PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

CC BY

Research

Cite this article: Lavrov DV, Gettle N. 2026 Extreme mitochondrial genome complexity in the calcaronean sponge *Sycon ciliatum* (Fabricius, 1780): fragmentation, gene duplication and mRNA editing. *Phil. Trans. R. Soc. B* **381**: 20250082.
https://doi.org/10.1098/rstb.2025.0082

Received: 22 July 2025
Accepted: 24 January 2026

One contribution of 14 to a theme issue 'Evolutionary genetics of mitochondria: on diverse and common evolutionary constraints across eukarya'.

Subject Areas:
genomics, evolution, cellular biology

Keywords:
mtDNA, linear chromosomes, mRNA editing, Calcarea

Author for correspondence:
Dennis V. Lavrov
e-mail: dlavrov@iastate.edu

Electronic supplementary material is available online at
https://doi.org/10.6084/m9.figshare.c.8381730.

THE ROYAL SOCIETY PUBLISHING

Extreme mitochondrial genome complexity in the calcaronean sponge *Sycon ciliatum* (Fabricius, 1780): fragmentation, gene duplication and mRNA editing

Dennis V. Lavrov[1] and Noah Gettle[2]

[1]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50011, USA
[2]Biological Sciences, University of Denver, Denver, CO 80222, USA

DVL, 0000-0002-2745-1704

Mitochondrial genomes (mt-genomes) of calcaronean sponges are among the most unusual in Metazoa. They are fragmented into multiple linear chromosomes (mt-chromosomes) and often rely on insertional mRNA editing to generate functional transcripts. These unusual features have precluded complete characterization of calcaronean mt-genomes using short-read sequencing technologies. Here, we assembled and analyzed the mt-genome of *Sycon ciliatum* (Fabricius, 1780) using HiFi PacBio data generated by the Aquatic Symbiosis Genomics Project. The mt-genome comprised several megabases of sequence distributed across thousands of chromosomes. While most protein-coding genes were preset in multiple copies, three typical animal mitochondrial protein-coding genes (*atp8*, *nad4L* and *nad6*) and multiple tRNA isotypes were not detected, and only short fragments of mt-rRNA genes were identified. We confirmed that mitochondrial mRNA editing in *S. ciliatum* occurred through single or double uridine insertions at 200+ sites, and that editing patterns were largely predictable from primary sequence motifs. Unexpectedly, editing sites varied among gene copies and were frequently disrupted by point mutations, leading to substantial changes in encoded amino acid sequences. Most mt-chromosomes were associated with repetitive elements that may function in genome maintenance and recombination. Together, our results reveal a distinctive mode of mt-genome evolution and function, shaped by extreme genome fragmentation, extensive gene duplication and pervasive RNA editing.

This article is part of the theme issue 'Evolutionary genetics of mitochondria: on diverse and common evolutionary constraints across eukarya'.

1. Introduction

Mitochondrial DNA (mtDNA) is a vestige of the genome of an α-proteobacterial partner acquired during the primary endosymbiotic event at the origin of eukaryotes [1]. As such, it probably inherited the circular genome architecture typical of most bacteria [2], as well as several thousand genes common in α-proteobacteria [3]. During eukaryotic evolution, mitochondrial genomes (mt-genomes) underwent massive gene loss and extensive gene transfer to the nucleus, such that no more than approximately 100 genes are retained there in extant eukaryotes [4]. The remaining genes typically encode core components of the oxidative phosphorylation (OXPHOS) system, as well as ribosomal and transfer RNAs (rRNAs and tRNAs, respectively), although the most gene-rich

© 2026 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

Downloaded from royalsocietypublishing.org/rstb/article-pdf/doi/10.1098/rstb.2025.0082/6130399/rstb.2025.0082.pdf
by Flanders Marine Institute user
on 28 September 2026

mt-genomes may also encode mitochondrial ribosomal proteins, a bacterium-like RNA polymerase, subunits of the Sec general secretory pathway and components of the cytochrome *c* maturation system [5,6]. This drastic reduction in gene content has been accompanied by diversification in mt-genome architecture, with linear and/or multipartite mt-genomes having evolved repeatedly across eukaryotes [4].

While mt-genomes are known to be diverse in some lineages, including land plants and fungi, those of animals have generally been regarded as highly conserved [7]. Indeed, in most bilaterian animals, mtDNA is a small, circular-mapping molecule encoding a canonical set of 37 genes, including 13 protein-coding genes for subunits of the electron transport chain and ATP synthase, two ribosomal RNA genes and 22 transfer RNA genes [8]. By contrast, most non-bilaterian animals—the phyla Cnidaria, Ctenophora, Placozoa and Porifera—depart from this 'typical' animal mitochondrial genome organization with respect to gene content, genome architecture, or both [9]. These departures include the presence of additional protein-coding genes and introns, frequent loss of tRNA genes and transitions to linear and/or multipartite genome organization [9,10]. Among the most unusual mitochondrial genome organizations described to date are those of calcaronean sponges, a subclass of marine sponges within the class Calcarea.

Calcaronea is a relatively small (578 species) but taxonomically diverse group of sponges comprising three orders, 15 families and 58 genera [11]. Together with their sister group, Calcinea, calcaroneans form the class Calcarea Bowerbank, 1864, a monophyletic lineage of marine sponges characterized by a mineral skeleton composed entirely of calcium carbonate [12]. Although typically minute and inconspicuous, calcaronean sponges have long served as important model systems in Porifera research (reviewed in [13,14]) and have played a central role in efforts to understand early metazoan evolution [15].

Calcaronean sponges display several unusual biological features, particularly with respect to reproduction and development. While oogenesis in *Sycon* and other calcareous sponges is commonly observed, spermatogenesis appears to be rare and remains poorly understood. In fact, *Sycon cf. ciliatum* has been considered putatively unisexual [16]. Despite the scarcity of observations of spermatogenesis, fertilization has been described in several species and involves specialized carrier cells that transport a structure interpreted as a spermiocyst into the oocyte. Embryonic development in Calcaronea is likewise highly unusual. The earliest embryonic stage is a coeloblastula in which the flagella are internal and oriented toward the central cavity. This blastula subsequently undergoes a complex inversion (also termed eversion) that turns the flagellated pole of the blastomeres outward, giving rise to an amphiblastula larva [16,17].

To our knowledge, only a single previous study has examined mitochondrial genomes in calcaronean sponges, namely that of Lavrov *et al.* [18]. This study demonstrated that calcaronean mt-genomes are composed of multiple linear chromosomes, each carrying a subset of the canonical mitochondrial gene complement. Remarkably, mitochondrial coding sequences in most calcaroneans undergo insertional RNA editing, in which uridines (Us) are added at specific sites within transcripts. In the absence of editing, these coding sequences would contain frameshifts, resulting in truncated or nonfunctional proteins. Such an atypical genome architecture raises numerous challenges for mt-genome maintenance, including chromosomal stability, coordinated replication and transcription, and faithful transmission of the complete mitochondrial gene repertoire.

Although the study by Lavrov *et al.* [18] described the general organization of calcaronean mt-genomes and their reliance on insertional RNA editing, it was constrained by the sequencing technologies available at the time. In particular, the high abundance of divergent repetitive sequences prevented complete assembly of the *Sycon ciliatum* mt-genome. As a result, the full extent of genome fragmentation, gene duplication and sequence divergence within and among mt-chromosomes remained unresolved. The advent of single-molecule long-read sequencing has provided an opportunity to revisit these questions. Here, we take advantage of genomic and transcriptomic data for *S. ciliatum* generated as part of the Aquatic Symbiosis Genomics Project (https://www.aquaticsymbiosisgenomics.org/) to investigate its mt-genome in greater detail, with a focus on gene content, the extent of gene duplication and divergence and the dynamics of RNA editing.

Our analyses reveal a highly divergent and complex mt-genome that rivals the most unusual systems described in ichthyosporeans [19], dinoflagellates [20] and trypanosomes [21]. These findings underscore the remarkable evolutionary plasticity of mitochondrial genomes and raise fundamental questions about the limits and mechanisms of mitochondrial genome maintenance in eukaryotes.

2. Material and methods

(a) Data sources

We used three datasets generated and made publicly available in GenBank by the Aquatic Symbiosis Genomics Project: PacBio HiFi Revio sequences (ERR12319363), Illumina NovaSeq 6000 poly(A)$^+$ RNA-seq data (ERR12318596) and Illumina NovaSeq 6000 genomic DNA data (ERR12318594) (https://www.ncbi.nlm.nih.gov/bioproject/PRJEB70259). Both PacBio and RNA-seq data were generated from the same individual of *Sycon ciliatum* (SAMEA8419667), whereas Illumina genomic DNA sequences were obtained from a different individual (SAMEA8419668).

