ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/372486411

Novel structures and evolution of tRNA genes: insight into the chloroplast tRNAs of family Sapindaceae

Article *in* Genetic Resources and Crop Evolution · July 2023

DOI: 10.1007/s10722-023-01671-z

CITATIONS
0

READS
60

12 authors, including:

Khurram Shahzad
Chinese Academy of Sciences
39 PUBLICATIONS **871** CITATIONS
SEE PROFILE

Zhonghu Li
Lanzhou University
201 PUBLICATIONS **2,422** CITATIONS
SEE PROFILE

Hanif Khan
Northwest University
20 PUBLICATIONS **75** CITATIONS
SEE PROFILE

Umar Zeb
Northwest University
45 PUBLICATIONS **288** CITATIONS
SEE PROFILE

All content following this page was uploaded by Umar Zeb on 03 August 2023.

The user has requested enhancement of the downloaded file.

Genet Resour Crop Evol
https://doi.org/10.1007/s10722-023-01671-z

RESEARCH ARTICLE

Novel structures and evolution of tRNA genes: insight into the chloroplast tRNAs of family Sapindaceae

Khurram Shahzad · Zhong-Hu Li · Hanif Khan · Azizullah Aziullah · Umar Zeb · Fengjie Cui · Junhu Dai · Saleha Ashfaq · Sajid Fiaz · Eman A. Mahmoud · Ihab Mohamed Moussa · Hosam O. Elansary

Received: 16 May 2023 / Accepted: 5 July 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract Transfer ribonucleic acids (tRNAs) are small non-coding ribonucleic acids that decode messenger RNA sequences and are directly involved in protein synthesis by carrying amino acids to the ribosome. However, the chloroplast genome needs to better understand tRNAs' phylogeny and evolutionary mechanisms. The present study aimed to delineate the novel structural variations and evolutionary characteristics in the chloroplast genome tRNAs of thirty-six Sapindaceae species. Several novel tRNA structures were identified in the Sapindaceae chloroplast genome. The length of tRNAs ranged from 64 to 93 nucleotides, containing 27–29 anticodons. Pair-wise sequence results showed the conserved nucleotide consensus sequence U-U-C-x-A–x-U in Sapindaceae. The structural analysis revealed that, except for a few tRNAs ($tRNA^{His}$, $tRNA^{Gly}$, $tRNA^{Thr}$, $tRNA^{Phe}$, $tRNA^{Try}$, $tRNA^{Met}$, and $tRNA^{Pro}$), all contained a G nucleotide at the 1st position in

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10722-023-01671-z.

K. Shahzad · J. Dai (✉)
Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Chaoyang District, Beijing, People's Republic of China
e-mail: daijh@igsnrr.ac.cn

K. Shahzad
e-mail: khurrams.khurram@gmail.com

K. Shahzad · Z.-H. Li · H. Khan
Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, College of Life Sciences, Northwest University, Xi'an 710069, People's Republic of China
e-mail: lizhonghu@nwu.edu.cn

H. Khan
e-mail: hanif.botany76@gmail.com

A. Aziullah · U. Zeb (✉) · S. Ashfaq
Department of Biology, The University of Haripur, Haripur, Khyber Pakhtunkhwa, Pakistan
e-mail: umar.zeb@uoh.edu.pk

A. Aziullah
e-mail: azizullah@uoh.edu.pk

U. Zeb · F. Cui
School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, People's Republic of China

S. Fiaz (✉)
Department of Plant Breeding and Genetics, The University of Haripur, Haripur 22620, Pakistan
e-mail: sfiaz@uoh.edu.pk

E. A. Mahmoud
Department of Food Industries, Faculty of Agriculture, Damietta University, Damietta, Egypt
e-mail: emanmail2005@yahoo.com

I. M. Moussa
Department of Botany and Microbiology, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia
e-mail: imoussa1@ksu.edu.sa

Published online: 20 July 2023

the acceptor's arm of tRNAs secondary structure. The rate of transition and transversion of tRNAs are Iso-acceptor-specific. Evolutionary analysis revealed that Sapindaceae chloroplast tRNAs might have evolved polyphyletically with a high percentage of gene loss. Phylogenetic analysis revealed that the chloroplast genome's tRNAs evolved from several common ancestors. At the same time, $tRNA^{Val}$ and $tRNA^{Met}$ appear to be the ancestral tRNAs that underwent duplication diversification to give rise to other tRNAs. Our findings will help us understand the evolution of the tRNA and suggest a key role in chloroplast tRNA biology.

Keywords Chloroplast genome · Conservation sequence · Phylogeny · Structure of tRNA · Sapindaceae · Transition · Transversion

Introduction

Transfer ribonucleic acids (tRNAs) are the single largest gene family with the most common ancestral short non-coding RNA belonging to an organism genome (Mallick et al. 2005; Michaud et al. 2011; Zuo et al. 2013). Around five decades ago, the universal genetic code and tRNAs were discovered, and they reshaped life science by giving a structured basis for understanding living systems. tRNA's essential and versatile molecules sustain and maintain the protein translation machinery. According to Robert Holley, they are distinguished by a clover leaf-like structure (Holley et al. 1965). With approximately 70–100 nucleotides, tRNA has a clover leaf-like structure with an acceptor's arm, variable loop or arm, anticodon arm, anticodon loop, D-arm, D-loop, TΨC-loop, and TΨC-arm (Kirchner and Ignatova 2015; Mohanta and Bae 2017). The D-arm is three or four base pairs long, while the acceptor's arm is seven base pairs long. The D-loop is four to twelve nucleotides long. The anticodon arm is five base pairs long, and the anticodon loop is seven nucleotides long. The variable loop is four to twenty-three nucleotides long, TΨC-loop is five base pairs, and TΨC-arm is seven nucleotides long (Kirchner and Ignatova 2015). The variable number of bases in the D-loop and the variable loop of tRNAs causes heterogeneity. There are 21 different iso-acceptors, 20 for each amino acid and one tRNA charged with pyrrolysine. (Seligmann and Warthi 2019). tRNA is also a multifunctional molecule that can be contained in the cytoplasm, chloroplasts, and mitochondria (Mallick et al. 2005; Michaud et al. 2011). Over the years, comprehensive research on nuclear tRNA's characterization has gotten much coverage (Kirchner and Ignatova 2015; Mohanta and Bae 2017). Maréchal-Drouard et al. (1991) used short sequence fragment markers to determine the structure and function of tRNAs and their genes in the chloroplast genome.

It is easy to isolate the complete chloroplast genome and the tRNAs to resurging in various areas beyond translation, including the environmental adaptation of translation (Schimmel 2018). These functions are related to iso-acceptors and iso-decoders in tRNA genes. Different tRNA base changes, protein binding partner affinity, and tRNA divergence events add up to a significant amount of complexity. tRNA is also involved in the tRNA-derived RNA silencing method and is a significant source of short interspersed nuclear elements (SINE) (Bermudez-Santana et al. 2010; Phizicky and Hopper 2010). The primary function of tRNAs is to decode the mRNA sequence and deliver specific amino acids from aminoacyl-tRNA synthetases (aaRSs) to the ribosome (Schimmel 2018). The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. However, within the chloroplast genome, tRNA machinery is poorly understood.

Sapindaceae (flowering plants) comprises 138 genera with 1,751 economically and ecologically important species. However, this family includes some important tree genera, e.g., *Acer* and *Dipteronia*, which are valuable economic, medicinal, and biofuel energy sustenance sources. Sapindaceae has a classic geographic and phylogeny disjunction trend covering Europe, northern Africa, Asia, and North America, with eastern Asia providing the most plant diversity (Nguyen 2006; Wen et al. 2016; Harris et al. 2017). The Sapindaceae family offers a rare platform to study evolution and genomic preservation. Gleiser and Verdu (2005) documented the Sapindaceae phylogenetic pattern using de Jong's (1976) method, they interpreted character transforms from previously representative

H. O. Elansary
Department of Plant Production, College of Food and Agriculture Sciences, King Saud University, P.O. Box 2460, Riyadh 11451, Saudi Arabia
e-mail: helansary@ksu.edu.sa

phylogenies that included 29, 37, and 39 species, respectively (with partial taxon overlap between investigations). This is an ideal model family to understand the tRNA evolution process and the origin and maintenance of biodiversity because of their rapid growth rates and great value for landscape, widely cultivated and exploited (Saeki and Murakami 2009).

Photosynthesis and metabolic processes in photoautotrophic plants to balance the internal and outer variations in weather, often not included into terrestrial vegetation models that control our biosphere (Knorr and Heimann 2001). Chloroplasts are multi-copy cellular organelles (Wise and Kenneth Hoober 2007). In plants, the chloroplast genome is more conserved than the nuclear genome, with two inverted repeats (IR) regions differentiated by large single-copy (LSC) and single small-copy (SSC) regions (Bendich 2004; Wang et al. 2008; Hereward et al. 2018). Recent research on the chloroplast genome of *Gossypium* and Adoxaceae plants analyzed to identify tRNA sequences resulted in several novel features (Zhong et al. 2021; Zhang et al. 2021). These observations of chloroplast tRNA contribute to a better understanding of tRNA science. However, earlier research needed to have conserved genomic specifics of tRNAs, phylogeny, and ancestors' tRNA relationship due to inadequate complete genomic knowledge of chloroplast tRNA. The tRNA sequences were studied to systematically analyze tRNAs, considering the evolutionary lineages to understand the detailed molecular aspects and novel tRNA structures of chloroplast tRNA in plants. As a result, thorough research of chloroplast tRNAs in the Sapindaceae family will shed light on tRNAs' genomics and evolution.

In the present study, thirty-six chloroplast genomes of the family Sapindaceae were compared, and tRNA sequences were isolated from delineating the genomic architecture, sequence conservation, and differences among tRNAs of each species. Here our primary aims were to identify the tRNAs in the chloroplast genome of family Sapindaceae, to determine the evolutionary relationships among tRNAs in Sapindaceae and investigate their general and unique characteristics of tRNA genes that experienced duplication replication and diversification.

Materials and methods

Annotation and extraction of chloroplast tRNAs of family Sapindaceae

The chloroplast genomes of 36 species from ten genera of the family Sapindaceae were used for tRNA analysis. Fifteen species of Genus Acer were used from our recent published study (Dong et al. 2021), and 21 species of the family Sapindaceae were downloaded from Gene bank (GB) format in the public database available at the National Center for Biotechnology Information (NCBI, HTTP:// www.ncbi.nlm.nih.gov/) (Hiratsuka et al. 1989; Shahid Masood et al. 2004; Saski et al. 2007). All analyzed sequences information has been given in Supplementary Table S1. The chloroplast genomes were annotated using Geneious R v9.0.5 (Kearse et al. 2012). After annotation, nucleotide sequences of 36 Sapindaceae species' chloroplast tRNA genes were collected and used in the rest of the analysis. The RNAalifold webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAalifold.cgi) was used to calculate the free energy of predicted novel tRNAs using default parameters (Bernhart et al. 2008).

tRNAscan-Se analysis of chloroplast tRNAs

The tRNA sequences of 36 Sapindaceae species were examined using ARAGORN and the tRNAscan-Se server (http://lowelab.ucsc.edu/tRNAscan-SE/) (Lowe and Chan 2016). In ARAGORN, the genomic tRNA sequences were analyzed using default parameters. The tRNAscan-Se server was used to examine genomic tRNA with the following parameters: the sequence source is bacterial; the query sequences are formatted (FASTA); the search mode is the default, and the universal genetic code is used to predict tRNA isotypes. The number and structure of nucleotides in various arms and loops were subsequently measured using the same parameters for all tRNAs. Putative novel tRNAs were identified as tRNAs with a structure different from the canonical clover leaf-like structure found in tRNA.

Alignment and recognition of conserved nucleotides.

Twenty tRNA isotypes were grouped separately to classify the conserved nucleotide sequences of

tRNA's. Later, the Multalin server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_multalin.html) was used to align multiple sequences of tRNA isotypes (Mohanta et al. 2019). The alignment evaluation used the following criteria: alignment matrix, Blosum61-12–2; gap penalty at opening and expansion, default; gap penalty at extremities, zero; gap penalty at extremities, null; and one iteration only, none. The maximum alignment consensus level (default) was maintained at 90%, and the lowest alignment consensus level was maintained at 50%. (default). Red indicates 90% sequence conservation in the displayed alignments, while blue indicates less than 90%. Black alignments show no conservation.

