence data. Aust Syst Bot. 2002;15(1):49–62. https://doi.org/10.1071/SB00039.
23. Nicolle D. Classification of the eucalypts, genus Eucalyptus.Version 7[EB/OL].(2025-4-16). http://www.dn.com.au/Classification-Of-The-Eucalypts.pdf

24. Zhang L, Morales-Briones DF, Li Y, Zhang G, Zhang T, Huang CH, Ma H. Phylogenomics insights into gene evolution, rapid species diversification, and morphological innovation of the apple tribe (Maleae, Rosaceae). New Phytol. 2023;240(5):2102–20. https://doi.org/10.1111/nph.19175.
25. Liu C, Tang L, Han L. Characterization of the chloroplast genome of Lindera setchuenensis and phylogenetics of the genus Lindera. Scientia Silvae Sinicae. 2021;57(7):167–74. 10.11707/j.1001-7488.20211217. http://www.linyekexue.net/EN/.
26. Wang WW, Schalamun M, Morales-Suarez A, Kainer D, Schwessinger B, Lanfear R. Assembly of chloroplast genomes with long- and short-read data: a comparison of approaches using *Eucalyptus pauciflora* as a test case. BMC Genomics. 2018;19(1):977. https://doi.org/10.1186/s12864-018-5348-8.
27. Gao Y, Zhang J, Tang H, Liu N, Li G, Yue D. The characteristics of the complete chloroplast genome for *Eucalyptus robusta* (Myrtaceae). Mitochondrial DNA Part B. 2021;6(12):3517–8. https://doi.org/10.1080/23802359.2021.2005491.
28. Steane DA. Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). DNA Res. 2005;12(3):215–20. https://doi.org/10.1093/dnares/dsi006.
29. Zhang H, Huang T, Zhou Q, Sheng Q, Zhu Z. Complete chloroplast genomes and phylogenetic relationships of *Bougainvillea spectabilis* and *Bougainvillea glabra* (Nyctaginaceae). Int J Mol Sci. 2024;24(16):13044. https://doi.org/10.3390/ijms241713044.
30. Gao L, Su Y, Wang T. Plastid genome sequencing, comparative genomics, and phylogenomics: Current status and prospects. J Syst Evol. 2010;48(1):77–93. https://doi.org/10.1111/j.1759-6831.2010.00071.x.
31. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17(1):134. https://doi.org/10.1186/s13059-016-1004-2.
32. Huang SQ, Zhang QG, Ye ZL, Qin XH, Wang QS, Lin XQ, Zheng ZW, Lin WF, Zou XX. Codon bias analysis of chloroplast genomes of 5 Cupressaceae plants. J Fujian Agric Forestry Univ (Natural Sci Edition). 2024;53(2):214–20.
33. Wang Z, Cai Q, Wang Y, Li M, Wang C, Wang Z, Jiao C, Xu C, Wang H, Zhang Z. Comparative analysis of codon bias in the chloroplast genomes of theaceae species. Front Genet. 2022;13:824610. https://doi.org/10.3389/fgene.2022.824610.
34. Luo YJ, Wang R, Zhao RF, Lu XX, Yin GK, Deng ZJ. Analysis of synonymous codon usage bias in the chloroplast genome of Davidia involucrata. J Beijing Forestry Univ. 2024;46(3):8–16.
35. Wang Z, Xu B, Li B, Zhou Q, Wang G, Jiang X, Wang C, Xu Z. Comparative analysis of codon usage patterns in chloroplast genomes of six Euphorbiaceae species. 2020;8:e8251. https://doi.org/10.7717/peerj.8251
36. Yang Y, Liu X, He L, Li Z, Yuan B, Fang F, Wang X. Comparative chloroplast genomics and codon usage bias analysis in Hevea Genus. Genes. 2025;16(2):201. https://doi.org/10.3390/genes16020201.
37. Kelchner SA, Group BP. Higher level phylogenetic relationships within the bamboos (Poaceae: Bambusoideae) based on five plastid markers. Mol Phylogenet Evol. 2013;67(2):404–13.
38. Huang H, Shi C, Liu Y, Mao S, Gao L. Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: Genome structure and phylogenetic relationships. BMC Evol Biol. 2014;14(1):151. https://doi.org/10.1186/1471-2148-14-151.
39. Jiang Z, Chen H, Bao H, Dai Y. Chloroplast genome characteristics and molecular marker development of Pennisetum. J Zhejiang A&F Univ. 2025;42(2):365–72.
40. Wang Y, Wang S, Liu Y, Yuan Q, Sun J, Guo L. Chloroplast genome variation and phylogenetic relationships of Atractylodes species. BMC Genomics. 2021;22(1):103. https://doi.org/10.1186/s12864-021-07394-8.

