Characteristics of Complete Chloroplast Genome of Grevillea robusta

guanghua liu (guanghualiu@126.com)
Huaihua University https://orcid.org/0009-0002-6716-0351
Jiaxi Li
Huaihua University
Yubing Yang
Southwest University
Jie Yu
Southwest University

Research Article

Keywords: Proteaceae, high-throughput sequencing, phylogenetic relationship, Protea kilimandscharica

Posted Date: August 24th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3250429/v1

Abstract

Grevillea robusta is an important plant in *Proteaceae*, and decoding and understanding the chloroplast genome of *Grevillea robusta* is of great theoretical significance and practical value to the genetic diversity and phylogenetic relationship of *Proteaceae.* On the basis of high-throughput sequencing data of *Grevillea robusta*, we assembled and annotated the sequencing results using GetOrganelle and CPGAVAS2 programs, and downloaded the chloroplast genome data of genera *Macadamia*, *Helicia* and *Protea* from NCBI database. The chloroplast genomes of four generas. The length of chloroplast genome of *Grevillea robusta* was 158,642 bp, consisting of 129 genes, including 84 protein-coding genes, 37 tRNA genes and 8 rRNA genes. 56 SSRs were obtained from *Grevillea robusta*, among which the single nucleotide repeats were the most (66.07%) and the six nucleotide repeats were the least (1). At the same time, 34 repeats were detected in chloroplast genome of *Grevillea robusta*, mainly are palindrome repeats (16). The IR region of *Grevillea robusta* didn't experience a significant contraction/expansion event, whereas *Protea kilimandscharica* showed a dramatic contraction. Gene selection pressure analysis showed that *ycf*1 genes showed positive selection signals. Analysis of RNA editing sites showed that there were 148 RNA editing sites in the protein-coding genes of chloroplast genome of *Grevillea robusta*, and most of them are C/U editing, up to 54.73%. Phylogenetic analysis confirmed that *Grevillea robusta* was belongs to *Proteaceae*, and grouped with *Helicia* and *Macadamia*, with a support rate of 100%. The chloroplast genome of *Grevillea robusta* was assembled successfully, which had high similarity with the chloroplast genome of *Helicia* and *Macadamia*, and was clustered into a branch during the phylogeny of *Proteaceae*. The results of this study laid a foundation for understanding the systematic evolution of the *Proteaceae* plants, and provide rich data supported for the development of molecular biological information such as molecular marker.

Introduction

Grevillea robusta A. Cunn. ex R. Br. is a tree of the genus *Grevillea* in the family *Proteaceae*, native to Australia, with beautiful shape, bright flowers, peculiar leaves and strong absorption ability of harmful gases, which is an ideal tree species for urban greening (Weng QJ et al. 2006). At the same time, *Grevillea robusta* has the characteristics of drought and infertile tolerance and strong adaptability, which is important for the comprehensive management of stone desertification mountain (Gong Zheng et al. 2013). In recent years, domestic and foreign scholars have focused on the analysis of extract components and activity assay of *Grevillea robusta*, for example, some studies found that leaf extracts of *Grevillea robusta* showed toxicity against grain storage pests and termites (Bhuvaneswari et al., 2014; Afzal et al. 2019), in addition, the phenolic substances in *Grevillea robusta* showed some anticancer activity and antimalarial In addition, phenolic substances in *Grevillea robusta* showed some anticancer and antimalarial activities (Ahmed et al. 2000; Chuang and Wu 2007). However, the genetic analysis and molecular phylogeny of *Grevillea robusta* have been less studied, and only one study has developed five SSR markers for it based on illumina low-depth high-throughput technology (Dabral et al. 2021), and no

chloroplast genomic information has been available for *Grevillea robusta*, which may limit our knowledge of this species.

Chloroplasts are important organelles of plants and a highly conserved tissue, whose genomic inheritance is mainly uniparental and the recombination rate of the evolutionary process is low, which can better reflect the evolutionary and developmental process of the population. In addition, the fast evolutionary rate of protein noncoding sequences in the chloroplast genome can provide some meaningful information loci that are important for the identification of plant species (Petit et al. 2005). Therefore, the chloroplast genome (cpDNA) is considered as an ideal barcode for studying plant phylogeny. In angiosperms, most chloroplast genomes are a double-stranded circular tetrad structure consisting of a large copy region (LSC), a small copy region (SSC), and two inverted repeat sequences (IRa/IRb) (Jansen and Ruhlman 2012). The chloroplast genome sequence ranges mostly within 150–170 kb (Terakami et al. 2012) and contains roughly 110–130 tRNA genes, rRNA genes, and protein-coding genes involved in photosynthesis (Choi et al. 2015). Currently, information from the chloroplast genome has been used in various studies, such as the development of DNA barcodes, chloroplast transformation and breeding of crops with superior agronomic traits (Bock and Khan 2004; Daniell et al. 2005)

In this paper, we sequenced, assembled and analyzed the chloroplast genomes of *Grevillea robusta* based on high-throughput sequencing (NGS), and compared them with those of *Helicia shweliensis* W. W. Sm., *Macadamia integrifolia* Maiden & Betche and *Protea kilimandscharica* Engl. from *Proteaceae*, with the aim of enriching the genomic resources of *Grevillea robusta* and providing information for the study of plant evolutionary patterns and phylogeny in the *Proteaceae*.

Material and methods

1.1 Test materials

Fresh *Grevillea robusta* leaves were collected from the campus of Southwest University, Beibei District, Chongqing (geospatial coordinates: N29.842889,E106.394527). Samples were preserved in the herbarium of Southwest University with voucher number YJ-066. Total genomic DNA was extracted by CTAB method (Arseneau et al. 2017) and total DNA fragments were detected using ultrasound. DNA libraries with an insert size of 350 bp were constructed with the help of the NEBNext (Emerman et al. 2017) library construction kit and sequenced using the HiSeq Xten PE150 sequencing platform. The sequencing yielded a total of 4.57 Gb of raw data, containing 15,249,412 reads, and clean data were obtained by Trimmomatic (Bolger et al. 2014): we removed more than 5% of low quality sequences with base N and more than 50% of base quality values Q < 19. Finally, 15,158. 092 clean reads were obtained after trimming.