Phylogenetic analysis

A phylogenetic tree was built using MEGA v X.0 (Kumar et al. 2018)software to examine chloroplast tRNAs' evolution in 36 Sapindaceae species (Tamura et al. 2013). Before constructing the phylogenetic tree, the Clustal omega server (https://www.ebi.ac.uk/Tools/msa/clustalo/) was used to build a Clustal file of all the tRNAs. The generated Clustal file of tRNAs was translated to a MEGA file type using the MEGAX program. Model selection was carried out before the construction of the phylogenetic tree. MEGAX program used the following statistical parameters to pick models: sample, model collection (ML); tree to use, automatic (neighbor-joining); statistical technique, maximum likelihood; replacement type, nucleotide; gaps/missing data treatment, partial deletion; and site coverage cutoff was 95%. The tree was rooted to an outgroup species belonging to mRNA of *Asteraceae*. A model selection study that yielded the lowest Bayesian information criterion (BIC) determined the best model for constructing the phylogenetic tree. The Kimura2 + G + I model had the lowest BIC score of 6,798.78, so it built a phylogenetic tree. Within the Kimura2 + G + I model, the following statistical parameters were examined for phylogeny reconstruction: gaps/missing data care, partial deletion; site coverage cutoff, 95%; and branch swap filter, very high. Rates across pages, Gamma-distributed with invariant sites (G + I), number of separate gamma groups, 5; gaps/missing data care, partial deletion; site coverage cutoff, 95%; and branch swap filter, very strong. Due to many branches and low resolution, we converted the final tree into Newick format (.nwk). Then, we compressed it with Figtree using the collapse option and made one branch of similar tRNAs.

Transition and transversion analysis

To evaluate the transition/transversion rate for all tRNAs, the same MEGA file format was used to build the phylogenetic tree. The transition/transversion rates of each of the 20 tRNA isotypes were also investigated separately. Multiple sequence alignment of the tRNA isotypes was performed using the Clustal omega server to create a Clustal file for each isotype. The substitution rate was calculated using MEGA X software (Kumar et al. 2018) after the produced Clustal files of tRNA isotypes were converted to a MEGA X file format. The transition/transversion rates of Sapindaceae plants were studied using the following statistical parameters: model/method, Kimura2-parameter model; rates among sites, Gamma distributed (G); many discrete Gamma categories, 5; gaps/missing data treatment, partial deletion; statistical method, maximum likelihood; substitution type, nucleotide; statistical method, maximum likelihood; substitution type, nucleotide; model/process, Kimura2-parameter model; rates among sites, Gamma distributed (G); many discrete Gamma categories, 5; branch swap buffer, very powerful.

Disparity index analysis

A disparity index test of the pattern heterogeneity was performed on all nucleotide substitutions to decide the ones that occurred homogeneously (at equivalent rates) throughout evolution (Mohanta et al. 2019). The statistical parameters used to analyze the determinants of the homogeneity study were the discrepancy index test of substitution pattern homogeneity, range, sequence pairs; Monte Carlo Replications, 1000; substitution type, nucleotide; gaps/missing data treatment, partial deletion; and site coverage cutoff 95%.

Duplication and loss of genes

Researchers used the NCBI taxonomy browser to investigate the replication and lack of tRNA genes to build an all-species tree (https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi). The

gene tree was based on the evolutionary study's phylogenetic tree, while the species tree was constructed using all species (Supplementary Table S1). Notung v 2.9 software (Chen et al. 2000) was used to investigate gene duplication/loss events. Due to low resolution and many branches, we use the Figtree option (Collapse) to compress the tree branches by making one clade of similar tRNAs.

Results

Features of chloroplast tRNA of Sapindaceae

Sapindaceae plants' whole chloroplast genome sequences were used to investigate the tRNA features and evolutions (Supplementary Table S1). After collecting sequences from Geneious software, the tRNAscan-Se server was used to check the genomic tRNA sequences. Types of tRNA, types of anticodons, and the total number of tRNA examined in 36 species of family Sapindaceae are in Supplementary Table S2. Results showed that *Acer ceriferum, Acer coriaceifolium, Acer buergerianum*, and *Koelreuteria paniculata* contain 37 tRNAs. In comparison, *Acer laevigatum* and *Sapindus mukorossi* contain 38 tRNAs except *Dodonaea viscosa* containing 34 tRNAs. All other species have a total number of 37 tRNAs (Supplementary Table S2). Types of anticodons vary in some species, such as *Acer ceriferum*, *Acer coriaceifolium*, *Acer sino-oblongum, Aesculus wangii*, *Dodonaea viscosa* contains 27 types of anticodons. Except for *Dimocarpus longan*, *Koelreuteria paniculate*, *Litchi chinensis*, *Sapindus mukorossi*, and *Xanthoceras sorbifolium* there are 29 types of anticodons. All other species contain 28 anticodon types. *Acer ceriferum, Acer coriaceifolium, Dodonaea viscosa,* and *Xanthoceras sorbifolium* comprise 19 types of tRNAs, while all other species have 20 different types of RNAs (Supplementary Table S2).

Conserved nucleotides length in chloroplast tRNAs

The chloroplast tRNAs' length ranged from 64 nucleotides ($tRNA^{Met}$ CAT) to 93 nucleotides ($tRNA^{Ser}$ TGA). $tRNA^{Gly}$ TCC and $tRNA^{Cys}$ GCA of *Acer burgarianum*, *Dimocarpus logan, Dodonaea viscosa*, *Koelreuteria paniculata,* and *Litchi chinensis* have 71 nucleotides (Supplementary Table S3). In contrast, the $tRNA^{Ser}$ TGA of *Acer burgarianum*, *Dimocarpus logan*, *Koelreuteria paniculate*, and *Litchi chinensis* contained 92 nucleotides $tRNA^{Ser}$ TGA and $tRNA^{Phe}$ GAA of *Dodonaea viscosa* that contained 93 and 71 nucleotides, respectively. The $tRNA^{Ser}$ GGA and $tRNA^{Ser}$ GCT have 87 and 88 nucleotides, respectively, except *Koelreuteria paniculate* lacks $tRNA^{Ser}$ GCT. On average, chloroplast tRNAs in the examined family Sapindaceae contain 74 nucleotides. $tRNA^{Cys}$ and $tRNA^{Gly}$ have 71 nucleotides, $tRNA^{Lys}$, $tRNA^{Gln}$, $tRNA^{Gly}$, $tRNA^{Arg}$, RNA^{Val}, and $tRNA^{Asn}$, contained 72 nucleotides, $tRNA^{Ala}$, $tRNA^{Glu}$, and $tRNA^{Phe}$ contained 73 nucleotides, $tRNA^{His}$, $tRNA^{Asp}$, $tRNA^{Thr}$, $tRNA^{fMet}$, $tRNA^{Trp}$, $tRNA^{Pro}$, $tRNA^{Arg}$, and $tRNA^{Ile2}$ were contained 74 nucleotides. The $tRNA^{Thr}$ and $tRNA^{Ile}$ contained 75 and 77 nucleotides, respectively. Additionally, $tRNA^{Leu}$ (TAG) contains 80 nucleotides, while $tRNA^{Leu}$ (CAA) contains 81 nucleotides in all examined species of the family Sapindaceae (Supplementary Table S3).

Number of anticodons found in chloroplast tRNAs

The investigated family Sapindaceae chloroplast genomes encode only 27–29 anti-codons (Supplementary Table S4). The most commonly found chloroplast genome's tRNA and their anticodons are listed in (Supplementary Table S4). The chloroplast genome of *Acer ceriferum, Acer coriaceifolium*, and *Dodonaea viscosa* contains 19 anticodons for 27 tRNAs (Supplementary Table S4). In contrast, the chloroplast genome of *Xanthoceras sorbifolium* contains 29 anticodons that encoded 19 tRNA. Amazingly, we find that the chloroplast genome of *Dimocarpus logan*, *Koelreuteria paniculata, Litchi chinensis*, and *Sapindus mukorossi* 29 anticodons that encoded 27 tRNAs. At the same time, *Acer sino-oblongum* and *Aesculus wangii* have 27 tRNAs that encode 27 anticodons. All other species of the family Sapindaceae contains 20 tRNA encodes 28 anticodons. Table 1 shows the distribution of tRNA isotypes in the chloroplast genome of 36 Sapindaceae plants.

The anticodon GAA ($tRNA^{Phe}$) was missing in the *Acer ceriferum* and *Acer coriaceifolium* chloroplast genome. In contrast, the anticodon GUG ($tRNA^{His}$) and CCA ($tRNA^{Trp}$) were missing in the chloroplast genome of the *Xanthoceras sorbifolium* and *Dodonaea viscosa* species, respectively (Supplementary Table S5). Similarly, the anticodons

Table 1 Distribution of tRNA isotypes in the chloroplast genome of 36 Sapindaceae plants

Species	Ala	Gly	Pro	Thr	Val	Ser	Arg	Leu	Phe	Asn	Lys	Asp	Glu	His	Gln	Ile	Met	Tyr	Sup	Cys	Trp
Acer amplum	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer burgarianum	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	1	1	0	1	1
Acer ceriferum	2	2	1	3	3	3	3	3	0	2	1	1	1	1	1	4	2	1	0	1	1
Acer caesium	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer cappadocicum	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer catapifolium	2	2	1	2	3	3	3	3	1	2	1	1	1	1	1	4	3	1	0	1	1
Acer caudatifolium	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer coriaceifolium	2	2	1	3	3	3	3	3	0	2	1	1	1	1	1	4	2	1	0	1	1
Acer davidii	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer fenzelianum	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer flabellatum	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer griseum	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer laevigatum	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	3	1	0	1	1
Acer lucidum	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer miaotaiense	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer morrisonense	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer oblongum	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer palmatum	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer pentaphyllum	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer pictum	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer sino-oblongum	1	1	2	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer stachyophyllum	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer tataricum	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer truncatum	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer wilsonii	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Acer yangbiense	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Aesculus wangii	2	2	1	3	2	3	3	3	1	2	1	1	1	1	1	4	3	1	0	1	1
Dipteronia dyeriana	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Dipteronia sinensis	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Dimocarpus logan	2	2	1	2	3	3	3	4	1	2	1	1	1	1	1	4	2	1	0	1	1
Dodonaea viscosa	2	1	1	2	3	3	3	4	1	2	1	1	1	1	1	4	1	1	0	1	0
Eurycorymbus cavaleriei	2	2	1	3	3	3	3	3	1	2	1	1	1	1	1	4	2	1	0	1	1
Koelreuteria paniculata	2	2	1	2	3	3	3	4	1	2	1	1	1	1	1	4	1	1	0	1	1

Table 1 (continued)

Species	Ala	Gly	Pro	Thr	Val	Ser	Arg	Leu	Phe	Asn	Lys	Asp	Glu	His	Gln	Ile	Met	Tyr	Sup	Cys	Trp
Litchi chinensis	2	2	1	2	3	3	3	4	1	2	1	1	1	1	1	4	2	1	0	1	1
Sapindus mukorossi	2	2	1	3	3	3	3	4	1	2	1	1	1	1	1	4	2	1	0	1	1
Xanthoceras sorbifolium	2	2	1	3	3	3	3	4	1	2	1	1	1	0	1	4	2	1	0	1	1

UCC ($tRNA^{Gly}$) was missing in the genome of *Acer sino-oblongum* and *Dodonaea viscosa*. In comparison, UAC ($tRNA^{Val}$) anticodon was found missing in the genome of *Aesculus wangii* (Supplementary Table S5). This absence refers to all tRNAs not encoded in the genome and others that show critical sequence variations that would prevent their activity as canonical translation adaptor molecules (Ehrlich et al. 2021). Besides, the anti-codons UAA ($tRNA^{Leu}$) was present only in *Dimocarpus logan*, *Dodonaea viscosa*, *Koelreuteria paniculate*, *Litchi chinensis*, *Sapindus mukorossi*, and *Xanthoceras sorbifolium*. Anticodon GGU ($tRNA^{Thr}$) was missing in *Acer catapifolium*, *Dimocarpus logan*, *Dodonaea viscosa*, *Koelreuteria paniculate,* and *Litchi chinensis*. Similarly, another repeated anticodon CAU ($tRNA^{Met}$) was missing in *Acer buergerianum, Dodonaea viscosa*, and *Koelreuteria paniculate* species, and found repentance in *Acer catapifolium*, *Acer laevigatum*, and *Aesculus wangii.* Outside of the 28 anticodons mentioned above, the 36 anticodons were not found in any tRNAs of the family Sapindaceae investigated chloroplast genomes (Seligmann 2015) (Supplementary Table S5).

Conservation sequences of chloroplast tRNAs

Numerous similar sequence alignments of all 20 Sapindaceae genes found small, strongly preserved consensus sequences in the pseudo-uridine (TΨC) loop, although not in the other parts of the tRNA (Table 2). In the TΨC -loop, a conserved U-U-C-x-A–x-U consensus nucleotide sequence was discovered. A G nucleotide was detected in the first place in most of the tRNAs. Instead of a G, $tRNA^{His}$, $tRNA^{Gly}$, $tRNA^{Thr}$, $tRNA^{Phe}$, $tRNA^{Try}$, $tRNA^{Met}$, and $tRNA^{Pro}$ were discovered to have a particular nucleotide (x) in the first place (Table 2). Previously, a primordial code was discovered in the tRNA acceptor stem of some tRNAs with presumed ancient cognate amino acids. The tRNAs for cognates A, D, G, and V, in which the 5′ acceptor stems have at positions 3–5 nucleotide triplets coding for the amino acid that is the tRNA's cognate amino acid (Demongeot and Seligmann 2020). The first position in the acceptor's arm of $tRNA^{Gln}$ and $tRNA^{Asn}$ comprises a U nucleotide, while the first position in the acceptor's arm of $tRNA^{Val}$ contains an A nucleotide. In the 5′-acceptor arms, however, there was no evidence of consensus sequence conservation.

