



OPEN

Anticodon table of the chloroplast genome and identification of putative quadruplet anticodons in chloroplast tRNAs

Tapan Kumar Mohanta^{1✉}, Yugal Kishore Mohanta² & Nanaocha Sharma^{3✉}

The chloroplast genome of 5959 species was analyzed to construct the anticodon table of the chloroplast genome. Analysis of the chloroplast transfer ribonucleic acid (tRNA) revealed the presence of a putative quadruplet anticodon containing tRNAs in the chloroplast genome. The tRNAs with putative quadruplet anticodons were UAUG, UGGG, AUAA, GCUA, and GUUA, where the GUUA anticodon putatively encoded tRNA^{Asn}. The study also revealed the complete absence of tRNA genes containing ACU, CUG, GCG, CUC, CCC, and CGG anticodons in the chloroplast genome from the species studied so far. The chloroplast genome was also found to encode tRNAs encoding N-formylmethionine (fMet), Ile², selenocysteine, and pyrrolysine. The chloroplast genomes of mycoparasitic and heterotrophic plants have had heavy losses of tRNA genes. Furthermore, the chloroplast genome was also found to encode putative spacer tRNA, tRNA fragments (tRFs), tRNA-derived, stress-induced RNA (tiRNAs), and the group I introns. An evolutionary analysis revealed that chloroplast tRNAs had evolved via multiple common ancestors and the GC% had more influence toward encoding the tRNA number in the chloroplast genome than the genome size.

The origin of the genetic code and the translation event is major transition points in evolutionary biology. The triplet genetic code is hailed as one of the most important and ultimate evolutionary anchors and indisputable evidence of life. The triplet genetic code understands the specific assignment of the amino acids in the translation machinery. Cells use the universal manual and a guided dictionary to translate the corresponding amino acids into the translating protein. The number of codon combinations on the mRNA can be an astounding number of feasible protein sequences, from which only a few can be found in nature. The triplet genetic code is believed to be universal and degenerate and accommodates twenty essential amino acids using sixty-one sense and three-stop codons. However, an emerging study has proved that the "Universal Genetic Code" is no longer universal and can be called as canonical^{1,2}. Sometimes nature enhances the protein functionalities through codon reassignment to incorporate new amino acids. This has led to the discovery of the role of selenocysteine (Sec) and pyrrolysine (Pyl) amino acids in the protein through the assignment of the stop codon as the sense codon. However, sense codon reassignment requires low-frequency codons, so stop codons are used for this purpose. It has been demonstrated that except for the triple codons, the *Escherichia coli* ribosome can accommodate codons and anticodons of variable sizes³.

Taking this opportunity, they have translated four base codon pairs, CCCU, AGGA, UAGA, and CUAG, using the four base anticodons³. Frame-shifting of the +1 nucleotide is most favorable in the absence of suppressor tRNA in *E. coli*³. The study also reveals that frame maintenance during translation is not absolute, and mutant tRNA can promote a frameshift, which can occur with high frequencies at the programmed site of the mRNA^{4,5}. Riddle and Carbon (1973) reported the presence of four base anticodons CCCC in tRNA^{Gly} instead of the wild-type CCC⁶. A study by Mohanta et al.¹⁴ revealed the presence of nine nucleotide anticodons instead of seven nucleotides⁷. These features in the tRNAs certainly explain the existence of extended codons and anticodons. Most likely, these evolutionary scenarios exist in codons and tRNAs to meet the novel translational demand.

The availability of enormous genome sequencing data is quite valuable for digging deep into the molecular features of the protein translation machinery. Significant studies have been performed in the field of codons

¹Natural and Medical Sciences Research Center, University of Nizwa, 616 Nizwa, Oman. ²Department of Applied Biology, School of Biological Sciences, University of Science and Technology Meghalaya, Baridua, Meghalaya 793101, India. ³Institute of Bioresources and Sustainable Development, Imphal, Manipur 795001, India. ✉email: nostoc.tapan@gmail.com; tapan.mohanta@unizwa.edu.om; sharma.nanaocha@gmail.com

and tRNAs (anticodons), yet, several things need to be explored. Taking this opportunity, we have conducted a large-scale study to deduce the anticodon table of the chloroplast genome to understand the existence of reduced or extended genetic codes/anticodons in tRNAs. Furthermore, we have also tried to understand the presence of Sec and Pyl tRNAs, which are part of the extended genetic code. Furthermore, an investigation has also been conducted to understand the presence of different introns and the presence of possible spacer tRNA and tRNA fragments.

Results

tRNAs with ACU, CUG, GCG, CUC, CCC, and CGG anticodons are absent in the chloroplast genome. Analysis of the chloroplast genome of the 5959 species from Algae (303), Bryophyte (69), Eudicot (3832), Gymnosperm (153), Magnoliids (182), Monocot (1177), Nymphaeales (34), protist (57), Pteridophyte (139), and unknown (13) led to the discovery of 215,966 tRNA genes. We did not find any tRNA encoded for ACU, CUG, GCG, CUC, CCC, and CGG anticodons from them (Table 1). Furthermore, we also found several anticodons, which were rare in the chloroplast genome. They were AGU (tRNA^{Thr}), AAG (tRNA^{Leu}), CGC (tRNA^{Ala}), UCA (tRNA^{Sup}), AGG (tRNA^{Pro}), AUU (tRNA^{Asn}), UAU (tRNA^{Ile}), AUA (tRNA^{Tyr}), CAG (tRNA^{Leu}), CUU (tRNA^{Lys}), CCU (tRNA^{Arg}), AAU (tRNA^{Ile}), and GAG (tRNA^{Leu}) (Table 1, Supplementary File 1). The tRNA with anticodons AAG, AGU, and CGC was found only once, whereas the tRNA with anticodon UCA, AGG, and AUU was found twice for each (Table 1, Supplementary Files 1, 2). However, the percentage of the CAU (5.47%, tRNA^{Met}) anticodon was the highest among all the 64 anticodons. The abundance of the CAU anticodon was followed by GUU, UGC, ACG, and others (Table 1).

Chloroplast genome encodes tRNA for N-formylmethionine, Ile2, selenocysteine, and pyrrolysine. A study revealed a chloroplast genome was found to encode tRNAs for tRNA^{fMet}, tRNA^{Ile2}, tRNA^{Sel}, and tRNA^{Pyl} (Table 1). The tRNA^{fMet} was encoded by the same CAU anticodon that coded tRNA^{Met}. We found 709 (0.33%) genes that encoded tRNA^{fMet} (Table 1). Also, tRNA^{Ile} encoded by the CAU anticodon was commonly referred to as tRNA^{Ile2} (Table 1). We found at least 10,575 (4.93%) tRNA genes encoding tRNA^{Ile2} (Table 1). Selenocysteine amino acid was encoded by a previously known stop codon UCA. At least 204 chloroplast genes were found to encode the UCA anticodon for tRNA^{Sel} (Table 1, Supplementary File 3). A chloroplast genome was also found to encode 197 genes for CUA anticodons that encoded tRNA^{Pyl} (Table 1). However, we did not find any CUA anticodon that encoded the suppressor tRNA (Table 1).

Chloroplast genome encodes putative duplet and quadruplet anticodons. We have already mentioned that the triplet genetic code is not universal; it is canonical. Therefore, the genome might have sup-

Ala	AGC: 214	GGC: 203	CGC: 1	UGC: 11,017		
Arg	ACG: 10,927	GCG: 0	CCG: 304	UCG: 103	CCU: 12	UCU: 5729
Asn	AUU: 2	GUU: 11,018				
Asp	AUC: 399	GUC: 6122				
Cys	ACA: 224	GCA: 6156				
Gln	CUG: 0	UUG: 6242				
Glu	CUC: 0	UUC: 5925				
Gly	ACC: 204	GCC: 4917	CCC: 0	UCC: 5684		
His	AUG: 405	GUG: 7148				
Ile	AAU: 22	GAU: 10,695	CAUIle2: 10,575	UAU: 5		
Leu	AAG: 1	GAG: 42	CAG: 6	UAG: 5745	CAA: 10,686	UAA: 5546
Lys	CUU: 7	UUU: 5312				
Met	CAU Met: 11,741					
CAUfMet: 709						
Phe	AAA: 207	GAA: 5982				
Pro	AGG: 2	GGG: 817	CGG: 0	UGG: 5883		
Ser	AGA: 203	GGA: 5520	CGA: 173	UGA: 5174	ACU: 0	GCU: 5980
Thr	AGU: 1	GGU: 5703	CGU: 454	UGU: 5669		
Trp	CCA: 5955					
Tyr	AUA: 6	GUA: 5964				
Val	AAC: 399	GAC: 10,687	CAC: 200	UAC: 4748		
Sec	UCA: 204					
Pyl	CUA: 197					
Sup	CUA:	UUA: 205	UCA: 2			

Table 1. Anticodon Table of the chloroplast genome. Study from 5959 chloroplast genomes shows several rare anticodons.

pressed or extended the genetic code, which is yet to be elucidated, to a greater extent. Our study found that the chloroplast genome encodes the putative duplet and quadruplet anticodons (Supplementary Files 4, 5). The annotation of tRNA with quadruplet anticodon was found when chloroplast genomes were annotated in the GeSeq chlorobox (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>). However, re-analysis of the tRNA with the quadruplet anticodon in tRNAscan-SE did not result in a tRNA with a quadruplet anticodon, which might be due to the default setting for identification of a tRNA with a triplet anticodon. We are the first to report the presence of duplet and quadruplet anticodons in the chloroplast genome of the plant kingdom. We found that at least 91 species were encoded quadruplet anticodons (Supplementary File 4). The quadruplet anticodons were UAUG, UGGG, AUA, GCUA, and GUUA (Supplementary File 4). The quadruplet anticodon GUUA found in *Gossypium sturtianum* (NC_023218.1) putatively encoded tRNA^{Asn}. Similarly, at least 13 species were found to encode duplet (two nucleotides) anticodons in the tRNAs of the chloroplast genome (Supplementary File 5). Among them, there were at least eight putative unique duplet anticodons, namely UG, AG, AU, CA, GA, GG, GU, and UA (Supplementary File 5). The putative duplet anticodons might have been caused by the loss of a nucleotide from the anticodon because, if there were duplet anticodons, the genome could encode only 16 anticodons in its genome and would not be able to accommodate all the 20 coding amino acids in the protein. However, there is a high possibility of having quadruplet anticodons in the tRNAs because, in a quadruplet anticodon table, there are 256 possibilities to encode different amino acids into the protein (Supplementary Table 1).

