Assembly and comparative analysis of the complete mitochondrial genome of Saussurea inversa (Asteraceae)

Wubin Dai (897624419@qq.com)
Qinghai University https://orcid.org/0009-0001-6435-2887
Xiuting Ju
Qinghai University https://orcid.org/0000-0001-7086-0653
Guomin Shi
Qinghai University
Tao He
Qinghai University https://orcid.org/0000-0002-2154-4029

Research Article

Keywords: Saussurea inversa, Mitochondrial genome, Comparative analysis

Posted Date: December 19th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3696705/v1

Abstract

Saussurea inversa is a perennial herb used in traditional Chinese medicine and is effective against rheumatoid arthritis. In this study, we sequenced the complete mitochondrial(mt) genome of *S. inversa* (GenBank accession number: ON584565.1). The circular mt genome of *S. inversa* was 335,372 bp in length, containing 62 genes, including 33 mRNAs, 22 tRNAs, six rRNAs, and one pseudo-gene, with 1,626 open reading frames. The GC content was 45.14%. Predictive analysis of substantial RNA-editing of its presence found that the most abundant RNA editing is *ccmFn*, with 36. Gene migration was observed to occur between the mt and chloroplast(cp) genomes of *S. inversa* via the detection of homologous gene fragments. Phylogenetic analysis revealed that *S. inversa* was clustered with *Arctium tomentosum* (Asteraceae). Our findings provide extensive information regarding the mt genome of *S. inversa* and help lay the foundation for future studies on the genetic variations, phylogeny, and breeding of *S. inversa* via an analysis of the mt genome.

INTRODUCTION

Saussurea inversa (Raab-Straube, 2011) is a perennial herb from the Asteraceae family, mainly found in the Qinghai-Tibet Plateau. It grows in alpine limestone flats at an altitude of 4700–5400 m (Shi et al., 2011) and is a typical alpine plant well-adapted to the extreme environment(Liu et al., 2006). It is a traditional Chinese medicine used for rheumatoid arthritis (Wang et al., 2009). There is limited literature on *S. inversa*, particularly in genomics research (Wang *et al.*, 2021). Mitochondria play a significant role in eukaryotic cells (Nass *et al.*, 1963). This study aimed to analyze the complete mt genome of *S. inversa* to provide a scientific basis for understanding the adaptability of alpine plants to extreme environments and help lay the foundation for future studies on the genetic variations, phylogeny, and breeding of *S. inversa*.

MATERIALS AND METHODS

In this study, fresh leaves of *S. inversa* were collected from the Daban Mountain (101°40'19" E, 37°35'21" N), a branch of the Qilian Mountain in the northeast of the Qinghai-Tibet Plateau, at an altitude of 4000 m. A specimen was deposited at the Plateau Plant Laboratory, College of Eco-Environmental Engineering, Qinghai University, Xining, China (Wubin Dai, szk05@qhu.edu.cn), under the voucher number DWB-2022-SH0615 (Fig. 1).

Second-generation sequencing, including sample quality detection, library construction, library quality detection, and library sequencing process, and sequencing was performed using the Illumina Novaseq 6000 platform (Illumina), according to the standard protocol provided by the manufacturer. The resulting readable raw data were uploaded to GenBank, BioProject, and BioSample, with the SRA numbers PRJNA863745, SAMN30061603, and SRR20747203, respectively.

The sequencing data were assembled using Canu (https://canu.readthedocs.io/), following the quality control. Annotations were performed for the mt genome using ORFF (www.ncbi.nlm.nih.gov/orffinder), tRNA using tRNAscanSE (Chan *et al.*, 2019), and encoding protein and rRNA using BLAST (blast.ncbi.nlm. nih.gov), RNA-editing sites were predicted using the PREP Suite (Mower, 2009). The mt genome map of *S. inversa* and structures of the genes that were difficult to annotate were prepared and analyzed using OGVIEW (http://www.1kmpg.cn/pmgmap; Liu et al., 2023).

Conserved CDS sequences of eight Asteraceae, two Rosaceae, and one Leguminosae (outgroup) were selected to draw a maximum likelihood evolutionary tree. Those sequences were compared with multiple sequences using MAFFT software (v7.427, Auto mode), and the compared sequences were joined head to tail, trimmed with trimAl (v1.4.rev15), and the model prediction was carried out using the Jmodeltest-2.1.10 software after the trimming to determine that the model

was of the GTR type. The Bootstrap test was set to 1000 times, and support values in percent are shown at nodes (Ju et al., 2021).

RESULTS

Genomic features of the S. inversa mt genome. The mt genome map showed that the circular mt genome of *S. inversa* was 335,372 bp in length. There were 1626 open reading frames containing 62 known genes, including 22 tRNAs, 6 rRNAs, and 1 pseudogene. Among the 62 known genes, 33 mRNAs were included, there are ATP synthases(*atp1, atp4, atp6, atp8, atp9*), Cytochrome c biogenesis(*ccmB, ccmC, ccmFc, ccmFn*), Ubiquinol cytochrome c reductase(*cob*), Cytochrome c oxidase(*cox1, cox2, cox3*), Maturases(*matR*), Transport membrane protein(*mttB*), NADH dehydrogenase(*nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9*), Ribosomal proteins (LSU:*rpl10, rpl16, rpl5*; SSU:*rps12, rps13, rps3, rps4*), Succinate dehydrogenase(*sdh4*), It is worth noting that *nad4L* and *nad5* were both two copies, and there were eight genes got introns (*ccmFc*, cox2*, nad1**** nad2****, nad4*, nad5****, nad7****, rps3**, '***' represents the number of introns). There are seven cis-splicing genes which were *ccmFc, cox2, nad4, nad7, rps3, trnS-GCT, trnT-TGT*, and three trans-splicing genes which were *nad1, nad2, nad5*. The minimum, maximum, and average depth of coverage of the assembled mt genome was 9, 4,227, and 76, respectively (Figure S1).