41. Zhou X, Zhang K, Peng Z, Sun S, Ya H, Zhang Y, Chen Y. Comparative Analysis of Chloroplast Genome Characteristics between *Paeonia jishanensis* and Other Five Species of Paeonia. CABI Digit Libr. 2020;56(4):82–8.
42. Li Q, Yan N, Song Q, Guo J. Complete chloroplast genome sequence and characteristics analysis of Morus multicaulis. Chin Bull Bot. 2018;l53(1):94–103. 10.11983/CBB16247. https://www.chinbullbotany.com/EN/.
43. Sun J, Wang Y, Garran TA, Qiao P, Wang M, Yuan Q, Guo L, Huang L. Heterogeneous genetic diversity estimation of a promising domestication medicinal motherwort Leonurus cardiaca based on chloroplast genome resources. Front Genet. 2021;12:721022. https://doi.org/10.3389/fgene.2021.721022.
44. Zhang X, Landis JB, Wang H, Zhu Z, Wang H. Comparative analysis of chloroplast genome structure and molecular dating in Myrtales. BMC Plant Biol. 2021;21(1):219. https://doi.org/10.1186/s12870-021-02985-9.
45. Li C, Lu J, Li Y, Zhu J, Zhang L. Comparative morphology of the leaf epidermis in Lobelia (Lobelioideae) from China. Microsc Res Tech. 2017;80(7):763–78. https://doi.org/10.1002/jemt.22862.
46. Ding Y, Bi G, Hu S, Zhou MY, Li H, Xia Z. Chloroplast genome characteristics and phylogenetic analysis of different flower color variation types in safflower (*Carthamus tinctorius*). Chin Traditional Herb Drugs. 2023;54(1):262–71.
47. Yu B, Sun Y, Liu X, Huang L, Xu Y, Zhao C. Complete chloroplast genome sequence and phylogenetic analysis of Camellia fraterna. Mitochondrial DNA Part B. 2020;5(3):3840–2. https://doi.org/10.1080/23802359.2020.1841576.
48. Li G, Xu J, Wang W, Wu S, Zhu Y, Wang YS, Pan J, Guo HY, Shi TY. Study on heterosis estimation and genetic analysis of Eucalyptus hybrids in cold area. J Nanjing Forestry Univ (Natural Sci Edition). 2017;41(4):55–63. http://njlydxxb.periodicals.net.cn/default.html.
49. Mo JY, Lan J, Luo JZ, Wu MFn, Peng ZB. Growth characteristics of hybrids between *Eucalyptus urophylla* and Section Exsertaria species. GuangXi Plant. 2021;41(4):631–9.
50. Luo JZ, Xie YJ, Cao JG, Lu WH, Ren SQ. Genetic variation in growth and wind resistance of two-year-old Eucalyptus hybrids. Acta Prataculturae Sinica. 2009;18(3):91–7.
51. Weng QJ, Lai QX, Li FG, Zhou CP, Li JW, Li M, Gan SM. Genetic analysis on early growth and cold tolerance of *Eucalyptus urophylla* × *E. dunnii* hybrids. J Nanjing Forestry Univ (Natural Sci Edition). 2015;39(5):33–8.

Figures

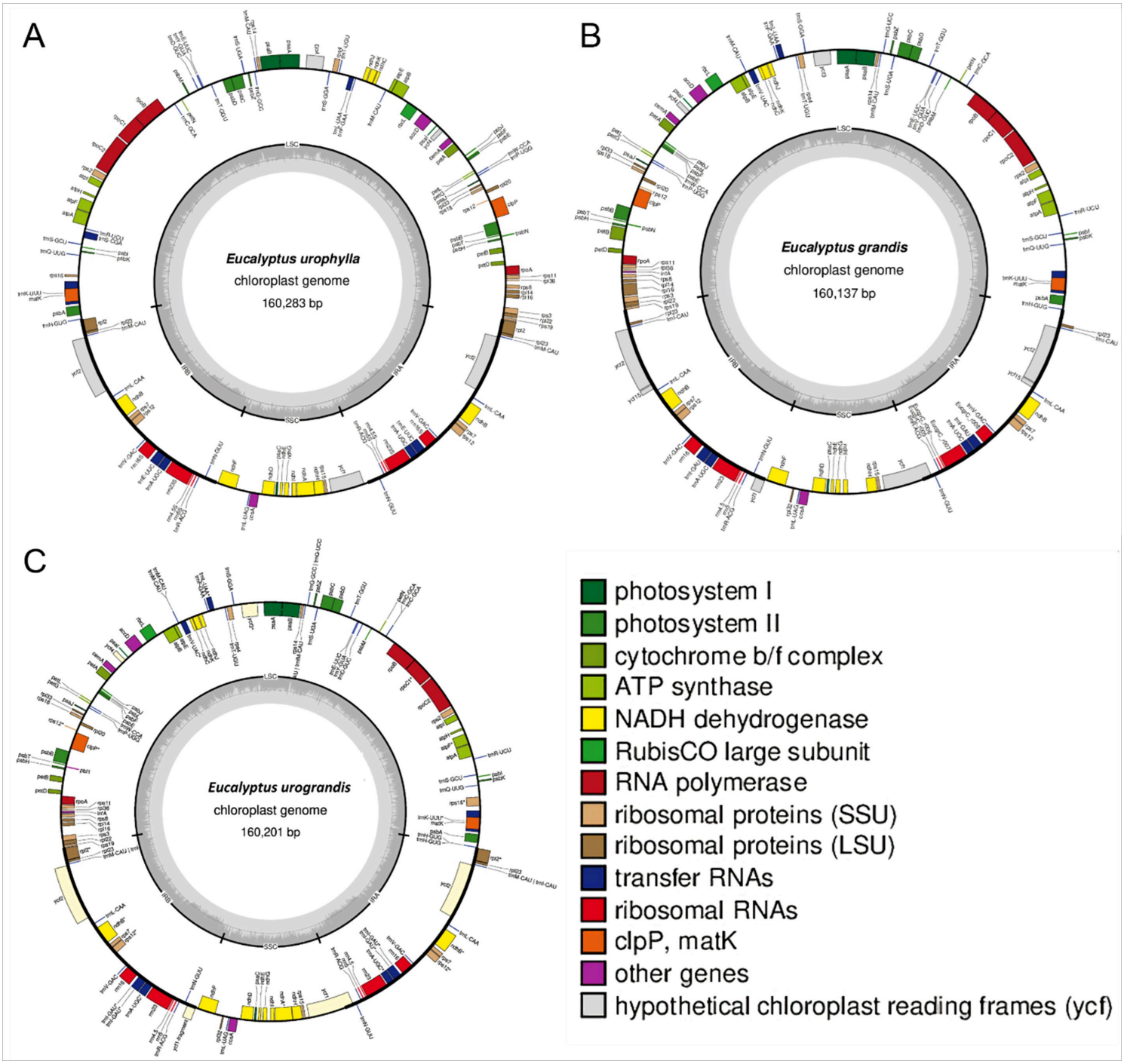

Figure 1

Chloroplastic genomes structure of *E. urophylla*, *E. grandis* and *E. urograndis.*

