BIODIVERSITAS
Volume 24, Number 7, July 2023
Pages: 3774-3784

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d240715

Comparative of the complete chloroplast genome and RNA editing of *Eucalyptus camaldulensis* T5 clone, an elite variety in Thailand

PHRUET RACHARAK[1,2,♥], ANONGPAT SUTTANGKAKUL[1], SUPACHAI VUTTIPONGCHAIKIJ[1,2,♥♥]

[1]Department of Genetics, Faculty of Science, Kasetsart University. 50 Ngamwongwan Rd, Chatuchak, Bangkok 10900, Thailand. Tel.: +66-2-5625444, [♥]email: fforprr@ku.ac.th, [♥♥]email: fsciscv@ku.ac.th

[2]Center for Advance Studies in Tropical Natural Resources, National Research University, Kasetsart University. 50 Ngamwongwan Rd, Chatuchak, Bangkok 10900, Thailand

Manuscript received: 24 April 2023. Revision accepted: 10 July 2023.

Abstract. *Racharak P, Suttangkakul A,* Vuttipongchaikij S. *2023. Comparative of the complete chloroplast genome and RNA editing of* Eucalyptus camaldulensis *T5 clone, an elite variety in Thailand. Biodiversitas 24: 3774-3784. Eucalyptus camaldulensis*, T5 clone, the excellent fast-growing tree plantation in Thailand, was analyzed for the complete chloroplast genome and RNA editing. The complete chloroplast genome revealed a total genome size of 160,204 bp that divided into a large single copy (LSC) (88,904 bp) and a small single copy region (SSC) (18,506 bp) by inverted repeat regions (IR) containing 26,397 bp. A circular mapping genome and gene order showed the circle antiparallel mapping gene of 135 genes, including 37 tRNAs, 10 rRNAs, and 1 pseudogene. GC content of the genome was 36.87%. The comparative genomes analysis between the T5 clone and *E. camaldulensis* from the NCBI database suggested that the thymine (T) and adenine (A) strongly impacted the indel and transversion process, which could be a point of mutation in the genome. Furthermore, 24 specific genes were used to investigate RNA editing. From all genes, only 11 genes were edited with C to U conversion. Triplet codons, tUA, tUt, tUg and Ugg were the most frequently edited codon and expressions; the crucial influence of amino acid alterations. Due to RNA editing events, the physicochemical properties of amino acids were changed from polar to nonpolar amino acids and from hydrophilic to hydrophobic amino acids. Physicochemical properties conversion is necessary to form complete amino acid sequences for several essential chloroplast proteins. The event might be the accumulation of amino acid alterations causing phenotypic variation for plant adaptation and evolution.

Keywords: *Eucalyptus camaldulensis*, complete genome, chloroplast, plastid, RNA editing

INTRODUCTION

Eucalyptus camaldulensis is an important species for Thai economics; it has been renowned extensively for raw materials in wood, pulp, paper industry, and bioenergy. In Thailand, *E. camaldulensis*, a T5 clone, has been improved by the breeding program to enhance growth and development to adapt and grow in unfavorable areas all over Thailand. This species tolerates various biotic and abiotic stresses more than other species imported for the breeding program, for example, *E. grandis*, *E. globulus*, *E. urophylla*, *E. pellita*, and *E. tereticornis*. *E. camaldulensis*, all T5 clones are the most outstanding salt and drought tolerance for abiotic stress compared to the other species. In the same way, for biotic stresses, a T5 clone can be capable of growing throughout in insect stress areas when it is high above 2-3 m from the ground, such as tolerance to a gall wasp in particular genus *Leptocybe* (Kumar et al. 2015; Mphahlele et al. 2021) Furthermore, they can tolerate to leaf spot, leaf blight, and rust disease. Moreover, *E. camaldulensis* is excellent for propagation in cutting and *in vitro* culture (Mendonça et al. 2016; García-Ramírez 2023).

The chloroplast genome contains approximately ±130 genes, and size varies, ranging from 15,553 to 521,168 base pairs (bp) in *Asarum minus* and *Floydiella terrestris*, respectively, chloroplast nucleoid copy range from 1,000 to 1,700 copies which highly variable number during the plant growth (Dobrogojski 2020). The chloroplast DNA (cpDNA) structure, inverted repeats (IR), the two identical regions separated by large single copy region (LSC) and small single copy region (SSC), have varied in size from 20,000 to 25,000 bp (Morley et al. 2019). In the *Eucalyptus* genus, the first complete chloroplast genome was reported in *E. globulus* (Steane 2005), following *E. grandis* (Paiva et al. 2011). The complete chloroplast genome of the *Eucalyptus* genus has been reported in 29 species, including *E. camaldulensis*. The chloroplast genome of *Eucalyptus* comprises 112 individual genes with different genome sizes ranging from 159,527 bp (*E. oblique*) to 161,071 bp (*E. spathulata*). The total RNA mostly contains 45, and 5 pseudogene, except for *E. grandis*, which only contains 44 genes for total RNA and 6 pseudogene. The total GC contents range from 36.82% to 36.98 %. Exon GC contents range from 39.03% to 39.11%, while intergenic GC contents were 31.28% to 31.61 % (Bayly et al. 2013).

RNA editing is a posttranscriptional event often occurring in plant mitochondria and chloroplast genomes. Conversion of cytidine (C) to uridine (U) is the most frequent expression in the chloroplast genome. Furthermore, RNA editing causes the creation of an initial codon, a vital substance for conserving amino acids for protein synthesis, but it is a rare occurrence; for instance, a change in an ATG codon to an ACG codon in the psbL gene. The stop codon from RNA editing probably causes short amino acid

sequences because the translation process cannot read across the stop codon (Yan et al. 2018; Knoop 2023). All the RNA editing events have shown a tendency of silent editing mutations, which do not change the final amino acid. In addition, the codon at the second position is frequently altered more than the first or third position. Nevertheless, RNA editing aims to improve the specific amino acid residues, protein function, and effective evolutional process (Hein et al. 2016; Huang et al. 2022).

Moreover, RNA editing sites were frequently found in groups of large and small ribosomal protein subunits, which directly function in amino acids produced in the genome. The result of the RNA editing site causes an alteration of physiochemical properties. The polarity was reported to be reversible, changing between polar and nonpolar properties. The hydrophobic amino acid increases from changing the hydrophilic amino acid (Jiang et al. 2012; Ichinose and Sugita 2016). The RNA editing process directly affects plant growth as specific changing in the nucleotide composition of the RNA molecule. The potential impacts of RNA editing on tree plants include gene expression regulation, protein diversification, adaptation to stress, regulation of organelle function, and reproductive pathways. These processes can influence plant growth, allowing trees to adapt to different environmental conditions, physiological requirements, or developmental stages (Booth et al. 2023). In *Arabidopsis*, RNA editing events in chloroplasts led to amino acid changes in specific proteins involved in photosynthesis. The RNA editing process could influence photosynthetic complexes' stability, structure, and function, potentially impacting the tree's photosynthetic efficiency and adaptation to different environmental conditions (Lv et al. 2014; Hackett et al. 2017; Ma et al. 2017). In *Populus trichocarpa*, a plant famous for its importance in phytoremediation has been reported the importance of pentatricopeptide repeat (PPR) proteins involved in the RNA editing event. The changing expression of PPR resulted in biotic and abiotic stress tolerance, such as salicylic acid (SA), methyl jasmonate (MeJA), mechanical wounding, drought, cold, and salinity stress (Xing et al. 2018).