Table 1
Functional classification of genes and physical location of the *S. inversa* mt genome

Group of genes	Gene name	Length	Start codon	Stop codon	Amino acid
ATP synthase	atp1	1533	ATG	TAA	511
	atp4	576	ATG	TAA	192
	atp6	774	ACG(ATG)	TAA	258
	atp8	480	ATG	TAA	160
	atp9	261	ATG	TAA	87
Cytohrome c biogenesis	ccmB	621	ATG	TGA	207
	ccmC	753	ATG	TGA	251
	ccmFc	1320	ATG	TAA	440
	ccmFn	1719	ATG	TAG	573
Ubichinol cytochrome c reductase	cob	1176	ATG	TGA	392
Cytochrome c oxidase	cox1	1584	ATG	TAA	528
	cox2	834	ATG	TAG	278
	cox3	798	ATG	TGA	266
Maturases	matR	1968	ATG	TAG	656
Transport membrance protein	mttB	801	ATT	TAG	267
NADH dehydrogenase	nad1	978	ATG	TAA	326
	nad2	1467	ATG	TAA	489
	nad3	357	ATG	TAA	119
	nad4	1488	ATG	TGA	496
	nad4L	273	ATG	TAA	91
	nad4L	273	ATG	TAA	91
	nad5	2013	ATG	TAA	671
	nad5	2013	ATG	TAA	671
	nad6	729	ATG	TAA	243
	nad7	1185	ATG	TAG	395
	nad9	573	ATG	TAA	191
Ribosomal proteins (LSU)	rpl10	489	ATG	TAA	163
	rpl16	423	ATG	TGA	141
	rpl5	564	ATG	TAA	188
Ribosomal proteins (SSU)	rps12	378	ATG	TGA	126
	rps13	351	ATG	TGA	117

Group of genes	Gene name	Length	Start codon	Stop codon	Amino acid
	rps3	1710	ATG	TAG	570
	rps4	1086	ATG	TAG	362
Succinate dehydrogenase	sdh4	387	ATG	CGA(TGA)	129
Ribosomal RNAs	rrn18	1937			
	rrn18	1937			
	rrn26	3619			
	rrn26	3619			
	rrn5	117			
	rrn5	117			
Transfer RNAs	trnC-GCA	71			
	trnD-GTC	74			
	trnE-TTC	72			
	trnF-GAA	74			
	trnG-GCC	73			
	trnH-ATG	69			
	trnH-GTG	74			
	trnK-TTT	73			
	trnM-CAT	73			
	trnM-CAT	74			
	trnM-CAT	73			
	trnN-GTT	72			
	trnP-TGG	74			
	trnP-TGG	75			
	trnQ-TTG	72			
	trnS-GCT	71			
	trnS-GCT	71			
	trnS-TGA	87			
	trnT-GGT	74			
	trnT-TGT	73			
	trnW-CCA	74			
	trnY-GTA	83			

Prediction of RNA editing sites. RNA-editing is a generic term comprising a variety of processes that alter the DNA-encoded sequence of a transcribed RNA by inserting, deleting, or modifying nucleotides in the transcript(Mower, 2009). There are large amounts of RNA-editing for the genes listed above and the number of RNA-editing for each gene was shown in Fig. 3, the most abundant RNA editing is *ccmFn*, with 36, and the least was *atp6*, with only one. After RNA editing, the hydrophobicity of 44.67% of amino acids remained unchanged, while 8.05% of hydrophobic amino acids became hydrophilic, and 47.28% of hydrophilic amino acids became hydrophobic(Table 2).

Table 2
Prediction of RNA editing sites

Type	RNA-editing	Number	Percentage
hydrophilic-hydrophilic	CAC (H) = > TAC (Y)	6	
	CAT (H) = > TAT (Y)	17	
	CGC (R) = > TGC (C)	10	
	CGT (R) = > TGT (C)	28	
	total	61	12.27%
hydrophilic-hydrophobic	ACA (T) = > ATA (I)	3	
	ACC (T) = > ATC (I)	2	
	ACG (T) = > ATG (M)	5	
	ACT (T) = > ATT (I)	4	
	CGG (R) = > TGG (W)	33	
	TCA (S) = > TTA (L)	69	
	TCC (S) = > TTC (F)	27	
	TCG (S) = > TTG (L)	40	
	TCT (S) = > TTT (F)	52	
	total	235	47.28%
hydrophilic-stop	CAG (Q) = > TAG (X)	2	
	CGA (R) = > TGA (X)	2	
	total	4	0.80%
hydrophobic-hydrophilic	CCA (P) = > TCA (S)	7	
	CCC (P) = > TCC (S)	11	
	CCG (P) = > TCG (S)	6	
	CCT (P) = > TCT (S)	16	
	total	40	8.05%
hydrophobic-hydrophobic	CCA (P) = > CTA (L)	49	
	CCC (P) = > CTC (L)	7	
	CCC (P) = > TTC (F)	6	
	CCG (P) = > CTG (L)	36	
	CCT (P) = > CTT (L)	19	
	CCT (P) = > TTT (F)	12	
	CTC (L) = > TTC (F)	6	
	CTT (L) = > TTT (F)	15	

Type	RNA-editing	Number	Percentage
	GCG (A) = > GTG (V)	4	
	GCT (A) = > GTT (V)	3	
	total	157	31.59%
	All	497	100%

Phylogenetic analysis. To understand the process of evolution of the *S. inversa* mt genome, this article conducted a phylogenetic analysis of the *S. inversa* mt genome and the published mt genomes of 12 plants. The clustering in the phylogenetic tree is consistent with the relationships of these species at the family and genus levels, indicating that mt genome-based clustering results are reliable. Based on the phylogenetic tree, results were obtained. The phylogenetic tree (Fig. 4) revealed that *S. inversa* was closely related to *Arctium tomentosum* (Asteraceae). The reliability of support values of each node was above 88%, indicating that the evolutionary tree can show the genetic distances between the listed species well.

Repeat sequence analysis. Interspersed repeat sequences are repetitive sequences that are scattered in the genome. In the *S. inversa* mt genome, we identified a total of 206 interspersed repeats with a length greater than or equal to 30 bp; of these, 112 were forward repeats, and 88 were palindrome repeats. The length of the longest forward repeat sequence was 16,421bp and that of the longest palindrome repeat sequence was 518bp. The distribution of the lengths of forward repeats and palindrome repeats is shown in Fig. 5; the abundance of both types of repeats was the highest when repeats were in the range of 30-39bp.

