

Assembly and comparative analysis of the complete mitochondrial genome of *Trigonella foenum-graecum* L.

Yanfeng He

Qinghai Minzu University

Wenya Liu

Qinghai Minzu University

Jiuli Wang (✉ wang_jiul@163.com)

Qinghai Minzu University

Research Article

Keywords: *Trigonella foenum-graecum* L., Mitochondrial genome, Sequence characteristics, Phylogenetic analysis

Posted Date: February 24th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-2593489/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Genomics on December 8th, 2023. See the published version at <https://doi.org/10.1186/s12864-023-09865-6>.

Abstract

Background

Trigonella foenum-graecum L. (*T. foenum-graecum*) is a Leguminosae plant, and the stems, leaves, and seeds of this plant are rich in chemical components that are of high research value. The chloroplast (cp) genome of *T. foenum-graecum* has been reported, but the mitochondrial (mt) genome remains unexplored.

Results

In this paper, we use second- and third-generation sequencing methods, which have the dual advantage of combining high accuracy and longer read length. The *T. foenum-graecum* mitochondrial genome was assembled and other analyses such as annotation of the assembled sequences were performed. The results showed that the mitochondrial genome of *T. foenum-graecum* was 345,604 bp in length and 45.28% in GC content. There are 59 genes, including: 33 protein-coding genes (PCGs), 21 tRNA genes, 4 rRNA genes and 1 pseudo gene. Among them, 11 genes contained introns. Significant AT preferences for codons in the mitochondrial genome of *T. foenum-graecum*. A total of 202 dispersed repetitive sequences, 96 simple repetitive sequences (SSRs) and 19 tandem repetitive sequences were detected. Nucleotide polymorphism analysis counted the variation in each gene, with *atp6* being the most notable. Both synteny and phylogenetic analyses showed that *T. foenum-graecum* was similar to *Trifolium pratense*, *Trifolium meduseum*, *Trifolium grandiflorum*, *Trifolium aureum*, *Medicago truncatula*, which are five species of Leguminosae with high similarity. Among them, the highest similarity with *Medicago truncatula* was 100%. The interspecies non-synonymous substitutions (K_a)/synonymous substitutions (K_s) results showed that 23 Protein-coding genes had $K_a/K_s < 1$, indicating that these genes would continue to evolve under purifying selection pressure. In addition, 23 homologous sequences were detected in the mitochondrial genome of *T. foenum-graecum*, and tRNAs were more conserved than PCGs during gene migration.

Conclusions

This paper explores the mitochondrial genome sequence information of *T. foenum-graecum* and advances the phylogenetic diversity of Leguminosae plants.

Background

T. foenum-graecum is a plant of Leguminosae *Trigonella*, which has been used for thousands of years[1]. Its leaves are used as a vegetable and the seeds are used to make spices[1]. In addition, *T. foenum-graecum* is rich in sugars, proteins, lipids and other nutrients[2]; the seeds and leaves contain a variety of active chemicals, such as flavonoids and alkaloids[3]. It is effective in a variety of ways, including analgesic and anti-inflammatory, antioxidant, and hypoglycaemic[4], which is a biologically active substance of great research value. The composition, structure and efficacy of *T. foenum-graecum* have been intensively studied

in the literature, but the study of in vivo organelles is relatively lacking, while mitochondria as energy converters have not been decoded so far.

Mitochondria consist of four parts: matrix, inner membrane, membrane gap, and outer module, and possess their own mitochondrial DNA (mtDNA), and mtDNA can participate in encoding some RNAs and polypeptides[5, 6]. Mitochondria are one of the energy converters in plant cells. In addition to providing energy, it can also serve as a hub for metabolism or signaling, and is closely related to apoptosis, necrosis, differentiation and other vital activities[7, 8]. It plays an important role in the growth and development of plants.

The size, structure, gene arrangement order and content of the mt genome[9], as well as the rate of evolution among different taxa, are highly variable. However, their gene numbers, types, and functions do not change much, and they again show relatively conserved characteristics[10]. It is generally accepted that plant mtDNA consists of a "master circle" conformation of the entire sequence content of all genomes and a set of subgenomic circle interconverted by repeat-mediated recombination[11]. Because of this, "master circle" and subgenomic circle can coexist in the cell, making the structure of plant mt genomes more complex and difficult to study. The mt genomes of angiosperms usually range from 200–750 kb[10], and the size varies significantly among plants. The number of editing sites for mt RNA in higher plants is greater than 400, which is about 13 times the number of editing sites for cp RNA[12]. Repeated sequences of plant mt genomes undergo frequent recombination, making their structures increasingly complex[13]. Based on the above multiple factors, sequence characterization and phylogenetic analysis of the mt genome of *T. foenum-graecum* were investigated in depth for a more comprehensive understanding of the genetic characteristics and affinities of Leguminosae *Trigonella*.

Results

Basic characteristics of *T. foenum-graecum* mt genome

The *T. foenum-graecum* mt genome was circular in structure with a total length of 345,604 bp and a GC content of 45.28%. The GC content of PCGs (42.72%) was lower than that of tRNA (52.41%) and rRNA (51.2%). The mt genome structure is shown in Fig. 1. There were 59 genes, including 33 protein-coding genes, 21 tRNA genes, 4 rRNA genes and 1 pseudo gene. The classification of genes in the mt genome of *T. foenum-graecum* is shown in Table 1. Among them, there are 11 genes with introns (ccmFC, nad1, nad2, nad4, nad5, nad7, rps3, rps7, rps10, trnP-CGG, trnT-TGT) containing a total of 25 introns. NADH dehydrogenase contains the largest number of introns, 19 in total. In addition, two copies of rrn26, trnF-GAA, trnG-GC and four copies of trnM-CAT were found in the *T. foenum-graecum* mt genome. The rps1 is a pseudo gene.

Table 1 Gene classification table

Group of genes	Gene name	Length	Start codon	Stop codon	Amino acid
ATP synthase	<i>atp1</i>	1518	ATG	TAA	506
	<i>atp4</i>	588	ATG	TAA	196
	<i>atp6</i>	663	ATG	CAA (TAA)	221
	<i>atp8</i>	483	ATG	TAA	161
	<i>atp9</i>	225	ATG	TAA	75
Cytochrome c biogenesis	<i>ccmB</i>	621	ATG	TGA	207
	<i>ccmC</i>	747	ATG	TGA	249
	<i>ccmFC*</i>	1431	ATG	TAG	477
	<i>ccmFN</i>	1728	ATG	TGA	576
Ubichinol cytochrome c reductase	<i>cob</i>	1179	ATG	TGA	393
Cytochrome c oxidase	<i>cox1</i>	1584	ATG	TAA	528
	<i>cox2</i>	906	ATG	TAA	302
	<i>cox3</i>	798	ATG	TGA	266
Maturases	<i>matR</i>	1986	ATG	TAG	662
Transport membrane protein	<i>mttB</i>	312	ATG	TGA	104
NADH dehydrogenase	<i>nad1****</i>	978	ACG (ATG)	TAA	326
	<i>nad2****</i>	1467	ATG	TAA	489
	<i>nad3</i>	357	ATG	TAA	119
	<i>nad4***</i>	1488	ATG	TGA	496
	<i>nad4L</i>	303	ACG (ATG)	TAA	101
	<i>nad5*****</i>	2010	ATG	TAA	670
	<i>nad6</i>	618	ATG	TAA	206
	<i>nad7****</i>	1185	ATG	TAG	395
	<i>nad9</i>	591	ATG	TAA	197
Ribosomal proteins (LSU)	<i>rpl16</i>	558	ATG	TAA	186

	<i>rpl5</i>	564	ATG	TAA	188
Ribosomal proteins (SSU)	<i>rps10*</i>	411	ATG	TAA	137
	<i>rps12</i>	372	ATG	TGA	124
	<i>rps14</i>	303	ATG	TAG	101
	<i>rps19</i>	138	ATG	TGA	46
	<i>rps3*</i>	1677	ATG	TAA	559
	<i>rps4</i>	1050	ATG	TAA	350
	<i>rps7*</i>	276	ATG	TAA	92
Ribosomal RNAs	<i>rrn18</i>	2008			
	<i>rrn26(2)</i>	(3137,3137)			
	<i>rrn5</i>	115			
Transfer RNAs	<i>trnC-GCA</i>	71			
	<i>trnD-GTC</i>	74			
	<i>trnE-TTC</i>	72			
	<i>trnF-GAA(2)</i>	(64,74)			
	<i>trnG-GCC(2)</i>	(72,72)			
	<i>trnH-GTG</i>	74			
	<i>trnK-TTT</i>	73			
	<i>trnL-CAA</i>	82			
	<i>trnM-CAT(4)</i>	(73,74,74,74)			
	<i>trnN-GTT</i>	72			
	<i>trnP-CGG*</i>	83			
	<i>trnP-TGG</i>	75			
	<i>trnQ-TTG</i>	72			
	<i>trnT-TGT*</i>	75			
	<i>trnW-CCA</i>	74			
	<i>trnY-GTA</i>	83			

Gene*: gene with one intron; Gene**: gene with two introns; Gene (2): copy number of multi-copy gene.

In protein-coding genes, the most used start codon is ATG and the most used stop codon is TAA. The 21 tRNAs involve 15 amino acids in the transport process, including: methionine (Met), lysine (Lys), glutamate (Glu), phenylalanine (Phe), proline (Pro), tryptophan (Trp), glutamine (Gln) glycine (Gly), aspartate (Asp), threonine (Thr), tyrosine (Tyr), asparagine (Asn), cysteine (Cys), histidine (His), and leucine (Leu). The difference between the number of tRNAs and amino acids indicates the existence of one amino acid being transported by multiple tRNAs.

Prediction Of Rna Editing Sites

RNA editing affects gene expression and RNA stability through base substitution, insertion or deletion and plays an important role in promoting transcriptional diversity and enriching the variety of proteins[14, 15]. RNA editing sites were predicted for the mt genome of *T. foenum-graecum*, and a total of 465 RNA editing sites were predicted in 33 PCGs, and all RNA editing sites were of C-T editing type. The relationship between the number of genes and editing sites is shown in Fig. 2. ATP synthase (except atp4), Transport membrane protein, Maturases, Ribosomal proteins (except rpl5) and Ribosomal proteins (except rps4, rps3) were found to have a relatively low number of RNA editing-derived substitutions (1–10 editing sites), while Cytochrome c biogenesis, Ubichinol cytochrome c reductase, Cytochrome c oxidase, and NADH dehydrogenase (except nad9) were significantly edited (10–41 editing sites). Among them, nad4 had the highest number of RNA editing sites.