Therefore, RNA editing in trees may contribute to their ability to adapt and survive in diverse ecological niches. It is important to note that the specific effects of RNA editing on tree plants can vary depending on the species, environmental factors, and the genes being edited (Wang et al. 2019). As for this research, *E. camaldulensis*, T5 clone, the famous crop plantation species in Thailand, was analyzed the complete chloroplast genome and RNA editing in levels of DNA, RNA, and protein molecule, which is a renowned application for DNA marker, phylogeny, and evolution study. All result databases would effectively enhance the *E. camaldulensis* breeding program, especially regarding inter and intraspecific hybridization investigation, clone screening selection, and phylogenetic relationship study.

MATERIALS AND METHODS

Study area and plant materials

Eucalyptus camaldulensis, T5 clone, was previously obtained from a *Eucalyptus* breeding program by the Department of Forest Biology, Faculty of Forestry, Kasetsart University, Thailand. Young leaves were harvested and kept in liquid nitrogen for chloroplast DNA and RNA isolation.

Chloroplast DNA isolation

Chloroplast DNA (cpDNA) isolation was modified from Sandbrink et al. (1989), Palmer (1986), and Triboush et al. (1998). Briefly, 50 g of young leaves were ground using a high-speed grinder with liquid nitrogen. The homogenate was ground again with 500 ml STE buffer (400 mM sucrose, 50 mM Tris pH 7.8, 20 mM EDTA-Na2, 0.2% bovine serum albumin, 0.2% β-mercaptoethanol) and filtered through four layers of bandage gauze, followed by filtering through two layers of Miracloth. The flow-through was centrifuged twice at 1,000 rpm at 4°C for 20 min. The supernatant was transferred to a new tube and centrifuged at 4,000 rpm at 4°C for 20 min. The sediment chloroplast was washed with three volumes of wash buffer (10 mM β-mercaptoethanol, 5 mM EDTA, 100 µg/ml proteinase K, 10 mM Tris-HCl (pH 8.0) and centrifuged at 4,000 rpm at 4°C for 20 min. The sediment chloroplast was diluted to 24 ml (depending on tube size) with wash buffer. The sucrose solution was diluted for the step gradient to 52%, 48%, and 30%, respectively. Firstly, 4 ml of 52% and 30% sucrose solutions were pipetted to step gradient tubes, followed by 4 mL sediment chloroplast. The solution was centrifuged at 25,000 rpm, 4°C, for 30 min. Secondly, the chloroplast layer between 52% and 30% sucrose solution was moved to a new tube containing each 3 ml of 52%, 48%, and 30% sucrose solutions, followed by a 3 mL chloroplast layer from the first step. Finally, the chloroplast pellet was transferred to a new tube, washed with three volumes of wash buffer, and centrifuged at 4,000 rpm, 4°C, and 15 min. The chloroplast pellet was kept at -20°C. Following the manufacturer's protocol, the chloroplast DNA was isolated from the chloroplast pellets using DNeasy Plant Mini Kit (QIAGEN).

DNA, RNA isolation, and complementary DNA

Young fresh leave covered with a plastic bag of *E. camaldulensis*, T5 clone, was ground in a mortar with liquid nitrogen to obtain 100 mg powder, and the sample was moved to a new tube. Genomic DNA isolation followed the manufacturer protocol with DNeasy Plant Mini Kit (QIAGEN). For RNA isolation, leave powder samples from the previous step were isolated using the FavorPrepTM Plant Total RNA Mini Kit, Favorgen, Taiwan. SuperScript® III First-Strand Synthesis System, Invitrogen, USA synthesized complementary DNAs (cDNA).

PCR amplification and sequencing for RNA editing

We reviewed and selected highly frequency RNA editing genes from the plant RNA database (Li et al. 2019a). The 24 gene regions were amplified by specific

primers designed by Primer 3 program (Koressaar 2018) and *E. grandis* as nucleotide template (primer data as Table 1). The polymerase chain reaction was performed in 30 µL reaction mixtures, including 50 ng of genomic DNA and cDNA template, 1x PCR buffer (KAPA HiFi), 2.0 mM dNTP, 0.2 pmol of forward and reverse primer, and 0.5 U Taq DNA Polymerase (KAPA HiFi) in a MiniAmp Plus Thermal Cycler (Thermo Fisher Scientific Inc. USA). The PCR thermal cycle for amplification was: 3 min at 94°C, followed by 35 cycles of 30 s at 94°C, 1 min at primer temperature vary and 1 min at 72°C, and a final elongation of 5 min at 72°C. A Gel/PCR purification mini kit (Favorgen) was applied for the PCR product selection and sequencing process carried out by Macrogen, Inc. (Korea).

Data analysis

Genome annotation and RNA editing analysis

The raw Illumina paired-end whole chloroplast genome data of X and Y were trimmed, and low-quality bases were removed using FASTP (Chen et al. 2018). The adapter sequence autodetection, length, and quality filters were enabled by default. The filtered reads were assembled into complete chloroplast genomes using GetOrganelle with organelle type and SPAdes kmer set with the following options: -F embplant_pt and -k 21, 45, 65, 85, 105 (Jin et al. 2020). The automatic annotator GeSeq was used to annotate the chloroplast genome with BLAST searches (default settings) against the chloroplast genomes of sibling species (Tillich et al. 2017). The online program OGDRAW was used to draw the circular chloroplast genome map (default settings), which was manually edited (Greiner et al. 2019). For genetic code translation, Table 2 was used for each coding sequence translation. The Clustal Omega and BLAST program was applied to compare all of the complete chloroplast genomes of *E. camaldulensis* (T5 clone) with the NCBI database (*E. camaldulensis*; NC_022398), which was used as a complete genome template (Freudenthal et al. 2020). Program PREP-Mt (http://prep.unl.edu) was used for RNA prediction. For editing site analysis, cDNA and genomic DNA were compared by applying Clustal Omega (Madeira et al. 2022) and BLAST Program (Boratyn et al. 2019).

RESULTS AND DISCUSSION

The chloroplast genome of the *E. camaldulensis* T5 clone

The T5 clone via ultracentrifugation and sucrose gradient method was used to obtain high-quality cpDNA for genome analysis; the cpDNA was isolated from leaves of *E. camaldulensis*. The cpDNA was subjected to Illumina HiSeq sequencing, and sequence assembly yielded a complete chloroplast genome with 160,204 bp. The complete genome was submitted to the NCBI database as accession number OQ355696. The structure of the chloroplast genome of the *E. camaldulensis* T5 clone included a section of large single copy (LSC) with 88,904 bp and small single copy (SSC) with 18,506 bp, in which these two were separated by inverted repeat regions (IRa and IRb) that contained 26,397 bp in each region (Figure 1). The genome has an antiparallel double-stranded circular DNA, and genes in the inner strand were transcribed counterclockwise. In contrast, those in the outer strand were transcribed clockwise. There are, in total, 135 genes in the genome, comprising 87 protein-coding genes, 10 rRNAs, 37 tRNAs, and 1 pseudogene. The gene sizes varied from 71 to 6,843 bp in trnC-GCA and ycf2, respectively. The GC content was 36.87%. The intron regions mostly appeared in transfer RNA, ribosomal protein large subunit, and NADH dehydrogenase genes, which generally included one intron appearing, except gene pafI and clpP1, which had two intron sections. Gene rps12, located in the LSC region and defined as a divided gene due to a splicing process, also contained two introns.