Parasitic organisms have lost the tRNA genes in their chloroplast genome. We found that some of the chloroplast genomes had lost the tRNA genes. The species that have been found to have lost the tRNA genes are *Pilostyles aethiopicus* (NC_029235.1) (Fig. 1) and *Pilostyles hamiltonii* (NC_029236.1) (Supplementary File 6). *Pilostyles aethiopicus* and *Pilostyles hamiltonii* are endoparasitic plants. Furthermore, some other plants have encoded fewer tRNAs in their chloroplast genome (Supplementary File 6). They are *Asarum minus* (5), *Gastrodia elata* (5), *Sciaphila densiflora* (6), *Epirixanthes elongata* (8), *Burmanningia oblonga* (8), *Lecanorchis japonica* (8), *Lecanorchis kiusiana* (9), and *Selaginella tamariscina* (9) (Supplementary File 6). The mentioned species encoded less than ten tRNA genes in their chloroplast genome. *Gastrodia elata* is a saprophyte, whereas, *Sciaphila densiflora*, *Epirixanthes elongata*, *Burmanningia oblonga*, *Lecanorchis japonica*, and *Lecanorchis kiusiana* are mycoheterotrophic, and *Cystopteris chinensis* is an endangered species.

The chloroplast genome of *Asarum minus* encoded UUU (tRNA^{Lys}), UUG (tRNA^{Gln}), GCU (tRNA^{Ser}), UCC (tRNA^{Gly}), and UCU (tRNA^{Arg}); *Gastrodia elata* encoded UUG (tRNA^{Gln}), GCA (tRNA^{Cys}), UUC (tRNA^{Glu}), CAU (tRNA^{Met}), and CCA (tRNA^{Trp}); *Sciaphila densiflora* encoded UUG (tRNA^{Gln}), CAU (tRNA^{Ile}), CCA (tRNA^{Trp}), CAU (tRNA^{Met}), UUC (tRNA^{Glu}), and GCA (tRNA^{Cys}); *Epirixanthes elongata* encoded CCA (tRNA^{Trp}), CAU (tRNA^{Met}), UUG (tRNA^{Gln}), GUC (tRNA^{Asp}), GUA (tRNA^{Tyr}), and UUC (tRNA^{Glu}); *Burmanningia oblonga* encoded UUG (tRNA^{Gln}), GCA (tRNA^{Cys}), GUA (tRNA^{Tyr}), UCC (tRNA^{Glu}), CAU (tRNA^{Met}), GUG (tRNA^{His}), and CAU (tRNA^{Ile}); *Lecanorchis japonica* encoded UUG (tRNA^{Gln}), GCA (tRNA^{Cys}), GUC (tRNA^{Asp}), CAU (tRNA^{Met}), GAA (tRNA^{Phe}), CAU (tRNA^{Ile}), and GUU (tRNA^{Asn}); *Lecanorchis kiusiana* encoded UUG (tRNA^{Gln}), GCA (tRNA^{Cys}), GUC (tRNA^{Asp}), UUC (tRNA^{Glu}), CAU (tRNA^{Met}), GAA (tRNA^{Phe}), CAU (tRNA^{Ile}), and GUU (tRNA^{Asn}); and *Selaginella tamariscina* encoded GUG (tRNA^{His}), GUC (tRNA^{Asp}), GUA (tRNA^{Tyr}), UUC (tRNA^{Glu}), GUU (tRNA^{Asn}), and CCA (tRNA^{Trp}). These species encoded only 14 anticodons CAU, CCA, GAA, GCA, GCU, GUA, GUC, GUG, GUU, UCC, UCU, UUC, UUG, and UUU.

Chloroplast genome encodes putative spacer tRNAs. Spacer RNA genes are usually found in bacterial genomes in the spacer region between the 16S and 23S rRNAs. When we focused our study on the spacer RNA in the chloroplast genome, we found that chloroplast genomes were also encoded in the putative spacer tRNAs between the 16 and 23S rRNA genes. tRNA^{Ala} (UGC) and tRNA^{Ile} (GAU) were the most predominant spacer tRNAs found in the chloroplast genome (Fig. 2). The percentages of the UCG and GAU anticodons in the chloroplast genome were 5.13 and 4.98, respectively. This showed that spacer tRNAs were more common in the chloroplast genome. Sometimes, it contained tRNA^{Met} (CAU) and tRNA^{Ser} (GCU) in the spacer region. All the chloroplast genomes did not encode the spacer tRNAs (Supplementary File 7). None of the mycoparasitic plants was found to encode the putative spacer tRNA in their chloroplast genome. However, the majority of the species encoded putative spacer tRNAs.

The majority of chloroplast tRNAs encode group I intron. It was found that the majority of chloroplast-encoding tRNAs encode introns. Except for tRNA^{Arg}, tRNA^{Asn}, tRNA^{Asp}, tRNA^{Gln}, tRNA^{His}, tRNA^{Pro}, tRNA^{Trp}, and tRNA^{Val} all other tRNA genes were found to contain group I introns (Table 2). The introns found in tRNA seem to be isotype-specific (Table 2). The introns are conserved within the tRNA isotype, and the conserved nucleotide sequences of the introns of one isotype do not match with the conserved introns of other isotypes (Table 2). When we cluster the conserved region of the introns, they form four groups (Supplementary Figure 1). We have named them groups A, B, C, and D. Group A contains tRNA^{Leu}, tRNA^{Tyr}, and tRNA^{Cys}; group B contains tRNA^{Ser}; group C contains tRNA^{Lys}, tRNA^{Met}, and tRNA^{Ala}; and group D contains tRNA^{Gly}, tRNA^{Ile}, tRNA^{Glu}, and tRNA^{Thr} (Supplementary Figure 1). However, the introns of tRNA^{Phe} do not group with any other introns (Supplementary Figure 1).

Chloroplast genome encodes putative novel tRNAs. Although we all are well-acquainted with the fact that tRNA makes a clover leaf-like structure, we found some variations in the tRNA structure. Analysis revealed the presence of a few novel tRNA structure/tRNA-like molecules (Fig. 3 and Fig. 4). Some putative novel tRNA-like structures seemed to lack the anticodon loop, whereas, in some cases, they had extra sequences near the anticodon arm region (Fig. 3). A tRNA-like structure contained an extended nucleotide sequence in the

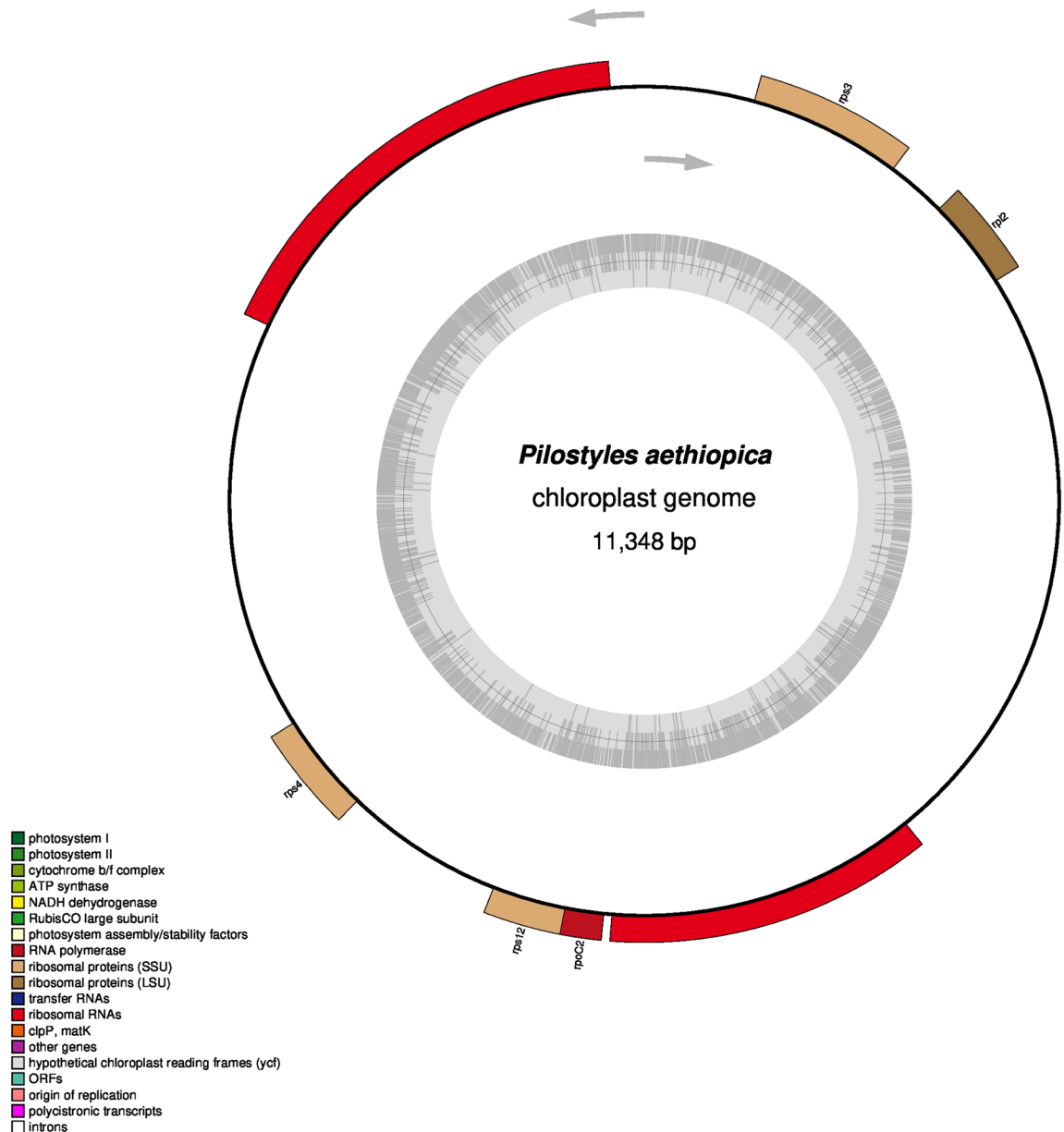


Figure 1. OGDRAW map of *Pilostyles aethiopica* (NC_029235.1) chloroplast genome. The map shows the loss of tRNA genes and inverted repeats.

region between the D-arm and anticodon arm (Fig. 4). At least 42 species were found to encode novel tRNA-like structures that contained extended nucleotide sequences between the D-arm and anticodon arm (Fig. 4). Furthermore, a few tRNAs were found to have lost the pseudouridine loop (Fig. 5), suggesting the presence of novel tRNAs/tRNA-like structures in the chloroplast genome.