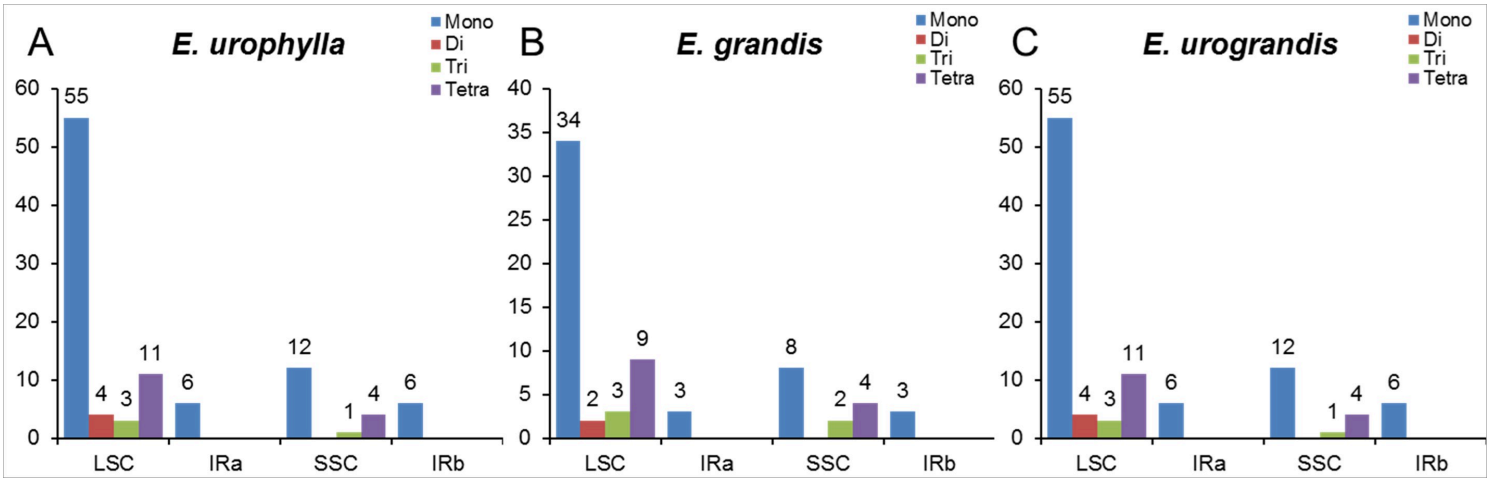

Figure 2

Analysis of SSRs and repeated sequences in the chloroplast genomes of three *Eucalyptus* species. Notes: Mono, Mononucleotide SSRs; Di, Dinucleotide SSRs; Tri, Trinucleotide SSRs; Tetre, Tetranucleotide SSRs; LSC, Large single-copy region; IR, Inverted repeats; SSC, Small single-copy region.

