

Extremely Divergent Repeats Made the Largest Bilaterian Mitogenome in Photosymbiotic Bivalves

Andy D. Y. Tan

Andy.Tan@colorado.edu

University of Colorado Boulder

Ruiqi Li

University of Southern California

Jingchun Li

University of Colorado Boulder

Research Article

Keywords: photosymbiosis, mitochondrial genome, transposable elements, long-read sequencing, host-symbiont interactions

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11188068/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

The prevailing view that bilaterian mitogenomes are compact and evolutionarily conserved in size (~ 16k) and structure has been largely shaped by a scarcity of truly complete assemblies, particularly for marine invertebrates. Advances in long-read sequencing now reveal widespread structural complexities and non-coding region (NCR) expansion, thus challenging traditional models of organellar streamlining. Mitogenomic gigantisms through NCRs are exceptionally pronounced in certain bivalves, yet their evolutionary drivers and regulatory roles remain poorly understood. Because mitochondria mediate oxidative adaptation, NCR architecture may be driven by extreme metabolic conditions, such as the severe diurnal hyperoxia-to-hypoxia fluctuations inherent to photosymbiotic bivalves (Cardiidae). Here, we employ long-read sequencing to uncover novel mitogenomic features in photosymbiotic cardiid, linking mitochondrial structural evolution to host-symbiont metabolic ecology within this bivalve lineage.

Results

We report a scale of mitogenomic gigantism with a great structural diversity previously undocumented in bilaterians, culminating in the 92.7-kb circular genome of *Fragum sueziense*, the largest known across Bilateria. Across photosymbiotic bivalve lineages, mitogenome sizes range from 22 to 92.7 kb, far surpassing the 14 to 21 kb genomes of non-photosymbiotic counterparts. These expansions arise from elaborate, species-specific repeats within the major NCR which cannot be resolved using short-read sequencing alone. The extreme NCR expansions showed a strong phylogenetic signal, being concentrated in closely related photosymbiotic bivalves and largely absent in non-photosymbiotic relatives. Despite rapid divergence that prevents interspecies alignment, these regions contain supernumerary tRNAs, transposable-element fragments and novel open reading frames.

Conclusions

Relaxed constraints on NCR compactness—driven by repetitive element accumulation and structural rearrangement—may be ancestral in photosymbiotic bivalves and reflects selective pressures associated with intimate host-algal photosymbiosis. Our findings reveal underappreciated flexibility in bilaterian mitogenomes, and challenge the paradigm of strict mtDNA structural conservation.

BACKGROUND

Arisen as an endosymbiont of an α -proteobacteria origin, mitochondria are essential eukaryotic organelles most known for their ATP production (1). Common practice in feature annotation and structural description of mitochondrial genomes (henceforth mitogenomes) has given rise to the ‘textbook description’ of the metazoan mitogenome (2): 37 invariantly arranged and compactly ordered genes — 13 Protein Coding Genes (PCGs), 2 ribosomal subunit RNA genes (rRNAs), 22 transfer RNA genes (tRNAs), alongside a non-coding ‘control region’ — all nestled within a 16kb, matrilineally inherited circularized molecule (3, 4). Countless studies involving phylogenetics, genetic diversity and populational variability have long persisted with, and perpetuated said canonical model. Yet, numerous deviations from this entrenched paradigm are now recognized (2, 5, 6), including but not limited to (i) varying numbers of linear and circular mitogenome chromosomes; (ii) non-clonal inheritance such as heteroplasmy (7, 8) and doubly uniparental inheritance (DUI) that appears unique to certain Bivalvia lineages (9, 10); (iii) variations in number and location of genes, especially tRNAs (3, 11); (iv) variance and instability in structure and gene order in non-chordates (11); (v) variation in number, size and location of the so-called non-coding regions (NCRs) which include control regions (12–14).

A decade of advances in long-read sequencing technologies revealed that mitogenomic structural complexities may be pervasive, but previously undetected due to technological limitations. Truly complete mitogenomes, particularly those with extensive NCRs, can now be confidently characterized (15), while earlier approaches, including Sanger sequencing and short-read whole-genome sequencing, frequently failed to resolve these intricate regions (16–18). A reassessment of 125 vertebrate species using PacBio and Nanopore sequencing underscores this shift, producing mitogenomes markedly more complete than previously published references (19). In light of novel gene region duplications and hypervariable repeats largely noted in the NCR(s) of vertebrate mitogenomes (19), we see the urgent need for equivalent long-read-based examinations for invertebrates – particularly molluscs, for which only ~ 50 *complete* long-read mitogenomes are currently available on GenBank (assessed May 2024), despite an estimated 150 000 extant species in the ocean alone (20).

Some studies consider the control region the “longest NCR” in the metazoan mitogenome (12, 21) while others report multiple origins of replications and regulatory elements in multiple control regions within NCRs (15, 22). Generally, a control or displacement-loop (d-loop)-containing region contains the origin of replication for either strand of the mitogenome, with the one for heavy strand replication often found in the major NCR of a mitogenome (12, 23). In this paper, we broadly define an NCR as any mitogenome region that does not encode the conventional 37 genes. Accordingly, the “major NCR” is designated as the longest NCR. For us, an NCR starts immediately after the last nucleotide encoding for a stop codon of a coding region and ends a position before the start codon of another coding region. Biological functions of the mitochondrial NCRs, especially in invertebrates, have not been thoroughly assessed. Studies in vertebrate systems have shown that the NCRs contain certain conserved regions, representing sites for origin of mitochondria replication, transcription initiation, or other gene regulatory elements (24, 25). Some of these functions are achieved through the formation of secondary structures (24,26). Despite growing research on molecular architectures of metazoan NCRs, our understanding of its sequence evolution and ecological role in animal adaptation remains minimal.

Bivalves exhibit a striking divergence from the canonical ≈ 16 kb blueprint; examples include the venerid clam *Ruditapes philippinarum* (≈ 20 kb), scallops such as *Placopecten magellanicus* (31–41 kb) and the ark-clam *Anadara broughtonii* which reaches ≈ 47 kb. These expansions are rarely driven by an increased repertoire of canonical protein coding genes. Instead, they result from the proliferation of NCRs, tandem repeats, lineage-specific ORFs and extensive rearrangements (27, 28), as well as DUI (29). While lineages within Bivalvia exhibiting such expansions suggest a hitherto untested phylogenetic propensity, such genomic gigantism indicate a significant relaxation of the purifying selection that typically maintains mitochondrial streamlining (30). Theoretical frameworks generally favour compactness for mitogenome sizes based on four primary non-mutually exclusive constraints:

1. **Mutational Burden:** Large organellar genomes present a greater target for deleterious mutation (31), in particular with close proximity to highly mutagenic reactive oxygen species (ROS) (32), which in turn augments
2. **Muller’s Ratchet:** in the absence of recombination, most asexual mitogenome lineages are prone to the rapid irreversible accumulation of deleterious mutation (32, 33)
3. **Energetics and Replicative Competition:** Smaller genomes replicate faster with fewer errors, and are less energetically costly, offering a “selfish” intracellular selective advantage (34)
4. **Symbiotic Reduction:** in (endo)symbiotic relationships, genome typically undergo massive reduction via gene loss or transfer or delegation of coding capacity to the partner (35–38)

As such, mitogenome expansions contradict our expectations of selection on genome size.

As mitochondria mediate oxidative metabolism, animals inhabiting extreme oxygen environments – whether hypoxic or hyperoxic – are expected to exhibit unique mitochondrial gene regulations reflecting their altered energetic demands (39, 40). Given the regulatory role NCRs may have, their sequence evolution can be shaped by environmental oxygen regimes. For instance, extreme oxygen levels will assert high oxidative stress to the mitochondria (41) and possibly increase DNA replication error and mutation rates, further impacting NCR evolution. Although a direct causal link between NCR evolution and oxygen regimes remains unestablished, recent complete metazoan mitogenomes offer emerging evidence: elapid sea snakes with low blood oxygen levels to support cutaneous respiration possess unusually longer NCRs characterized by

tandem repeats and lack of conserved sequence blocks compared to their terrestrial relatives (15, 42) and the deep sea chemosymbiotic bivalve *Calyptogena marissinica* exhibits expanded NCR elements, which were hypothesized to aid in mitochondrial energy metabolism in the deep sea (43).

Photosymbiosis, where a heterotrophic host harbours photosynthetic symbionts, presents a uniquely extreme oxygen microenvironment with dramatic diurnal shifts in oxygen levels. High rates of net photosynthesis create daytime hyperoxia within the host, while holobiont respiration results in nocturnal hypoxia or even anoxia (44, 45). This extreme fluctuation poses formidable challenges on mitochondrial gene regulation, raising the question whether NCRs in photosymbiotic animals bear unique evolutionary signatures. Extant, obligate photosymbiotic bivalves are confined to shallow-water Cardiidae – giant clams (Tridacninae) and certain heart cockles (Fraginae) maintain photosymbiotic relationship with dinoflagellates from the family Symbiodiniaceae (Li et al, 2018). These photosymbiotic bivalves occupy overlapping ecological niches to their shallow-water, infaunal to epifaunal non-photosymbiotic cardiid relatives (46, 47), making them the prime system for comparative genomic studies. Moreover, comparative transcriptomics already hint at unique molecular responses to ROS for one of these symbiotic bivalves (48). While transposable elements (TE) are reported in their nuclear genomes (49), their extent and impact on mitogenome architecture and performance remain largely unexplored. This study represents an initial foray into this uncharted territory, uncovering novel metazoan mitogenomic features and structures potentially linked to photosymbiotic ecology.

MATERIALS AND METHODS

Study System and Complete Mitogenome Sourcing

Complete mitogenomes of seven photosymbiotic cardiids – *Fragum fragum*, *Fr. whitleyi*, *Fr. sueziense* (subfamily Fraginae), *Tridacna crocea*, *T. derasa*, *T. gigas* and *Hippopus hippopus* (subfamily Tridacninae), and two non-photosymbiotic cardiids – *Cerastoderma edule* and *Cerastoderma glaucum* (subfamily Lymnocardinae), were sequenced and assembled through long-read sequencing as part of the Aquatic Symbiosis Genomics Project by the Wellcome Sanger Institute (50). Detailed methods and genome statistics are described in the published genome notes (Table S1). Mitogenomes of three more non-photosymbiotic cardiids: *Acanthocardia tuberculata* (subfamily Lymnocardinae), *Fulvia mutica* (subfamily Cardiinae) and *Laevicardium crassum* (subfamily Laevicardiinae) were obtained from NCBI Genbank for analyses. For phylogenetic signal testing analyses, five long-read assembled mitogenomes from five Cardida families were also downloaded from NCBI Genbank as outgroups – *Macomangulus tenuis* (Tellinidae), *Abra segmentum* (Semelidae), *Donax variegatus* (Donacidae), *Azorinus chamasolen* (Solencurtiidae), and *Gari tellinella* (Psammobiidae). NCBI GenBank accession numbers, sequencing technology, genome coverage, taxonomic classification and justification of use for all species can be found in Table S1.