The RNA editing sites were classified according to the hydrophilicity of amino acids, as shown in Table 2. It includes five types of edits: hydrophilic-hydrophilic, hydrophobic-hydrophobic, hydrophilic-hydrophobic, hydrophobic-hydrophilic and hydrophilic-stop. Among them, 13.12% of the amino acids remained hydrophilic; 31.83% of the amino acids remained hydrophobic; 47.53% of the amino acids changed from hydrophilic to hydrophobic; 6.45% of the amino acids changed from hydrophobic to hydrophilic; and 1.08% of the amino acids were prematurely terminated during the coding process. Premature termination occurred in atp6, ccmFc, and cox1 in the *T. foenum-graecum* mt genome. In addition, a total of 32 codon transitions were involved, with TCA (S) = > TTA (L) being the most common, with 68 editing sites.

Table 2 Classification table of RNA editing sites

Type	RNA-editing	Number	Percentage
hydrophilic-hydrophilic	CAC (H)=>TAC (Y)	9	13.12%
	CAT (H)=>TAT (Y)	14	
	CGC (R)=>TGC (C)	11	
	CGT (R)=>TGT (C)	27	
	total	61	
hydrophobic-hydrophobic	CCA (P)=>CTA (L)	39	31.83%
	CCC (P)=>CTC (L)	12	
	CCC (P)=>TTC (F)	6	
	CCG (P)=>CTG (L)	28	
	CCT (P)=>CTT (L)	23	
	CCT (P)=>TTT (F)	10	
	CTC (L)=>TTC (F)	7	
	CTT (L)=>TTT (F)	14	
	GCA (A)=>GTA (V)	1	
	GCC (A)=>GTC (V)	1	
	GCG (A)=>GTG (V)	4	
	GCT (A)=>GTT (V)	3	
	total	148	
hydrophilic-hydrophobic	ACA (T)=>ATA (I)	5	47.53%
	ACC (T)=>ATC (I)	1	
	ACG (T)=>ATG (M)	7	
	ACT (T)=>ATT (I)	3	
	CGG (R)=>TGG (W)	34	
	TCA (S)=>TTA (L)	68	
	TCC (S)=>TTC (F)	25	
	TCG (S)=>TTG (L)	40	
	TCT (S)=>TTT (F)	38	
	total	221	
hydrophobic-hydrophilic	CCA (P)=>TCA (S)	4	6.45%

	CCC (P)=>TCC (S)	7	
	CCG (P)=>TCG (S)	3	
	CCT (P)=>TCT (S)	16	
	total	30	
hydrophilic-stop	CAA (Q)=>TAA (X)	1	1.08%
	CAG (Q)=>TAG (X)	2	
	CGA (R)=>TGA (X)	2	
	total	5	

Discussion in terms of amino acid conversion revealed that 151 (32.47%) of these editing sites were located on the first base of the triplet codon and 298 (64.09%) on the second base of the triplet codon. In addition, the first and second bases of one codon were edited and the amino acid changed from the original proline (CCT) to phenylalanine (TTT). In the study it was also found that the highest number of leucine was present after RNA editing. This includes: 108 sites converted from serine to leucine and 102 sites converted from proline to leucine.

Codon Preference

A study of *T. foenum-graecum* codon preference showed that when a certain codon for which the relative synonymous codon usage (RSCU) > 1, it indicates that the codon was used relatively frequently and had a preferences[16]. Among these codons, a total of 32 codons were biased, and 29 of them ended with A or T, accounting for 90.63% of the codons. In addition, the 96 bases that make up the 32 codons contain 30 A bases and 32 T bases, indicating that codons with preferences use more A/T bases in their composition. Thus, the *T. foenum-graecum* mt genome has a significant AT preferences. When a certain codon for which the RSCU = 1, it indicates that there is no preferences for that codon[16]. In the *T. foenum-graecum* mt genome, tyrosine has no preferences. The schematic diagram of codon preference is shown in Fig. 3.

Repeated Sequences

Dispersed repetitive sequences are repetitive units that are present in a scattered form throughout the genome[17]. A total of 202 dispersed repeat sequences were detected in the *T. foenum-graecum* mt genome, including 108 forward repeats (F) and 94 palindrome repeats (P) of two repeat types, with repeat lengths mostly concentrated between 30–60 (83). The total length of the scattered repetitive sequences was 47506 bp, accounting for 13.75% of the total length of the mt genome. The length of each repeat sequence and the number of repeat types are detailed in Table 3.

Table 3
Distribution of dispersed repeat
sequences

Length	Dispersed type	Number
20–29	P	2
	F	2
30–39	P	16
	F	19
40–49	P	14
	F	9
50–59	P	11
	F	14
60–69	P	2
	F	3
70–79	P	5
	F	11
80–89	P	3
	F	1
90–99	P	9
	F	9
100–199	P	18
	F	31
≥ 200	P	14
	F	9

SSRs are 1–6 bp DNA fragments with the advantages of high variability, covariance and reproducibility, which are resources for establishing polymorphic DNA markers and can be widely used in plant genetic breeding[18–21]. A total of 96 SSRs were detected in the *T. foenum-graecum* mt genome, including 11 monomers, 21 dimers, 10 trimers, 34 tetramers, 16 pentamers and 4 hexamers. Among them, tetramers had the highest number of repeats, accounting for 35.42% of the total SSRs, and hexamers had the lowest number of repeats, accounting for only 4.17% of the total SSRs. Each SSRs is shown in Table 4.

Table 4
Distribution of SSRs

SSR type	Repeats	total
monomer	A/T	10
	C/G	1
dimer	AC/GT	1
	AG/CT	15
	AT/AT	5
trimer	AAC/GTT	1
	AAG/CTT	4
	AAT/ATT	4
	ATC/ATG	1
tetramer	AAAC/GTTT	2
	AAAG/CTTT	8
	AAAT/ATTT	3
	AACC/GGTT	1
	AAGC/CTTG	3
	AAGT/ACTT	3
	AATG/ATTC	3
	AATT/AATT	1
	ACAG/CTGT	1
	ACAT/ATGT	1
	ACCG/CGGT	1
	ACGG/CCGT	2
	ACTG/AGTC	1
	AGCC/CTGG	1
	AGCT/AGCT	1
	AGGC/CCTG	1
	CCCG/CGGG	1
pentamer	AAAAG/CTTTT	1
	AAAAT/ATTTT	1

SSR type	Repeats	total
	AAACC/GGTTT	4
	AAACT/AGTTT	1
	AAATT/AATTT	1
	AACTG/AGTTC	1
	AACTT/AAGTT	2
	AAGAT/ATCTT	1
	AAGCT/AGCTT	1
	AATTC/AATTG	1
	ACACC/GGTGT	1
	ACGGC/CCGTG	1
hexamer	AAACTT/AAGTTT	2
	AAATGG/ATTTCC	1
	AGATAT/ATATCT	1

Tandem repetitive repeat are formed by the tandem arrangement of repetitive DNA units of 1-200 bp and are widely found in eukaryotes and some prokaryotes[22]. A total of 19 tandem repeats were detected in the *T. foenum-graecum* mt genome, with length distributions ranging from 5–57, and 13 tandem repeats had a match rate of > 97%, as shown in Table 5. The distribution of repetitive sequences on the genome is shown in Fig. 4.

Table 5
Distribution of tandem repeat sequences

NO.	Size	Repeat sequence	Percent Matches
1	36	TAACATAGACCCTCTTTACTTACAGTCGAGCTCTAT	98
2	57	ATATGAAGTTCTAATATTATCTGCACTAAGAAGTGATTACGACTTGTTGTAGATGA	89
3	32	GAGAGGTATGAAAGCGATACTCGACTGATAAG	82
4	22	TTCGATGTAATTGATTTGCGCA	100
5	36	AGGGTCTATGTTAATAGAGCTCGACTGTAAGTAAAG	100
6	30	CGGAGGTTGAGGAGGAGTTTCGGGCTGCTG	64
7	16	CTTGTTATTAGTAAAG	100
8	27	TCTGTATCACTTCTTTACTTGGCTTAT	100
9	27	ATTCTCAATCCACGACGACTATTAACG	100
10	25	TTGATGAACAAGAAGGAACGAAGTG	100
11	12	ATTTATAGCAGC	100
12	15	TCTGACGTCCTTCCT	100
13	19	AATTATCTTATCTAAAATA	70
14	19	CACCTGCAGTTTGGTGCAG	88
15	28	TGCAGGCGAATAGAAAGAGCCCGGCACC	100
16	25	GGGTGAGGGATTAATAAACTAGCTC	100
17	5	ATTCA	100
18	9	GAGACTTTTG	90
19	36	CTTTACTTACAGTCGAGCTCTATTAACATAGACCCT	100

Nucleotide Polymorphism

When a gene or gene spacer varies, causing DNA sequence polymorphism. Analysis of nucleotide polymorphism in the mt genome of *T. foenum-graecum* showed that the range of its was 0-0.03891. The corresponding nucleotide polymorphism values for rps12, rps3, rpl5, cox2, and atp6 were 0.01174, 0.01288, 0.01314, 0.02692, and 0.03891, respectively. Their higher nucleotide polymorphism indicates that these genes or gene spacers have undergone higher variation. Nucleotide polymorphism values of each gene are shown in Fig. 5.

Synteny And Phylogenetic Analysis

T. foenum-graecum and five other Leguminous species (*Trifolium pratense*, *Trifolium meduseum*, *Trifolium grandiflorum*, *Trifolium aureum*, *Medicago truncatula*) were subjected to synteny analysis to tentatively determine their affinities. The results showed that *T. foenum-graecum* was the most similar to *Medicago truncatula*. Schematic diagrams of the covariance and mt structures of these six plants are shown in Fig. 6 and Fig. 7. Among them, *Trifolium meduseum* had the largest length of 348,724 bp and *Medicago truncatula* had the smallest length of 271,618 bp. They all had a GC content of about 45%, further indicating that the plant mt genomes is relatively conserved.

T. foenum-graecum and 25 other Leguminosae species were subjected to phylogenetic analysis. In the comparison between *T. foenum-graecum* and other Papilionoideae plants, *T. foenum-graecum* (*Trigonella*) was first linked to *Medicago truncatula* (Medicago) in a group with a maximum similarity of 100%. In a group connected with *Trifolium pratense*, *Trifolium meduseum*, *Trifolium grandiflorum*, and *Trifolium aureum* (Trifolium), the similarity was high at 93%. Caesalpinioideae, Cercidoideae and Detarioideae were compared as outgroups of the phylogenetic tree. The phylogenetic tree is shown in Fig. 8. There are 24 nodes in the phylogenetic tree, 18 of which have 100% support and 22 of which have more than 80% support.

