Assembly and comparative analysis of the complete mitochondrial genome of *Suaeda glauca*

Yan Cheng
Fujian Agriculture and Forestry University
Xiaoxue He
Fujian Agriculture and Forestry University
S. V. G. N. Priyadarshani
Fujian Agriculture and Forestry University
Yu Wang
Fujian Agriculture and Forestry University
Li Ye
Fujian Agriculture and Forestry University
Chao Shi
Fujian Academy of Agricultural Sciences
Kangzhuo Ye
Fujian Agriculture and Forestry University
Qiao Zhou
Fujian Agriculture and Forestry University
Ziqiang Luo
Fujian Agriculture and Forestry University
Fang Deng
Fujian Agriculture and Forestry University
Ling Cao
Fujian Agriculture and Forestry University
Ping Zheng
Fujian Agriculture and Forestry University
Mohammad Aslam
Fujian Agriculture and Forestry University
Yuan Qin (yuanqin@fafu.edu.cn)
Fujian Agriculture and Forestry University https://orcid.org/0000-0002-4814-3049

Research article

Keywords: Suaeda glauca, mitochondrial genome, repeats, phylogenetic analysis

Posted Date: November 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-112827/v1

Version of Record: A version of this preprint was published at BMC Genomics on March 9th, 2021. See the published version at https://doi.org/10.1186/s12864-021-07490-9.

Abstract

Background

Suaeda glauca is a halophyte widely distributed in saline and sandy beaches, with strong saline-alkali tolerance. It is also a beautiful landscape plant with high development prospects and scientific research value. The *S. glauca* chloroplast genome has recently been reported; however, the mitochondria genome is still unexplored.

Results

This study assembled the mitochondria genome and annotated the mitochondrial genes of *S. glauca* based on the Pacbio long reads. The circular mitochondrial genome of *S. glauca* has a length of 474,330 bp. The base composition of the *S. glauca* mt genome showed A (27.96%), T (28.01%), C (21.64%), G (21.64%). *S. glauca* mt genome has 51 genes, including 26 protein-coding genes, 22 tRNA genes, and 3 rRNA genes. Phylogenetic analysis with common genes of 28 species resulted in similar morphological classification.

Conclusions

As a *Chenopodiaceae* species, *S. glauca* mt genome will provide insights into the missing pieces in the evolution of sex determination and improve genomic breeding in the future.

Background

Chenopodiaceae family is among the large families of angiosperms that mainly include *Spinacia oleracea L*. *Chenopodium quinoa Willd*, and *Beta vulgaris* [1–3]. *Chenopodiaceae* plants are mostly annual herbs, half shrubs, shrubs, living in the desert and saline soil areas. Therefore, they often show xerophytic adaptation. As an annual herb of *Chenopodiaceae*, *S. glauca* grows in saline-alkali land and beaches. It displays a strong salt tolerance and drought tolerance capacity and has high value as medicine and food material [4–6]. Moreover, *S. glauca* possesses immense ecological value as it can tolerate heavy metals at higher levels and can be used as a super accumulator of heavy metals. The environmental protection and remediation of contaminated soil make it a natural resource with great economical and environmental importance [7].

Plant mitochondria are involved in many cellular functions such as nucleotide biosynthesis, generation of protein clusters of iron and sulfur, and apoptosis [8]. Margulis' endosymbiosis theory suggests that mitochondria originated from archaea living in nucleated cells when eukaryotes swallowed the bacteria. Later it evolved into organelles with special functions during the long-term symbiosis [9–11], justifying additional mitochondrial genome. Mitochondria convert biomass energy into chemical energy through phosphorylation and provide energy for life activities of the body. Besides, it is involved in cell differentiation, apoptosis, can regulate cell growth and cell cycle [12–15]. Therefore, mitochondria play a crucial role in plant productivity and development [16]. For most seed plants, nuclear genetic information is inherited from both parents, while chloroplasts (cp) and mitochondria are inherited from the maternal parent. This genetic mechanism eliminates the paternal line's influence, thus reducing the difficulty of genetic research and facilitating the study of genetic mechanisms [17].

With the development of sequencing technology, more and more mitochondrial genomes have been reported. Up to July 2020, 335 complete mt genomes have been deposited in GenBank Organelle Genome Resources. Long periods of mutualism leave mitochondria with some of their original DNA gone, and some of them transferred, leaving only the

DNA that codes for it [18, 19]. Mt DNA has long been recognized as tending to integrate DNA from various sources through intracellular and horizontal transfer [20]. Therefore, the mt genome in plants has significant differences in length, gene sequence, and gene content [21]. The mt genome length of the smallest known terrestrial plant is about 66 Kb, and the largest terrestrial plant mt genome length is 11.3 Mb [22, 23]. As a result, the amount of genes in terrestrial plants varies widely, typically between 32 and 67 [24]. In this study, we sequenced and annotated the mt genome of *S. glauca* and compared it with the genomes of other angiosperms (as well as gymnosperms) to better understand the evolution of mt genomes of terrestrial plants. More importantly, this study could provide additional information for a better understanding of genetic transformation, molecular breeding, and sex-determined evolutionary missing segments.

Results

Genomic features of the *S. glauca* mt genome

The *S. glauca* mt genome is circular with a length of 474,330 bp. The base composition of the genome is A (28.55%), T (27.93%), C (21.62%), G (22.45%). There are 51 genes annotated in the mt genome, including 26 protein-coding genes, 22 tRNA genes, and 3 rRNA genes. The functional categorization and physical locations of the annotated genes on the mt genome were presented in Figure 1. According to our findings, there were 26 protein-coding genes in the mt genome of *S. glauca*, which could be divided into 9 classes (Table 1): NADH dehydrogenase (7 genes), ATP Synthase (5 genes), Cytochrome C Biogenesis (4 genes), Cytochrome C oxidase (3 genes), Ribosomal proteins (SSU) (3 genes), Ribosomal proteins (SSL) (1 gene), Succinate dehydrogenase (1 gene), Maturases (1 gene), and Ubichinol Cytochrome c Reductase (1 gene). All of the protein-coding genes used ATG as starting codon, and all three stop codons TAA, TGA, and TAG were found in *S. glauca* mt genome. TAA and TGA the same utilization rate of 44.44%, followed by TGA (40.74%). The introns in mitochondrial genes are rarely observed. There are only 6 intron-containing genes (*nad2, nad5, nad7, cox2, ccmFc, TrnA-UGC)* in *S. glauca* mt genome harboring 11 introns in total with a total length of 12,084 bp. The intron lengths varied from 105bp (*nad5*) to 2,103 bp (*nad2*). The gene *nad7* contains 4 introns, which is the highest intron number. While the *trna-UGC* only 1 intron with a length of 105bp, which is the smallest intron in the *S. glauca* mt genome.

It has been reported that most of the land plants contain 3 rRNA genes [9, 11]. Consistently, three rRNA genes *rrn5* (119 bp), *rrnS* (1303 bp), and *rrnL* (1369 bp) were annotated in *S. glauca* mt genome. In addition, 22different transfer RNAs transport 19 amino acids were identified in this *S. glauca* mt genome. Different transfer RNAs might transport the same amino acid for different codons. For example, *trnS-TGA* and *trnS-CGT* transport Ser for synonymous codons AGU and AGC, respectively, and *trnF-GAA, trnM-CAU* and *trnN-GUU* have two different structures of the same antisense codon. Taking *trnM*-CAU as an example, both structures A and B can transport Met amino acids (Figure S1).

Table 1. Gene profile and organization of the *S. glauca* mt genome.

Group of genes	Gene name	Length	Start codon	Stop codon	Amino acid
NADH dehydrogenase	*Nad1*	327	ATG	TGA	108
	Nad2[a]	915	ATG	TAA	304
	Nad3	357	ATG	TAA	118
	Nad4L	273	ATG	TAA	90
	Nad5[a]	1452	ATG	TGA	483
	Nad7[a]	1092	ATG	TAG	363
	Nad9	579	ATG	TAA	192
ATP synthase	*Atp1*	1521	ATG	TAA	506
	Atp4	597	ATG	TAG	108
	Atp6	741	ATG	TAA	246
	Atp8	480	ATG	TGA	159
	Atp9	240	ATG	TGA	59
Cytochrome c biogenesis	*ccmB*	621	ATG	TGA	206
	ccmC	744	ATG	TAA	247
	ccmFC[a]	1338	ATG	TAG	445
	ccmFN	1635	ATG	TGA	544
Cytochrome c oxidase	*Cox1*	1575	ATG	TAA	524
	Cox2[a]	768	ATG	TAA	255
	Cox3	798	ATG	TGA	256
Maturases	*matR*	1968	ATG	TAG	355
Ubichinol cytochrome c reductase	*cob*	1182	ATG	TGA	393
Ribosomal proteins（LSU）	*Rpl5*	555	ATG	TAA	184
Ribosomal proteins（SSU）	*Rps3*	1680	ATG	TAA	559
	Rps7	447	ATG	TAA	148
	Rps12	381	ATG	TGA	126
Transport membrane protein	*Sdh4*	294	ATG	TGA	97
Ribosomal RNAs	*Rrn5*	119			
	RrnS	1303			
	RrnL	1369			
Transfer RNAs	*TrnA-UGC*[a,b]	73			
	TrnC-GCA	76			

