

A high-quality mitochondrial genome of *Suriana maritima* reveals tremendous gene transfers between the plastid and mitochondrial genomes

Jian-Xin Li

Jishou University

Runxian Yu

University of Chinese Academy of Sciences

Ze-Long Nie

Jishou University

Meng hua Zhang

Jishou University

Miao-Miao Shi

Chinese Academy of Sciences

Zhongtao Zhao

Chinese Academy of Sciences

Shijin Li

Chinese Academy of Sciences

Xiang-Ping Wang

Chinese Academy of Sciences

Ying Meng

Jishou University

Tieyao Tu

tut.tieyao@scbg.ac.cn

Chinese Academy of Sciences

Dianxiang Zhang

Chinese Academy of Sciences

Research Article

Keywords: Gene transfer, Fabales, mitochondrial genome, *Suriana maritima*, Surianaceae

Posted Date: February 20th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-3948491/v1>

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

Additional Declarations: No competing interests reported.

Abstract

Background

Gene transfer between the plastid and mitochondrial genomes has been reported in several lineages of the legume family (Leguminosae or Fabaceae). However, it is not clear whether these events happened in other families of the order Fabales. We herein generated a high-quality mitochondrial genome of *Suriana maritima* in the family Surianaceae, which is closely related to Leguminosae and distributed mainly in extreme environments of tropical coral islands, to understand the gene transfers between organelle genomes of *S. maritima* and the structural and functional evolution of mitogenomes in Fabales.

Results

We combined Illumina and Nanopore technologies to assemble the mitogenome of *S. maritima*. The mitogenome was 458,738 bp in length and contained 39 protein-coding, three ribosomal RNA, and 20 transfer RNA genes. A number of sequences derived from the chloroplast genome of *S. maritima* were detected in the mitogenome, including 41 plastid genes among 83 mitochondrial plastid DNA sequences with a total length of 30,834 bp. Transferred genes accounted for 18.8% of the plastid genome and 6.7% of the mitogenome, and these estimates were much higher than those in previous studies of Fabales.

Conclusions

In addition to providing a high-quality mitogenome of an additional species in Fabales and the first species in Surianaceae, *S. maritima*, our findings advance our understanding of gene transfer between organelle genomes.

Background

The plant mitogenome is mainly assembled into a circular structure containing all genes [1], generally known as the “master circle”, with a few exceptions as linear, branched, overlapping, or multiple coexisting structures [2–5]. Redundant sequences and extensive genomic recombination make it difficult to recover the conformation of the plant mitogenome. In particular, it is difficult to correctly assemble the mitogenome based on second-generation sequencing data and accurately evaluate mitogenome evolution, including codon usage, RNA editing, horizontal transfer, gene loss and acquisition [6]. Among these, intracellular gene transfer has been found in several lineages of plants. For example, the genome transferred from plastid to mitochondrion span 1.67 kb in *Silene* (Caryophyllaceae) and 3.52 kb in *Vigna radiata* (Fabaceae) [7]. The total length of *Bupleurum chinense* (Apiaceae) sequence transferred from the plastid genome to the mitogenome is 11,144 bp, accounting for 2.56% of the

mitogenome [8]. However, analyses of gene transfer between the plastid, mitochondrial, or nuclear genomes in a broader range of taxa are needed, including species closely related to legumes are lacking.

Fabales contains four families, namely Leguminosae, Polygallaceae, Quillajaceae and Surianaceae [10]. Surianaceae is a small family with only five woody genera and eight species distributed in pantropical areas, and, in particular, in Australia and Mexico [11, 12]. *Suriana maritima* L. is the member of Surianaceae and shows with an extensive distribution in the tropical coral islands and coastal zones. In China, it is only found on the Xisha Islands and Dongsha Islands (coral islands) in the South China Sea, mainly growing on the sandy land or rock cracks, as a pioneer species [13, 14].

In this study, we assembled a high-quality mitogenome of *S. maritima* using both Illumina and Nanopore technologies and analyzed its genomic and structural features as well as the role of gene transfer from plastids. This study provides new resources for evolutionary studies in Fabales and improves our general understanding of mitogenome evolution.

Results

Features of the *Suriana maritima* mitogenome

In total, 5.6 Gb of Illumina reads and 43.31 Gb of nanopore clean reads were generated. The average read length of nanopore was 29,915 bp, the read N50 was 36,836 bp, and the longest read was 227,816 bp. The mitogenome of *Suriana maritima* contained a circular contig with 458,738 bp in length, including 39 protein-coding genes, 20 transfer RNA (tRNA) genes, and three ribosomal RNA (rRNA) genes (Fig. 1). The average GC content of the *S. maritima* mitogenome was 43.5%. The mitogenome contained 24 core and 15 variable genes (Table 1). Ten genes contained introns, among which *cox1*, *rpl2*, *rps3*, and *ccmFc* had one intron; *cox2* had two introns; *nad4* had three introns; and *nad1*, *nad2*, *nad5*, and *nad7* had four introns respectively (Table 1).

Table 1
Composition of the mitogenome of *Suriana maritima*

Group of genes	Genes
ATP synthase	<i>atp1, atp4, atp6, atp8, atp9</i>
Cytochrome C biogenesis	<i>ccmB, ccmC, ccmFc*, ccmFn</i>
Ubichinol Cytochrome C reductase	<i>cob</i>
Cytochrome C oxidase	<i>cox1*, cox2**, cox3</i>
Maturases	<i>matR</i>
Transport membrane protein	<i>mttB</i>
NADH dehydrogenase	<i>nad1****, nad2****, nad3, nad4***, nad4L, nad5****, nad6, nad7****, nad9</i>
Large subunit of ribosome	<i>rpl2*, rpl5, rpl10, rpl16</i>
Small subunit of ribosome	<i>rps1, rps3*, rps4, rps7, rps10, rps12, rps13, rps14, rps19</i>
Succinate dehydrogenase	<i>sdh3, sdh4</i>
Ribosomal RNAs	<i>rrn5, rrn18, rrn26</i>
Transfer RNA	<i>trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnFM-CAU, trnG-GCC, trnH-GUG, trnI-CAU, trnK-UUU, trnL-CAA, trnM-CAU, trnN-GUU, trnP-UGG, trnQ-UUG, trnS-GCU, trnS-GGA, trnS-UGA, trnV-GAC, trnW-CCA, trnY-GUA</i>

Note: Gene* indicates that the gene had intron, * indicates the number of introns

Repeat sequence analysis

Simple sequence repeats (SSRs), known as microsatellite sequences, are tandem repeats, consisting of several nucleotides (usually 1–6) up to several tens of nucleotides [15]. A total of 318 SSR sequences were identified in the mitogenome of *S. maritima*. Mononucleotide A/T repeats were the most prevalent with 199 mononucleotide repeats accounting for 63% of the total SSRs. Additionally, 32 dinucleotide repeats, accounted for 10% of the total SSRs, while trinucleotide and hexanucleotide repeats were less numerous and occurred only in intergenic or intronic regions. There were 16 trinucleotide repeats, 56 tetranucleotide repeats, 56 tetranucleotide repeats, 11 pentanucleotide repeats and 4 hexanucleotide repeats, accounting for 5%, 18%, 3% and 1% of the total SSRs, respectively (Fig. 2). Repeats of each base in SSRs (Fig. 3) were also counted. A and T repeats were the most frequent, appearing 86 times, followed

by C (15 times), G (12 times), and AT (10 times). TA and AT had similar counts (11). Repetitions of the remaining base were all below 5, and there were 69 types of base repeats, and A and T repeated 8 times was frequently detected. There were two types of dispersed repeats in the mitogenome, forward repeats (28) and palindrome repeats (22). Thirty-three long tandem repeats (LTRs), were identified in the mitogenome of *Suriana maritima* (Table S1). The size of these tandem repeat units, which ranged from 5 bp to 27 bp in length. The longest duplicate unit was 114 bp (13,865–13,979), the shortest was only 26 bp (414,564–414,590).