Analysis of codon composition. Due to codon concatenation, each amino acid corresponds to a minimum of one codon and a maximum of six codons. There is a great variation in the rate of codon usage in the genomes of different species and organisms. This inequality in synonymous codon usage is called Relative Synonymous Codon Usage (RSCU). This preference is believed to be the combined result of natural selection, species mutation, and genetic drift. We used a self-coded Perl script to analyze the codon composition of the *S.inversa* mt genome. The results are shown in Table 3, the number of codons in all coding genes was 9,881, and the relative RSCU in the S.inversa mt genome is shown in Fig. 6. There were 31 codons with RSCU > 1, indicating that the usage frequency of these codons is greater than that of other synonymous codons. Among these, 28 codons ending with the A/T base were identified, and these accounted for 90.32% of the codons, indicating that frequently used codons tend to end with the A/T base.

Analysis of homologous fragments of mitochondria and chloroplast. Using the BLAST tool, we screened the fragments of the S.inversa mt genome and cp genome(Wang *et al.*, 2021) exhibiting > 70% similarity, and performed homologous fragment analysis (Fig. 7). We screened out 15 homologous fragments with a total length of 9,078 bp, which accounted for 2.70% of the mt genome (Table 3). These homologous fragments contained 14 annotated genes, of which 6 were tRNA genes, namely, tRNA-CCA, tRNA-UGG,tRNA-GUC, tRNA-GUG, tRNA-GUU, tRNA-CAU, and others were rrn16, psaB, psaA, ycf2, petL, petG, rbcL, rps14.

Table 3
Comparison of a homologous fragment in the *S.inversa* cp genome to that in the mt genome

Matches (%)	length	gap	mt start	mt end	cp start	cp end	genes
99.922	2556	1	36401	38955	307643	310198	psaB(partical:98.64%);psaA(partical:15.76%)
91.348	601	5	87770	88358	5938	5347	ycf2(partical:8.57%)
91.348	601	5	147195	147783	5347	5938	ycf2(partical:8.57%)
80.285	913	40	65237	66122	190138	191003	petL;petG;tRNA-CCA;tRNA-UGG
73.874	888	39	99807	100670	235198	234340	rrn16(partical:57.99%)
73.874	888	39	134883	135746	234340	235198	rrn16(partical:57.99%)
73.874	888	39	99807	100670	303750	302892	rrn16(partical:57.99%)
73.874	888	39	134883	135746	302892	303750	rrn16(partical:57.99%)
93.939	198	3	55269	55461	207773	207576	rbcL(partical:13.52%)
89.163	203	0	35948	36150	77007	76805	rps14(partical:61.52%)
90.476	126	1	11104	11226	162819	162944	tRNA-GUC
97.531	81	0	2	82	86537	86617	tRNA-GUG
94.048	84	1	127807	127889	316935	316852	tRNA-GUU
94.048	84	1	107664	107746	316852	316935	tRNA-GUU
94.937	79	0	51506	51584	326025	326103	tRNA-CAU

DISCUSSION

Mitochondria have a relatively independent genetic transcription system (Pan *et al.*, 2005). However, the mt genome not only contains its genes but also some genes from the cp genome, they have multiple types of repetitive sequences and relatively conserved coding sequences. This suggests that these genes may have migrated during evolution, and further investigation is needed to understand the specific mechanism. Rapid advances in genome sequencing technology have accelerated the study of the mt genome. This study describes the basic features of the mt genome of *Saussurea inversa* for the first time, which provides an important foundation for understanding the function, inheritance, and evolutionary trajectory of the mt genome. The mt genome of the *Saussurea inversa* is a circular sequence with a length of 335,372 bp and a GC content of 45.14%. We performed BLAST analysis of the mt sequence and annotated the sequence using software, and found 33 protein-coding genes, 22 tRNA genes, 4 rRNA genes, and 1, 626 orf in the mt genome. Since sequence duplications can lead to intermolecular recombination in mitochondria, it is particularly important to perform repetitive sequence analysis, and we analyzed the dispersed repetitive sequences of the mt genome of Saussurea inversa. The results showed that a total of 206 dispersed repeat sequences with lengths greater than or equal to 30 bp were identified, the length of the longest forward repeat sequence was 16,421 bp and that of the longest palindrome repeat sequence was 518 bp.

RNA editing occurs in the post-transcriptional processes of the cp and mt genomes of higher plants and can alter genetic information at the mRNA level, thus making protein folding easier. In this study, 497 RNA editing sites were identified in 34 coding genes of *S. inversa* genome with 29 codon shift types. Among the codon transfer types, TCA = > TTA was the most common, which was consistent with the findings of Qiao (Qiao et al., 2022), with 69 editing sites. After RNA editing, 8.12%

of hydrophobic amino acids became hydrophilic and 47.28% of hydrophilic amino acids became hydrophobic. Consistent results existed in the *Bupleurum chinense* DC. mt genome, where the most abundant transfer type in the plant was TCA = > TTA, number 78, which was edited to change the hydrophobicity of more than half of the amino acids (Qiao et al., 2022). The selection of editing sites in the genome of the *S. inversa* was strongly biased, with all editing sites being C-T edited, which is the most common type of editing in plant mt genomes. In the previous version of the study, it was shown that RNA editing occurring at the second position of the codon accounted for more than half of the total number of edits. In the *S. inversa* mt genome, 78.87% of the editing sites were also located at the second base of the triplet codon, which is consistent with previous findings. In addition, after RNA editing, the encoded amino acids were partially converted into termination codons (TAG, TGA). In the mt genome of the *S. inversa*, 0.80% of the amino acids were edited into termination codons, resulting in an early stop of the coding process, which altered the function of the gene.

Plant DNA transfer between organellar and nuclear genomes and between species occurs frequently, and sequencing analyses have identified many DNA transfer events between different plant genomes (mitochondrial, nuclear, and chloroplast). Previous studies have found that DNA transfer events are mainly the transfer of DNA fragments from the organelle genome to the nuclear genome, followed by the transfer of nuclear and plastid genomes to the mt genome. In the present study, we found that the total length of homologous fragments transferred from the cp genome to the mt genome of the *S. inversa* was 9,078 bp, and these homologous fragments contained eight annotated genes, six of which were tRNA genes. This result is similar to that of Ma (Ma et al. 2022) who found that the fragment of *A. truncatum* cp genome transferred to the mt genome contained 6 integrative genes, 5 of which were tRNA genes. The percentage of transferred fragments in the mt genome of the *S. inversa* was 2.7%, which was similar to the data previously reported for *Acer truncatum* (2.36%)(Ma et al., 2022)and *Salix suchowensis* (2.8%)(Ye et al., 2017), but lower than that of *Suaeda glauca* (5.18%)(Cheng et al., 2021).