1.2Assembly and annotation of genes

Chloroplast assembly was done using GetOrganelle (v1.7.3) to assemble the chloroplast genome with default parameters based on clean reads. The original sequencing data were mapped to the assembled

chloroplast genome using Bowite2 (v2.96 0.1) (Langmead et al. 2009) under default settings to verify the accuracy of the assembly. Based on the assembly results annotation of the *Grevillea robusta* chloroplast genome was completed with the help of the CPGAVAS2 (Tillich et al. 2017) program and the reference genome (*Macadamia ternifolia*, GenBank: NC_036416.1), and annotation errors were manually edited and corrected using Appolo software. The genome sequence has been saved in GenBank with the accession number OK054586.1.

1.3 Repeated sequences and SSR analysis

Simple repeat sequences (SSRs) were obtained through the online website MISA (https://webblast.ipk-gatersleben.de/misa/) and include single, two, three, four, five and six nucleotide repeats with minimum numbers of 10, 5, 4, 3, 3 and 3, respectively. REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer) was used for the calculation of palindromic repeats, reverse repeats, and dispersed repeats, etc. The parameters were set as follows:Hamming distance of 3, the minimum repeat size is 30 bp, and the e-value is limited to less than 1e-05.

1.4 Sequence diversity analysis

We used the online website IRscop (https://irscope.shinyapps.io/irapp/) to visualize the chloroplast genome IR/SC boundary and its adjacent genes. Shared protein-coding gene sequences were extracted in Phylosuite software based on annotation files in .gb format, and sequences were aligned using Mafft software. The CODEML model in PAML v4.9 (Yang 1997) was used to calculate the nucleotide substitution rate of protein-coding genes for each chloroplast, including the non-synonymous substitution rate (Ka), the synonymous substitution rate (Ks) and the ratio of both (Ka / Ks). The detailed parameters are as follows: verbose = 0; icode = 0; weighting = 0; commonf3x4 = 0 ; ndata = 1. For each protein-coding gene, we constructed a single-gene tree as a background tree using the maximum likelihood (ML) method implemented in RAxML v8.2.4.

1.5 Identification of RNA editing sites

We downloaded the *Grevillea robusta* transcriptome data from NCBI's SRA database (accession number: ERR2040150) and mapped the raw transcriptome data to the protein-coding gene sequences of *Grevillea robusta* chloroplasts, using our custom Python script to count the base coverage and specific base composition of each locus to identify loci where RNA editing events had occurred. To exclude false positive loci triggered by natural variation, we also downloaded another DNA-seq data set for *Grevillea robusta* (SRR11309581) and mapped all reads from this data set to the nucleotide sequences of our assembled *Grevillea robusta* chloroplast protein-coding genes to identify potential single nucleotide polymorphism loci (SNPs). Among the RNA editing sites identified by the transcriptome, we only retained sites with coverage greater than 5 and carefully screened for reads that may have been mapped up incorrectly due to sequence homology, and finally excluded previously identified SNP sites, and the remaining sites will be retained as high-quality RNA editing sites.

1.6 Construction of phylogenetic tree

Download from NCBI repository *Helicia shweliensis* NC_045942.1, *Helicia nilagirica* NC_057271.1, *Macadamia integrifolia* KF862711.1, *Macadamia ternifolia* NC_036416.1, *Protea kilimandscharica*, *Platanus occidentalis* DQ923116.1, *Platanus racemosa* MZ128519.1, *Sabia parviflora* MW566751.1, *Meliosma aff. cuneifolia* NC_029430.1, *Nelumbo lutea* JQ336992.1, *Nelumbo nucifera* JQ336993.1, *Coptis chinensis* NC_036485.1 and *Coptis teeta* NC_054331.1 chloroplast genomes for phylogenetic analysis.

The chloroplast genomes of the 16 species were multiplexed and saved as .phy format files. *Coptis chinensis* and *Coptis teeta* were selected as outgroups, and maximum likelihood trees were constructed using IQtree v2.0 with the command "iqtree2 -s sequences.phy --alrt 1000 -B 1000 ", and finally embellish the tree results with ITOL.

Results

1.1 Chloroplast genome characteristics of *Grevillea robusta*

The chloroplast genome of *Grevillea robusta* is a typical tetrameric ring structure (Fig. 1), with a total genome sequence length of 158,642 bp, and a GC content of 38.13%, of which the LSC, SSC and IR regions were 87,021 bp, 18,961 bp and 26,330 bp, respectively .The GC content of the IR region (43.24%) was significantly higher than that of the SSC region (The GC content of the IR region (43.24%) was significantly higher than that of the SSC region (31.78%) and the LSC region (36.43%). The *Grevillea robusta* chloroplast genome contained 129 genes, including 37 tRNAs, 8 rRNAs and 84 protein-coding genes. *Grevillea robusta* did not differ significantly from the other three genus species of the *Proteaceae* in terms of genome length, gene content and total GC content (Table 1). *Grevillea robusta* contained 16 intron-containing genes, clpP and ycf3 genes had two introns, and 14 genes contained one intron (Table 2).

Table 1
Comparison of chloroplast genome characteristics of four genera in *Proteaceae*

Species		***G.robusta***	***H.shweliensis***	***M. integrifolia***	***P.kilimandscharica***
Accession number		OK054586.1	NC_045942.1	KF862711.1	MH362765.1
Length (bp)	Total	158,642	157,151	159,714	158,657
	LSC	87,021	85,490	88,092	87,241
	SSC	18,961	18,256	18,812	18,534
	IR	26,330	26,701	26,405	26,441
GC content (%)	Total	38.13	38.11	38.12	38.03
	LSC	36.43	36.46	36.54	36.31
	SSC	31.76	31.77	31.67	31.56
	IR	43.24	42.93	43.03	43.14
Gene numbers	Total	129	129	129	128
	tRNA	37	36	36	36
	rRNA	8	8	8	8
	Protein-coding	84	85	85	84