Codon usage bias

The codon usage statistics for protein-coding genes (PCGs) are shown in Table 5. A total of 10,963 codons are used for 39 PCGs in the mitogenome of *Suriana maritima*. There exist AT mutation bias in the coding genes given that the base A and T were used in high frequency in the third-codon position compared to other bases (C v.s. G). The most frequent codon was UUU (Phe), which appeared 403 times, followed by AAA (Lys) with 293 times. The least used CGC encoding arginine (Arg) appeared 65 times, followed by GGC encoding glycine (Gly) (79 times). The three stop codons UAA, UGA, and UAG were recorded 158, 181, and 207 times in the whole mitogenome, respectively, demonstrating a preference for UAG.

The Relative synonymous codon usage (RSCU) is usually used to describe the codon usage preferences of each amino acid. RSCU was visually evaluated using ggplot2 [16] in R, as shown in Table S2 and Fig. 4. The codons for methionine (Met), tryptophan (Trp), serine (Ser), and arginine (Arg) showed no preference (RSCU = 1). Their codon counterparts are AUG (also the start codon), UGG, UCC, and CGG. There were 29 codons with RSCU > 1 (including a termination codon UAG), indicating a preference for these codons (Fig. 4 and Table S2). For example, arginine had a strong preference for AGA (RSCU = 1.49) and glycine for GGA (RSCU = 1.44).

Gene transfer between mitochondrial and plastid genomes

Using BLAST v2.10.1 to screen fragments of the mitogenome and plastid genomes of *Suriana maritima* showing over 80% identity and performed a homology analysis [17]. we found tremendous genetic residues of plastids in the mitogenome, which means that sequence transfer is quite common between the two organelle genomes. We identified a total of 83 homologous sequences (Table S3), of which 30 were longer than 1000 bp and the longest was 4379 bp. As determined using TBtools [18] (Fig. 5), we detected substantial collinearity between fragments on mitochondria and plastids. Using the plastid genome as a reference for annotation and found identified 41 plastid genes in mitochondria (Table S4), including 28 CDS and 13 tRNA, with a total of 30,834 bp of homologous sequences.

The homologous sequences accounted for 18.8% of the plastid genome and 6.7% of the mitogenome, indicating that the plastid genome of *Suriana maritima* is a rich source of foreign sequences in the mitogenome. In particular, 16 plastid genes (*ycf3*, *psaA*, *psaB*, *matK*, *atpB*, *atpl*, *rps2*, *rps4*, *atpF*, *rps7*, *rpl23*, *atpH*, *petN*, *trnC-GCA*, *trnI-CAU*, and *trnS-GGA*) have been completely transferred into the mitogenome (Fig. 6).

Mapping the 29 genes transferred from the plastids to mitochondria with the transcriptome data, 20 genes can distinguish between mitochondria and plastid genome, the highest expression is in 24 transcripts, and the average expression of normal plastid genes and mitochondrial genes is 100, so it is assumed that the expression of the genes that have been transferred is low (Table S7).

Discussion

Assembly of a high-quality mitogenome of *Suriana maritima*

Compared with plastid genome assembly, the assembly of the mitochondrial genome is more difficult. Firstly, there is no effective method to enrich plant mitochondria before DNA extraction [19]. Second, the mitochondrial structure is extremely diverse, and the plant mitogenome shows substantial variation in size, with frequent structural rearrangements, even within closely related species [20, 21]. For example, in the family Cucurbitaceae, the mitogenome of *Citrullus lanatus* is 379 kb [20], while that of *Cucumis sativus* is 2.74 Mb [22], which is seven times the size of the former. Even greater variation has been detected within the genus *Silene*, with estimates ranging from 253 kb to 11 Mb [23]. Several technical limitations using the data of second genome sequencing may affect the quality of mitogenome assembly as well. Therefore, it is challenging to ascertain whether an assembly is complete or not.

In this study the mitogenome was assembled from total DNAs with a hybrid assembly strategy, combining the unitig sequences assembled from short Illumina reads and contigs assembled from long Nanopore reads. This strategy can avoid false positives caused by the polishing strategy. During assembly of the *S. maritima* mitogenome, the depth at several sites determined by mapping to Illumina reads was relatively low (data not shown). This may be explained by assembly errors or may reflect differences among tissue types, as has been found in other plant species (Fig. 7). We thus performed error correction with the second-generation data (Fig. 8); for two loci with low depth we used PCR and Sanger sequencing to obtain a high-quality mitogenome. We obtained the first complete mitogenome for *S. maritima* in *Suriana* (family Surianaceae), which is expected to provide a useful reference for follow-up studies of the function and structure of plant mitochondria. Bioinformatic predictions must be further validated by quantitative PCR experiments, Sanger sequencing, Southern blot, and electronic microscopy [24, 25].

Structural characteristics of the *Suriana maritima* mitogenome

The mitogenome of *S. maritima* is a circular structure (Fig. 1). The structure of plant mitogenomes shows great instability [26], and various structures have been reported, including linear, branching, and polychromosomal structures. The mitogenome of the garden rocket contains six main ring structures and four small subring molecules, which may be due to the complex structure of the repetitive sequences [27]. A multichromosomal mitogenome structure has been verified in various angiosperms, including *Cucumis sativus*, *Silene gallica*, *Amborella trichopoda*, and parasitic plants [5, 28–30]. However, the typical mitogenome structure consists of ring molecules [31], as observed in *Leucaena trichandra* [32], and *Dalbergia odorifera* [9].

Repeat sequence-mediated homologous recombination is universal in plant mitogenomes [33, 34]. The large number of repeats (> 1000 bp) may lead to a more complex structure, which can also explain why there are fewer genes in the mitogenome than the plastid genome, despite its much larger length. In the assembly of the *S. maritima* mitogenome, we also identified repetitive sequences suggesting substantial recombination (Fig. 2). Plant mitogenomes are rich in repetitive sequences, accounting for up to 38% of the genome size [26]; however, the origin of repetitive sequences remains unclear. There were 318 simple repeats, 33 long tandem repeats, and 50 dispersed repeats in *S. maritima* (Fig. 2, Table S1). Based on plant mitogenome sequences available to date, the content and composition of plant mitochondrial repeat sequences differ substantially [35], indicating that these sequences were obtained independently during evolution [36]. Therefore, repeated sequences in the plant mitogenome are closely related to recombination [31], resulting in a highly complex plant mitochondrial structure and causing male abortion in plants.

Mitochondrial functional genes are highly conserved [37], and the majority of genomic variation is found in intergenic regions [38]. Intergenic sequences are characterized by a large number of repeated sequences and imported sequences [30] as well as intron length variation and pseudogene factors [39]. Repetitive sequences appear in highly variable intergenic regions [31], frequent structural rearrangements [40], substantial gene loss, and the transfer of endogenous and exogenous DNA [41]. The long repeats in the mitogenome also had a high frequency of recombination. Short repeats had a low recombination frequency, consistent with previous results [42].

In addition, mitochondrial PCGs have strong codon preference. UUU encoding phenylalanine (Phe) is the most frequently used codon, while the CGC encoding arginine (Arg) is the least codon in the PCGs in the mitogenomes of *S. maritima*.

Gene transfer between mitochondria and plastid genomes

Genes transferred from the plastid genome are commonly found in plant mitogenomes, typically representing 1–12% of the total length [43]. In contrast, the transfer of sequences from the plant mitogenome into plastid genomes is nearly impossible, since the plastid genome is relatively conserved and does not accept the integration of foreign DNA fragments [44]. The transfer of plastid DNA to the mitogenome occurred at least 300 million years ago [45]. No insertions of plastid and nuclear genes were found in the mitogenome of bryophytes, indicating that the absorption of foreign DNA by plant mitochondria occurred during the evolution of vascular plants [46].

Compared with previous studies of other taxa, the mitogenome of *S. maritima* had the highest rates of transfer from the plastid genome. Homologous sequences with a total length of 30,834 bp accounted for 18.8% of the plastid genome and 6.7% of the mitogenome of *S. maritima* (Table S3). The smaller mitogenome of *Citrullus lanatus* contains about 23 kb of transferred plastid sequences. More than 5% of the mitogenomes of *Cucurbita pepo* is derived from plastid sequences of (approximately 113 kb) [47]. Plastid-to-mitochondrion sequences are 1.67 kb in *Silene* and 3.52 kb in *Vigna radiata* [7]. In *Bupleurum*

chinense, the total length of sequences transferred from the plastid genome to the mitogenome was found to be 11,144 bp, accounting for 2.56% of the mitogenome [8].