Table 2 Multiple sequence alignment and the presence of isotype-specific conserved nucleotide consensus sequence in chloroplast tRNA of Sapindaceae family. Asterisks (*) indicate no conserved sequence

Type	AC arm	D arm	D loop	ANC arm	ANC loop	Variable region	Ψ-arm	Ψ-loop
Histidine	X_2-G-X_3-G	G-C-C	A-A-G-U-G-G-A-U-C-A-A	G-U-G-G-A	U-U-G-U-G-A-A	C-A-U-G-C	G-C-G-G-G	U-U-C-A-A-U-U
Lysine	G-G-G-U-U-G-C	A-C-U-C	A-A-C-G-G-U-A	U-C-G-G	C-U-U-U-U-A-A	C-U-A-G-U-U	C-C-G-G-G	U-U-C-G-A-G-U
Glutamine	U-G-G-G-G-C-G	G-C-C	A-A-G-X-G-G-U-A-A	X_2-G-G	U-U-U-U-G-G-U	X_3-U-X-C	G-G-A-G-G	U-U-C-G-A-A-U
Glycine	X_2-G-G-X_2-A	G-U-X_{0-2}	X-G-X_6-A-A-A	X_3-C-U	X-U-X-C-C-A-A	A-X_{3-4}	G-C-G-G-G	U-U-C-G-A-U-U
Cysteine	G-G-C-G-G-C-A	G-C-C	G-A-G-C-G-G-U-A-A	G-G-G-G-A	C-U-G-C-A-A-A	U-U-U-C	C-C-C-A-G	U-U-C-A-A-A-U
Aspartate	G-G-G-A-U-U-G	G-U-U-C	A-A-U-U-G-G-U-C-A	C-C-G-C-C	C-U-G-U-C-A-A	A-A-G-C-U	X-C-G-G-G	U-U-C-G-A-G-C
Tyrosine	G-G-G-U-C-G-A	C-C-C-G	A-G-C-G-G-U-U-A-A	X-C-G-G-A	C-U-G-U-A-A-A	G-G-C-A	G-C-U-G-G	U-U-C-A-A-A-U
Glutamate	G-C-C-C-C-X-A	G-U-C-U	A-G-X-G-G-U-U-C-A	U-C-U-C-U	C-U-U-U-C-A-A	C-A-G-C	G-G-G-G-A	U-U-C-G-A-C-U
Threonine	X_{4-7}-U	X-C-U-C	A-G-X-G-G-X-U-A	X-C-G-C-X	X_3-G-U-A-A	X2-G-U-C	X-U-C-G-X_{1-2}	X-U-C-X_3-U
Phenylalanine	X_3-G-X-G-A	G-C-U-C	A-G-X_2-G-G-U-A	G-A-G-G-A	C-U-G-A-A-A-A	G-U-G-U-C	A-C-C-A-G	U-U-C-A-A-A-U
Valine	A-G-G-G-C-U-A	X-C-U-C	A-G-X_{3-4}-G-G-U-A	*****	U-U-X-A-C-X_{2-3}-C	A-A-G-X-U-C	X_2-C-X-G	U-U-C-G-A-G-X
Tryptophan	X-C-G-C-U-C-U	G-U-U-C	A-G-U-U-C-G-G-U-A	U-G-G-G-U	C-U-C-C-A-A-A	A-U-G-U-C	G-U-A-G-G	U-U-C-A-A-A-U
Proline	X-G-X_3-U-G	G-C-G-C	A-G-C-U-U-G-G-U-A	U-U-U-G-U	U-U-U-G-G-G-U	A-U-G-U-C	A-C-G-G-G	U-U-C-A-A-A-U
Asparagine	U-C-C-U-C-A-G	G-C-U-C	A-G-U-G-G-U-A	G-U-C-G-G	C-U-G-U-U-A-A	U-G-G-U-C	G-U-A-G-G	U-U-C-G-A-A-U
Leucine	G-X_6	G-X-G	A-A-A-U-X-G-X_4-A	*****	C-U-X-A-X_{2-4}-A	**********	X_3-G-G	X_5-G-U
Alanine	G-G-G-G-A-U-A	G-C-U-C	A-G-U-U-G-G-U-A	C-C-G-C-U	C-U-U-G-C-A-A	A-U-G-U-C	A-G-C-G-G	U-U-C-G-A-G-U
Serine	G-X_5-A	G-C-X_{1-2}	X_4-G-X_{4-5}-A	*****	X-U-X_3-A-X	X_2-A-X_{3-5}	*****	X_3-G-X_{1-3}
Arginine	G-X-G-X-C-X_2	G-X_3	A-X_2-G-G-A-U-X_{0-1}-A	X_3-G-X	C-U-X-C-X-A-A	X_2-G-U-X_{0-1}	X_3-G-G	U-U-C-X-A-A-U
Methionine	X-C-X_4	X_3-C	A-G-U-X_{3-4}-U-A	*****	X-U-C-A-U-A-X	X_2-G-U-C	A-X_2-G-G	U-U-C-A-A-A-U
Isoleucine	G-X_6	G-C-U-X_{0-1}	X_{3-4}-G-G-U-X_{1-3}	C-X-C-X_2	C-U-X-A-U-A-A-X_{0-4}	A-X_2-U-C-X_{0-1}	X_3-G-G	U-U-C-A-A–X-U

The D-arm's first position contained G nucleotide except for the $tRNA^{Lys}$, $tRNA^{Tyr}$, $tRNA^{Thr}$, $tRNA^{Val}$, $tRNA^{Met}$. Similarly, D-arm had a conserved C nucleotide at the 4th position of the arm. In contrast, some tRNA such as $tRNA^{Glu}$, $tRNA^{Gly}$, $tRNA^{Leu}$, $tRNA^{Ser}$, $tRNA^{Tyr}$, $tRNA^{Arg}$, and $tRNA^{Ile}$ do not possess a C nucleotide at the 4th position of the D-arm. The A box formed by nucleotides 7 to 16 of canonical tRNA has been found to contain two conserved consensus sequences, 7GUGGCNNAGU16- and -GGU-AGNGC15 (- stands for gap and N stands for any nucleotide). According to our findings, just three of

the 20 tRNAs examined have a conserved G nucleotide, and six have an A nucleotide at the 7th position. In most tRNAs, the 12th position (1st nucleotide of D-loop) was conserved with A nucleotide. Except for $tRNA^{Ile}$, the D-last loop's nucleotide was found to be a conserved A nucleotide (Table 2). Box B was formed by the consensus sequence 52GGUUCG ANUCC62, which begins at the 52nd position and ends at the 62nd position of tRNA (Dieci and Sentenac 2003).

According to our findings, the conservation nucleotide sequences in tRNA is family specific. Except for $tRNA^{Cys}$, $tRNA^{Glu}$, $tRNA^{Phe}$, and $tRNA^{Ser}$, the G-G nucleotide at the 52nd and 53rd positions were conserved in most tRNAs the nucleotide sequence U-U-C-x-A–x-U was found to be conserved at the 54th, 55th, 56th, 58th, and 60th positions. At the 54th, 55th, 56th, 58th, 59th, and 60th positions of $tRNA^{Cys}$, a conserved U-U-C-G-A-G-C consensus sequence was discovered instead of U-U-C-x-A–x-U consensus sequence (Table 2). U-U-C-G-A-G-x conserved nucleotides were found in $tRNA^{Val}$. At the 60th position of $tRNA^{Val}$, no conserved nucleotides were detected. Conserved C-U or U-U nucleotides were discovered in the anticodon loop at the 32nd and 33rd positions. Instead of C-U nucleotides, $tRNA^{Gln}$, $tRNA^{His}$, $tRNA^{Pro}$, and $tRNA^{Val}$ contained conserved U-U nucleotides. In most situations, the anticodon loop had a conserved A-A nucleotide at the 38th and 39th positions. Except for nucleotides A-A, $tRNA^{Gln}$, $tRNA^{Pro}$, and $tRNA^{Val}$ had conserved G-U, G-U, and x-C nucleotides. Instead of nucleotide A-A, $tRNA^{Ser}$, $tRNA^{Met}$, and $tRNA^{Ile}$ held a conserved A–x nucleotide at the 38th and 39th positions (Table 2). In the tRNA gene, the chloroplast genome encodes a predefined CCA tail. According to the current report, the 3′-end Sapindaceae family chloroplast genomes of tRNAAla, tRNAArg, tRNAIle, tRNALys, and tRNATyr contain CCA nucleotides. However, the remaining tRNA's do not possess a CCA consensus sequence at their 3′-end (Table 2).

Nucleotide variation in chloroplast tRNA arms and loops

In this analysis, the acceptor arm of chloroplast tRNA was discovered to contain 1–7 nucleotides. There were 35 tRNA sequences with one nucleotide and three nucleotides among the 1,327 tRNA sequences representing 36 Sapindaceae species. In comparison, 27 tRNAs have five nucleotides, 72 have six nucleotides, and the remaining 1,192 (89.82%) tRNAs have seven nucleotides (Table 2). The D-arm of each tRNA was found to have 2–4 nucleotides, with no tRNA having fewer than two or more than four nucleotides in the D-arm. Two, three, and four nucleotides were found in 37 tRNA, 368 (27.73%) tRNA, and 922 (69.48%) tRNA. Seven to eleven nucleotides made up the D-loop, which is part of the A box. 328 (24.72%) of the 1,327 tRNAs have seven nucleotides, 239 (18.01%) have eight, 429 (32.32%) have nine, 146 (11.00%) have ten, and 185 (13.94%) have eleven nucleotides (Table 2). The chloroplast tRNAs' anticodon arm had 4–5 nucleotides. The anticodon arms of 142 (10.27%) of the 1,327 tRNAs contain four nucleotides, whereas 922 (89.20%) contain five nucleotides (Supplementary Table S6). Except for 37 tRNAs, which have nine nucleotides in the anticodon loop, all the tRNAs had seven nucleotides in the anticodon loop. Frameshifting protein translation can occur when tRNAs with enlarged anticodons insert amino acids at isolated mono- or dinucleotide-expanded codons (Seligmann and Warthi 2019). The variable loop was discovered to contain a wide range of nucleotides (3–16), such as 72 (5.43%) tRNA containing three nucleotides, 109 (8.21%) tRNA containing four nucleotides, 890 (67.1%) tRNA containing five nucleotides, 179 (13.50%) tRNA contain six nucleotides, 36 (2.71%) tRNA contain eight nucleotides, 36 (2.71%) tRNA contain eleven nucleotides, and 5 (0.40%) tRNA contain sixteen nucleotides. None of the chloroplast tRNAs possessed 1, 2, 7, 9, 10, 12, 13, 14, 15, or more nucleotides in the variable loop (Supplementary Table S1). $tRNA^{Leu}$ (TAA) $tRNA^{Leu}$ (TAG) had 11 or more nucleotides, respectively, whereas the other tRNAs possessed less than ten nucleotides in the variable loop (Supplementary Table S1). Among 1,327 examined tRNA sequences, only 72 $tRNA^{Leu}$ (CAA) genes and 36 $tRNA^{Ser}$ (GGA) genes had four nucleotides in the TΨC-arm, while the remaining tRNA sequences had five nucleotides. Similarly, in the TΨC-loop region in all the 1,327 tRNAs, only 72 $tRNA^{Leu}$ (CAA) genes and 72 $tRNA^{Ser}$ (GGA) genes had five nucleotides, while the remaining tRNA sequences had seven. The acceptor's arm had seven base pairs, and the D-arm had three to four base pairs, with significant differences in the other parts. The anticodon loop is 7 or 9 nucleotides

long, and the anticodon arm is 4–5 base pairs long. The TΨC-loop had several nucleotides ranging from 3 to 16, with no tRNA containing more than 16 nucleotides. The TΨC-arm had 4–5 nucleotides (Tables 2, Supplementary Table S6).