Table 2
Gene composition in the chloroplast genome of *Grevillea robusta*

Category of Genes	Group of Genes	Name of Genes
Ribosomal RNA	rRNA	*rrn*16S(x2),*rrn*23S(x2),*rrn*5S(x2),*rrn4.*5S(x2)
Transfer RNA	tRNA	37 unique trna genes
Photosynthesis	Subunits of ATP synthase	*atp*A, *atp*B, *atp*E, *atp*F*, *atp*H, *atp*I
	Subunits of photosystem II	*psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*I, *psb*J, *psb*K, *psb*M, *psb*N, *psb*T, *psb*Z, *ycf*3**
	Subunits of NADH-dehydrogenase	*ndh*A*, *ndh*B*(x2), *ndh*C, *ndh*D, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J, *ndh*K
	Subunits of cytochrome b/f complex	*pet*A, *pet*B*, *pet*D, *pet*G, *pet*N
	Subunits of photosystem I	*psa*A, *psa*B, *psa*C, *psa*J
	Subunit of rubisco	*rbc*L
Self-replication	Large subunit of ribosome	*rpl*14, *rpl*16, *rpl*2*(x2), *rpl*20, *rpl*22, *rpl*23(x2), *rpl*32, *rpl*33, *rpl*36
	DNA dependent RNA polymerase	*rpo*A, *rpo*B, *rpo*C1*, *rpo*C2
	Small subunit of ribosome	*rps*11, *rps*12(x2), *rps*14, *rps*15, *rps*16*, *rps*18, *rps*19, *rps*2, *rps*3, *rps*4, *rps*7(x2), *rps*8
Other genes	Subunit of Acetyl-CoA-carboxylase	*acc*D
	c-type cytochrom synthesis gene	*ccs*A
	Envelop membrane protein	*cem*A
	Protease	*clp*P**
	Translational initiation factor	*inf*A
	Maturase	*mat*K

Note: (×2) indicates that the gene is located in IRs and therefore has two complete copies, * and ** indicate that the gene contains 1 / 2 introns.

Category of Genes	Group of Genes	Name of Genes
Unknown	Conserves open reading frames	*ycf*1, *ycf*15*(x2), *ycf*2 (x2), *ycf*4
Note: (×2) indicates that the gene is located in IRs and therefore has two complete copies, * and ** indicate that the gene contains 1 / 2 introns.		

1.2 Replication and SSR analysis

Microsatellites, also known as simple repeat sequences (SSRs), are usually 6 bp tandem sequences in the genomes of eukaryotic organisms (Lovin et al. 2009), and the high polymorphism and co-dominant inheritance of SSRs have led to a wide range of applications in different types of molecular markers (Morgante et al. 2002). There are 34 oligonucleotide repeats in the chloroplast genome of *Grevillea robusta*, of which palindromic repeat sequences are the most abundant with 16, followed by dispersed repeats and tandem repeats, both with 8, and lastly, reverse repeat sequences with 2. We screened 56 SSR loci from the *Grevillea robusta* chloroplast genome, among which the single nucleotide repeat types were the most (37), and were mainly A/T sequence repeats, which accounted for about 62.50% of the total number of SSRs; there were three dinucleotide repeat types, two trinucleotide repeats, eight tetranucleotide repeats, five pentanucleotide repeats, and one hexanucleotide sequence type.

1.3 Boundary Characterization of the Chloroplast Genomes of *Proteaceae*

There are four boundaries in the tetrad structure of the chloroplast genome, namely, JLB(LSC/IRb), JSB(IRb/SSC), JSA(SSC/IRa), and JLA(IRa/LSC).Expansion and contraction phenomena of the IR region are very common in the process of the evolution of the chloroplast genome. In the chloroplast genomes of the six species of *Proteaceae*, the lengths of the IR regions were similar, the types of boundary genes were not significantly different, and the boundary genes of *Grevillea robusta* were similarly located to those of the four species of *Macadamia* and *Helicia*, which suggests that the IR regions of the chloroplast genome of *Grevillea robusta* did not undergo drastic contraction/expansion events, however, *Protea kilimandscharica* IRa/LSC boundary gene, trnH, entered the IRa region and was found to be in the IRa/LSC region. the IRa region and generated another copy of trnH in the IRb region. The IR region of *Protea kilimandscharica* underwent a small expansion event compared to the other five species of *Proteaceae*.

1.3 Analysis of gene selection pressures

The chloroplast protein coding genes of *Grevillea robusta* with *Helicia shweliensis*, *Macadamia integrifolia* and *Protea kilimandscharica* were analyzed for gene selection pressure by PAML software, and the results showed that rpl22 and matK had high ka values, and rpl33, ndhF, rpl22 and psaC had relatively high ks values. The Ka/Ks values of many genes are less than 1. It is evident that most of the protein-coding genes are in a state of purifying selection during the course of evolution. However, two

genes (ycf1 and ycf2) had Ka/Ks values greater than 1, with Ka/Ks = 1.30 for ycf1 and 1.03 for ycf2. This suggests that they may have been positively selected during evolution.

1.4 RNA editing sites

The statistics of RNA editing sites are shown in (Fig. 5a). In the protein coding sequences of the *Grevillea robusta* chloroplast genome, a total of 148 RNA editing sites were detected for 42 protein coding genes, among which ndhB possessed the most editing sites (18), followed by psbA (13). In terms of editing efficiency, the number of sites possessing 100% editing efficiency reached 25, and the number of other low-frequency editing sites was also high (Fig. 5b), with a total of 47 sites with editing efficiency of less than 10%. In terms of editing types, all 12 possible editing types were found (Fig. 5c), among which, the highest number of C/U editing was found, accounting for 54.73% of the total RNA editing sites, i.e., 81 sites were edited from base C of the genome to base U of the mRNA strand, followed by T/C, i.e., 17 sites were edited from base T of the genome to base C of the mRNA strand.

1.5 Phylogenetic analysis

We selected representatives at the level of *Proteales* and constructed a phylogenetic tree based on the entire plastid genome sequence. The results showed that the four families in the order *Proteales* (*Proteaceae*, *Sabiaceae*, *Nelumbonaceae*, and *Platanaceae*) were well clustered. Among them, *Sabiaceae* is at the most basal position and is the earliest one to be differentiated, followed by *Nelumbonaceae*. *Proteaceae* and *Platanaceae* were closer in affinity and had high support for each node. In the *Proteaceae*, *Grevillea* is in the same branch as *Macadamia* and *Helicia* with 100% support. Due to the scarcity of chloroplast genome resources in the *Proteaceae*, we were unable to make more in-depth and specific comparisons of the phylogenetic relationships of *Grevillea robusta* with species of other genera in the family.