Early studies have shown that plastid-derived tRNA genes are required for the translation of plant mitogenomes [23]. The plastid fragments that were transferred to mitochondria mostly include genes, pseudogenes, and intergenomic sequences [3]. Most functional genes are tRNA genes, while other genes are generally not expressed and many become pseudogenes gradually over time [48, 49]. Here, we identified 16 full-length genes that have been completely transferred from the plastid of *S. maritima* to the mitogenome (i.e., *ycf3*, *psaA*, *psaB*, *matK*, *atpB*, *atpI*, *rps2*, *rps4*, *atpF*, *rps7*, *rpl23*, *atpH*, *petN*, *trnC-GCA*, *trnI-CAU*, and *trnS-GGA*) (Fig. 6). However, these genes might not function in mitochondria, and they might undergo pseudogenization [50]. Pseudogenes mainly encode important proteins; however, complete plastid protein genes have also been found in *Vitis vinifera* [51]. For example, the *rpl32* gene from the plastid has been transferred into the subfamily Thalicthroideae (family Ranunculaceae)[52]. Three complete genes in the mitogenomes migrated from the plastid genome of okra, including *psaA*, *rps7*, and *psbJ*, in addition to several plastid-derived gene fragments [53].

In addition, we found that *S. maritima* has substantial differences in fragment position caused by the insertion of plastid genes into the mitogenome. Intracellular gene transfer may be responsible for the high degree of rearrangement [54]. Since the plastid genome fragments are well-aligned with the original plastid genome sequence, this gene transfer may cause the disassembly of the mitogenome. The mitochondrial and nuclear genomes can transfer DNA sequences in both directions. However, most studies have found that plastid genome fragments are only transferred to the nuclear and mitogenomes, while other genomes are not transferred to the plastid, and the plastid genome tends not to integrate foreign DNA [55, 56].

Conclusions

In this study, we successfully assembled the mitogenome of *S. maritima* with a high quality by using both Nanopore and Illumina reads, enabling a comprehensive comparison of the organelle genomes of *S. maritima*. A single circular master molecule represents the mitogenome of *S. maritima*. The chloroplast genome of *S. maritima* provided abundant foreign sequences in the mitogenome, including 41 plastid genes among 83 mitochondrial plastid DNA sequences with a total length of 30,834 bp. The transferred sequences accounted for 18.8% of the plastid genome and 6.7% of the mitogenome, which is much higher than previous estimates in Fabales. This study provided strong evidence for gene transfer between the mitochondria and plastids genes. Additional sampling from all families in the order are needed to more fully understand their complex history and the functional interplay between plastid and mitochondrial genetics.

Materials and methods

Taxon sampling, DNA extraction and sequencing

We collected *S. maritima* in Xisha Islands (China) and cultivated plants at the South China Botanical Garden, CAS. Total genomic DNA was extracted from fresh leaves using a modified cetyltrimethylammonium bromide (CTAB) method [57]. Second-generation sequencing was conducted using the Illumina HiSeq-2500 platform with a read length of 150 bp. Sequencing produced a total of 5.6 Gb of raw data. For Oxford Nanopore sequencing, purified DNA was prepared for long-read sequencing following the protocol provided with the SQK-LSK109 genomic sequencing kit (Oxford Nanopore Technologies, Oxford, UK). Extraction, database construction, and sequencing were completed by Beijing Baimaike Biotechnology Co., Ltd.]\

Mitogenome assembly

We initially assembled the mitogenome of *S. maritima* using Illumina short reads. GetOrganelle v1.7.7.0 was used for assembly with the parameter setting “embplant_mt” and a custom mitochondrial database as references [58]. The output contained 238 contigs of 276–89,145 bp in length, with a total length of 637,303 bp. When visualizing the raw GFA file produced by GetOrganelle in Bandage v0.8.1 [59], the mitogenome consisted of a network of connected contigs (Figure S1). Unnecessary contigs were manually deleted. After many rounds of verification, a complete closed-loop structure was obtained.

The published mitochondrial CDS sequences of *Leucaena trichandra* and *Vicia faba* were used as references (among which *Leucaena trichandra* contained 56 CDS sequences and *Vicia faba* contained 61 CDS sequences). Minimap2 [60] was used to identify putative mitochondrial Nanopore reads of *S. maritima*. Then, Canu [61] to perform self-correction and trimming of the identified putative mitochondrial Nanopore reads. The parameters are set to: genomeSize = 2m, corThreads = 64; For the assembly result, LoRDEC [62] was used to correct the reads twice and trim once, with the following parameter settings: -k = 19, -s = 3; The parameters for retrimming and assembly were genomeSize = 2m, minReadLength = 5000, and correctedErrorRate = 0.134 to obtain 11 contigs with lengths of 26,663–514,650 bp (Table S5).

The Canu assembly result showed that contig1 was “suggestCircular.” Considering that the length of this contig was the longest among the 11 contigs, it might represent the true mitogenome sequence of *S. maritima*. To further verify whether contig1 is a complete mitogenome, Illumina reads (5.6 Gb) were mapped to the 41 genes (24 core genes and 17 variable genes) of angiosperm mitogenomes using Bowtie2 [63], and the mapped reads were extracted using Samtools v1.9 [64] with the parameter -bF 4, generating a much smaller subset of Illumina mitochondrial reads. We then calculated their depth of coverage using Samtools. It was found that 39 genes had a depth of coverage of > 5000×, except *rps2* and *rps11* (2–16 for *rps2* and *rps11*) (Table S6). These 39 genes except *rps2* and *rps11* exist in the mitogenome of *S. maritima*. The above 39 genes were used as a reference genome to annotate contig1, revealing that contig1 contained 39 genes except for *rps2* and *rps11*. Therefore, contig1 may contain the complete information of the mitogenome of *S. maritima*. Then, Pilon [65] was used for gap filling and sequence correction with default settings until its depth reached 5000–8000, and this was used in subsequent analyses.

Genome annotation

The web-based tool Public MITOFY Analysis was used to identify genes [20]. We also revised the start and stop codons of genes based on similar genes, manually checked by MITOFY against other sequenced mitogenomes. Mitogenome annotation was performed using the BLASTN alignment tools of the National Center for Biotechnology Information (NCBI) database (<https://www.ncbi.nlm.nih.gov/>) with the gene bank of angiosperm mitogenomes as reference sequences. The tRNAs were annotated using tRNAscan-SE [66] with default settings. The circular mitogenome map was drawn using OGDRAW [67]. The final mitogenome of *S. maritima* has been deposited in GenBank (Accession number: OQ933798).

Analysis of mitogenome characteristics

The GC content was determined using the cusp program provided by EMBOSS [68]. The SSRs of the mitogenome of *S. maritima* were analyzed using Misa [69] (<https://webblast.ipk-gatersleben.de/misa/>), with a size of one to six nucleotides, thresholds of eight, five, four, three, three, and three, separately, and a minimum distance between two SSRs of 100 bp. REPuter [70, 71] (<https://bibiserv.cebitec.uni-bielefeld.de/reputer/>) was used to calculate palindromic repeats, forward repeats, reverse repeats, and complement repeats with the following settings: hamming distance of three and minimal repeat size of 30 bp. LTR identification was performed using the "Advanced" module of Tandem Repeats Finder [72] (<https://tandem.bu.edu/trf/trf.html>) with default settings. Codon preference (synonymous codon usage, RSCU) of mitochondrial PCGs of the *S. maritima* mitogenome was calculated using CodonW.

Identification of plastid gene transfer to the mitogenome

The complete plastids genome of *S. maritima* (NC_047313.1) [73, 74] was downloaded from the NCBI Organelle Genome Resources database. Plastid gene transfer to the mitogenome were identified using BLAST [75], with the following parameter settings:- evaluate 1e-5, -word_size 9, -gapopen 5, -gapextend 2, -reward 2, -penalty-3, and -dust no [42]. Screening criteria were set with an identity threshold of $\geq 80\%$. The identification results were visualized with reference to collinearity analysis using the "Advanced Circos" module of TBtools [18].

Abbreviations

PCG
Protein-coding genes
tRNA
Transfer RNA
rRNA
Ribosomal RNA
Mitogenome
Mitochondrial genome
NCBI

National Center for Biotechnology Information
SSR
Simple sequence repeat
BLAST
Basic Local Alignment Search Tool
LTR
Long tandem repeats
RSCU
Relative Synonymous Codon Usage CDS: Coding sequence.