	TrnE-UUC	72
	TrnF-GAA(2)[b]	74
	TrnG-GCC	74
	TrnH-GUG[b]	76
	TrnI-GAU[b]	79
	TrnK-UUU	73
	TrnL-CAA	83
	TrnM-CAU(2)[b]	76
	TrnN-GUU(2)[b]	74
	TrnP-UGG	90
	TrnQ-UUG	72
	TrnR-ACG[b]	75
	TrnS-GCU	91
	TrnS-UGA	88
	TrnV-GAC[b]	72
	TrnW-CCA	74
	TrnY-GUA	84

Notes. a.Genes containing introns.

b. Cp-derived genes. The only one form in *trnM-CAU* is the Cp-derived genes

ND, not determined

Repeat sequences in the *S. glauca* mt genome

Microsatellites, or simple sequence repetitions (SSRs), are DNA fragments consisting of short units of sequence repetition of 1-6 base pairs in length [37]. The uniqueness and the value of microsatellites are due to their polymorphism, codominant inheritance, relative abundance, extensive genome coverage, and simplicity in PCR detection [38]. There are 361 SSRs in the mt genome of *S. glauca*, and the proportion of different forms are shown in Figure S1. SSRs in monomer and dimer forms accounted for 78.67% of the total SSRs present in the mt genome of *S. glauca*. Adenine (A) monomer repeats accounted for 46.28% (56) of the 121 SSR monomers, and AT repeats is the highest type among the dimeric SSRs, accounting for 58.15% of all the dimeric type. There are only two hexameric SSRs, located between *nad4l* and *cox2*, and between trnQ and trnM. The specific locations of pentamer and hexamer are shown in Table 2. Tandem repeats, also named as satellite DNA, refer to the core repeating units of about 1 to 200 bases, repeated several times in tandem. They are widely found in eukaryotic genomes and in some prokaryotes [39]. A total of 9 tandem repeats with a matching degree greater than 95% and a length ranging from 13 bp to 38 bp were present in the mitochondrial genome of *S. glauca*, as shown in Table 3.

Table 2. Distribution of penta and hexa SSRs in the *S. glauca* mt genome.

No.	Type	SSR	Start	End	Location
1	pentamer	(tatac)×3	3006	3020	*cox1*
2	pentamer	(agaat)×3	49581	49595	*nad7*
3	pentamer	(taagt)×3	78725	78739	IGS(*nad7,trnI*
4	pentamer	(ggaaa)×3	107921	107935	IGS(trnQ-UUG,trnM-CAU)
5	pentamer	(cgggc)×3	139703	139717	IGS(nad2,nad9)
6	pentamer	(cttct)×3	168170	168184	IGS(trnW-CCA,atp1)
7	pentamer	(tcttg)×3	201546	201560	IGS(trnV-GAC,trnA-UGC)
8	pentamer	(agaat)×3	225057	225071	nad7
9	pentamer	(ttctt)×3	316091	316105	IGS(trnF-GAA.trnS-UGU
10	pentamer	(actag)×3	330081	330095	*matR*
11	pentamer	(caaaa)×3	388600	388614	IGS(atp8,atp9(
12	pentamer	(agaaa)×3	401486	401500	IGS(atp9(rrnS(
13	hexamer	(caaaat)×3	92262	92279	IGS(nad4L(cox2(
14	hexamer	(tagaaa)×3	106488	106505	IGS(trnQ-UUG,trnM-CAU(

Table 3. Distribution of perfect tandem repeats in the *S. glauca* mt genome.

No.	Size	Repeat sequence	Copy	Percent Matches	Start	End
1	32	CCATACTTGTTCCAAGTAAGTGAATTGCATTA	6	99	48018	48212
2	31	GAGACAAGTCTAGTATAGACGCAGGGTCGAA	5	98	104348	104524
3	38	TTTCGGAAGTTTTATCCTATAAGAATTGGCTTTTCCTT	2	95	168613	168711
4	13	TCTAATAGAAAAT	2	100	201473	201497
5	16	AATGTGTATTATCCAT	2	100	294569	294601
6	18	ATATCGTCACTAGCATCA	2	100	296770	296808
7	18	AGTCTATCAACGCTACTG	2	100	335715	335749
8	32	GGTAATGCCAATTCACTTACTTGGAACAAGTAT	6	99	454228	454422

RNA editing in the *S. glauca* mt genome

RNA editing refers to the addition, loss, or conversion of the base in the coding region of the transcribed RNA [40], which is commonly found in terrestrial plants other than mosses [41]. RNA editing alters genomic information by translating specific cytosine sites of chloroplasts and mitochondrial transcripts into uridine [42]. This process improves protein preservation in plants by modifying codons. According to this principle, 216 RNA editing sites within 26 protein-

coding genes in *S. glauca* mt genome were identified with PREP-MT program (Figure 2). Among those protein-coding genes, *COX1* does not have any editing sites, and *ccmB* has the most editing sites (29). Among all the editing sites, 35.19% (76) was located at the first base of the codon, 63.89% (138) was located at the second base of the codon, and there was a special editing situation where the first and second positions of the codon changed from the original proline (CCC) to phenylalanine (TTC). The affinity of most (54.16%) amino acids to water did not change, 43.98% of the amino acids were hydrophobic to hydrophilic, and only two amino acids changed from hydrophobic to hydrophilic. The occurrence of RNA editing can lead to the premature termination of protein-coding genes, and it is occurring in *ATP4* and *ATP9* of *S. glauca* mt genome. The amino acids of 48.15% (106 sites) editing codons showed a tendency to leucine. Other RNA-editing cases are shown in Table 4.

Table 4. The RNA-editing locations

Type	RNA -editing	Number	Percentage
hydrophilic	GCT (A) => GTT (V)	7	8.33%
	CTT (L) => TTT (F)	11	
hydrophobic	CAT (H) => TAT (Y)	12	45.83%
	CCC (P) => TCC (S)	24	
	CGT (R) => TGT (C)	12	
	TCT (S) => TTT (F)	8	
	TCA (S) => TTA (L)	37	
	ACT (T) => ATT (I)	6	
Hydrophilic to hydrophobic	CCC (P) => TTC (F)	2	0.93%
Hydrophobic to hydrophilic	CCC (P) => CTC (L)	47	43.98%
	CGG (R) => TGG (W)	15	
	TCC (S) => TTC (F)	11	
	TCG (S) => TTG (L)	20	
	ACG (T) => ATG (M)	2	
Hydrophilic to stop	CAA (Q) => TAA (X)	1	0.93%
	CGA (R) => TGA (X)	1	

The substitution rates of protein-coding genes in the *S. glauca* mt genome

The calculation of non-synonymous substitutions (Ka) and synonymous substitutions (Ks) is of great significance for the reconstruction of phylogeny and the understanding of the evolutionary dynamics of protein-coding sequences in closely related but different species [20]. In genetics, Ka/Ks can determine whether there is selective pressure on the

protein-coding gene. If Ka/Ks>1, it is considered as a positive selection that occurred during evolution. If Ka/Ks=1, it is regarded as a neutral selection, and if Ka/Ks<1, it is considered to have a negative selection effect. The 24 protein-coding genes of *S. glauca* mt genome were compared with the mitochondrial genome of *C.papaya* for Ka/Ks calculation, and the results are shown in Figure 3. The ratios of Ka/Ks were significantly less than 1 in most of the protein-coding genes. *CCmB* was close to 1, indicating that it may have experienced neutral evolution since their common ancestor's divergence. The Ka/Ks values of all the other genes are less than 1, indicating their negative selections during evolution. It is suggested that these genes may be highly conservative in the evolutionary process of higher plants.

Migration of chloroplast DNA in mitochondrial genomes

A total of 32 fragments were with a total length of 26.87 kb were found to be transferred from chloroplast to mt in *S. glauca*, accounting for 5.18% of the mitochondrial genome. There are 8 annotated genes located on those fragments, all of which are tRNA genes, namely *trnA-UGC, trnF-GAA(2), trnH-GUG, trnI-GAU, trnR-ACG, trnM-CAU, trnN-GUU(2), and trnV-GACI*. Additionally, our data also supported that the cp protein-coding genes, such as *atpA, rrn16, rrn23, rpoC2, ndhA, psaB and psbB,* also had been transferred from cp to mt (Table S1). However, most of them lost their integrities during evolution, and only partial sequences of those genes could be found in the mt genome nowadays. The different destinations of transferred protein-coding genes and tRNA genes suggested that tRNA genes were much more conserved in the mt genome than the protein-coding genes, indicating their indispensable roles in mitochondrial.

Phylogenetic analysis within higher plant mt genomes

Phylogenetic analyses were performed based on an aligned data matrix of 23 protein-coding genes for 29 species. As shown in Figure 4, 29 plant species were divided into three categories, including 24 eudicots, 4 monocots, and 2 gymnosperms. The two gymnosperms were designated as outgroups. Abbreviations of all these observed plant mt genomes are listed in Table S1. The phylogenetic tree strongly supported the separation of eudicots from monocots and the separation of angiosperm from gymnosperm. The branches of *Leguminosae*, *Cruciferae*, *Salicaceae*, *Cucurbitaceae,* and *Chenopodiaceae* were highly reliable, indicating the consistency of traditional taxonomy with the molecular classification.