We analyzed the codons in the *S. inversa* mt genome, and the RSCU value reflects the ratio of the actual frequency of use of a codon to the theoretical frequency of use in the absence of usage bias; if RSCU = 1 it means that there is no bias in the use of the codon, and if RSCU < 1 it means that the actual frequency of use of the codon is lower than that of the other synonymous codons, and conversely it is higher than the usage frequency of other synonymous codons(Sharp *et al.*, 1986). The results of the analysis showed that there were 31 codons with RSCU > 1, indicating that the usage frequency of these codons is greater than that of other synonymous codons. Among these, 28 codons ending with the A/T base were identified, and these accounted for 90.32% of the codons, indicating that frequently used codons tend to end with the A/T base.

CONCLUSIONS

In this study, the mt genome of *S. inversa* was sequenced, assembled, and annotated, and the DNA and amino acid sequences of annotated genes were analyzed. The mt genome of the *S. inversa* is circular and has a total length of 335,372 bp. 62 genes were annotated in the mt genome, including 33 protein-coding genes, 22 tRNA genes, and 6 rRNA genes. We analyzed repeat sequences, RNA editing processes, and codon preferences in the mt genome of the *S. inversa*. We observed gene transfer between the mt and cp genomes of the *S. inversa* by examining homologous fragments. In addition, our results showed that although the size of plant mt genomes varies greatly, their GC content is relatively conserved during evolution, suggesting that the mt gene is conserved during evolution. This study provides extensive information about the mt genome of the golden monkey. Importantly, this study lays the foundation for future studies on genetic variation, phylogeny, and plateau adaptation of the *S. inversa* using the mt genome.

Declarations

Acknowledgement:

We thank to all listed authors for their valuable help. Wubin Dai and Tao He has substantial contributions to the conception or design of the work, Xiuting Ju and Guomin Shi draft the work or review it critically for important intellectual content and final approval of the version to be published. All of the authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding acknowledgement statement:

This work was supported by The National Natural Science Foundation of China [31960222].

References

1. Chan P P, Lowe T M. 2019. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol*. 1-14.
2. Chang S, Yang T, Du T, *et al.* 2011. Mitochondrial genome sequencing helps show the evolutionary mechanism of mitochondrial genome formation in Brassica. *BMC Genomics*. 12: 497.
3. Chang S, Wang Y, Lu J, *et al.* 2013. The mitochondrial genome of soybean reveals complex genome structures and gene evolution at intercellular and phylogenetic levels. *PLoS One.* 8(2): e56502.
4. Cheng Y, He X. 2021. Priyadarshani SVGN, Wang Y, Ye L, Shi C, *et al.* Assembly and comparative analysis of the complete mitochondrial genome of *Suaeda glauca*. *BMC Genomics*.22(1):167.
5. Eckhard V R. 2011.The genus Saussurea (Compositae, Cardueae) in China: taxonomic and nomenclatural notes. *Willdenowia.* 41(1): 83-95.
6. Handa H. 2003. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (Brassica napus L.): comparative analysis of the mitochondrial genomes of rapeseed and Arabidopsis thaliana. *Nucleic Acids Res.* 31(20):5907-16.
7. Ju X T, Shi G M, Chen S R, *et al.* 2021. Characterization and phylogenetic analysis of the complete chloroplast genome of *Tulipa patens* (Liliaceae). *Mitochondrial DNA. Part B*: *Resources.* 6(9).
8. Koren S, Walenz B P, Berlin K, *et al.* 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*. 27: 722-736.
9. Liu J Q, Wang Y J, Wang A L, *et al*. 2006. Radiation and diversification within the Ligularia–Cremanthodium–Parasenecio complex (Asteraceae) triggered by the uplift of the Qinghai-Tibetan Plateau. *Mol Phylogenet Evol.* 38(1): 31–49.
10. Liu S, Ni Y, Li J, *et al.* 2023. CPGView: a package for visualizing detailed chloroplast genome structures. *Mol Ecol Resour*. 23(3):694–704
11. Luo Z P, Pan LW. 2019. Characterization of the complete mitochondrial genome of Ageratum conyzoides. *Mitochondrial DNA. Part B*. 4(2):3540-3541.
12. Ma Q, Wang Y, Li S, *et al.* 2022. Assembly and comparative analysis of the frst complete mitochondrial genome of Acer truncatum Bunge: a woody oiltree species producing ervonic acid. *BMC Plant Biol*. 22(1):29.
13. Mower J P. 2009. The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res*.37.
14. Nass M M, Nass S. 1963. Intramitochondrial fibers with DNA characteristics I. Fixation and electron staining reactions. *Journal of Cell Biology*. 19(3): 593.
15. Pan B P, Bu W J. 2005. Research progress on inheritance and evolution of the mitochondrial genome. *Chinese journal of biology*. 40(8): 1-3.
16. Qiao Y G,Zhang X R,Li Z, *et al*. 2022. Assembly and comparative analysis of the complete mitochondrial genome of Bupleurum chinense DC. *BMC Genomic*s, (1)664-664.

17. Reineke A, Karlovsky P, Zebitz C. 1998. Preparation and purification of DNA from insects for AFLP analysis. *Insect Mol Biol*. 7(1): 95–99.
18. Sharp P M, Tuohy T M. 1986. Mosurski K R. Codon usage in yeast: cluster analysis clearly diferentiates highly and lowly expressed genes. *Nucleic Acids Res*.14(13):5125–43.
19. Shi Z, Chen Y L, Chen Y S, *et al.* 2011. Flora of China (Asteraceae). *Missouri Botanical Garden Press*. 62: 20-21.
20. Wang Y J, Susanna A, Von R E, *et al.* 2009. Island-like radiation of Saussurea (Asteraceae, Cardueae) triggered by uplifts of the Qinghai-Tibetan Plateau. *Biol J Linnean Soc*. 97(4): 893–903.
21. Ye N, Wang X, Li J, *et al*. 2017. Assembly and comparative analysis of complete mitochondrial genome sequence of an economic plant *Salix suchowensis*. *Peer J*. 5(1):e3148.