Discussion

Chloroplast genome sequences evolve at a fast rate, but their gene content and gene order are relatively highly conserved in green plants, making chloroplast genomes ideal for studying the phylogenetic and evolutionary relationships in plants.IR regions are usually present in the chloroplast genomes of most angiosperms, but there are exceptions, such as *Fabaceae*, where about 95% of the chloroplast genome lacks IR regions (or only very short residues are present), and some species of the *Cactaceae* family have also been shown to lack IR regions, which have been interpreted as a novel evolutionary and structural change (Lavin et al. 1990; Sanderson et al. 2015). In this study, none of the IR regions of the *Proteaceae* species analyzed showed deletions and there were no significant differences in their lengths, which were relatively conserved in all species except for *Protea kilimandscharica*, which had a relatively small expansion of their IR regions. Changes at the junction of IR regions of chloroplast genomes can help to understand the evolutionary patterns of species (Tillich et al. 2017), and studies have shown that species with a high degree of similarity in the boundary region have a lower degree of differentiation at the

chloroplast boundaries (Liu et al. 2018), therefore, from the results of the IR boundary analysis, it can be found that the *Grevillea robusta* is more closely related to *Macadamia* and *Helicia .*

SSRs in plant chloroplasts have been widely used in the study of plant population genetics and evolution and are thought to occur frequently in the non-coding regions of the chloroplast genome, especially in the gene spacer region (Singh et al. 2015). In this paper, a total of 56 SSRs of *Grevillea robusta* were screened, among which the single nucleotide repeat type was the most abundant, accounting for 66.07% of the total SSRs, and was mainly dominated by A/T. In this paper, four types of oligonucleotide repeats were screened: palindromic repeats, inverted repeats, tandem repeats and scattered repeats. Among them, the proportion of palindromic repeat sequences is as high as 47%. Oligonucleotide repeat sequences are used as a resource for polymorphic region identification when there are fewer resources for mutation hotspots in the chloroplast genome (Ahmed et al. 2012). Oligonucleotide repeats in *Grevillea robusta* play an important role in the evolution of the genome and will provide a theoretical basis for its molecular marker development, phylogeny of the *Proteaceae* or species identification.

The calculation of non-synonymous substitutions (Ka) and synonymous substitutions (Ks) can be used to detect gene conservation, and they are important for understanding the evolutionary dynamics of protein coding sequences and reconstructing the phylogeny of closely related species. According to the results of gene selection pressure analysis, most of the chloroplast protein-coding genes of *Grevillea robusta* and *Helicia shweliensis*, *Macadamia integrifolia* and *Protea kilimandscharica* had low Ka/Ks ratios (< 0.5), suggesting that these genes are under purifying selection. However, ycf1 had a high Ka/Ks ratio of 1.30, suggesting that the ycf1 gene may be under low selection pressure or higher positive selection conditions. ycf1 genes are DNA barcoding candidates in many species (Jia et al. 2017). The ycf2 gene of *Grevillea robusta* has Ka/Ks = 1.03, a value that is more likely to be considered as a sign that the gene was subjected to neutral selection rather than positive selection. Strong purifying selection pressure on genes can lead to higher chances of synonymous substitutions arising than non-synonymous ones (Lawrie et al. 2013). In the chloroplast genome of *Grevillea robusta*, we found that synonymous substitutions were more common than nonsynonymous substitutions, and other studies have found this trend in chloroplast genomes (Tae et al. 2014; Xu et al. 2015), which further supports our results.

In this study, a phylogenetic tree was constructed based on the chloroplast genomic data, and the support rate of each of its branches was high, indicating the reliability of the phylogenetic results, which provided strong evidence for the phylogenetic classification of *Proteaceae* and the understanding of the position of *Grevillea robusta* in *Proteaceae*. Although the development of high-throughput technology has brought the phylogenetic study to the genomic level, the current chloroplast genomic data are still relatively scarce for the species of *Proteaceae*, and as the study of the medicinal value of the species of *Proteaceae* gradually deepens, more genomic information is still needed to explore the phylogenetic relationships of the family in the future.

In this paper, the whole chloroplast genome of *Grevillea robusta* was sequenced, assembled and annotated, and its structure, gene composition, repeat sequences and RNA editing sites of protein-coding genes were analyzed. The total length of the chloroplast genome of silver birch was 158,642 bp, 129 genes were annotated, 56 SSR sites were screened, and 148 RNA editing events located in the protein coding sequences of chloroplasts were identified. Meanwhile, a comparative analysis between *Grevillea robusta* and four genera of *Proteaceae* revealed that they were not significantly different in terms of chloroplast genome length, gene number and composition, and the IR region of *Grevillea robusta* did not undergo any obvious contraction/expansion events as revealed by the boundary analysis. The results of phylogenetic analysis indicated that *Grevillea* was more closely related to the genera *Macadamia* and *Helicia* .*Grevillea robusta* has a high potential for exploitation both as timber and edible nuts. The sequencing and analysis of its chloroplast genome not only provides genetic resources for this important forest tree, but also provides a data base for the phylogenetic analysis of *Proteaceae*.

Declarations

Data availability

All data generated or analyzed during this study are included in this published article.

Acknowledgements

Thanks to Mr. Zuo Youwei from Southwest University for their help.

Funding

The authors did not receive support from any organization for the submitted work.

Author information

Authors and Affiliations

College of Biological and Food Engineering, Huaihua University, Huaihua, Hunan, 418000, China.

Jiaxi Li

Key Laboratory of Research and Utilization of Ethnomedicinal Plant Resources of Hunan Province, Huaihua University, Huaihua, Hunan,418008, China.

Guanghua Liu

School of life sciences, Southwest University, Chongqing, 400715, China.

Yubing Yang

College of Horticulture and Landscape Architecture, Southwest University, Chongqing, 400715, China.

Jie Yu

Contributions

All the authors participated in compiling and analyzing the data and preparing the manuscript.