Comparison of genomic features with other mitochondrial genomes

With the rapid development of sequencing technology, more and more complete plant mt genomes were reported, facilitating the comparison analysis of the mt genome features among multiple plant species [43]. As described by Richardson et al., the mt genomes of in plants varied considerably in size, gene content, and gene order, only the same genus sharing the similar mt genomic features [21]. We found that the coding genes were conserved by comparing the genome features in six higher plant mt genomes (Table 5). However, the proportions of coding regions were significantly different from each other, probably due to their different sizes of mt genomes. As shown in Table 5, protein-coding genes and cis-introns regions account for 4.34% and 3.31% of the whole *S. glauca* mt genome sequence, respectively. In comparison, the proportions of rRNA and tRNA regions only represent 0.53% and 0.34%, respectively. Simultaneously, the whole *S. glauca* genome sequence protein-coding genes and cis-introns regions account for 4.34% and 3.31%.

Table 5. Comparison of genome features in six higher plant mt genomes

Plant species	Coding regions(%)				Non-coding regions (%)
	Protein-coding genes	Cis-spliced introns	rRNAs	tRNAs	
G.biloba	9.95	11.31	1.44	0.5	76.8
Z.mays	6.06	4.06	0.99	0.28	88.61
G.max	8.48	8.09	1.31	0.35	81.77
G.raimondii	5.14	5.28	1.61	0.28	87.69
S.suchowensis	4.68	4.21	0.83	0.27	90.01
S.glauca	4.34	3.31	0.53	0.34	91.48

Additionally, we also compared the sizes and GC contents of mt genomes of 22 land plants (Figure 5), including 4 phycophyta (*Chlamydomonas reinhardtii*, *Chlorella heliozoae*, *Chara vulgaris*, *Nitella hyalina*), 3 bryophytes (*Buxbaumiaaphylla*, *Bartramiapomiformis*, and *Sphagnum palustre*), 2 gymnosperms (*Gink gobiloba* and *Cycas taitungensis*), 3 monocots (*Triticumaestivum*, *Sorghumbicolor*, and *Oryzasativa*), and 10 dicots (*Arabidopsis thaliana*, *Brassica napus*, *Glycine max*, *Vitis vinifera*, *Nicotiana tabacum*, *Suaeda glauca*, *Beta vulgaris subsp. vulgaris*, *Gossypium hirsutum*, *Cucurbita pepo*, and *Cucumis sativus*). The abbreviation of all these observed plant mt genomes are listed in Table S1. The mt genome sizes of taxa observed vary from 15,758 bp (*C.reinhardtii*) to 1,555,935 bp (*C.sativus*), while that of *S. glauca* (474,330 bp) is in the middle of the vacation range. The mt genomes of phycophyta and bryophytes are generally small. Similarly, the GC contents are also variable, ranging from 32.24% in *S.palustre* to 50.36% in *G.biloba*. Additionally, the GC contents of angiosperms, including monocots and dicots, are larger than that of bryophytes but smaller than that of gymnosperms, suggesting that the GC contents frequently changed since the divergence of bryophytes, gymnosperms, and angiosperms. However, in phycophyta, GC content fluctuates wildly. While the GC contents in angiosperms were conserved during the evolution, despite their genome sizes varied tremendously.

Discussion

Mitochondria are the powerhouse of the plants that produce the required energy to carry out life processes. Plant mitochondria possess a more complex genome than the animal mitochondrial genome with extensive variations in size, sequence arrangements, and repeat content and have a highly conserved coding sequence [44]. In the current study, we studied the characteristics of the mitochondrial genome of *S. glauca*, which is important as a salt stress tolerance plant with great value as a food source and phytoremediation agent. In this study, we found that the mt genome of *S. glauca* was linear. According to the reported data, we found that most of the mt genome is circular, and few mt genomes are linear such as the rice mt genome (*Oryza sativa* Japonica Group) [45]. Therefore, this study provided data for the subsequent study of mitochondrial genome structure. In addition, the mt genome characteristics of *S. glauca* were consistent with those of other terrestrial green plants. The understanding of mt genome structure is required to unravel its function, replication, inheritance, and its evolutionary trajectories[44]. The presence of repeat sequences plays a pivotal role in shaping mt genome[46] [48]. These repeats include tandem, short and large repeats[47, 48]. In the current study, we found that monomers and dimers are highly present with 9 tandem repeats. Previous studies have shown that repeats in mitochondria have much genetic information and are vital for the intramolecular recombination[46].

RNA-editing is a posttranscriptional process that occurs in chloroplast and mitochondrial genomes in higher plants, and it contributes to the better folding of proteins [49]. Understanding of the RNA-editing sites is beneficial for understanding the gene expression of the cp and mt genes in plants. Previous studies reported that there are 441 RNA-editing sites within 36 genes in *Arabidopsis*, and 491 RNA-editing sites within 34 genes in rice [45, 50]. In the study, 233 RNA-editing sites within 37 genes were identified in *S. glauca*. The identification of RNA editing sites will provide important clues for predicting gene functions with novel codons.

The Ka/Ks analysis and the comparison of genome features with other plant's mt genomes provide a comprehensive understanding of plant mt evolution. Generally, most of the results in this study were consistent with previous reports. The genes have undergone neutral, and negative selections were also identified in *S. glauca*. Migration of cp DNA to the mt genome occurred during the plant evolution. Our analysis found that thirty-six fragments were transferred from the cp genome to mt with ten genes. Other than protein-coding genes, 10 tRNA genes were also transferred from the cp genome to the mt genome. Transfer of tRNA genes from cp to mt is common in angiosperms [49].

Further, we have analyzed the *S. glauca* mt genome's phylogenetic relationship with the other species mt genomes. The resulted phylogenetic tree reflected a clear taxological l relationship among the taxa. We also analyzed GC content of the mt genome in S. glauca. The result supported the conclusion that GC content is highly conserved in higher plants. In crop plants, the Mitochondria are very important for plant breeding. With the accomplishment of the mt genome of *S. glauca* and further developing the next and third-generation sequencing platforms, we also have an opportunity to conduct further genomic studies in *S. glauca*. Understanding of mt genome will set a foundation for the evolutionary analysis, cytoplasmic male sterility, and molecular biological information for plant breeding. Therefore, our study provides an essential background study for future molecular breeding of this plant [49].

Conclusion

In this study, we sequenced and annotated the mt genome of *S. glauca* and performed extensive analyses based on the DNA sequences and amino acid sequences of the annotated genes. The *S. glauca* mt genome is circular, with a length of 474,330 bp. 51 genes, including 26 protein-coding genes, 22 tRNA genes, and 3 rRNA genes, were annotated in the genome. The repeats sequences and RNA editing in *S. glauca* mt genome were analyzed subsequently. The Ka/Ks analysis based on code substitution revealed that most of the coding genes underwent positive selections, indicating the conservation of mt genes during the evolution. The gene conversation between mt and cp genome was also observed in *S. glauca* by detecting gene migration. Moreover, our result also indicated the consistency of traditional taxonomy with the molecular classification and that the GC contents in angiosperms were conserved during the evolution, despite their genome sizes varied tremendously. This study could provide additional information for a better understanding of genetic transformation, molecular breeding, and sex-determined evolutionary missing segments.

Methods

Plant growth conditions, DNA extraction, and sequencing

Plants germinated from the seeds of *S. glauca* (preserved in Qin Lab, Fujian Agriculture and Forestry University) were first treated with gibberellin (GA3, 1mg/L) for 24 h to accelerate germination. All experimental materials were planted in an artificial intelligence greenhouse, and growth conditions were ensured as 16 h light, 8 h dark; the temperature was 26 °C ~ 28°C with 70% humidity. Leaves from about 40 days old plant were used to extract DNA using the CTAB method [25]. The quality of DNA samples was examined with agarose-gel electrophoresis, and Nanodrop (2000c UV-Vis) and qualified samples were sent to the company for sequencing.

Assembly and annotation of the mitochondrial genome

Mitochondrial genome reads were first extracted from whole-genome sequencing reads of *S. glauca* with Pilon v1.18. Then the mitochondrial sequences of *S. glauca* were selected with blast software using the conserved mitochondrial sequences of *Beta vulgaris*, *Spinacia oleracea L*, and *Chenopodium quinoa Willd* as queries. The mitochondria genome was assembled using Canu v1.8 with those selected reads [26]. The assembled contigs were polished with second-generation sequencing reads to avoid the read errors.

The GE-Seq function of MPI-MP CHLOROBX [27] (https://chlorobox.mpimp-golm.mpg.de) was used for the mt genome annotation with the mt genomes of *Arabidopsis thaliana* (NC_037304), *Beta vulgaris* (NC_002511), *Brassica napus* (NC_008285), *Carica papaya* (NC_012116), *Chenopodium quinoa Willd* (NC_041093), *Daucus Carota* (NC_017855), *Glycine max* (NC_020455), *Nicotiana tabacum* (NC_006581), *Spinacia oleracea L.* (NC_035618), and *Salix suchowensis* (NC_029317) as references. The threshold for Protein Search Identity was 55%, and that of rRNA, tRNA, and DNA Search Identity was 85%. The software's annotation results were manually adjusted with Mega v7.0 [28]. The output GB format file was manually modified, and the mitochondrial circular map was drawn using OrganellarGenomeDRAW (OGDRAW) [29].