Declarations

Ethics approval and consent to participate

The material involved in the article does not involve ethical conflicts. Field studies complied with local legislation, and appropriate permissions were granted before the samples were collected from the South China Botanical Garden.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The complete mitogenome sequence of *Suriana. maritima* has been deposited in the GenBank database under the accession number OQ933798 (these numbers were automatically generated by NCBI and refer to the same sample). The associated BioProject, SRA, and Bio-Sample numbers are PRJxxxxx, SRRxxxx, and SAMNxxxxx, respectively.

Funding

This work was financially supported by the National Natural Science Foundation of China (32070222, 32170232), the National Key Research and Development Program of China (2021YFC3100405) and the Guangdong Provincial Special Fund for Natural Resource Affairs on Ecology and Forestry Construction (GDZZDC20228704).

Authors' contributions

In this research, Jian-Xin Li conceived and designed the study. She has completed the data analysis and the first draft of the paper. Runxian Yu helped assemble and annotate the mitogenome, Meng hua

Zhang, Miao-Miao Shi, Zhongtao Zhao, Shijin Li, Xiang-Ping Wang, Dianxiang Zhang, Ying Meng, Ze-Long Nie and Tie-Yao Tu have been responsible for supervising the experimental design, data analysis, and the writing and revision of the manuscript. All authors approved the final version and agreed to be accountable for all aspects of the work.

References

1. Gualberto, J.M. and K.J. Newton, Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annual Review of Plant Biology*, 2017. 68: p. 225-252.
2. Bendich, A.J., Reaching for the ring : the study of mitochondrial genome structure. *Current Genetics*, 1993. 24(4): p. 279-290.
3. Notsu, Y., et al., 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): p. 434-445.
4. Bendich, A.J., The size and form of chromosomes are constant in the nucleus, but highly variable in bacteria, mitochondria and chloroplasts. *Bioessays*, 2007. 29(5): p. 474-83.
5. Sloan, D.B., et al., Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLOS Biology*, 2012. 10(1): p. e1001241.
6. Kovar, L., et al., PacBio-based mitochondrial genome assembly of *Leucaena trichandra* (Leguminosae) and an intrageneric assessment of mitochondrial RNA editing. *Genome Biology and Evolution*, 2018. 10(9): p. 2501-2517.
7. Alverson, A.J., et al., The mitochondrial genome of the legume *Vigna radiata* and the analysis of recombination across short mitochondrial repeats. *PLoS One*, 2011. 6(1): p. e16404.
8. Qiao, Y., et al., Assembly and comparative analysis of the complete mitochondrial genome of *Bupleurum chinense* DC. *BMC Genomics*, 2022. 23(1): p. 664.
9. Hong, Z., et al., A complete mitochondrial genome for fragrant Chinese rosewood (*Dalbergia odorifera*, Fabaceae) with comparative analyses of genome structure and intergenomic sequence transfers. *BMC Genomics*, 2021. 22(1): p. 672.
10. APG IV, An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, 2016. 181: p. 1-20.
11. Crayn, D.M., et al., A reassessment of the familial affinity of the Mexican genus *Recchia* Mocino & Sesse ex DC. *Brittonia*, 1995. 47(4): p. 397-402.
12. Christenhusz, M.J.M. and J.W. Byng, The number of known plants species in the world and its annual increase. *Phytotaxa*, 2016. 261(3): p. 201-217.
13. Claxton, F., et al., Pollen morphology of families Quillajaceae and Surianaceae (Fabales). *Review of Palaeobotany and Palynology*, 2005. 133(3): p. 221-233.
14. Liu, J., et al., A karyological study of *Suriana maritima* L. (Surianaceae) from Xisha Islands of South China Sea. *Caryologia*, 2018. 71(2): p. 109-112.

15. Li, Q., et al., Development of genic SSR marker resources from RNA-seq data in *Camellia japonica* and their application in the genus *Camellia*. *Scientific Reports*, 2021. 11(1): p. 9919.
16. Wickham H, Ggplot2:Elegant Graphics for Data Analysis (Springer-Verlag). 2016.
17. Camacho, C., et al., BLAST plus: architecture and applications. *Bmc Bioinformatics*, 2009. 10: p. 421.
18. Chen, C., et al., TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, 2020. 13(8): p. 1194-1202.
19. Ni, Y., et al. Comparative analysis of the chloroplast and mitochondrial genomes of *Saposhnikovia divaricata* revealed the possible transfer of plastome repeat regions into the mitogenome. *BMC genomics*, 2022. 23, 570 DOI: 10.1186/s12864-022-08821-0.
20. Alverson, A.J., et al., Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Molecular Biology and Evolution*, 2010. 27(6): p. 1436-48.
21. Tang, M., et al., Rapid evolutionary divergence of *Gossypium barbadense* and *G. hirsutum* mitochondrial genomes. *BMC Genomics*, 2015. 16: p. 770.
22. Rodríguez-Moreno, L., et al., Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. *BMC Genomics*, 2011. 12: p. 424.
23. Sloan, D.B., et al., Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. *BMC Evolutionary Biology*, 2010. 10: p. 274.
24. Yu, R., et al., The minicircular and extremely heteroplasmic mitogenome of the holoparasitic plant *Rhopalocnemis phalloides*. *Curr Biol*, 2021.
25. Yang, H., et al., De Novo Hybrid Assembly of the *Salvia miltiorrhiza* Mitochondrial Genome Provides the First Evidence of the Multi-Chromosomal Mitochondrial DNA Structure of *Salvia* Species. *Int J Mol Sci*, 2022. 23(22).
26. Mower, J.P., et al., Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. *Genome Biology and Evolution*, 2012. 4(5): p. 670-86.
27. Weng, M.L., et al., Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Molecular Biology and Evolution*, 2014. 31(3): p. 645-659.
28. Alverson, A.J., et al., Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *Plant Cell*, 2011. 23(7): p. 2499-513.
29. Sanchez-Puerta, M.V., et al., Unparalleled replacement of native mitochondrial genes by foreign homologs in a holoparasitic plant. *New Phytologist*, 2017. 214(1): p. 376-387.
30. Rice, D.W., et al., Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science*, 2013. 342(6165): p. 1468-73.

31. Kitazaki, K. and T. Kubo, Cost of having the largest mitochondrial genome: evolutionary mechanism of plant mitochondrial genome. *Journal of Botany*, 2010. 2010: p. 620137.
32. Kovar, L., et al., PacBio-Based Mitochondrial Genome Assembly of *Leucaena trichandra* (Leguminosae) and an Intrgeneric Assessment of Mitochondrial RNA Editing. *Genome Biol Evol*, 2018. 10(9): p. 2501-2517.
33. Vasupalli, N., et al., Analysis of mitochondrial recombination in the male sterile *Brassica juncea* cybrid Og1 and identification of the molecular basis of fertility reversion. *Plant Molecular Biology*, 2021. 106(1-2): p. 109-122.
34. Wang, S., et al., Evolution and Diversification of Kiwifruit Mitogenomes through Extensive Whole-Genome Rearrangement and Mosaic Loss of Intergenic Sequences in a Highly Variable Region. *Genome Biol Evol*, 2019. 11(4): p. 1192-1206.
35. Gualberto, J.M., et al., The plant mitochondrial genome: dynamics and maintenance. *Biochimie*, 2014. 100: p. 107-20.
36. André, C., A. Levy, and V. Walbot, Small repeated sequences and the structure of plant mitochondrial genomes. *Trends Genet*, 1992. 8(4): p. 128-32.
37. Clifton, S.W., et al., Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiology*, 2004. 136(3): p. 3486-3503.
38. Cupp, J.D. and B.L. Nielsen, Minireview: DNA replication in plant mitochondria. *Mitochondrion*, 2014. 19 Pt B: p. 231-7.
39. Knoop, V., Plant mitochondrial genome peculiarities evolving in the earliest vascular plant lineages. *Journal of Systematics and Evolution*, 2013. 51.
40. Galtier, N., The intriguing evolutionary dynamics of plant mitochondrial DNA. *BMC Biology*, 2011. 9: p. 61.
41. Bock, R., The give-and-take of DNA: horizontal gene transfer in plants. *Trends in Plant Science*, 2010. 15(1): p. 11-22.
42. Li, J., et al., Assembly of the complete mitochondrial genome of an endemic plant, *Scutellaria tsinyunensis*, revealed the existence of two conformations generated by a repeat-mediated recombination. *Planta*, 2021. 254(2): p. 36.
43. Mower, J.P., D.B. Sloan, and A.J. Alverson, Plant mitochondrial genome diversity: the genomics revolution, in *Plant Genome Diversity Volume 1: Plant Genomes, their Residents, and their Evolutionary Dynamics*. 2012, Springer Vienna: Vienna. p. 123-144.
44. Leister, D., Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends in Genetics*, 2005. 21(12): p. 655-63.
45. Wang, D., et al., Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. *Molecular Biology and Evolution*, 2007. 24(9): p. 2040-8.
46. Grewe, F., et al., A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the *lycophyte* *Isoetes engelmannii*. *Nucleic Acids Research*, 2009. 37(15): p. 5093-5104.