Chloroplast genome encodes putative tRNA Fragments (tRFs). Analysis revealed the presence of at least 55 tRFs in the chloroplast genome. The tRFs are small 14–32 nucleotides novel class of small, non-coding RNAs, derived from the mature or precursor tRNAs that are different from the tRNA-derived, stress-induced tRNAs (tiRNAs)^{8,9}. The tRFs found were for tRNA^{Glu}, tRNA^{Arg}, tRNA^{Gly}, tRNA^{His}, tRNA^{Val}, tRNA^{Ile}, tRNA^{Thr}, tRNA^{Leu}, tRNA^{Lys}, and tRNA^{Ala} (Supplementary File 8). The tRFs of tRNA^{Glu} were found to contain conserved nucleotide sequence GGCCTTATCGTCTAGTGAT, whereas those of tRNA^{Gly} were found to contain conserved GCGGGTATAGTTT TAGTGGTAAA nucleotides (Supplementary File 8). As such, we did not find conserved nucleotide sequences for the other tRFs. The tRFs of tRNA^{Ala}, tRNA^{Gly}, tRNA^{Ile}, tRNA^{Lys}, and tRNA^{Leu} were 5'-tRFs, whereas the tRFs of tRNA^{His}, tRNA^{Thr}, and tRNA^{Val} were 3'-tRFs. The tRFs of tRNA^{Glu} did not match either the 5'- or 3'-end of the tRNA, which might have originated from the precursor tRNA transcript. Therefore, they can be classified as tRF-1.

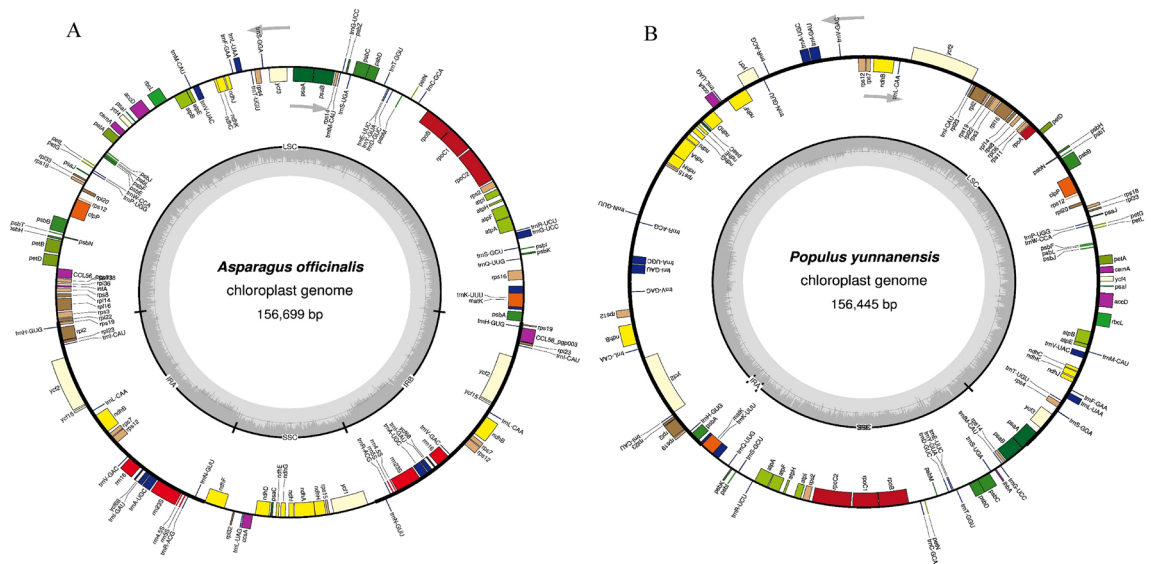


Figure 2. OGDRAWM map of (A) *Asparagus officinalis* (NC_034777.1) and (B) *Populus yunnanensis* (NC_037421.1) chloroplast genomes. (A) *Asparagus officinalis* shows the presence of a putative spacer tRNAs. tRNA^{Ala} (UGC) and tRNA^{Ile} (GAU) are present between 16 and 23S rRNA in *A. officinalis*. No putative spacer tRNA is found in the chloroplast genome of *Populus yunnanensis*.

tRNA Isotype	Conserved Consensus sequence of Intron
Alanine	A-U-U-G-G-G-U-C-G-U-U-G-C-G-A-U-U-A-C-G-G-x-G-U-x-U-G-G-A-U-G-U-C-U-A-A-U-U-G
Arginine	Not found
Asparagine	Not found
Aspartic acid	Not found
Cysteine	G-C-G-C-G-C-C-C-A-A-U-G-U-U-U-U-U-x-C-A-G-x-G-G-A-x-G-U-x-C-A-U-C-A-U-G-x-A-A-U-C-A-A-A-x ₃ -U-x-A-U G-x ₈ -U-x-U-x _{4,5} -C-A-G/A-x ₃ -A-x ₂ -U-C-x ₂ -U-x ₅ -A-A-U-x-A-x ₂ -A-x _{2,3} -U-U-G-A-U-C/U-x ₂ -U-U-U-A
Glutamic acid	A-U-U-G-C-G-U-C-G-U-U-G-U-U-G-C-x-G-G-G-C-U-G-U-G-A-x-G-G-C-U-C-U-C-A U-x ₂ -U/C-G-U/C-x-G-x-U-G-x-G-x ₈ -C-U
Glutamine	Not found
Glycine	G-x-G-x-C-x ₃ -G-C-x ₂ -U-U-x _{1,5} -C-x ₃ -U-A-U-A-x ₂ -C
Histidine	Not found
Isoleucine	A/C-G/U-U-G-C-G-x-C-A/G-U-G-U-U/G-U/G-U/C-U/C-U-x _{1,3} -C-x-G-x ₃ -A/G-G-U/G-x ₂ -A/C-U-C/U-A-x ₂ -U/G-x-C-A-x ₅ -A/U-x ₄ -U
Leucine	A-A-C-x ₅ -A-A-x-U-x ₃ -A-G-x-A-x ₂ -A-x ₂ -A-A
Lysine	A-G-U-G-C-G-x-C-U-x ₄ -U-x-U-U-U-x-A-C-A-C-A-U-U-U-x ₂ -A-U-G-A-A
Methionine	U-x-U-G-x-A-x ₂ -A-G-A-G-x-U-U-U-x _{9,10} -C-G-A-C-U-x ₂ -A-A-U-A
Phenylalanine	C-x ₂ -G-C-G-C-C-A-A-U-G-x _{1,2} -U-U-x-U-C-A-x ₂ -G-x-A-G-U-C-x-A-U-x-A-U-G-x-A-A-U-x-A-x-A-A-A-x-A
Proline	Not found
Serine	A-C-G-U-U-x-A-A-A-x-A-x-U-x _{2,7} -G-U-C-G-A-A-C-C-C-C A-x ₃ -A-x _{2,5} -G-U-C-G-A-A-C-C-C A-x _{2,7} -A-U-x ₂ -A-C-x ₄ -G-x _{1,2} -C-x ₂ -C C-x ₅ -A-A-x ₆ -A-x ₈ -U-C-x ₅ -C-x _{1,2} -U-x ₂ -A-x ₃ -C
Threonine	A-U-U-G-C-G-U-C-G-U-U-G-U-U-G-C-C-U-G-G-G-C-U-G-U-G-A-G-G-G-C-U-C-U-C-A-G-C-C-A-C-A-U-G-G-A-U-A-G-U-U-C
Tryptophan	Not found
Tyrosine	G-U-U-G-G-G-U-x-U/C-C/U-U-x ₂ -A-A-C-A-G-U-U-C-A-A-A-U-x-A-U-U-U-G-A-U-A-A-U-A-A-x-A-x-C-U-U-U-G-A-U-C-U-G-U-U-x-U-A G-x-U-U-U-x ₄ -C-x ₅ -A-x ₅ -U
Valine	Not found

Table 2. Conservation of introns in chloroplast tRNAs. From the mentioned tRNA isotypes, at least eight isotypes do not encode any intron in their tRNA genes or its not conserved.

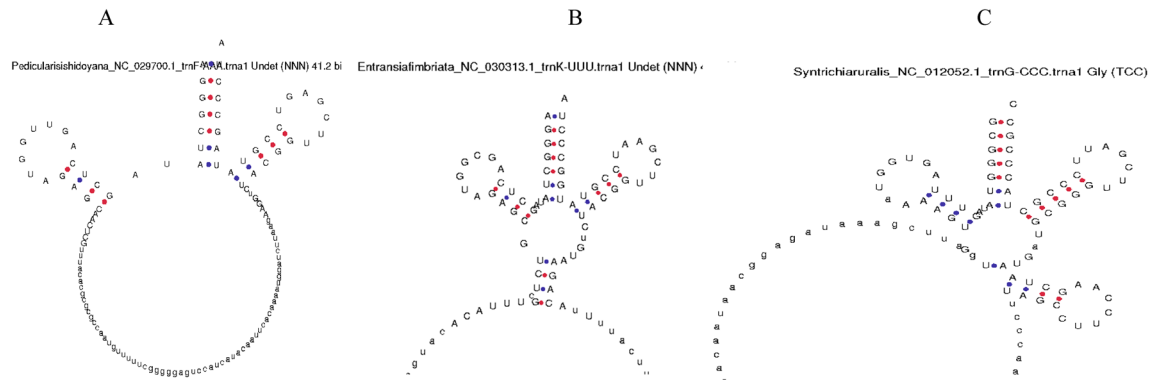


Figure 3. Putative novel tRNAs in the chloroplast genome. (A) In the tRNA of *Pedicularis ishidoyana* (NC_029700.1) there is a long nucleotide sequence present in between the D arm and anticodon arm. (B) In *Entransia fimbriata* (NC_030313.1) tRNA (tRNA^{Lys}_{UUU}) a long nucleotide sequence is present in the anticodon loop region that masks the anticodon loop. (C) In *Syntrichia ruralis* (NC_012052.1) tRNA^{Gly}_{UCC}, a long nucleotide sequence is found in between the D-arm and anticodon arm.