(b) Assembly and identification of mitochondrial chromosomes

Three subsets of putative mitochondrial sequences were identified from the PacBio dataset using BLAST v.2.16.0 [22] and assembled with the `cap3` program [23].

Downloaded from royalsocietypublishing.org/rstb/article-pdf/doi/10.1098/rstb.2025.0082/6130399/rstb.2025.0082.pdf
by Flanders Marine Institute user
on 28 September 2026

First, we searched for sequences similar to previously published and conceptually de-edited *Sycon ciliatum* mitochondrial transcript sequences [18]. Conceptual de-editing was necessary to remove uridine insertions introduced by RNA editing and was performed with the `de-edit_transcripts.py` script. Command-line BLAST searches were performed using: `blastn -query $query -db ERR12319363_blast_db -out $outfile -num_threads 16 -evalue 1e-10 -outfmt '6 qseqid sseqid sstart send qcovs qcovhsp qcovus evalue bitscore'`. Results were sorted by Query Coverage Per Subject (`qcovus`), duplicates were removed and only hits with `qcovus > 80` were retained. A total of 1029 sequences containing these hits were extracted from the PacBio dataset using `seqtk subseq` and used for the first assembly.

Second, we identified two common repeat motifs in the initial assembly that occurred in multiple copies, often in inverted orientation outside of coding regions (electronic supplementary material, file S1). These motifs were used as queries to identify additional putative mt-chromosomes using BLAST: `blastn -query repeat_elements.fa -db ERR12319363_blast_db -out reads_with_repeat.out -num_threads 16 -evalue 1e-5 -max_target_seqs 10000 -outfmt '6 qseqid sseqid sskingdoms staxids sscinames scomnames qcovs qcovhsp qcovus evalue bitscore'`. This search identified 2933 sequences with similarity to the repeat motifs, including 662 sequences already identified by gene-based searches. The remaining 2271 sequences were assembled separately.

Finally, all 2981 sequences identified by similarity to either coding sequences or repeat motifs were combined and assembled to generate the final mt-genome dataset. All assemblies, scripts and intermediate output files from the tools used for the analyses are located in the accompanying GitHub repository: https://github.com/dlavrov/Sycon_ciliatum_mt-genome_analysis.

(c) Transcriptome assembly and identification of protein-coding regions

For de novo transcriptome assembly, RNA-seq reads were trimmed using Trimmomatic v. 0.39 [24] and assembled using both rnaSPAdes v. 3.15.5 [25] and Trinity [26]. Trimmomatic was run with the following parameters: `ILLUMINACLIP:$adapters_dir/TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50`. rnaSPAdes was run with default parameters. For Trinity, the options `-min_kmer_cov 2 -no_normalize_reads` were used to remove erroneous k-mers and retain all reads.

To assess assembly robustness and transcript expression levels, trimmed RNA-seq reads were mapped back to assembled transcripts using Bowtie v. 1.3.0 with parameters: `-a -v 2 -best -strata -X 1000 -fr`. Alignment files were sorted using Samtools v. 1.19.2 [27], and per-base coverage was calculated using `samtools depth`. Average sequencing depth was computed for each mt-chromosome and each protein-coding gene to assess copy number variation and coverage uniformity. These steps were combined in the `bowtie1.sh` script.

Protein-coding regions in the assembled transcriptome were identified using TransDecoder2 (TD2) [28]. Open reading frames (ORFs) longer than 50 amino acids were predicted using `TD2.LongOrfs` with the `-m 50` and `-M 50` options. Probable coding sequences were then identified using `TD2.Predict`. To accommodate the non-standard mitochondrial genetic code, translation table 4 (Mould, Protozoan and Coelenterate Mitochondrial Code) was specified using the `-G 4` option. Although this code does not account for one of the two codon reassignments inferred for calcaronean sponges [9], it prevented premature ORF termination caused by reassignment of the TGA codon from stop to tryptophan.

(d) Mitochondrial genome annotation

(i) Genome statistics and coverage analysis

Basic composition statistics for the mt-genome were calculated using `seqtk comp` (https://github.com/lh3/seqtk). For comparison, identical statistics were calculated for 100 000 random PacBio reads selected using `seqtk sample -s 753` (https://github.com/lh3/seqtk). AT-skew and GC-skew were calculated as $(A-T)/(A+T)$ and $(G-C)/(G+C)$, respectively, and absolute values were averaged across sequences.

To estimate sequence coverage, Illumina genomic DNA reads (ERR12318594) were adapter- and quality-trimmed with Trimmomatic v. 0.39 using the same parameters as for RNA-seq reads and mapped to the mt-genome assembly. A Bowtie 2 index was built using Bowtie 2 v. 2.5.4 [29] and mapping was done using strict alignment parameters (`-end-to-end -no-unal -no-mixed -no-discordant`) to avoid partial or ambiguous alignments. Alignments were processed using Samtools v. 1.19.2 [27], and coverage was calculated using `samtools depth`. For each mt-chromosome, average depth was calculated as the mean of per-base coverage values across all positions with defined depth. These procedures were combined in the `bowtie2.sh` script.

(ii) Identification of coding sequences

Identification of protein-coding genes in the *S. ciliatum* mt-genome was complicated by insertional RNA editing, which converts cryptogenes into functional coding sequences. Two complementary approaches were used.

Mitochondrial genes identified in the previous study [18] were searched for using conceptually de-edited mRNA sequences as queries in BLAST v. 2.16.0 [22]. BLAST searches were run as described above, retaining only hits with `qcovus > 80`. BLAST output was converted to GFF3 format using the `blast2gff.py` script.

Genes not identified previously (*atp8*, *nad4L* and *nad6*) were searched in conceptually edited mt-chromosomes and the assembled transcriptome using `mmseqs2` [30], with protein queries from *Clathrina clathrus* [31] and *Oscarella carmela* [32]. Conceptual editing of genomic sequences was performed using the `annotate_editing.py` script. Additionally, profile HMMs were built

Downloaded from royalsocietypublishing.org/rstb/article-pdf/doi/10.1098/rstb.2025.0082/6130399/rstb.2025.0082.pdf
by Flanders Marine Institute user
on 28 September 2026

from UniProt Porifera sequences of *atp8*, *nad4L* and *nad6* using `hmmbuild` and searched against the transcriptome using `hmmsearch` (HMMER v. 3.4; [33]) with default parameters. A gene was considered undetected if no significant match was recovered by either `mmseqs2` [30] searches or profile-HMM searches, and no corresponding transcript-supported ORF was identified by TransDecoder2 [28].

Editing sites were confirmed by aligning assembled transcripts to their corresponding genomic copies and scoring positions requiring insertion of one or two uridines to restore reading frame at the canonical NUUU and CCCUU motifs. Exceptions were recorded when transcript evidence supported a different insertion state or absence of editing at an otherwise motif-matching site.

(iii) Identification of rRNA and tRNA genes

Mitochondrial tRNA genes were searched using tRNAscan-SE v. 2.0.12 [34] with the `-O` option (organellar mode). Because tRNA genes were not identified in the previous study [18] and their editing status is unknown, searches were conducted on both unedited and conceptually edited mt-sequences. Nuclear-encoded tRNA genes were identified using default parameters on the published nuclear genome assembly of *S. ciliatum*.

Mitochondrial rRNA genes were identified using `nhmmer` from the HMMER3 package [33] with profile HMMs employed by the MFannot pipeline https://github.com/BFL-lab/MFannot_data. Significant hits were manually inspected to verify rRNA identity, boundaries and orientation.

(iv) Repeat motifs

Tandem repeats were identified using the `etandem` program from the EMBOSS suite v. 6.6.0.0 [35]. Searches were performed separately for each mt-chromosome using `-minrepeat 20 -maxrepeat 200` and the `run_etandem.sh` script. Etandem output was converted to GFF3 format using `etandem2gff3.py`, and repeat orientation was updated using `update_gff_etandem_strand.py`. Homopolymeric nucleotide runs were identified using the `polyAT2gff.py` script.