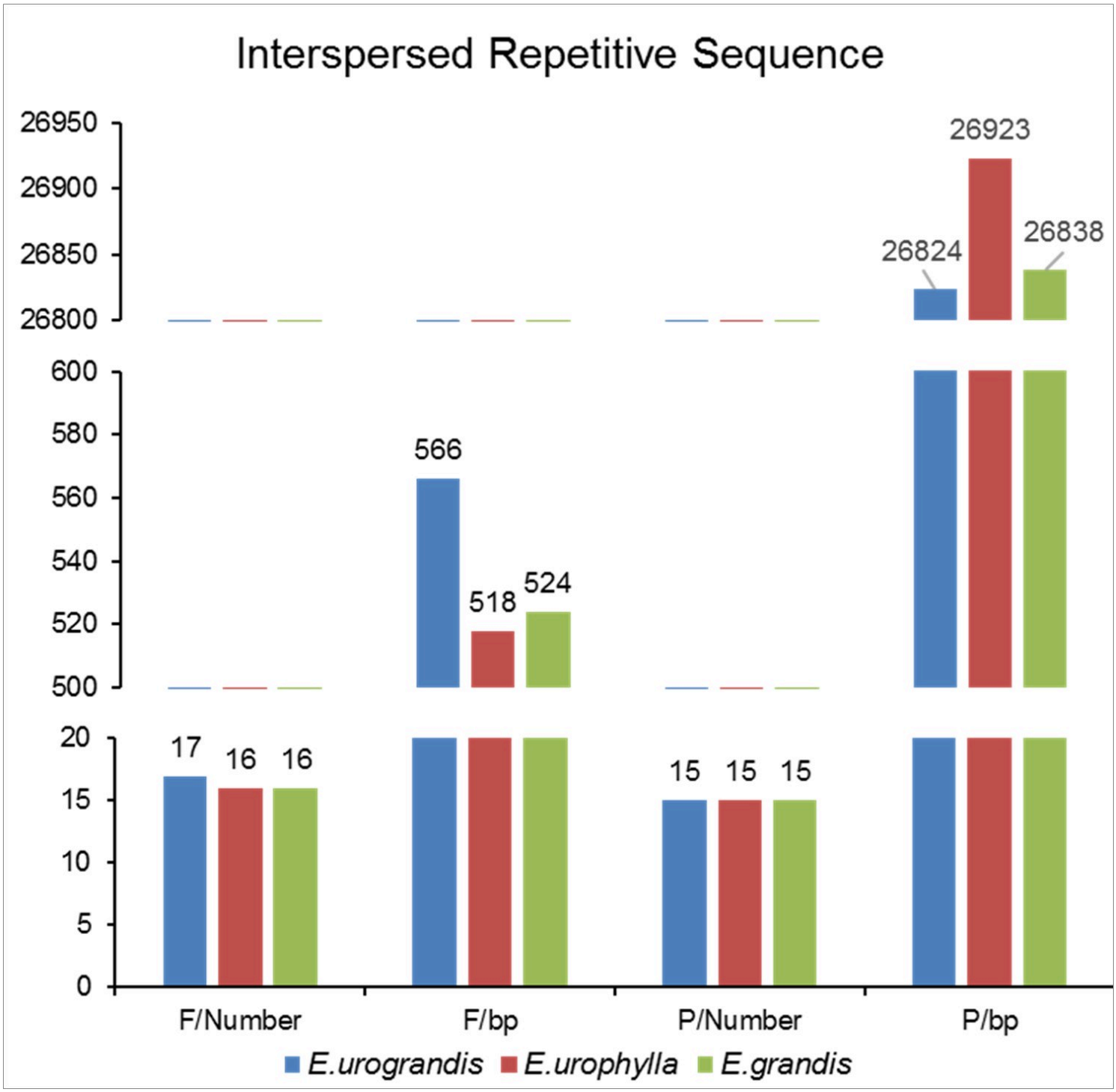

Figure 3

Scattered repeating sequences in three *Eucalyptus* species. Notes: F, Forward sequence; P, Palindromic sequence.

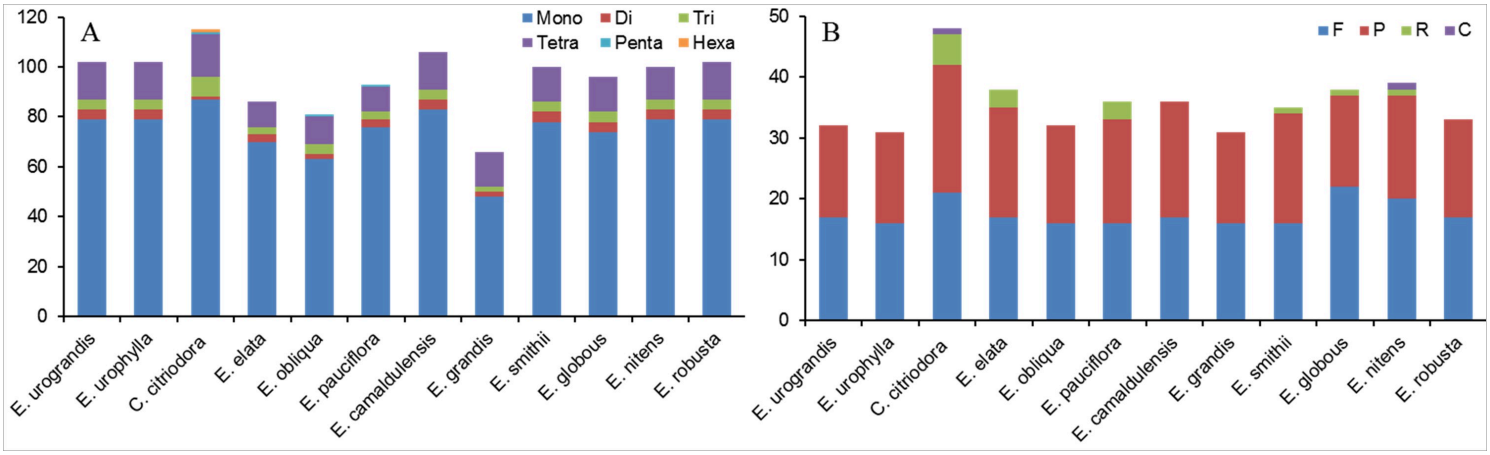

Figure 4

A. Analysis of SSRs and repeated sequences in the chloroplast genomes of 12 eucalypt species. **B.** Scattered repeating sequences in 12 eucalypt species. Notes: F, Forward sequence; P, Palindromic sequence.