Novel tRNAs structure encoded in the chloroplast genome

The chloroplast genome was discovered to express a few prospective novel tRNAs (Fig. 1, 2, 3). $tRNA^{Thr}$ (GGU) (ΔG = −19.20 (joules/mol)) contains five nucleotides in their acceptor arm (Fig. 1a). In some cases, e.g., the *Acer palmatum* acceptor arm contains three nucleotides in this tRNA. On the other hand, some species, such as *Acer sino-Oblongum, Acer wilsonii Sapindus mukorossi,* and *Xanthoceras sorbifolium* have only one nucleotide in the accept arm. Another, $tRNA^{Thr}$ (UGU) (ΔG = −30.50 (joules/mol) acceptor arm contains eight nucleotides (Fig. 1b). The $tRNA^{Met}$ (CAU) (ΔG = −14.40 (joules/mol)) total length is 64 nucleotides. Simultaneously, the accept arm contains four nucleotides at the 3′-end but does not contain any nucleotide at the 5′-end (Fig. 1c), in some species such as *Acer wilsonii*, *Acer catalpifolium, Aesculus wangii*, and *Acer laevigatum*; this tRNA is doubled. The duplicate tRNA total length is 70 nucleotides, and the acceptor's arm contains seven nucleotides. This tRNA contains 73 nucleotides in total length and seven in the acceptor arm in two species, i.e., *Dimocarpus logan* and *Litchi chinensis;* this tRNA was found very novel. Similarly, we found a repeated $tRNA^{Pro}$ (UGG) (ΔG = -22.78 (joules/mol)) found only in one species, i.e., *Acer sino-Oblongum* contains seven nucleotides in the 3′-end of his acceptor's arm but five nucleotides in the 5′-end (Fig. 1d). The $tRNA^{Ser}$ (UGA) (ΔG = -32.00 (joules/mol)) found in all species of the family Sapindaceae that contains eight nucleotides in their variable loop contains more nucleotides than the anticodon loop (Fig. 2a). This tRNA acceptor arm contains seven nucleotides at 3′-end, and five nucleotides at 5′-end,

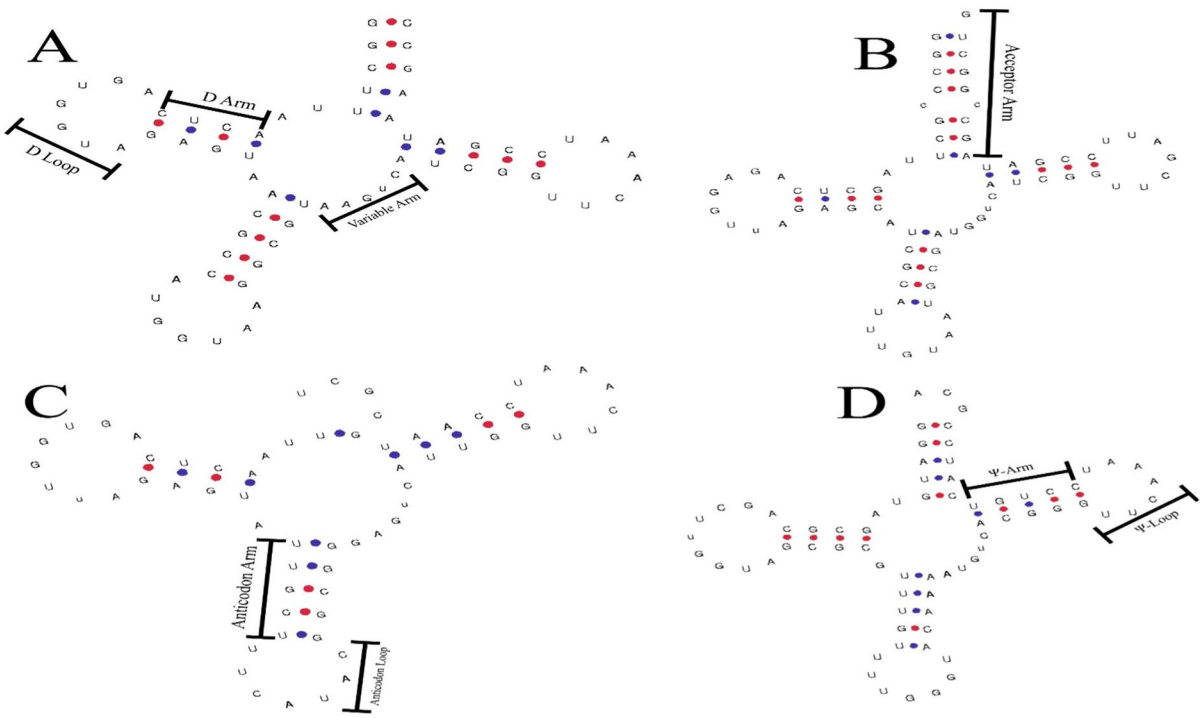

Fig. 1 Structure of novel chloroplast tRNAs containing acceptor arm nucleotides more than usual. **a** The $tRNA^{Thr}$ (GGU) acceptor arm contains five nucleotides, **b** The $tRNA^{Thr}$ (UGU) accept arm contains eight nucleotides, **c** The $tRNA^{Met}$ (CAU) accept arm contains four nucleotides at 3′-end. Still, there is no nucleotide at the 5′-end, **d** The $tRNA^{Pro}$ (UGG) present in *Acer sino-Oblongum* that contains seven nucleotides in the 3′-end but five nucleotides in the 5′-end of accept the arm

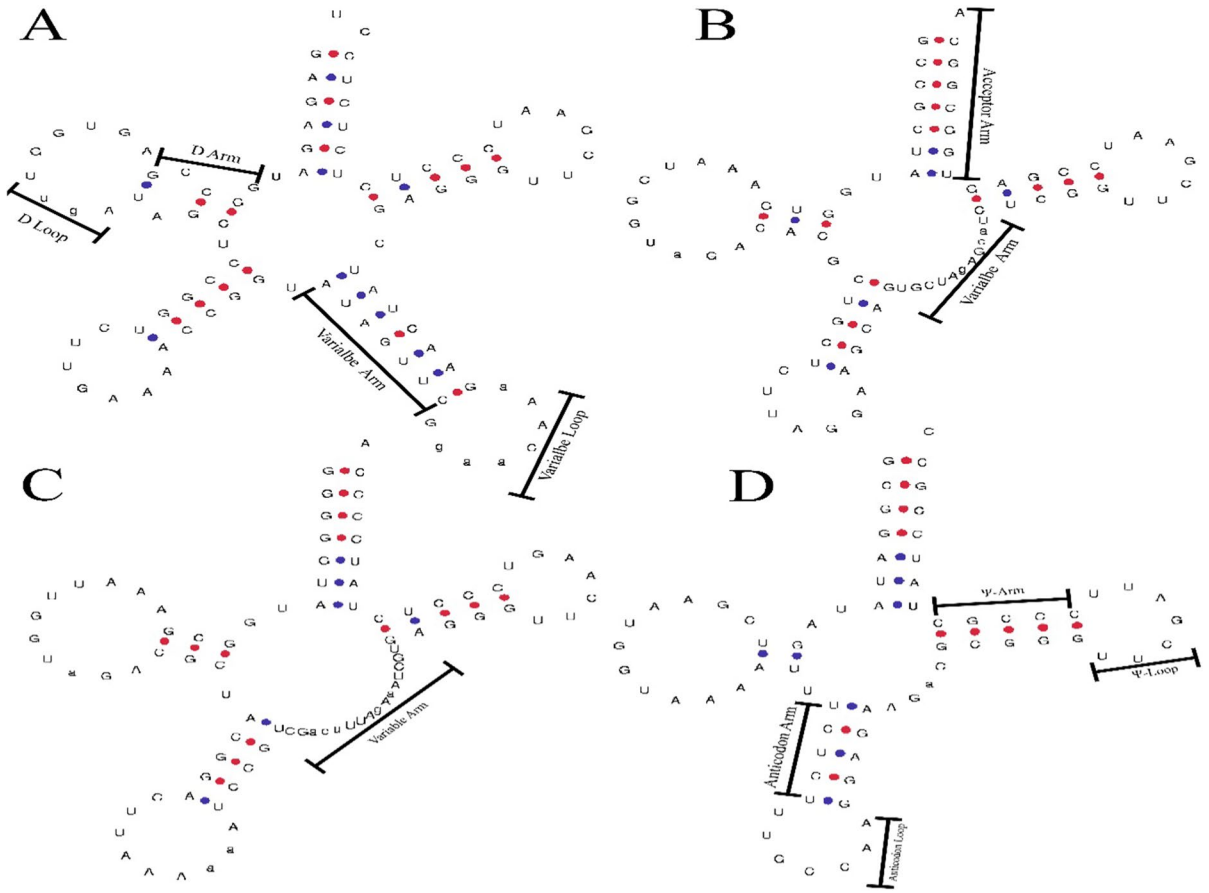

Fig. 2 Structure of chloroplast tRNAs shows novel characters by more than standard nucleotides in their loops and arms structures. **a** The $tRNA^{Ser}$ (UGA) found in all Sapindaceae species contains more nucleotides than the anticodon loop, e.g., eight nucleotides, **b** The $tRNA^{Leu}$ (UAG) contains eleven nucleotides in its D-loop and variable loop, **c** The $tRNA^{Leu}$ (UAA) contains eleven and sixteen nucleotides in its D-loop and variable loop, respectively, **d** The $tRNA^{Gly}$ (GCC) comprises eleven nucleotides with a bigger D-loop in all Sapindaceae species

normally the acceptor arm has seven nucleotides in every tRNA. The $tRNA^{Leu}$ (UAG) ($\Delta G = -30.46$ (joules/mol)) D-loop and variable loop contain eleven nucleotides each, which is called a novel character (Fig. 2b). Additionally, there is only one $tRNA^{Leu}$ found with (UAA) ($\Delta G = -27.71$ (joules/mol)) anticodon with amazing features (Fig. 2c). This tRNA's total length is 87 nucleotides, while D-loop and variable loop contain eleven and sixteen nucleotides, respectively. Another feature of this tRNA contained nine nucleotides in the anticodon loop. Similarly, another $tRNA^{Leu}$ (CAA) contains 81 nucleotides. It contains the largest variable loop with 16 nucleotides. This tRNA contains different characteristics because the D-loop and variable loop contain eleven and nine nucleotides, respectively, different from $tRNA^{Leu}$ (CAA). This novel tRNA was found only in six species among all examined species, i.e., *Dimocarpus logan*, *Dodonaea viscosa*, *Litchi chinensis, Koelreuteria paniculate*, *Sapindus mukorossi*, and *Xanthoceras sorbifolium*. Another $tRNA^{Gly}$ (GCC) ($\Delta G = -25.47$ (joules/mol)) found in all examined species with a bigger D-loop contains eleven nucleotides (Fig. 2d). The nucleotide-based anticodon arm, anticodon loop, and variable arm are different from $tRNA^{Gly}$ UCC. We discovered four nucleotides in the anticodon arm and nine nucleotides in the anticodon loop in the $tRNA^{Val}$ (UAC) ($G = -27.20$ (joules/mol)). (Fig. 3a). We checked whether or not the presence of nine nucleotides were introns or not?

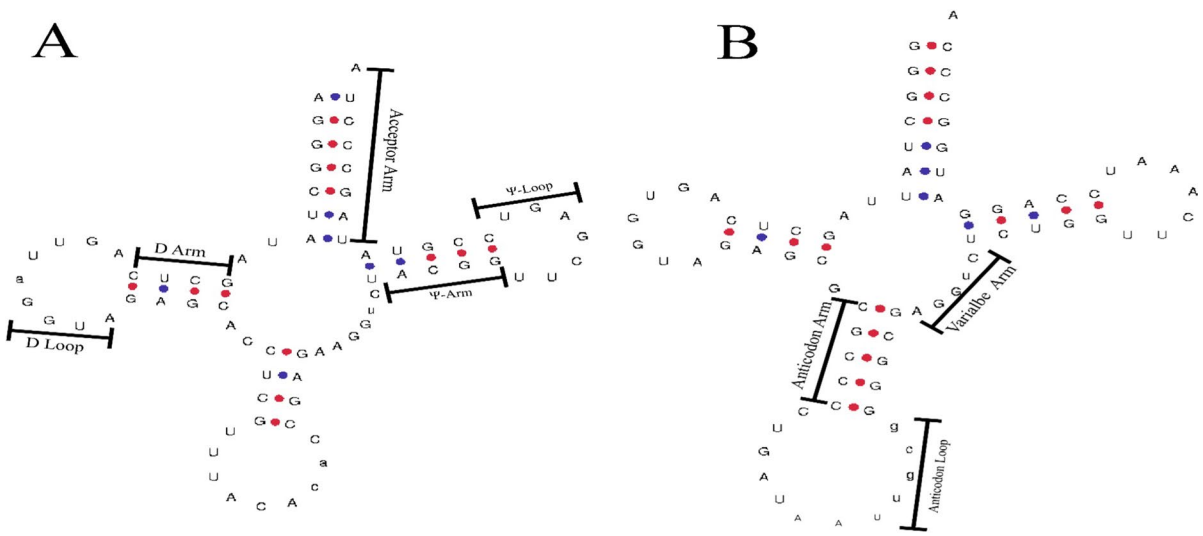

Fig. 3 tRNA structures with novel anticodon loops. **a** The anticodon loop of $tRNA^{Val}$ (UAC) contains nine nucleotides while only four nucleotides are present in its anticodon arm, **b** The $tRNA^{Ile}$ (GAU) contains 12 nucleotides in the anticodon loop

We visualized the predicted genes/exons match with detected RNAs. Similarly, the anticodon loop of $tRNA^{Ile}$ (GAU) ($\Delta G = -33.40$ (joules/mol)) was found to contain 12 nucleotides (Fig. 3b). All these tRNA's repeat in all species with the above amazing features. Many related tRNA structures have been discovered in cyanobacteria and plants' genomic tRNA (unpublished data).