Comparative analysis of the chloroplast genome of *E. camaldulensis*

Sequence alignment between the chloroplast genome of *E. camaldulensis*, T5 clone, and *E. camaldulensis* (NC_022398; 160,164 bp) used as template comparison showed a total of 239 sites for nucleotide differences (Figure 2). Among these, 87 sites (36.40%) and 152 sites (63.60%) were in the coding and non-coding regions, respectively. The nucleotide differences included 38 sites (15.90%) of base transition, 61 sites (25.52%) of a base transversion, 107 sites (44.77%) of insertion, and 33 sites (13.81%) of deletion. Considering mutations within the coding regions, 20 out of 135 genes had nucleotide variations consisting of atpI, rpoC2, rpoB, rpoA, psbC, psaA, psaC, rpl33, rpl16, rpl22, clpP1, petB, ndhF, ndhD, ccsA, rrn16, ycf2, and ycf1 (two copies of ycf2 and ycf1 were in the IR regions). Transition and transversion played an important role in nucleotide variations for coding regions indicating from higher variation position than insertion and deletion in this study.

Nucleotide variation of transition and transversion events

The transition process of base A (adenine), T (thymine), C (cytosine), and G (guanine) within a group of purine (base A and G) and pyrimidine (base C and T) derivatives of chloroplast genome was examined. The base substitution was more significant within the purine group (21 sites) than in the pyrimidine group (17 sites). In the purine group, base A was changed to G more frequently than changing base G to A. While in pyrimidine, base T was changed to base C. Transversion event, one of the most important for point mutation, exhibited a high ratio of nucleotide variation because bases could be replaced reversely between a group of purine and pyrimidine derivatives resulting in nucleotide changing. Substitution of a base in the purine group by the pyrimidine group was higher than substitution from the pyrimidine to purine group at 52.46% and 47.54%, respectively. Base A and T strongly expressed involvement with base conversion. Base A (purine group) showed conversion to base C (pyrimidine group) more often than base T to G, C to A, and G to T, respectively (Figure 2).

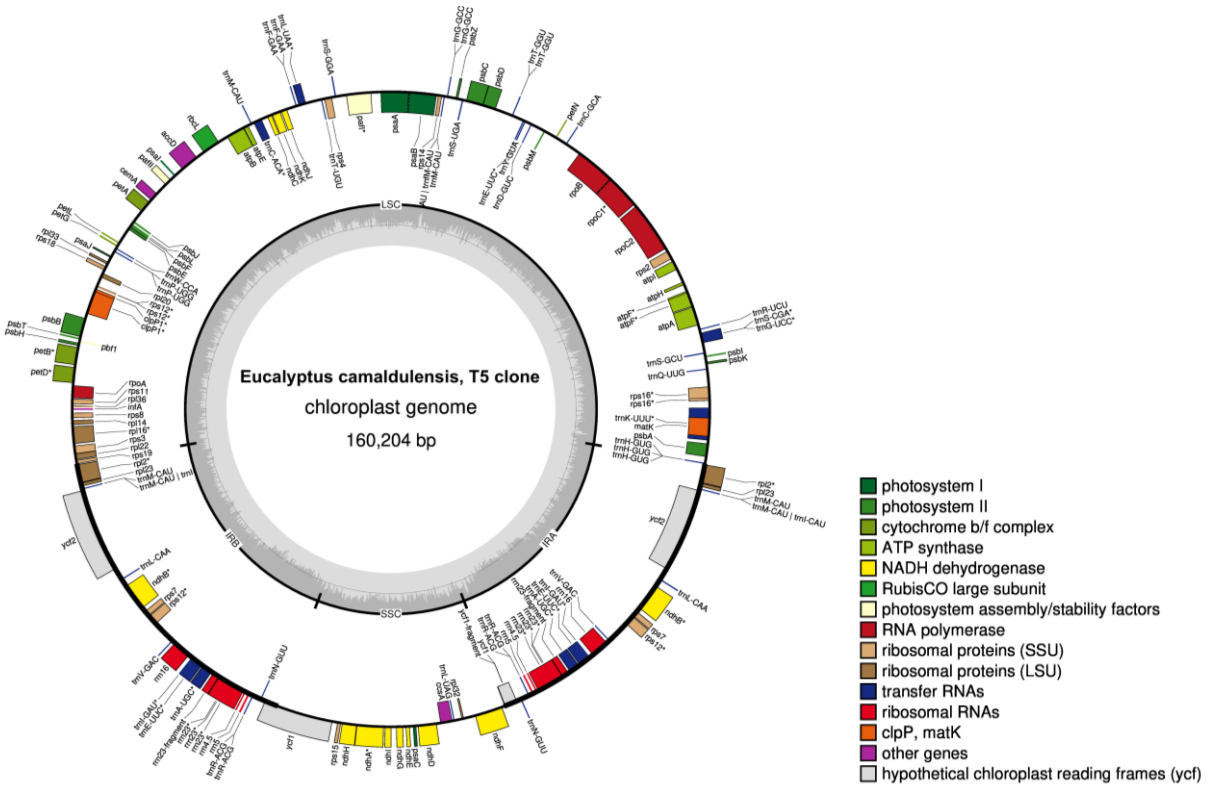

Figure 1. The circular complete chloroplast genome map of *E. camaldulensis*, T5 clone (accession number: OQ355696) drawn by OGDRAW program (Greiner et al. 2019)

Nucleotide variation from indel insertion and deletion events

The insertion process had a high effect on nucleotide variation, especially in the borders of two inverted repeat regions (IRa and IRb) with large single copies (LSC) and small single copies (SSC). From comparative analysis, inserting 35 bp in the IR junction impacted the genome difference in *E. camaldulensis* (NC_022398) and *E. camaldulensis*, T5 clone. Insertions with T and A were highly observed, and insertion of base T and A in the position of poly A and poly T-tail caused nucleotide differences. However, poly A and poly T-tail occurred similarly in deletion and insertion (Figure 3).

Chloroplast RNA editing site analysis

A total of 24 genes of *E. camaldulensis*, T5 clone, 9 genes from genetic system gene groups, 14 genes from photosynthetic system gene groups, and 1 gene from another gene group were analyzed for RNA editing sites. The analysis showed that 11 of 24 genes had edited transcripts: matK, rpl23, rps2, rps14, accD, ndhG, ndhD, ndhB, psbE, psbF, and psbL. Cytidine (C) was edited to uridine (U) in the transcription of all edited genes. Genes in the groups of ribosomal protein large subunit, small subunit, and NADH dehydrogenase had highly frequent RNA editing sites. The highest number of RNA editing sites was found in rpl23 (2.84%), which belongs to the ribosomal protein large subunit group. Genes encoding NADH dehydrogenases (ndh) possess RNA editing, and in this work, the transcripts of 3 out of 4 ndh genes were found to be edited from C to U, including ndhG, ndhD, and ndhB. The effect of RNA editing on psbL and ndhD gene transcribed ACG codon as the initiation for translation codon replacing with ATG codon. The silent editing codons were identified in 3 positions, including 2 positions in gene rpl23 and 1 position in gene ndhD (Table 2).

Regarding codon editing, the second codon position was the highest editing at 54.00%, followed by 20.00% at the first codon and 11.00% at the third codon, in term of start and silent editing site was found at 6.00% and 9.00%, respectively. Of the total 35 codons found in RNA editing, the codon tUa frequently showed higher codon editing than the other codons. The tUa edited codon importantly caused amino acids to change from serine to leucine. The other codons included tUt, tUg and Ugg found often change serine to phenylalanine, serine to leucine and arginine to tryptophan, respectively. The amino acid changing of proline to leucine resulted in 3 types of editing codons comprising cUg, cUa and cUt codon (Table 2). The serine was found highly editing in this study which shows the importance of this amino acid.