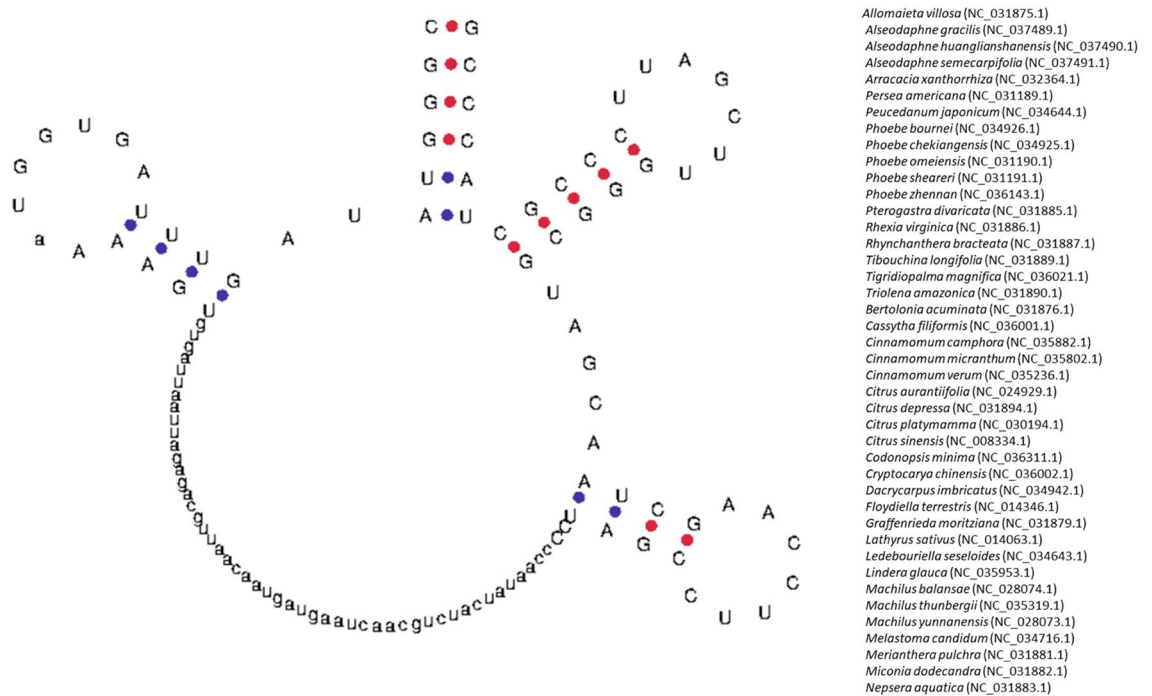


Figure 4. Putative novel tRNA of chloroplast tRNA. The tRNA contains a long nucleotide sequence in between the D-arm and anticodon arm. At least 42 chloroplast genomes are found to encode a similar tRNA structure in it. The structure was predicted using the tRNAscan-SE 2.0 program.

Chloroplast genome encodes putative tiRNAs. The longer tRFs (tRNA fragments) of 30–50 nucleotide-long sequences are called tRNA-derived, stress-induced RNAs (tiRNAs)⁸. Therefore, we searched for the presence of 30–50 nucleotide tRFs. We found at least 244 tRNA sequences, which encoded the 30–50 nucleotides (Supplementary File 9). The tiRFs were part of putative tRNA^{Ala} (UGC), tRNA^{Phe} (GAA), tRNA^{fMet} (CAU), tRNA^{Gly} (GCC, UCC), tRNA^{His} (GUG), tRNA^{Ile} (CAU, GAU), tRNA^{Lys} (UUU), tRNA^{Leu} (UAA), tRNA^{Asn} (GUU), and tRNA^{Val} (GAC, UAC) (Supplementary File 9). Among them, tiRFs of tRNA^{His} (GUG) and tRNA^{fMet} (CAU) were found only once, whereas tRNA^{Lys} (UUU) was the highest (72) encoding tiRF. The tiRFs of tRNA^{Lys} (UUU) were followed by tRNA^{Ile} (GAU) and tRNA^{Ala} (UGC), which were found to contain 51 and 52 putative tiRFs, respectively (Supplementary File 9).

Machine machine-learning approach showed GC% influences the tRNA number in the chloroplast genome. We grouped the chloroplast genomes of all the species according to their clade and conducted a comparative study. The analysis revealed that the average tRNA gene number in monocot (37.80%)

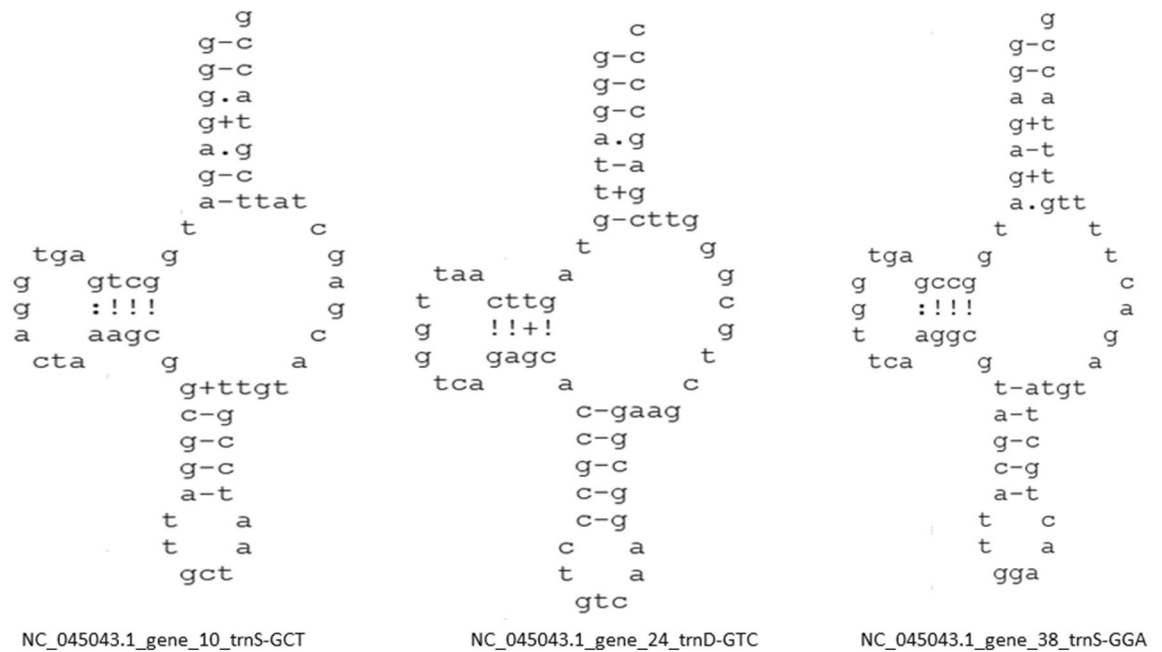


Figure 5. Fig. 5 shows the presence of putative nematode mitochondrial tRNA in the chloroplast genome. The tRNAs have been seen to lose the Ψ -arm and Ψ -loop. The presence of nematode mitochondrial genome in the chloroplast genome shows that the truncated tRNAs are shared in between the chloroplast and mitochondria. The structure has been predicted using the Aragorn software.

plants is comparatively higher than in other plants (Supplementary File 6). The protists showed the lowest (29.5%) average tRNA gene number, followed by algae (30.12%) (Supplementary File 6). A correlation analysis of the GC% with the tRNA number showed a positive correlation ($r=0.362$) for the monocot clade (Fig. 6). The chloroplast genomes of the species *Isolepis setacea* and *Vitis rotundifolia* were found to encode the highest number of tRNAs, that is, 52 each (Supplementary File 6). On average, the chloroplast genomes were found to encode 36 tRNA genes per genome. A machine-learning approach was used to understand the role of the GC content and genome size in the tRNA number in the chloroplast genome. The boosting analysis revealed that the relative influence of the GC% was more than the genome size (Fig. 7). A principal component analysis was conducted to see their association with different clades.

Chloroplast tRNAs evolve from multiple common ancestors. We conducted a phylogenetic analysis by considering the tRNA genes of the chloroplast genome. The phylogenetic analysis revealed clear and distinct phylogenetic clusters of tRNAs. The phylogenetic tree showed two major distinct clusters suggesting their origin from multiple common ancestors (Fig. 8). In cluster I, anticodons GCU, GGA, UGA, GCC, UCC, CGU, CGA, GCC, CGA, CGU, UUC, UCU, CAU, UAA, CAA, GUA, UAG, UAU, UAUG, CAA, GCU, UCG, UCU, GAA, CUA, UAG, and GAG, grouped together, whereas, in cluster II, anticodons UUG, GUG, GCA, GAA, UUU, GUU, UGG, GGG, CCA, UGU, GGU, CAU, UAC, GCC, GUC, GAC, GAU, UUC, CGU, ACG, CCG, ACA, and UGC, grouped together (Fig. 8). The anticodons GAA (tRNA^{Phe}), CAU (tRNA^{Met}), GCC (tRNA^{Gly}), UUC (tRNA^{Glu}), and CGU (tRNA^{Thr}) were shared in both clusters. The phylogenetic analysis of quadruplet anticodons revealed that quadruplet anticodon AUAA shares a phylogenetic relationship with UAUG anticodons, whereas the UGGG and GUUA anticodons fall in a distinct cluster (Fig. 9).

Genes undergo mutation, which is a common phenomenon. Although it was a common phenomenon in coding genes, non-coding genes also showed frequent mutation. Therefore, a transition/transversion bias study was conducted for the chloroplast tRNAs. The analysis revealed that transition predominates transversion (Supplementary Table 2). The transition/transversion bias was found to be the highest for tRNA^{Asn} ($R=13.71$), whereas tRNA^{Ser} (1.22) had the lowest bias (Supplementary Table 2). The transition/transversion bias of tRNA^{Asn} was followed by tRNA^{Tyr} (11.51) and tRNA^{Trp} (8.63). Although tRNA^{Arg}, tRNA^{Leu}, and tRNA^{Ser} encoded six Isoacceptors, their transition/transversion bias was comparatively lower than others (Supplementary Table 2).