(e) Phylogenetic analysis

Phylogenetic analyses were conducted for individual gene families, with and without corresponding transcripts, as well as for repetitive sequences. When appropriate, RNA and DNA sequences were combined using the `cat_dna_rna.sh` script. Alignments were generated using MAFFT v. 7.526 with the `-auto` and `-adjustdirection` options via the `do4all` script. Uncorrected pairwise (p) distances were calculated using `dist_calc.py`. Phylogenetic trees were inferred using ParGenes [36] with RAxML-NG v. 1.2.2 [37] under the GTR+G model.

(f) Data visualization

Sequence length distributions were visualized in R using the `Ggplot2` package with the `geom_density` function [38]. Selected mitochondrial chromosomes were visualized using the `gff2pdf.py` script. Sequence coverage profiles were plotted with the `plot_bowtie_coverage.py` script. Phylogenetic trees were visualized using FigTree v. 1.4.4 (https://github.com/rambaut/figtree).

3. Results

(a) Multipartite linear genome of *Sycon ciliatum*

We detected 710 BLAST hits in the PacBio dataset to previously published and conceptually de-edited coding sequences of *Sycon ciliatum*, ranging from a single hit for *nad2* to 363 hits for *atp9*. Assembly of these sequences produced 121 contigs and 282 singlets, with an average of 3.5 sequences per contig (the 'CDS dataset'). Most CDS-associated sequences contained degenerate tandem repeats of either 105 or 121 bp, often present in multiple copies outside coding regions.

BLAST searches using two of these repeat motifs identified 2933 PacBio reads, including 662 previously recovered by the gene-based search. The remaining 2271 reads were assembled into 330 contigs and 1109 singlets (the 'repeat dataset'), again averaging 3.5 sequences per contig. Combining both datasets yielded 443 contigs and 1341 singlets (2981 sequences total), with an average of 3.7 sequences per contig (the 'combined dataset').

Sequences in the combined dataset ranged from 1104 to 35 580 bp in length (mean 5793 bp; median 4549 bp) and were moderately GC-rich (mean 55%; s.d. 0.04), with an average absolute AT-skew of 0.14 and an average absolute GC-skew of 0.03. Assembled contigs accounted for 2.53 Mbp, whereas the combined length of contigs and singlets totaled 10.47 Mbp. Length distributions of contigs and singlets were broadly similar (figure 1), although singlets showed a stronger right skew, consistent with the presence of chimeric reads. Accordingly, singlets had higher mean and median lengths (5922 and 4504 bp) than contigs (4886 and 4025 bp). Randomly sampled PacBio reads showed a substantially right-shifted distribution (mean 8610 bp; median 7955 bp) and had lower GC content (47%) and lower AT-skew (0.056), supporting a mitochondrial origin for the combined dataset.

Most sequences in the combined dataset had moderate sequencing depth (minimum 1; median 14) (electronic supplementary material, file S2). A small number of mt-chromosomes exhibited exceptionally high depth (maximum 3034), largely owing to microsatellite-like sequences, resulting in a mean depth of 48 across the dataset. Coverage distributions were similar for contigs and singlets (median depths 13.5 and 14.5, respectively).

Downloaded from royalsocietypublishing.org/rstb/article-pdf/doi/10.1098/rstb.2025.0082/6130399/rstb.2025.0082.pdf
by Flanders Marine Institute user
on 28 September 2026

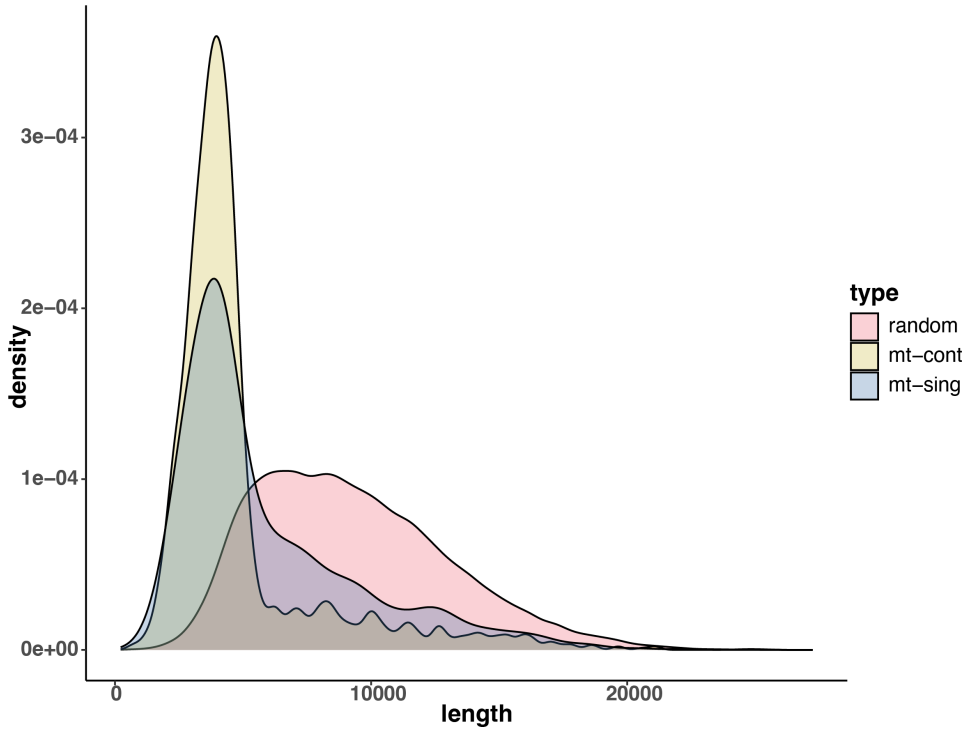

Figure 1. Length (bp) distribution of *S. ciliatum* PacBio sequences identified as mitochondrial versus randomly selected from the dataset. Mitochondrial sequences were subdivided into those that were assembled into contigs (mt–cont) and those that remained as singlets (mt–sing). 100 000 randomly selected sequences were used to approximate the distribution in the whole dataset (random).

Table 1. Number of full-size coding sequences identified in *S. ciliatum* mt-genome. Copy #—the number of mt-chromosomes with complete or nearly complete (>80%) copies of a gene; % divergence—the largest p-distance among the copies including gaps; # of transcripts—the number of assembled transcripts by rnaSpades [25] and Trinity [26], respectively.

gene	copy #	% divergence	# of transcripts
atp6	29	0.10	4/1
atp9	195	0.11	42/21
cob	20	0.18	9/1
cox1	25	0.200	5/1
cox2	11	0.050	1/2
cox3	30	0.140	13/1
nad1	1	—	1/0
nad2	1	—	2/1
nad3	10	0.170	1/1
nad4	3	0.030	5/2
nad5	2	0.002	2/1

(b) Reduced set of mt protein-coding genes with extensive gene duplication

We identified the same set of mitochondrial protein-coding genes in the full genomic dataset as in our previous transcriptomic study [18]: *atp6*, *atp9*, *cob*, *cox1–3* and *nad1–5*. Three additional small genes typically present in animal mtDNA (*atp8*, *nad4L* and *nad6*) were not detected in either genomic or transcriptomic data.

Gene copy number varied widely, from a single full-length copy of *nad1* and *nad2* to 195 copies of *atp9* (table 1). Genes encoding subunits of complex I generally had fewer copies than genes from other respiratory complexes. Sequence divergence among near full-length copies ranged from 0% to 17% and included point mutations, insertions, deletions, truncations and frameshifts. After correction for predicted RNA editing and translation using the inferred mitochondrial genetic code, most variants still contained frameshifts owing to indels or mutations disrupting editing motifs. As a result, typically only one copy per gene family was predicted to encode a full-length functional protein.