47. Cui, H., et al., Comparative analysis of nuclear, chloroplast, and mitochondrial genomes of watermelon and melon provides evidence of gene transfer. *Scientific Reports*, 2021. 11(1): p. 1595.
48. Anderson, B.M., K. Krause, and G. Petersen, Mitochondrial genomes of two parasitic *Cuscuta* species lack clear evidence of horizontal gene transfer and retain unusually fragmented *ccmF(C)* genes. *BMC Genomics*, 2021. 22(1): p. 816.
49. Choi, I.-S., et al., In and out: Evolution of viral sequences in the mitochondrial genomes of legumes (Fabaceae). *Molecular Phylogenetics and Evolution*, 2021.
50. Choi, K.S. and S. Park, Complete Plastid and Mitochondrial Genomes of *Aeginetia indica* Reveal Intracellular Gene Transfer (IGT), Horizontal Gene Transfer (HGT), and Cytoplasmic Male Sterility (CMS). *International journal of molecular sciences*, 2021. 22(11).
51. Goremykin, V.V., et al., Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Molecular Biology and Evolution*, 2009. 26(1): p. 99-110.
52. Park, S., R.K. Jansen, and S. Park, Complete plastome sequence of *Thalictrum coreanum* (Ranunculaceae) and transfer of the *rpl32* gene to the nucleus in the ancestor of the subfamily Thalicthroideae. *BMC Plant Biology*, 2015. 15(1): p. 40.
53. Li, J., et al., The complete mitochondrial genome of okra (*Abelmoschus esculentus*): using nanopore long reads to investigate gene transfer from chloroplast genomes and rearrangements of mitochondrial DNA molecules. *BMC Genomics*, 2022. 23(1): p. 481.
54. Shidhi, P.R., et al., Genome characterization, comparison and phylogenetic analysis of complete mitochondrial genome of *evolvulus alsinoides* reveals highly rearranged gene order in *Solanales*. *Life (Basel)*, 2021. 11(8).
55. Huang, C.Y., et al., Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to Angiosperm nuclear chromosomes. *Plant Physiology*, 2005. 138(3): p. 1723-1733.
56. Han, A.D., et al., pRRD9, a repetitive DNA sequence that transfers from the chloroplast genome to the nuclear genome with diverse amplification across species, in *Rice Genetics III*. 2008. p. 373-376.
57. Doyle, J.J. and J.L. Doyle, A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 1987. 19(1): p. 11-15.
58. Jin, J.J., et al., GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology*, 2020. 21(1): p. 241.
59. Wick, R.R., et al., Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*, 2015. 31(20): p. 3350-3352.
60. Li, H., Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 2018. 34(18): p. 3094-3100.
61. Koren, S., et al., Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, 2017. 27(5): p. 722-736.
62. Salmela, L. and E. Rivals, LoRDEC: accurate and efficient long read error correction. *Bioinformatics*, 2014. 30(24): p. 3506-14.

63. Langmead, B. and S.L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 2012. 9(4): p. 357-359.
64. Danecek, P., et al., Twelve years of SAMtools and BCFtools. *Gigascience*, 2021. 10(2).
65. Walker, B.J., et al., Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, 2014. 9(11): p. e112963.
66. Chan, P.P. and T.M. Lowe, tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods in Molecular Biology*, 2019. 1962: p. 1-14.
67. Greiner, S., P. Lehwark, and R. Bock, OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research*, 2019: p. 545509.
68. Rice P, L.I., Bleasby A. , EMBOSS the european molecular biology open software suite *Trends Genet*, 2000. 16: p. 276-7.
69. Beier, S., et al., MISA-web: a web server for microsatellite prediction. *Bioinformatics*, 2017. 33(16): p. 2583-2585.
70. Kurtz S, S.C., REPuter fast computation of maximal repeats in complete genomes. *Bioinformatics*, 1999. 15: p. 426-7.
71. Kurtz S, C.J., Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R., REPuter the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research*, 2001. 29(22): p. 4633-42.
72. G., B., Tandem repeats finder a program to analyze DNA sequences. *Nucleic Acids Research*, 1999. 27: p. 573-80.
73. Lai, Q., et al., Complete plastid genome of *Suriana maritima* L. (Surianaceae) and its implications in phylogenetic reconstruction of Fabales. *Journal of Genetics*, 2019. 98(5).
74. Zhang, R., et al., Exploration of Plastid Phylogenomic Conflict Yields New Insights into the Deep Relationships of Leguminosae. *Systematic Biology*, 2020.
75. Johnson, M., et al., NCBI BLAST: a better web interface. *Nucleic Acids Res*, 2008. 36(Web Server issue): p. W5-9.