Discussion

The chloroplast genome harbors several coding and non-coding sequences, including rRNA and tRNA¹⁰. These genetic elements and their potential to translate codons make them semi-autonomous organelles of the plant cell. A detailed genomic analysis of the chloroplast tRNA reveals that it does not encode all the 64 anticodons required for the tRNAs. The tRNAs with anticodons ACU (tRNA^{Ser}), CUG (tRNA^{Gln}), GCG (tRNA^{Arg}), CUC (tRNA^{Glu}), CCC (tRNA^{Gly}), and CGG (tRNA^{Pro}), are absent in the chloroplast genome of the studied species. Therefore, these anticodons can be classified as rare anticodons of the chloroplast genome. The ACU anticodon of tRNA^{Ser} and the GCG anticodon of tRNA^{Arg} are from the hexa-isoacceptor group. In contrast, the CCC anticodon

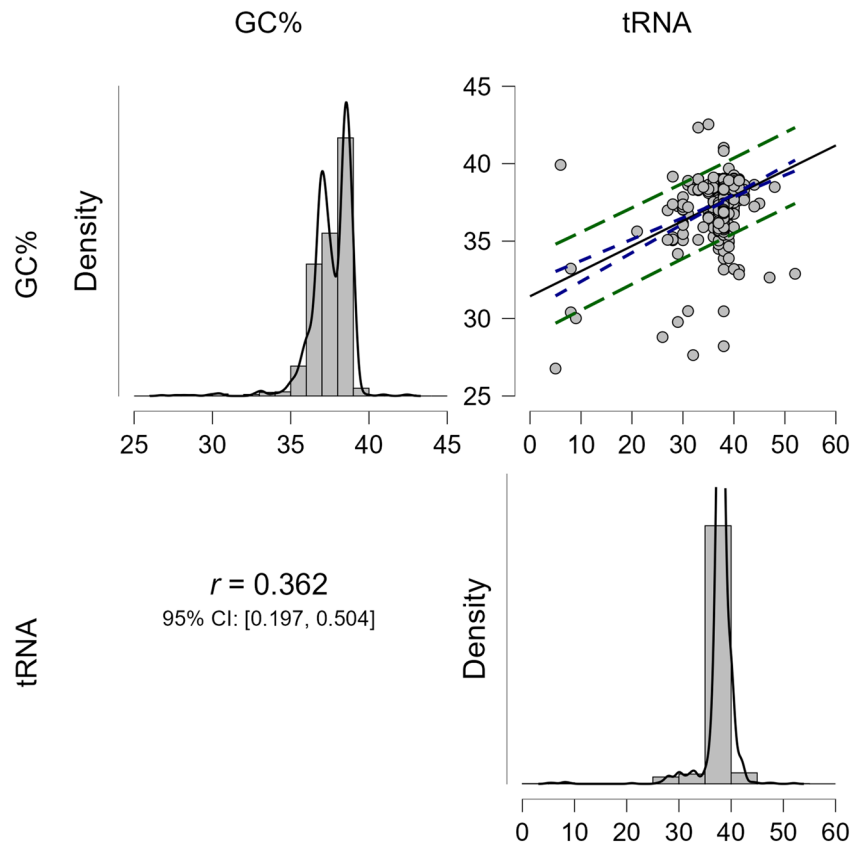


Figure 6. Correlation regression analysis ($r = 0.362$) of GC % and tRNA gene number in the chloroplast genome. Analysis showed a slight positive correlation between the chloroplast genome's GC% and tRNA gene number. The analysis was conducted at a p value < 0.05 . Correlation analysis was conducted using the JASP 0.16.1.0 version software.

of tRNA^{Gly} and the CGG anticodon of tRNA^{Pro} are from the tetra-isoacceptor group. Therefore, a lack of these anticodons from their isoacceptor group does not make any difference in the genome, as other isoacceptors are available for their use to encode the codon. However, tRNA^{Gln} is encoded only by CUG and UUG anticodons, whereas tRNA^{Glu} is encoded by the CUC and UUC anticodons. The lack of the CUG anticodon from tRNA^{Gln} and the CUC anticodon from tRNA^{Glu} in the chloroplast genome has left these tRNA isotypes with only one choice of anticodon (Table 1). The lack of the CUG anticodon in tRNA^{Gln} and the CUC anticodon in tRNA^{Glu} in the chloroplast genome may be due to a strong selection pressure to establish UUG (tRNA^{Gln}) and UUC (tRNA^{Glu}) anticodons as the dominant anticodons. The tRNA anticodons followed by nucleotides CUX ($x = \text{any nucleotide}$) may have undergone a strong evolutionary pressure. Hence anticodons CUA, CUU, CUG, and CUC, encode only 197, 7, 0, and 0 anticodons, respectively, in the chloroplast genome (Table 1). However, the CAU anticodon encoding tRNA^{Met} has been seen to have the highest percentage (5.47%) in the chloroplast genome (Supplementary File 1). The CAU anticodon of tRNA^{Met}, of the nuclear-encoded genome has also been found in the highest (5.03%) abundance¹¹, thus corroborating CAU, as the most abundant anticodon in the nuclear and chloroplast genomes. The anticodons CAU (tRNA^{Met}), GUU (tRNA^{Asp}), UGC (tRNA^{Ala}), and ACG (tRNA^{Arg}) have been found to encode more than 5% each of the total anticodons, suggesting the role of positive selection pressure in these anticodons (Supplementary File 1). However, at the isotype/isodecoder level, tRNA^{Leu} (10.27%) has been found to contain the highest percentage of anticodons, followed by tRNA^{Ile} (9.93%) and tRNA^{Arg} (7.96%) (Supplementary File 1). A similar level of abundance has been found for tRNA^{Leu} (7.80%) for the nuclear-encoded tRNA genes, reflecting a similarity in the anticodon abundance in the nuclear and chloroplast genomes¹¹. However, an abundance of the nuclear-encoded anticodons tRNA^{Leu} is followed by tRNA^{Ser} (7.66%), tRNA^{Gly} (7.52%), and tRNA^{Arg} (7.28%)¹¹. Although tRNA^{Leu} is the highest encoding isotype/isodecoder in nuclear- (7.80%) and chloroplast (10.27%)-encoded genomes, there is a great difference in their percentage. The chloroplast-encoded CAU anticodon also encodes tRNA^{Ile2} (4.93%). The CAU anticodon for tRNA^{fMet} (0.33%) is also relatively abundant in the chloroplast genome. The tRNA^{fMet} acts as an initiation anticodon in protein synthesis in mitochondria, bacteria, and chloroplasts, and the presence of tRNA^{fMet} in the chloroplast genome is quite justified. However, only 709 tRNA^{fMet} genes were found during the analysis suggesting that tRNA^{fMet} is not a universal tRNA of the chloroplast genome. A majority percentage of the chloroplast genome does not encode tRNA^{fMet}. A few chloroplast genomes encode the tRNAs for selenocysteine and pyrrolysine amino acids (Table 1). However, Zhao et al.¹² has reported the absence of tRNA^{Sec} in gymnosperm plants. The Sec amino acid specified by the UGA codon requires the presence of the selenocysteine insertion sequence (SECIS) element, and the Pyl amino acid

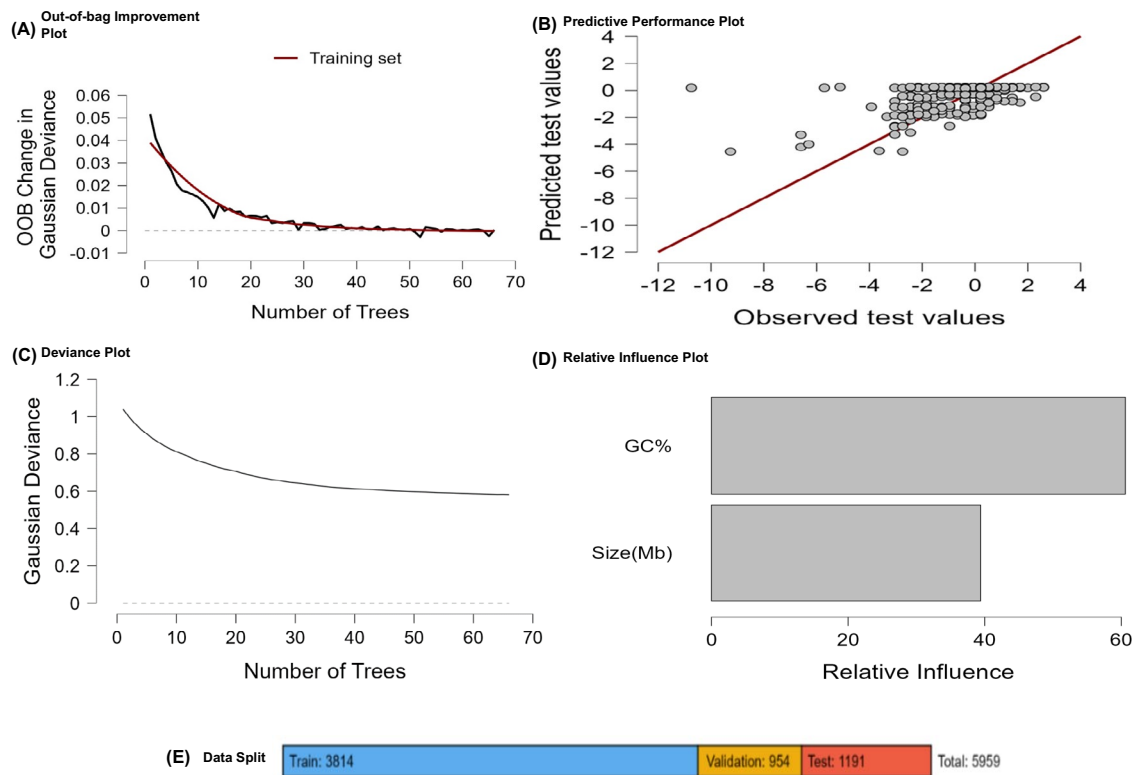


Figure 7. Machine-learning analysis of GC % content and genome size in the tRNA gene number in the chloroplast genome; the random forest approach was used to run the analysis. In the study, from 5959 species, 3814 species were used as training sets, 954 for validation, and 1191 as test sets. All the analysis was conducted at $p < 0.05$. Analysis revealed that the GC% content had more influence toward the number of tRNA gene numbers than the genome size.

encoded by the UAG codon requires the pyrrolysine insertion sequence (PYLIS)¹³. The presence of tRNA for encoding Sec and Pyl reflects that the chloroplast genome may have SECIS and PYLIS.