Table 1. Specific primer designed for RNA editing investigation

Primer no.	Primer name	Sequence (5'-3')	Product size (bp)	Tm °C	Ta °C	Gene target
1	atpH_246F	TCTGGATGAATCTTAGTGATGGAA	341	60	60	atpH
	atpH_246R	TCGGATTACACAAAGGGATTC		58.9		
2	psbJLFE_791F	AAATTGCGTTGCTGTGTCAG	1033	59.9	60	psbJ, psbL, pdbF, psbE
	psbJLFE_791R	AGGATCCCCTTGCCTTCTTA		60		
3	accD_1495F	TGCTTATTCCGATAGAAAATGAAA	1699	59.2	60	accD
	accD_1495R	TTGACTTTGATTCCGATAAACTACC		59.8		
4	atpI_744F	AAGCAAGTCGAAAAAGAGATGG	970	59.9	60	atpI
	atpI_744R	CAATTCCCTAAAATCGAAATATCA		58.5		
5	ndhG_531F	GACCCATTTTTAATCCATTCG	854	58.3	60	ndhG
	ndhG_531R	TGAACAAGAATTCAACGTATTCA		57.4		
6	ccsA_960F	GTACGCCGCTATGGTGAAAT	1289	60	60	ccsA
	ccsA_960R	AATGGTTCTCGACAGCCTACA		59.8		
7	petLG_383F	AGCTCCTTCATCTTCATGCTT	580	58.2	60	petL, petG
	petLG_383R	GGTAGAACGTGGGTCTCCAA				
8	ndhB_2217F1	GGATCAACTAAGCCCTCTCG	1491	60	60	ndhB
	ndhB_2217R1	ACCGAATCCATTCGATTTCT				
	ndhB_2217F2	CAGGGTCAGGAACAACGAAT	1457	60	60	
	ndhB_2217R2	GGTCCGGTATGGAATGAACT				
9	ndhD_1503F1	GCATGAAACAACTCGAAGCA	1052	60	60	ndhD
	ndhD_1503R1	GATTATTTGGATTGCGCCTA				
	ndhD_1503F2	TTACCAGACACCCATGGAGA	1003	60	60	
	ndhD_1503R2	AGGCTGTCGAGAACCATTTG				
10	rpoB_3219F1	CATTCCATATACGGGGTGAG	1038	60	60	rpoB
	rpoB_3219R1	CCATTCCGAATTTCATTCCA				
	rpoB_3219F2	GTAGGCGGAGATCCGGTATT	1036	60	60	
	rpoB_3219R2	CTGAATCTAGAGCCGCTTGG				
	rpoB_3219F3	CGACACGTCTGAAGGAATCA	1031	60	60	
	rpoB_3219R3	GAATGCCAAGTATGGCTCGT				
	rpoB_3219F4	CGGCGAAAGAATCATCGTAT	1123	60	60	
	rpoB_3219R4	GGGTTTTGTCACCTCTCCAA				
11	matK_1512F1	CGAATTCGATTGAAAAAGAGAGG	1707	60	60	matK
	matK_1512R1	CCATGATCTCATAGACAATCATTCA				
12	ndhF_2206F1	TCATTTGACCAATTCTAACTTCTTG	1214	60	60	ndhF
	ndhF_2206R1	TAAAGCGGCTCGATAAGACC				
	ndhF_2206F2	GGTAGCAGCAGGCATTTTTC	1500	60	60	
	ndhF_2206R2	GACTATTAAATCGAATTAAGAAT				
13	rps2_711F1	GGGGTTTGGATTGTGTATCG	944	60	60	rps2
	rps2_711R1	TCCGCAGTAATTTGGATCTCTT				
14	rps14_303F1	GCTTTCTTGATTGCCTCCAC	638	60	60	rps14
	rps14_303R1	TACCCCGCGATGAAGATAAG				
15	rpl23_282F	TGGAATTGGCTCTGTATCAATG	484	60	60	rpl23
	rpl23_282R	TGCTCGGGGTAGAAGTTTTG				
16	rpl20_354F	CATTGGAAGAATCCGTTTTG	532	60	60	rpl20
	rpl20_354R	ACCTTCCCGGAGTTCATTCT				
17	psbM_105F	TCTCGACGATGAGTTGATTTG	371	60	60	psbM
	psbM_105R	CGAGTCCGCTCATTTCATTT				
18	rpoA_1014F	CCTATGCCACATAATGGCTGT	1284	60	60	rpoA
	rpoA_1014R	TCGCTTCAAAGTGAATTTTCC				
19	rpsl814_962F	TCGGGATAGGAACGTAGAACC	1142	60	60	rps8, rpl 14
	rpsl814_962R	TAGCAAATTCCGGATCCAAT				

Physicochemical property alteration

The physicochemical properties of amino acids are essential for protein structure and interaction, comprised of hydrophilicity, hydrophobicity, and polarity. For this study, the polarity of all amino acids tended to change from a polar amino acid (neutral and basic amino acid form) to a nonpolar amino acid of 76.00% in both aliphatic and aromatic amino acid forms. The polar (neutral amino acid form; PNA) was changed to the nonpolar in both aromatic amino acid form (NAO) and aliphatic amino acid form (NAL) at 26.67% and 73.33%, respectively, except only the polar (basic amino acid form; PBA) changed to NAO at 100%. Moreover, the polar groups (PNA and PBA) were not found to change each other. In contrast, the nonpolar (aliphatic amino acid form, NAL) highly changed within the group to NAO and NAL at 33.33% and 50.00%, respectively. The change of NAL to PNA was only found at 16.17% (Table 2). For example, in the highly changing of serine, polar neutral amino acid, caused to be leucine, nonpolar aliphatic amino acid, was 26.92%, while changing

to be phenylalanine, nonpolar aromatic amino acid, of 11.54% and to serine at 3.85%. Even though the RNA editing event caused codon changes, we found that some amino acid polarity, such as tyrosine, serine, and phenylalanine, were unchanged.

Interestingly, in the overall analysis, most amino acids were changed to nonpolar amino acids in both aliphatic and aromatic amino acids. However, only proline was rarely changed from nonpolar aliphatic amino acids to polar neutral amino acids. The alteration of hydrophilic and hydrophobic properties of amino acids resulted from the codon alterations found that serine, a group of hydrophilic amino acids, was highly changed to leucine and phenylalanine (a group of hydrophobic amino acids). The alteration of hydrophilic amino acid to hydrophobic amino acid was highly observed at 56.25%. The alterations within the hydrophilic amino acid and the hydrophobic amino acid group were observed at 21.88% and 6.25%, respectively. The property of hydrophobic amino acid had inclined to increase.

