



OPEN

Unprecedented frequency of mitochondrial introns in colonial bilaterians

Helen Louise Jenkins^{1,2}✉, Rachael Graham¹, Joanne Sara Porter^{1,3}, Leandro Manzoni Vieira^{1,4}, Ana Carolina Sousa de Almeida^{4,5}, Andrea Hall⁶, Aaron O'Dea⁷, Simon Edward Coppard⁸ & Andrea Waeschenbach¹✉

Animal mitogenomes are typically devoid of introns. Here, we report the largest number of mitochondrial introns ever recorded from bilaterian animals. Mitochondrial introns were identified for the first time from the phylum Bryozoa. They were found in four species from three families (Order Cheilostomatida). A total of eight introns were found in the complete mitogenome of *Exechonella vieirai*, and five, 17 and 18 introns were found in the partial mitogenomes of *Parantropora penelope*, *Discoporella cookae* and *Cupuladria biporosa*, respectively. Intron-encoded protein domains reverse transcriptase and intron maturase (RVT-IM) were identified in all species. Introns in *E. vieirai* and *P. penelope* had conserved Group II intron ribozyme domains V and VI. Conserved domains were lacking from introns in *D. cookae* and *C. biporosa*, preventing their further categorization. Putative origins of metazoan introns were explored in a phylogenetic context, using an up-to-date alignment of mitochondrial RVT-IM domains. Results confirmed previous findings of multiple origins of annelid, placozoan and sponge RVT-IM domains and provided evidence for common intron donor sources across metazoan phyla. Our results corroborate growing evidence that some metazoans with regenerative abilities (i.e. placozoans, sponges, annelids and bryozoans) are susceptible to intron integration, most likely via horizontal gene transfer.

A crucial step in early eukaryote evolution was the origin of mitochondria, which arose by incorporating a Proteobacteria endosymbiont as cellular organelles¹. Since this origin, the various evolutionary trajectories of eukaryotes have produced mitochondrial (mt) genomes (mitogenomes) that vary substantially in their sizes, gene content, genetic code, organization and physical shape². Plant, fungal and other non-metazoan eukaryote mitogenomes range in size from 12 to 236 Kb, 66–11.3 Mb and 34–113 Kb, respectively and harbour a large proportion of non-coding DNA, including introns. In contrast, bilaterian mitogenomes are typically circular, streamlined molecules of approximately 16 kb (range: 11–50 Kb), which are typically devoid of introns³. The lack of introns in bilaterian mitogenomes has been attributed to the relatively high mutation rate of animal mtDNA, which can be ~9 to 25 times faster than that of corresponding nuclear DNA⁴. Introns, like all non-coding DNA, provide a substrate for potentially harmful mutations in the presence of high mutation rates, and are therefore hypothesised to have been purged from most animal mitogenomes^{4,5}.

In addition to spliceosomal introns, which occur in the nuclear genomes of eukaryotes and require ribonucleoprotein complexes called spliceosomes to facilitate their excision from precursor mRNA, there are two other types of introns. Group I introns occur in bacterial, archaeal, viral and organellar genomes as well as some eukaryote nuclear genomes^{6–9}. Group II introns occur in bacterial, archaeal and organellar genomes (particularly of plants, fungi, algae and protists), but are not found in nuclear genomes^{10,11}. Both types are self-splicing, mobile ribozymes, but they are distinguished by their splicing mechanisms. Group I introns use external guanosine

¹Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK. ²Marine Biological Association, The Laboratory, Citadel Hill, Plymouth PL1 2PB, Devon, UK. ³International Centre Island Technology, Heriot Watt University, Orkney Campus, Robert Rendall Building, Franklin Road, Stromness, Orkney KW16 3AW, Scotland, UK. ⁴Laboratório de Estudos de Bryozoa—LAEBry, Departamento de Zoologia, Centro de Biociências, Universidade Federal de Pernambuco, Recife, PE 50670-901, Brazil. ⁵Museu de História Natural, Setor da Zoologia, Universidade Federal da Bahia, Salvador, BA 40170-115, Brazil. ⁶Core Research Labs, Natural History Museum, Cromwell Road, London SW7 5BD, UK. ⁷Smithsonian Tropical Research Institute, Balboa, Ancon, Republic of Panama. ⁸Bader International Study Centre, Queen's University (Canada), Herstmonceux Castle, Hailsham BN27 1RN, East Sussex, UK. ✉email: helen.jenkins@nhm.ac.uk; a.waeschenbach@nhm.ac.uk

nucleotides as cofactors, whereas Group II introns use a mechanism more akin to that found in spliceosomal introns⁸ and are believed to be precursors of eukaryote spliceosomal introns, retroelements and spliceosome components^{11,12}. Both types frequently harbour intron-encoded proteins (IEPs) that facilitate intron mobility; Group I introns typically encode homing endonucleases, whereas Group II IEPs include reverse transcriptase (RVT) and intron maturase (IM) domains¹². In terms of their secondary structures, Group II introns form six domains (DI—DVI) that radiate out from a central loop. Based on the intricacies of their secondary structures, five Group II introns have been identified (IIA1, IIA2, IIB1, IIB2, IIC)^{13,14}. Conversely, Group I introns typically form secondary structures of P1–P10 stems¹⁵.