Corresponding author

Correspondence to guanghualiu

Ethics declarations

Conflict of interest

The authors declare no conflict of interest.

References

1. Gong Z, Zhang WH, Zhang FQ et al (2013) Preliminary report on afforestation experiment of two Species of Grevillea. Guangdong Forestry Science and Technology 29:60–63
2. https:/ /doi.org/CNKI:SUN:GDLY.0.2013-04-013
3. Weng QJ, Liu YC (2006) Biological characteristics and cultivation techniques of Grevillea robusta. Guangdong Forestry Science and Technology 22:101–103
4. https://doi.org/10.3969/j.issn.1006-4427.2006.01.027
5. Afzal M, Farman M, Rasib KZ, Qureshi NA (2019) Biocidal action of silver oak (*Grevillea robusta*) leaf extract on the termite Heterotermes indicola Wasmann (*Blattodea: Rhinotermitidae*). Int Biodeterior Biodegrad 139:1–10
6. https://doi.org/10.1016/j.ibiod.2019.02.001
7. Ahmed AS, Nakamura N, Meselhy MR, Makhboul MA, El-Emary N, Hattori M (2000) Phenolic constituents from Grevillea robusta. Phytochemistry 53:149–154
8. https://doi.org/10.1016/S0031-9422(99)00484-7
9. Ahmed I, Biggs PJ, Matthews PJ, Collins LJ, Hendy MD, Lockhart PJ (2012) Mutational dynamics of aroid chloroplast genomes. Genome Biol Evol 4:1316–1323
10. https://doi.org/10.1093/gbe/evs110
11. Arseneau JR, Steeves R, Laflamme M (2017) Modified low-salt CTAB extraction of high-quality DNA from contaminant-rich tissues. Mol Ecol Resour 17:686–693
12. https://doi.org/10.1111/1755-0998.12616
13. Bhuvaneswari S, Sripriya N, Balamurugan A, Siril A, Prakash N (2014) Studies on the phytochemistry and bioefficacy of industrial crops - Coffea canephora and Gravillea robusta from Kolli hills. Int J Res Pharm Sci 5:147–151

14. Bock R, Khan MS (2004) Taming plastids for a green future. Trends Biotechnol 22:311–318. https://doi.org/10.1016/S0167-7799(04)00079-4
15. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
16. Choi KS, Son O, Park S (2015) The chloroplast genome of Elaeagnus macrophylla and trnH duplication event in Elaeagnaceae. PLoS ONE 10:e0138727
17. https://doi.org/10.1371/journal.pone.0138727
18. Chuang T-H, Wu P-L (2007) Cytotoxic 5-Alkylresorcinol Metabolites from the Leaves of Grevillea robusta. J Nat Prod 70:319–323. https://doi.org/10.1021/np0605687
19. Dabral A, Shamoon A, Meena RK, Kant R, Pandey S, Ginwal HS, Bhandari MS (2021) Genome skimming-based simple sequence repeat (SSR) marker discovery and characterization in Grevillea robusta. Physiol Mol Biology Plants 27:1623–1638. https://doi.org/10.1007/s12298-021-01035-w
20. Daniell H, Ruiz ON, Dhingra A (2005) Chloroplast genetic engineering to improve agronomic traits. Transgenic Plants: Methods and Protocols :111–137
21. https://doi.org/10.1385/1-59259-827-7:111
22. Emerman AB, Bowman SK, Barry A, Henig N, Patel KM, Gardner AF, Hendrickson CL (2017) NEBNext Direct: A Novel, Rapid, Hybridization-Based Approach for the Capture and Library Conversion of Genomic Regions of Interest. Curr Protoc Mol Biol 119. 7.30.31–37.30.24
23. Jansen RK, Ruhlman TA (2012) Plastid genomes of seed plants. Genomics of Chloroplasts and Mitochondria. Springer, Dordrecht, pp 103–126
24. https://doi.org/10.1007/978-94-007-2920-9_5
25. Jia Y, Vázquez L, Xiaodan C, Huimin L, Hao Z, Zhanlin L, Guifang Z (2017) Development of Chloroplast and Nuclear DNA Markers for Chinese Oaks (*Quercus Subgenus Quercus*) and Assessment of Their Utility as DNA Barcodes. Front Plant Sci 8:816
26. https://doi.org/10.3389/fpls.2017.00816
27. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:1–10
28. https://doi.org/10.1186/gb-2009-10-3-r25
29. Lavin M, Doyle JJ, Palmer JD, EVOLUTIONARY SIGNIFICANCE OF THE LOSS OF THE CHLOROPLAST-DNA INVERTED REPEAT IN THE LEGUMINOSAE SUBFAMILY PAPILIONOIDEAE (1990) Evolution 44(2). https://doi.org/10.1111/j.1558-5646.1990.tb05207.x
30. Lawrie DS, Messer PW, Hershberg R, Petrov DA, Plotkin JB (2013) Strong Purifying Selection at Synonymous Sites in D. melanogaster. PLoS Genet 9:261–270
31. https://doi.org/10.1371/journal.pgen.1003527
32. Liu L, Wang Y, He P, Li P, Lee J, Soltis DE, Fu C (2018) Chloroplast genome analyses and genomic resource development for epilithic sister genera Oresitrophe and Mukdenia (*Saxifragaceae*), using genome skimming data. BMC Genomics 19:235. https://doi.org/10.1186/s12864-018-4633-x