CAU anticodon codes for $tRNA^{Ile}$ in chloroplast tRNAs

The CAU anticodon is unique to $tRNA^{Met}$ because it has only one iso-acceptor. We also found a CAU anticodon in $tRNA^{Met}$, and the $tRNA^{Ile}$ of chloroplast tRNA. The $tRNA^{Ile}$ found in all family Sapindaceae chloroplast genomes encodes a CAU anticodon. Another study documented a CAU anticodon in chloroplast $tRNA^{Ile}$ of the monocot family (Mohanta et al. 2019).

Chloroplast tRNAs descended from several shared ancestors.

The tRNA sequences in the chloroplast genomes of all the Sapindaceae plants studied were used to build a phylogenetic tree. Three major clades, each with 30 tRNAs grouped, were discovered through phylogenetic analysis. Clade I contain $tRNA^{Cys}$, $tRNA^{Tyr}$, $tRNA^{His}$, $tRNA^{Gln}$, $tRNA^{Ser}$, and $tRNA^{Leu}$. Clade II contains $tRNA^{Gly}$, $tRNA^{Glu}$, $tRNA^{Ala}$, $tRNA^{Asp}$, $tRNA^{Val}$, and $tRNA^{Pro}$. Clade III contains $tRNA^{Phe}$, $tRNA^{Trp}$, $tRNA^{Lys}$, $tRNA^{Ile}$, $tRNA^{Arg}$, $tRNA^{Thr}$, $tRNA^{Met}$, and $tRNA^{Asn}$. There are 10, 8, and 12 groups in clade I, II, and III, respectively (Fig. 4). In clade I, $tRNA^{Leu}$ and $tRNA^{Ser}$ are grouped three times and present in the same clade. Each tRNA from both is continuously grouped in the same branch. While the other tRNAs present once in clade I, are grouped once and present on the other branch of Clade I. In clade II, $tRNA^{Gly}$ is grouped twice in the same branch as $tRNA^{Glu}$. $tRNA^{Val}$ has also grouped twice sister with $tRNA^{Pro}$, but all other groups are grouped once. In group III, $tRNA^{Ile}$, $tRNA^{Arg}$, $tRNA^{Thr}$, and $tRNA^{Met}$ are grouped twice. $tRNA^{Ile}$, once grouped with $tRNA^{Lys}$ and other presents as single brach near $tRNA^{Asn}$. Similarly, $tRNA^{Thr}$ is grouped twice in the same branch sister with $tRNA^{Met}$, while the other $tRNA^{fMet}$ is grouped with $tRNA^{Phe}$. $tRNA^{Arg}$ grouped twice, once a sister branch containing $tRNA^{Trp}$ and the other presenting as a single branch (Fig. 4). No tRNA is repeated in clades I, II, and III of the gene genealogy trees. The tRNAs with the anticodon G-A-C and U-A-C of $tRNA^{Val}$, G-G-U and U-G-U of $tRNA^{Thr}$, C-A-U of $tRNA^{Met}$, U-G-A, G-C-U, and G-G-A of $tRNA^{Ser}$, G-C–C and

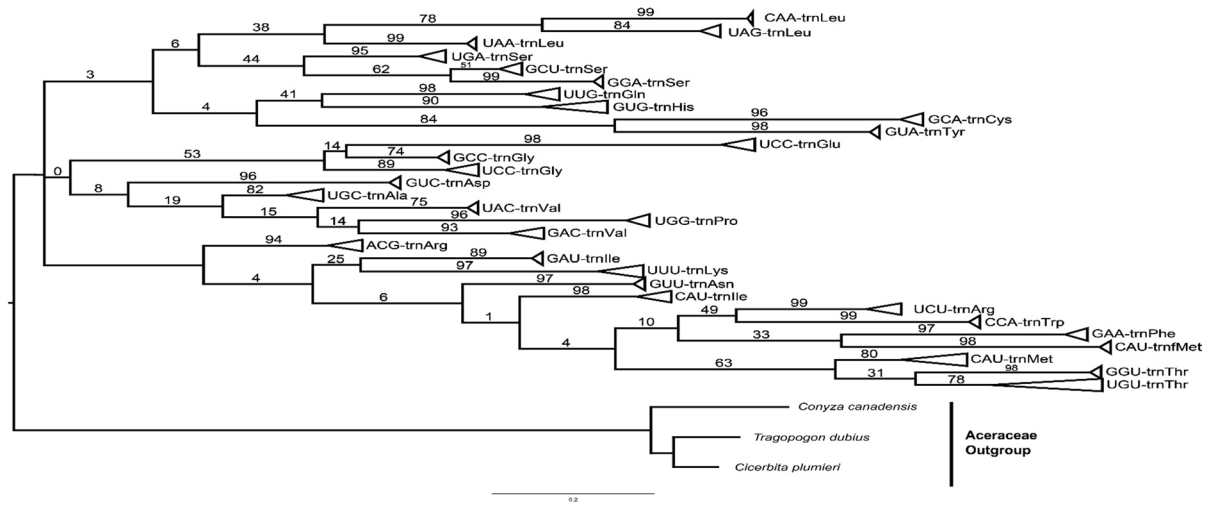

Fig. 4 Phylogenetic diagram of chloroplast tRNAs revealed the duplication and evolutionary process to give rise to other tRNAs from ancestors. MEGA software was used to construct a phylogenetic tree of chloroplast tRNAs of Sapindaceae species

U-C–C of $tRNA^{Gly}$, U-A-A, U-A-G, and C-A-A of $tRNA^{Leu}$ all grouped in the same branch. On the other hand, all these groups were dispersed in all three clades. While C-A-U and G-A-U of $tRNA^{Ile}$, U-C-U, and A-C-G of $tRNA^{Arg}$ were grouped separately in clade III.$tRNA^{Ile}$ (GAU and CAU) is closely grouped with $tRNA^{Lys}$ (UUU) and $tRNA^{Asn}$ (GUU), suggesting the evolution of $tRNA^{Lys}$ (UUU) and $tRNA^{Asn}$ (GUU) from $Trna^{Ile}$ (GAU and CAU). The grouping of $tRNA^{Glu}$ (UCC) with $tRNA^{Gly}$ (UCC and GCC) suggested that $tRNA^{Gly}$ (UCC and GCC) evolved directly from $tRNA^{Glu}$ (UCC) may have evolved from a common ancestor or by a gene duplication event. $tRNA^{His}$ (GUG) and $tRNA^{Gln}$ (UUG) were grouped, suggesting that these tRNAs are closely related and have common ancestors. $tRNA^{Ser}$ (GGA, GCU, UGA) is grouped with $tRNA^{Leu}$ (UAA, UAG, and CAA), suggesting that $tRNA^{Ser}$ evolved from $tRNA^{Leu}$. Notably, $tRNA^{Leu}$ and $tRNA^{Ser}$ were in the same branch of clade I. This indicates that $tRNA^{Leu}$ (UAA, UAG, CAA) underwent a base replacement to give rise to $tRNA^{Ser}$, and that $tRNA^{Ser}$ underwent more replication and diversification (GGA, GCU, UGA). The grouping of $tRNA^{Val}$ (UAC and GAC) with $tRNA^{Pro}$ (UGG) indicates that $tRNA^{Pro}$ (UGG) evolved from $tRNA^{Val}$ (UAC and GAC). The fact that $tRNA^{Met}$ and $tRNA^{Thr}$ (UGU and GGU) are related shows that $tRNA^{Thr}$ (UGU and GGU) developed from $tRNA^{Met}$ (Fig. 4). A disparity index test of homogeneity with the Monte Carlo duplication method was performed to decide whether all the substitutions and the nucleotide substitution rate are homogeneous. The null hypothesis was dismissed for $tRNA^{Arg}$, $tRNA^{Gln}$, $tRNA^{Ala}$, $tRNA^{Met}$, $tRNA^{Thr}$, and $tRNA^{Val}$, indicating that the substitution rate of nucleotides in these classes is homogeneous. Fourteen tRNA isotypes did not exhibit pattern homogeneity outside of these six.

The transition and transversion rate depended on Iso-acceptor

The determined major transition or transversion events examined species are large, indicating that tRNAs are evolutionarily conserved molecules. The transition rate (8.33) and transversion rate (8.34) of $tRNA^{Try}$ are nearly identical, meaning that the transversion rate is marginally higher than the transition rate. In terms of transition and transversion, these tRNAs have developed approximately simultaneously (Table 3). Additionally, $tRNA^{Asn}$, $tRNA^{Asp}$, $tRNA^{Glu}$, $tRNA^{Lys}$, $tRNA^{Phe}$, and $tRNA^{Tyr}$ had identical transition rates (25.0) and transversion rates (0.0). Notably, however, $tRNA^{Asn}$, $tRNA^{Asp}$, $tRNA^{Glu}$, $tRNA^{Lys}$, $tRNA^{Phe}$, and $tRNA^{Tyr}$ in the chloroplast genome of examined Sapindaceae plants have undergone the highest rate of transition but have not experienced any transversion. Another group was also found with a high percentage of transition (12.00), but a

Table 3 Transition and transversion rate of the chloroplast tRNAs. A bold letter indicates transition

	A	T	C	G		A	T	C	G
Alanine					Lysine				
A	–	12.50	12.50	0.00	A	–	0.00	0.00	25.00
T	12.50	–	0.00	12.50	T	0.00	–	25.00	0.00
C	12.50	0.00	–	12.50	C	0.00	25.00	-	0.00
G	0.00	12.50	12.50	–	G	25.00	0.00	0.00	–
Arginine					Methionine				
A	–	12.50	12.50	0.00	A	–	4.94	4.94	15.13
T	12.50	–	0.00	12.50	T	4.94	–	15.13	4.94
C	12.50	0.00	–	12.50	C	4.94	15.13	–	4.94
G	0.00	12.50	12.50	–	G	15.13	4.94	4.94	–
Asparagine					Phenylalanine				
A	–	0.00	0.00	25.00	A	–	0.00	0.00	25.00
T	0.00	–	25.00	0.00	T	0.00	–	25.00	0.00
C	0.00	25.00	–	0.00	C	0.00	25.00	–	0.00
G	25.00	0.00	0.00	–	G	25.00	0.00	0.00	–
Aspartate					Proline				
A	–	0.00	0.00	25.00	A	–	6.22	6.22	12.57
T	0.00	–	25.00	0.00	T	6.22	–	12.57	6.22
C	0.00	25.00	–	0.00	C	6.22	12.57	–	6.22
G	25.00	0.00	0.00	–	G	12.57	6.22	6.22	–
Cystine					Serine				
A	–	6.05	6.05	12.89	A	–	4.29	4.29	16.41
T	6.05	–	12.89	6.05	T	4.29	–	16.41	4.29
C	6.05	12.89	–	6.05	C	4.29	16.41	–	4.29
G	12.89	6.05	6.05	–	G	16.41	4.29	4.29	–
Glutamine					Threonine				
A	–	6.78	6.78	11.44	A	–	6.16	6.16	12.69
T	6.78	–	11.44	6.78	T	6.16	–	12.69	6.16
C	6.78	11.44	–	6.78	C	6.16	12.69	–	6.16
G	11.44	6.78	6.78	–	G	12.69	6.16	6.16	–
Glutamate					Tryptophan				
A	–	0.00	0.00	25.00	A	–	8.34	8.34	8.33
T	0.00	–	25.00	0.00	T	8.34	–	8.33	8.34
C	0.00	25.00	–	0.00	C	8.34	8.33	–	8.34
G	25.00	0.00	0.00	–	G	8.33	8.34	8.34	–
Glycine					Tyrosine				
A	–	3.99	3.99	17.01	A	–	0.00	0.00	25.00
T	3.99	–	17.01	3.99	T	0.00	–	25.00	0.00
C	3.99	17.01	–	3.99	C	0.00	25.00	–	0.00
G	17.01	3.99	3.99	–	G	25.00	0.00	0.00	–
Histidine					Valine				
A	–	6.16	6.16	12.68	A	–	4.55	4.55	15.90
T	6.16	–	12.68	6.16	T	4.55	–	15.90	4.55
C	6.16	12.68	–	6.16	C	4.55	15.90	–	4.55
G	12.68	6.16	6.16	–	G	15.90	4.55	4.55	–
Isoleucine					Overall				
A	–	6.58	6.58	11.84	A	–	4.84	4.84	15.31

Table 3 (continued)

	A	T	C	G		A	T	C	G
T	6.58	–	11.84	6.58	T	4.84	–	15.31	4.84
C	6.58	11.84	–	6.58	C	4.84	15.31	–	4.84
G	11.84	6.58	6.58	–	G	15.31	4.84	4.84	–
Leucine									
A	–	6.12	6.12	12.76					
T	6.12	–	12.76	6.12					
C	6.12	12.76	–	6.12					
G	12.76	6.12	6.12	–					

transversion rate is zero, i.e., $tRNA^{Ala}$ and $tRNA^{Arg}$. In contrast, the rate of transversion in $tRNA^{Cys}$ (12.89), $tRNA^{Gln}$ (11.44), $tRNA^{His}$ (12.68), $tRNA^{Ile}$ (11.84), $tRNA^{Leu}$ (12.76), $tRNA^{Pro}$ (12.57), and $tRNA^{Thr}$ (12.69) found to be almost equal but higher relative to the rate of transition for $tRNA^{Cys}$ (6.05), $tRNA^{Gln}$ (6.78), $tRNA^{His}$ (6.16), $tRNA^{Ile}$ (6.58), $tRNA^{Leu}$ (6.12), $tRNA^{Pro}$ (6.22), and $tRNA^{Thr}$ (6.16), respectively. A higher transition rate was observed in $tRNA^{Gly}$ (17.01), $tRNA^{Met}$ (15.13), $tRNA^{Ser}$ (16.41), and $tRNA^{Val}$ (15.90) with a low rate of transversions such as $tRNA^{Gly}$ (3.99), $tRNA^{Met}$ (4.94), $tRNA^{Ser}$ (4.29), and $tRNA^{Val}$ (4.55), respectively. However, the average rate of transition (15.31) is more than the rate of transversion (4.84) in the collective examination of the tRNAs (Table 3).