Figures

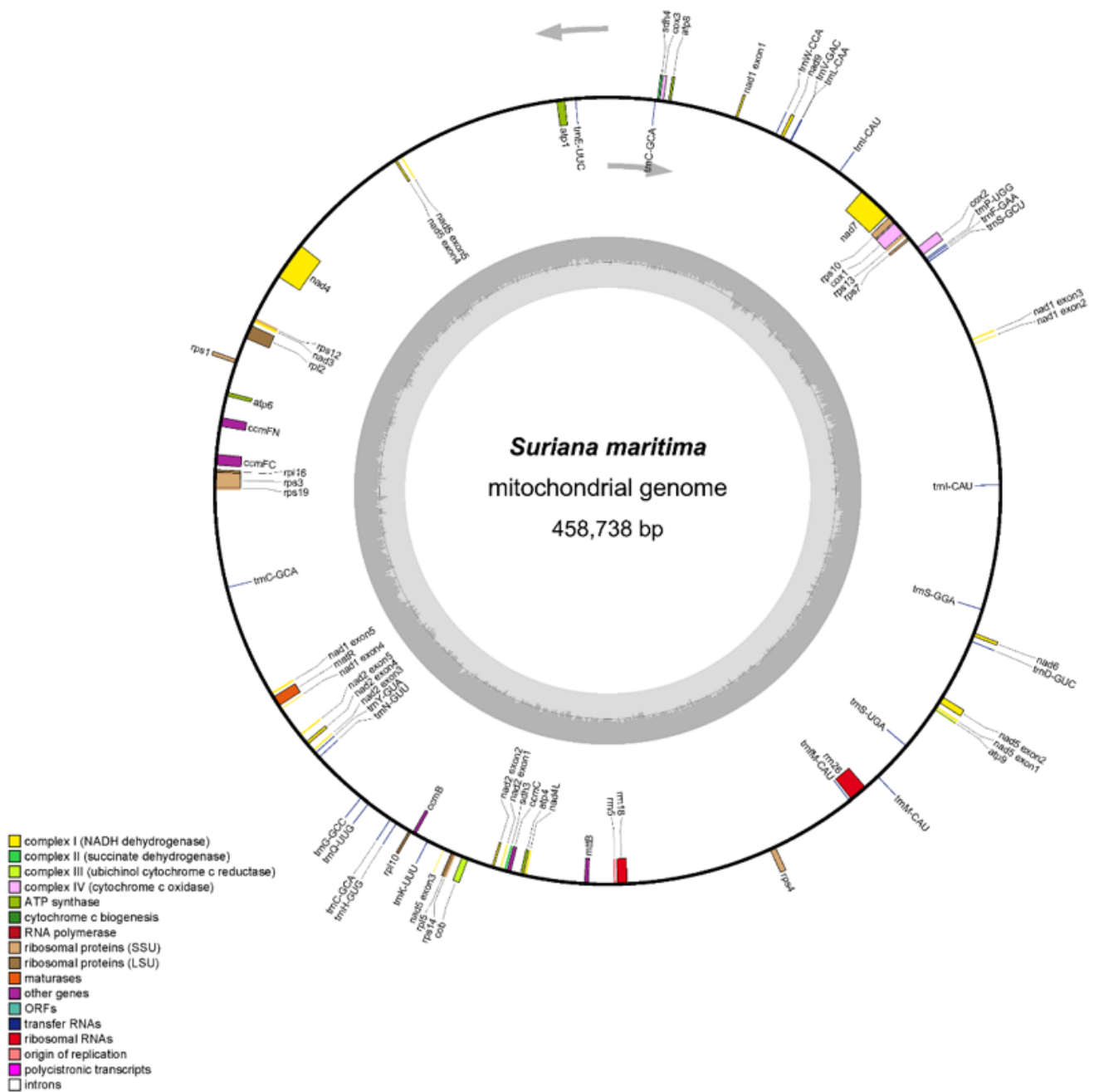


Figure 1

Mitogenome map of *Suriana maritima*. Genes in the circle are transcribed clockwise and genes outside the circle are transcribed counterclockwise. Genes belonging to different functional groups are color-coded. The darker gray area in the inner ring corresponds to GC content, while the lighter gray area corresponds to AT content.

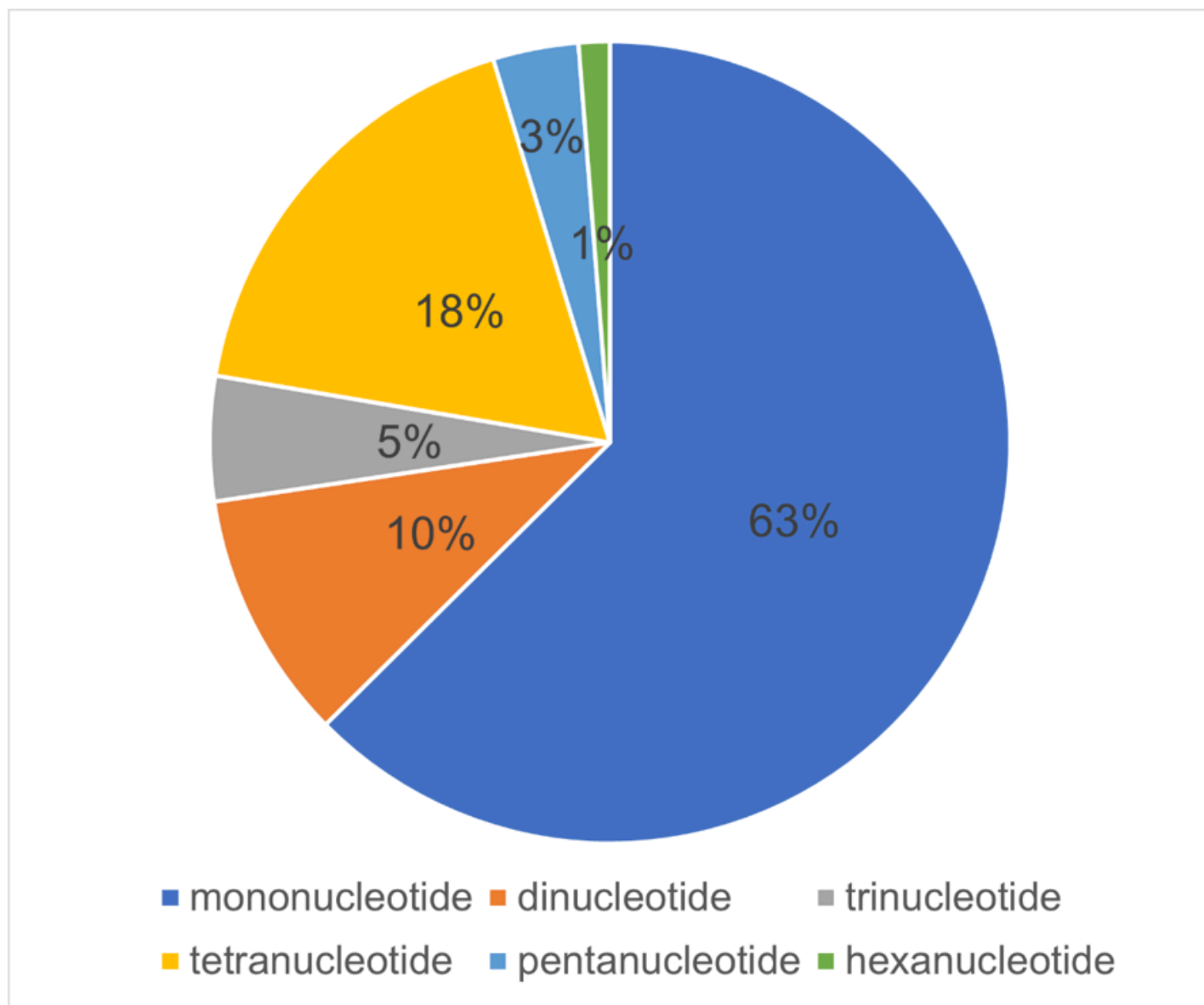


Figure 2

Proportions of simple sequence repeats (SSRs) types in the mitogenome of *Suriana maritima*.

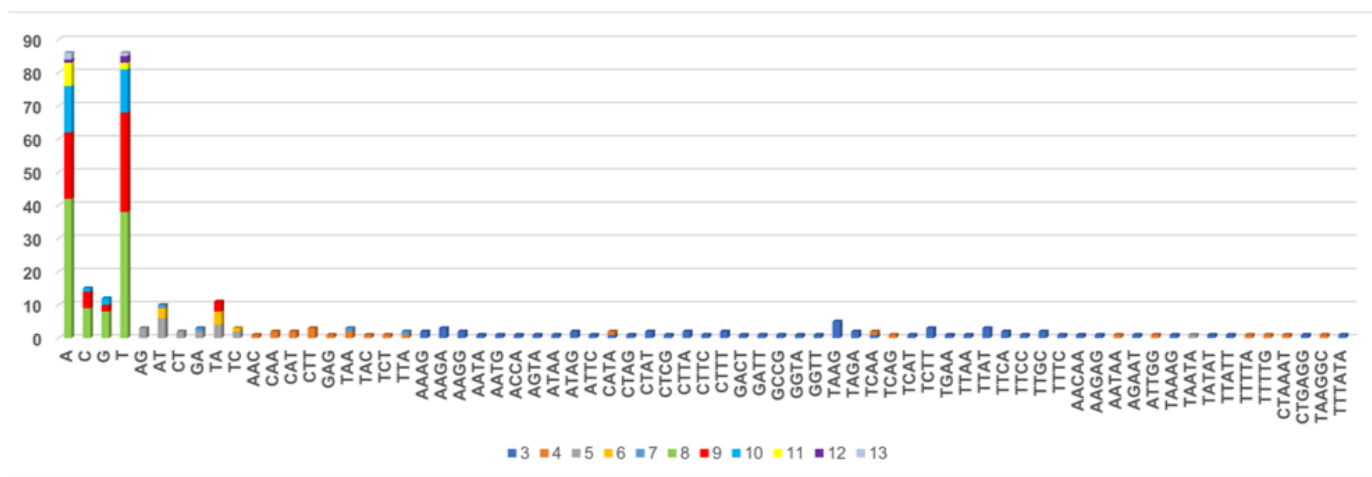


Figure 3

Comparison of mitochondrial SSRs repeat units in the mitogenome of *Suriana maritima*.