We also identified sequences resembling non-mitochondrial elements, including retroviral and transposable element fragments. These sequences did not require editing to maintain open reading frames, were often located on contigs substantially

Downloaded from royalsocietypublishing.org/rstb/article-pdf/doi/10.1098/rstb.2025.0082/6130399/rstb.2025.0082.pdf
by Flanders Marine Institute user
on 28 September 2026

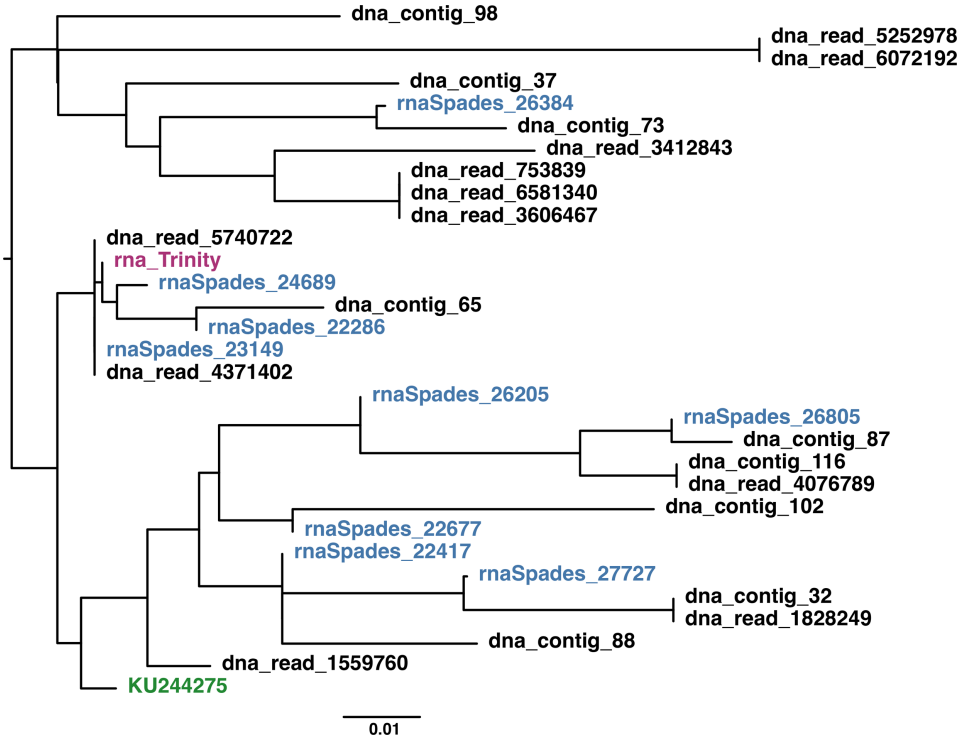

Figure 2. Phylogenetic relationships among gene and transcript variants for *cob*. Unrooted maximum likelihood tree reconstructed with the GTR+G substitution model is shown. 'dna_read' and 'dna_contig' labels correspond to singlets and contigs in the assembly. 'rnaSpades' and 'rna_Trinity' labels indicate transcripts assembled by the corresponding programmes. KU244275.1 transcriptomic sequence was identified in our previous study [18].

longer than typical mt-chromosomes, and frequently lacked characteristic tandem repeat arrays. Moreover, most had nearly identical matches on assembled nuclear chromosomes, suggesting that they reside on chimeric contigs and/or represent assembly artefacts.

(c) Expression of coding sequences

To assess expression of gene variants, we assembled a transcriptome of *S. ciliatum* and analysed genomic and transcriptomic sequences together (figure 2). Assembly outcomes differed substantially between rnaSPAdes and Trinity. For all gene families, rnaSPAdes produced more transcript variants, several of which clustered with distinct genomic gene copies, indicating their expression. By contrast, Trinity produced one or two transcripts for all gene families except *atp9*, for which 21 variants were recovered (table 1). A subset of mt-chromosomes containing coding sequences most closely matching Trinity transcripts is shown in figure 3.

Coverage analysis revealed that Trinity transcripts generally showed high and uniform RNA-seq support (electronic supplementary material, figure S1), whereas many rnaSPAdes transcripts exhibited uneven coverage, particularly near 3′ ends (electronic supplementary material, figure S2). Similarly, only one Trinity-assembled *atp9* contig showed coverage comparable with other mitochondrial genes. Together, these results indicate that only one or a few copies per gene family are robustly expressed. Comparison with transcriptomic data from a previous study revealed high similarity for most genes, although several *nad* sequences differed at up to 10% of positions, suggesting that the corresponding mt-chromosomes may have been lost in the individual sampled for this project.

(d) mRNA editing and evolution of editing sites

In addition to its unusual architecture, the mt-genome of *S. ciliatum* stores genetic information in an abbreviated form, with coding sequences completed at the RNA level by insertional RNA editing. Previously, we showed that RNA editing in *S. ciliatum* occurs at NUUU and CCCUU motifs (where N denotes any nucleotide), with GUUU and CCCUU sequences expanded by single-uridine insertions and AUUU, CUUU and UUUU sequences expanded by double-uridine insertions [18]. The present study confirms these patterns and enables a more comprehensive assessment of their extent. In total, we identified 219 editing sites in sequences inferred to encode functional copies of mt-genes (table 2). Editing was also observed across divergent gene copies, and single-nucleotide substitutions at editing motifs sometimes prevented editing and led to translational frameshift in otherwise similar transcripts. These results support the conclusion that editing is largely determined by short primary sequence motifs.

We also observed a small number of exceptions to the previously described rules. Most occurred when two editing sites were immediately adjacent or separated by a single nucleotide; in such cases, the second site was not edited. Accordingly, editing was absent (or reduced) at position 10 in *atp6*, which contained the [TTTT]CCCTT sequence; at position 98 of *cox3*, containing either [CTTTA]TTTT or [CTTTA]CTTT; and at positions 82 and 883 in *nad5*, containing [CCCTT]CTTT and [CTTTC]ATTT, respectively

Downloaded from royalsocietypublishing.org/rstb/article-pdf/doi/10.1098/rstb.2025.0082/6130399/rstb.2025.0082.pdf
by Flanders Marine Institute user
on 28 September 2026

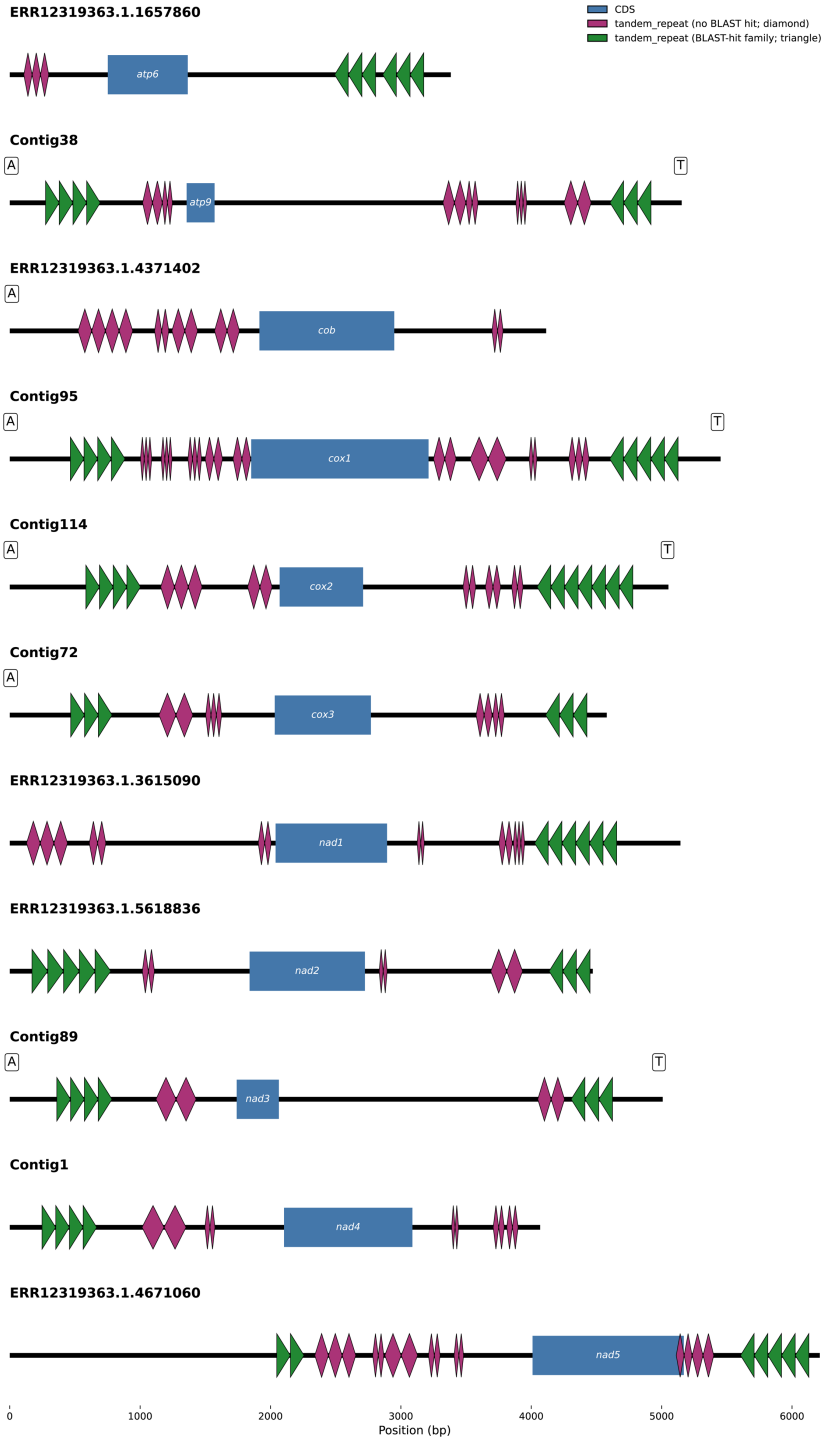

Figure 3. Maps of a representative set of mt-chromosomes containing protein-coding genes. Chromosomes were selected by sequence similarity to Trinity transcripts. Rectangles (blue) represent coding sequences, rhombuses (purple)—units of tandem repeats, triangles (green)—units of tandem repeats matching repetitive sequences used for identification of additional mt-chromosomes. The direction of a triangle indicates relative orientation of a repeat unit. Squared letters A and T signify the presence of homopolymeric A or T runs at the end of mt-chromosomes.