Substitution Rates Of PcgS

The six Leguminosae plants (*T. foenum-graecum*, *Trifolium pratense*, *Trifolium meduseum*, *Trifolium grandiflorum*, *Trifolium aureum*, *Medicago truncatula*) were compared two by two to analyze Ka/Ks values between species, as shown in Fig. 9. Among the 28 PCGs counted, 23 genes (atp1, atp4, atp6, atp8, ccmB, ccmC, ccmFC, ccmFn, ccmFn2, cob, cox1, cox2, cox3, mttB, nad1, nad2, nad4, nad4L, nad5, nad6, rpl16, rps10, rps14) had Ka/Ks < 1. When Ka/Ks < 1, it indicates that these genes will continue to evolve under purifying selection; when Ka/Ks > 1, it indicates that positive selection of genes has occurred and proteins have been changed; when Ka/Ks = 1, it indicates that there is neutral selection[23].

Chloroplast And Mitochondrial Homologous Sequences

Annotation of the *T. foenum-graecum* chloroplast genome using the same leaf. Homology analysis of the mitochondria and chloroplasts of *T. foenum-graecum* observed a transfer of DNA sequences from the cp genome to the mt genome. *T. foenum-graecum* mitochondria contain 23 cp insertions, ranging from 35 to 2427 bp in length, for a total length of 10,023 bp, or 2.9% of the total genome length, as shown in Table 6. Annotation of these homologous sequences revealed that some genes were lost during the migration of PCGs from chloroplasts to mitochondria, and only some sequences could be found in mitochondria. However, tRNA genes are able to retain their integrity during transfer to mitochondria, such as: trnW-CCA, trnN-GUU, trnD-GUC, trnH-GUG, trnM-CAU. Therefore, it is inferred that tRNA genes are more conserved and have better gene integrity than PCGs during migration. The analysis of homologous fragments of cp and mt sequences is shown in Fig. 10.

Table 6
Cp insertions in the mt genome of *T. foenum-graecum*

	Identity%	Length	Mismatches	Gap openings	gene
1	100	2427	0	0	<i>rrn4.5</i> (partical:6.73%) <i>rrn23</i> (partical:82.89%)
2	99.916	1188	1	0	<i>psbC</i> (partical:83.54%)
3	100	1140	0	0	<i>psaB</i> (partical:51.70%)
4	100	1016	0	0	<i>psbC</i> (partical:10.34%) <i>psbD</i> (partical:86.82%)
5	99.741	772	1	1	<i>rrn23</i> (partical:11.68%) <i>trnA-UGC</i> (partical:30.86%)
6	100	426	0	0	<i>trnI-GAU</i> (partical:55.72%)
7	92.708	384	4	5	<i>rrn23</i> (partical:13.68%)
8	99.288	281	2	0	<i>trnI-GAU</i> (partical:34.17%)
9	98.252	286	4	1	<i>trnW-CCA</i> ; <i>petG</i> (partical:10.53%)
10	74.972	887	172	38	<i>rrn16</i> (partical:57.98%)
11	87.547	265	30	3	<i>rrn16</i> (partical:17.69%)
12	99.167	120	1	0	<i>psaA</i> (partical:5.27%)
13	91.27	126	11	0	<i>psaA</i> (partical:5.53%)
14	98.824	85	0	1	<i>trnN-GUU</i>
15	96.429	84	3	0	<i>trnD-GUC</i>
16	96.154	78	3	0	<i>trnH-GUG</i>
17	93.59	78	5	0	<i>trnM-CAU</i>
18	98.077	52	1	0	<i>rrn16</i> (partical:3.49%)
19	96.296	54	0	2	<i>ycf2</i> (partical:0.85%)
20	80.412	97	19	0	<i>rrn23</i> (partical:3.47%)
21	80.412	97	19	0	<i>rrn23</i> (partical:3.47%)
22	95.556	45	2	0	<i>rrn23</i> (partical:1.61%)
23	97.143	35	1	0	<i>rrn23</i> (partical:1.25%)

Discussion

Mitochondria are double-membrane organelles commonly found in eukaryotes and play an important role in life activities. Plant mt genomes exhibit complex and relatively conserved properties^[10], and this structural feature creates conditions for occurring genomic rearrangements. In recent years, with the continuous development of sequencing technology, the plant mt genomes has been studied more deeply.

Plant mt genomes are conformationally diverse, with co-existing molecular forms in addition to single conformations, but most of the mt genome maps sequenced so far exist as ring-shaped molecules[24]. The *T. foenum-graecum* mt genome is 345,604 bp in length with 59 genes, 11 of which have introns. In addition, some of the genes in NADH dehydrogenase contain multiple introns, such as nad1, nad2, nad3, nad5, nad7 containing four introns and nad4 containing three introns. The presence of introns in the plant mt genome, which is one of the characteristics of higher plant mt genes, can be distinguished from other species[25]. GC content is an important factor in the evaluation of species[26]. *T. foenum-graecum* mt genome GC content was 45.28%, which was similar to *Trifolium pratense* (NC_048499.1) 45.20%, *Trifolium meduseum* (NC_048500.1) 44.99%, *Trifolium grandiflorum* (NC_048501.1) 45.09%, *Trifolium aureum* (NC_048502.1) 44.88%, *Medicago truncatula* (NC_029641.1) 45.39% were similar.

It has been suggested that the origin of the RNA editing sites was to repair mutations produced by themselves and UV irradiation during the evolution of plants[27–29]. There are significant differences in the role of RNA editing in the coding and non-coding regions of genes. RNA editing on the coding region of a gene often occurs in the first 2 bases of the codon, which can change the hydrophilicity and hydrophobicity of amino acids and ultimately affect the function of the protein[30, 31]. And RNA editing on non-coding regions plays an important role in mRNA splicing[32]. A total of 465 RNA editing sites were predicted in the 33 PCGs of the *T. foenum-graecum* mt genome, and all RNA editing sites were of the C-T editing type. The C-T editing type is the most common type of editing in plant mt genomes[33, 34], and the results of the study are the same as those previously reported.

Codon preference refers to favor to use one or more fixed codons in a given species or gene[35, 36]. When RSCU > 1, it indicates that the codon is used more frequently than other synonymous codons, which means that the codon generates bias; when RSCU = 1, both are used with the same frequency and the codon is unbiased[16]. During gene encoding, codons with RSCU < 1 should be avoided. A total of 32 codons were found to be biased in the *T. foenum-graecum* mt genome and these codons used a higher number of A/T bases.

Repeated sequences are nucleic acid sequences that occur in multiple copies in the genome, are an important part of eukaryotic genomes, and play an important role in genetic evolution of the genome[37–39]. The repetitive sequences of *T. foenum-graecum* mt genome were analyzed, and a total of 202 dispersed repeats were detected with a maximum length of 5133 bp for forward repeats and 5503 bp for palindrome repeats. A total of 96 SSRs were detected, with the largest number of tetrameric repeats and each repeat sequence consisting mainly of A and T bases. A total of 19 tandem repeats were detected, with a maximum repeat length of 57 bp, of which 12 tandem repeats had a 100% match rate.

Nucleotide polymorphism revealed the magnitude of variation in nucleic acid sequences of different species. In the *T. foenum-graecum* mt genome, rps12, rps3, rpl5, cox2, and atp6, the nucleotide polymorphism values of these five genes were relatively high, all greater than 0.01, indicating a higher degree of variability. The

regions where these genes are located could provide potential molecular markers for Leguminosae plant genetics.

The synteny analysis revealed that *T. foenum-graecum* was very similar to the other five Leguminosae plants (*Trifolium pratense*, *Trifolium meduseum*, *Trifolium grandiflorum*, *Trifolium aureum*, *Medicago truncatula*), and it was inferred that gene rearrangement might have occurred in the mitochondria of Leguminosae species. They differed greatly in size from each other, but all had a GC content of about 45%. Further indicating the complex yet relatively conserved characteristics that plant mitochondria have exhibited during their evolution[10]. In this study, the phylogenetic relationships of *T. foenum-graecum* were analyzed, and the results showed that *T. foenum-graecum* had the highest similarity to *Medicago truncatula*, which was consistent with the results of the synteny analysis.

The ratio of Ka/Ks can determine the type of selection on genes and is important for reconstructing phylogenies and understanding the evolutionary dynamics of protein-coding sequences in closely related species[40]. In this study, 23 genes had Ka/Ks < 1, indicating that the genes are well conserved and will continue to evolve under purifying selective pressure.

Migration of DNA sequences is frequently observed in the mt genome of plants[41]. Length and sequence similarity of migrating fragments vary by species[42]. The *T. foenum-graecum* mt genome contains 23 homologous segments with the cp genome and the mt genome, and most of the genes have only partially migrated and lost their integrity. The tRNA genes are more conserved than the PCGs and rRNA genes[42]. In the present study, tRNA genes were able to retain their integrity during transfer and their migration results were consistent with those reported in the literature.

Conclusions

In this paper, the *T. foenum-graecum* mt genome was sequenced, assembled and annotated, and the annotation results were analyzed. The *T. foenum-graecum* mt genome was 345,604 bp in length with 45.28% GC content. There were 59 genes, including 33 protein-coding genes, 21 tRNA genes, 4 rRNA genes, and 1 pseudo gene. Specific analyses of RNA editing sites, codon bias, three repeat types, nucleotide polymorphism, cp and mt homologous sequences were also performed. Synteny and phylogenetic analysis according to *T. foenum-graecum* relatives revealed that *T. foenum-graecum* had the highest similarity to *Medicago truncatula*. The size and GC content of the mt genomes of the six closest related Leguminosae plants were compared, and the GC content of *T. foenum-graecum* was found to be more conserved during the evolutionary process. Ka/Ks analysis revealed that most PCGs would continue to evolve under purifying selection pressure. In summary, a comprehensive analysis of the *T. foenum-graecum* mt genome was conducted, which laid the foundation for further in-depth studies of Leguminosae *Trigonella*.

Materials And Methods

Plant materials and DNA sequencing

T. foenum-graecum was grown at the medicinal herb planting base of the College of Pharmacy, Qinghai Minzu University (Xining, Qinghai, China). After growing into seedlings, they were scrubbed with 70% alcohol to remove dust and soil from the surface of the fenugreek, frozen in liquid nitrogen, and placed in pre-chilled 50 ml sealed bags. DNA of *T. foenum-graecum* was extracted from the samples and sequenced using Illumina, Oxford Nanopore PromethION for second and third generation sequencing, respectively. This sequencing is technically supported by GENEPIIONEER (Nanjing, China). Using fastp v0.20.0 (<https://github.com/OpenGene/fastp>, Accessed 30 October 2022) software, the raw data of second-generation sequencing was filtered to obtain high quality reads. The third-generation sequencing data was filtered using Filtlong v0.2.1 (<https://github.com/rrwick/Filtlong>, Accessed 30 October 2022) software.