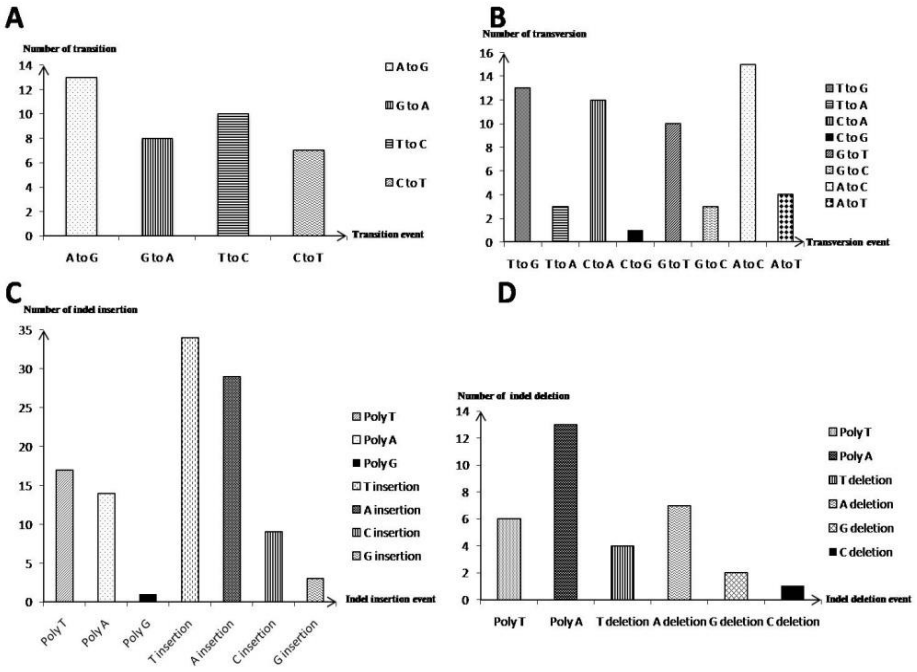

Figure 2. Nucleotide comparison between *E. camaldulensis* from NCBI database using as template comparison and *E. camaldulensis*, T5 clone showed site differences, including the number of transitions (A), number of transversions (B), number of indel insertions (C), and number of indel deletion (D) events that occurred in the complete chloroplast genome

Inverted repeat A section

Eucalyptus camaldulensis TCTTAATAAATGATTCGCTACAAAAGGATTTTTTTTTAGTGAACGTGTCA 88872
Eucalyptus camaldulensis (T5) TCTTAATAAATGATTCGCTACAAAAGGATTTTTTTTTAGTGAACGTGTCA 88869
****

IRa section

Eucalyptus camaldulensis CG ATTAATTACTCCTATTTTTTTTTT---------- 88898
Eucalyptus camaldulensis (T5) CG ATTAATTACTCCTATTTTTTTTTTTATAATTAATAAATTAATTACTCCT 88919
** **********

Eucalyptus camaldulensis ---------TATAATTAATAAAAAGAAATAAATTCGATTTTATCCCCTA 88938
Eucalyptus camaldulensis (T5) ATTTTTTTTTTATAATTAATAAAAAGAAATAAATTCGATTTTATCCCCTA 88969
****

Inverted repeat B section

IRb section

Eucalyptus camaldulensis GATTCCATCGTTTATAGTCGAAAAGAAGAGTCACAAGAG GTTTTTCAAAC 133776
Eucalyptus camaldulensis (T5) GATTCCATCGTTTATAGTCGAAAAGAAGAGTCACAAGAG GTTTTTCAAAC 133819
********** **********

Eucalyptus camaldulensis CATAAAAAAAAATGGATCTTCTTTC---------- 133800
Eucalyptus camaldulensis (T5) CATAAAAAAAAATGGATCTTCTTTCGTTTTTCAAACCATAAAAAAAAATGGA 133869
****

Eucalyptus camaldulensis ---------AATTTAAATATTTCTAACTTCGAATTCTCTTTATTTTTCTT 133841
Eucalyptus camaldulensis (T5) TCTTCTTTCAATTTAAATATTTCTAACTTCGAATTCTCTTTATTTTTATT 133919
********** **

Figure 3. Comparative analysis in inverted repeat region (IRa and IRb) between *E. camaldulensis*, T5 clone, and *E. camaldulensis* (NC_022398) from the NCBI database as a template comparison

Table 2. RNA editing analyses and alterations of amino acid properties of *E. camaldulensis*, T5 clone

Functions	Gene	Base pair	Editing (%)	Base position	Codon position	C-to-U	Codon position 1	Codon position 2	Codon position 3	Start site	Silent site	Edited codons	Amino acid changes	Polarity changes
Intron maturase	matK	1512	0.26	155	52	1		1				tUt	Ser to Phe	PNA to NAO
				161	54	1		1				tUa	Ser to Leu	PNA to NAL
				704	235	1		1				tUt	Ser to Phe	PNA to NAO
				1171	391	1	1					Ugg	Arg to Trp	PBA to NAO
RNA polymerase	rpoB	3219	0											
	rpoA	1014	0											
Ribosomal	rpl23	282	2.84	11	4	1		1				aUc	Thr to Ile	PNA to NAL
proteins				18	6	1			1		1	taU	Tyr to Tyr	not change
(Large subunit)				26, 27	9	1		1	1			tUU	Ser to Phe	PNA to NAO
				39	13	1			1		1	agU	Ser to Ser	not change
				46	16	1	1					Uat	His to Tyr	PBA to NAO
				70, 71	24	1	1	1				UUt	Pro to Phe	NAL to NAO
	rpl20	354	0											
	rpl14	369	0											
Ribosomal	rps2	711	0.28	134	45	1		1				aUa	Thr to Ile	PNA to NAL
proteins				248	83	1		1				tUa	Ser to Leu	PNA to NAL
(small subunit)	rps14	303	0.99	47	16	1		1				cUg	Pro to Leu	NAL to NAL
				110	37	1		1				tUg	Ser to Leu	PNA to NAL
				121	41	1	1					Ugg	Arg to Trp	PBA to NAO
	rps8	405	0											
Acetyl-CoA	accD	1473	0.2	806	269	1		1				tUg	Ser to Leu	PNA to NAL
carboxylase				1397	466	1		1				cUa	Pro to Leu	NAL to NAL
				1415	472	1		1				cUt	Pro to Leu	NAL to NAL
ATP synthase	atpH	246	0											
	atpI	744	0											
Cytochrome *b/f*	petL	96	0											
	petG	114	0											
Cytochrome *c*	ccsA	960	0											
NADH	ndhG	531	0.19	50	17	1		1				tUa	Ser to Leu	PNA to NAL
dehydrogenase	ndhB	2216	0.05	149	50	1		1				tUa	Ser to Leu	PNA to NAL
	ndhD	1503	0.33	2	1	1		1		1		aUg	Thr to Met	PNA to NAL
				313	105	1	1					Ugg	Arg to Trp	PBA to NAO
				383	128	1		1				tUg	Ser to Leu	PNA to NAL
				1033	345	1	1					Utt	Leu to Phe	NAL to NAO
				1041	347	1			1		1	ttU	Phe to Phe	not change
	ndhF	2244	0											
Photosystem II	psbE	252	0.4	214	72	1	1					Uct	Pro to Ser	NAL to PNA
	psbF	120	0.83	77	26	1		1				tUt	Ser to Phe	PNA to NAO
	psbL	117	0.85	2	1	1		1		1		aUg	Thr to Met	PNA to NAL
	psbJ	123	0											
	psbM	105	0											

Note: NAL: Nonpolar (aliphatic amino acid); NAO: Nonpolar (aromatic amino acid); PNA: Polar (neutral amino acid); PBA: Polar (basic amino acid)

Discussion

The complete chloroplast genome of *E. camaldulensis*, T5 clone, comprised 160,204 bp of genome size, which was between genome sizes of the genus *Eucalyptus* (15,553 to 521,168 bp) (Dobrogojski 2020). The genome structure was similar to other genomes of *Eucalyptus* spp, including 10 rRNAs, 37 tRNAs, and 90 protein-coding genes, except for gene rps16, which was not observed in *E. grandis* (Paiva et al. 2011). The rRNA and tRNA genes were the most conserved regions, and in contrast, intergenic spacer and introns were the most variables (Bayly et al. 2013).