Figures

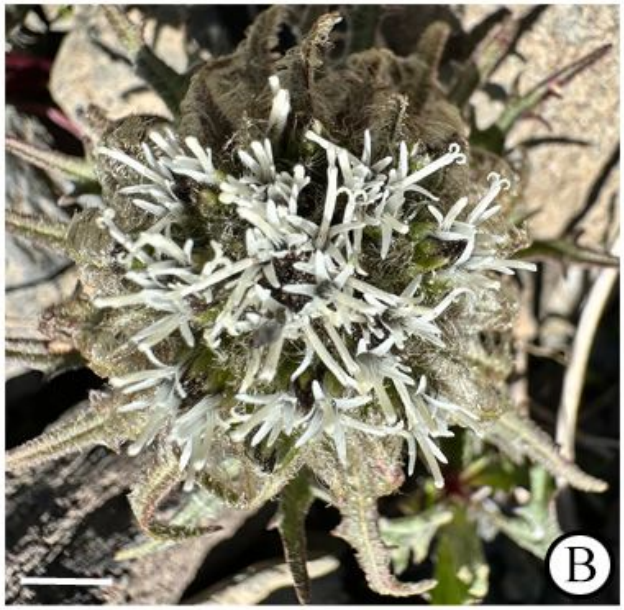

Figure 1

Reference image of *S. inversa.* (A) The specimen of *S. inversa* was made by Wubin Dai. (B) The inflorescence (C) Acrial part. Scale bar =1 cm. (B) and (C) were taken by Wubin Dai at Daban Mountain, China (101°40'19"E, 37°35'21"N).

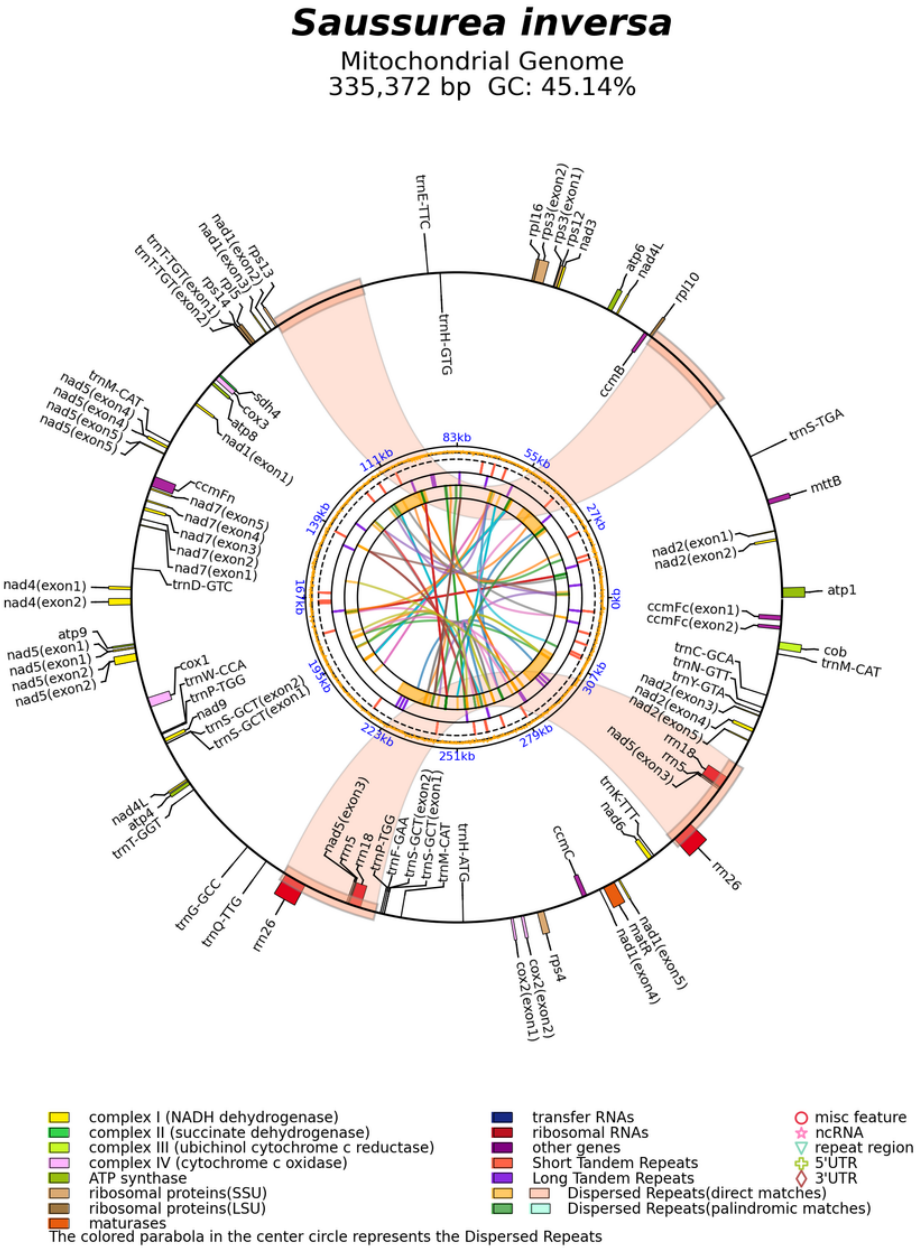

Figure 2

Mt genome map of *S. inversa*. The forward coding gene is on the outside of the circle and the reverse coding gene is inside the circle. The colored parabola in the center circle represents the Dispersed Repeats.