Assembly And Annotation Of The Mt Genome

The sequences with higher quality (more complete core genes covered) were selected as seed sequence using Minimap2 v2.1 [43] software, and the original triple sequencing data were compared to the seed sequence to obtain all triple sequencing data of the mt genome. The resulting tri-generation data are then corrected using the tri-generation assembly software Canu v2.2[44], the second-generation data are compared to the corrected sequence using Bowtie2 v2.3.5.1[45], and then the second-generation data on the pair and the corrected tri-generation data are stitched together using the default parameters of Unicycler v0.4.8 (<https://github.com/rrwick/Unicycler>, Accessed 30 October 2022). Due to the complex physical structure of the mt genome, at this point, the corrected triple sequencing data were compared to the contig obtained in the second step of Unicycler v0.4.8 (<https://github.com/rrwick/Unicycler>, Accessed 30 October 2022) using Minimap2 v2.1 [43] to manually determine the branching direction to obtain the final assembly results.

The encoded proteins and rRNAs were compared to published and used as ref plant mt sequences using BLAST v2.6 (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>, Accessed 30 October 2022). The tRNA was annotated using tRNAscanSE v2.0 [46] (<http://lowelab.ucsc.edu/tRNAscan-SE/>, Accessed 30 October 2022). ORFs were annotated using Open Reading Frame Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>, Accessed 30 October 2022). Use OGDRAW [47] (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>, Accessed 30 October 2022) to create a mt genome map.

Analysis of RNA editing sites and codon bias

RNA editing sites were analyzed using the Plant Predictive RNA Editor (PREP) suite[48]. Screening of unique CDS and calculation of codon preference using a self-encoded Perl script.

Analysis Of Repeat Sequences

SSRs were identified using MISA v1.0 (<http://pgrc.ipk-gatersleben.de/misa/misa.html>, Accessed 23 November 2022) [49] software, tandem repeat sequences was identified using TRF v4.09 (<https://github.com/Benson-Genomics-Lab/TRF>, Accessed 23 November 2022) software, and dispersed repeat sequences was identified using BLASTN v2.10.1 (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome, Accessed 23 November 2022) software.

Use Circos v0.69-5 [50] (<http://circos.ca/software/download/>, Accessed 23 November 2022) to visualize and analyze the repeated sequences.

Nucleotide Polymorphism Analysis

Homologous gene sequences of different species were compared globally using MAFFT v7 [51] software (-auto mode) and nucleotide polymorphism values were calculated for each gene using DnaSP v5[52].

Synteny Analysis

Using the BLASTN v2.10.1 (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome, Accessed 1 November 2022) software, fragments with a comparative length greater than 300 bp were screened, and the assembled species and the selected species were compared sequentially to plot the covariance.

Comparative Analysis Of The Mt Genome Structure

Comparative analysis of mt genome structure for closely related species using the software CGView [53] default parameters.

Phylogenetic Analysis

CDS was used to make the maximum likelihood evolutionary tree, and the sequences between species were compared by MAFFT v7 [51] software, and the sequences that were compared were concatenated first and last, trimmed by trimAl (v1.4.rev15) (parameter: -gt 0.7), and then the RAxML v8 [54] software was used to construct the maximum likelihood evolutionary tree by choosing GTRGAMMA model, rapid Bootstrap analysis, bootstrap = 1000 likelihood evolutionary tree.

Kaks Values Analysis

KaKs values of genes were calculated using KaKs_Calculator v2.0 [55] software, and MLWL was chosen as the calculation method.

Homologous Sequence Analysis

Homologous sequences between chloroplasts and mitochondria were found using BLAST v2.6 software, setting the similarity to 70% and the E-value to 10E-5. Mapping of cp and mt sequence homologous fragments using Circos v0.69-5 [50] (<http://circos.ca/software/download/>, Accessed 1 November 2022).

Abbreviations

T. foenum-graecum

Trigonella foenum-graecum L

mt

Mitochondria

cp

Chloroplast

PCG

Protein-coding genes

tRNA

Transfer RNA

rRNA

Ribosomal RNA

SSRs

Simple sequence repeat

Ka/Ks

nonsynonymous-to-synonymous substitution ratio

Met

Methionine

Lys

Lysine

Glu

Glutamate

Phe

Phenylalanine

Pro

Proline

Trp

Tryptophan

Gln

Glutamine

Gly

Glycine

Asp

Aspartate

Thr

threonine

Tyr

Tyrosine

Asn

Asparagine

Cys

Cysteine

His

Histidine

Leu

Leucine

RSCU

Relative synonymous codon usage.

Declarations

Acknowledgments

Not applicable.

Sample storage location

The sample was collected from the traditional Chinese medicinal planting base, College of Pharmacy, Qinghai Minzu University,, identified by Lu Yongchang. A voucher specimen was deposited at Medicinal Herbarium, College of Pharmacy, Qinghai Minzu University (Yongchang Lu, qhlych@126. com) under the voucher number 20200428.

License statement

This research was carried out within a legal scope and did not violate local laws and ethics. Our samples do not require ethical approval.

Authors' contributions

YFH conceived and designed the research. YFH and WYL performed the experiments and wrote the paper. JLW helped with a critical discussion on the work revised the paper. The author(s) read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China [81960785], the Applied Basic Research Project of Qinghai Province [2020-ZJ-717].

Availability of data and materials

The sequence and annotation of *T. foenum-graecum* mt and cp genome was submitted to the NCBI. The accession numbers of mt and cp in Gene Banks are OP605625 (<https://www.ncbi.nlm.nih.gov/nuccore/OP605625>) and OP747310 (<https://www.ncbi.nlm.nih.gov/nuccore/OP747310>), respectively.

Ethics approval and consent to participate

The study is conducted with plant material complies with relevant institutional, national, and international guidelines and legislation. Also, the study did not use any endangered or protected species. *T. foenum-graecum* is widely cultivated throughout China. The *T. foenum-graecum* plants used in this experiment were grown at the medicinal herb planting base of the College of Pharmacy, Qinghai Minzu University.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Altuntaş E, Özgöz E, Taşer ÖF. Some physical properties of fenugreek (*Trigonella foenum-graecum* L.) seeds. J Food Eng. 2005;71(1):37–43.
2. Murlidhar M, Goswami TK. A review on the functional properties, nutritional content, medicinal utilization and potential application of fenugreek. J Food Process Tech. 2012;3(9):1–10.
3. Khan F, Negi K, Kumar T. Effect of sprouted fenugreek seeds on various diseases: a review. J Diabetes Metab Disord Control. 2018;5(4):119–25.
4. Chaubey PS, Somani G, Kanchan D, Sathaye S, Varakumar S, Singhal RS. Evaluation of debittered and germinated fenugreek (*Trigonella foenum graecum* L.) seed flour on the chemical characteristics, biological activities, and sensory profile of fortified bread. J Food Process Pres. 2018;42(1):1–11.
5. Mukherjee I, Ghosh M, Meinecke M. MICOS and the mitochondrial inner membrane morphology – when things get out of shape. FEBS Lett. 2021;595(8):1159–83.
6. Mclean JR, Cohn GL, Brandt IK, Simpson MV. Incorporation of labeled amino acids into the protein of muscle and liver mitochondria. J Biol Chem. 1958;233(3):657–63.
7. Sedlackova L, Korolchuk VI. Mitochondrial quality control as a key determinant of cell survival. BBA-Mol Cell Res. 2019;1866(4):575–87.
8. Amorim JA, Coppotelli G, Rolo AP, Palmeira CM, Ross JM, Sinclair DA. Mitochondrial and metabolic dysfunction in ageing and age-related diseases. Nat Rev Endocrinol. 2022;18(4):243–58.
9. Yu X, Duan Z, Wang Y, Zhang Q, Li W. Sequence Analysis of the Complete Mitochondrial Genome of a Medicinal Plant, *Vitex rotundifolia* Linnaeus f. (Lamiales: Lamiaceae). Genes. 2022;13(5):1–13.
10. Kubo T, Newton KJ. Angiosperm mitochondrial genomes and mutations. Mitochondrion. 2008;8(1):5–14.
11. Sloan DB. One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. New Phytol. 2013;200(4):978–85.
12. Wilson RK, Hanson MR. Preferential RNA editing at specific sites within transcripts of two plant mitochondrial genes does not depend on transcriptional context or nuclear genotype. Curr Genet. 1996;30(6):502–8.

13. Ogiwara Y, Yamazaki Y, Murai K, Kanno A, Terachi T, Shiina T, et al. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. *Nucleic Acids Res.* 2005;33(19):6235–50.
14. Alon S, Garrett SC, Levanon EY, Olson S, Graveley BR, Rosenthal JJ, et al. The majority of transcripts in the squid nervous system are extensively recoded by A-to-I RNA editing. *eLife.* 2015;4:e05198.
15. Funkhouser SA, Steibel JP, Bates RO, Raney NE, Schenk D, Ernst CW. Evidence for transcriptome-wide RNA editing among *Sus scrofa* PRE-1 SINE elements. *BMC Genomics.* 2017;18(1):1–9.
16. Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* 1986;14(13):5125–43.
17. Qiao Y, Zhang X, Li Z, Song Y, Sun Z. Assembly and comparative analysis of the complete mitochondrial genome of *Bupleurum chinense* DC. *BMC Genomics.* 2022;23(1):1–17.
18. Li Q, Su X, Ma H, Du K, Yang M, Chen B, et al. Development of genic SSR marker resources from RNA-seq data in *Camellia japonica* and their application in the genus *Camellia*. *Sci Rep.* 2021;11(1):1–11.
19. Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 1989;17(16):6463–71.
20. Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 2005;23(1):48–55.
21. Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK. Microsatellite markers: an overview of the recent progress in plants. *Euphytica.* 2011;177(3):309–34.
22. Paço A, Freitas R, Vieira-da-Silva A. Conversion of DNA sequences: From a transposable element to a tandem repeat or to a gene. *Genes.* 2019;10(12):1014.
23. Zhang Z, Li J, Zhao X, Wang J, Gane Wong KJ, Yu J. KaKs_Calculator: Calculating Ka and Ks through model selection and model averaging. *Genom Proteom Bioinf.* 2006;4(4):259–63.
24. Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T. The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNA^{Cys} (GCA). *Nucleic Acids Res.* 2000;28(13):2571–6.
25. Dombrovskaya O, Qiu Y. Distribution of introns in the mitochondrial gene *nad1* in land plants: phylogenetic and molecular evolutionary implications. *Mol Phylogenet Evol.* 2004;32(1):246–63.
26. Ma Q, Wang Y, Li S, Wen J, Zhu L, Yan K, et al. Assembly and comparative analysis of the first complete mitochondrial genome of *Acer truncatum* Bunge: a woody oil-tree species producing nervonic acid. *BMC Plant Biol.* 2022;22(1):1–17.
27. Maier UG, Bozarth A, Funk HT, Zauner S, Rensing SA, Schmitz-Linneweber C, et al. Complex chloroplast RNA metabolism: just debugging the genetic programme? *BMC Biol.* 2008;6(1):1–9.
28. Takahashi A, Ohnishi T. The significance of the Study about the biological effects of solar ultraviolet radiation using the exposed facility on the international space station. *Bio Sci Space.* 2004;18(4):255–60.
29. Fujii S, Small I. The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytol.* 2011;191(1):37–47.