It was also very peculiar to see the loss of tRNA genes in the chloroplast genome of heterotrophic and mycoparasitic plants. Our previous study reported the loss of several other genes in the chloroplast genome in mycoparasitic and heterotrophic plants¹⁴. A similar is true for the tRNA genes as well. In the absence of tRNA genes in the chloroplast genome, the cell most probably uses the tRNA genes from the nuclear-encoded genome. However, the loss of tRNA genes in the chloroplast genome seems independent of the nuclear genome. The parasitic and heterotrophic plants require less effort to complete their lifecycle, as they are completely dependent on their host. Hence, they do not need a lot of genes for their function and hence, may be under constant pressure to eliminate genes. Therefore, these mycoparasitic and heterotrophic plants contain only 14 (CAU, CCA, GAA, GCA, GCU, GUA, GUC, GUG, GUU UCC, UCU, UUC, UUG, and UUU) anticodons in their chloroplast genome.

It is well known that the triplet genetic code is canonical and not universal¹⁵. The genetic code can be expanded, where specific codons can be reallocated to encode non-proteogenic amino acids. The tRNA genes undergo rapid changes to meet the translational demand of the cell¹⁶. Therefore, it is highly possible that tRNA can expand its anticodon nucleotide number. Our study helped us to discover the presence of quadruplet anticodons in the chloroplast genome of 91 plant species (Supplementary File 4). The quadruplet anticodons found in our study were UAUG, UGGG, AUAA, GCUA, and GUUA. Studies regarding functional quadruplet anticodons are reported in a few cases^{17–25}. DeBenedictis et al., (2022) reported the translation of four-base codons in natural and synthetic systems²⁵. They reported 20 isoacceptors can be changed to functional quadruplet tRNAs²⁵. Anderson et al.¹⁷ reported the role of the quadruplet codon AGGA through changes in the tRNA anticodon loop to CUUCCUAAA in a suppressor tRNA_{CUA}. The suppression of the amber tRNA led to the encoding of homoglutamine (hGln), using the AGGA codon¹⁷. They also reported that quadruplet codons CCCU or CUAG could suppress the amber tRNA and allow the incorporation of unnatural amino acids into the protein in *Escherichia coli*¹⁷. Neumann et al.¹⁸, reported the encoding of unnatural amino acids through the evolution of the quadruplet anticodon in response to the amber codon tRNA_{CUA}. Chloramphenicol resistance was achieved when tRNA_{UCCU}^{Ser2} translated the AAGA codon, and tRNA_{UCCU}^{Ser2} translated the AGGA codon¹⁸. Niu et al.¹⁹ replaced tRNA_{CUA}^{Pyl} with the UCCU anticodon and generated tRNA_{UCCU}^{Pyl}, which recognized and suppressed the quadruplet codon AGGA. This provided a qualitative notion for suppressing the quadruplet codon through tRNA_{UCCU}¹⁹. Most specifically, the presence of the quadruplet anticodon was associated with the suppression of the amber tRNA and the incorporation of the unnatural amino acid into the protein chain. The tRNA_{GCUA} contained an additional G nucleotide prior to the tRNA_{CUA} anticodon, suggesting its role in the suppression of

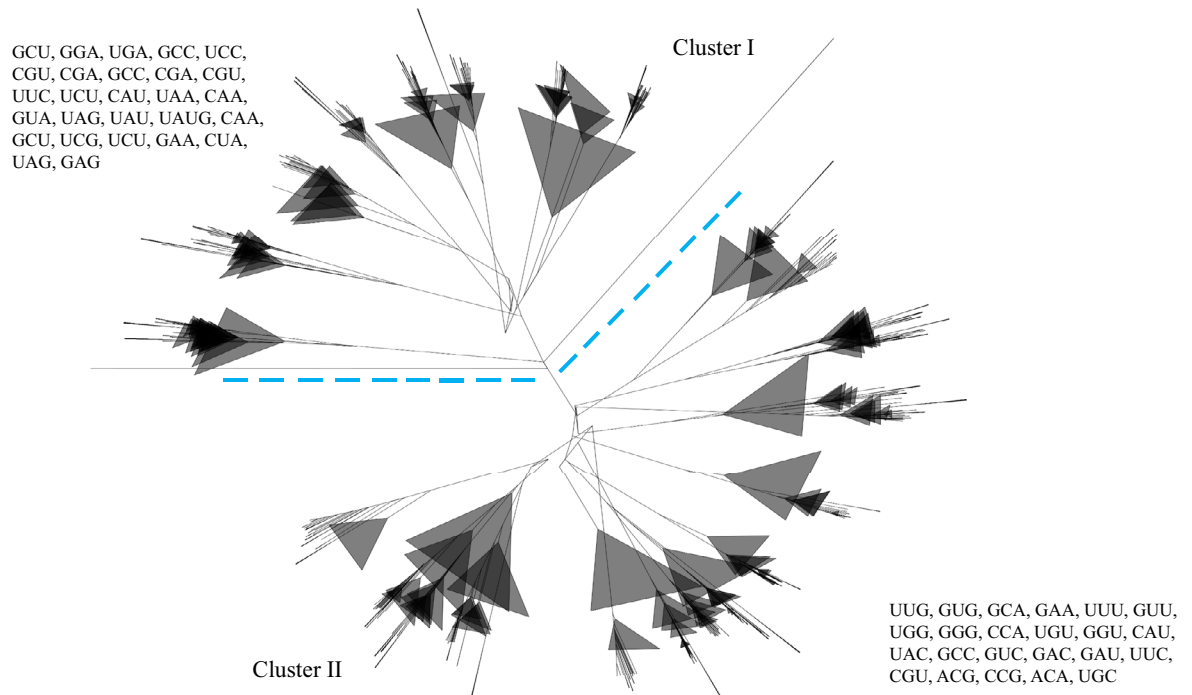


Figure 8. Phylogenetic tree of chloroplast tRNAs. The phylogenetic tree shows two distinct major clusters named clusters I and II. The phylogenetic tree shows that chloroplast tRNAs have evolved from multiple common ancestors. In cluster I, anticodons GCC, CGU, CGA, UCU, CAA, and UAG, are found in more than one group, and the anticodons GCC, GCU, UUC, CAU, and GAA are found in both clusters, showing their evolution via duplication. The phylogenetic tree has been constructed using the neighbor-joining method using the Clustal W program.

the amber codon. In the tRNA^{Asn}_{GUUA} anticodon, most probably, nucleotide A was incorporated after the GUU anticodon, as the tRNA with the GUU anticodon was grouped with the GUUA anticodon in the phylogenetic tree (Fig. 9). Similarly, in the UGG anticodon, the G nucleotide got incorporated in the UGG anticodon, as they grouped with the UGG anticodon (Fig. 9). The GCUA anticodon was grouped with GCU anticodon, suggesting that the A nucleotide was incorporated at the fourth position of the GCU anticodon, which gave rise to the GCUA anticodon (Fig. 9). However, no such clue was found in the case of the UAUG and AUAA anticodons. Considering the incorporation of the additional nucleotide at the fourth position, we could speculate that the G nucleotide was most probably incorporated in the UAU anticodon and gave rise to the UAUG anticodon. Similarly, the A nucleotide was incorporated at the fourth position of the AUA anticodon to give rise to the AUAA anticodon. Although we found only five putative quadruplet anticodons, the genome could accommodate at least 256 quadruplet anticodons/codons in the cell (Table 2). We also found the presence of tRNAs, with only duplet anticodon, where one nucleotide was possibly deleted from the anticodon (Supplementary File 5). At least 13 species contained duplet anticodons in the tRNA of the chloroplast genome (Supplementary File 5).

The chloroplast encoding tRNAs were also found to encode the group I introns. These group I introns were conserved in their respective isotype/isodecoder groups (Table 2). From a total of 20 isotypes, 12 of them were found to encode the group I introns (Table 2). However, the group I intron of one isotype was not conserved with the intron of another isotype, reflecting the isotype-based conservation of the group I intron, in the tRNA.

It is well-reported that group I introns are found in tRNAs, bacteria, lower eukaryotes, and higher plants^{26–28}. Some of the group I intron encode homing endonucleases catalyze intron mobility, thus facilitating the movement of the intron from one location to another and from one organism to another²⁷. However, the incorporation of the group I intron in the tRNA gene is isotype-specific, as only 12 isotypes have been found to encode the intron, while eight isotypes do not have any intron in their tRNAs (Table 2). From the eight isotypes, tRNA^{His}, tRNA^{Gln}, tRNA^{Asp}, tRNA^{Asn}, and tRNA^{Arg} belong to the polar group, whereas, tRNA^{Trp}, tRNA^{Pro}, and tRNA^{Val} belong to the non-polar group. This shows that the presence of the type I intron tends to be more toward the tRNA that encodes polar amino acids. Furthermore, it is seen that the chloroplast genome also encodes the putative spacer tRNAs (Fig. 2). It is reported that *E. coli* contains a spacer tRNA (tRNA^{Ala} and tRNA^{Ile}) that is present in the spacer region of the 16S and 23S rRNA²⁹. The tRNAs, tRNA^{Ala}, and tRNA^{Ile}, have also been found in the spacer region of 16S and 23S rRNA, suggesting the presence of a spacer tRNA in the chloroplast genome. Although, in a majority of cases, tRNA^{Ala} and tRNA^{Ile} are the predominant spacer tRNAs; tRNA^{Glu} can be the third most possible spacer tRNA of the chloroplast genome.