Chloroplast tRNA duplication and deletion

Plant genomes have more duplicated genes than animal genomes, and whole-genome replication events have occurred several times over the last 200 million years (Lyons et al. 2008; Soltis et al. 2009; Lee et al. 2013; Renny-Byfield and Wendel 2014). Given the chloroplast genome's cyanobacterial roots, the rate of duplication and loss events can vary from genes in the nuclear-encoded genome. Duplication/loss research of chloroplast tRNA showed that 1,015 genes experienced duplication events, 1,991 genes underwent losses, and 234 genes underwent conditional duplication in the 36 Sapindaceae species studied (Supplementary Table S7). During evolution, the majority of chloroplast tRNA genes were lost. Even though all the tRNAs originated from the same family (Sapindaceae), the genes lost more than duplicated ones (Fig. 5).

Discussion

The conservation of nucleotide sequences is a significant phenomenon that indicates a conserved functional role. As a result, it is essential to consider the conserved nucleotide consensus sequences in plant chloroplast tRNAs. tRNAs are a family of conserved genes that carry out protein translation (Kanai 2015). The chloroplast genome's existence leads to the genome's semi-autonomous structure. Multiple sequence analyses of chloroplast tRNAs revealed a variety of conserved genetic characteristics. Each tRNA sequence analysis in the Sapindaceae chloroplast genome showed that $tRNA^{Leu}$, $tRNA^{Ser}$, and $tRNA^{Tyr}$ encode for the most extended tRNA sequences, ranging from 84 to 92 base pairs. Our results indicated that $tRNA^{Tyr}$ contains the most extended nucleotide sequence in Sapindaceae plants. A previous study found that plants' chloroplast genome includes 80 or more nucleotides in $tRNA^{Leu}$ and $tRNA^{Ser}$ (Mohanta and Bae 2017). Results also revealed the absence of $tRNA^{Phe}$, $tRNA^{His}$, $tRNA^{Trp}$, and $tRNA^{Try}$ in the chloroplast genome of these Sapindaceae plants. A previous study shows that tRNA sequences partially overlap with flanking genes. Some tRNA pairs seem templated by sense-antisense strands (Barthélémy and Seligmann 2016). While some anticodons of $tRNA^{Gly}$, $tRNA^{Val}$, $tRNA^{Thr}$, and $tRNA^{Met}$ were also missing in some monocot plants (Mohanta et al. 2019). This is the first report regarding the complete absence of $tRNA^{Phe}$, $tRNA^{His}$, and $tRNA^{Trp}$ in some species. The partial lack of some anticodons of $tRNA^{Ala}$, $tRNA^{Gly}$, $tRNA^{Val}$, $tRNA^{Thr}$, $tRNA^{Leu}$, and $tRNA^{Met}$ the chloroplast genome. Direct evidence for these exceptional tRNAs, predicted by purely computational means, has been lacking so far. However, some research demonstrate that several of

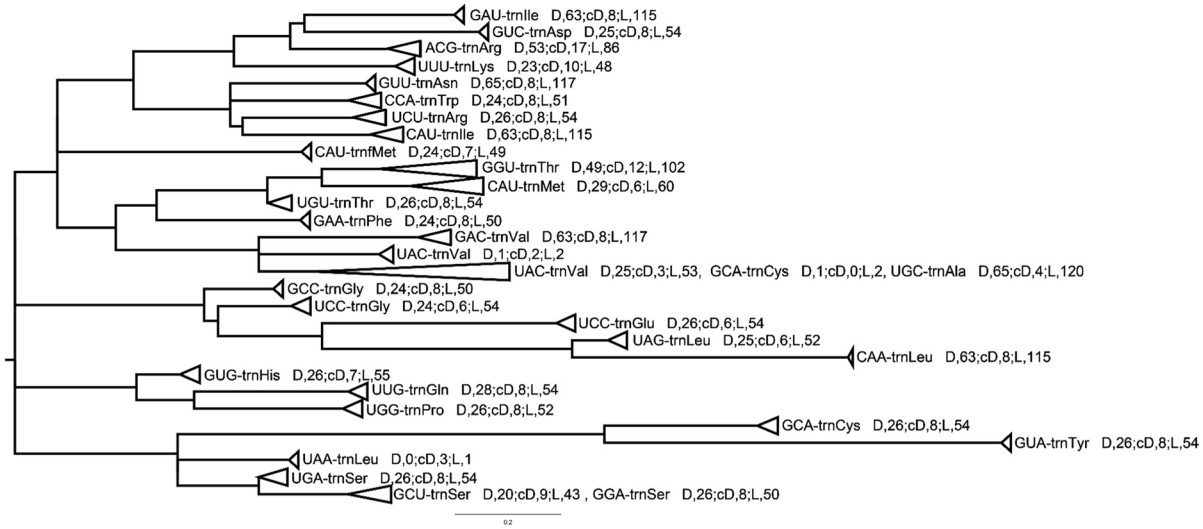

Fig. 5 The Notung software was used to examine the duplication and loss of chloroplast tRNAs. The findings suggest that chloroplast tRNAs experienced extensive gene loss during development, followed by diversification. Duplication, complete duplication, and loss events are denoted by D, cD, and L, respectively

these miniaturized armless tRNAs consisting only of acceptor- and anticodon-arms are indeed transcribed and correctly processed by non-encoded CCA based on mitochondrial sequence (Jühling et al. 2012, 2018; Wende et al. 2014; Pons et al. 2019; Hennig et al. 2020). Besides, the $tRNA^{Phe}$, $tRNA^{His}$, and $tRNA^{Trp}$ in Sapindaceae plants' chloroplast genomes also lack selenocysteine, pyrrolysine, and suppressor tRNA in all species studied (Table S7). The lack of any important tRNA genes encoding in the chloroplast genome is interesting, and understanding how protein translation is carried out in Sapindaceae plants without them is critical. That is probably why genomic tRNA compensates for the lack of plastid tRNAs, or that other organellar genome tRNAs play numerous protein translation functions.

The Sapindaceae chloroplast genome contains the most significant number of genes, "4" encoded by $tRNA^{Leu}$ and $tRNA^{Ile}$, followed by $tRNA^{Thr}$, $tRNA^{Val}$, $tRNA^{Arg}$, and $tRNA^{Ser}$ has "3". There are 64 codons in the universal genetic table. Of these 61 codons are sense codons, three of which are stop codons. Consequently, tRNAs with 61 distinct anticodons can exist to code for 61 sense codons. We also found that the Sapindaceae chloroplast genome's tRNAs are missing about 32 anticodons. For example, the absence of UCC anticodons in *A. sino-oblongum* species of $tRNA^{Gly}$ is balanced by UGG anticodons of $tRNA^{Pro}$. In species *Aesculus wangii*, however, the absence of anticodon UAC of $tRNA^{Val}$ is balanced by the presence of GAC anticodon. Some species' anticodon GGU of tRNAThr is absent in some species (*Dimocarpus logan*, *Dodonaea viscosa*, *Koelreuteria paniculate*, and *Litchi chinensis*). Surprisingly, our results showed that the anticodon UAA in $tRNA^{Leu}$ is present in Sapindaceae plant species, which was compensated by the absence of GGU of $tRNA^{Thr}$. Similarly, another anticodon GUG of $tRNA^{His}$ was absent in *Xanthoceras sorbifolium* species, and results showed that it was replaced with anticodon UAA in $tRNA^{Leu}$. It is concluded that the presence of replaced anticodon UAA of $tRNA^{Leu}$ in the species, as mentioned earlier, plays an essential role in the missing tRNA's characters. The complete absence of a $tRNA^{Trp}$ (CCA) in *Dodonaea viscosa*, and $tRNA^{Phe}$ (AAA and GAA) in *A. ceriferum,* and *A. coriaceifolium* is challenging to understand. In contrast, these genes were compensated with other tRNA anticodons.

However, genomic tRNAs and perhaps other chloroplast or nuclear tRNAs could compensate for the deficit caused by the absence of these tRNAs in the chloroplast genome. Our analysis indicated that $tRNA^{Met}$ and $tRNA^{fMet}$ encoded the anticodon CAU in the Sapindaceae plant genome. The involvement of $tRNA^{fMet}$ in the chloroplast genome has

been documented in previous studies (Howe 1985; Hiratsuka et al. 1989; Mohanta et al. 2019). The prokaryotic genome's $tRNA^{fMet}$ mediates the start of protein translation, while $tRNA^{Met}$ mediates the eventual attachment of methionine to the polypeptide chain (Kozak 1999). The existence of $tRNA^{Met}$ and $tRNA^{fMet}$ in prokaryotic and organellar genes is a common feature (Salinas-Giegé et al. 2015). According to our findings, including $tRNA^{fMet}$ in Sapindaceae plants' chloroplast genome suggests a prokaryotic origin.

The nucleotide composition of the tRNA is maintained by the structure of the translated tRNA. As a result, conserved coding sequences should represent the conserved functions of tRNA. In tRNAs with only the -loop, a previous study discovered a conserved nucleotide consensus sequence (Mohanta and Bae 2017). The existence of the U-U-C-x-A–x-U nucleotide consensus sequence in the -loop was found in our research. On the other hand, other tRNAs had no conserved consensus sequences but did have several conserved nucleotides. A U nucleotide is present in the first position of nuclear-encoded $tRNA^{Gln}$ and $tRNA^{Asn}$ (Table 2). (Mohanta and Bae 2017). However, a multiple sequence alignment analysis showed that chloroplast tRNAs' sequence conservation is family-specific (Table 2). The D-stem is supposed to have a rigid structure due to its high G/C content, but it appears to be relatively weak due to more A-U and G-U base pairing (Hardt et al. 1993; Mohanta and Bae 2017). According to a previous study, D-arm is vital in identifying aminoacyl-tRNA synthetase and has rare confirmation due to G residue (Smith and Yarus 1989; Hardt et al. 1993). For example, our results suggested that the D-arm's first position was conserved with G nucleotide except for some tRNAs.

The fourth position of the D-arm, on the other side, was filled by a conserved C nucleotide. The nucleotide of the D-first loop was found to be conserved with A nucleotide in most tRNA. The anticodon loop preserved the U nucleotide at its 2^{nd} position and A-A nucleotides at its 6th and 7th positions. These results suggested the unique conservation of nucleotide sequences in the chloroplast genome of Sapindaceae plants. Chloroplast tRNA genes are transcribed by the bacterial-like RNA polymerase of the organelle during protein translation. Consensus sequences are found in these two bins. The signal sequence for transcription activation in the chloroplast genome's tRNAsonly sometimesally conserved.

At the 32nd position, the anticodon loop was conserved (Sharp et al. 1985). Nucleotide conservation was found only at the 33rd position in the current study, while the anticodon loop at the 38th and 39th positions had a conserved A-A nucleotide in most cases. Furthermore, a 3'-CCA tail was discovered in several tRNA sequences. Inserting a CCA tail to the 3′-end of a tRNA is more accessible by tRNA-nucleotidyl transferases. Chloroplast genomes, on the other hand, do not contain tRNA-nucleotidyl transferases. In the absence of nucleotidyl transferases, adding a CCA tail to the 3′-end of tRNA would be challenging. The few tRNAs' that lacks a CCA tail at the 3′ ends indicate their recent evolution, as most nuclear tRNAs did not have a 3' CCA tail.