30. Wu B, Chen H, Shao J, Zhang H, Wu K, Liu C. Identification of symmetrical RNA editing events in the mitochondria of *salvia miltiorrhiza* by strand-specific RNA sequencing. *Sci Rep.* 2017;7(1):1–11.
31. Wang M, Liu H, Ge L, Xing G, Wang M, Song W, et al. Identification and analysis of RNA editing sites in the chloroplast transcripts of *Aegilops tauschii* L. *Genes.* 2016;8(1):13.
32. Guo W, Grewe F, Mower JP. Variable frequency of plastid RNA editing among ferns and repeated loss of uridine-to-cytidine editing from vascular plants. *PLoS ONE.* 2015;10(1):e0117075.
33. Edera AA, Sanchez-Puerta MV. Computational detection of plant RNA Editing Events. *Methods Mol Biol.* 2021;218:13–34.
34. Verhage L. Targeted editing of the Arabidopsis mitochondrial genome. *Plant J.* 2020;104(6):1457–8.
35. Grantham R, Gautier C, Gouy M. Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Res.* 1980;8(9):1893–912.
36. Grantham R, Gautier C, Gouy M, Jacobzone M, Mercier R. Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res.* 1981;9(1):213.
37. Klein SJ, O'Neill RJ. Transposable elements: genome innovation, chromosome diversity, and centromere conflict. *Chromosome Res.* 2018;26(1):5–23.
38. Garrido-Ramos MA. Satellite DNA in plants: More than just rubbish. *Cytogenet Genome Res.* 2015;146(2):153–70.
39. Graur D, Zheng Y, Azevedo RBR. An evolutionary classification of genomic function. *Genome Biol Evol.* 2015;7(3):642–5.
40. Fay JC, Wu CI. Sequence divergence, functional constraint, and selection in protein evolution. *Annu Rev Genom Hum G.* 2003;4(1):213–35.
41. Straub SCK, Cronn RC, Edwards C, Fishbein M, Liston A. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (apocynaceae). *Genome Biol Evol.* 2013;5(10):1872–85.
42. Cheng Y, He X, Priyadarshani SVGN, Wang Y, Ye L, Shi C, et al. Assembly and comparative analysis of the complete mitochondrial genome of *Suaeda glauca*. *BMC Genomics.* 2021;22(1):1–15.
43. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34(18):3094–100.
44. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* 2017;27(5):722–36.
45. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357–9.
46. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol.* 2019;162:1–14.
47. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 2019;47(W1):W59–64.
48. Mower JP. The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res.* 2009;37(suppl2):W253–9.
49. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics.* 2017;33(16):2583–5.

50. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009;19(9):1639–45.
51. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80.
52. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009;25(11):1451–2.
53. Stothard P, Wishart DS. Circular genome visualization and exploration using CGView. *Bioinformatics.* 2005;21(4):537–9.
54. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post- analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–3.
55. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteom Bioinf.* 2010;8(1):77–80.

Figures

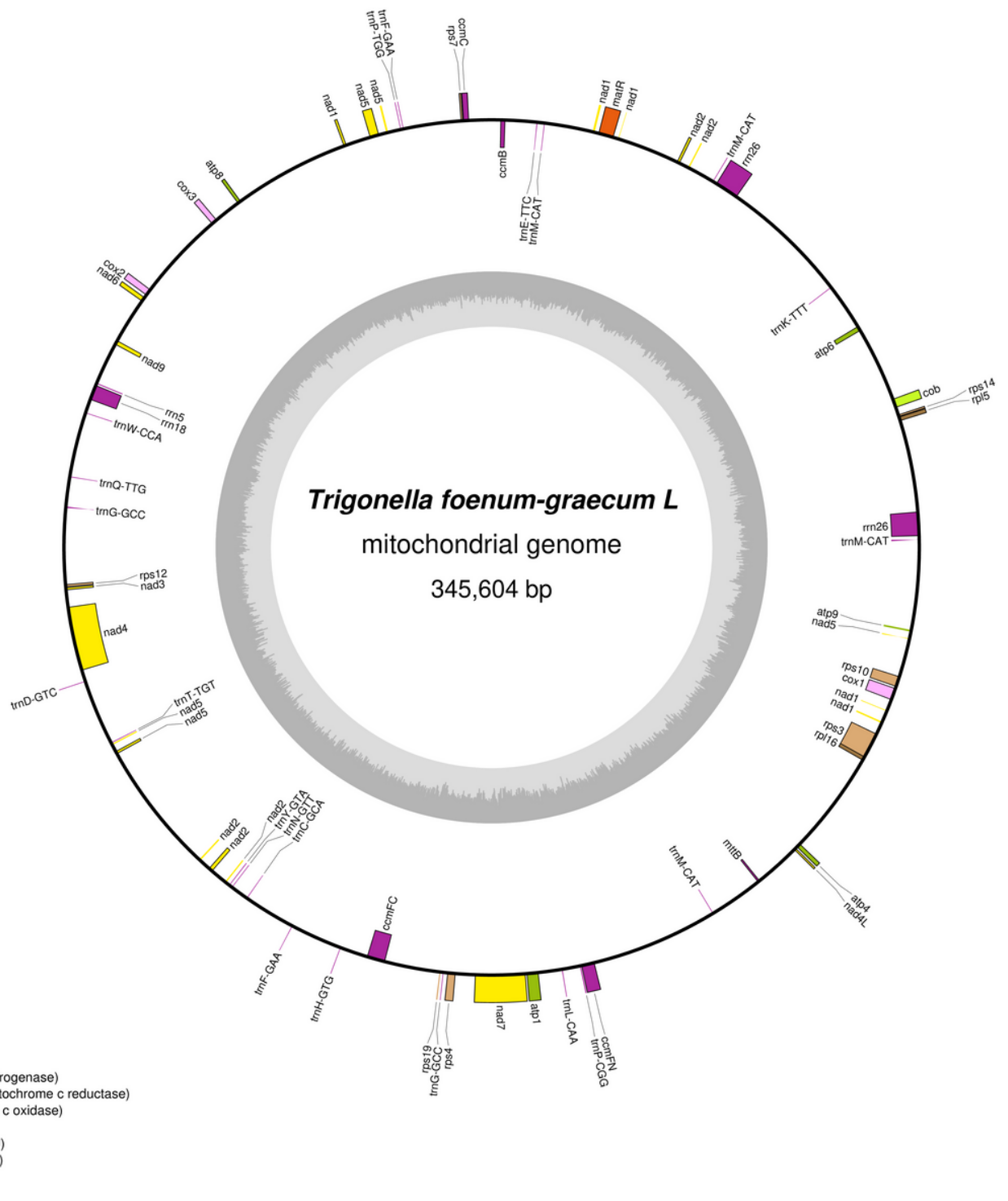


Figure 1

The circular map of the *T. foenum-graecum* mt genome. The gray circles represent the GC content of the mt genome, and different colors represent different gene types

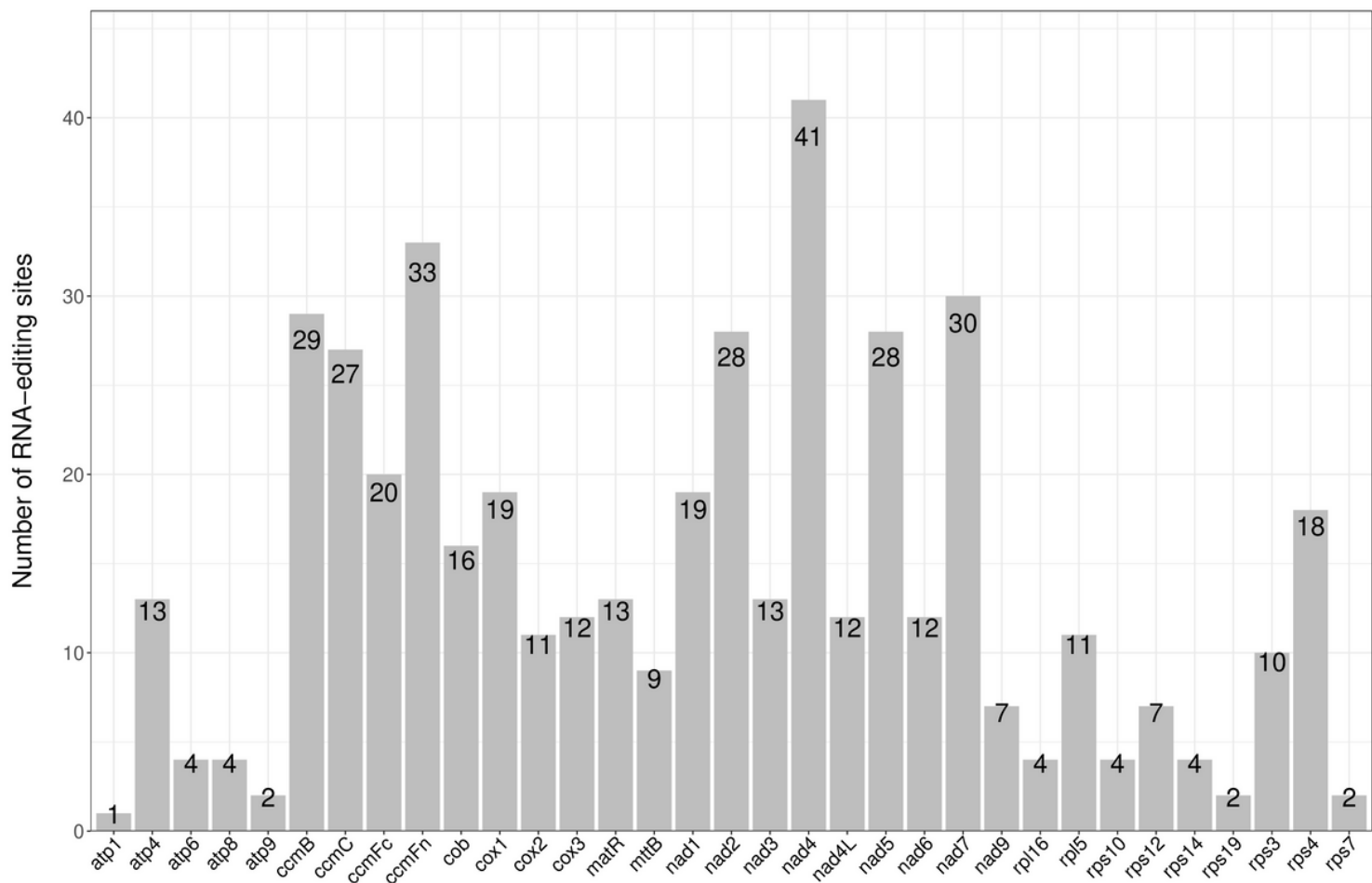


Figure 2

Distribution of RNA editing sites. The horizontal coordinate represents the gene name and the vertical coordinate represents the number of RNA edits