The analysis also revealed the presence of tRNA fragments (tRFs) in the chloroplast genome. We found at least 55 tRFs that belonged to ten tRNA isotypes (Supplementary File 8). These tRFs were putatively derived

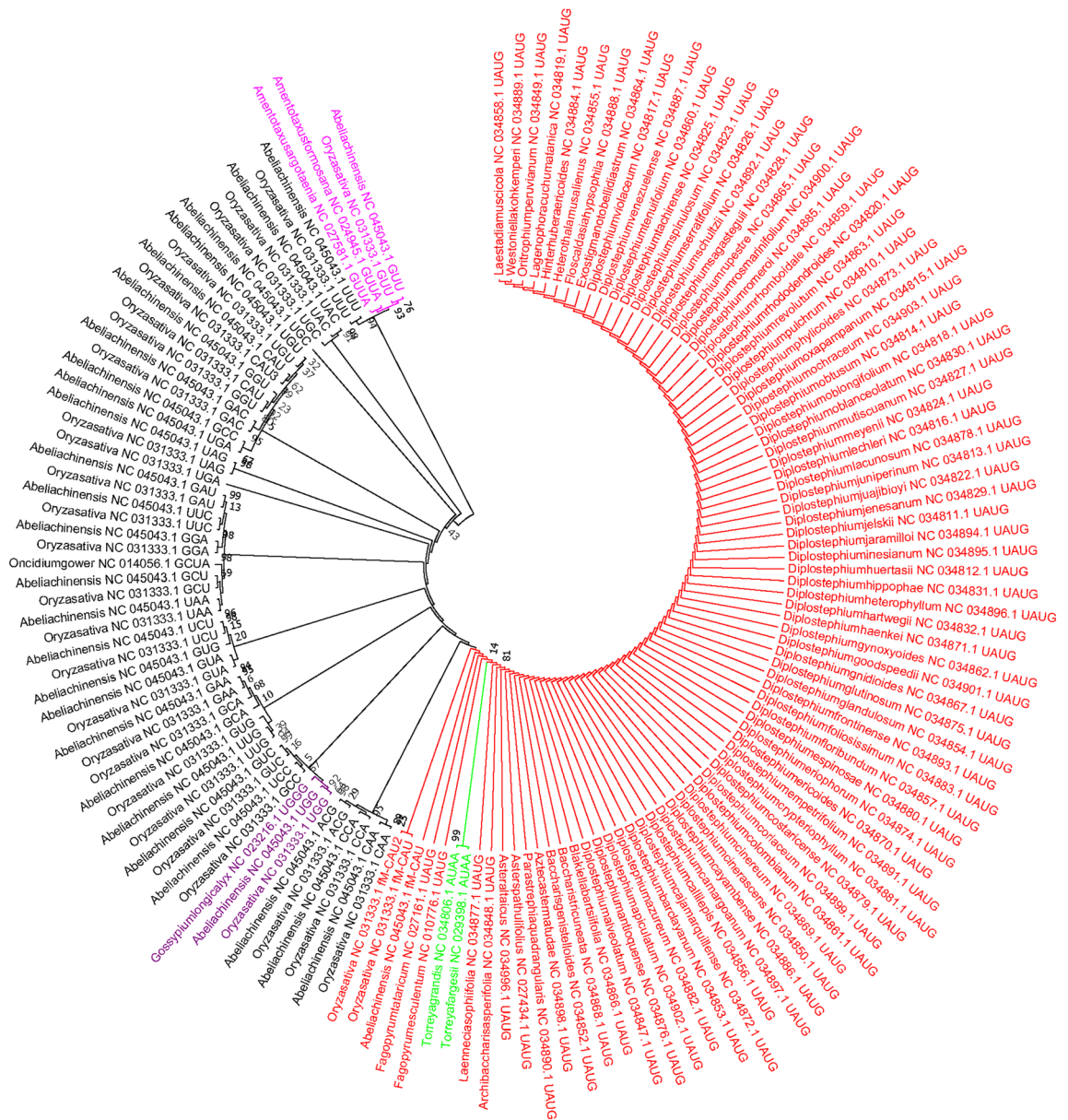


Figure 9. Phylogenetic tree of a putative quadruplet anticodon containing tRNAs with triplet codon-containing tRNAs. The phylogenetic grouping revealed that the quadruplet anticodons had evolved via the addition of a nucleotide preceding the third nucleotide of the triplet anticodons. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura–Nei model. The tree with the highest log likelihood (−1053.93) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying the Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances, estimated by using the Maximum Composite Likelihood (MCL) approach and then selecting the topology with the superior log likelihood value. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 147 nucleotide sequences. All positions with less than 95% site coverage were eliminated; that is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 26 positions in the final dataset. Evolutionary analyses were conducted using the MEGA 7.

from the tRNA precursors or from the cleavage of mature tRNAs³⁰. The tRFs were reported to control gene expression, translation control, transposon control, ncRNA, and DNA damage response^{8,30–32}. Although we found ten different chloroplast-derived tRFs, the majority of them belonged to tRNA^{Glu} and tRNA^{Gly} (Supplementary File 8). Among them are the, tRNA^{Glu} are tRF-1 type, tRNA^{Gly} are tRF-5'-type, and tRNA^{His}, tRNA^{Thr}, and tRNA^{Val} are tRF-3' type (Supplementary File 8). Furthermore, we also noted the presence of a few putative tRNA-derived, stress-induced RNA (tiRNAs) fragments (tiRFs) in the chloroplast genome. The majority of the tiRFs were from tRNA^{Lys} (UUU). For the first time, tiRFs were reported in the human fetus hepatic tissue and osteosarcoma cells^{33,34}. These tiRFs could be generated in the cell under different stress conditions via cleavage of mature tRNAs³³. However, their presence as independent nucleotide fragments in the annotated genome

sequence reflected their independent presence in the genome. Although the cleavage of tRNAs to tiRFs was brought about by the enzyme angiogenin (an RNase superfamily)³⁴ in the human cell, its counterpart in plants needs to be identified to understand its detailed functions. The 5'-tRNA^{Ala} and tRNA^{Cys} were reported to inhibit translation in rabbit reticulocytes³⁴ suggesting their inhibitory role in protein translation.

This study also found a putative novel tRNA structure encoded by the chloroplast genome (Fig. 4). The tRNA^{Gly} (UCC) was found to contain a long nucleotide sequence between the D-arm and anticodon arm in several species. This long arm could most probably be an intron that might have been incorporated in between these two arms. The chloroplast tRNAs, which had lost the pseudouridine loop (Ψ), seemed to be metazoan mitochondrial-specific (Fig. 5). The loss of the Ψ -loop in tRNA was first reported in the 1970s^{35–37}. Previous studies also reported the loss of the Ψ -arm and loop in nematode mitochondrial tRNA³⁷. However, in the nematode mitochondrial tRNA, the Ψ -arm and loop were present in the tRNA^{Ser} (GCU), whereas it had lost the Ψ -arm and loop in tRNA^{Ser} (GCU and GGA) in the chloroplast genome (Fig. 5). The elongation factor (EF) Tu combined with GTP to form a complex that delivered the amino-acyl tRNA to the ribosome A site through binding of the acceptor, s arm and Ψ -arm³⁸. In the absence of the Ψ -arm and loop in the tRNAs, it might use some alternative binding mode for EF-Tu^{39,40}. *Caenorhabditis elegans* mitochondrial EF-Tu, it has around 60 amino acid extensions at the C-terminal end that might play an essential role in binding tRNAs that lack the Ψ -arm^{41,42}. This also suggested that the mitochondrial ribosomal protein might have alternate binding sites for the truncated tRNA. Furthermore, the metazoan, mitochondria-specific, truncated tRNA in the chloroplast genome suggested that these tRNA genes might be shared by sub-cellular organelle chloroplast and mitochondria.

Evolutionary analysis revealed that chloroplast tRNAs are derived from multiple common ancestors (Fig. 8). The phylogenetic tree of the chloroplast tRNA shows two distinct clusters, which reflect their evolution from multiple common ancestors. In cluster I, anticodons GCC, CGU, CGA, UCU, CAA, and UAG make more than one group, whereas none of the anticodons from cluster II are found to make more than one group (Fig. 8). The anticodons GCC, GCU, UUC, CAU, and GAA are also found in both clusters (Fig. 8). This suggests that tRNAs with anticodons GCC, CGU, CGA, UCU, CAA, and UAG, of cluster I may have undergone vivid duplication and produced more than one anticodon group.

Conclusions

Chloroplast is a semiautonomous organelle of the plant and protist kingdom with a great potential to encode its own genome and protein translation machinery. The important tRNA molecules required for the protein translation process are well documented. The chloroplast genome encodes putative duplet, triplet, and quadruplet anticodons suggesting their role in recognizing duplet, triplet, and quadruplet codons in the mRNA. Mycoparasitic plants have lost their chloroplast genome to a large extent, thereby losing several chloroplast-encoded tRNA genes. Further, several of the chloroplast-encoded tRNA genes were found to encode introns, and the presence of intron in the chloroplasts genome suggests the presence of introns in the gene of their prokaryotic ancestor cyanobacteria. Further, the chloroplast genome is very selective and encoded only a few isoacceptors abundantly, while GCG, CUG, CUC, CCC, CGG, and ACU anticodons were found to be the rarest form of anticodons in the chloroplast genome. It is important to understand why chloroplast genomes do not encode tRNA with such anticodons. The tRNAs with quadruplet anticodons will enable us to provide a platform for the synthetic biologist to engineer tRNAs with quadruplet anticodons to understand the expansion of quadruplet genetic code.

Materials and methods

All the chloroplast genomes were downloaded from the National Center for Biotechnology Information (NCBI) database. In total, 5959 chloroplast genomes were used in this study. The downloaded chloroplast genomes were subjected to tRNA annotation. tRNA annotation was conducted using tRNAscan-SE 2.0, Aragorn, and the GeSeq-Annotation of the organellar genomes^{43–45}. The Linux-based approach was used to annotate the chloroplast tRNA for tRNAscan-SE 2.0 and Aragorn. In the GenSeq-annotation of the organellar genome, the chloroplast genome files were uploaded with the following parameters; sequence source: plastid; annotation option: Annotate plastid inverted repeats; blat search: Default; annotate: CDS, tRNA, and rRNA; and third-party tRNA annotator: Aragorn v1.2.38, tRNAscan-SE v2.0.7. All the tRNA sequences generated from these three annotation pipelines were corroborated and used for further analysis. All the data obtained from tRNAscan-SE and Aragorn were further processed in an excel worksheet. The Organellar Genome Draw (OGDRAW) was used to draw the organellar genome map of the chloroplast genome⁴⁶. The Genbank file was used to draw the chloroplast genome map in OGDRAW⁴⁶.

Multiple sequence alignment. The intron sequences retrieved from the chloroplast tRNA were aligned to find the possible conserved structure. Multiple sequence alignment was conducted using the Multalin software (<http://multalin.toulouse.inra.fr/multalin/>) that uses hierarchical clustering⁴⁷. Default parameters were used to construct the alignment.

Machine learning approach and statistical analysis. A machine learning approach was used to understand the role of the genome size and GC% content in the number of tRNA genes in the chloroplast genome. The random forest regression approach was used for this purpose. The following parameters were used in the random forest analysis: target tRNA gene number, predictor's genome size, and GC% content; Plots: data split, out-of-bag error, predictive performance, the mean decrease in accuracy, and the total increase in node purity; tables: evaluation matrix; data split preference: sample 20% of all data; training and validation of data: 20% validation data. The training parameters were as follows; training data used per tree: 50%; predictor per split: auto; and the max tree: 100%. The machine-learning approach was studied using the JASP software version

0.16.1.0⁴⁸. The correlation plot for GC% content and tRNA was also conducted using the JASP 0.16.1.0 software. The following parameters were used for the correlation analysis, sample correlation coefficient: Pearson's r and confidence interval: 95% ($p < 0.05$)⁴⁸.