33. Lovin DD, Washington KO, deBruyn B, Hemme RR, Mori A, Epstein SR, Harker BW, Streit TG, Severson DW (2009) Genome-based polymorphic microsatellite development and validation in the mosquito Aedes aegypti and application to population genetics in Haiti. BMC Genomics 10:590. https://doi.org/10.1186/1471-2164-10-590
34. Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200
35. https://doi.org/10.1038/ng822
36. Petit JR, Duminil J, Fineschi S, Hampe A, Salvini D, Ven-dramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Mo-lecular Ecol 14:689–701. https://doi.org/10.1111/j.1365-294X.2004.02410.x
37. Sanderson MJ, Copetti D, Alberto B, Bustamante E, Charboneau JLM, Eguiarte LE, Kumar S, Lee HO, Lee J, McMahon M (2015) Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): Loss of the ndh gene suite and inverted repeat. Am J Bot 102:1115–1127. https://doi.org/10.3732/ajb.1500184
38. George B, Bhatt BS, Awasthi M, George B, Singh AK (2015) Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. Curr Genet 61:665–677. https://doi.org/10.1007/s00294-015-0495-9
39. Tae KH, Ki-Joong K, Marcel Q (2014) Chloroplast Genome Differences between Asian and American Equisetum arvense (*Equisetaceae*) and the Origin of the Hypervariable trnY-trnE Intergenic Spacer. PLoS ONE 9:e103898. https://doi.org/10.1371/journal.pone.0103898
40. Terakami S, Matsumura Y, Kurita K, Kanamori H, Katayose Y, Yamamoto T, Katayama H (2012) Complete sequence of the chloroplast genome from pear (*Pyrus pyrifolia*): genome structure and comparative analysis. Tree Genet Genomes 8:841–854
41. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S (2017) GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45:W6–W11. https://doi.org/10.1093/nar/gkx391
42. Xu J-H, Liu Q, Hu W, Wang T, Xue Q, Messing J (2015) Dynamics of chloroplast genomes in green plants. Genomics 106:221–231. https://doi.org/10.1016/j.ygeno.2015.07.004
43. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555–556
44. https://doi.org/10.1093/bioinformatics/13.5.555