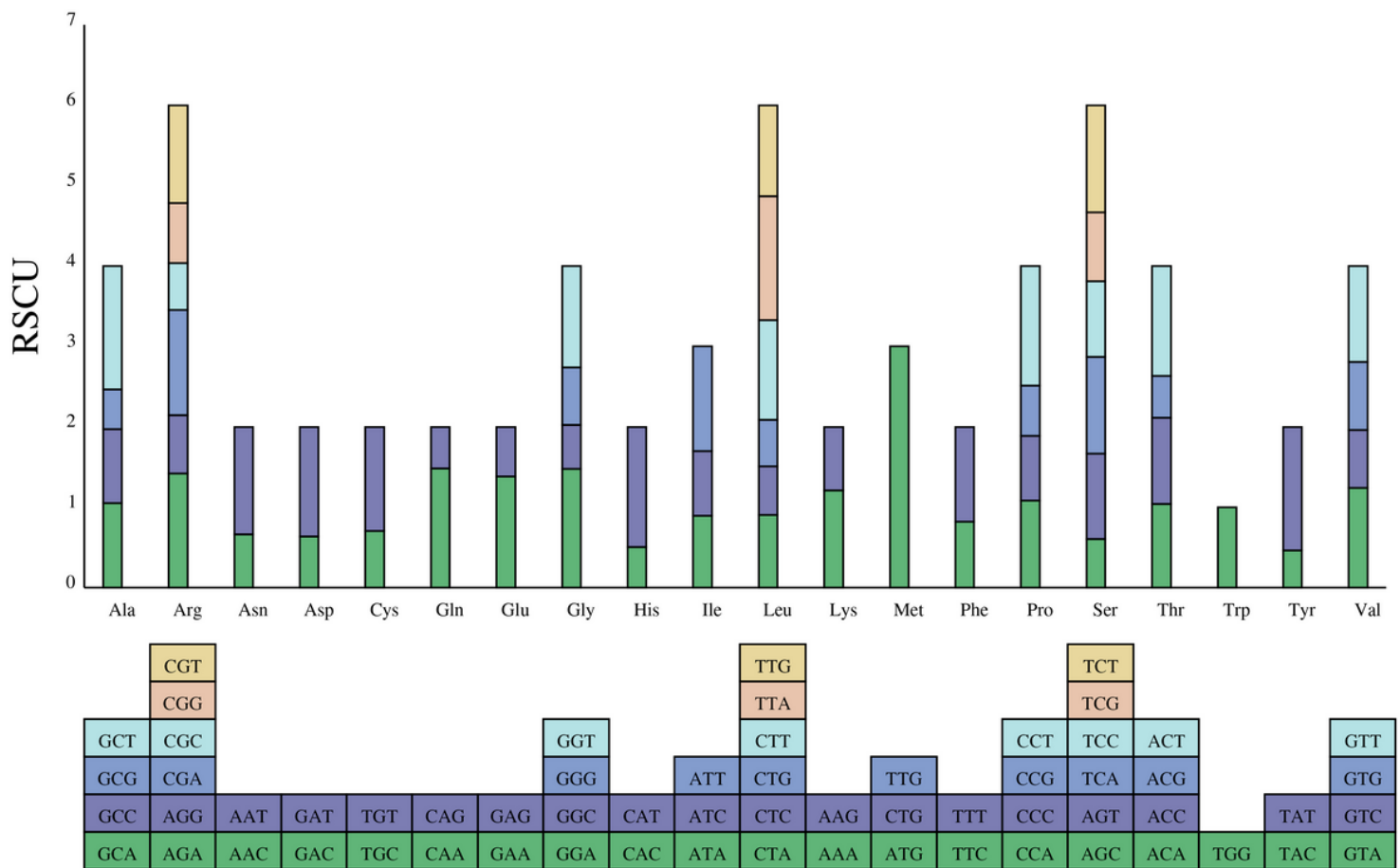


Figure 3

RSCU histogram. The square below represents all codons encoding each amino acid and the height of the bar above represents the sum of all codon RSCU values

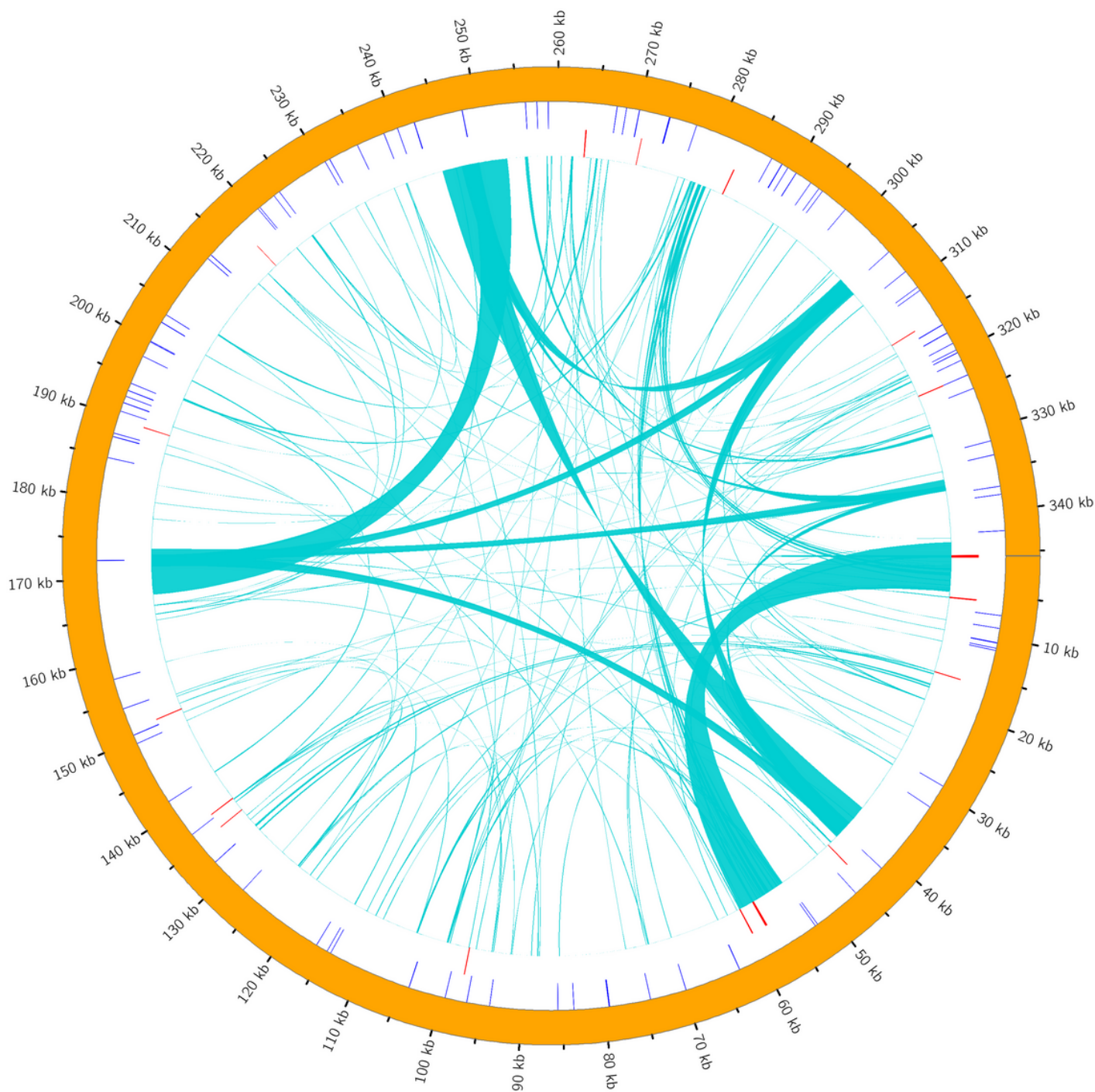


Figure 4

Distribution of repetitive sequences on the genome. The outermost circle is the SSRs, followed by the tandem repeating sequence, and the innermost concatenation is the dispersed repeating sequence