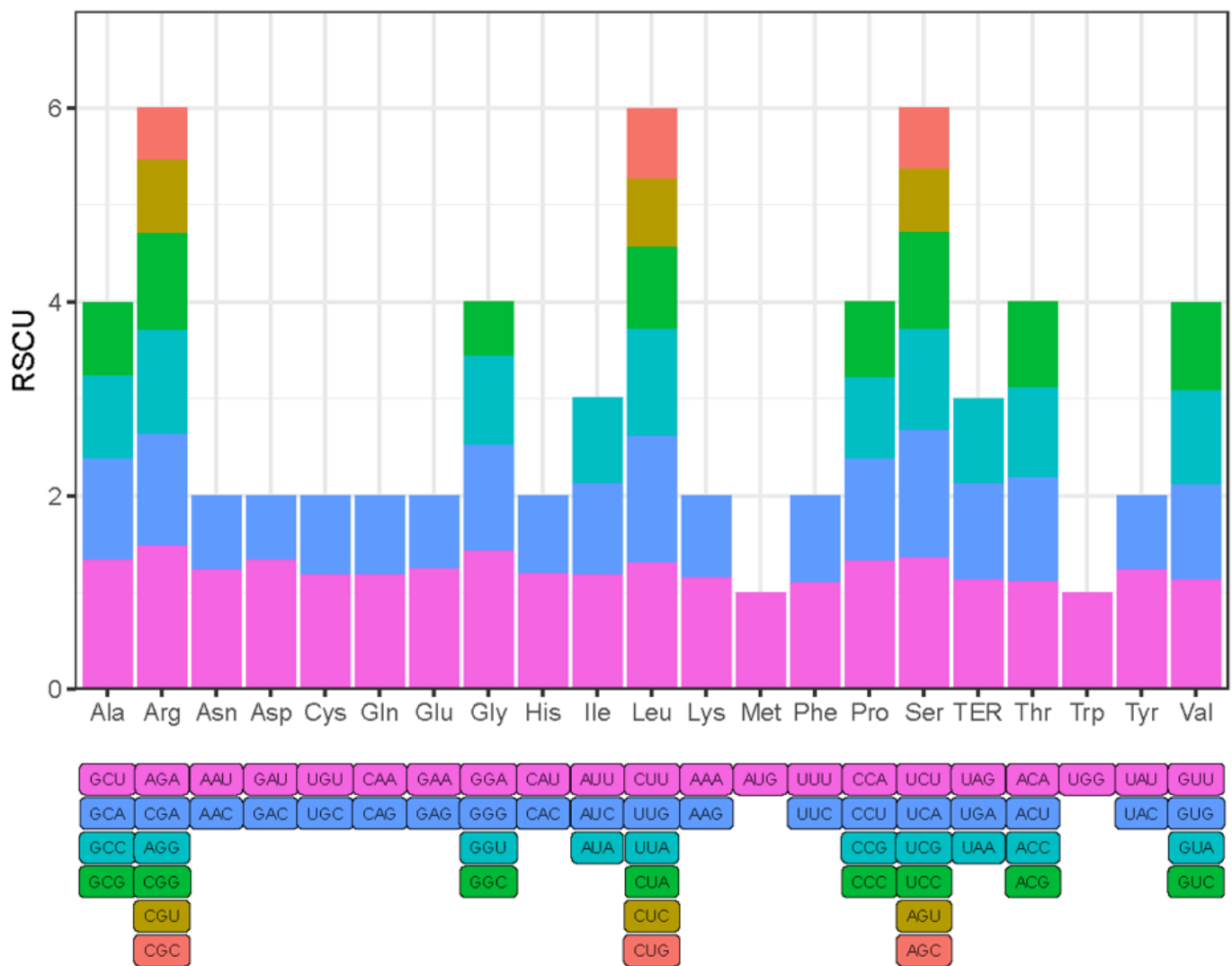


Figure 4

Codon usage preferences for mitochondrial PCGs in *Suriana maritima*

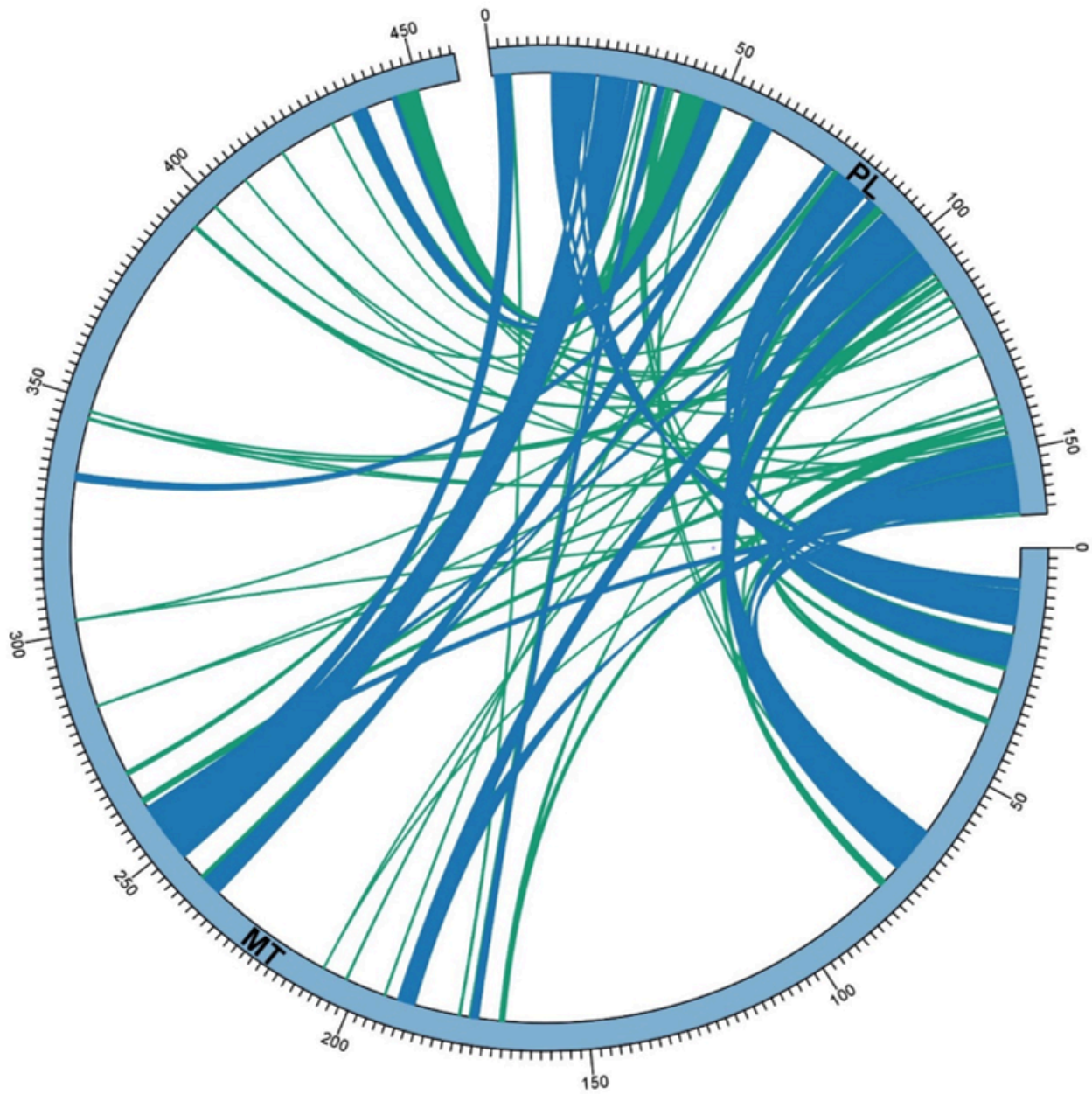


Figure 5

Visualization of Blastn results for the mitogenome and plastid genome of *Suriana maritima*.