Analysis of repeated sequences

The repeats of 1-6 bases with 8, 4, 4, 3, 3, and 3 repeats numbers were considered simple repeats, which were investigated using the Microsatellite identification tool [30] (https://webblast.ipk-gatersleben.de/misa/index.php). Tandem repeats were analyzed using Tandem Repeats Finder v4.09 [31] (http://tandem.bu.edu/trf/trf.submit.options.html). Repeat sequences with a length of 10-38 bp and a matching degree of repeating units higher than 95% were approved.

Analysis of RNA editing and substitution rate

Carica papaya genome was used as a reference to analyze the synonymous (Ks) and non-synonymous (Ka) substitution rates of the protein-coding genes in *S. glauca* and other higher plants. Mega software (version 7.0) was used to extract and compare the protein sequences of corresponding protein-coding genes in *S. glauca* [28], and DNAsP v.6.12 was used to calculate Ka/Ks [32]. The Plant Predictive RNA Editor (PREP) suite [33](http://prep.unl.edu/) was used to identify possible editing sites in protein-coding genes with a cut off value of 0.2. The editing sites in the mitochondrial RNA of *S. glauca* were revealed by taking the gene encoding protein of mt genome of other plants as reference.

Chloroplast to mitochondrion DNA transformation analysis

DNA migration is common in plants and varies from species to species [34]. This phenomenon occurs during autophagy, gametogenesis and fertilization [35]. The chloroplast genome of *S. glauca* (NC_045302.1) was downloaded from NCBI Organelle Genome Resources Database. Blastn software on NCBI was used to identify the protein-coding and tRNA genes transferred from chloroplasts to mitochondria. Screening criteria were set as the matching rate ≥ 80%, E-value ≤ -10, and length ≥ 40.

Phylogenetic analysis

The conserved protein-coding genes from mt genomes of *S. glauca* and other 28 taxa were used for phylogenetic tree construction. The mitochondrial genomes were downloaded from NCBI, and the conserved protein-coding genes (*atp1 atp4, atp6, atp8, atp9, ccmB, ccmC, ccmFc, ccmFn, cob, cox1, cox2, cox3, matR, nad1, nad2, nad3, nad4L, nad5, nad6,*

nad7 and nad9) were extracted using TBtool software, and then aligned using Muscle software [36]. Subsequently, a Neighbor-joining (NJ) tree was constructed by Mega v.7.0 software using the Poisson model with a bootstrap of 1000 [28]. *C.taitungensis* and *G. biloba* were designated as the outgroup in this analysis.

Abbreviations

S. glauca : Suaeda glauca; mt: mitochondria; cp: chloroplast

Declarations

Ethics approval and consent to participate

Not applicable.

Consent to publish

Not applicable.

Availability of data and materials

The sequence and annotation of *S. glauca* mt genome was provided as Additional file 2. We are on the process of submitting the datat to NCBI. The Gene Banks ID will be provided when the sumbssion is done.

Competing interests

The authors declare that they have no competing interests.

Funding

Y.C. is supported by a grant from National Natural Science Foundation, China (31671267). A grant from the Natural Science Foundation of Fujian Province (2018J01704). Y.Q. is supported by a grant from National Natural Science Foundation, China (31970333) and Guangxi Distinguished Experts Fellowship. The Funding bodies were not involved in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' Contributions

YC and YQ concieved and designed the research. XH, YW, LY, CS, KY, QZ, ZL, FD and LC performed the experiments. MA helped with a critical discussion on the work. XH and YC wrote the paper. SP, MA, and YQ revised the paper. All authors discussed the results and approved the final version of the manuscript.

Acknowledgements

We thank Chunyin Zhang for providing the original seeds of *Suaeda Glauca*.

References

1. Cai X, Jiao C, Sun H, Wang X, Xu C, Fei Z, Wang Q: **The complete mitochondrial genome sequence of spinach, Spinacia oleracea L.** *Mitochondrial DNA Part B* 2017, **2**(1):339-340.
2. Maughan PJ, Chaney L, Lightfoot DJ, Cox BJ, Tester M, Jellen EN, Jarvis DE: **Mitochondrial and chloroplast genomes provide insights into the evolutionary origins of quinoa (Chenopodium quinoa Willd.)**. *Scientific reports*

2019, **9**(1):1-11.

3. 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 tRNACys (GCA)**. *Nucleic acids research* 2000, **28**(13):2571-2576.
4. Huang W, Li W, Niu Z, Xie Z, Liu X: **Interactive Effect of Salinity and Drought on the Germination of Dimorphic Seeds of Suaeda salsa**. In: *Sabkha Ecosystems.* Springer; 2014: 143-153.
5. Song J, Fan H, Zhao Y, Jia Y, Du X, Wang B: **Effect of salinity on germination, seedling emergence, seedling growth and ion accumulation of a euhalophyte Suaeda salsa in an intertidal zone and on saline inland**. *Aquatic Botany* 2008, **88**(4):331-337.
6. Huang W, Li Z-G, Qiao H-L, Li C-Z, Liu X-J: **Interactive effect of sodium chloride and drought on growth and osmotica of Suaeda salsa**. *Chin J Eco Agric* 2008, **16**:173-178.
7. Zhang X, Li M, Yang H, Li X, Cui Z: **Physiological responses of Suaeda glauca and Arabidopsis thaliana in phytoremediation of heavy metals**. *Journal of environmental management* 2018, **223**:132-139.
8. Shtolz N, Mishmar D: **The mitochondrial genome–on selective constraints and signatures at the organism, cell, and single mitochondrion levels**. *Frontiers in Ecology and Evolution* 2019, **7**:342.
9. Cavalier-Smith T: **The origin of nuclei and of eukaryotic cells**. *Nature* 1975, **256**(5517):463-468.
10. Berry S: **Endosymbiosis and the design of eukaryotic electron transport**. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 2003, **1606**(1-3):57-72.
11. Archibald JM: **Origin of eukaryotic cells: 40 years on**. *Symbiosis* 2011, **54**(2):69-86.
12. Bonora M, De Marchi E, Patergnani S, Suski J, Celsi F, Bononi A, Giorgi C, Marchi S, Rimessi A, Duszyński J: **Tumor necrosis factor-α impairs oligodendroglial differentiation through a mitochondria-dependent process**. *Cell Death & Differentiation* 2014, **21**(8):1198-1208.
13. van Loo G, Saelens X, Van Gurp M, MacFarlane M, Martin S, Vandenabeele P: **The role of mitochondrial factors in apoptosis: a Russian roulette with more than one bullet**. *Cell Death & Differentiation* 2002, **9**(10):1031-1042.
14. Kroemer G, Reed JC: **Mitochondrial control of cell death**. *Nature medicine* 2000, **6**(5):513-519.
15. Rehman J, Zhang HJ, Toth PT, Zhang Y, Marsboom G, Hong Z, Salgia R, Husain AN, Wietholt C, Archer SL: **Inhibition of mitochondrial fission prevents cell cycle progression in lung cancer**. *The FASEB Journal* 2012, **26**(5):2175-2186.
16. Ogihara Y, Yamazaki Y, Murai K, Kanno A, Terachi T, Shiina T, Miyashita N, Nasuda S, Nakamura C, Mori N: **Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome**. *Nucleic acids research* 2005, **33**(19):6235-6250.
17. Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza A, Elsas LJ, Nikoskelainen EK: **Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy**. *Science* 1988, **242**(4884):1427-1430.
18. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P: **Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers**. *Annals of the entomological Society of America* 1994, **87**(6):651-701.
19. Knoop V: **The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective**. *Current genetics* 2004, **46**(3):123-139.
20. Schuster W, Brennicke A: **Plastid, nuclear and reverse transcriptase sequences in the mitochondrial genome of Oenothera: is genetic information transferred between organelles via RNA?** *The EMBO Journal* 1987, **6**(10):2857-2863.

21. Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD: **The “fossilized” mitochondrial genome of Liriodendron tulipifera: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate**. *BMC biology* 2013, **11**(1):1-17.
22. Skippington E, Barkman TJ, Rice DW, Palmer JD: **Miniaturized mitogenome of the parasitic plant Viscum scurruloideum is extremely divergent and dynamic and has lost all nad genes**. *Proceedings of the National Academy of Sciences* 2015, **112**(27):E3515-E3524.
23. Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR: **Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates**. *PLoS Biol* 2012, **10**(1):e1001241.
24. Hsu CL, Mullin BC: **Physical characterization of mitochondrial DNA from cotton**. *Plant molecular biology* 1989, **13**(4):467-468.
25. Doyle J, Doyle J: **CTAB total DNA isolation**. *Molecular techniques in taxonomy* 1991:283-293.
26. 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 research* 2017, **27**(5):722-736.
27. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S: **GeSeq–versatile and accurate annotation of organelle genomes**. *Nucleic acids research* 2017, **45**(W1):W6-W11.
28. Kumar S, Stecher G, Tamura K: **MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets**. *Molecular biology and evolution* 2016, **33**(7):1870-1874.
29. Greiner S, Lehwark P, Bock R: **OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes**. *Nucleic Acids Research* 2019, **47**(W1):W59-W64.
30. Beier S, Thiel T, Münch T, Scholz U, Mascher M: **MISA-web: a web server for microsatellite prediction**. *Bioinformatics* 2017, **33**(16):2583-2585.
31. Benson G: **Tandem repeats finder: a program to analyze DNA sequences**. *Nucleic acids research* 1999, **27**(2):573-580.
32. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A: **DnaSP 6: DNA sequence polymorphism analysis of large data sets**. *Molecular biology and evolution* 2017, **34**(12):3299-3302.
33. Mower JP: **The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments**. *Nucleic acids research* 2009, **37**(suppl_2):W253-W259.
34. Chang S, Wang Y, Lu J, Gai J, Li J, Chu P, Guan R, Zhao T: **The mitochondrial genome of soybean reveals complex genome structures and gene evolution at intercellular and phylogenetic levels**. *PLoS One* 2013, **8**(2):e56502.
35. Huang CY, Ayliffe MA, Timmis JN: **Direct measurement of the transfer rate of chloroplast DNA into the nucleus**. *Nature* 2003, **422**(6927):72-76.
36. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput**. *Nucleic acids research* 2004, **32**(5):1792-1797.
37. Liu Y-c, Liu S, Liu D-c, Wei Y-x, Liu C, Yang Y-m, Tao C-g, Liu W-s: **Exploiting EST databases for the development and characterization of EST-SSR markers in blueberry (Vaccinium) and their cross-species transferability in Vaccinium spp**. *Scientia Horticulturae* 2014, **176**:319-329.
38. Powell W, Machray GC, Provan J: **Polymorphism revealed by simple sequence repeats**. *Trends in plant science* 1996, **1**(7):215-222.
39. GAO H, KONG J: **Distribution characteristics and biological function of Tandem Repeat Sequences in the genomes of different organisms**. *Zoological Research* 2005, **26**(5):555-564.

40. Brennicke A, Marchfelder A, Binder S: **RNA editing**. *FEMS microbiology reviews* 1999, **23**(3):297-316.
41. Malek O, Lättig K, Hiesel R, Brennicke A, Knoop V: **RNA editing in bryophytes and a molecular phylogeny of land plants**. *The EMBO journal* 1996, **15**(6):1403-1411.
42. Schallenberg-Rüdinger M, Knoop V: **Coevolution of organelle RNA editing and nuclear specificity factors in early land plants**. In: *Advances in Botanical Research.* vol. 78: Elsevier; 2016: 37-93.
43. Wei S, Wang X, Bi C, Xu Y, Wu D, Ye N: **Assembly and analysis of the complete Salix purpurea L.(Salicaceae) mitochondrial genome sequence**. *SpringerPlus* 2016, **5**(1):1894.
44. Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, Christensen AC: **The alternative reality of plant mitochondrial DNA: One ring does not rule them all**. *PLoS genetics* 2019, **15**(8):e1008373.
45. Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K: **The complete sequence of the rice (Oryza sativa L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants**. *Molecular Genetics and Genomics* 2002, **268**(4):434-445.
46. Dong S, Zhao C, Chen F, Liu Y, Zhang S, Wu H, Zhang L, Liu Y: **The complete mitochondrial genome of the early flowering plant Nymphaea colorata is highly repetitive with low recombination**. *BMC genomics* 2018, **19**(1):1-12.
47. Guo W, Zhu A, Fan W, Mower JP: **Complete mitochondrial genomes from the ferns Ophioglossum californicum and Psilotum nudum are highly repetitive with the largest organellar introns**. *New Phytologist* 2017, **213**(1):391-403.
48. Gualberto JM, Mileshina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A: **The plant mitochondrial genome: dynamics and maintenance**. *Biochimie* 2014, **100**:107-120.
49. Bi C, Paterson AH, Wang X, Xu Y, Wu D, Qu Y, Jiang A, Ye Q, Ye N: **Analysis of the complete mitochondrial genome sequence of the diploid cotton Gossypium raimondii by comparative genomics approaches**. *BioMed Research International* 2016, **2016**.
50. Unseld M, Marienfeld JR, Brandt P, Brennicke A: **The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides**. *Nature genetics* 1997, **15**(1):57-61.

Supplementary Information

Additional file 1 Supplementary tables and figures in a word file.

Table S1. The abbreviations and NCBI accession numbers of all used plant mt genomes.

Figure S1. The distribution of SSRs in the *S. glauca* mt genome. The colors represent different types of SSRs. The percentages of SSRs are also provided on the pie chart.

Figure S2. The secondary structure of tRNA. A and B are two different structure of TrnM-CAU.

Additional file 2 The sequence and annotation of *S. glauca* mt genome.

Figures

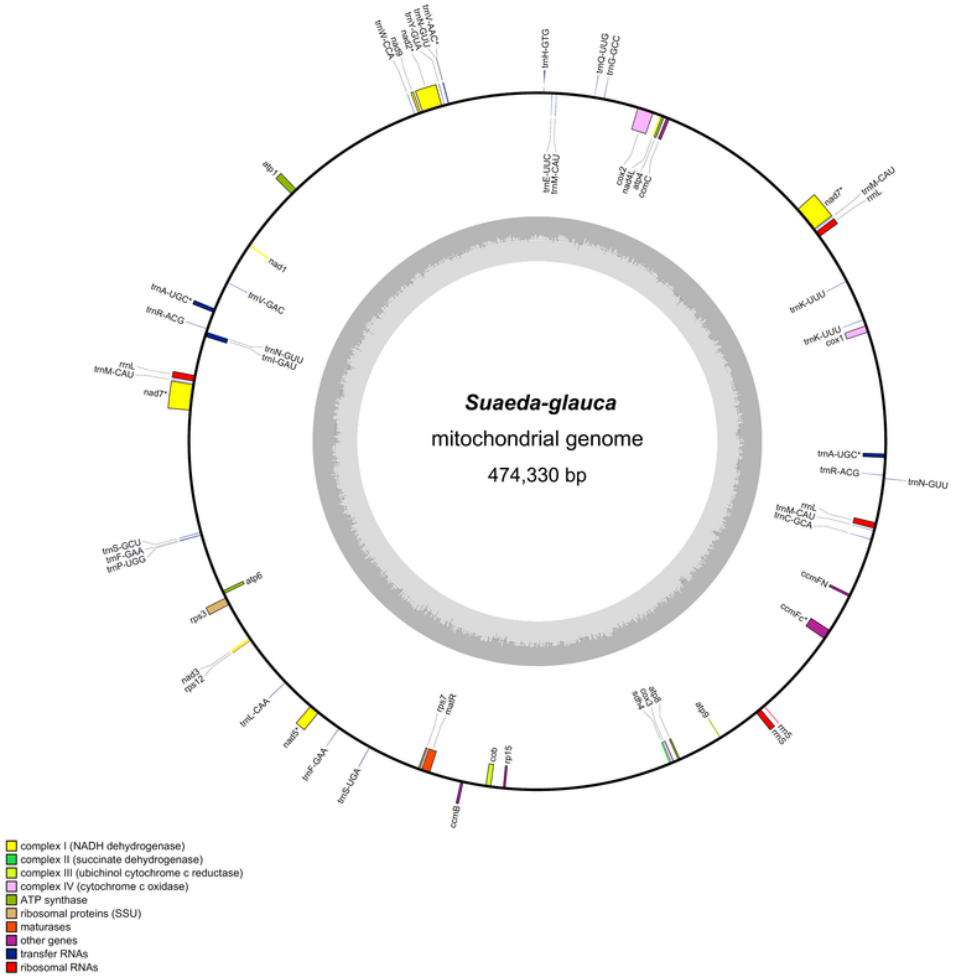

Figure 1

The circular map of S. glauca mitochondrial genome. Gene map showing 53 annotated genes with different functional groups.