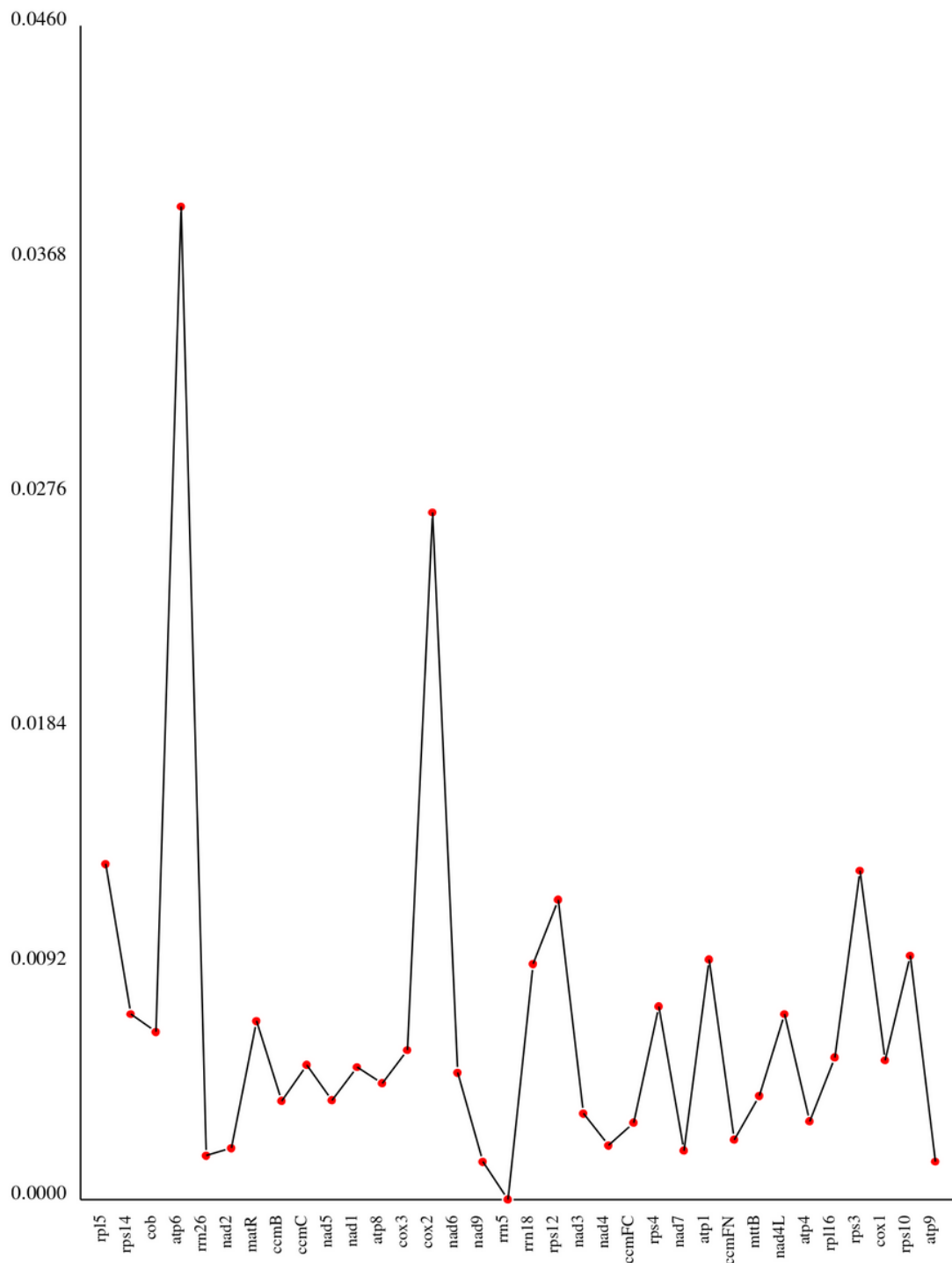


Figure 5

Folding graph of gene nucleotide polymorphism values. The horizontal coordinate indicates the gene name and the vertical coordinate indicates the nucleotide polymorphism value

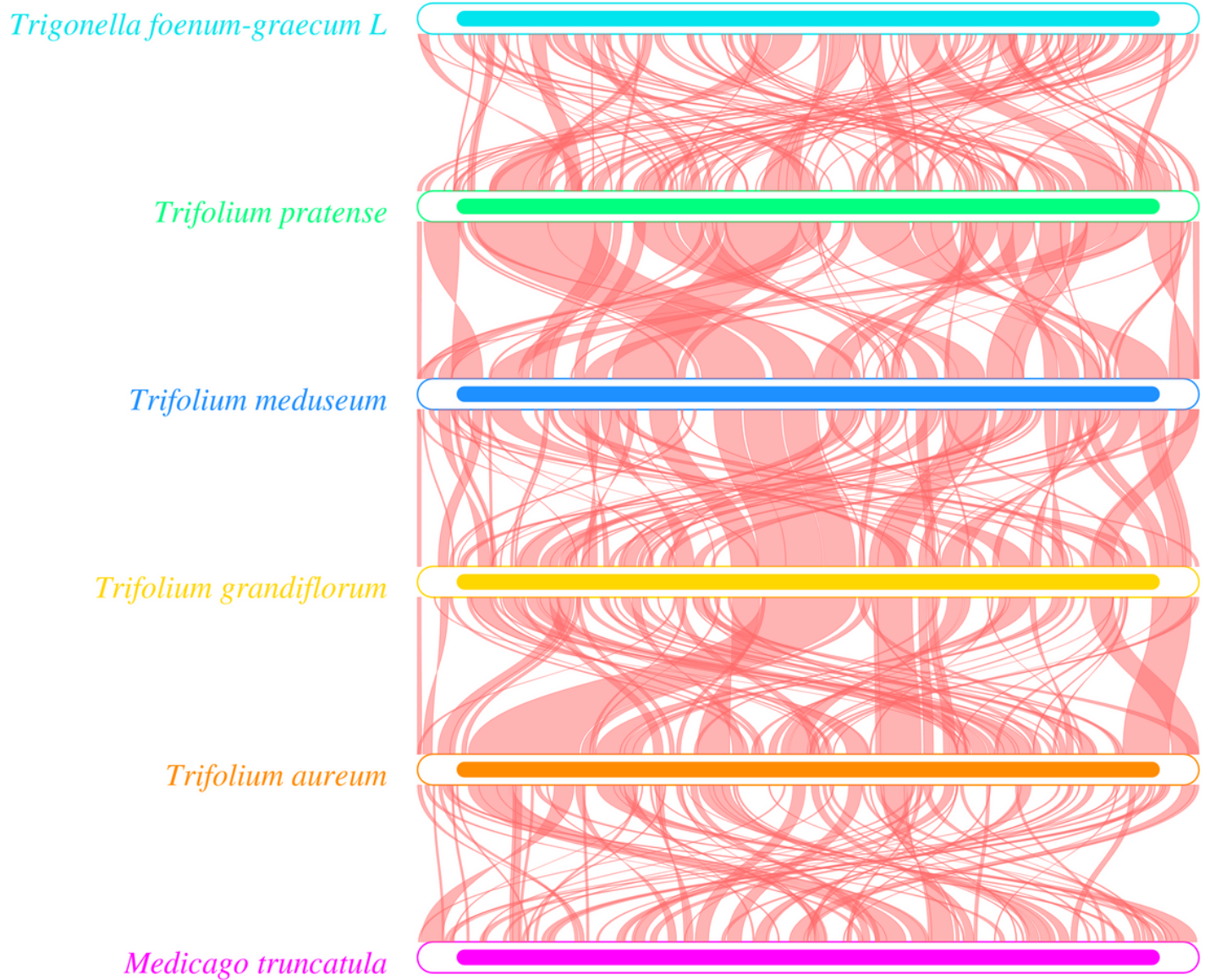


Figure 6

Mt sequence synteny analysis. The boxes in each row indicate a genome, and the connecting lines in the middle indicate homologous regions

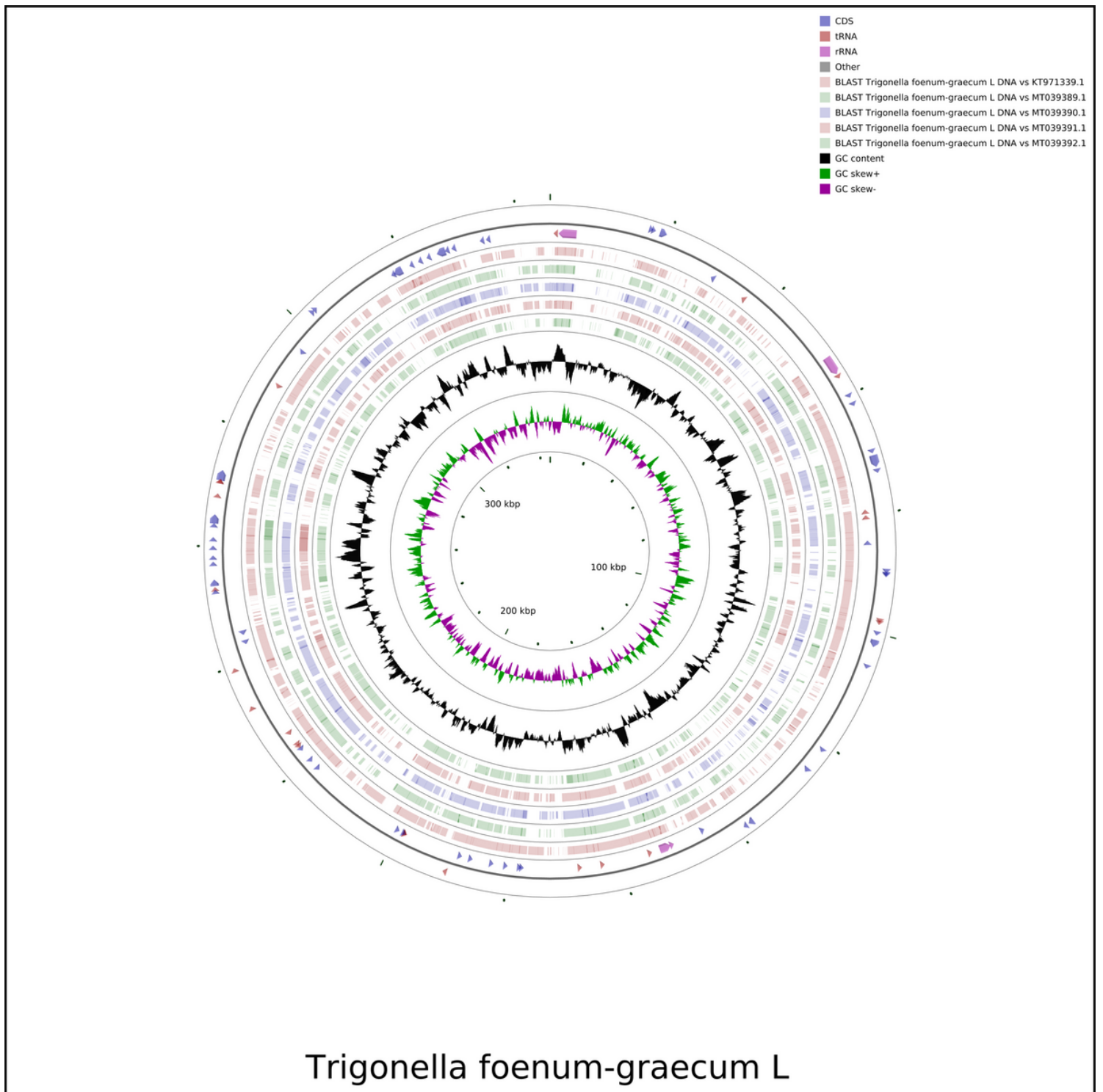


Figure 7

Comparison of mt structures. The two outermost circles depict the gene length and orientation of the genome; the inner circles represent the similarity results with other reference genomes; the black circles represent the GC content