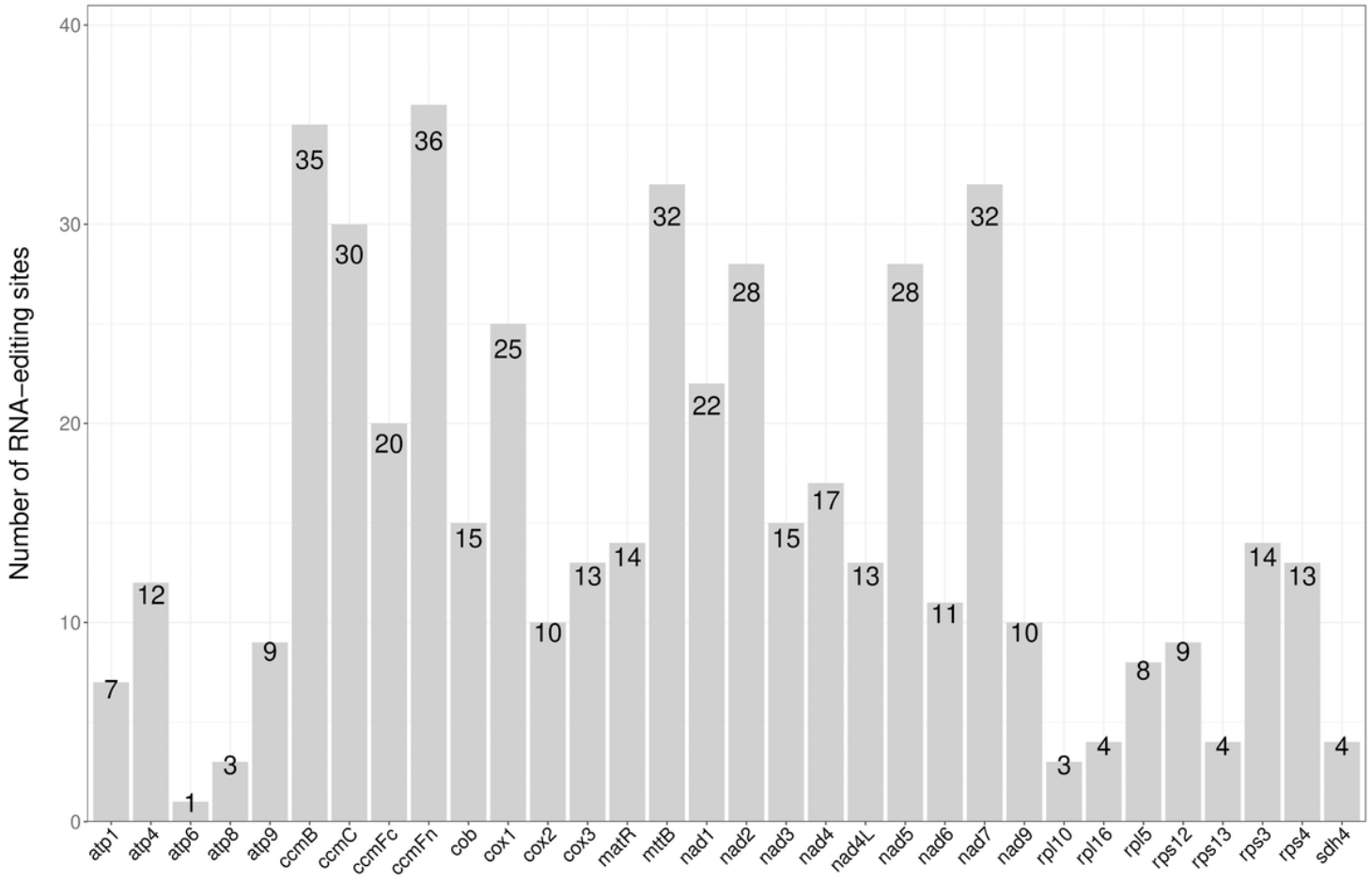

Figure 3

The number of RNA-editing sites.

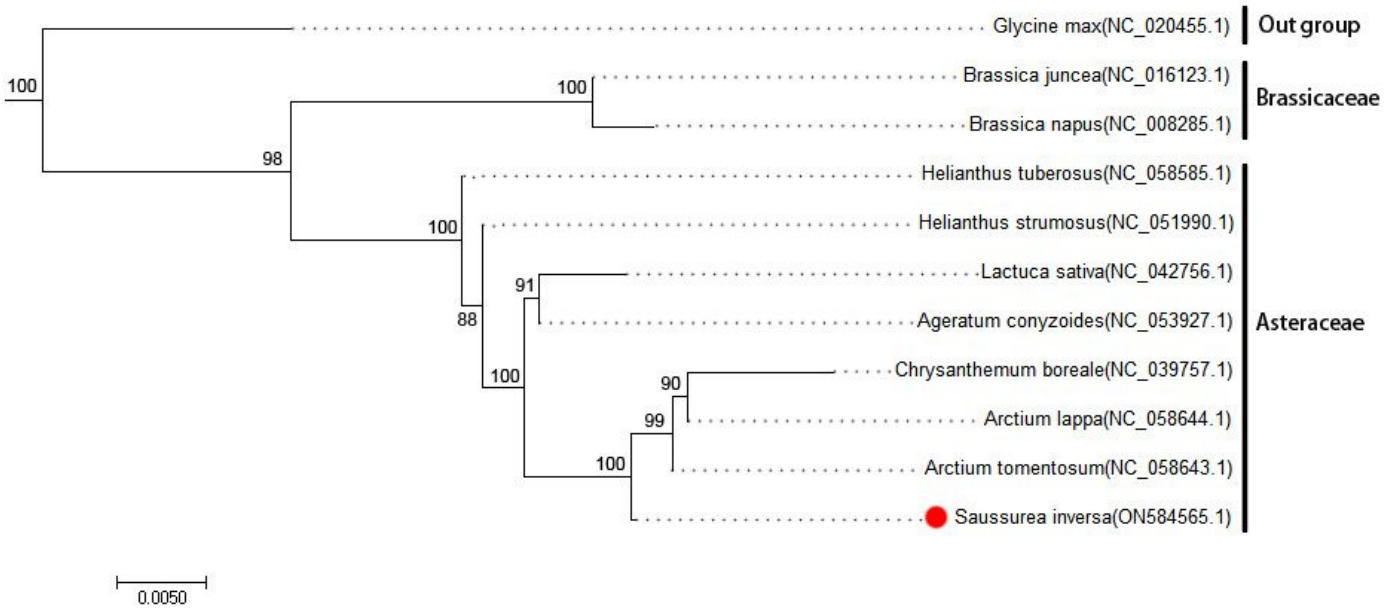

Figure 4

Phylogenetic tree of *S. inversa*. The following sequences were used to establish the phylogenetic tree: *G. max* NC_020455.1 (Chang *et al.*, 2013), *B. juncea* NC_016123.1 (Chang *et al.*, 2011), *B. napus* NC_008285.1 (Handa *et al.*, 2003), *H. tuberosus* NC_058585.1, *H. strumous* NC_051990.1, *L. sativa* NC_042756.1, *A. conyzoides* NC_053927.1 (Luo *et al.*,2019), *C. boreale* NC_039757.1, *A. lappa* NC_058644.1, *A. tomentosum* NC_058643.1. Among them, *G. max* as an outgroup, *B. juncea* and *B. napus* as two species of Brassicaceae, they both clustered separately into a cluster. The GenBank accession numbers of each species are shown in parentheses. Bootstrap support values in percent are shown at nodes.