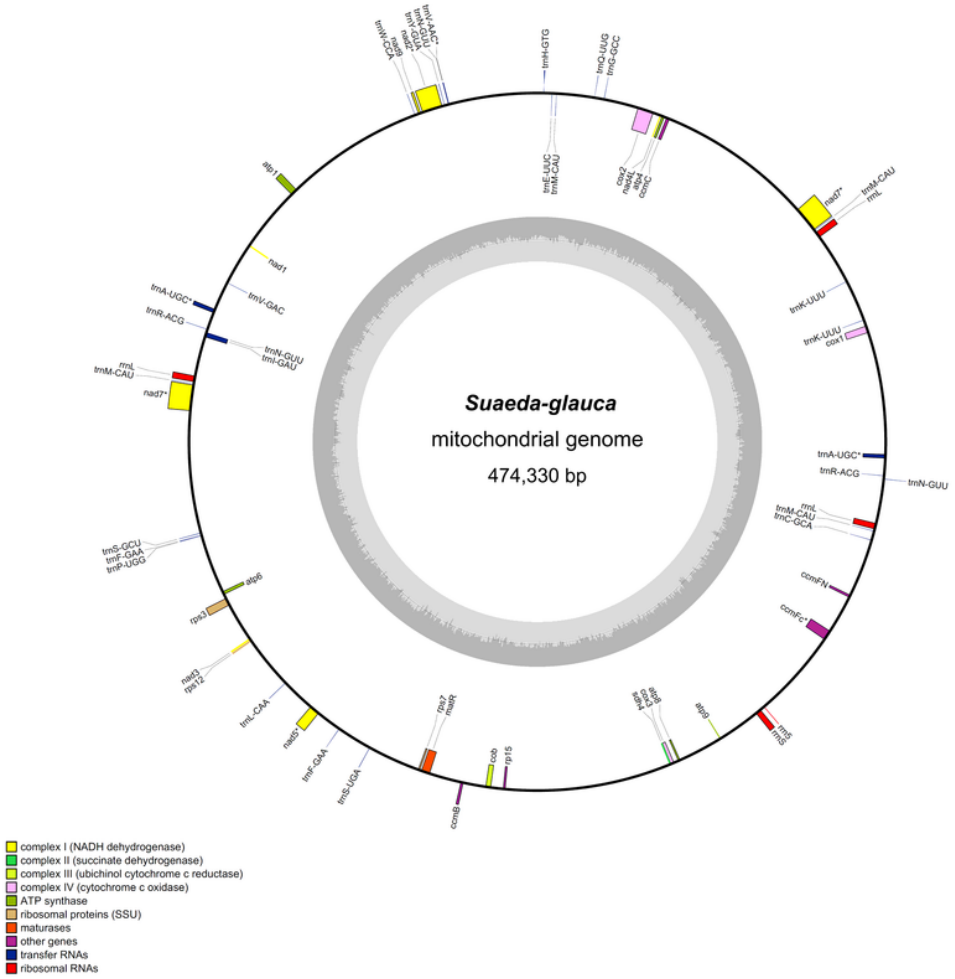

Figure 1

The circular map of S. glauca mitochondrial genome. Gene map showing 53 annotated genes with different functional groups.

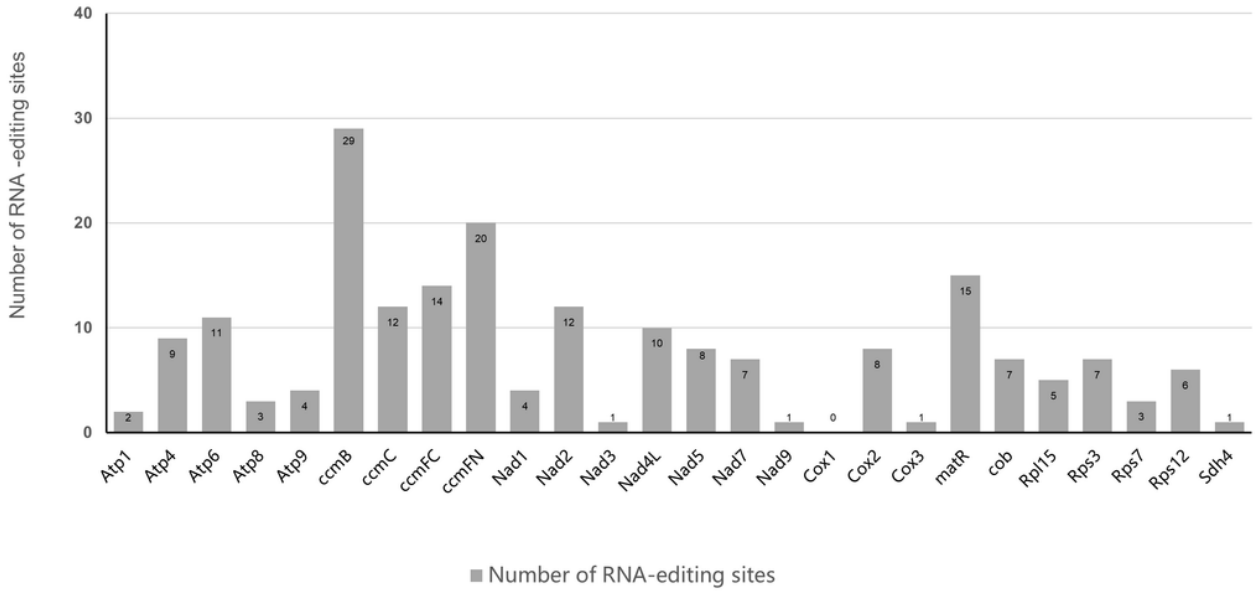

Figure 2

The distribution of RNA-editing sites in the S. glauca mt protein-coding genes. The number shown by gray box represents the RNA-editing sites of each gene.

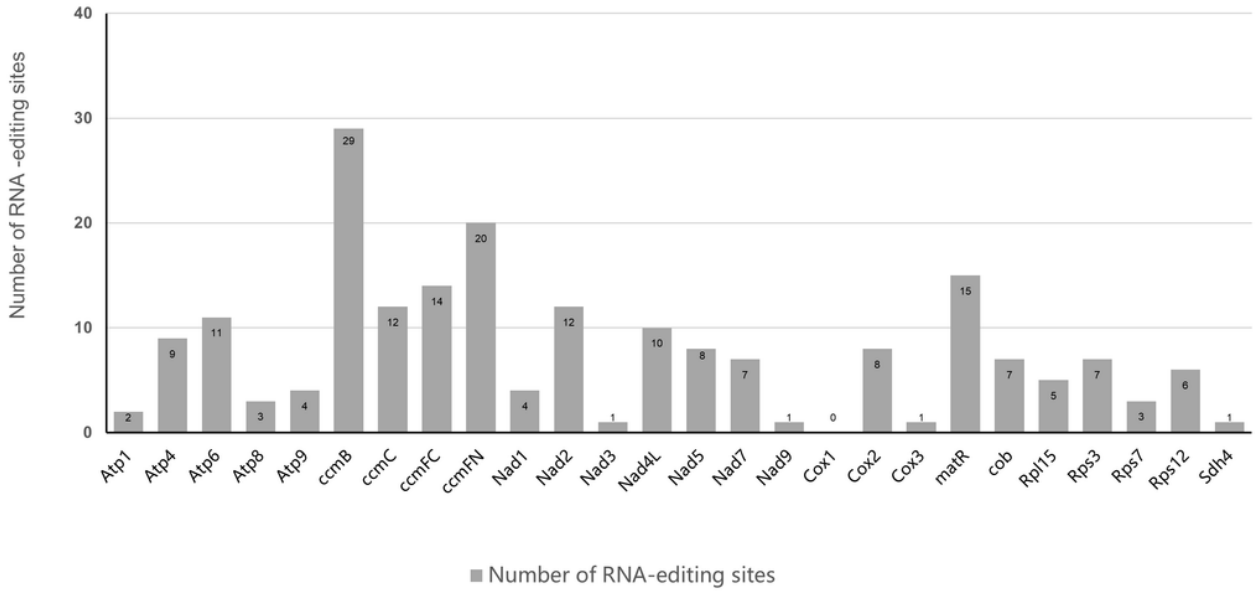

Figure 2

The distribution of RNA-editing sites in the S. glauca mt protein-coding genes. The number shown by gray box represents the RNA-editing sites of each gene.

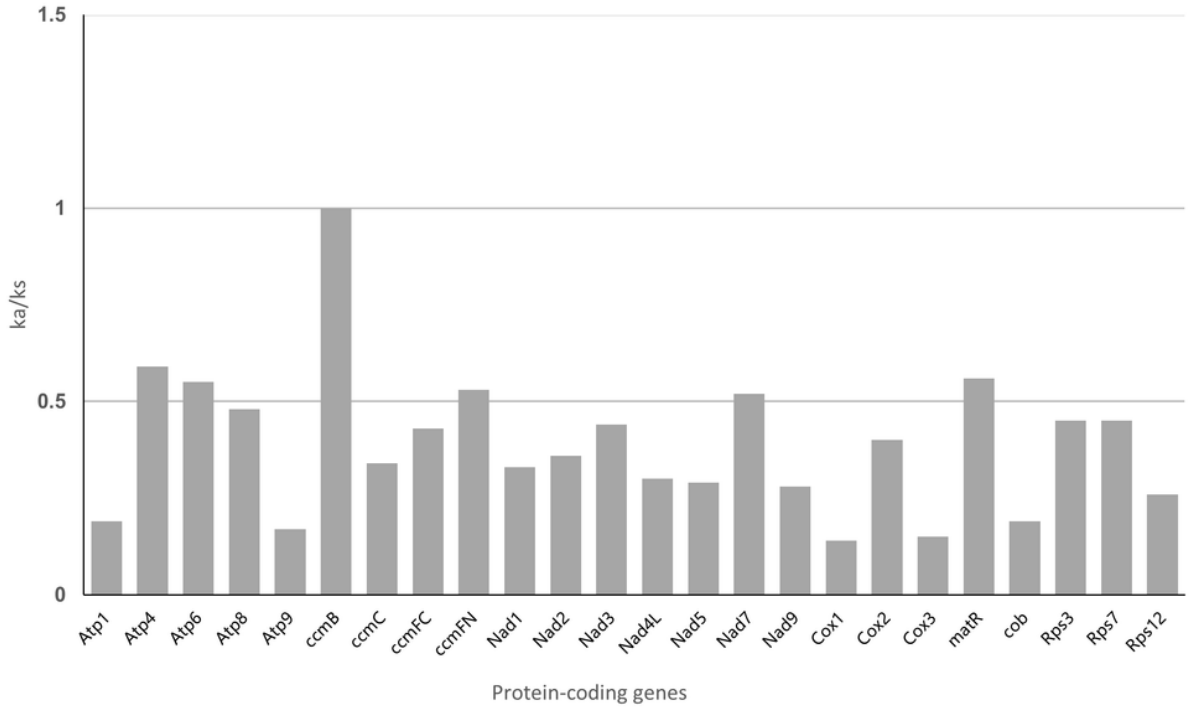

Figure 3

The Ka/Ks values of 24 protein-coding genes of S. glauca versus C. papaya.