(nucleotides upstream of the site in question are shown in square brackets). Other exceptions did not follow an obvious pattern. In *cox1*, the [GG]CTTT motif at position 272 was predominantly edited by insertion of a single uridine, although transcripts with two insertions were also observed. In addition, three sites in *atp9* showed no editing, despite matching the motif rules and being well separated from other editing sites. These exceptions suggest that additional, as yet unrecognized, constraints influence RNA editing in calcaronean sponges.

Downloaded from royalsocietypublishing.org/rstb/article-pdf/doi/10.1098/rstb.2025.0082/6130399/rstb.2025.0082.pdf
by Flanders Marine Institute user
on 28 September 2026

Table 2. Number of editing sites in coding sequences. An asterisk indicates the presence of an additional unedited site or a site with non-standard editing.

gene	ATTT	CTTT	TTTT	GTTT	CCCTT	total	unedited
atp6	1	10	3	4	7	25	1
atp9	0	3	0	0	0	3	3
cob	1	11	7	1	6	26	0
cox1	4	15*	5	3	4	31*	1
cox2	3	2	2	7	1	15	0
cox3	2	11*	6	1	1	21*	1
nad1	0	12	6	1	2	21	0
nad2	1	14	11	1	9	36	0
nad3	0	4	3	1	1	9	0
nad4	1	15	11	3	3	33	0
nad5	1*	14*	6	4	5	30	2

(e) Putative tRNA and rRNA genes

(i) tRNA genes

Across 1784 putative mtDNA sequences, tRNAscan identified 108 putative tRNA genes, including 26 with undetermined isotypes. The distribution of predicted tRNAs was uneven across both isotypes and sequences. Among 443 assembled contigs, we identified only 13 putative tRNA genes, all with low Infernal scores (16.5–31.7) and belonging to only two isotypes (serine and leucine); two additional predictions had undetermined anticodons. Five of these tRNA predictions occurred on an unusually long contig (Contig352). By contrast, among 1341 singlets, we identified 95 putative tRNA genes with Infernal scores ranging from 15.3 to 70.2. This difference in the proportion of tRNA predictions between contigs and singlets was statistically significant (Fisher's exact test, $p < 0.005$) and may reflect sequencing or assembly artefacts. Among the 95 singlet-derived predictions, nine isotypes were still missing, whereas a few isotypes were represented by multiple copies. The most common predicted isotypes were serine (42 copies), leucine (13 copies) and methionine (13 copies).

We previously inferred two changes in the mitochondrial genetic code of *Sycon ciliatum* relative to the standard code: UGA (Stop→Trp) and AGR (Arg→Ser/Gly) [18]. The first reassignment was probably present in the common ancestor of animals, whereas the second appears to have evolved within calcaronean sponges, potentially in parallel with bilaterians. Both changes would require mt-specific tRNAs for these codons. However, no such mt-tRNA genes were identified on mt-chromosomes.

The lack of multiple tRNA genes could either indicate that tRNA genes rely heavily on RNA editing, which makes their primary sequences unrecognizable, or that they are encoded in the nuclear genome and their products are imported to mitochondria. To evaluate the former possibility, we made conceptual editing of predicted editing patterns on both the forward and reverse-complement strands of each predicted mtDNA sequence and repeated tRNAscan searches. Although the total number of predicted tRNAs and their scores changed slightly, the overall pattern remained similar to that obtained from unedited sequences. To evaluate tRNA import, we searched for mt-like tRNA genes in the nuclear genome assembly of *S. ciliatum*. We identified more than 12 000 tRNA and tRNA-like sequences, most located on assembled nuclear chromosomes. Many predicted tRNAs had high Infernal scores, consistent with functional cytosolic tRNA genes, whereas the majority had low scores and could include mt-tRNAs. Notably, we identified 11 *trnW(cca)* predictions with Infernal scores of 21.0–39.3 (whereas inferred cytosolic *trnW(cca)* copies had scores of 75.9) and 45 *trnR(ucu)* predictions with scores of 20.1–48.2 (putative cytosolic copies had scores of 66.4–71.0). All but one of these were encoded on assembled nuclear chromosomes; the only exception was a *trnR(ucu)* prediction on NW_027044754.1.

(ii) rRNA genes

We identified several short fragments with similarity to mitochondrial rRNA genes in the combined assembly. Fragments matching the 3′ end of the large-subunit rRNA gene (*rnl*) were found on 7 contigs and 12 singlets. The best and longest matches were located on singlets longer than 7 kbp and showed near-complete identity to the *Sycon ciliatum* nuclear *lrRNA* gene, suggesting a nuclear origin or chimeric assembly rather than bona fide mitochondrial chromosomes. Shorter and more weakly scoring hits showed partial similarity to mitochondrial *rnl* sequences in other metazoans. They also occurred on contigs comparable in size and structural features to other mt-chromosomes (e.g. containing terminal poly-A/poly-T runs) and showed evidence of RNA editing.

Fragments matching the 3′ end of the small-subunit rRNA gene (*rns*) were identified on 20 contigs and 36 singlets. As with *rnl*, the strongest four hits matched the nuclear *srRNA* gene and were located on unusually large contigs or singlets. By contrast, shorter and more weakly scoring hits were found on contigs resembling typical mt-chromosomes. However, there was little sequence similarity between these mt-chromosomes and known metazoan srRNA genes.

Downloaded from royalsocietypublishing.org/rstb/article-pdf/doi/10.1098/rstb.2025.0082/6130399/rstb.2025.0082.pdf
by Flanders Marine Institute user
on 28 September 2026

Finally, four matches to the 5′ region of *rnl* were detected, all of which exceeded the typical size of mitochondrial chromosomes and matched the *Sycon ciliatum* nuclear *lrRNA* gene. No comparable hits were recovered for the 5′ region of *rns*.

Overall, these results indicate that mitochondrial rRNA genes in *S. ciliatum* are fragmented and divergent, with only short recognizable rRNA-like regions present, whereas longer high-identity rRNA matches represented either imported nuclear loci, contamination or, more likely, chimeric sequences/assemblies.

(f) Repetitive sequences

Coding sequences—when present—constituted only a small fraction of their corresponding mt-chromosomes. By contrast, arrays of imperfect tandem repeats were abundant. The `Tandem` program identified 34 132 tandem repeats (20–200 bp) grouped into 10 549 clusters across 1784 putative mtDNA sequences, with major peaks at repeat sizes of 105 bp (7460 repeats), 121 bp (1716 repeats) and 106 bp (1666 repeats). While a comprehensive analysis of repetitive sequences is beyond the scope of this study, we conducted an exploratory analysis of tandem repeat clusters in the representative set of mt-chromosomes identified above (figure 3).

These repeat clusters displayed two prominent patterns. First, individual repeats often showed greater similarity to repeats on other chromosomes than to repeats within the same cluster. Second, repeats flanking coding regions frequently occurred in opposite orientations, resembling terminal inverted repeats characteristic of linear organellar chromosomes.

Multiple repeat clusters showed similarity to the motifs used for BLAST-based chromosome identification, indicating that these motifs are part of larger repeat families (figure 3). Additional repeat clusters not detected by motif-based searches were also present. Beyond tandem repeats, we detected homopolymeric nucleotide runs of As and Ts and minisatellites composed primarily of GT and CA repetitions.