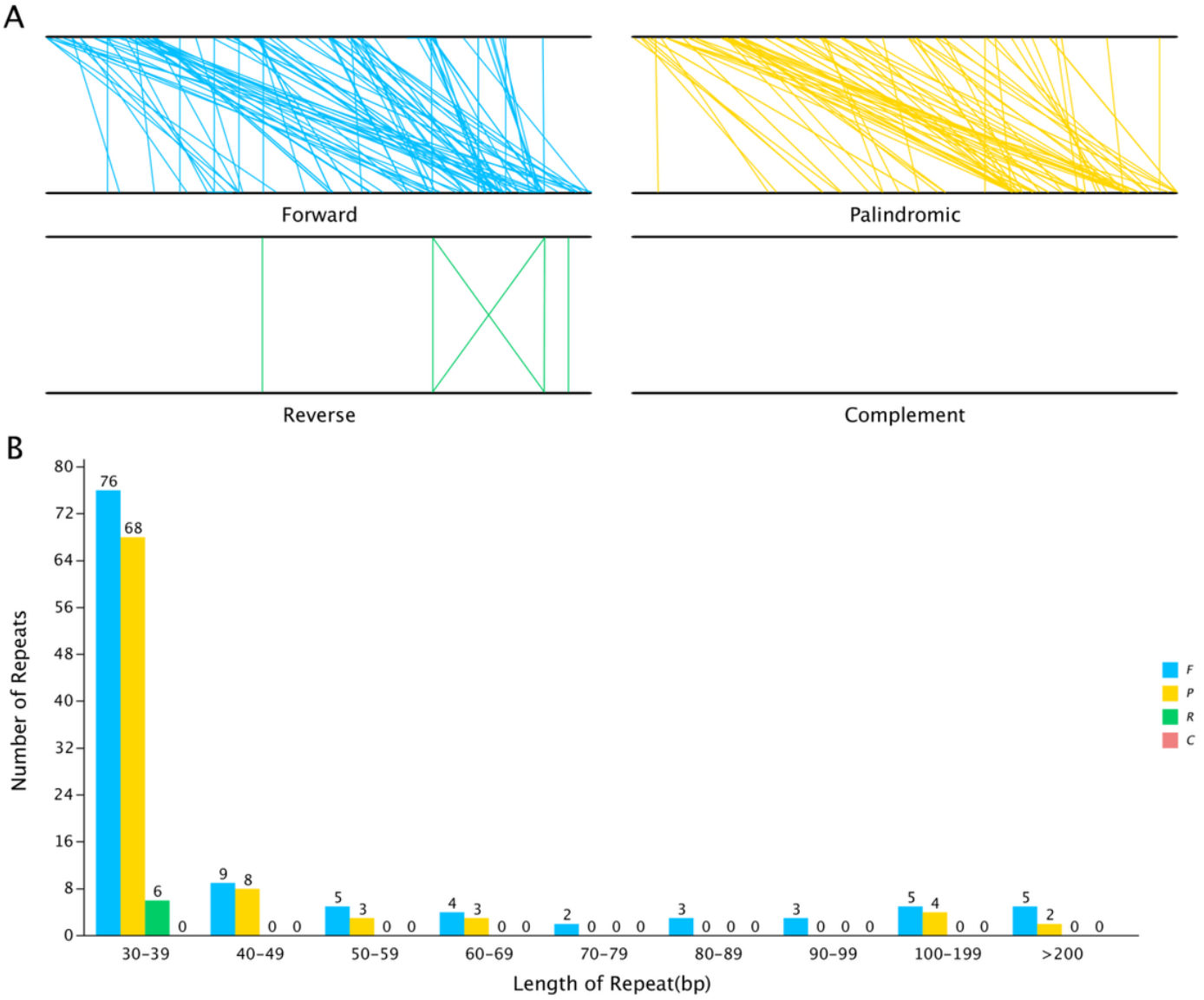

Figure 5

The interspersed repeat sequences in the Saussurea inversa mt genome. **A** The four interspersed repeat types are distributed throughout the genome; the two black lines represent the mt genome, and the same repeats are associated with the line segments. **B** Distribution of lengths of interspersed repeats in the mt genome. The abscissa indicates the type of interspersed repeats, and the ordinate indicates the number of scattered repeats.