NCR Verification through Illumina Sequencing

Genomic DNA for two other distinct individuals of *Fr. fragum* and *Fr. whitleyi* collected from Guam (deposited in the University of Colorado Museum of Natural History collection, USA) were extracted using the E.Z.N.A.® Mollusc DNA Kit, per manufacturer's instructions. Sample libraries were prepped and sequenced as paired-end 150 bp reads on the Illumina Novaseq S4 platform at the University of Colorado BioFrontiers Institute NGS Facility (NCBI Bioproject PRJNA1162570). Reference-guided assembly mapping of short reads to PacBio mitogenome using *SAMtools* and *BCFtools* were conducted to verify the presence, position, alignment and NCR coverage (51, 52).

Annotation and Comparative Analysis of Mitogenomes

Mitogenome annotations were performed and cross-checked using MITOS2 v2.1.8 (usegalaxy.org)(53), and GeSeq (Chlorobox)(54), incorporating the re-annotation of *atp8*, previously reported as missing. PCG boundaries were verified using NCBI ORFfinder. Transfer RNAs, identified via tRNAscan-SE (55) and ARWEN (56), were screened for canonical clover-shaped

secondary structures using UNAFold v2.3 (25°C folding temperature set)(57). Linearised and circular mitogenomic maps were generated through custom R scripts and OGDRAW (58), respectively (Figs. S1-S12; complete annotations in Data S1). Comparative gene block synteny analyses (order and rearrangement) were visually inspected and illustrated in Adobe Creative Suite. Purple ribbons on the synteny plot connect the most parsimonious deduction of interspecies mapping for homologous blocks to illustrate syntenic conservation and rearrangement points.

Duplications and repeat motifs were identified and visualized via NCBI Blast dotplots and detected using Tandem Repeat Finders v4.09 (59). TE-encoded proteins were identified using RepeatMasker's Interspersed Repeat Protein Masking Based on Protein Similarity program (60). Candidate is Open Reading Frames (ORFs) were delineated using NCBI ORFfinder (≥ 75 -nucleotide cut-off; invertebrate genetic code; canonical and alternative initiation codons). The primary dataset retains only ORFs matching TEs/TE fragments and prominent tandem repeats satisfying specific criteria: (i) consensus size ≥ 100 nt OR (ii) copy number ≥ 10 , AND (iii) percent matches $\geq 98\%$.

Phylogenetic Signal Tests

To characterize phylogenetically informed mitogenome trait evolution, we quantified phylogenetic signals for five parameters: (i) log-transformed mitogenome size, (ii) log-transformed NCR size, (iii) logit-transformed NCR proportion, (iv) ORF count, and (v) TE counts. Pagel's λ was estimated via maximum-likelihood (*phytools* v2.4.4; *phylosig*), assessing deviations from phylogenetically independence ($\lambda = 0$) via likelihood ratio tests. Bloomberg's K was computed, against Brownian motion expectations employing 10,000 randomized tip permutations.

To examine whether photosymbiosis drives genome architecture, Phylogenetic Generalized Least Squares (PGLS) regressions were fitted (*caper* v1.0.4; *pgls*) using the model: $Trait \sim \beta_0 + \beta_1 \times Photosymbiosis$, enabling assessment whether the presence of photosymbiosis explains phylogenetically structured trait variations. Analyses were executed in R v4.5.2 (RStudio v2025.09.2) employing a fully ultrametric, time calibrated phylogeny derived by pruning existing dated bivalve phylogenies (47, 61).

RESULTS

Complete circular mitogenomes were verified, annotated and evaluated across structural parameters, including overall genome and NCR sizes, gene annotations, order and rearrangements, repeat patterns, ORFs and TEs. Annotation elements for both *Cerastoderma edule* mitogenomes (OX401423.1 and OZ022487.1) virtually identical; hence the latter was excluded from subsequent analyses to balance species representation. Cardiid phylogeny and relative mitogenomic NCR proportions are depicted in Fig. 1a.

Large Mitogenome and NCR Expansion in Photosymbiotic Species

Cardiid mitogenomes appears to have abandoned the architectural austerity, departing starkly from the 15–16 kb mitochondrial blueprints expected for bilaterians, reaching $\sim 6\times$ the expected size at 92,770 kb in *Fr. sueziense*, which we report as the largest known circular mitogenome in bilaterians. Photosymbiotic species ($n = 7$) range from 21.8 to 92.8 kb (median: 25.3 kb; mean: 35.2 kb). These expansions are almost entirely driven by a single major NCR (> 5.99 kb across all photosymbiotic taxa; mean NCR size ~ 18.4 kb; Table 1). The largest NCRs occur in *Fr. sueziense* (75.7 kb; 82% of total genome) and *Fr. whitleyi* (13.8 kb; 45%), whereas *H. hippopus* exhibits the smallest major NCR among photosymbiotic species (5.99 kb; 27.5%).

Table 1
Comparison of mitogenome parameters.

Ecology	Species	(Sub)family	Mtg Size	NCR Size	NCR %	ORF count	Repeats
P	<i>Fragum fragum</i> (Linnaeus, 1758)	Fraginae	22 357	6 638	29.69	368	Compound Pattern A
P	<i>Fragum whitleyi</i> Iredale, 1929	Fraginae	30 336	13 752	45.33	580	Compound Pattern A-like
P	<i>Fragum sueziense</i> (Issel, 1869)	Fraginae	92 770	75 654	81.55	924	Compound Pattern A-like
P	<i>Tridacna crocea</i> Lamarck, 1819	Tridacninae	29 009	11 083	38.21	x476	Compound Pattern B
P	<i>Tridacna derasa</i> (Röding, 1798)	Tridacninae	24 955	7 110	28.49	402	Compound Pattern B
P	<i>Tridacna gigas</i> (Linnaeus, 1758)	Tridacninae	25 335	8 785	34.68	381	Compound Pattern B
P	<i>Hippopus hippopus</i> (Linnaeus, 1758)	Tridacninae	21 808	5 992	27.48	359	Compound Pattern B
NP	<i>Laevicardium crassum</i> (Gmelin, 1791)	Laevicardiinae	21 427	6 517	30.41	329	3 minisatellites
NP	<i>Fulvia mutica</i> (Reeve, 1844)	Cardiinae	19 110	2 612	13.67	293	1 minisatellite
NP	<i>Acanthocardia tuberculata</i> (Linnaeus, 1758)	Lymnocardiinae	16 104	1 101	6.84	281	1 minisatellite
NP	<i>Cerastoderma edule</i> (Linnaeus, 1758)	Lymnocardiinae	14 943	267	1.79	237	Absent
NP	<i>Cerastoderma glaucum</i> (Bruguère, 1789)	Lymnocardiinae	14 950	268	1.79	244	Absent
O	<i>Donax variegatus</i> (Gmelin, 1781)	Donacidae	22 790	7 261	31.86	412	NA
O	<i>Macomangulus tenuis</i> (da Costa, 1778)	Tellinidae	17 125	1 469	8.58	242	NA
O	<i>Abra segmentum</i> (Récluz, 1843)	Semelidae	16 851	1 271	7.54	264	NA
O	<i>Gari tellinella</i> (Lamarck, 1818)	Psammobiidae	18 464	2 652	14.36	289	NA
O	<i>Azorinus chamasolen</i> (da Costa, 1778)	Solencurtiidae	17 135	1 314	7.67	264	NA
Genome parameters include mitogenome overall sizes (Mtg Size), major NCR sizes, major NCR to full mitogenome proportion (NCR %), ORF count (for ORF > 75 nucleotide in length) and repeat types found in each mitogenome. PB : PacBio Long-Read Sequencing with Hi-C sequencing; IL : Illumina Short-Read Sequencing; PW-PCR : Primer-Walking PCR; ONT : Oxford Nanopore Technologies. P : photosymbiotic, NP : non-photosymbiotic, O : outgroups (none photosymbiotic).							

Conversely, non-photosymbiotic cardiid mitogenomes span 14.9 to 21.4 kb, with NCRs comprising 1% to 14% of total size, except in *L. crassum* (6.5 kb NCR; 30%). The major NCRs in *Cerastoderma* spp. remain below 300 bp. Notably, major NCRs in *Fu. mutica* and *H. hippopus* are interrupted by a small coding blocks encoding for two tRNAs and *cox2*, respectively. Comparative parameters are summarised in Table S1, Fig. 1b and Fig. 2.

Additional Verification of NCR Presence/Absence

Long-read sequencing confirmed expansive major NCR containing long repeats (> 200 bp) in the mitogenomes of all photosymbiotic species. Reference-guided assemblies mapping Illumina short reads (150 bp) to PacBio assemblies for *Fr. fragum* and *Fr. whitleyi* validated the presence and length of the major NCR (> 350 bp repeats at > 11 times fold coverage). Elevated read coverage in these repetitive regions introduces ambiguous assembly, illustrating the limitations of short reads alone in resolving exact repeat identities.

Mitogenome Composition

All taxa harbour the complete set of standard 37 mitochondrial genes –13 PCGs (*cox1-3*, *atp6*, *atp8*, *nad1-6*, *nad4l*, *cytb*), 2 rRNAs (*lrRNA* and *srRNA*) and 22 tRNAs (including duplicate *trnL* and *trnS*; **Table S2**, **Fig. S14**). Nucleotide compositions and biases are detailed in Table S5–S7. Reconstructed 15-gene maximum-likelihood phylogenies validated annotation accuracy and taxon identity (**Fig. S16**).

All cardiid species in this study possess an additional *trnM*, with *Fu. mutica* and *H. hippopus* containing two extra copies. Supernumerary tRNAs (3 to 24 copies) were annotated across all species, except *T. derasa* and *A. tuberculata*. Unprecedented tRNA expansions occur within the major NCRs of two *Fragum* species: *Fr. fragum* exhibits 11 supernumerary *trnS* copies evenly distributed within the long NCR repeats (alongside two canonical coding copies; Fig. 2A, **Fig. S1**), whereas *Fr. sueziense* contains 10 additional *trnN* and 11 additional *trnS* copies. These tRNA duplications events have not been previously reported in the literature. Five *trnC* copies occur within the NCR of non-photosymbiotic *Fu. mutica*, consistent with previous study (17).