4. Discussion

(a) A multipartite, linear mitochondrial genome with extensive gene duplication and divergence

The mt-genome of *Sycon ciliatum* represents a striking departure from the small, circular architecture typical of most metazoans. It is both multipartite and linear and probably comprises thousands of distinct mt-chromosomes, each functioning as a semi-autonomous genetic element. Although linear and/or multipartite mt-genomes have evolved independently several times in Metazoa—most notably in medusozoan cnidarians [39]—and are also present in Calcinea, the sister group to calcaronean sponges [31], the extreme degree of fragmentation observed in *S. ciliatum* appears to be restricted to Calcaronea [18].

In addition to extensive fragmentation, most mt-coding sequences in *S. ciliatum* are present in multiple, highly divergent copies that often show evidence of mutation-induced degradation. Despite high copy number, typically only one or a few variants per gene family are predicted to produce full-length, correctly edited and presumably functional transcripts. This pattern contrasts with other calcaronean sponges, in which mt-gene diversification appears more limited and it helps to explain the difficulty of assembling *S. ciliatum* mt-coding sequences from short-read genomic data [18]. Notably, many mutated gene copies appear to be transcribed, albeit at reduced levels. Such transcription is expected to be deleterious because frameshifted coding sequences are likely to produce aberrant proteins with generic cytotoxic effects [40].

Outside animals, linear mitochondrial genomes occur in several lineages of green algae, protists and fungi [4]. Among these, the ichthyosporean *Amoebidium parasiticum*, one of the closest unicellular relatives of animals, exhibits the only other known mt-genome with complexity approaching that of *S. ciliatum* [19]. Like *S. ciliatum*, its mt-genome consists of hundreds of linear chromosomes containing multiple gene copies and fragments. However, mt-chromosomes in *A. parasiticum* are more heterogeneous in organization, comprising small gene-free molecules, single-gene chromosomes and larger multi-gene chromosomes [19].

(b) Insertional mRNA editing and its evolutionary parallels

Another distinctive feature of the *S. ciliatum* mt-genome is its reliance on insertional mRNA editing to restore coding capacity through the addition of one or two uridines at specific sequence motifs. Although insertional editing has evolved independently in viruses, bacteria, myxomycetes and kinetoplastid protists [41–43], the system in calcaronean sponges is unique in that both the position and the number of inserted nucleotides are largely predictable from short primary sequence motifs [18]. As a result, editing sites in calcaroneans function as genetic abbreviations that are expanded into complete coding sequences at the RNA level.

Mitochondrial mRNA editing is rare in animals, possibly reflecting high rates of mtDNA sequence evolution [44]. Outside calcaronean sponges, extensive editing has been reported only in the tunicate *Oikopleura dioica*, where long poly-A/T tracts are replaced by polyuridine stretches in transcripts [45]. Other reported cases are limited in scope and phylogenetic distribution [46–48], and some have subsequently been attributed to amplification artefacts [49].

Superficially, editing in *S. ciliatum* resembles that of kinetoplastid protists such as trypanosomes [50]. However, the underlying mechanisms appear fundamentally different. Editing in kinetoplastids is post-transcriptional, guided by small RNAs and mediated by a large multi-protein complex [51]. By contrast, no guide RNAs have been identified in *S. ciliatum*, and the presence of well-defined sequence motifs suggests a simpler, potentially co-transcriptional mechanism, possibly involving mitochondrial RNA polymerase.

Downloaded from royalsocietypublishing.org/rstb/article-pdf/doi/10.1098/rstb.2025.0082/6130399/rstb.2025.0082.pdf
by Flanders Marine Institute user
on 28 September 2026

(c) Repetitive sequences and genome architecture

The mt-genome of *S. ciliatum* is further distinguished by the abundance and diversity of repetitive sequences. Most occur as tandem repeats, although inverted repeats, homopolymeric tracts and interspersed minisatellites are also present. Tandem repeats are frequently arranged in opposite orientations upstream and downstream of coding regions, resembling terminal inverted repeats (TIRs) found in linear mt-genomes of fungi, protists and some plants [52]. In these systems, TIRs often play roles in chromosome replication, stability and transcription initiation, as shown for *Amoebidium parasiticum* [19]. The high similarity among repeat units across mt-chromosomes in *S. ciliatum* suggests ongoing recombination or concerted evolution.

(d) Loss or replacement of mitochondrial structural RNA genes

Despite extensive searches, we identified only short fragments of mt-rRNA genes and a limited number of low-confidence mt-tRNA candidates in the *S. ciliatum* mt-genome. It remains unclear whether this reflects true gene loss or extreme sequence divergence that prevents detection by standard approaches. However, the presence of numerous low-scoring tRNA and tRNA-like sequences in the nuclear genome of *S. ciliatum* raises the possibility that mitochondrial tRNAs are imported from the cytosol. Loss of mt-tRNA genes accompanied by tRNA import has occurred repeatedly during eukaryotic evolution and is relatively common among non-bilaterian animals [53].

By contrast, complete loss of mt-rRNA genes has not been conclusively demonstrated in organisms that retain mitochondria and mtDNA. Nevertheless, severe degeneration and fragmentation of mt-rRNA genes have been documented in apicomplexans [54], dinoflagellates [55], diplonemids [56], ctenophores [57] and appendicularians [45]. Our findings are consistent with an extreme extension of this trend in calcaronean sponges.

(e) Evolutionary and mechanistic implications

The multipartite, repetitive and linear architecture of the *Sycon ciliatum* mt-genome, together with its reliance on insertional mRNA editing, challenges prevailing models of animal mitochondrial genome evolution, maintenance and function.

A central question concerns how a complete and functional mitochondrial gene complement is maintained, and whether the mechanism of maintenance itself promotes heteroplasmy. In the absence of a genome-segregation pathway analogous to mitosis, inheritance of mt-chromosomes probably depends on the presence of multiple copies of each chromosome, such that they are transmitted statistically rather than deterministically. Although seemingly wasteful, comparable strategies occur in other systems, such as trypanosomes, where organellar DNA constitutes a substantial fraction of total cellular DNA [58]. Such inheritance is expected to promote the accumulation of pseudogenes and selfish chromosomes, as extensive duplication provides functional buffering. At the same time, the one-gene-per-chromosome organization characteristic of *S. ciliatum* should eliminate genetic linkage among loci, potentially increasing the efficiency of positive selection. This framework may explain the rapid evolution of editing sites [18] and the apparent transcriptional discrimination against mutated gene copies required to avoid cytotoxic effects. In this context, it is striking that other calcaronean sponges, such as *Petrobiona massiliana* and *Leucosolenia complicata*, maintain relatively homogeneous mitochondrial gene complements despite similarly fragmented mt-genomes.

A second key question concerns the evolutionary sequence in which these unusual features arose. Because multipartite organization, shared repeat motifs and rRNA fragmentation are also present in *C. clathrus*, a representative of Calcinea [31], these traits were probably inherited from the common ancestor of calcareous sponges. Subsequent divergence between Calcinea and Calcaronea therefore provides a powerful comparative framework for dissecting the evolution of mt-genome architecture, RNA editing and gene retention. Further fragmentation, extreme rRNA degeneration, tRNA gene loss and the evolution of insertional RNA editing in Calcaronea may have been facilitated by aspects of their biology, including hermaphroditism, unisexuality and rapid population cycles. Such traits could impose repeated population bottlenecks, reducing effective population size and relaxing purifying selection. Consistent with this view, the nuclear genome of *S. ciliatum* is itself highly repetitive and exhibits extensive gene family expansion [59]. Under this model, genome fragmentation precedes duplication and degeneration, with RNA editing evolving secondarily to preserve gene function.

(f) Limitations and future studies

Although the mt-genome size of *Sycon ciliatum* was estimated to be around 10 Mbp, several factors could have biased this estimate. Some of the longest sequences classified as mitochondrial may be chimeric, as they contain regions resembling nuclear DNA, including cytosolic rRNA and nuclear repetitive elements. Although these sequences could represent genuine mt-chromosomes, we consider this unlikely for two reasons. First, previous Southern blot analyses did not detect mitochondrial gene-containing fragments larger than 7 kbp [18]. Second, larger chromosomes would be expected to be outcompeted by smaller ones during replication, generating selective pressure toward reduced chromosome size. In our assembly, 57 contigs and 376 singlets longer than 7 kbp contributed 4.8 Mbp to the total length. Even if all such sequences were excluded, the estimated mt-genome size would still exceed 5.6 Mbp and include more than 1300 distinct chromosomes.