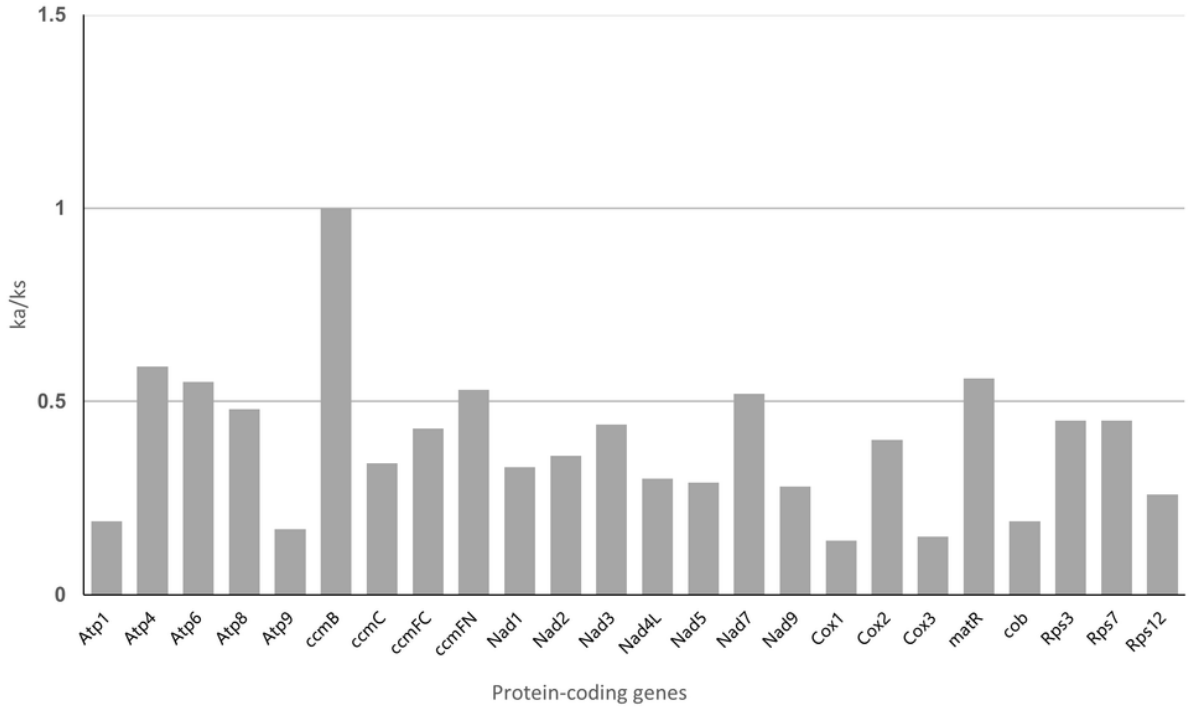

Figure 3

The Ka/Ks values of 24 protein-coding genes of S. glauca versus C. papaya.

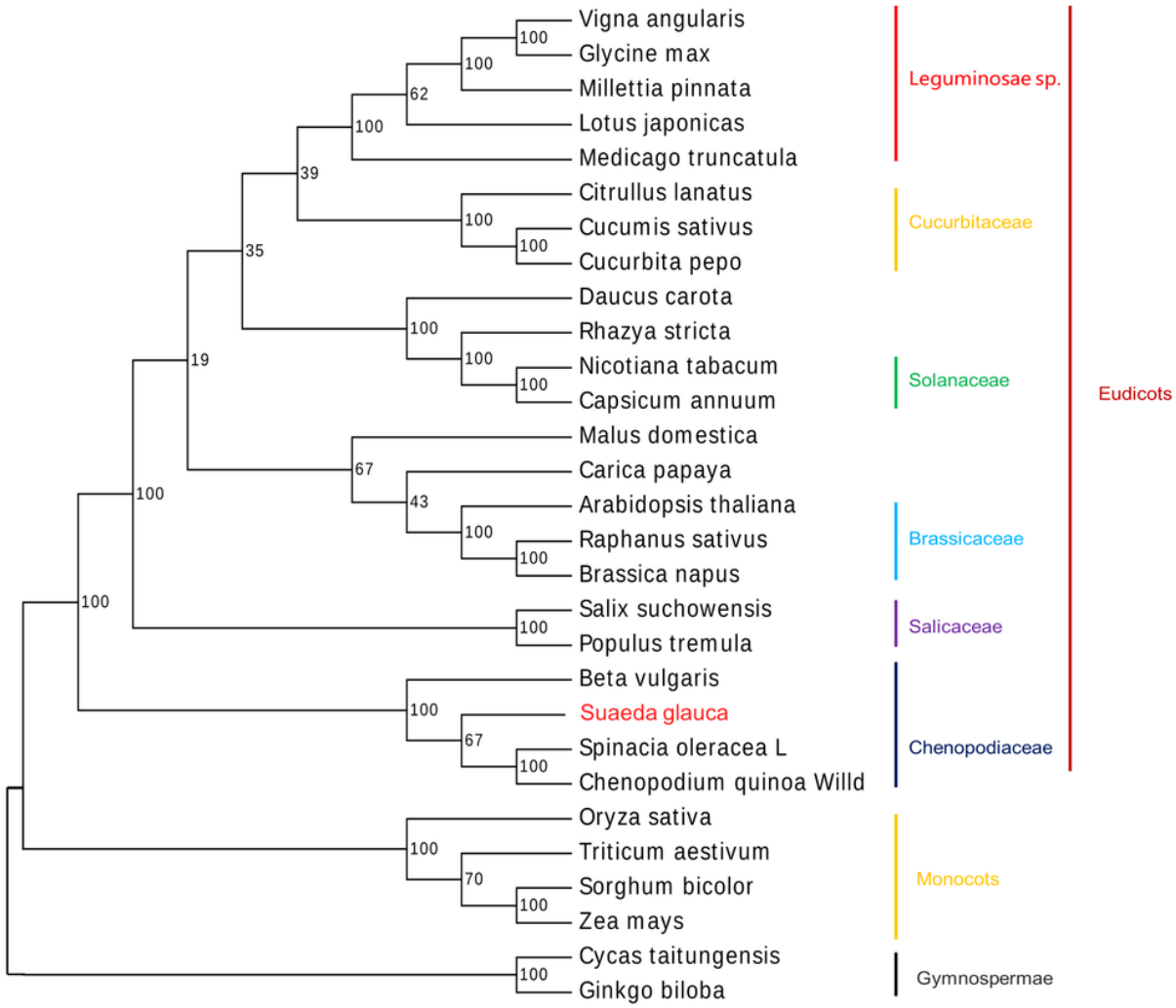

Figure 4

Neighbor-Joining tree was constructed based on 23 conserved genes of 29 representative plant mt genomes.