High Levels of Mitochondrial Gene Rearrangements within Cardiida

Extensive gene rearrangements have been reported in bivalves (3, 62, 63), and our study supports the phenomenon not only for minor elements (tRNAs only), but also for PCGs and rRNAs (Fig. 3). Intrageneric gene order is more conserved than intergeneric order. Between sister clades Fraginae and Tridacninae, *cox2* exhibits four distinct genomic positions, whereas *nad3*, *nad4*, *nad4l*, and *atp8* each occupy two.

Major NCR loci are conserved within genera but diverge across Cardiidae, arising at ≥ 6 distinct loci: between *nad1* and *cox1* in Fraginae, between *atp8/nad4* and *cox2* in Tridacninae, and differently placed for each non-photosymbiotic genera. These distributions suggest up to six independent origination events, or fewer rearrangement events prior to lineage-specific NCR expansions echoing patterns reported in various pectinid lineages (64).

Repeat patterns in the major NCRs

Major NCRs are dominated by sequence repeats (Table 1, Fig. 2). Repeats in non-photosymbiotic species, if present, are restricted to minisatellites, and are much simpler compared to those in photosymbiotic species (Fig. 1c).

In *Fr. Fragum*, the NCR displays a compound intersperse repeat sequence, in a “(repeat motif A – individual/species specific sequence repeat motif B)_n” form, which we term ‘compound pattern A’. We observe a ~ 350 bp-motif repeated > 11 times, each separated by > 100 bp of species-specific sequences (Fig. 2D). This repeat pattern appear to have small proportions of interspecies match, and are expected to be overlooked between two closely related species (65).

In *Tridacna*, repeat patterns are more conserved, featuring two distinct repetitive regions, in the form of “(repeat motif K)_x – (repeat motif L)_y, where $y > x$ ”, which we term ‘compound pattern B’. Both photosymbiotic subfamilies exhibit complex repeat patterns combining compound interspersed elements and satellite blocks (Fig. 1c(B-F), Fig. 2), containing nested tandem repeats of varying period length and copy numbers (Data S2). To our knowledge, these types of repeat patterns have never before been described.

TE in the mitogenomes and ORFs in major NCRs

Identified TE fragments spanned 9 to 3318 codons (median: 25; mean: 50; mode: 21; **Data S3**) and frequently coincided with ORFs of 75 to 4263 nt (median: 126; mean: 180; mode: 81; Fig. 2, **Data S4**). Non-photosymbiotic cardiid species contain 30 to 39 total masked TE fragments (≤ 13 within NCRs; **Table S2**). Conversely, photosymbiotic species exhibit wider TE range (29 to 332), peaking in *Fr. sueziense* (332 total; 318 in NCR) and *T. crocea* (94 total; 62 in NCR).

Long Terminal Repeats (LTR) retrotransposons and Long Interspersed Nuclear Elements (LINEs) predominate, notably LTR/*Copia*, LTR/*Gypsy* and LINE/*L1* types (**Table S3**). Unlike plants (e.g., *Arabidopsis*), where retrotransposons show no integration hotspots (66) photosymbiotic cardiid NCRs are heavily enriched for specific TE fragments. LTR/*Gypsy*-types dominate *Fragum* NCRs. LTR/*Copia*-types are co-dominant with LTR/*Gypsy*-types in *Fr. fragum* and *Fr. sueziense*, and with LINE/*L1*-types in *T. crocea*. LTR/*Pao*, LINE/*L1* and RC/*Helitron*-type TEs are also abundant in *Fr. sueziense*. RC/*Helitron*, LTR-*ERV*K and LINE/*Ambal*-types represent the primary repeat classes in *T. derasa*, *T. gigas* and *Fu. mutica*, respectively. Lacking extensive NCRs with repetitive sequences, *A. tuberculata* and *C. edule* display no localized TE enrichment.

Phylogenetic Signal Testing

Phylogenetic signal analyses revealed contrasting evolutionary patterns among mitochondrial traits. NCR size and proportion exhibited strong phylogenetic signal (Pagel's $\lambda > 0.9$, $P < 0.05$; Bloomberg's $K \approx 0.55$, $P < 0.01$; **Table S4**, Fig. 4). Conversely, overall mitogenome size, ORF counts, and TE counts showed negligible signal (Pagel's $\lambda \approx 0$, $P > 0.05$), indicating evolutionarily lability (effectively independent of shared ancestry). In PGLS models (Trait $\sim$ Photosymbiosis), photosymbiotic state was a significant predictor for all traits ($\lambda = 0$ in model residuals). This demonstrates that while NCR size and proportion are phylogenetically structured, this structure is fully accounted for by photosymbiosis, with no remaining phylogenetic autocorrelation. NCR traits appear tightly associated with a phylogenetically clustered ecological transition to a photosymbiotic lifestyle.

DISCUSSIONS

Repetitive Elements and Gene Duplications in Expanded NCR as the Norm

The application of long-read sequencing reveals that the architectural austerity typically attributed to metazoan mitogenomes is largely an artefact of short-read assembly limitations. By bridging previously undetected, or unresolvable repetitive elements, more complete reference mitogenomes could be assembled. For example, we recovered mitogenomes for *T. crocea*, *T. derasa* and *T. gigas* (25–29 kb) with > 7 kb NCR, which are significantly larger than prior short-read references of the same species (18–20 kb) (67–69). The inflation is driven almost exclusively by massive NCRs in photosymbiotic cardiid mitogenomes. While similar expansions have been noted across diverse taxa (from vertebrates (19) to other invertebrates including freshwater amphipod (70), sawflies (71), molluscs (2), and demosponges (72)), the magnitude of expansion in photosymbiotic bivalves is exceptional.

Our findings suggest that expanded NCRs may be a widespread metazoan feature rather than an anomaly. Consequently, we propose a re-evaluation of completeness in presently published references, and advocate for long-read sequencing as the gold standard for verifying mitogenome completeness. Similarly, *atp8*, which has been repeatedly reported absent in bivalve mitogenomes (73–75), was annotated for all species in this study. Consistent with recent analyses finding *atp8* in various bivalves (76), including Mytilidae (76, 77) and Pectinidae (63), our findings support its presence in Cardiidae, and flags the need rigorous annotation pipelines before asserting mitochondrial gene loss (78).

NCR Origin and Evolution

Photosymbiotic bivalves from two cardiid subfamilies exhibit exceptionally long NCRs compared to non-photosymbiotic cardiid species. Our finding of a strong phylogenetic signal for NCR size and proportion suggests that these “genomic

deserts” are not merely stochastic accumulations, but are phylogenetically structured features that are likely plesiomorphic to the two photosymbiotic lineages, and potentially their relatives, the clade containing *Fulvia* and *Laevicardium*. Interestingly, while the NCR itself remains a conserved hallmark, the overall mitogenome size, ORFs and TEs show a weak correspondence with phylogenetic structure and behave as a more labile genomic feature. This discrepancy suggests that while the “core” of the NCR expansion was likely a singular ancestral event, subsequent lineage-specific turnover have obscured the phylogenetic signal in the overall genome size, effectively diluting the signal of the ancestral expansion with the noise of rapid genomic flux.

The inextricable link between NCR expansion and the photosymbiotic state raises a compelling question of causality. While our PGLS models indicate that the photosymbiotic lifestyle is a significant predictor of NCR traits, with the ecological transition accounting for the observed phylogenetic autocorrelation, the sister-clade relationship of Fraginae and Tridacninae complicates the decoupling of phylogeny from ecology, making it hard to tease out the causality between photosymbiosis and NCR expansion. Directly evaluating the functional significance of expanded mitogenomic NCRs in photosymbiosis may ultimately require experimental genetic manipulation in model systems such as *Exaiptasia* anemones. Nevertheless, given the hypothesized independent evolution of photosymbiosis in Fraginae and Tridacninae despite their sister relationship (47), their convergent acquisitions of expansive NCRs strongly point to a functional association(s) with their specialized niche. Such mitogenomic reorganisation may reflect altered selective pressures unique to their solar-powered lifestyle, where relaxed constraints on mitochondrial streamlining or a genomic response to the metabolic shifts inherent to hosting intracellular symbionts permit the evolutionary persistence of sprawling NCRs.

Decoupling NCR Expansion from the DUI hypothesis

While DUI is a known driver of mitogenomic complexities in several bivalve families (e.g. Mytilidae, Veneridae and Unionidae), it is an unlikely explanation for the architectural shifts observed here. Under DUI, maternally transmitted (F-type) mitochondria exist in both sexes, while paternally transmitted (M-type) mitochondria are sequestered in male germline, resulting in distinct male heteroplasmy (79, 80). In these systems, large NCRs with complex genetic elements and gender specific functions often emerge from duplication and recombination between divergent mitogenomic lineages in heteroplasmic males (10, 81, 82), or via relaxed selection and rapid evolution on the M-type genome (83). However, sex-linked heteroplasmy is reported absent in *Cerastoderma glaucum* (84), and DUI that tend to be taxonomically conserved has never been reported within Cardiidae. Furthermore, we did not observe any evidence of strong heteroplasmy in our mitogenome assemblies, therefore concluding that cardiid NCR expansion is an independent phenomenon and propose three additional non-mutually exclusive ecological hypotheses that could potentially explain the association between NCR expansion and a photosymbiotic lifestyle. Further comparative genomic and experimental studies, particularly of non-photosymbiotic Fraginae and across additional photosymbiotic systems, will be needed to evaluate these hypotheses.

Hypothesis I: Solar-Powered Scars from ROS and UV Catalysed NCR Expansion

Photosymbiosis imposes a chronic oxidative burden on host tissues, as algal photosynthesis generates hyperoxic conditions and elevated ROS levels (85). Symbiont-hosting tissues also endure prolonged sun exposure to sustain ample light supply to the algae. Unlike the nuclear DNA, the mtDNA lacks protective histones and resides in immediate proximity to ROS production sites, rendering it acutely vulnerable to oxidative damage and subsequent ATP depletion (86, 87). While cardiids employ optical strategies (such as shell/tissue optical arrangement (88, 89)) to mitigate UV and ROS exposure, we postulate that these stressors saturate endogenous repair mechanisms, inducing higher replication errors and slippage events that fuel NCR expansion (90). Under this hypothesis, the highly variable, species-specific repeat sequences observed are the genomic scars of a solar-powered existence, beyond a mere neutral “bloat”; the necessity of light exposure overrides the typical selective pressure for mitochondrial streamlining.