Conversely, standard PacBio library preparation protocols favour longer DNA fragments and are therefore expected to bias against shorter mt-chromosomes in *S. ciliatum*. The large number of singlets in the assembly suggests that additional mt-chromosomes may not have been sequenced, implying that the true genome size could be larger than estimated. Future studies combining long-read sequencing with library preparation methods that capture short DNA molecules will be required to refine

Downloaded from royalsocietypublishing.org/rstb/article-pdf/doi/10.1098/rstb.2025.0082/6130399/rstb.2025.0082.pdf
by Flanders Marine Institute user
on 28 September 2026

genome size estimates. Furthermore, comparative analyses across calcaronean species, together with biochemical and genetic characterization of the RNA editing machinery, will be essential for understanding the origin, maintenance and potential reversibility of this extraordinary mitochondrial system.

Finally, the exceptionally large number of mt-chromosomes raises a conceptual challenge. In such a highly fragmented system, it becomes unclear when sequence variants should be considered distinct chromosomes rather than alternative forms of the same chromosome. As a result, the concept of a single mitochondrial 'genome size' may lose much of its traditional meaning.

5. Conclusion

Our analysis of the mitochondrial genome of *Sycon ciliatum* reveals an unprecedented combination of structural and functional features that challenges prevailing views of mitochondrial genome evolution in animals. Its multipartite and linear organization, the apparent loss or extreme reduction of structural RNA genes and the extensive use of predictable insertional mRNA editing together define a highly derived organellar system that has evolved novel mechanisms to maintain gene expression. These features place calcaronean sponges among a small number of eukaryotic lineages, including kinetoplastids and dinoflagellates, that exhibit extreme mitochondrial genome architectures and demonstrate a surprising degree of evolutionary plasticity in animal mitochondria. Our results raise fundamental questions about the mechanisms of mitochondrial genome maintenance, inheritance and gene regulation in *S. ciliatum* and establish a framework for future experimental and comparative studies. As additional mitochondrial genomes from calcareous sponges and other non-bilaterian animals become available, comparative analyses will be essential to distinguish lineage-specific innovations from ancestral traits and to clarify the evolutionary forces that shape mitochondrial genome complexity.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. We used only publicly available datasets in this study, all of which are cited in the manuscript. All scripts and intermediate files are available in the accompanying GitHub repository: https://github.com/dlavrov/Sycon_ciliatum_mt-genome_analysis.

Supplementary material is available online [60].

Declaration of AI use. AI (ChatGPT 5.2) was used to create scripts for data processing and visualization as well as to check for grammar and spelling errors in the manuscript.

Authors' contributions. D.V.L.: conceptualization, formal analysis, methodology, visualization, writing—original draft; N.G.: data curation, formal analysis, methodology, resources, software, writing—review and editing.

Both authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. We acknowledge grants from Gordon and Betty Moore Foundation (GBMF8897) and Wellcome Trust (206194) that made previous work by N.G. possible.

Acknowledgements. We also acknowledge the Wellcome Sanger Institute Tree of Life Program and the whole team working on the Aquatic Symbiosis Genomics Project. The preliminary version of this manuscript was written by D.L. while visiting his mother in Ukraine. We thank the Armed Forces of Ukraine for holding the Russian army and making normal life activities possible in the country. Finally, we are grateful to the Editor and the three anonymous referees for thorough reading of the previous versions of the manuscript and for their many helpful suggestions.

References

1. Roger AJ, Muñoz-Gómez SA, Kamikawa R. 2017 The origin and diversification of mitochondria. *Curr. Biol.* **27**, R1177–R1192. (doi:10.1016/j.cub.2017.09.015)
2. Bobay LM, Ochman H. 2017 The evolution of bacterial genome architecture. *Front. Genet.* **8**, 72. (doi:10.3389/fgene.2017.00072)
3. Boussau B, Karlberg EO, Frank AC, Legault BA, Andersson SGE. 2004 Computational inference of scenarios for α-proteobacterial genome evolution. *Proc. Natl Acad. Sci. USA* **101**, 9722–9727. (doi:10.1073/pnas.0400975101)
4. Smith DR, Keeling PJ. 2015 Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proc. Natl Acad. Sci. USA* **112**, 10177–10184. (doi:10.1073/pnas.1422049112)
5. Burger G, Gray MW, Forget L, Lang BF. 2013 Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. *Genome Biol. Evol.* **5**, 418–438. (doi:10.1093/gbe/evt008)
6. Moreira D, Blaz J, Kim E, Eme L. 2024 A gene-rich mitochondrion with a unique ancestral protein transport system. *Curr. Biol.* **34**, 3812–3819. (doi:10.1016/j.cub.2024.07.017)
7. Lang BF, Gray MW, Burger G. 1999 Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* **33**, 351–397. (doi:10.1146/annurev.genet.33.1.351)
8. Boore JL. 1999 Animal mitochondrial genomes. *Nucleic Acids Res.* **27**, 1767–1780. (doi:10.1093/nar/27.8.1767)
9. Lavrov DV, Pett W. 2016 Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in nonbilaterian lineages. *Genome Biol. Evol.* **8**, 2896–2913. (doi:10.1093/gbe/evw195)
10. Kayal E, Lavrov DV. 2025 One ring does not rule them all: linear mtDNA in metazoa. *Gene* **933**, 148999. (doi:10.1016/j.gene.2024.148999)
11. de Voogd N. 2026 *World porifera database*. See https://www.marinespecies.org/porifera (accessed 22 January 2026).
12. Manuel M, Borojevic R, Boury-Esnault N, Vacelet J. 2002 Class Calcarea Bowerbank, 1864. In *Systema Porifera. a guide to the classification of sponges* (eds JNA Hooper, RWM Soest, P Willenz), pp. 1103–1116, vol. 2. New York, NY; Boston, MA; Dordrecht, The Netherland; London, UK; Moscow, Russian Federation: Kluwer Academic/Plenum Press. (doi:10.1007/978-1-4615-0747-5_115)
13. Ereskovsky AV. 2010 *The comparative embryology of sponges*, pp. 3–36. Dordrecht, The Netherlands: Springer. (doi:10.1007/978-90-481-8575-7_1)
14. Adamska M. 2021 Porifera. In *Handbook of marine model organisms in experimental biology*, pp. 67–77. Boca Raton, FL; London, UK; New York, NY: CRC Press, Taylor & Francis Group. (doi:10.1201/9781003217503-4)
15. Haeckel E. 1874 Die Gastraea-Theorie, die phylogenetische Classification des Thierreichs und die Homologie der Keimblätter. *Jenaische Zeitschrift Für Naturwissenschaft* **NF_1**, 1–55.
16. Franzen W. 1988 Oogenesis and larval development of *Scypha ciliata* (Porifera, Calcarea). *Zoomorphology* **107**, 349–357. (doi:10.1007/BF00312218)

Downloaded from royalsocietypublishing.org/rstb/article-pdf/doi/10.1098/rstb.2025.0082/6130399/rstb.2025.0082.pdf
by Flanders Marine Institute user
on 28 September 2026