Figures

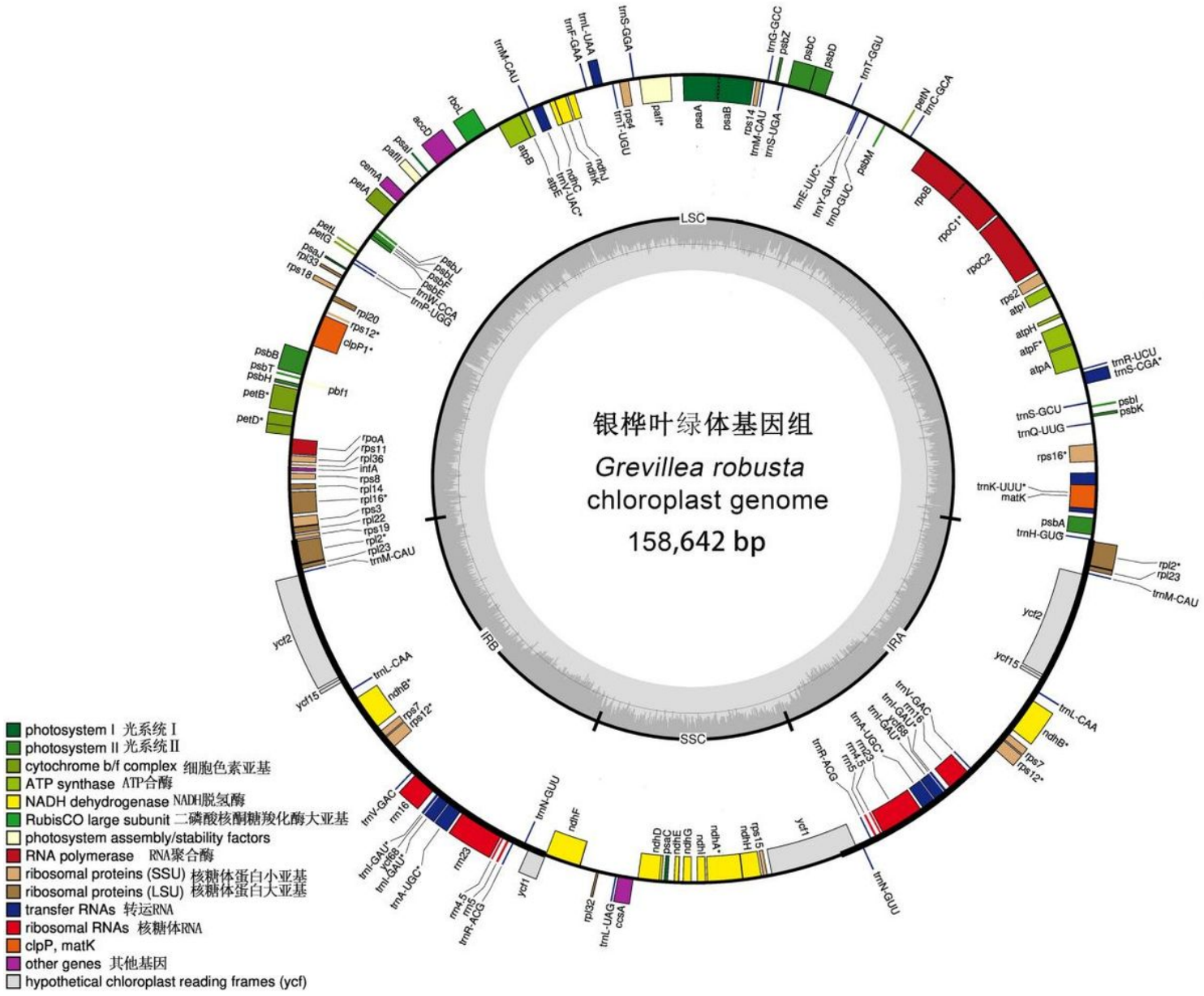

Figure 1

Chloroplast genome map of *Grevillea robusta*

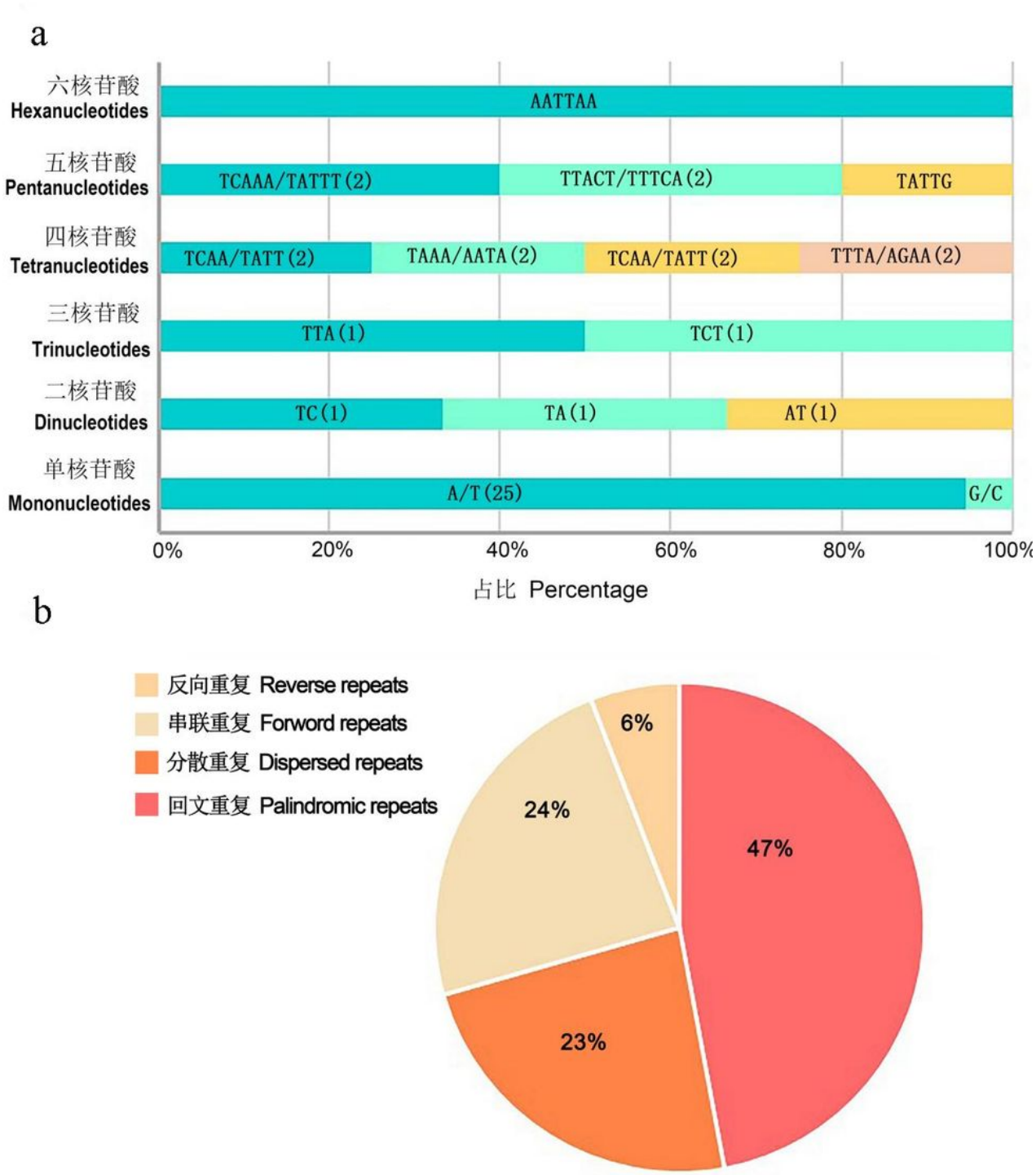

Figure 2

Composition of repeats in the chloroplast genomes of *Grevillea robusta*

a Type and number of SSRs; (b) Percentage of four type repeats

Figure 3

Comparison of chloroplast genome boundaries in *Proteaceae*

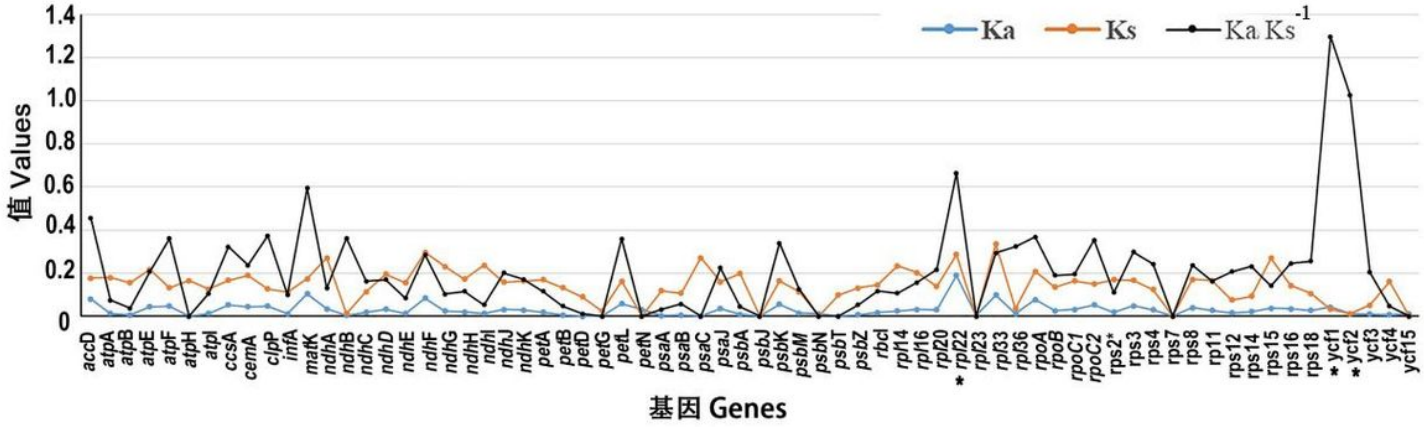

Figure 4

Estimation of synonymous Ks non-synonymous Ka and Ka Ks^{-1} of 4 species in *Proteaceae*