Robert Holley (Holley et al. 1965) proposed that tRNAs have a cloverleaf-like structure, but specific tRNAs have separate secondary structures (Mohanta and Bae 2017). tRNAs have several arms and loops that assist with protein translation. The nucleotide composition of each arm and loop is different. The acceptor's arm has seven base pairs (7 bp), the D-arm 3–4 bp, and the D-loop 4–12 nucleotides long. The anticodon arm five bp, the anticodon loop seven nucleotides, the variable arm 4–23 nucleotides, the TΨC -arm five bp, and the TΨC-loop seven nucleotides, according to a previous report (Kirchner and Ignatova 2015). This suggests that the nucleotide composition of the D-loop has a significant variance. The A1 nucleotide, on the other hand, was conserved in most instances, while nucleotide conservation in the different sections of the D-loop was unique to each tRNA family. Long-range interactions connect the D-loop to the 9-loop (Hanawa-Suetsugu et al. 2001). According to the previous research and the current analysis, there are plenty of differences in chloroplast tRNAs' arms and loops (Mohanta et al. 2019).

The acceptor's arm contains specific information for tRNA-nucleotidyltransferases (Mohanta et al. 2019). However, an acceptor's arm is absent in chloroplast $tRNA^{Met}$ (CAT) in all Sapindaceae species except *Acer catalpifolium*, *Acer laevigatum*, *Acer wilsonii*, *Aesculus wangii*, *Dimocarpus logan*, and *Litchi chinensis*. Similarly, $tRNA^{Thr}$ (GGT) has a missing acceptor arm in *Acer sino-oblongum*, *Acer wilsonii*, *Sapindus mukorossi*, *Xanthoceras*

sorbifolium. Surprisingly, $tRNA^{Leu}$ (TAA) was only found in *Dimocarpus logan, Koelreuteria paniculate*, *Dodonaea viscosa*, *Sapindus mukorossi*, and *Xanthoceras sorbifolium* species. We suggested that $tRNA^{Leu}$ (TAA) characterize some extra functions in these species. Some tRNAs contain novel loop structures identified in the present study. The earlier study on the chloroplast genome also raised whether these loops mimicked the tRNA anticodon loop and played a critical role in the chloroplast protein translation machinery (Mohanta et al. 2019). Some tRNAs were also discovered to have nine nucleotides in the anticodon ring, a novel tRNA phenomenon. The anticodon loop's practical effect of having nine nucleotides has yet to be determined. Previously it was noted that occasional anti-sense tRNAs with predicted expanded anticodons (depending on taxon) suggest complex tetra-decoding mechanisms. Transcripts of anti-sense tRNAs with unusual anticodons are more abundant than those of homologs with regular anticodons (Seligmann 2012, 2013, 2014; Seligmann and Labra 2013; Barthélémy and Seligmann 2016; Seligmann and Warthi 2019). Furthermore, some tRNA contained putative novel tRNA systems, as seen in Figs. 1, 2, and 3. We hypothesized that these tRNAs contain extra characters for chloroplast genome translation and are responsible for evolution. A C-A-U anticodon has been discovered in specific tRNAs, which codes for $tRNA^{Ile}$. However, previously, a C-A-U anticodon in $tRNA^{Ile}$ has been found in the Bacillus subtilis and monocot chloroplast genomes (Köhrer et al. 2014; Mohanta et al. 2019).

Phylogenetic analysis of chloroplast tRNAs showed three main groups in Sapindaceae plants. There is no mix-up of the group I, II, and III tRNA members. However, anticodon GAC, UAC, UGC, GAA, UGG, GCA, GUA, UGA, UAA, CAU, and ACG are found in separate branches phylogenetic tree, implying that they evolved from several shared ancestors. A previous study of chloroplast tRNA also suggested that anticodon GAC, UAC, UGA, UAA, CAU, and ACG were found independently and arose from multiple common ancestors (Mohanta et al. 2019). The overlapping tRNA family members suggest that the tRNAs with these anticodon groups may derive from numerous shared ancestors or have been caused by duplication. Our results indicated that $tRNA^{Glu}$ (UCC) presented in the same lineage with $tRNA^{Gly}$ (UCC) with the same anticodons in group II as proof of evolution in the chloroplast genome. We suggested that $tRNA^{Glu}$ (UCC) evolved directly from $tRNA^{Gly}$ (UCC). Similarly, tRNAThr (GGU and UGU) grouping with $tRNA^{Met}$ (CAU) in group III clustered in the same branch strongly suggests that $tRNA^{Thr}$ evolved directly from $tRNA^{Met}$. Previously it was also reported that $tRNA^{Met}$ was a significant driver in the evolution of tRNAs in the chloroplast genome (Mohanta et al. 2019). Phylogenetic analysis further revealed that $tRNA^{Ser}$ (UGA, GGA, GCU) and $tRNA^{Leu}$ (UAA, UAG, CAA) presented in group I same clade, suggesting an evolutionary relationship. We indicated that $tRNA^{Ser}$ evolved from $tRNA^{Leu}$, the most primitive form of tRNAs and the basal evolutionary ancestor (Mohanta et al. 2019). Similarly, $tRNA^{Pro}$ (UGG) and $tRNA^{Val}$ (UAC and GAC) presented in the same lineage in group II have a close relationship. We suggested that $tRNA^{Pro}$ evolved directly from $tRNA^{Val}$ by duplication event(s). The appearance of $tRNA^{Val}$ twice in group II in the same tRNA cluster meant that $tRNA^{Val}$ was one of the tRNA families that had undergone significant replication events to produce other tRNAs.

A gene's evolution happens because of random mutations and genetic drift. The pattern and frequency of nucleotide substitution are primarily determined by the gene's mutational events (Zhang and Gerstein 2003; Arnheim and Calabrese 2009). The transition/transversion bias remains strong at low genetic divergence levels, while the transition/transversion bias remains low at high genetic divergence levels (Yang and Yoder 1999). When paired with selection/genetic drift or novel adaptation, Silent mutation may substantially affect genetic variation (Palumbi 1994). According to our findings, chloroplast tRNAs' transitions were more significant than the transversion rate. $tRNA^{Asn}$ (25.00), $tRNA^{Tyr}$ (25.00) $tRNA^{Met}$ (15.13), and $tRNA^{Ser}$ (16.41) belong to a polar R group with a high transition rate, and the rate of a transversion is almost zero in tRNAs that carry polar amino acids. While some tRNA contained high transition rate and transversion rate is zero from a charged group, e.g., $tRNA^{Arg}$ $tRNA^{Asp}$, $tRNA^{Glu}$, and $tRNA^{Lys}$ that act as side chains often form salt bridges. Polar amino acids dissolve easily in water and form strong hydrogen bonding with other molecules. Transition substitutions were strongly favored over transversion substitutions during the evolution of chloroplast $tRNA^{Asn}$, $tRNA^{Tyr}$, $tRNA^{Met}$, and

$tRNA^{Ser}$. The transition has occurred more often in some tRNA Iso-acceptors than transversion (Mohanta et al. 2019). However, a few tRNAs had a higher rate of transversion than transition, suggesting that the rate of tRNA development, transition, and transversion is Isoacceptor-specific and that tRNAs have not evolved at the same rate. (Mohanta et al. 2019).

It is crucial to understand how new genes evolve and act in genomics and evolutionary biology. Gene duplication is an important mechanism for developing new gene functions by sub-functionalization and neo-functionalization, while gene deletion significantly affects the gene family. (Rasmussen and Kellis 2012; Magadum et al. 2013; Teufel et al. 2016). Gene replication is a crucial step in evolving new organisms, particularly in eukaryotes, where it is critical to developing novel gene functions (Zhao et al. 2015). Similarly, segmental deletion results in the absence of a gene, and pseudogenization retains a minimal number of functional gene copies (Cotton and Page 2005; Blomme et al. 2006; Demuth et al. 2006). Gene replication events, such as genome duplication, retrotransposons, and unequal mixing, have developed most novel gene functions (Ohta 2010). Ancient replication activities and the preservation of extant pairs of duplicated genes have greatly aided the evolution and adaptive diversification of gene families (Panchy et al. 2016). Plant genomes change quickly, resulting in greater genetic diversity than other animals (Kejnovsky et al. 2009). Previous research has shown that tRNAs are a textbook example of gene families duplicated and lost over time (Bermudez-Santana et al. 2010; Rogers et al. 2010). Previous experiments on chloroplast tRNAs found that the incidence of deletion is far higher than the rate of tRNA replication (Mohanta et al. 2019). This suggests the cyanobacterial chloroplast genome's maternal inheritance is more stable than the nuclear-encoded plant genomes. Specific genes were already missing within each species, and these species were still part of the maternal lineage. This adds to the increasing evidence that cyanobacterial tRNAs have come from typical polyphenylene ancestors with higher replication losses. Most of the tRNAs examined are lost in each of the 36 Sapindaceae plants. (Fig. 5, Table S5). Our research gives a general overview of the evolutionary history of tRNA in the Sapindaceae family.

We suggested a new method for sequencing and reading tRNA, so they can better understand individual genetic variation. There is a need for more investigation into the acceptor's arms in detail. More research should be needed on how tRNAs contribute to disease or if the process can be reversed. In this research, we did not investigate protein mutations and focused on tRNA instead. More research should be needed to contribute to how they influence the development of genetic diseases.

Conclusions

Sapindaceae chloroplast tRNAs have lengths ranging from 64 nucleotides ($tRNA^{Met}$ CAT) to 93 nucleotides ($tRNA^{Ser}$ TGA), and they encode 27–29 anticodon types and 36–37 unique tRNAs. The Sapindaceae family's cp genome lacks 36 anticodon types. The CAU anticodon in $tRNA^{Met}$, and the $tRNA^{Ile}$ of chloroplast tRNA. The acceptor arm of Sapindaceae chloroplast tRNA is 1–7 nt long, whereas the D-arm is 2–4 nt long, the D-loop is 7–11 nt long, the anticodon loop is 7–9 nt long, the -arms are 4–5 nt long, and the -loop is 3–16 nt long. The majority of tRNAs shared the G-G nucleotide at positions 52 and 53. It was discovered that the nucleotide sequence U-U-C-x-A-x-U is conserved at positions 54, 55, 56, 58, and 60. Furthermore, phylogenetic analysis suggests that tRNAs possibly have several inferred ancestors, including $tRNA^{Leu}$ and $tRNA^{Ser}$, $tRNA^{Ile}$ and $tRNA^{Gly}$, etc., in evolutionary history. However, compared to their transversion rate, the average transition rate of all the implicated cp tRNAs was more significant. In Sapindaceae chloroplast tRNAs, gene duplication events (1,015 genes underwent duplication events, and 234 genes underwent conditional duplication) have been seen more frequently than gene loss events (1,991 genes underwent losses). These findings shed light on the tRNA family's specific traits and evolutionary diversity.

Acknowledgements The authors thank researchers who helped and supported the data collection and bioinformatics analysis.

Authors contribution It is stated that each Author contributes substantially to the article. KS was a significant contributor to writing and organizing the research plan in the manuscript. HK, AA, UZ, FC, JD and SA guided the interpretation and discussion. SF, EAM, IMM, HOE provided technical expertise, editing, and funding acquisition. JD and KS helped in writing the manuscript. All authors have seen and approved

the final manuscript and have taken due care to ensure the work's integrity.

Funding Researchers Supporting Project number (RSPD2023R741), King Saud University. This work was supported by the National Key R & D Program of China (No. 2018YFA0606102), the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA19020303, No. XDA26010202), and the Natural Science Foundation of China (No. 41771056).

Declarations

Conflict of interest There is no conflict of interest.

References

Arnheim N, Calabrese P (2009) Understanding what determines the frequency and pattern of human germline mutations. Nat Rev Genet 10

Barthélémy RM, Seligmann H (2016) Cryptic tRNAs in chaetognath mitochondrial genomes. Comput Biol Chem 62. https://doi.org/10.1016/j.compbiolchem.2016.04.007

Bendich AJ (2004) Circular chloroplast chromosomes: The grand illusion. Plant Cell 16

Bermudez-Santana C, Attolini CS, Kirsten T, et al (2010) Genomic organization of eukaryotic tRNAs. BMC Genomics 11. https://doi.org/10.1186/1471-2164-11-270

Bernhart SH, Hofacker IL, Will S, et al (2008) RNAalifold: Improved consensus structure prediction for RNA alignments. BMC Bioinformatics 9. https://doi.org/10.1186/1471-2105-9-474

Blomme T, Vandepoele K, De Bodt S, et al (2006) The gain and loss of genes during 600 million years of vertebrate evolution. Genome Biol 7. https://doi.org/10.1186/gb-2006-7-5-r43

Chen K, Durand D, Farach-Colton M (2000) NOTUNG: A Program for Dating Gene Duplications and Optimizing Gene Family Trees. Mary Ann Liebert, Inc.