Hypothesis II: Suppressed Immune System Lacking in Ability to Curb TE Proliferation

The rapid expansion of repetitive elements in photosymbiotic cardiid is not limited to the mitochondrial genomes; parallel expansions occur within their nuclear genomes (91). This cross-compartmental expansion appears phylogenetically widespread among photosymbiotic metazoans, from giant clams (exhibiting ~ 70% repeat content) to zoantharians, anemones and stony corals (92, 93). Consequently, oxidative stress alone is unlikely the sole driver of rapid genomic gigantism. Instead, we propose that the physiological “truce” required for endosymbiosis provides a uniquely permissive window for TE proliferation.

A compelling mechanism involves the necessary systemic downregulation of host immune defences to accommodate algal endosymbionts, as documented in corals (94) and heart cockles (48). Immune-related pathways involving epigenetic modifications like DNA methylations are not only critical in response to bacterial infections, but also for TE suppression (95, 96). Furthermore, prevalent symbionts like Symbiodiniaceae actively modulate host immunity (97). Beyond endogenous proliferation, intimate host-symbiont associations can serve as highways for horizontal transposon transfers, as observed in other host-parasite and symbiotic systems (98, 99). As such, long-term immune suppression for the establishment and maintenance of a stable photosymbiotic relationship may thus impair host TE silencing, providing an additional TE source that drives expansion across both nuclear and mitochondrial landscapes.

Within the examined mitogenomes, dominant TE sequences appear fragmentary, likely originating from partial duplications or ancestral insertions that subsequently degenerated (66). A single degeneration event followed by duplication represents the more parsimonious trajectory, in contrast to concomitant degenerations of each repeated TE fragment. The presence of plant-associated elements (e.g., Gypsy- and Copia-like LTRs) within metazoan mitogenomes highlights a remarkable architectural shift, likely permitted by relaxed genomic surveillance inherent to the symbiotic lifestyle. Comparative studies across multiple independent photosymbiotic and non-photosymbiotic sister clades are warranted to test this hypothesis.

Hypothesis III: Solar Subsidy for Functional Innovation

While mitochondrial streamlining is typically driven by the energetic advantages of rapid replication (34), the long-term maintenance of hypertrophied mitogenome suggests these regions are not merely deleterious “bloat”. We propose that the metabolic surplus provided by photosymbiosis offsets the cost of replicating large mitogenomes, allowing these sprawling regions to serve as laboratories for functional innovation. Our annotation of the NCRs identified various additional RNA components, including 11 supernumerary tRNAs in the NCR of *Fr. fragum* that retained canonical cloverleaf secondary structures, which is highly unusual for Metazoan. tRNAs are crucial for mitochondrial gene translation and many organisms need to import nucleus-encoded tRNAs to facilitate the process (100). Additional tRNA repeats may confer a selective advantage by enhancing mitochondrial gene translation and replication efficiency (101, 102), as a compensatory mechanism to maintaining proteostasis in their hyperoxic, ROS-rich environments.

Alternatively, these supernumerary tRNAs may serve dual functions via molecular exaptations; non-canonical processing of these RNAs could yield general small regulatory RNAs, such as small mitochondrial highly transcribed RNAs (smRNA) and mitochondrial transfer RNA fragments (mt-tRFs), which could modulate downstream transcription within or outside mitochondrion (103, 104). Such expanded NCRs provide an ideal, vast physical space for RNA-based regulatory mechanisms, an asset that may be particularly desirable for managing host-endosymbiont metabolic crosstalk. This phenomenon could also generally be the case with increased mitogenome size, independent of photosymbiosis.

Beyond translational and regulatory machinery, these NCRs exhibit substantial coding potential. The prevalence of ORFs indicates roles in expression modulation or *de novo* protein origination. Notably, overlaps between ORFs and TEs suggest that TEs may be transcribed, or co-opted as promoters, enhancers, transcriptional effectors, or gene silencers (105). Given the numerous TE insertions in the photosymbiotic bivalves, certain elements may have been “domesticated”, serving as an available genomic reservoir for novel protein coding genes and non-coding RNAs (106). Furthermore, although primary repeat sequences are highly divergent, selection may conserve their secondary structures to provide a structural framework for gene regulation (25, 107). Collectively, these features suggest that the NCR expansion could arise as a mutational byproduct

turned functional “hotbed”, where relaxed selective pressure for brevity has permitted the evolution of complex mitochondrial regulatory landscapes.

Disentangling the exact evolutionary mechanisms -- such as recent transposition bursts, horizontal transfer, or migration from the host nuclear genome -- remain an important priority. Future work integrating paired nuclear and mitogenomic resources will be necessary to unravel these complex cross-compartmental dynamics. Ultimately, our proposed hypotheses are not mutually exclusive; metabolic subsidies, relaxed constraints and molecular exaptations presented herein are compatible and likely act in synergy to drive and maintain mitogenomic gigantisms.

CONCLUSIONS

Our study sheds light on previously underappreciated yet ecologically significant aspects of animal mitogenome evolution: NCR structures and functional annotations. Using cardiid as an exemplar, our findings implicate a compelling association between endosymbiotic lifestyles and mitogenome complexity. Apart from contributing to one of the largest mitogenomes documented amongst Metazoa (*Frugum sueziense*), the inflated major NCRs in photosymbiotic bivalves, with either several independent origins and frequent rearrangements, have complex repeat topologies that warrant deeper characterizations. Many ORFs in the NCRs might contain more unrecognized genetic components, but with hence far unknown functions, including active TEs and other regulatory elements. While we have explored several plausible causes for having expanded NCRs, comparative phylogenetic analyses incorporating other bivalve lineages with expanded NCRs (such as actively swimming or periodically oxygen-deprived pectinids) will be essential to disentangle whether, and how, photosymbiosis uniquely governs mitogenome evolution.

Abbreviations

ATP

adenosine triphosphate

DNA

Deoxyribonucleic acid; **mtDNA**:mitochondrial DNA

DUI

Doubly Uniparental Inheritance

Mitogenome

mitochondrial genome

NCR

Non-Coding Region

ORF

Open Reading Frames

PCG

Protein Coding Genes

RNA

ribonucleic acid; **rRNA**:Ribosomal RN; **tRNA**:Transfer RNA; **smithRNA**:small mitochondrial highly transcribed RNA; **mt-tRFs**:mitochondrial transfer RNA fragments

ROS

reactive oxygen species

TE

Transposable Elements; **LTR**:Long Terminal Repeats retrotransposons; **LINEs**:Long Interpreted Nuclear Elements

PGLS

Phylogenetic Generalized Least Squares

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for Publication

Not applicable.

Competing Interests

The authors declare no competing interests.

FUNDING

This research was supported by funding from The David and Lucile Packard Foundation, Packard Fellowship for Science and Engineering 2019–69653 to J.L.

Author Contribution

A.D.Y.T. and J.L. conceptualized the project. A.D.Y.T. led and implemented the methodology and formal analyses for this study with R.L. out performing repeatmasking for *F. sueziense* and DNA extractions of *F. whiteleyi* and *F. fragum* for WGS sequencing. A.D.Y.T. visualized and prepared all the main and supplementary figures and tables. J.L. acquired funding for this study. All authors wrote the original draft of the main manuscript text. A.D.Y.T. and J.L. reviewed and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgement

Many thanks to the Aquatic Symbiosis Genomics Project for enabling large-scale long-read genomic sequencing of symbiotic life forms including our photosymbiotic cockles and clams. We would like to express our gratitude to Dr. Nolan C. Kane, Dr. Kyle Keepers and Emily Curcio for help and support in genomic sequencing and analyses, and Dr. Edward Chuong for insightful conversations pertaining to TE. We are extending our appreciation to Kelly R. Martin, along with the University of Colorado Museum of Natural History, for the support in our use of specimens and databased records, and to Yu Kai Tan and Andressa Viol for aiding in the polygon visualizations of mitogenome segments in R.

Data Availability

All sequence data used in this study can be assessed on NCBI GenBank and all gene annotation datasets will be available as online supplementary material. Table S1 contains GenBank accession numbers for all mitogenome sequences. R codes for linearized visualization of mitogenomes are available on GitHub repository: https://github.com/Andy-TanDY/mitog_st8vis.

References

1. Roger AJ, Muñoz-Gómez SA, Kamikawa R. The Origin and Diversification of Mitochondria. *Curr Biol*. 2017;27(21):R1177–92. 10.1016/j.cub.2017.09.015.
2. Ghiselli F, Gomes-dos-Santos A, Adema CM, Lopes-Lima M, Sharbrough J, Boore JL. Molluscan mitochondrial genomes break the rules. *Philosophical Trans Royal Soc B: Biol Sci*. 2021;376(1825):20200159. 10.1098/rstb.2020.0159.
3. Boore JL. Animal mitochondrial genomes. *Nucleic Acids Res*. 1999;27(8):1767–80. 10.1093/nar/27.8.1767.