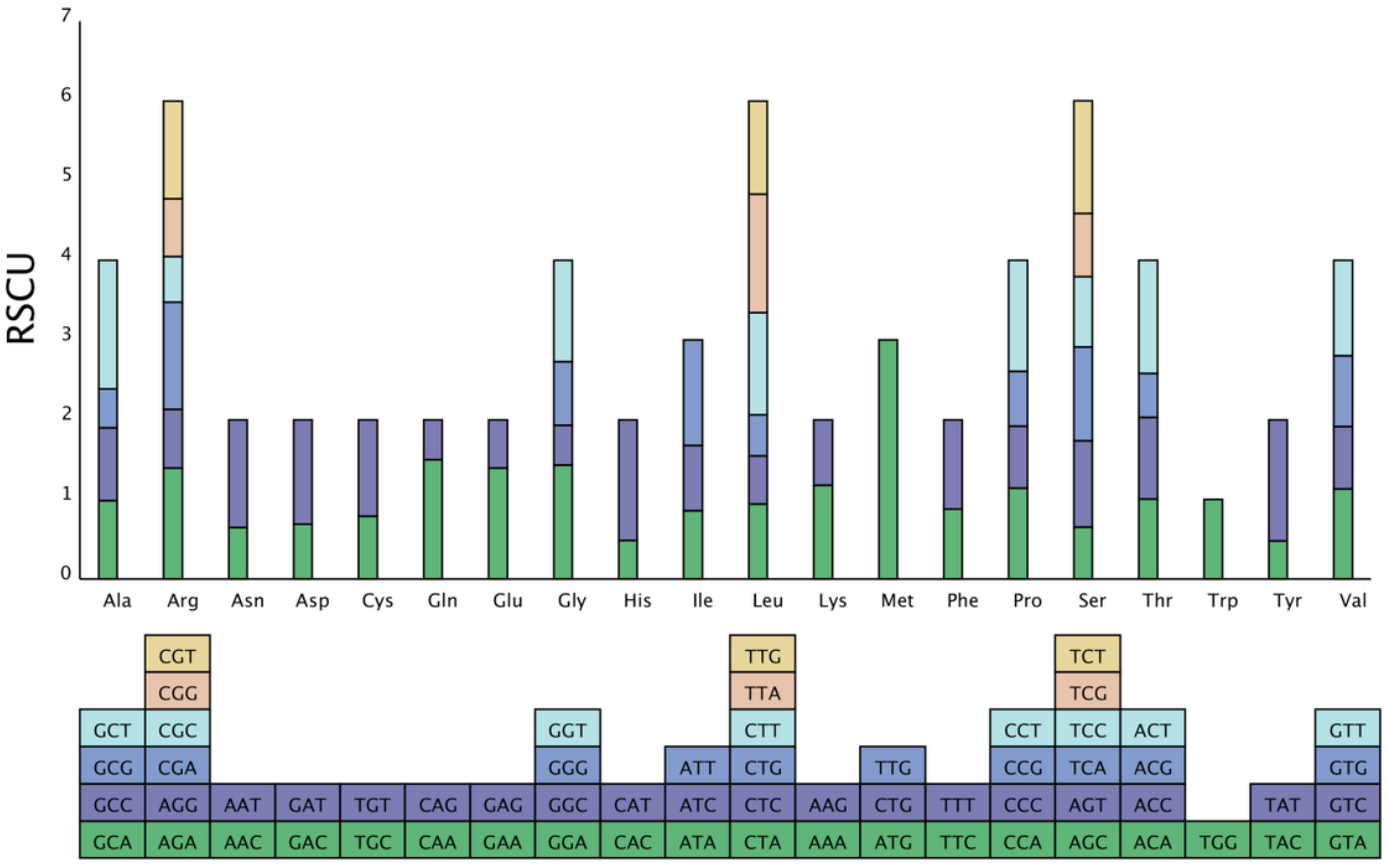

Figure 6

RSCU in the *S.inversa* mt genome. The different amino acids are shown on the x-axis. RSCU values are the number of times a particular codon is observed relative to the number of times that codon would be expected for a uniform synonymous codon usage.

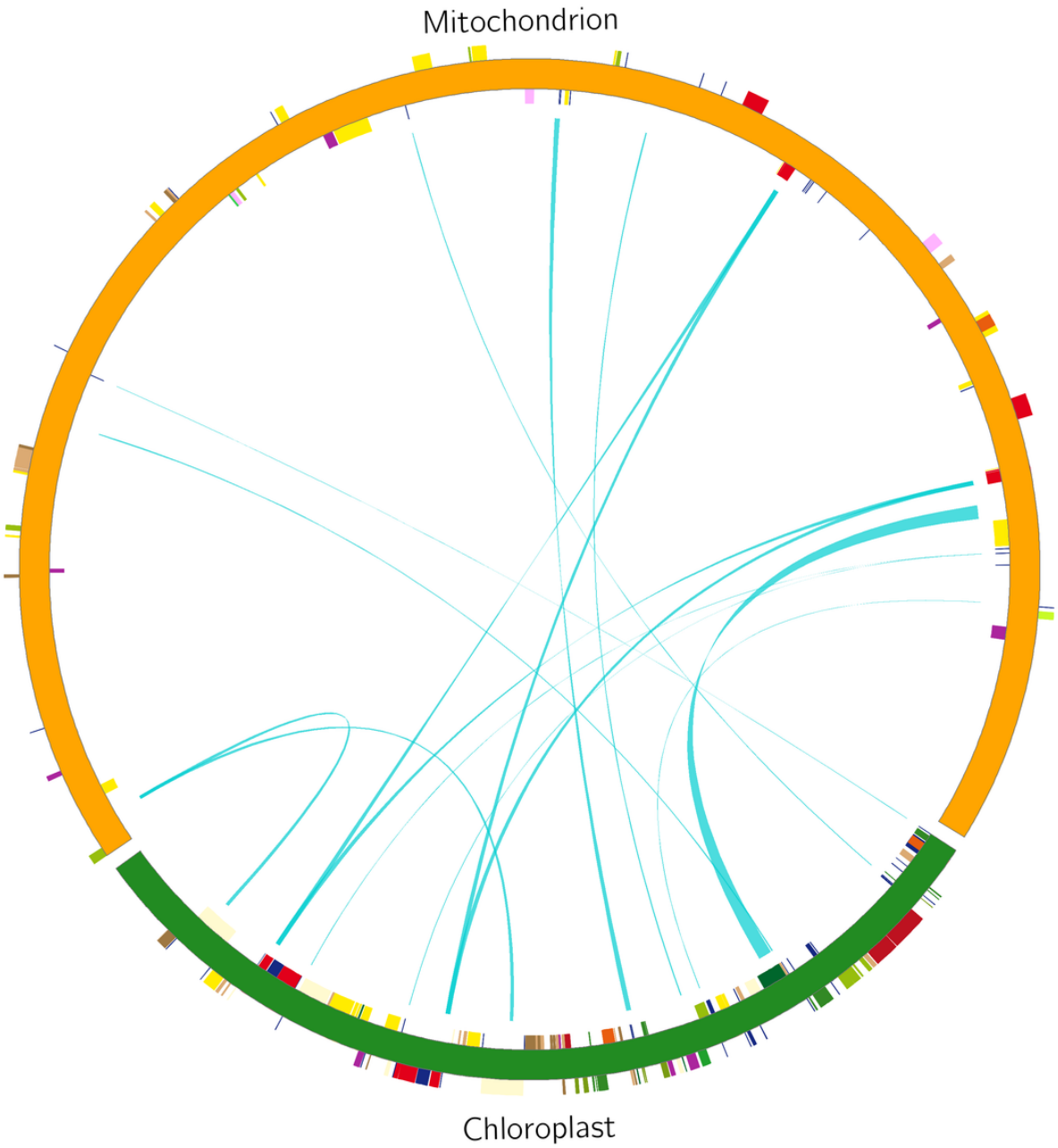

Figure 7

Comparison of a homologous fragment in the *S.inversa* cp genome to that in the mt genome. The green line segment of the circle represents the *S. inversa* cp genome, and the red line segment represents the *S. inversa* mt genome. The green line segment in the circle connects the start and end points of the transferred gene fragments. The width of the green line segment represents the size of the transferred fragment.

Supplementary Files

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

- Fig.S1.tif