The comparative genome analysis between *E. camaldulensis*, T5 clone, and *E. camaldulensis* (NC_022398) found the difference of nucleotide in the border of the two inverted repeats (IRa and IRb) from the indel process. Most of these indels might have originated from slipped-stand mispairing of the surrounding sequence. These are the crucial evidential region that created the nucleotide variation between the species and supported the event of the expansion and contraction of the chloroplast genome (Li et al. 2022). The border between the inverted repeats usually varies between species as one of the highly variable in the chloroplast genome that frequently uses for population level and phylogeographic study in the *Eucalyptus* genus (Morley et al. 2019).

From the comparison of whole genomes, the indel and the transversion process of base thymine (T) and adenine (A) were involved with the difference of nucleotide in the genome, even though the indel was found in the coding region less than the transversion and transition process. The variation in poly A and poly T-tail was common in many genomes. The indel we have discovered might have numerous essential applications in systematic and evolutionary biology, such as elucidating the original distribution species tracking biogeographic moments, and clarifying complex relationships among species (Shi et al. 2023). The indel and transversion of base A and T in the non-coding region occurred more than in the coding region because the coding region has a mechanism to conserve and protect nucleotides from the loss or addition of genes. However, the divergence could be estimated based on the number of nucleotide insertions, deletions, transitions, and transversion in the chloroplast genome (Ichinose and Sugita 2016; Li et al. 2019b; Miao 2022).

The nucleotide variation in coding, non-coding and even variation in inverted repeat regions of chloroplast genomes have affected distribution and evolution. Because from the previous study of phylogenetic trees among *E. globulus*, *E. grandis*, and *E. camaldulensis* exhibited intense support branch lengths that indicated deep divergence. Moreover, these species are distributed in various regions and geographically isolated from each together in Australia (Thornhill et al. 2019). In addition, *E. globulus*, *E. grandis*, and *E. camaldulensis* were classified into sections and series, followed by the difference in morphological phenotype, including section Maidenaria (series Globulares), section Latoangulatae (series Transversae) and section Exsertaria (series Rostratae), respectively. Furthermore, the phylogenetic relationship revealed their relationship of *E. camaldulensis* and *E. globulus* congruently that these the two species were the sister taxon to the *E. grandis* clades with accordant previously molecular (in chloroplast nucleotide variation) and morphological analyses (Bayly et al. 2013).

The comparative genome analysis between *E. camaldulensis*, T5 clone, and *E. camaldulensis* (NC_022398) found the difference of nucleotide in the border of the two inverted repeats (IRa and IRb) from the indel process. Most of these indels might have originated from slipped-stand mispairing of the surrounding sequence. These are the crucial evidential region that created the nucleotide variation between the species and supported the event of the expansion and contraction of the chloroplast genome (Li et al. 2022). The border between the inverted repeats usually varies between species as one of the highly variable in the chloroplast genome that frequently uses for population level and phylogeographic study in the *Eucalyptus* genus (Morley et al. 2019).

From the comparison of whole genomes, the indel and the transversion process of base thymine (T) and adenine (A) were involved with the difference of nucleotide in the genome, even though the indel was found in the coding region less than the transversion and transition process. The variation in poly A and poly T-tail was common in many genomes. The indel we have discovered might have numerous essential applications in systematic and evolutionary biology, such as elucidating the original distribution species tracking biogeographic moments, and clarifying complex relationships among species (Shi et al. 2023). The indel and transversion of base A and T in the non-coding region occurred more than in the coding region because the coding region has a mechanism to conserve and protect nucleotides from the loss or addition of genes. However, the divergence could be estimated based on the number of nucleotide insertions, deletions, transitions, and transversion in the chloroplast genome (Ichinose and Sugita 2016; Li et al. 2019b; Miao 2022).

The nucleotide variation in coding, non-coding and even variation in inverted repeat regions of chloroplast genomes have affected distribution and evolution. Because from the previous study of phylogenetic trees among *E. globulus*, *E. grandis*, and *E. camaldulensis* exhibited intense support branch lengths that indicated deep divergence. Moreover, these species are distributed in various regions and geographically isolated from each together in Australia (Thornhill et al. 2019). In addition, *E. globulus*, *E. grandis*, and *E. camaldulensis* were classified into sections and series, followed by the difference in morphological phenotype, including section Maidenaria (series Globulares), section Latoangulatae (series Transversae) and section Exsertaria (series Rostratae), respectively. Furthermore, the phylogenetic relationship revealed their relationship of *E. camaldulensis* and *E. globulus* congruently that these the two species were the sister taxon to the *E. grandis* clades with accordant previously molecular (in chloroplast nucleotide variation) and morphological analyses (Bayly et al. 2013).

The transition and transversion events can introduce changes in the genotype of a tree, potentially leading to alterations in its phenotype. Depending on the specific

genes and genetic pathways involved, these mutations can affect various aspects, such as leaf color, growth patterns, or responses to environmental stimuli. In *Dendrobium*' Burana Green', a hybrid orchid with variegated leaf phenotype was investigated for VAR2 gene variation. The results showed that transition and transversion events might be involved with variegation leaf patterns and confirmed the hypothesis that these events lead to amino acid modification caused by genetic variations. *E. camaldulensis*, T5 clone, a result of a breeding program, possibly tend to affect transition and transversion events (Putri et al. 2023). Moreover, the variant nucleotides that occurred from transition and transversion events have been used for different genetic studies, from the understanding of the genetic structure of populations to the development of breeding selection markers (Muñoz-Espinoza 2023)

For this research, we first reported RNA editing of *E. camaldulensis*, T5 clone, in which 11 out of 24 genes were investigated for RNA editing. The editing of base cytosine (C) to uracil (U) has commonly appeared in green plants and this research. The member of NADH dehydrogenase, 3 out of 4 genes investigated, were highly found editing in RNA molecules, especially gene ndhD, and ndhC, according to a report in *Arabidopsis thaliana*, including ndhB, ndhD, ndhF, and ndhG, were found in this study (Dobrogojski 2020). Gene matK was found editing in RNA molecule in only *E. camaldulensis*, T5 clone compared with *A. thaliana*, *Cucumis sativus*, and *Amborella trichopoda* (Table 3) (Hein et al. 2016; Ishibashi et al. 2019).

The mutation loci were examined in the coding region totaling 20 out of 135 genes. We found that only the ndhD gene expressed an RNA editing event. Although the group of genes that showed RNA editing was found a similar to the group of mutation loci that occurred in the coding region, it is not affected by RNA editing as the group of NADH dehydrogenase, photosystem II, and ribosomal protein large subunit. RNA editing seemed to be the specific mechanism process for the specific target. The group of genes that showed mutation loci look like conserve genes when considering RNA editing. The genes in which the RNA editing event disappeared included rpoB, rpoA, atpI, ccsA, and ndhF, found in mutation loci in the coding region. The mutation of these genes in the coding region might affect the breeding program for this research. Genes in a group of NADH dehydrogenase have been reported to have highly frequent RNA editing occurring as it regulates gene expression, generates a translational initiation codon, and potentially regulates translation efficiency. However, RNA editing of the NADH dehydrogenase group may depend on the tissue and developmental stage (Shikanai 2015).

RNA editing analysis found that amino acids changed from polar to nonpolar properties. From these changes, the hydrophilic property of amino acids was altered to hydrophobic property. These physicochemical properties directly changed to protein structures, particularly the folding of tertiary and quaternary structures, which involved hydrophobic effect reaction in protein folding processes. These reactions are directly derived from nonpolar amino acids (Qulsum et al. 2019). Therefore, the RNA editing process in post-transcription increased the specification of amino acid properties and effectively enhanced protein formation in the protein structure process (Small et al. 2020).