4. Wolstenholme DR, Animal Mitochondrial DNA. Structure and Evolution. In: Wolstenholme DR, Jeon KW, editors. International Review of Cytology [Internet]. Academic Press; 1992 [cited 2024 Sep 17]. pp. 173–216. Available from: <https://www.sciencedirect.com/science/article/pii/S0074769608620665> 10.1016/S0074-7696(08)62066-5
5. Formaggioni A, Luchetti A, Plazzi F. Mitochondrial Genomic Landscape: A Portrait of the Mitochondrial Genome 40 Years after the First Complete Sequence. *Life*. 2021;11(7):7. 10.3390/life11070663.
6. Lavrov DV, Pett W. Animal Mitochondrial DNA as We Do Not Know It: mt-Genome Organization and Evolution in Nonbilaterian Lineages. *Genome Biol Evol*. 2016;8(9):2896–913. 10.1093/gbe/evw195.
7. Kmiec B, Woloszynska M, Janska H. Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Curr Genet*. 2006;50(3):149–59. 10.1007/s00294-006-0082-1.
8. Wang H, Zhang S, Li Y, Liu B. Complete mtDNA of *Meretrix lusoria* (Bivalvia: Veneridae) reveals the presence of an atp8 gene, length variation and heteroplasmy in the control region. *Comp Biochem Physiol D: Genomics Proteomics*. 2010;5(4):256–64. 10.1016/j.cbd.2010.07.003.
9. Breton S, Ghiselli F, Milani L. Mitochondrial Short-Term Plastic Responses and Long-Term Evolutionary Dynamics in Animal Species. *Genome Biol Evol*. 2021;13(7):evab084. 10.1093/gbe/evab. 084 PubMed PMID: 33892508; PubMed Central PMCID: PMC8290114.
10. Zouros E. Biparental Inheritance Through Uniparental Transmission: The Doubly Uniparental Inheritance (DUI) of Mitochondrial DNA. *Evol Biol*. 2013;40(1):1–31. 10.1007/s11692-012-9195-2.
11. Gissi C, Iannelli F, Pesole G. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*. 2008;101(4):301–20. 10.1038/hdy.2008.62.
12. Larizza A, Pesole G, Reyes A, Sbisà E, Saccone C. Lineage Specificity of the Evolutionary Dynamics of the mtDNA D-Loop Region in Rodents. *J Mol Evol*. 2002;54(2):145–55. 10.1007/s00239-001-0063-4.
13. Saccone C, Gissi C, Reyes A, Larizza A, Sbisà E, Pesole G. Mitochondrial DNA in metazoa: degree of freedom in a frozen event. *Gene*. 2002;Papers presented at the International Meeting Mitochondria:Evolution, Genomics, Homeostasis and Pathology, Selva di Fasano (BR, Italy), 9–12 May 2001286(1):3–12. 10.1016/S0378-1119(01)00807-1
14. Wu X, Xu X, Yu Z, Kong X. Comparative mitogenomic analyses of three scallops (Bivalvia: Pectinidae) reveal high level variation of genomic organization and a diversity of transfer RNA gene sets. *BMC Res Notes*. 2009;2(1):69. 10.1186/1756-0500-2-69.
15. Xiaokaiti X, Hashiguchi Y, Ota H, Kumazawa Y. Evolution of the Noncoding Features of Sea Snake Mitochondrial Genomes within Elapidae. *Genes*. 2022;13(8):8. 10.3390/genes13081470.
16. Ekblom R, Smeds L, Ellegren H. Patterns of sequencing coverage bias revealed by ultra-deep sequencing of vertebrate mitochondria. *BMC Genomics*. 2014;15(1):467. 10.1186/1471-2164-15-467.
17. Imanishi Y, Tanaka M, Fujiwara M. Complete mitochondrial genome sequence of Japanese cockle *Fulvia mutica* (Cardiidae). *Fish Sci*. 2013;79(6):949–57. 10.1007/s12562-013-0662-1.
18. Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX et al. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences*. 1989;86(16):6196–200. 10.1073/pnas.86.16.6196
19. Formenti G, Rhie A, Balacco J, Haase B, Mountcastle J, Fedrigo O, et al. Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. *Genome Biol*. 2021;22(1):120. 10.1186/s13059-021-02336-9.
20. Bouchet P, Bary S, Virginie H, Marani G. How many species of molluscs are there in the world's oceans, and who is going to describe them? *Mémoires du Muséum. Natl d'Histoire Naturelle*. 2016;208:9–24.
21. Bronstein O, Kroh A, Haring E. Mind the gap! The mitochondrial control region and its power as a phylogenetic marker in echinoids. *BMC Evol Biol*. 2018;18:80. 10.1186/s12862-018-1198-x . PubMed PMID: 29848319; PubMed Central PMCID: PMC5977486.
22. Jiang ZJ, Castoe TA, Austin CC, Burbrink FT, Herron MD, McGuire JA, et al. Comparative mitochondrial genomics of snakes: extraordinary substitution rate dynamics and functionality of the duplicate control region. *BMC Evol Biol*.

- 2007;7(1):123. 10.1186/1471-2148-7-123.
23. Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C. Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene*. 1997;Junk DNA: The Role and the Evolution of Non-Coding Sequences 205(1):125–40. 10.1016/S0378-1119(97)00404-6
 24. Jemt E, Persson Ö, Shi Y, Mehmedovic M, Uhler JP, Dávila López M, et al. Regulation of DNA replication at the end of the mitochondrial D-loop involves the helicase TWINKLE and a conserved sequence element. *Nucleic Acids Res*. 2015;43(19):9262–75. 10.1093/nar/gkv804.
 25. Sebastian W, Sukumaran S, Gopalakrishnan A. The signals of selective constraints on the mitochondrial non-coding control region: insights from comparative mitogenomics of Clupeoid fishes. *Genetica*. 2021;149(3):191–201. 10.1007/s10709-021-00121-x.
 26. Pereira F, Soares P, Carneiro J, Pereira L, Richards MB, Samuels DC, et al. Evidence for Variable Selective Pressures at a Large Secondary Structure of the Human Mitochondrial DNA Control Region. *Mol Biol Evol*. 2008;25(12):2759–70. 10.1093/molbev/msn225.
 27. Liu YG, Kurokawa T, Sekino M, Tanabe T, Watanabe K. Complete mitochondrial DNA sequence of the ark shell *Scapharca broughtonii*: An ultra-large metazoan mitochondrial genome. *Comp Biochem Physiol D: Genomics Proteomics*. 2013;8(1):72–81. 10.1016/j.cbd.2012.12.003.
 28. Smith DR, Snyder M. Complete Mitochondrial DNA Sequence of the Scallop *Placopecten magellanicus*: Evidence of Transposition Leading to an Uncharacteristically Large Mitochondrial Genome. *J Mol Evol*. 2007;65(4):380–91. 10.1007/s00239-007-9016-x.
 29. Passamonti M, Plazzi F. Doubly Uniparental Inheritance and beyond: The contribution of the Manila clam *Ruditapes philippinarum*. *J Zoological Syst Evolutionary Res*. 2020;58(2):529–40. 10.1111/jzs.12371.
 30. Mamirova L, Popadin K, Gelfand MS. Purifying selection in mitochondria, free-living and obligate intracellular proteobacteria. *BMC Evol Biol*. 2007;7(1):17. 10.1186/1471-2148-7-17.
 31. Lynch M. Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes. *Mol Biol Evol*. 1997;14(9):914–25. 10.1093/oxfordjournals.molbev.a025834.
 32. Saccone C, Gissi C, Lanave C, Larizza A, Pesole G, Reyes A. Evolution of the mitochondrial genetic system: an overview. *Gene*. 2000;Papers presented at the Anton Dohrn Workshop 261(1):153–9. 10.1016/S0378-1119(00)00484-4
 33. Moran NA. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences*. 1996;93(7):2873–8. 10.1073/pnas.93.7.2873
 34. Diaz F, Bayona-Bafaluy MP, Rana M, Mora M, Hao H, Moraes CT. Human mitochondrial DNA with large deletions repopulates organelles faster than full-length genomes under relaxed copy number control. *Nucleic Acids Res*. 2002;30(21):4626–33. 10.1093/nar/gkf602.
 35. Smith SD, Pennell MW, Dunn CW, Edwards SV. Phylogenetics is the New Genetics (for Most of Biodiversity). *Trends Ecol Evol*. 2020;35(5):415–25. 10.1016/j.tree.2020.01.005.
 36. McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol*. 2012;10(1):13–26. 10.1038/nrmicro2670.
 37. George EE, Husnik F, Tashyreva D, Prokopchuk G, Horák A, Kwong WK, et al. Highly Reduced Genomes of Protist Endosymbionts Show Evolutionary Convergence. *Curr Biol*. 2020;30(5):925–e9333. 10.1016/j.cub. 2019.12.070 PubMed PMID: 31978335.
 38. Schneider A, Ebert D. Covariation of Mitochondrial Genome Size with Gene Lengths: Evidence for Gene Length Reduction During Mitochondrial Evolution. *J Mol Evol*. 2004;59(1):90–6. 10.1007/s00239-004-2607-x.
 39. Hourdez S, Lallier FH. Adaptations to hypoxia in hydrothermal-vent and cold-seep invertebrates. In: Amils R, Ellis-Evans C, Hinghofer-Szalkay H, editors. *Life in Extreme Environments* [Internet]. Dordrecht: Springer Netherlands; 2007 [cited 2024 Sep 23]. pp. 297–313. Available from: https://doi.org/10.1007/978-1-4020-6285-8_19 doi:10.1007/978-1-4020-6285-8_19.

40. Welker AF, Moreira DC, Campos ÉG, Hermes-Lima M. Role of redox metabolism for adaptation of aquatic animals to drastic changes in oxygen availability. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*. 2013;First International Conference on Oxidative Stress in Aquatic Ecosystems165(4):384–404. 10.1016/j.cbpa.2013.04.003
41. Xuefei Y, Xinyi Z, Qing C, Dan Z, Ziyun L, Hejuan Z, et al. Effects of Hyperoxia on Mitochondrial Homeostasis: Are Mitochondria the Hub for Bronchopulmonary Dysplasia? *Front Cell Dev Biol*. 2021;9. 10.3389/fcell.2021.642717.
42. Palci A, Seymour RS, Van Nguyen C, Hutchinson MN, Lee MSY, Sanders KL. Novel vascular plexus in the head of a sea snake (Elapidae, Hydrophiinae) revealed by high-resolution computed tomography and histology. *Royal Soc Open Sci*. 2019;6(9):191099. 10.1098/rsos.191099.
43. Yang M, Gong L, Sui J, Li X. The complete mitochondrial genome of *Calypptogena marissinica* (Heterodonta: Veneroida: Vesicomyidae): Insight into the deep-sea adaptive evolution of vesicomyids. *PLoS ONE*. 2019;14(9):e0217952. 10.1371/journal.pone.0217952.
44. Gardella DJ, Edmunds PJ. The oxygen microenvironment adjacent to the tissue of the scleractinian *Dichocoenia stokesii* and its effects on symbiont metabolism. *Mar Biol*. 1999;135(2):289–95. 10.1007/s002270050626.
45. Linsmayer LB, Deheyn DD, Tomanek L, Tresguerres M. Dynamic regulation of coral energy metabolism throughout the diel cycle. *Sci Rep*. 2020;10(1):19881. 10.1038/s41598-020-76828-2.
46. Herrera ND, ter Poorten JJ, Bieler R, Mikkelsen PM, Strong EE, Jablonski D, et al. Molecular phylogenetics and historical biogeography amid shifting continents in the cockles and giant clams (Bivalvia: Cardiidae). *Mol Phylogenet Evol*. 2015;93:94–106. 10.1016/j.ympev.2015.07.013.
47. Li J, Lemer S, Kirkendale L, Bieler R, Cavanaugh C, Giribet G. Shedding light: a phylotranscriptomic perspective illuminates the origin of photosymbiosis in marine bivalves. *BMC Evol Biol*. 2020;20(1):50. 10.1186/s12862-020-01614-7.
48. Li R, Zarate D, Avila-Magaña V, Li J. Comparative transcriptomics revealed parallel evolution and innovation of photosymbiosis molecular mechanisms in a marine bivalve. *Proceedings of the Royal Society B: Biological Sciences*. 2024;291(2023):20232408. 10.1098/rspb.2023.2408
49. Li R, Leiva C, Lemer S, Kirkendale L, Li J. Photosymbiosis shaped animal genome architecture and gene evolution as revealed in giant clams. *Commun Biol*. 2025;8(1):7. 10.1038/s42003-024-07423-8.
50. McKenna V, Archibald JM, Beinart R, Dawson MN, Hentschel U, Keeling PJ et al. The aquatic symbiosis genomics project: probing the evolution of symbiosis across the tree of life. *Wellcome Open Research* [Internet]. 2021 [cited 2024 Sep 18];6(254):254. Available from: <https://wellcomeopenresearch.org/articles/6-254/v1?src=rss>
51. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;10(2):giab008. 10.1093/gigascience/giab008.
52. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987–93. 10.1093/bioinformatics/btr509.
53. Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, et al. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Res*. 2019;47(20):10543–52. 10.1093/nar/gkz833 . PubMed PMID: 31584075; PubMed Central PMCID: PMC6847864.
54. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45(W1):W6–11. 10.1093/nar/gkx391.
55. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol Biol*. 2019;1962:1–14. 10.1007/978-1-4939-9173-0_1 PubMed PMID: 31020551; PubMed Central PMCID: PMC6768409.
56. Laslett D, Canbäck B. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics*. 2008;24(2):172–5. 10.1093/bioinformatics/btm573.
57. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 2003;31(13):3406–15. 10.1093/nar/gkg595.

58. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 2019;47(W1):W59–64. 10.1093/nar/gkz238.
59. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999;27(2):573–80. 10.1093/nar/27.2.573.
60. Smit A, Hubley R, Green P. RepeatMasker Open-4.0 [Internet]. 2015. Available from: <http://www.repeatmasker.org>
61. Crouch NMA, Edie SM, Collins KS, Bieler R, Jablonski D. Calibrating phylogenies assuming bifurcation or budding alters inferred macroevolutionary dynamics in a densely sampled phylogeny of bivalve families. *Proc Biol Sci.* 2021;288(1964):20212178. 10.1098/rspb.2021.2178
62. Hoffmann RJ, Boore JL, Brown WM. A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. *Genetics.* 1992;131(2):397–412. 10.1093/genetics/131.2.397.
63. Malkócs T, Viricel A, Becquet V, Evin L, Dubillot E, Pante E. Complex mitogenomic rearrangements within the Pectinidae (Mollusca: Bivalvia). *BMC Ecol Evol.* 2022;22(1):29. 10.1186/s12862-022-01976-0.
64. Gjetvaj B, Cook D, Zouros E. Repeated Sequences and Large-Scale Size Variation of Mitochondrial DNA: A Common Feature among Scallops (Bivalvia: Pectinidae). *Mol Biol Evol.* 1992;9(1):106. 10.1093/oxfordjournals.molbev.a040711.
65. Xiong Y, Yu Q, Xiong Y, Zhao J, Lei X, Liu L, et al. The Complete Mitogenome of *Elymus sibiricus* and Insights Into Its Evolutionary Pattern Based on Simple Repeat Sequences of Seed Plant Mitogenomes. *Front Plant Sci.* 2022;12. 10.3389/fpls.2021.802321.
66. Knoop V, Unseld M, Marienfeld J, Brandt P, Sünkel S, Ullrich H, et al. *copia*-, *gypsy*- and *LINE*-Like Retrotransposon Fragments in the Mitochondrial Genome of *Arabidopsis thaliana*. *Genetics.* 1996;142(2):579–85. 10.1093/genetics/142.2.579.
67. Ma H, Lin L, Zhang Y, Chen S, Shi W, Yu Z. The complete mitochondrial genome sequence of the giant clam *Tridacna derasa* (Tridacnidae: Tridacna). *Mitochondrial DNA B Resour.* 2018;3(2):911–2. PubMed PMID: 33474363; PubMed Central PMCID: PMC7799873.
68. Ma H, Xiang Z, Zhang Y, Li J, Qin Y, Zhang Y et al. First report of the mitochondrial DNA sequences of the giant clam, *Tridacna gigas* (Tridacnidae Tridacna). *Mitochondrial DNA Part B* [Internet]. 2020 Jul 2 [cited 2024 Sep 18]. Located at: world. Available from: <https://www.tandfonline.com/doi/abs/10.1080/23802359.2020.1830729>
69. Baeza JA, Neo ML, Huang D. Genomic Survey and Resources for the Boring Giant Clam *Tridacna crocea*. *Genes.* 2022;13(5):5. 10.3390/genes13050903.
70. Romanova EV, Bukin YS, Mikhailov KV, Logacheva MD, Aleoshin VV, Sherbakov DY, et al. The Mitochondrial Genome of a Freshwater Pelagic Amphipod *Macrohectopus branickii* Is among the Longest in Metazoa. *Genes.* 2021;12(12). 10.3390/genes12122030.
71. Olli S, Lam NT, Hiljanen S, Kettunen T, Haikonen L, Hyvönen HM, et al. Large mitochondrial genomes in tenthredinid sawflies (Hymenoptera, Tenthredinidae). *Mitochondrial DNA Part A.* 2025;35(3–4):84–92. 10.1080/24701394.2024.2427206 . PubMed PMID: 39526637.
72. Wang X, Lavrov DV. Seventeen New Complete mtDNA Sequences Reveal Extensive Mitochondrial Genome Evolution within the Demospongiae. *PLoS ONE.* 2008;3(7):e2723. 10.1371/journal.pone.0002723 . PubMed PMID: 18628961; PubMed Central PMCID: PMC2444032.
73. Serb JM, Lydeard C. Complete mtDNA Sequence of the North American Freshwater Mussel, *Lampsilis ornata* (Unionidae): An Examination of the Evolution and Phylogenetic Utility of Mitochondrial Genome Organization in Bivalvia (Mollusca). *Mol Biol Evol.* 2003;20(11):1854–66. 10.1093/molbev/msg218.
74. Liao D, Zhou Y, Tong J, Cao S, Yu X, Fu B, et al. Characterization and phylogenetic analysis of the complete mitochondrial genome from Rock Scallop (*Crassadoma gigantea*) using next-generation sequencing. *Mitochondrial DNA Part B.* 2018;3(2):827–8. 10.1080/23802359.2018.1483752 . PubMed PMID: 33474337.
75. McElroy KE, Masonbrink R, Chudalayandi S, Severin AJ, Serb JM. A chromosome-level genome assembly of the disco clam, *Ctenoides ales*. *G3 Genes[Genomes]Genetics.* 2024;14(9):jkae115. 10.1093/g3journal/jkae115

76. Plazzi F, Puccio G, Passamonti M. Comparative Large-Scale Mitogenomics Evidences Clade-Specific Evolutionary Trends in Mitochondrial DNAs of Bivalvia. *Genome Biol Evol.* 2016;8(8):2544–64. 10.1093/gbe/evw187.
77. Zhao B, Gao S, Zhao M, Lv H, Song J, Wang H, et al. Mitochondrial genomic analyses provide new insights into the missing atp8 and adaptive evolution of Mytilidae. *BMC Genomics.* 2022;23(1):738. 10.1186/s12864-022-08940-8.
78. Gan HM, Grandjean F, Jenkins TL, Austin CM. Absence of evidence is not evidence of absence: Nanopore sequencing and complete assembly of the European lobster (*Homarus gammarus*) mitogenome uncovers the missing nad2 and a new major gene cluster duplication. *BMC Genomics.* 2019;20(1):335. 10.1186/s12864-019-5704-3.
79. Breton S, Beaupré HD, Stewart DT, Hoeh WR, Blier PU. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? *Trends Genet.* 2007;23(9):465–74. 10.1016/j.tig.2007.05.011 . PubMed PMID: 17681397.
80. Skibinski DOF, Gallagher C, Beynon CM, Mitochondrial. *DNA Inherit Nat.* 1994;368(6474):817–8. 10.1038/368817b0.
81. Cao L, Ort BS, Mizi A, Pogson G, Kenchington E, Zouros E, et al. The Control Region of Maternally and Paternally Inherited Mitochondrial Genomes of Three Species of the Sea Mussel Genus *Mytilus*. *Genetics.* 2009;181(3):1045–56. 10.1534/genetics.108.093229.
82. Guerra D, Ghiselli F, Passamonti M. The largest unassigned regions of the male- and female-transmitted mitochondrial DNAs in *Musculista senhousia*. *Gene.* 2014;536(2):316–25. 10.1016/j.gene.2013.12.005.
83. Stewart DT, Kenchington ER, Singh RK, Zouros E. Degree of Selective Constraint as an Explanation of the Different Rates of Evolution of Gender-Specific Mitochondrial DNA Lineages in the Mussel *Mytilus*. *Genetics.* 1996;143(3):1349–57. 10.1093/genetics/143.3.1349.
84. Lucentini L, Plazzi F, Sfriso AA, Pizzirani C, Sfriso A, Chiesa S. Additional taxonomic coverage of the doubly uniparental inheritance in bivalves: Evidence of sex-linked heteroplasmy in the razor clam *Solen marginatus* Pulteney, 1799, but not in the lagoon cockle *Cerastoderma glaucum* (Bruguière, 1789). *J Zoological Syst Evolutionary Res.* 2020;58(2):561–70. 10.1111/jzs.12386.
85. Wuitchik DM, Aichelman HE, Atherton KF, Brown CM, Chen X, DiRoberts L, et al. Photosymbiosis reduces the environmental stress response under a heat challenge in a facultatively symbiotic coral. *Sci Rep.* 2024;14(1):15484. 10.1038/s41598-024-66057-2.
86. He H, Xiong L, Jian L, Li L, Wu Y, Qiao S. Role of mitochondria on UV-induced skin damage and molecular mechanisms of active chemical compounds targeting mitochondria. *J Photochem Photobiol B.* 2022;232:112464. 10.1016/j.jphotobiol.2022.112464.
87. Starkov AA. The Role of Mitochondria in Reactive Oxygen Species Metabolism and Signaling. *Ann N Y Acad Sci.* 2008;1147(1):37–52. 10.1196/annals.1427.015.
88. McCoy DE, Burns DH, Klopfer E, Herndon LK, Ogunlade B, Dionne JA et al. Heart cockle shells transmit sunlight to photosymbiotic algae using bundled fiber optic cables and condensing lenses [Internet]. *bioRxiv*; 2024 [cited 2024 Sep 22]. p. 2022.10.28.514291. Available from: <https://www.biorxiv.org/content/10.1101/2022.10.28.514291v2> doi:10.1101/2022.10.28.514291.
89. Rossbach S, Subedi RC, Ng TK, Ooi BS, Duarte CM. Iridocytes Mediate Photonic Cooperation Between Giant Clams (Tridacninae) and Their Photosynthetic Symbionts. *Front Mar Sci.* 2020;7. 10.3389/fmars.2020.00465.
90. Shokolenko I, Venediktova N, Bochkareva A, Wilson GL, Alexeyev MF. Oxidative stress induces degradation of mitochondrial DNA. *Nucleic Acids Res.* 2009;37(8):2539–48. 10.1093/nar/gkp100.
91. Li R. Elucidating Evolution of Bivalve Photosymbiosis: Insights From Phylogenomics, Transcriptomics, and Genomics [Ph.D.] [Internet]. [United States – Colorado]: University of Colorado at Boulder; 2024 [cited 2024 Oct 18]. Available from: <https://www.proquest.com/docview/3058337753?parentSessionId=Fs2lcJEBb3wGr096Xom1RtcUn%2B9CyaBXzcrtj483lfo%3D&sourcetype=Dissertations%20%20Theses>
92. Baumgarten S, Simakov O, Esherrick LY, Liew YJ, Lehnert EM, Michell CT et al. The genome of Aiptasia, a sea anemone model for coral symbiosis. *Proceedings of the National Academy of Sciences.* 2015;112(38):11893–8. 10.1073/pnas.1513318112