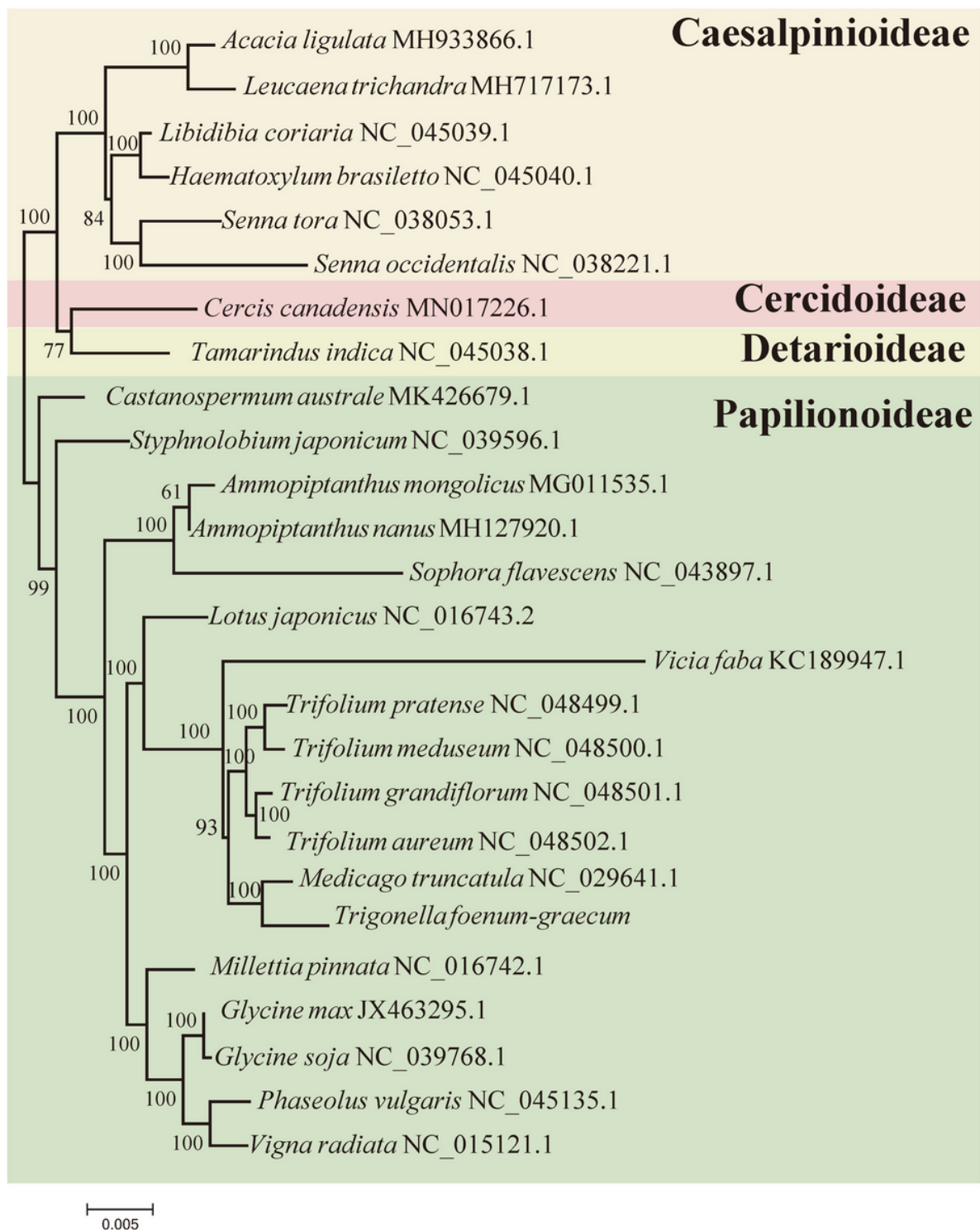


Figure 8

Phylogenetic relationships of *T. foenum-graecum* with 25 other plant species

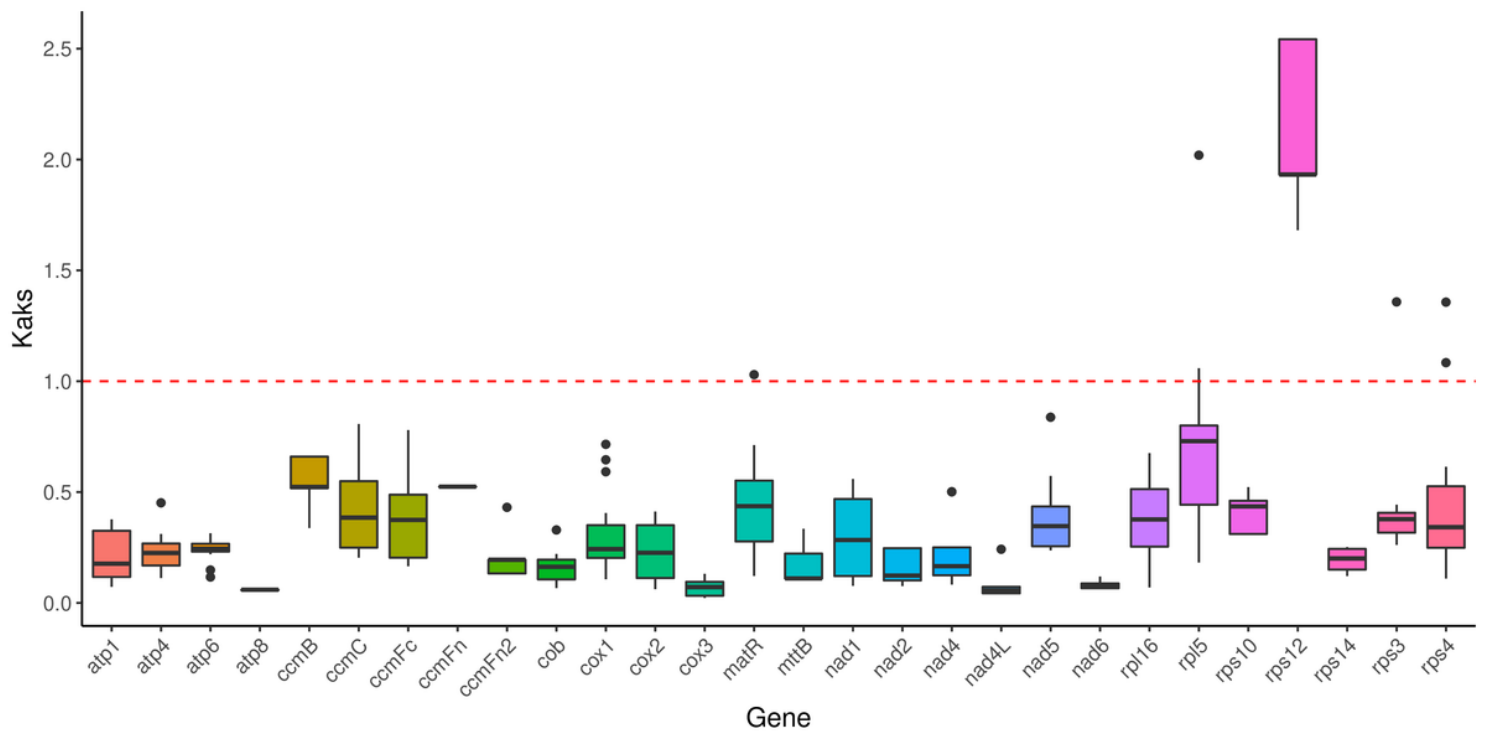


Figure 9

Boxplots of Kaks between species. The horizontal coordinates indicate the gene names and the vertical coordinates indicate the Kaks values

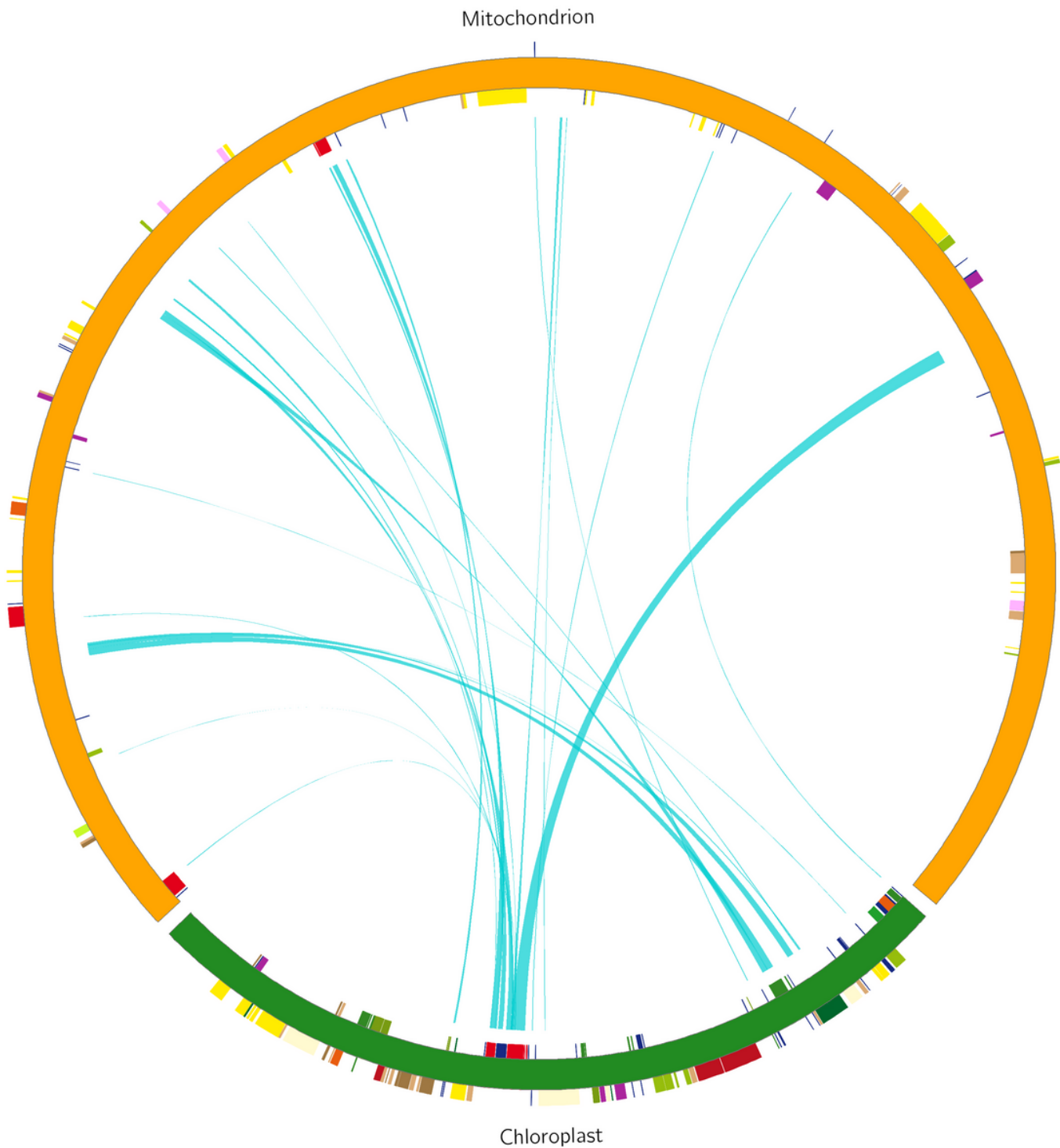


Figure 10

Cp and mt sequence homologous fragments. Chloroplast are cp sequences, others are mt sequences. Genes from the same complex are labeled with the same color, and the middle line connection indicates homologous sequences