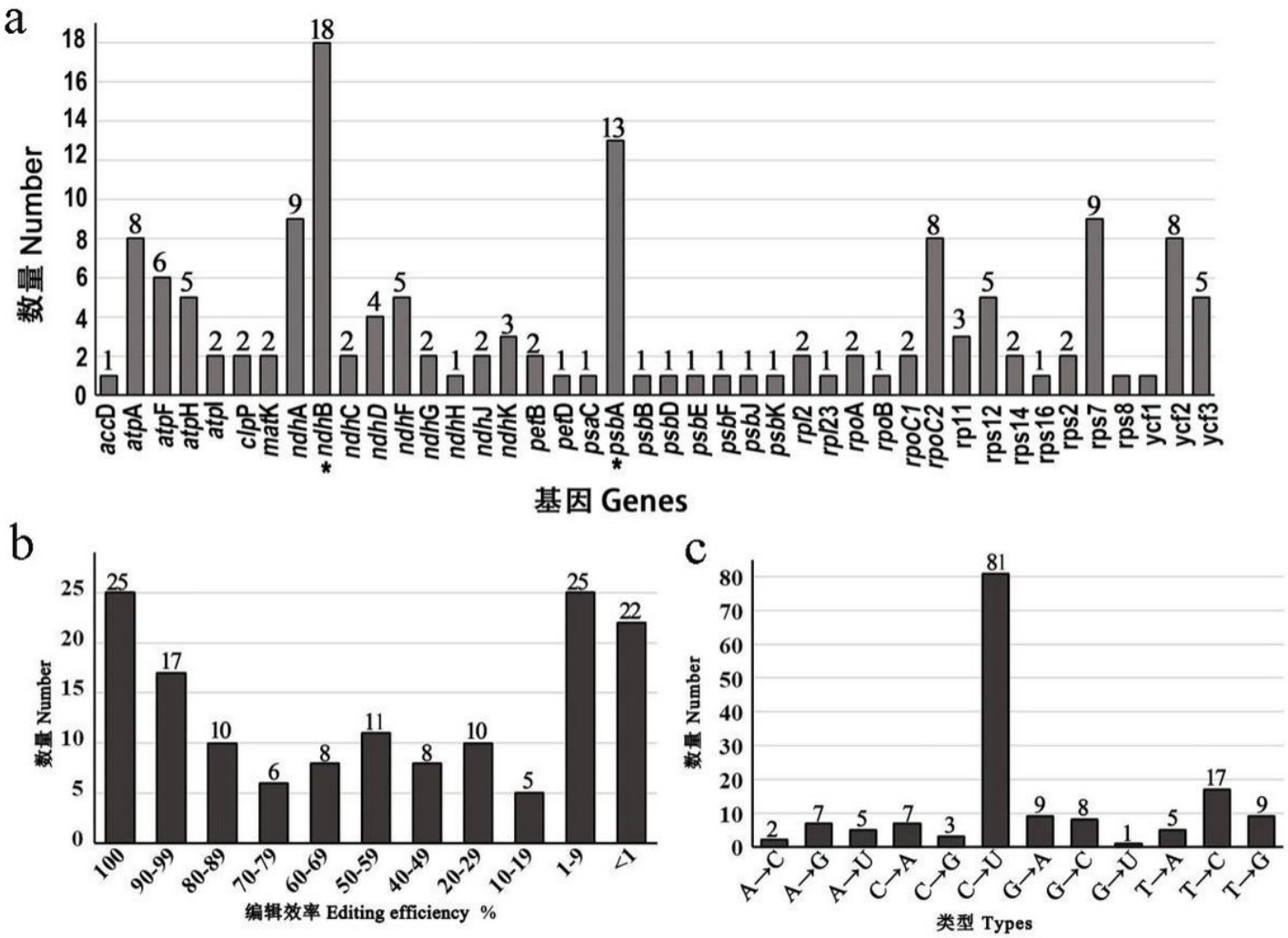

Figure 5

RNA editing sites of the protein-coding gene in chloroplast genome of *Grevillea robusta*

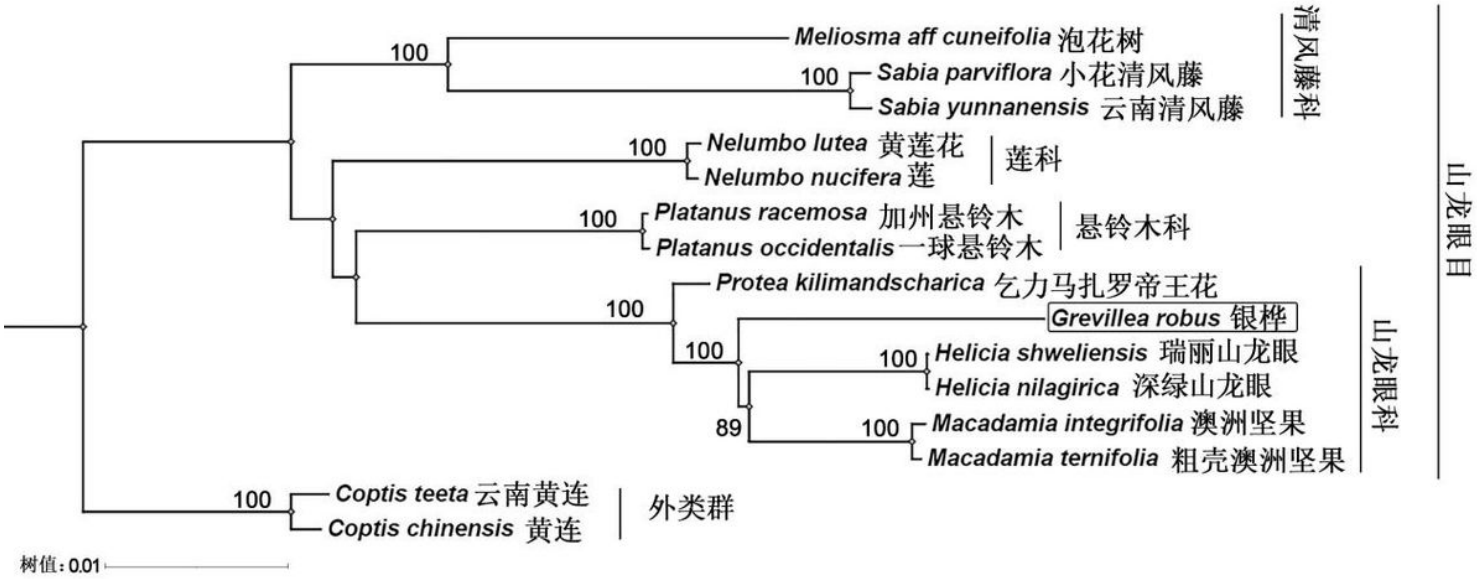

Figure 6

Phylogenetic relationships between *Grevillea robusta* and 15 other plant species