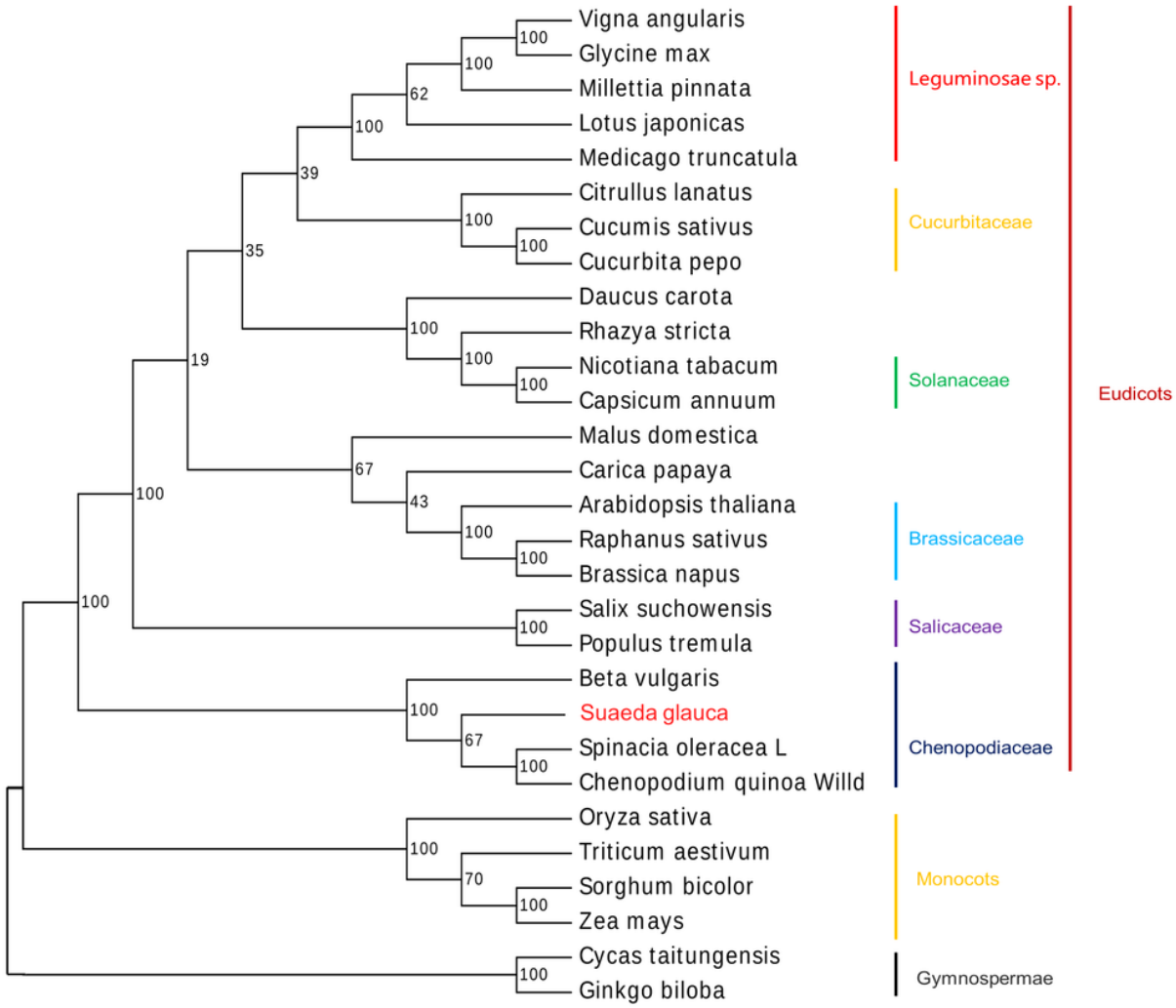

Figure 4

Neighbor-Joining tree was constructed based on 23 conserved genes of 29 representative plant mt genomes.

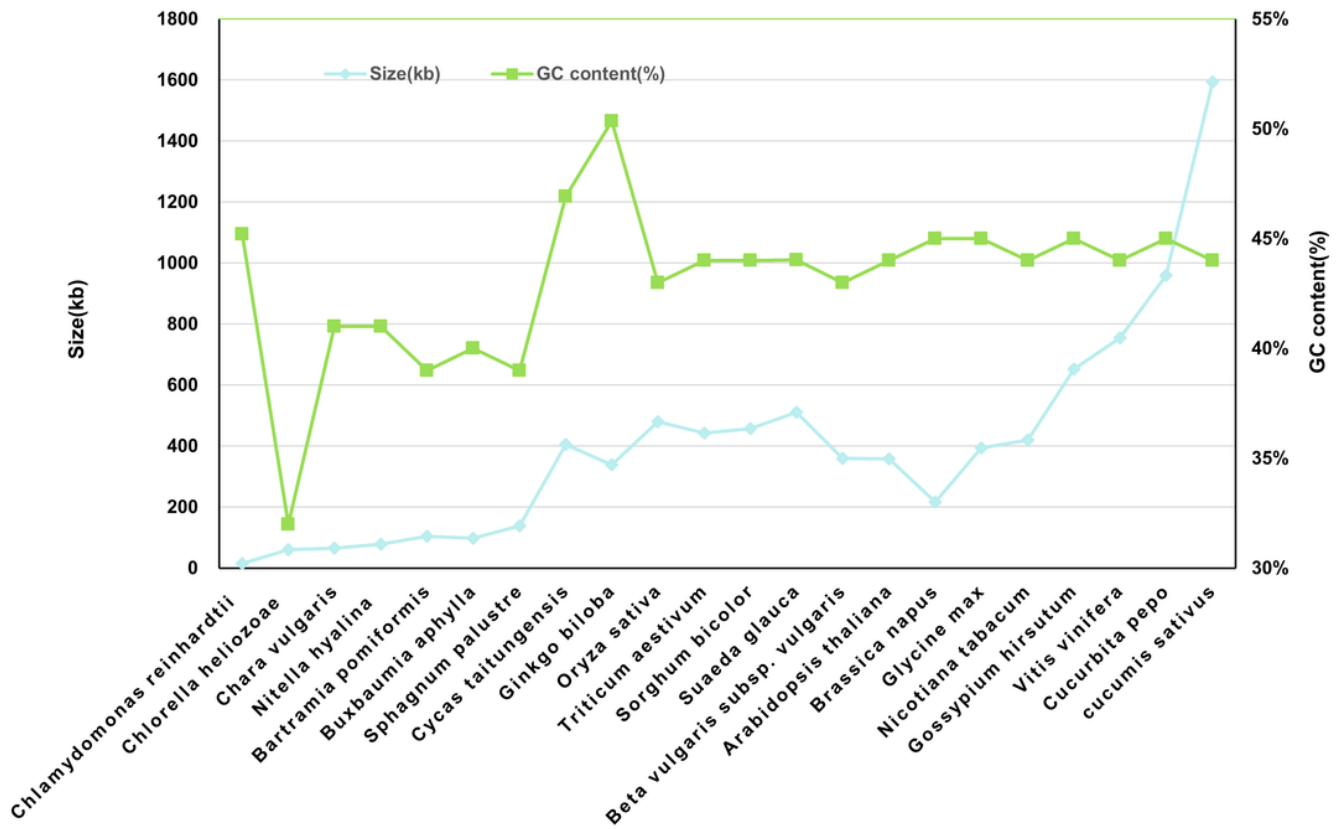

Figure 5

Comparison of sizes and GC contents within 22 plant mt genomes.

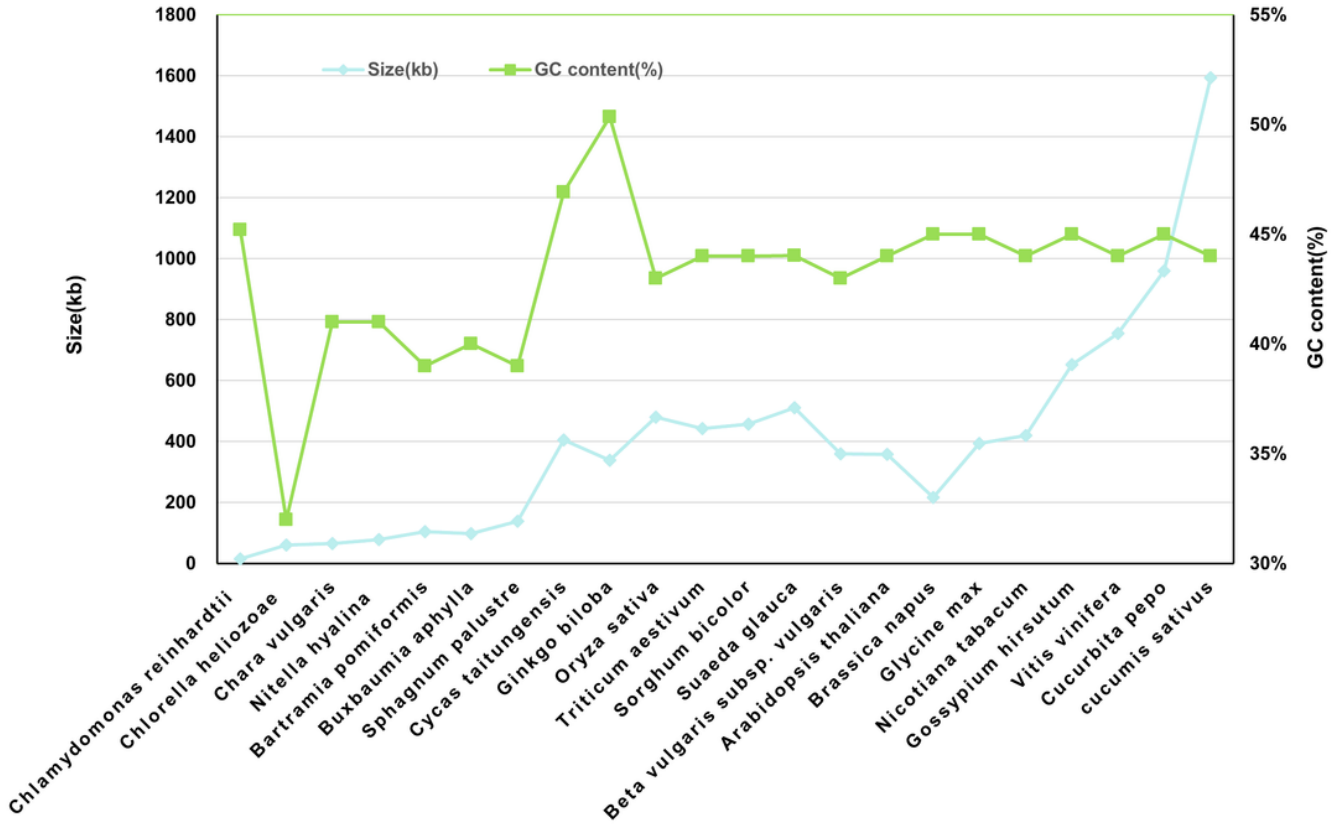

Figure 5

Comparison of sizes and GC contents within 22 plant mt genomes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Sgmitochondria.gb.docx
- Sgmitochondria.gb.docx
- SupplementaryFiguresandTables20201012.doc
- SupplementaryFiguresandTables20201012.doc