Phylogenetic tree. The tRNA sequences of the chloroplast genomes were taken to construct the phylogenetic tree. The phylogenetic tree was constructed using the Clustalw program (version 2.1) in a Linux-based environment. A neighbor-joining tree was constructed with 100 bootstrap replicates. The resulting file was saved in nwk file format and later uploaded in the iTOL Interactive Tree of life (version 6) to view tree⁴⁹. The phylogenetic tree of the tRNA quadruplet anticodons, with other anticodons, was constructed using the MEGA software version 7⁵⁰. Prior to the construction of the phylogenetic tree, the tRNA sequences were subjected to multiple sequence alignments. Multiple sequence alignments were conducted using the MUSCLE software version 1⁵¹. The resulting clustal file was converted to the MEGA file format (aln) using the MEGA 7 software⁵⁰. The converted file was subjected to construct the phylogenetic tree in the MEGA 7 software, using the maximum-likelihood approach. The phylogenetic tree of the tRNA introns was also constructed using the MEGA 7 software with the same statistical parameters⁵⁰. The Tamura-Nei model, with 500-bootstrap replicates was used for the analysis.

Data availability

All the data used during this study was taken from National Center for Biotechnology Information database, and all the data are available in the public domain. Also, the accession numbers are provided in the supplementary files.

Received: 1 July 2022; Accepted: 10 January 2023

Published online: 14 January 2023

References

- Lehman, N. Molecular evolution: Please release me, genetic code. *Curr. Biol.* **11**, R63–R66 (2001).
- Keeling, P. J. & Doolittle, W. F. A non-canonical genetic code in an early diverging eukaryotic lineage. *EMBO J.* **15**, 2285–2290 (1996).
- Magliery, T. J., Anderson, J. C. & Schultz, P. G. Expanding the genetic code: Selection of efficient suppressors of four-base codons and identification of 'shifty' four-base codons with a library approach in *Escherichia coli*. *J. Mol. Biol.* **307**, 755–769 (2001).
- Atkins, J. F., Weiss, R. B., Thompson, S. & Gesteland, R. F. Towards a genetic dissection of the basis of triplet decoding, and its natural subversion: Programmed reading frame shifts and hops. *Annu. Rev. Genet.* **25**, 201–228 (1991).
- Kurland, C. G. Translational accuracy and the fitness of bacteria. *Annu. Rev. Genet.* **26**, 29–50 (1992).
- Riddle, D. & Carbon, J. Frameshift suppression: A nucleotide addition in the anticodon of a glycine transfer RNA. *Nat. New Biol.* **242**, 230–234 (1973).
- Mohanta, T. K. *et al.* Analysis of genomic tRNA revealed presence of novel genomic features in cyanobacterial tRNA. *Saudi J. Biol. Sci.* **27**, 124–133 (2019).
- Yu, M. *et al.* tRNA-derived RNA fragments in cancer: Current status and future perspectives. *J. Hematol. Oncol.* **13**, 121 (2020).
- Kumar, P., Mudunuri, S. B., Anaya, J. & Dutta, A. tRFdb: A database for transfer RNA fragments. *Nucl. Acids Res.* **43**, D141–D145 (2015).
- Song, W. *et al.* Comparative analysis the complete chloroplast genomes of nine *Musa* species: Genomic features, comparative analysis, and phylogenetic implications. *Front. Plant Sci.* **13**, 832884 (2022).
- Mohanta, T. K. *et al.* Construction of anti-codon table of the plant kingdom and evolution of tRNA selenocysteine (tRNA^{Sec}). *BMC Genomics* **21**, 804 (2020).
- Zhao, Y.-H. *et al.* Evolution and structural variations in chloroplast tRNAs in gymnosperms. *BMC Genomics* **22**, 750 (2021).
- Zhang, Y., Baranov, P. V., Atkins, J. F. & Gladyshev, V. N. Pyrrolysine and selenocysteine use dissimilar decoding strategies. *J. Biol. Chem.* **280**, 20740–20751 (2005).
- Mohanta, T. K. *et al.* Gene loss and evolution of the plastome. *Genes* **11**, 1133 (2020).
- Karasev, V. A. The canonical table of the genetic code as a periodic system of triplets. *Biosystems* **214**, 104636 (2022).
- Yona, A. H. *et al.* tRNA genes rapidly change in evolution to meet novel translational demands. *Elife* **2013**, 1–17 (2013).
- Anderson, J. C. *et al.* An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7566–7571 (2004).
- Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441–444 (2010).
- Niu, W., Schultz, P. G. & Guo, J. An expanded genetic code in mammalian cells with a functional quadruplet codon. *ACS Chem. Biol.* **8**, 1640–1645 (2013).
- Watanabe, T., Muranaka, N. & Hohsaka, T. Four-base codon-mediated saturation mutagenesis in a cell-free translation system. *J. Biosci. Bioeng.* **105**, 211–215 (2008).
- Anderson, J. C. & Schultz, P. G. Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. *Biochemistry* **42**, 9598–9608 (2003).
- de la Torre, D. & Chin, J. W. Reprogramming the genetic code. *Nat. Rev. Genet.* **22**, 169–184 (2021).
- deBenedictis, E. A., Carver, G. D., Chung, C. Z., Söll, D. & Badran, A. H. Multiplex suppression of four quadruplet codons via tRNA directed evolution. *Nat. Commun.* **12**, 5706 (2021).
- Wang, K., Schmied, W. H. & Chin, J. W. Reprogramming the genetic code: From triplet to quadruplet codes. *Angew. Chemie Int. Ed.* **51**, 2288–2297 (2012).
- DeBenedictis, E., Söll, D. & Esvelt, K. Measuring the tolerance of the genetic code to altered codon size. *Elife* **11**, 1–19 (2022).
- Hausner, G., Hafez, M. & Edgell, D. R. Bacterial group I introns: Mobile RNA catalysts. *Mob. DNA* **5**, 8 (2014).
- Zhou, Y. *et al.* GISSD: Group I intron sequence and structure database. *Nucleic Acids Res.* **36**, D31–D37 (2008).
- Mohanta, T., Syed, A., Ameen, F. & Bae, H. Novel genomic and evolutionary perspective of cyanobacterial tRNAs. *Front. Genet.* **8**, 200 (2017).
- Lund, E. & Dahlberg, J. E. Spacer transfer RNAs in ribosomal RNA transcripts of *E. coli*: Processing of 30S ribosomal RNA in vitro. *Cell* **11**, 247–262 (1977).
- Molla-Herman, A. *et al.* tRNA Fragments populations analysis in mutants affecting tRNAs processing and tRNA methylation. *Front. Genet.* **11**, 518949 (2020).

31. Goodarzi, H. *et al.* Endogenous tRNA-derived fragments suppress breast cancer progression via YBX1 displacement. *Cell* **161**, 790–802 (2015).
32. Kescu, C. *et al.* tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *RNA* **24**, 1093–1105 (2018).
33. Fu, H. *et al.* Stress induces tRNA cleavage by angiogenin in mammalian cells. *FEBS Lett.* **583**, 437–442 (2009).
34. Yamasaki, S., Ivanov, P., Hu, G. & Anderson, P. Angiogenin cleaves tRNA and promotes stress-induced translational repression. *J. Cell Biol.* **185**, 35–42 (2009).
35. Baer, R. J. & Dubin, D. T. The sequence of a possible 5S RNA-equivalent in hamster mitochondria. *Nucleic Acids Res.* **8**, 3603–3610 (1980).
36. Dubin, D. T. & Friend, D. A. Comparison of cytoplasmic and mitochondrial 4 s RNA from cultured hamster cells: Physical and metabolic properties. *J. Mol. Biol.* **71**, 163–175 (1972).
37. Watanabe, Y.-I., Suematsu, T. & Ohtsuki, T. Losing the stem-loop structure from metazoan mitochondrial tRNAs and co-evolution of interacting factors. *Front. Genet.* **5**, 109 (2014).
38. Poul, N. *et al.* Crystal structure of the ternary complex of phe-tRNA^{Phe}, EF-Tu, and a GTP analog. *Science* **270**, 1464–1472 (1995).
39. Ohtsuki, T., Sato, A., Watanabe, Y. & Watanabe, K. A unique serine-specific elongation factor Tu found in nematode mitochondria. *Nat. Struct. Biol.* **9**, 669–673 (2002).
40. Arita, M. *et al.* An evolutionary ‘intermediate state’ of mitochondrial translation systems found in *Trichinella* species of parasitic nematodes: co-evolution of tRNA and EF-Tu. *Nucl. Acids Res.* **34**, 5291–5299 (2006).
41. Ohtsuki, T. *et al.* An “Elongated” translation elongation factor Tu for truncated tRNAs in nematode mitochondria. *J. Biol. Chem.* **276**, 21571–21577 (2001).
42. Sakurai, M., Watanabe, Y., Watanabe, K. & Ohtsuki, T. A protein extension to shorten RNA: elongated elongation factor-Tu recognizes the D-arm of T-armless tRNAs in nematode mitochondria. *Biochem. J.* **399**, 249–256 (2006).
43. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: Improved detection and functional classification of transfer RNA genes. *Nucl. Acids Res.* **49**, 9077–9096 (2021).
44. Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucl. Acids Res.* **32**, 11–16 (2004).
45. Tillich, M. *et al.* GeSeq - versatile and accurate annotation of organelle genomes. *Nucl. Acids Res.* **45**, W6–W11 (2017).
46. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucl. Acids Res.* **47**(W1), W59–W64 (2019).
47. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.* **16**, 10881–10890 (1988).
48. Team, J. JASP (Version 0.16.1). <https://jasp-stats.org/> (2022).
49. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucl. Acids Res.* **49**, W293–W296 (2021).
50. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
51. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* **32**, 1792–1797 (2004).

Acknowledgements

The authors would like to extend their sincere thanks to the Natural and Medical Sciences Research Center, University of Nizwa, Oman, for providing required facility to conduct the research.

Author contributions

T.K.M.: Conceived the idea, ran the experiment, analyzed data, drafted and revised the manuscript; Y.K.M.: analyzed data; N.S.: revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-27886-9>.

Correspondence and requests for materials should be addressed to T.K.M. or N.S.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023