93. Fourreau CJL, Kise H, Santander MD, Pirro S, Maronna MM, Polisenio A, et al. Genome sizes and repeatome evolution in zoantharians (Cnidaria: Hexacorallia: Zoantharia). *PeerJ*. 2023;11:e16188. 10.7717/peerj.16188.
94. Pinzón JH, Kamel B, Burge CA, Harvell CD, Medina M, Weil E, et al. Whole transcriptome analysis reveals changes in expression of immune-related genes during and after bleaching in a reef-building coral. *Royal Soc Open Sci*. 2015;2(4):140214. 10.1098/rsos.140214.
95. Hollister JD, Gaut BS. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res*. 2009;19(8):1419–28. 10.1101/gr.091678.109 . PubMed PMID: 19478138.
96. Qin W, Scicluna BP, van der Poll T. The Role of Host Cell DNA Methylation in the Immune Response to Bacterial Infection. *Front Immunol*. 2021;12. 10.3389/fimmu.2021.696280.
97. Bove CB, Ingersoll MV, Davies SW. Help Me, Symbionts, You're My Only Hope: Approaches to Accelerate our Understanding of Coral Holobiont Interactions. *Integr Comp Biol*. 2022;62(6):1756–69. 10.1093/icb/icac141.
98. Gilbert C, Schaack S, Pace IJJK, Brindley PJ, Feschotte C. A role for host–parasite interactions in the horizontal transfer of transposons across phyla. *Nature*. 2010;464(7293):1347–50. 10.1038/nature08939.
99. Hotopp JCD. Horizontal gene transfer between bacteria and animals. *Trends Genet*. 2011;27(4):157–63. 10.1016/j.tig.2011. 01.005 PubMed PMID: 21334091.
100. Salinas-Giegé T, Giegé R, Giegé P. tRNA Biology in Mitochondria. *Int J Mol Sci*. 2015;16(3):4518–59. 10.3390/ijms16034518.
101. McDonald MJ, Chou CH, Swamy KB, Huang HD, Leu JY. The evolutionary dynamics of tRNA-gene copy number and codon-use in *E. coli*. *BMC Evol Biol*. 2015;15(1):163. 10.1186/s12862-015-0441-y.
102. Raval PK, Ngan WY, Gallie J, Agashe D. The layered costs and benefits of translational redundancy. Pilpel Y, Barkai N, Schirman D, editors. *eLife*. 2023;12:e81005. 10.7554/eLife.81005
103. Muneretto G, Plazzi F, Passamonti M. Mitochondrion-to-nucleus communication mediated by RNA export: a survey of potential mechanisms and players across eukaryotes. *Biol Lett*. 2024;20(7):20240147. 10.1098/rsbl.2024.0147.
104. Shaukat AN, Kaliatsi EG, Stamatopoulou V, Stathopoulos C. Mitochondrial tRNA-Derived Fragments and Their Contribution to Gene Expression Regulation. *Front Physiol*. 2021;12. 10.3389/fphys.2021.729452.
105. Fueyo R, Judd J, Feschotte C, Wysocka J. Roles of transposable elements in the regulation of mammalian transcription. *Nat Rev Mol Cell Biol*. 2022;23(7):481–97. 10.1038/s41580-022-00457-y.
106. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. *Genome Biol*. 2018;19(1):199. 10.1186/s13059-018-1577-z.
107. Nardi F, Carapelli A, Frati F. Repeated regions in mitochondrial genomes: Distribution, origin and evolutionary significance. *Mitochondrion*. 2012;12(5):483–91. 10.1016/j.mito.2012.07.105.

Figures

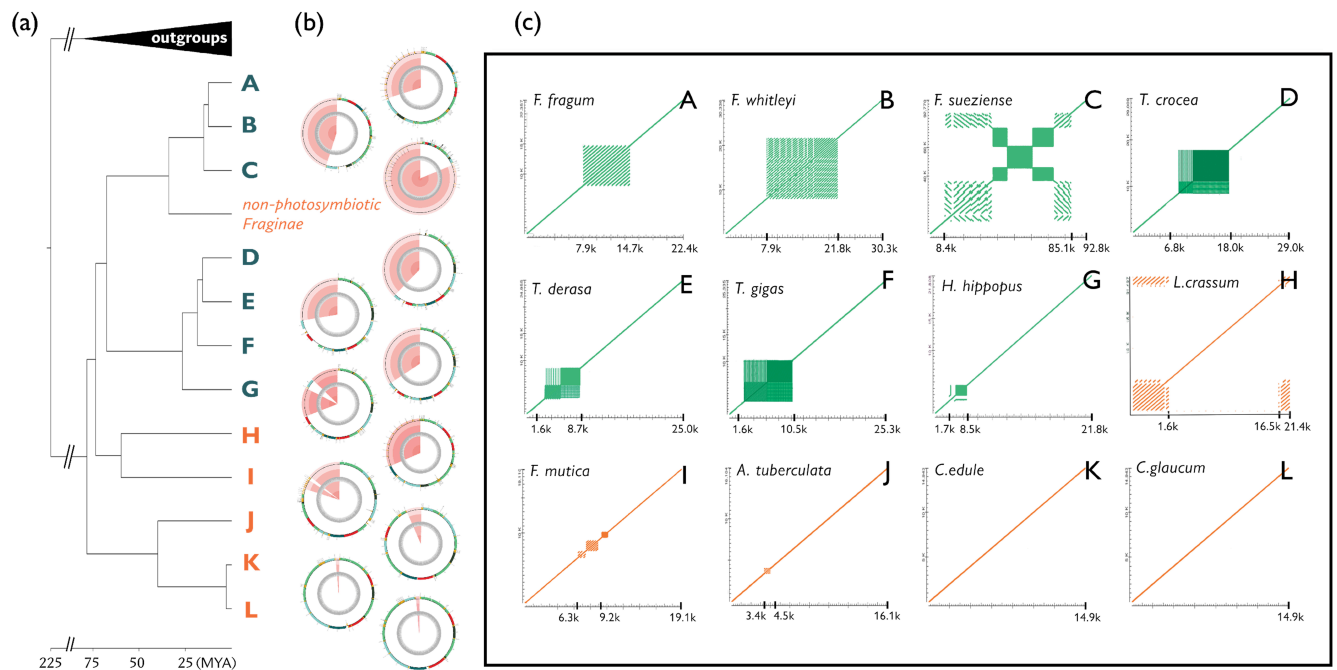


Figure 1

Phylogenetic context and mitochondrial architecture of focal cardiid species. (a) Time-calibrated phylogeny of twelve cardiid bivalves (time scale in MYA), categorized by ecological state: photosymbiotic (green) and non-photosymbiotic (orange). (b) Circularized mitogenome maps (A–L; expanded in **Figs. S1–S12a**), corresponding to phylogenetic identities in panel (a); red arcs denote the major NCR. (c) Self-comparative dot plots illustrating genomic complexity and repetitive landscapes. Dense green patterns in photosymbiotic lineages (A–G) reflect marked expansions of tandem and interspersed repeats, contrasting with the streamlined architectures of non-photosymbiotic species (H–L; outgroup dot plots in **Fig. S13**). Alphabetical labels are consistent across panels.