17. Duboscq O, Tuzet O. 1935 Un nouveau stade du développement des éponges calcaires. *CR Acad. Sci. Paris* **200**, 1788–1790.
18. Lavrov DV, Adamski M, Chevaldonné P, Adamska M. 2016 Extensive mitochondrial mRNA editing and unusual mitochondrial genome organization in calcaronean sponges. *Curr. Biol.* **26**, 86–92. (doi:10.1016/j.cub.2015.11.043)
19. Burger G, Forget L, Zhu Y, Gray MW, Lang BF. 2003 Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc. Natl Acad. Sci. USA* **100**, 892–897. (doi:10.1073/pnas.0336115100)
20. Nash EA, Nisbet RER, Barbrook AC, Howe CJ. 2008 Dinoflagellates: a mitochondrial genome all at sea. *Trends Genet*. **24**, 328–335. (doi:10.1016/j.tig.2008.04.001)
21. Lukes J, Guilbride DL, Votýpka J, Zíková A, Benne R, Englund PT. 2002 Kinetoplast DNA network: evolution of an improbable structure. *Eukaryotic Cell* **1**, 495–502. (doi:10.1128/EC.1.4.495-502.2002)
22. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990 Basic local alignment search tool. *J. Mol. Biol*. **215**, 403–410. (doi:10.1016/S0022-2836(05)80360-2)
23. Huang X, Madan A. 1999 CAP3: a DNA sequence assembly program. *Genome Res*. **9**, 868–877. (doi:10.1101/gr.9.9.868)
24. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170)
25. Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. 2019 rnaSPAdes: a *de novo* transcriptome assembler and its application to RNA-Seq data. *Gigascience* **8**, 1–13. (doi:10.1093/gigascience/giz100)
26. Grabherr MG *et al*. 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol*. **29**, 644–652. (doi:10.1038/nbt.1883)
27. Li H *et al*. 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)
28. Mao A, Ji HJ, Haas BJ, Salzberg SL, Sommer MJ. 2025 TD2: finding protein coding regions in transcripts. BioRxiv 2025.04.13.648579 (doi:10.1101/2025.04.13.648579)
29. Langmead B, Salzberg SL. 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. (doi:10.1038/nmeth.1923)
30. Steinegger M, Söding J. 2017 MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol*. **35**, 1026–1028. (doi:10.1038/nbt.3988)
31. Lavrov DV, Pett W, Voigt O, Wörheide G, Forget L, Lang BF, Kayal E. 2013 Mitochondrial DNA of *Clathrina clathrus* (Calcarea, Calcinea): Six Linear Chromosomes, Fragmented rRNAs, tRNA Editing, and a Novel Genetic Code. *Mol. Biol. Evol*. **30**, 865–880. (doi:10.1093/molbev/mss274)
32. Wang X, Lavrov DV. 2006 Mitochondrial genome of the homoscleromorph *Oscarella carmela* (Porifera, Demospongiae) reveals unexpected complexity in the common ancestor of sponges and other animals. *Mol. Biol. Evol*. **24**, 363–373. (doi:10.1093/molbev/msl167)
33. Eddy SR. 2011 Accelerated profile HMM searches. *PLoS Comput. Biol*. **7**, e1002195. (doi:10.1371/journal.pcbi.1002195)
34. Chan PP, Lin BY, Mak AJ, Lowe TM. 2021 tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res*. **49**, 9077–9096. (doi:10.1093/nar/gkab688)
35. Rice P, Longden I, Bleasby A. 2000 EMBOSS: the European molecular biology open software suite. *Trends Genet*. **16**, 276–277. (doi:10.1016/S0168-9525(00)02024-2)
36. Morel B, Kozlov AM, Stamatakis A. 2019 ParGenes: a tool for massively parallel model selection and phylogenetic tree inference on thousands of genes. *Bioinformatics* **35**, 1771–1773. (doi:10.1093/bioinformatics/bty839)
37. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019 RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455. (doi:10.1093/bioinformatics/btz305)
38. Wickham H. 2016 *Ggplot2: elegant graphics for data analysis*. New York, NY: Springer-Verlag.
39. Kayal E, Bentlage B, Collins AG, Kayal M, Pirro S, Lavrov DV. 2012 Evolution of linear mitochondrial genomes in medusozoan cnidarians. *Genome Biol. Evol.* **4**, 1–12. (doi:10.1093/gbe/evr123)
40. Bucciantini M *et al*. 2002 Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**, 507–511. (doi:10.1038/416507a)
41. Brennicke A, Marchfelder A, Binder S. 1999 RNA editing. *FEMS Microbiol. Rev*. **23**, 297–316. (doi:10.1111/j.1574-6976.1999.tb00401.x)
42. Shaw JM, Feagin JE, Stuart K, Simpson L. 1988 Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons. *Cell* **53**, 401–411. (doi:10.1016/0092-8674(88)90160-2)
43. Bundschuh R, Altmüller J, Becker C, Nürnberg P, Gott JM. 2011 Complete characterization of the edited transcriptome of the mitochondrion of *Physarum polycephalum* using deep sequencing of RNA. *Nucleic Acids Res*. **39**, 6044–6055. (doi:10.1093/nar/gkr180)
44. Lynch M, Koskella B, Schaack S. 2006 Mutation pressure and the evolution of organelle genomic architecture. *Science* **311**, 1727–1730. (doi:10.1126/science.1118884)
45. Klirs Y, Novosolov M, Gissi C, Garić R, Pupko T, Stach T, Huchon D. 2024 Evolutionary insights from the mitochondrial genome of Oikopleura dioica: sequencing challenges, RNA editing, gene transfers to the nucleus, and tRNA loss. *Genome Biol. Evol* **16**, evae181. (doi:10.1093/gbe/evae181)
46. Riepsamen AH, Blok VC, Phillips M, Gibson T, Dowton M. 2008 Poly(T) variation within mitochondrial protein-coding genes in *Globodera* (Nematoda: Heteroderidae). *J. Mol. Evol*. **66**, 197–209. (doi:10.1007/s00239-007-9064-2)
47. Burger G, Yan Y, Javadi P, Lang BF. 2009 Group I-intron trans-splicing and mRNA editing in the mitochondria of placozoan animals. *Trends Genet.* **25**, 381–386. (doi:10.1016/j.tig.2009.07.003)
48. Vanfleteren JR, Vierstraete AR. 1999 Insertional RNA editing in metazoan mitochondria: the cytochrome b gene in the nematode *Teratocephalus lirellus*. *RNA* **5**, 622–624. (doi:10.1017/s135583829999009x)
49. Riepsamen AH, Gibson T, Rowe J, Chitwood DJ, Subbotin SA, Dowton M. 2011 Poly(T) variation in heteroderid nematode mitochondrial genomes is predominantly an artefact of amplification. *J. Mol. Evol*. **72**, 182–192. (doi:10.1007/s00239-010-9414-3)
50. Simpson L, Shaw J. 1989 RNA editing and the mitochondrial cryptogenes of kinetoplastid protozoa. *Cell* **57**, 355–366. (doi:10.1016/0092-8674(89)90911-2)
51. Aphasizhev R, Aphasizheva I. 2014 Mitochondrial RNA editing in trypanosomes: small RNAs in control. *Biochimie* **100**, 125–131. (doi:10.1016/j.biochi.2014.01.003)
52. Nosek J, Tomáška Ľ, Kucejová B. 2004 The chromosome end replication: lessons from mitochondrial genetics. *J. Appl. Biomed*. **2**, 71–79. (doi:10.32725/jab.2004.008)
53. Pett W, Lavrov DV. 2015 Cytonuclear interactions in the evolution of animal mitochondrial tRNA metabolism. *Genome Biol. Evol*. **7**, 2089–2101. (doi:10.1093/gbe/evv124)
54. Feagin JE, Mericle BL, Werner E, Morris M. 1997 Identification of additional rRNA fragments encoded by the *Plasmodium falciparum* 6 kb element. *Nucleic Acids Res*. **25**, 438–446. (doi:10.1093/nar/25.2.438)
55. Waller RF, Jackson CJ. 2009 Dinoflagellate mitochondrial genomes: stretching the rules of molecular biology. *Bioessays* **31**, 237–245. (doi:10.1002/bies.200800164)
56. Kaur B, Záhonová K, Valach M, Faktorová D, Prokopchuk G, Burger G, Lukeš J. 2020 Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplonemids. *Nucleic Acids Res*. **48**, 2694–2708. (doi:10.1093/nar/gkz1215)
57. Pett W, Ryan JF, Pang K, Mullikin JC, Martindale MQ, Baxevanis AD, Lavrov DV. 2011 Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi*: Insight from mtDNA and the nuclear genome. *Mitochondrial DNA* **22**, 130–142. (doi:10.3109/19401736.2011.624611)
58. Simpson L. 1973 Structure and function of kinetoplast DNA. *J. Protozool*. **20**, 2–8. (doi:10.1111/j.1550-7408.1973.tb05993.x)
59. Leininger S *et al*. 2014 Developmental gene expression provides clues to relationships between sponge and eumetazoan body plans. *Nat. Commun*. **5**, 3905. (doi:10.1038/ncomms4905)

Downloaded from royalsocietypublishing.org/rstb/article-pdf/doi/10.1098/rstb.2025.0082/6130399/rstb.2025.0082.pdf
by Flanders Marine Institute user
on 28 September 2026

60. Lavrov DV, Gettle N. 2026 Supplementary material from: Extreme Mitochondrial Genome Complexity in the Calcaronean Sponge Sycon ciliatum: Fragmentation, Gene Duplication, and mRNA Editing. Figshare (doi:10.6084/m9.figshare.c.8381730)

Downloaded from royalsocietypublishing.org/rstb/article-pdf/doi/10.1098/rstb.2025.0082/6130399/rstb.2025.0082.pdf
by Flanders Marine Institute user
on 28 September 2026