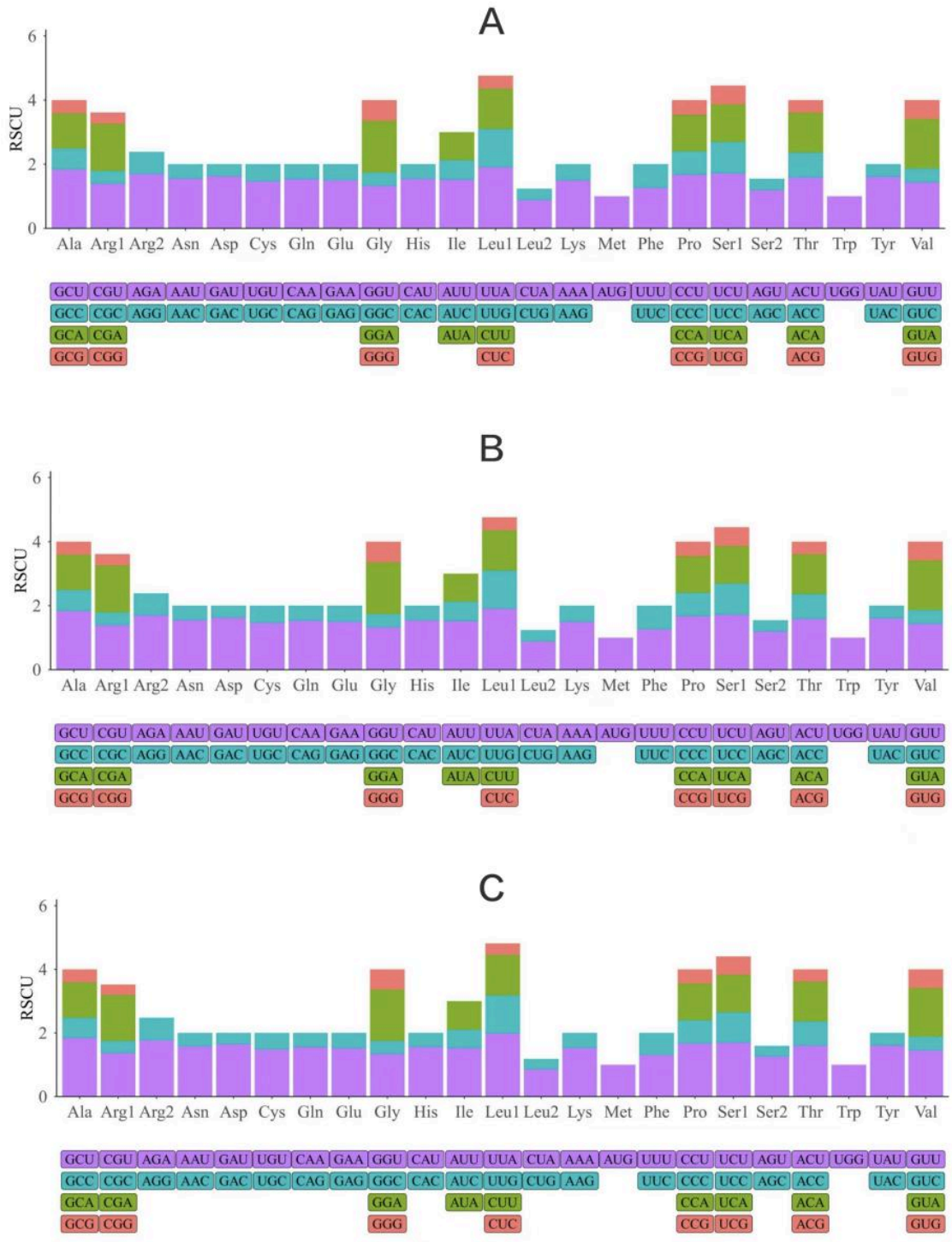

Figure 5

Usage of codon preferences in three *Eucalyptus* species. Notes: A, *E. urograndis*; B, *E. urophylla*; C, *E. grandis*.

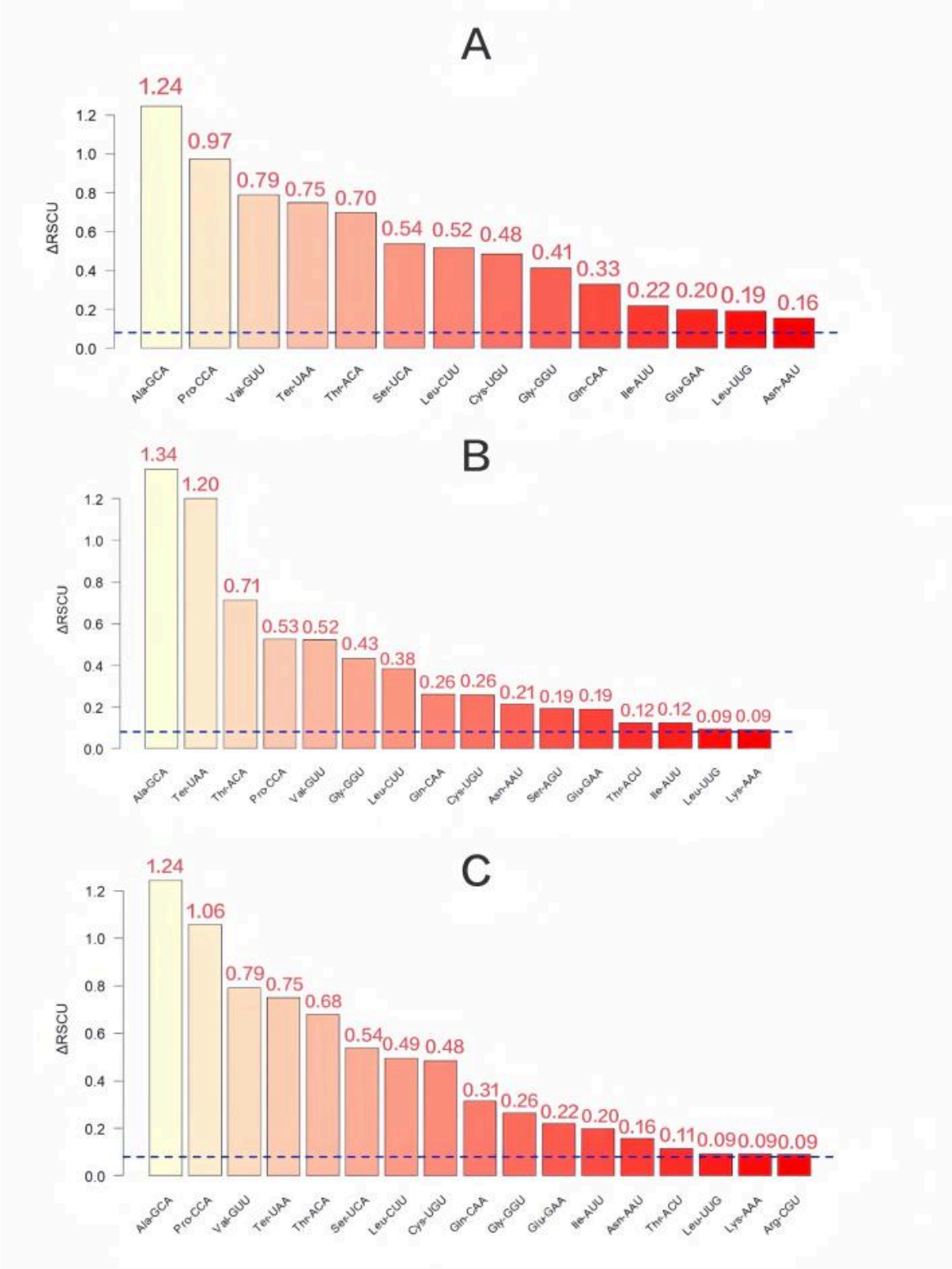

Figure 6

Ranking of ΔRSCU values for each amino acid in the three *Eucalyptus* species. Notes: A, *E. urograndis*; B, *E. urophylla*; C, *E. grandis.*