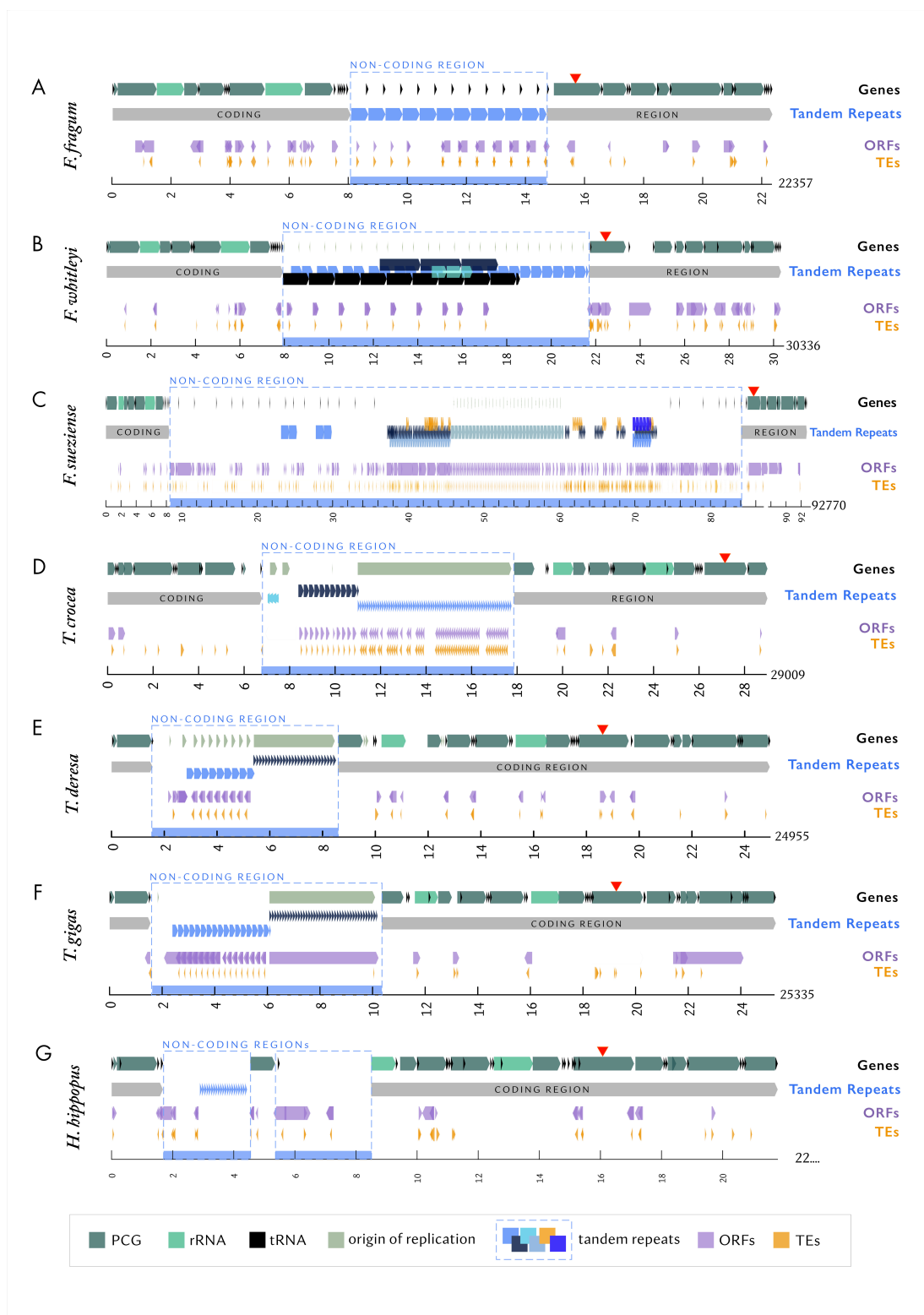


Figure 2

Comparative linearized mitochondrial architectures and annotation landscapes. Multi-track annotations for photosymbiotic (A–G) and non-photosymbiotic species (H–L; in **Fig. S14**). Track depicts: **(top row)** PCGs and putative origins of replication; **(second row)** structural partitioning between coding regions and the major NCR (dashed blue box) alongside tandem repeats patterns (blue blocks); **(bottom two rows)** relative density/overlap of predicted ORFs (purple) and TEs (orange). Red inverted triangles denote *cox1* positions across all assemblies. Horizontal axis indicates nucleotide position in base pairs (bp).

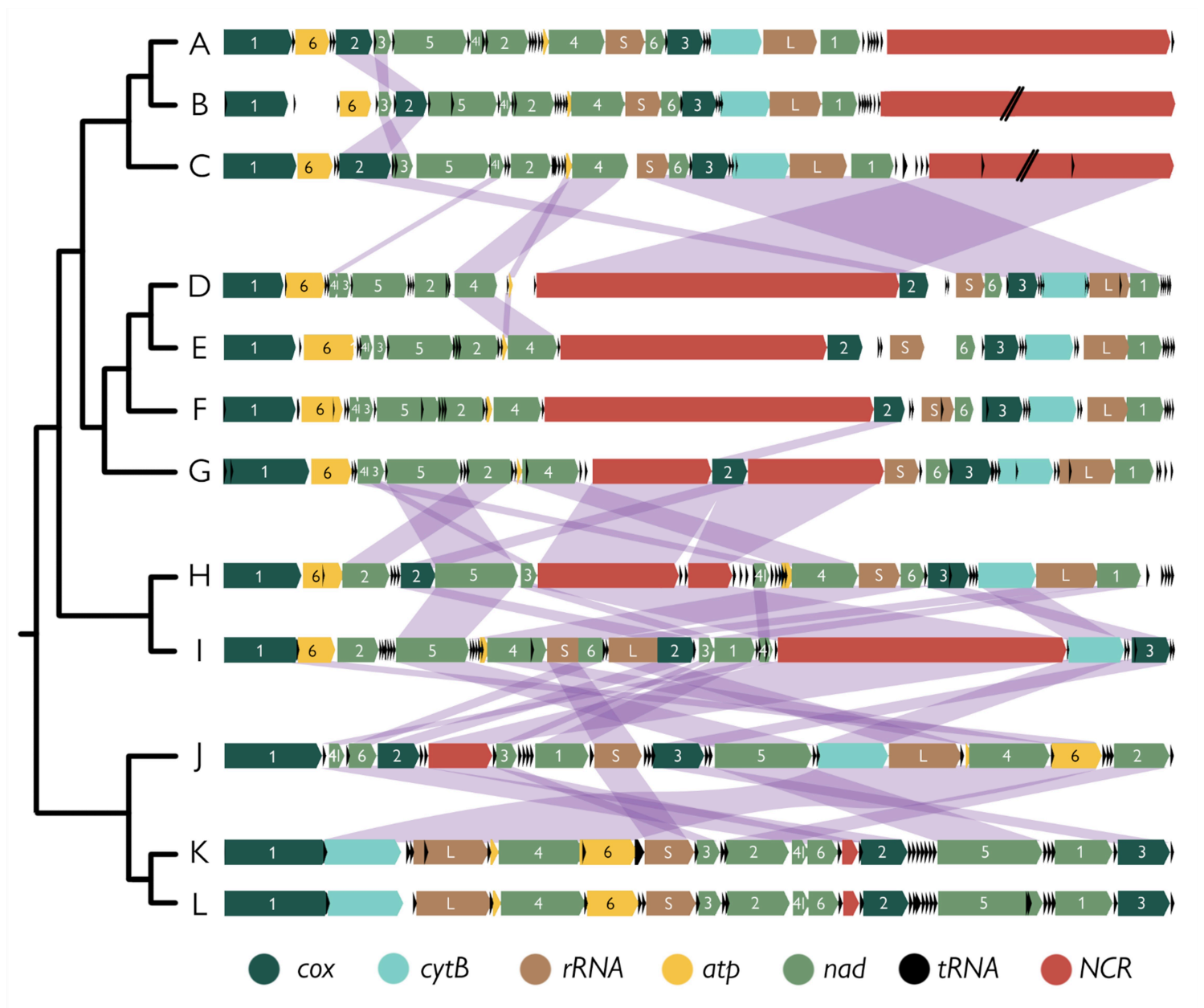


Figure 3

Mitogenomic synteny and gene arrangement across focal cardiid species. Linearized representations of mitogenome (A-L) presented in a phylogenetic framework, aligned at the *cox1* start position. Purple ribbons on the synteny plot connect the most parsimonious deduction of interspecies mapping for homologous blocks to illustrate syntenic conservation and rearrangement points. Numbers within gene blocks denote PCG subunits (e.g., '3' on *cox* = *cox3*, '3' on *nad* = *nad3*); red blocks designate the major NCR (truncated for taxa B and C for clarity). All elements are scaled according to their relative length and position. Taxa: A: *Fr. fragum*, B: *Fr. whitleyi*, C: *Fr. sueziense*, D: *T. crocea*, E: *T. derasa*, F: *T. gigas*, G: *H. hippopus*, H: *Fu. mutica*, I: *L. crassum*, J: *A. tuberculata*, K: *C. edule*, and L: *C. glaucum*.

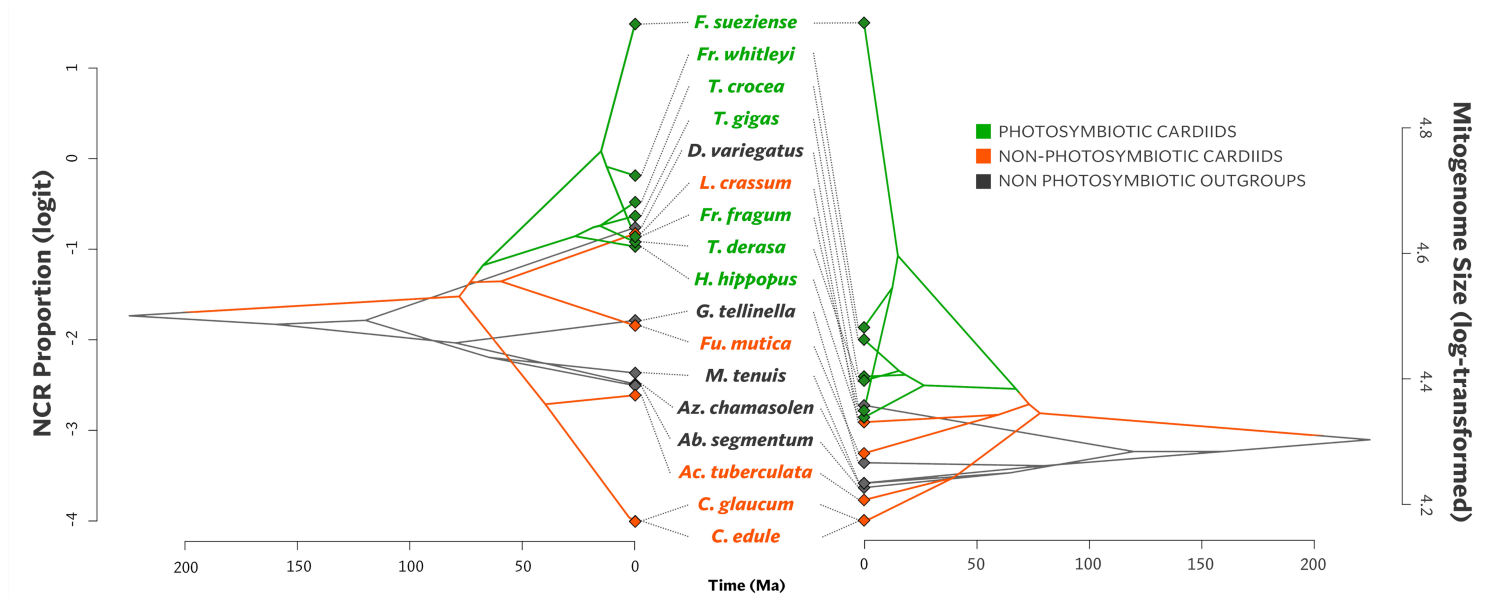


Figure 4

Phylomorphospace projections of mitogenomic trait evolution. Evolutionary trajectories of mitochondrial traits over time (Ma) projected across the cardiid phylogeny: logit-transformed NCR proportion (left) and log-transformed total mitogenome size (right). Lineages are color-coded by ecological state: photosymbiotic cardiids (green), non-photosymbiotic cardiids (orange), and non-photosymbiotic outgroups (grey). These phenograms demonstrate pronounced divergence in NCR proportion and genome size within photosymbiotic clades, with maximal expansions in *Fragum* and *Tridacna*.

Supplementary Files

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

- [20260928Table1.docx](#)
- [202601mtgsupp3completeTEannotation.xlsx](#)
- [202601mtgsupp2completeTandomRepeatannotation.xlsx](#)
- [202601mtgsupp1annotation.xlsx](#)
- [202601mtgsupp4completeORFannotation.xlsx](#)
- [202607mtgsupp515genesupermatrix.fasta](#)
- [20260928mtgsuppBMC.docx](#)