Cotton JA, Page RDM (2005) Rates and patterns of gene duplication and loss in the human genome. Proceedings of the Royal Society B: Biological Sciences 272. https://doi.org/10.1098/rspb.2004.2969

De Jong P (1976) Flowering and Sex Expression in Acer L., A Biosystematic Study. Mededelingen Landbouwhogeschool Wageningen 76

Demongeot J, Seligmann H (2020) RNA Rings Strengthen Hairpin Accretion Hypotheses for tRNA Evolution: A Reply to Commentaries by Z.F. Burton and M. Di Giulio. J Mol Evol 88

Demuth JP, Bie T De, Stajich JE, et al (2006) The evolution of mammalian gene families. PLoS One 1. https://doi.org/10.1371/journal.pone.0000085

Dieci G, Sentenac A (2003) Detours and shortcuts to transcription reinitiation. Trends Biochem Sci 28

Dong P Bin, Wang RN, Afzal N, et al (2021) Phylogenetic relationships and molecular evolution of woody forest tree family Aceraceae based on plastid phylogenomics and nuclear gene variations. Genomics 113. https://doi.org/10.1016/j.ygeno.2021.03.037

Ehrlich R, Davyt M, López I, et al (2021) On the Track of the Missing tRNA Genes: A Source of Non-Canonical Functions? Front Mol Biosci 8. https://doi.org/10.3389/FMOLB.2021.643701/FULL

Gleiser G, Verdú M (2005) Repeated evolution of dioecy from androdioecy in Acer. New Phytologist 165. https://doi.org/10.1111/j.1469-8137.2004.01242.x

Hanawa-Suetsugu K, Bordeau V, Himeno H, et al (2001) Importance of the conserved nucleotides around the tRNA-like structure of Escherichia coli transfer-messenger RNA for protein tagging. Nucleic Acids Res 29. https://doi.org/10.1093/nar/29.22.4663

Hardt WD, Schlegl J, Erdmann VA, Hartmann RK (1993) Role of the D arm and the anticodon arm in tRNA recognition by Eubacterial and Eukaryotic RNase P enzymes. Biochemistry 32. https://doi.org/10.1021/bi00211a014

Harris AJ, Frawley E, Wen J (2017) The utility of single-copy nuclear genes for phylogenetic resolution of acer and dipteronia (Acereae, Sapindaceae). Ann Bot Fenn 54. https://doi.org/10.5735/085.054.0603

Hennig O, Philipp S, Bonin S, et al (2020) Adaptation of the romanomermis culicivorax cca-adding enzyme to miniaturized armless trna substrates. Int J Mol Sci 21. https://doi.org/10.3390/ijms21239047

Hereward JP, Werth JA, Thornby DF, et al (2018) Complete chloroplast genome of glyphosate resistant Sonchus oleraceus L. from Australia, with notes on the small single copy (SSC) region orientation. Mitochondrial DNA B Resour 3. https://doi.org/10.1080/23802359.2018.1450682

Hiratsuka J, Shimada H, Whittier R, et al (1989) The complete sequence of the rice (Oryza sativa) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. MGG Mol General Genet 217. https://doi.org/10.1007/BF02464880

Holley RW, Apgar J, Everett GA et al (1965) Structure of a ribonucleic acid. Science (1979) 147:. https://doi.org/10.1126/science.147.3664.1462

Howe CJ (1985) The endpoints of an inversion in wheat chloroplast DNA are associated with short repeated sequences containing homology to att-lambda. Curr Genet 10. https://doi.org/10.1007/BF00636479

Jühling F, Pütz J, Florentz C, Stadler PF (2012) Armless mitochondrial tRNAs in enoplea (nematoda). RNA Biol 9. https://doi.org/10.4161/rna.21630

Jühling T, Duchardt-Ferner E, Bonin S, et al (2018) Small but large enough: Structural properties of armless mitochondrial tRNAs from the nematode Romanomermis culicivorax. Nucleic Acids Res 46. https://doi.org/10.1093/nar/gky593

Kanai A (2015) Disrupted tRNA genes and tRNA fragments: A perspective on tRNA gene evolution. Life 5

Kearse M, Moir R, Wilson A et al (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kejnovsky E, Leitch IJ, Leitch AR (2009) Contrasting evolutionary dynamics between angiosperm and mammalian genomes. Trends Ecol Evol 24

Kirchner S, Ignatova Z (2015) Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. Nat Rev Genet 16

Knorr W, Heimann M (2001) Uncertainlies in global terrestrial biosphere modeling 1. A comprehensive sensitivity analysis with a new photosynthesis and energy balance scheme. Global Biogeochem Cycles 15. https://doi.org/10.1029/1998GB001059

Köhrer C, Mandal D, Gaston KW, et al (2014) Life without tRNAIle-lysidine synthetase: Translation of the isoleucine codon AUA in Bacillus subtilis lacking the canonical tRNA 2Ile. Nucleic Acids Res 42. https://doi.org/10.1093/nar/gkt1009

Kozak M (1999) Initiation of translation in prokaryotes and eukaryotes. Gene 234

Kumar S, Stecher G, Li M et al (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096

Lee TH, Tang H, Wang X, Paterson AH (2013) PGDD: A database of gene and genome duplication in plants. Nucleic Acids Res 41. https://doi.org/10.1093/nar/gks1104

Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44. https://doi.org/10.1093/nar/gkw413

Lyons E, Pedersen B, Kane J, et al (2008) Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. Plant Physiol 148. https://doi.org/10.1104/pp.108.124867

Magadum S, Banerjee U, Murugan P, et al (2013) Gene duplication as a major force in evolution. J Genet 92. https://doi.org/10.1007/s12041-013-0212-8

Mallick B, Chakrabarti J, Sahoo S et al (2005) Identity elements of Archaeal tRNA. DNA Research 12. https://doi.org/10.1093/dnares/dsi008

Maréchal-Drouard L, Guillemaut P, Pfitzingzer H, Weil JH (1991) Chloroplast tRNAs and tRNA genes: structure and function. In: The translational apparatus of photosynthetic organelles

Michaud M, Cognat V, Duchêne AM, Maréchal-Drouard L (2011) A global picture of tRNA genes in plant genomes. Plant J 66. https://doi.org/10.1111/j.1365-313X.2011.04490.x

Mohanta TK, Bae H (2017) Analyses of genomic trna reveal presence of novel tRNAs in oryza sativa. Front Genet 8. https://doi.org/10.3389/fgene.2017.00090

Mohanta TK, Khan AL, Hashem A, et al (2019) Genomic and evolutionary aspects of chloroplast tRNA in monocot plants. BMC Plant Biol 19. https://doi.org/10.1186/s12870-018-1625-6

Nguyen MLT (2006) Flora of China. Vol. 14. Apiaceae through Ericaceae. Econ Bot 60. https://doi.org/10.1663/0013-0001(2006)60[95:focvat]2.0.co;2

Ohta T (2010) Gene conversion and evolution of gene families: an overview. Genes (Basel) 1

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. Annu Rev Ecol Syst 25

Panchy N, Lehti-Shiu M, Shiu SH (2016) Evolution of gene duplication in plants. Plant Physiol 171. https://doi.org/10.1104/pp.16.00523

Phizicky EM, Hopper AK (2010) tRNA biology charges to the front. Genes Dev 24

Pons J, Bover P, Bidegaray-Batista L, Arnedo MA (2019) Armless mitochondrial tRNAs conserved for over 30 millions of years in spiders. BMC Genomics 20. https://doi.org/10.1186/s12864-019-6026-1

Rasmussen MD, Kellis M (2012) Unified modeling of gene duplication, loss, and coalescence using a locus tree. Genome Res 22. https://doi.org/10.1101/gr.123901.111

Renny-Byfield S, Wendel JF (2014) Doubling down on genomes: Polyploidy and crop plants. Am J Bot 101. https://doi.org/10.3732/ajb.1400119

Rogers HH, Bergman CM, Griffiths-Jones S (2010) The evolution of tRNA genes in Drosophila. Genome Biol Evol 2. https://doi.org/10.1093/gbe/evq034

Saeki I, Murakami N (2009) Chloroplast DNA phylogeography of the endangered Japanese red maple (Acer pycnanthum): The spatial configuration of wetlands shapes genetic diversity. Divers Distrib 15. https://doi.org/10.1111/j.1472-4642.2009.00609.x

Salinas-Giegé T, Giegé R, Giegé P (2015) TRNA biology in mitochondria. Int J Mol Sci 16

Saski C, Lee SB, Fjellheim S, et al (2007) Complete chloroplast genome sequences of Hordeum vulgare, Sorghum bicolor and Agrostis stolonifera, and comparative analyses with other grass genomes. Theor Appl Genet 115. https://doi.org/10.1007/s00122-007-0567-4

Schimmel P (2018) RNA Processing and Modifications: The emerging complexity of the tRNA world: Mammalian tRNAs beyond protein synthesis. Nat Rev Mol Cell Biol 19

Seligmann H (2013) Pocketknife tRNA hypothesis: Anticodons in mammal mitochondrial tRNA side-arm loops translate proteins? BioSystems 113. https://doi.org/10.1016/j.biosystems.2013.07.004

Seligmann H (2012) Overlapping genetic codes for overlapping frameshifted genes in Testudines, and Lepidochelys olivacea as special case. Comput Biol Chem 41. https://doi.org/10.1016/j.compbiolchem.2012.08.002

Seligmann H (2014) Putative anticodons in mitochondrial tRNA sidearm loops: Pocketknife tRNAs? J Theor Biol 340. https://doi.org/10.1016/j.jtbi.2013.08.030

Seligmann H, Labra A (2013) Tetracoding increases with body temperature in Lepidosauria. BioSystems 114. https://doi.org/10.1016/j.biosystems.2013.09.002

Seligmann H, Warthi G (2019) Chimeric Translation for Mitochondrial Peptides: Regular and Expanded Codons. Comput Struct Biotechnol J 17. https://doi.org/10.1016/j.csbj.2019.08.006

Shahid Masood M, Nishikawa T, Fukuoka SI et al (2004) The complete nucleotide sequence of wild rice (Oryza nivara) chloroplast genome: First genome wide comparative sequence analysis of wild and cultivated rice. Gene 340. https://doi.org/10.1016/j.gene.2004.06.008

Sharp SJ, Schaack J, Cooley L, et al (1985) Structure and transcription of eukaryotic tRNA gene. Crit Rev Biochem Mol Biol 19. https://doi.org/10.3109/10409238509082541

Smith D, Yarus M (1989) Transfer RNA structure and coding specificity. II. A D-arm tertiary interaction that restricts coding range. J Mol Biol 206. https://doi.org/10.1016/0022-2836(89)90497-X

Soltis DE, Albert VA, Leebens-Mack J, et al (2009) Polyploidy and angiosperm diversification. Am J Bot 96. https://doi.org/10.3732/ajb.0800079

Tamura K, Stecher G, Peterson D, et al (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30. https://doi.org/10.1093/molbev/mst197

Teufel AI, Liu L, Liberles DA (2016) Models for gene duplication when dosage balance works as a transition state to subsequent neo-or sub-functionalization. BMC Evol Biol 16. https://doi.org/10.1186/s12862-016-0616-1

Wang RJ, Cheng CL, Chang CC, et al (2008) Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol Biol 8. https://doi.org/10.1186/1471-2148-8-36

Wen J, Nie ZL, Ickert-Bond SM (2016) Intercontinental disjunctions between eastern Asia and western North America in vascular plants highlight the biogeographic importance of the Bering land bridge from late Cretaceous to Neogene. J Syst Evol 54

Wende S, Platzer EG, Jühling F, et al (2014) Biological evidence for the world's smallest tRNAs. Biochimie 100. https://doi.org/10.1016/j.biochi.2013.07.034

Wise RR, J. Kenneth Hoober (2007) The structure and function of plastids. Springer Science & Business Media

Yang Z, Yoder AD (1999) Estimation of the transition/transversion rate bias and species sampling. J Mol Evol 48. https://doi.org/10.1007/PL00006470

Zhang TT, Yang Y, Song XY, et al (2021) Novel structural variation and evolutionary characteristics of chloroplast trna in gossypium plants. Genes (Basel) 12. https://doi.org/10.3390/genes12060822

Zhang Z, Gerstein M (2003) Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. Nucleic Acids Res 31. https://doi.org/10.1093/nar/gkg745

Zhao J, Teufel AI, Liberles DA, Liu L (2015) A generalized birth and death process for modeling the fates of gene duplication. BMC Evol Biol 15. https://doi.org/10.1186/s12862-015-0539-2

Zhong QY, Fu XG, Zhang TT, et al (2021) Phylogeny and evolution of chloroplast tRNAs in Adoxaceae. Ecol Evol 11. https://doi.org/10.1002/ece3.7133

Zuo Z, Peng D, Yin X, et al (2013) Genome-wide analysis reveals origin of transfer RNAl tRNA halves. Mol Biol Evol 30. https://doi.org/10.1093/molbev/mst107

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

View publication stats