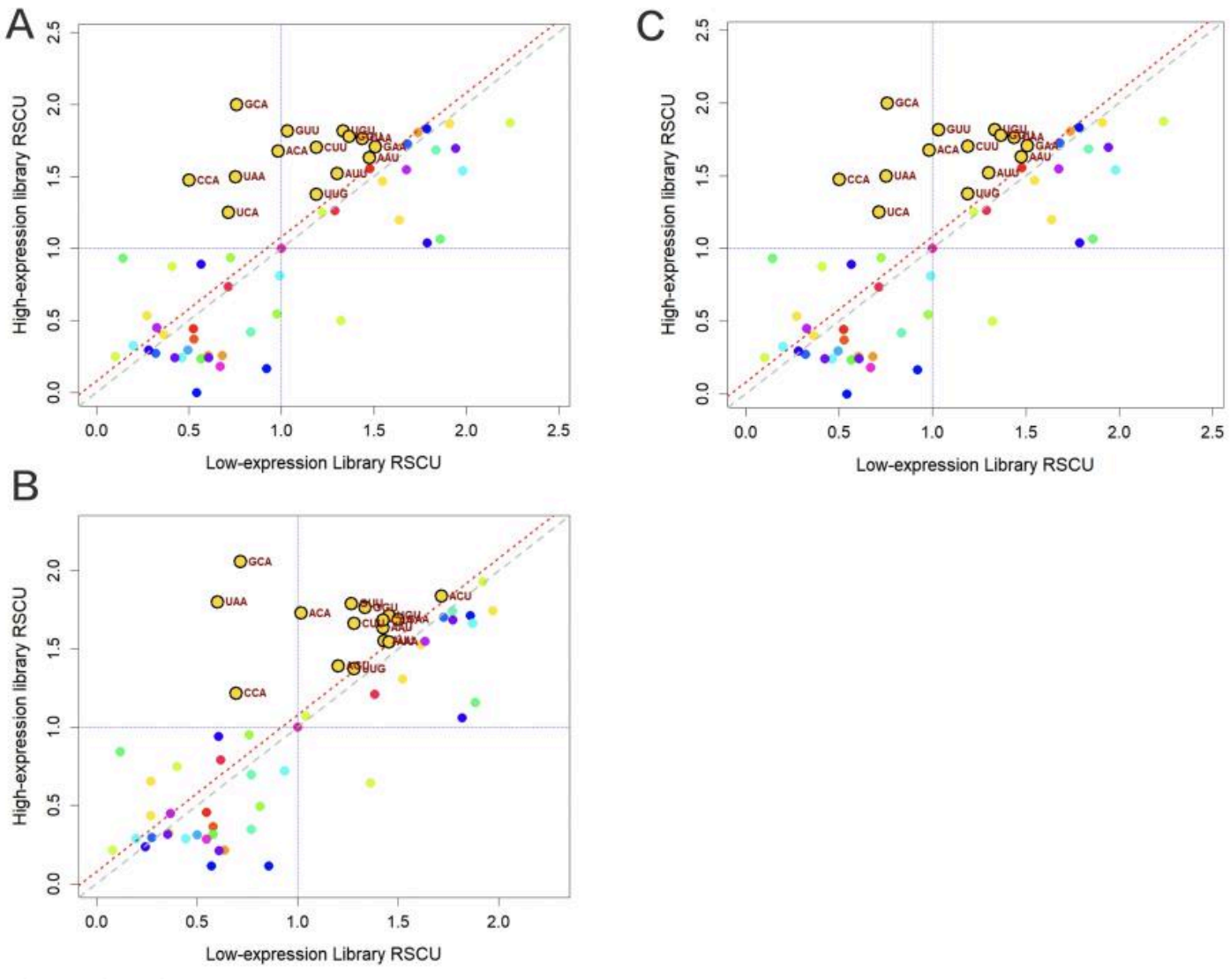

Figure 7

Distribution of optimal codons for each amino acid in the three *Eucalyptus* species. Note: A, *E. urograndis*; B, *E. urophylla*; C, *E. grandis*. The horizontal axis indicates the low-expression library preference region, and the vertical axis indicates the high-expression library preference region. Yellow dots represent optimal codons, the red dashed line indicates the threshold of ΔRSCU > 0.08, and the blue dashed line indicates RSCU > 1.