Although alteration of amino acid can affect physicochemical properties, causing the amino acid sequence to lack the property of protein-protein interactions, especially in the protein domain region, for this study, we were not observed the variation in the protein domain from which amino acid sequence alterations. Because the position of amino acids changing was not posited in the protein domain sequencing; furthermore, the number of amino acids changing appeared in a low-level position that needed more effect on protein domain changing (Edera et al. 2018). In *Eucalyptus* spp. the higher complexity plant, have tended to be more robust due to increasing functional redundancy; at the same time, genes act as the capacity that buffers genotypic variation under normal conditions like RNA editing. For this study, RNA editing was less influence on protein domain alteration. However, the effects of these events might be the accumulation of phenotypic variations and individual-specific variation, in particular protein levels, that all result in functional and evolutionary processes. Consequently, adaptive evolution based on protein sequence alteration would be minimal if mutational events that affected genomic and protein-coding sequences were the only available molecular mechanism to generate new variants. In addition, adaptation due to phenotypic selection is the fast acclimatization of individuals to handle the sudden change in the environment that is usually not heritable to the next generation (Zhang et al. 2020; Knoop 2023). The results revealed a high similarity of amino acid sequence and protein domain. These sequences did not signal the difference in the chloroplast genome regarding species distribution.

Table 3. Comparison of genes in *E. camaldulensis*, T5 clone, resulted from RNA editing with *Arabidopsis thaliana, Cucumis sativus*, and *Amborella trichopoda*

Genes	Species RNA editing occurred			
	***Eucalyptus camaldulensis* (T5 clone)**	***Arabidopsis thaliana****	***Cucumis sativus****	***Amborella trichopoda****
matK	✓			
rpl23	✓	✓	✓	✓
rps2	✓		✓	✓
rps14	✓	✓	✓	✓
accD	✓	✓	✓	✓
ndhG	✓	✓	✓	✓
ndhD	✓	✓	✓	✓
ndhB	✓	✓	✓	✓
psbE	✓	✓		
psbF	✓	✓	✓	
psbL	✓			✓

Note: *Data was modified from Hein et al. (2016)

Eucalyptus camaldulensis, T5 clone, was the complete chloroplast genome with a total genome size of 160,204 bp that indicated to LSC, SSC, and IR section. Genome comparative analysis of *E. camaldulensis*, T5 clone, revealed the significant point mutation comprised transitions, transversions, indel insertions, and deletions caused nucleotide variation and phenotypic evolution. These are a point for further application, which increases efficient improvement in the *E. camaldulensis* breeding program. The RNA editing showed the result of amino acid alteration, which changed physicochemical properties. These editing amino acids might be used for protein study in response to phenotypic variation from the environmental effect.

ACKNOWLEDGEMENTS

We thank the Center for Advanced Studies in Tropical Natural Resources, Kasetsart University Institute for Advanced Studies, Thailand for financial support.

REFERENCES

Bayly MJ, Rigault P, Spokevicius A, Ladiges PY, Ades PK, Anderson C, Bossinger G, Merchant A, Udovicic F, Woodrow IE, Tibbits J. 2013. Chloroplast genome analysis of Australian eucalypts–*Eucalyptus*, *Corymbia*, *Angophora*, *Allosyncarpia* and *Stockwellia* (Myrtaceae). Mol Phylogenet Evol 69: 704-716. DOI: 10.1016/j.ympev.2013.07.006.

Booth BJ, Nourreddine S, Katrekar D, Savva Y, Bose D, Long TJ, Huss DJ, Mali P. 2023. RNA Editing: Expanding the Potential of RNA Therapeutics. Mol Ther 31 (6): 1533-1549. DOI: 10.1016/j.ymthe.2023.01.005.

Boratyn GM, Thierry-Mieg J, Thierry-Mieg D, Busby B, Madden TL. 2019. Magic-BLAST, an accurate RNA-seq aligner for long and short reads. BMC Bioinformatics 20 (1): 405. DOI: 10.1186/s12859-019-2996-x.

Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: An ultra-fast all-in-one FASTQ preprocessor. J Bioinform 34 (17): 884-890. DOI: 10.1093/bioinformatics/bty560.

Dobrogojski J, Adamiec M, Luciński R. 2020. The chloroplast genome: A review. Acta Physiologiae Plantarum 42: 98. DOI: 10.1007/s11738-020-03089-x.

Edera AA, Gandini CL, Sanchez-Puerta MV. 2018. Towards a comprehensive picture of C-to-U RNA editing sites in angiosperm mitochondria. Plant Mol Biol 97:215-231. DOI: 10.1007/s11103-018-0734-9.

Freudenthal JA, Pfaff S, Terhoeven N, Korte A, Ankenbrand MJ, Förster F. 2020. A systematic comparison of chloroplast genome assembly tools. Genome Biol 21: 254. DOI: 10.1186/s13059-020-02153-6.

García-Ramírez Y. 2023. Temporary immersion system for in vitro propagation via organogenesis of forest plant species. Trees 37: 611-626. DOI: 10.1007/s00468-022-02379-w.

Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res 47 (W1): W59-64. DOI: 10.1093/nar/gkz238.

Hackett JB, Shi X, Kobylarz AT, Lucas MK, Wessendorf RL, Hines KM, Bentolila S, Hanson MR, Lu Y. 2017. An organelle RNA recognition motif protein is required for photosystem II subunit psbF transcript editing. Plant Physiol 173: 2278-2293. DOI: 10.1104/pp.16.01623.

Hein A, Polsakiewicz M, Knoop V. 2016. Frequent chloroplast RNA editing in early-branching flowering plants: Pilot studies on angiosperm-wide coexistence of editing sites and their nuclear specificity factors. BMC Evol Biol 16: 23. DOI: 10.1186/s12862-016-0589-0.

Huang G, Ding Q, Xie D, Cai Z, Zhao Z. 2022. Technical challenges in defining RNA modifications. In Seminars in Cell and Developmental Biol 127: 155-165. DOI: 10.1016/j.semcdb.2021.11.009.

Ichinose M, Sugita M. 2016. RNA editing and its molecular mechanism in plant organelles. Genes 8 (1): 5. DOI: 10.3390/genes8010005.

Ishibashi K, Small I, Shikanai T. 2019. Evolutionary model of plastidial RNA editing in angiosperms presumed from genome-wide analysis of *Amborella trichopoda*. Plant Cell Physiol 60 (10): 2141-2151. DOI: 10.1093/pcp/pcz111.

Jiang Y, Fan SL, Song MZ, Yu JN, Yu SX. 2012. Identification of RNA editing sites in cotton (*Gossypium hirsutum*) chloroplasts and editing events that affect secondary and three-dimensional protein structures. Genet Mol Res 11 (2): 987-1001. DOI: 10.4238/2012.

Jin JJ, Yu WB, Yang JB, Song Y, DePamphilis CW, Yi TS, Li DZ. 2020. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol 21: 241. DOI: 10.1186/s13059-020-02154-5.

Knoop, V. 2023. C-to-U and U-to-C: RNA editing in plant organelles and beyond. J Exp Bot 74 (7): 2273-2294. DOI: 10.1093/jxb/erac488.

Koressaar T, Lepamets M, Kaplinski L, Raime K, Andreson R, Remm M. 2018. Primer3_masker: Integrating masking of template sequence with primer design software. Bioinformatics 34 (11): 1937-1938. DOI: 10.1093/bioinformatics/bty036.