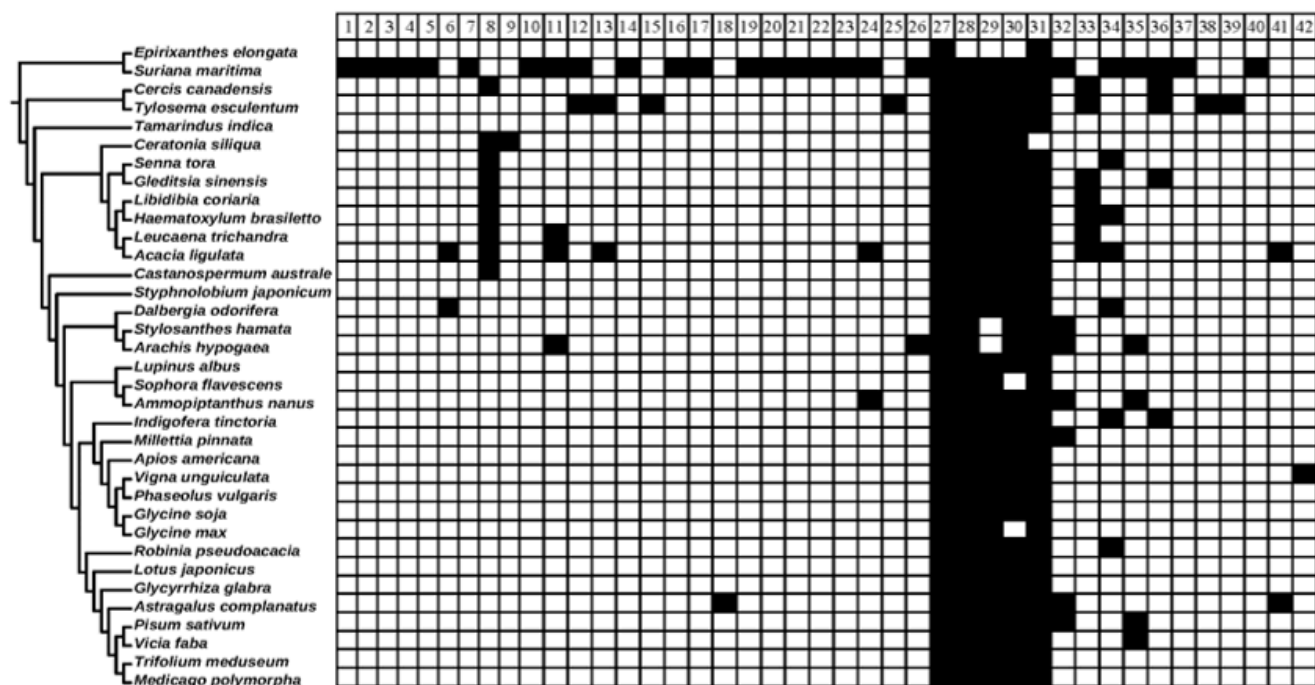


Figure 6

Genes transferred from the plastid genome to the mitogenome.

Note 1–42 represent *atpB*, *atpE*, *atpF*, *atpH*, *atpI*, *ndhB*, *matK*, *petG*, *petL*, *petN*, *psaA*, *psbB*, *psbC*, *psbD*, *psbZ*, *rbcL*, *rpl14*, *rpl2*, *rpl22*, *rpl23*, *rps2*, *rps3*, *rps4*, *rps7*, *rps14*, *ycf3*, *trnD-GUC*, *trnH-GUG*, *trnM-CAU*, *trnN-GUU*, *trnW-CCA*, *trnI-CAU*, *trnP-UGG*, *trnV-GAC*, *trnA-UGC*, *trnS-GGA*, *trnC-GCA*, *trnfM-CAU*, *trnG-GCC*, *trnK-UUU*, *trnL-CAA*, and *trnT-GGU*

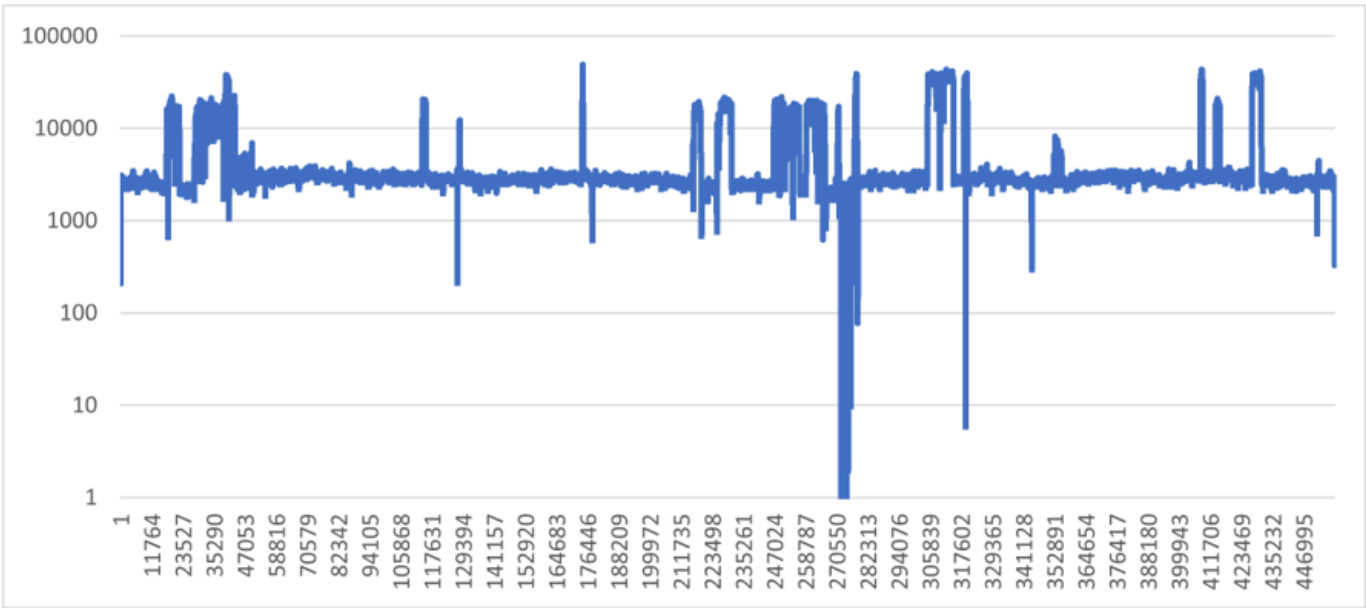


Figure 7

Depth of coverage of Illumina reads along the Nanopore assembly.

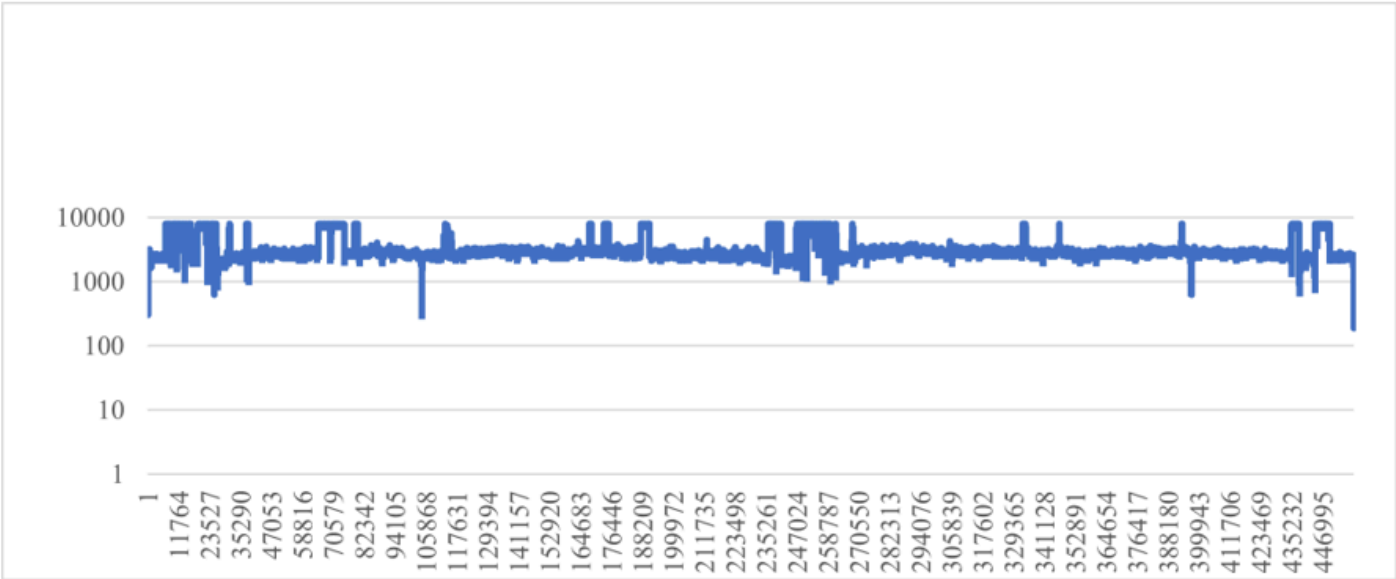


Figure 8

Depth of coverage of Illumina reads along the corrected Nanopore assembly.

Supplementary Files

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

- [SuppTable.docx](#)
- [FigureS1Structureddiagramofthesecond.docx](#)