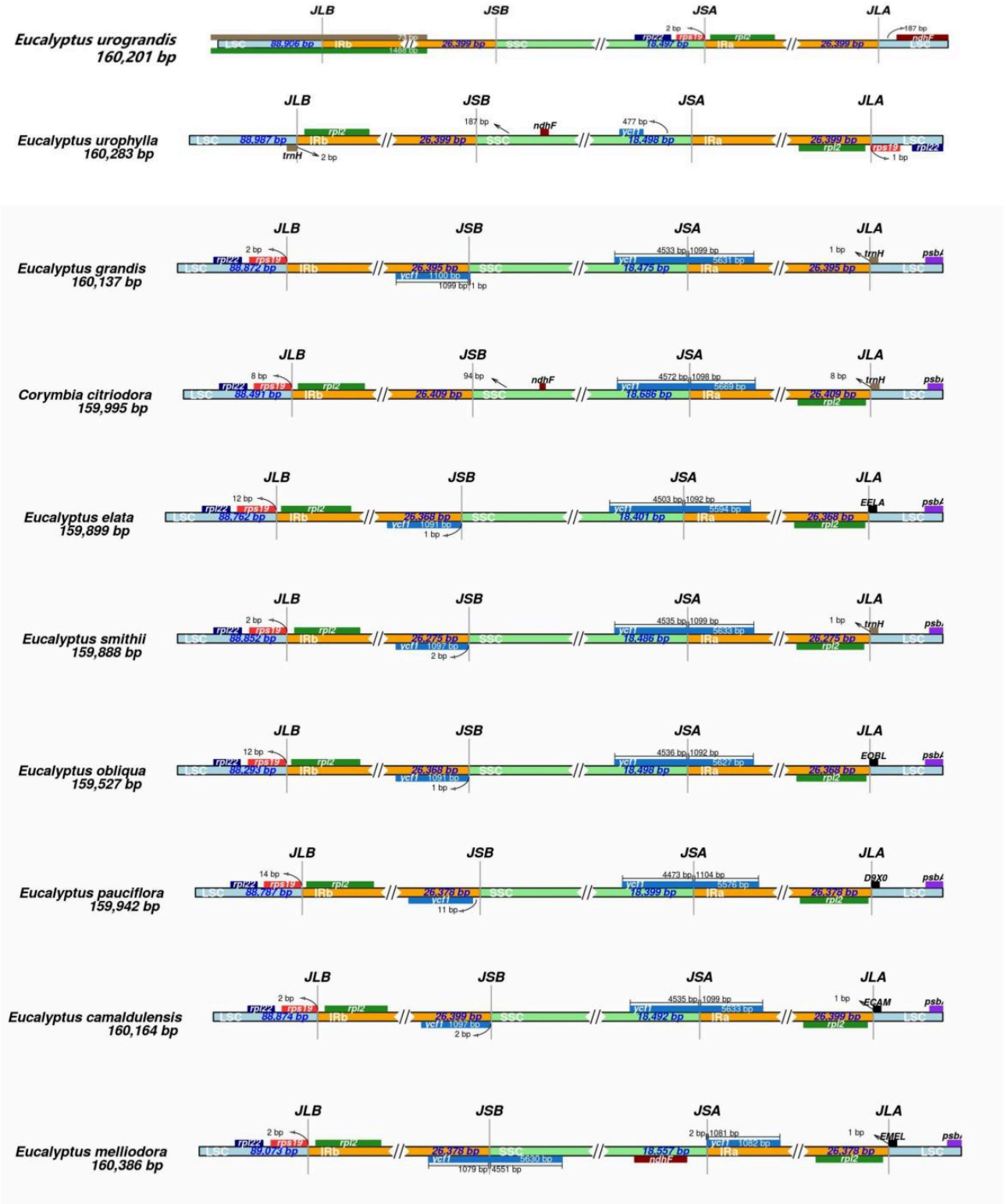

Figure 8

Divergence in boundary regions of chloroplast genomes among 10 eucalypt species. JLB,LSC/IRb; JSB, IRb/SSC; JSA, SSC/IRa; JLA, IRa/LSC. The light blue/yellow/light green regions correspond to the LSC/IRa and IRb/SSC regions, respectively.