Kumar A, Sangha KS, Dhillon GP. 2015. Screening of 19 genotypes of *Eucalyptus* spp. against gall wasp (*Leptocybe invasa*) in North-western India. J For Res 26: 355-359. DOI: 10.1007/s11676-015-0052-x.

Li M, Xia L, Zhang Y, Niu G, Li M, Wang P, Zhang Y, Sang J, Zou D, Hu S, Hao L. 2019a. Plant editosome database: A curated database of RNA editosome in plants. Nucleic Acids Res 47 (D1): D170-174. DOI: 10.1093/nar/gky1026.

Li W, Zhang C, Guo X, Liu Q, Wang K. 2019b. Complete chloroplast genome of *Camellia japonica* genome structures, comparative and phylogenetic analysis. PloS ONE 14 (5): e0216645. DOI: 10.1371/journal.pone.0216645.

Li C, Liu Y, Lin F, Zheng Y, Huang P. 2022. Characterization of the complete chloroplast genome sequences of six *Dalbergia species* and its comparative analysis in the subfamily of Papilionoideae (Fabaceae). PeerJ 10: e13570 DOI: 10.7717/peerj.13570.

Lv HX, Huang C, Guo GQ, Yang ZN. 2014. Roles of the nuclear-encoded chloroplast SMR domain-containing PPR protein SVR7 in photosynthesis and oxidative stress tolerance in Arabidopsis. J Plant Biol 57: 291-301. DOI: 10.1007/s12374-014-0041-1.

Ma F, Hu Y, Ju Y, Jiang Q, Cheng Z, Zhang Q, Sodmergen. 2017. A novel tetratricopeptide repeat protein, WHITE TO GREEN1, is required for early chloroplast development and affects RNA editing in chloroplasts. J Exp Bot 68: 5829-5843. DOI: 10.1093/jxb/erx383.

Madeira F, Pearce M, Tivey AR, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A, Lopez R. 2022. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res 50: W276-279. DOI: 10.1093/nar/gkac240.

Mendonça EG, Stein VC, de Carvalho HH, Santos BR, Beijo LA, Paiva LV. 2016. The use of continuous, temporary immersion bioreactor system and semisolid culture medium for the production of *Eucalyptus camaldulensis* clones. Ciência Florestal 26 (4): 1211-1224. DOI: 10.5902/1980509825112.

Miao H, Bao J, Li X, Ding Z, Tian X. 2022. Comparative analyses of chloroplast genomes in 'Red Fuji'apples: Low rate of chloroplast genome mutations. PeerJ 10: e12927. DOI: 10.7717/peerj.12927.

Morley SA, Peralta-Castro A, Brieba LG, Miller J, Ong KL, Ridge PG, Oliphant A, Aldous S, Nielsen BL. 2019. *Arabidopsis thaliana* organelles mimic the T7 phage DNA replisome with specific interactions between Twinkle protein and DNA polymerases Pol1A and Pol1B. BMC Plant Biol 19 (1): 241. DOI: 10.1186/s12870-019-1854-3.

Mphahlele MM, Isik F, Hodge GR, Myburg AA. 2021. Genomic breeding for diameter growth and tolerance to Leptocybe gall wasp and *Botryosphaeria*/*Teratosphaeria* fungal disease complex in *Eucalyptus grandis*. Front Plant Sci 12: 638969. DOI: 10.3389/fpls.2021.638969.

Muñoz-Espinoza C, Meneses M, Hinrichsen P. 2023. Transcriptomic approach for global distribution of SNP/Indel and plant genotyping. In: Plant Genotyping: Methods and Protocols. Springer, New York.

Paiva JA, Prat E, Vautrin S, Santos MD, San-Clemente H, Brommonschenkel S, Fonseca PG, Grattapaglia D, Song X, Ammiraju JS, Kudrna D. 2011. Advancing *Eucalyptus* genomics: Identification

and sequencing of lignin biosynthesis genes from deep-coverage BAC libraries. BMC Genomics 12: 137. DOI: 10.1186/1471-2164-12-137.

Palmer JD. 1986. Isolation and structural analysis of chloroplast DNA. Methods Enzymol 118: 167-186. DOI: 10.1016/0076-6879(86)18072-4.

Putri SU, Linggabuwana A, Kurniawan FY, Lawrie MD, Semiarti E. 2023. Morphological and molecular characterization of variegated-leaf pattern in *Dendrobium*' Burana Green'Based on the structure of VAR2 gene. In: 3rd International Conference on Smart and Innovative Agriculture (ICoSIA 2022), 22 May 2022.

Qulsum U, Azad MT, Tsukahara T. 2019. Analysis of tissue-specific RNA editing events of genes involved in RNA editing in *Arabidopsis thaliana*. J Plant Biol 62: 351-358. DOI: 10.1007/s12374-018-0452-5.

Sandbrink JM, Vellekoop P, Ham RV, Brederode JV. 1989. A method for evolutionary on RFLP of chloroplast DNA, application to a range of plant species. Biochem Syst Ecol 17: 45-49. DOI: 10.1016/0305-1978(89)90041-0.

Shi W, Song W, Chen Z, Cai H, Gong Q, Liu J, Shi C, Wang S. 2023. Comparative chloroplast genome analyses of diverse *Phoebe* (Lauraceae) species endemic to China provide insight into their phylogeographical origin. PeerJ 11: e14573. DOI: 10.7717/peerj.14573.

Shikanai T. 2015. RNA editing in plants: Machinery and flexibility of site recognition. Biochimica et Biophysica Acta 1847 (9): 779-785. DOI: 10.1016/j.bbabio.2014.12.010.

Small ID, Schallenberg-Rüdinger M, Takenaka M, Mireau H, Ostersetzer-Biran O. 2020. Plant organellar RNA editing: What 30 years of research has revealed. Plant J 101 (5): 1040-56. DOI: 10.1111/tpj.14578.

Steane DA. 2005. Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). DNA Res 12: 215-220. DOI: 10.1093/dnares/dsi006.

Thornhill AH, Crisp MD, Külheim C, Lam KE, Nelson LA, Yeates DK, Miller JT. 2019. A dated molecular perspective of eucalypt taxonomy, evolution and diversification. Australian Syst Bot 32 (1): 29-48. DOI: 10.1071/SB18015.

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 (W1): W6-11. DOI: 10.1093/nar/gkx391.

Triboush SO, Danilenko NG, Davydenko OG. 1998. A method for isolation of chloroplast DNA and mitochondrial DNA from sunflower. Plant Mol Biol Rep 16: 183-189. DOI: 10.1023/A:1007487806583.

Wang D, Meng S, Su W, Bao Y, Lu Y, Yin W, Liu C, Xia X. 2019. Genome-wide analysis of multiple organellar RNA editing factor family in poplar reveals evolution and roles in drought stress. Intl J Mol Sci 20 (6): 1425. DOI: 10.3390/ijms20061425.

Xing H, Fu X, Yang C, Tang X, Guo L, Li C, Xu C, Luo K. 2018. Genome-wide investigation of pentatricopeptide repeat gene family in poplar and their expression analysis in response to biotic and abiotic stresses. Sci Rep 8: 2817. DOI: 10.1038/s41598-018-21269-1.

Yan J, Zhang Q, Yin P. 2018. RNA editing machinery in plant organelles. Sci China Life Sci 61: 162-169. DOI: 10.1007/s11427-017-9170-3.

Zhang J, Fu XX, Li RQ, Zhao X, Liu Y, Li MH, Zwaenepoel A, Ma H, Goffinet B, Guan YL, Xue JY. 2020. The hornwort genome and early land plant evolution. Nat Plants 6: 107-118. DOI: 10.1038/s41477-019-0588-4.