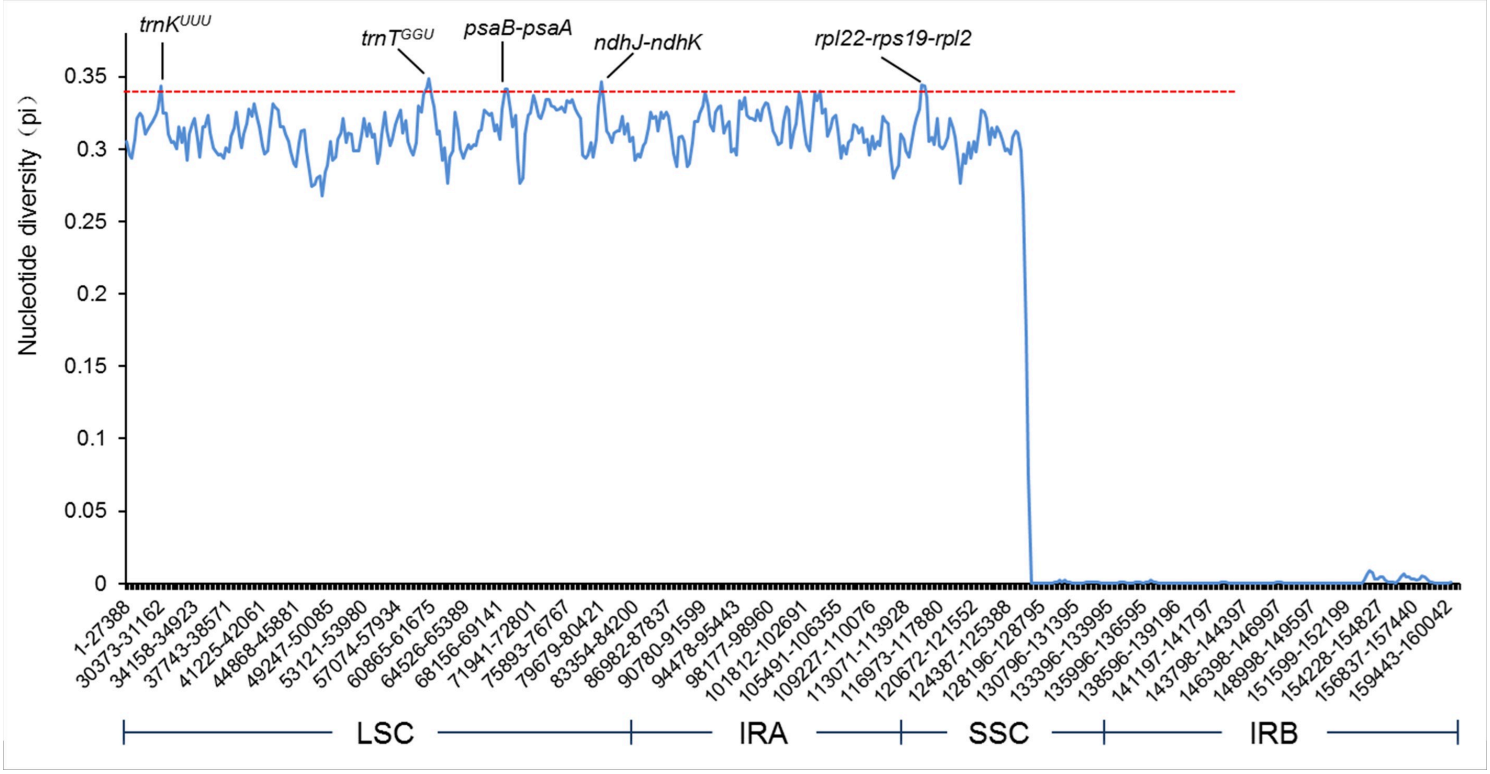

Figure 9

Sliding window test of Pi in the Hybrid tail *E.urograndis*, *E.urophylla* and *E.grandis*chloroplast genomes.Window length: 600 bp; step size: 200 bp. X-axis: the position of the midpoint of a window. Y-axis:nucleotide diversity of each window. LSC, IRA, SSC, and IRB represent the four regions of chloroplast genome structure.

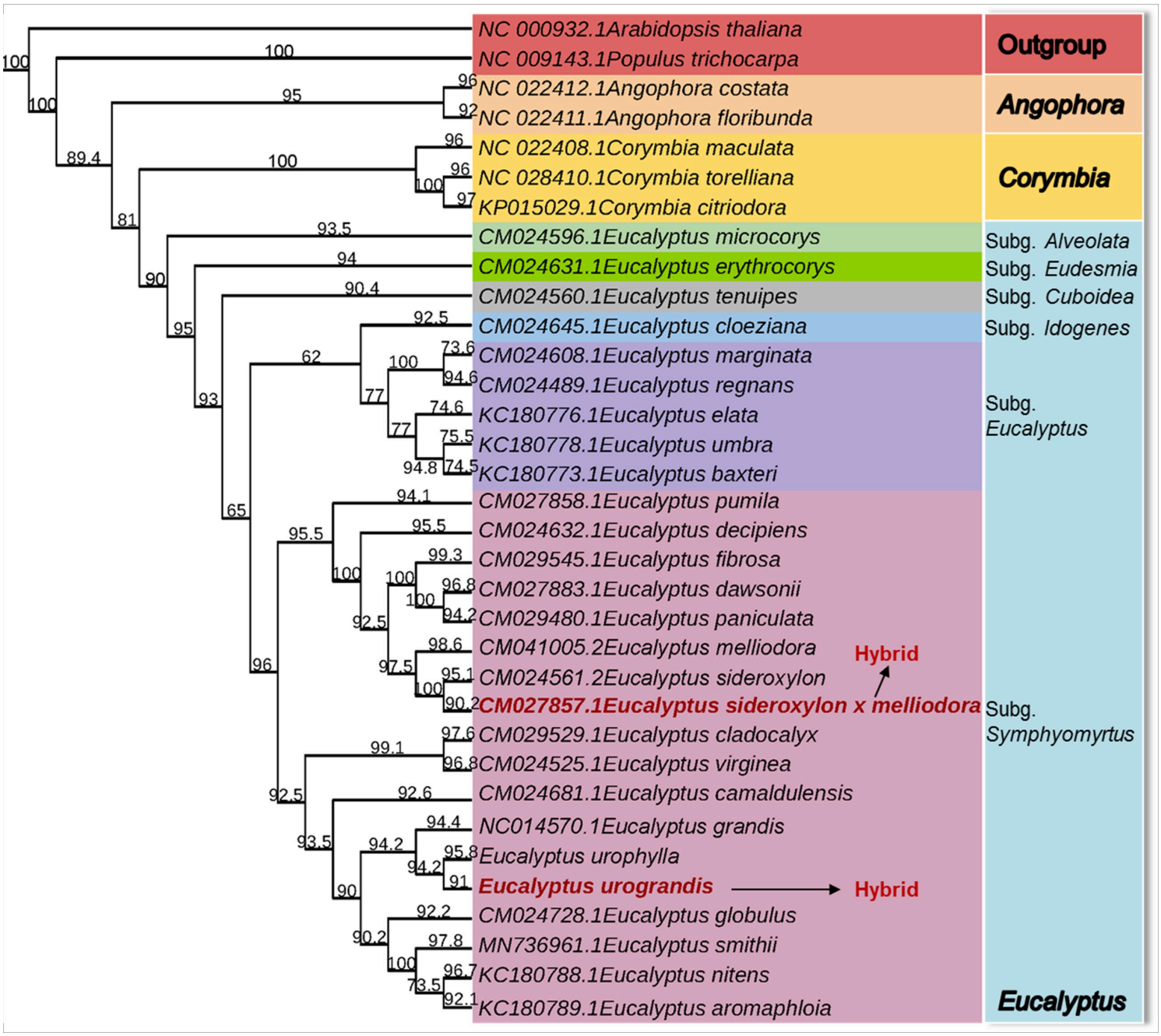

Figure 10

Phylogenetic tree of 32 eucalypt species and two outgroup species based on the complete chloroplast genome data. Numbers near the nodes are bootstrap percentages.

Supplementary Files

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

- SupplementaryTables.xlsx