Although typically not found in metazoans, both intron types have been recorded in the mitogenomes of non-bilaterian phyla Porifera^{16–22} and Placozoa^{23–26}, whereas Cnidaria are only known to harbour Group I introns^{27–29}. Conversely, in bilaterians, mitogenome introns are apparently very rare, but Group II introns have been recognised in the *cox1* gene of annelids *Nephtys* (Phyllococida)³⁰, *Glycera fallax*, *G. unicornis* (Phyllococida)³¹, *Decemunciger* sp. (Terebellida)³² and the myxostomid *Endomyxostoma* sp.³³. Furthermore, high-throughput DNA sequencing has revealed additional introns of unknown type(s) in other protein-coding genes (PCGs) of *Decemunciger* sp. (*nad1* and *nad4*)³² and the mollusc *Cucullaea labiata* (Arcoida) (*cox1*)^{34,35}.

Here, we reveal the widespread occurrence of introns in the mitogenomes of taxa of the lophotrochozoan phylum Bryozoa. Bryozoans, whose fossil record goes back as far as the Cambrian³⁶, are modular invertebrates, found worldwide in aquatic habitats, including freshwater bodies, shallow coastal waters, and the deep sea^{37,38}. Almost exclusively colonial, they form encrustations on a variety of substrates, construct erect, three-dimensional structures or produce free living disks^{39,40}. There are three classes of bryozoans. The least diverse of these is the exclusively freshwater class Phylactolaemata with ~86 Recent species⁴¹. The Phylactolaemata forms the sister group to all remaining bryozoans⁴². The class Stenolaemata, whose only surviving order is the exclusively marine Cyclostomatida, constitutes ~543 Recent species⁴¹ and they form the sister group to the class Gymnolaemata⁴². Although predominantly marine, the Gymnolaemata also includes a few brackish and freshwater species. It is the most speciose class with their ~5240 Recent species, of which ~319 belong to the order Ctenostomatida and ~4921 belong to the order Cheilostomatida⁴¹; the Cheilostomatida nest within a paraphyletic Ctenostomatida⁴². Apart from phylactolaemates, which can disperse via asexually produced propagules called statoblasts, and some gymnolaemates which can form resting structures called hibernacula⁴³, bryozoan colonies are typically established by sexually produced larvae, which settle and metamorphose into the founding module (zooid) of the colony, the ancestrula³⁷. The ancestrula then buds the adjacent zooids, which in turn continue the colony growth via asexual budding of further zooids⁴⁴. Reparative growth of damaged colony parts frequently occurs⁴⁵, and some species propagate largely via colony fragmentation and subsequent regeneration⁴⁶.

At present, there are eight verified published bryozoan mitogenomes available from NCBI (<https://www.ncbi.nlm.nih.gov/>), none of which contain introns (see GenBank accessions NC_008192, NC010197, NC_011820, NC_015646, NC_016722, NC_018344, NC_018355, NC_038192). As part of a wider study sequencing hundreds of mitogenomes from species of the bryozoan order Cheilostomatida, genome-skimming has revealed the occurrence of introns in multiple PCGs in single mitogenomes, multiple introns within single genes, as well as in intergenic regions. Here, we document introns discovered in the mitogenomes of four cheilostome bryozoans: two encrusting species – *Parantropora penelope* (AW2102; Calloporoidea, Antroporidae) and *Exechonella vieirai* (AW1260; Arachnopusioidea, Exechonellidae) and two free-living species from the family Cupuladriidae (Calloporoidea), *Cupuladria biporosa* (AW817) and *Discoporella cookae* (AW3739) (Fig. 1). Given the high frequency of introns recorded, we verify their presence using polymerase chain reactions (PCRs), in order to rule out mis-assemblies of Illumina reads being the cause of the fragmented open reading frames (ORFs). Furthermore, we analyse the PCGs in a phylogenetic context together with published RVT-IM domains to identify possible intron donor organisms. This is the largest number of mt introns reported from a bilaterian phylum, to date.

Results and discussion

Mitogenome information. Numbers of raw Illumina reads per species, contig sizes and coverage, and GenBank accession numbers are given in Table 1. Only the mitogenome of *E. vieirai* could be circularised. In this mitogenome, genes *nad1* and *nad4* were terminated by abbreviated stop codon U (Supplementary Table S1); stop codon UAA is presumably completed by polyadenylation of the cleaved polycistronic transcript mRNA⁴⁷. The mitogenome of *C. biporosa*, although complete with regards to gene contents, could not be circularised. Both these mitogenomes had the full complement of 13 PCGs, two rRNA genes and 22 tRNA genes. The mitogenomes of *P. penelope* and *D. cookae* were incomplete. Genes missing in the *P. penelope* contig were tRNAs *trnW* and *trnE*. Furthermore, we infer that gene *nad4L* starts on the non-standard initiation codon GUU, which encodes for valine (Supplementary Table S1). Codon GUG, also encoding for valine, has been shown to be a functional initiation codon for *atp6* in some humans⁴⁸ and has therefore been accepted as alternative initiation codon. However, no evidence for the functionality of GUU as initiation codon has yet been shown. The mitogenome data for *D. cookae* was in two separate contigs: Contig A (size: 5,283 bp) and Contig B (size: 15,602 bp). Missing genes were: PCGs *atp6* and *nad4* and tRNAs *trnA*, *trnC*, *trnE*, *trnH*, *trnR*, *trnT*, *trnV* and *trnW*; genes *nad2* and *nad6* were incomplete at their 5' ends. For diagrammatic representation of gene order of mt fragments, see Fig. 2. For more detailed mitogenome information (gene boundaries/lengths, initiation/stop codons, GC content), see Supplementary Table S1.

Intron verification and characterization. A large number of introns were found within several PCGs, sometimes multiple introns per gene, across multiple species of cheilostome bryozoans (Fig. 2, Supplementary Table S2). We eliminate the possibilities of them being assembly artefacts by verifying 22 of them with PCRs and

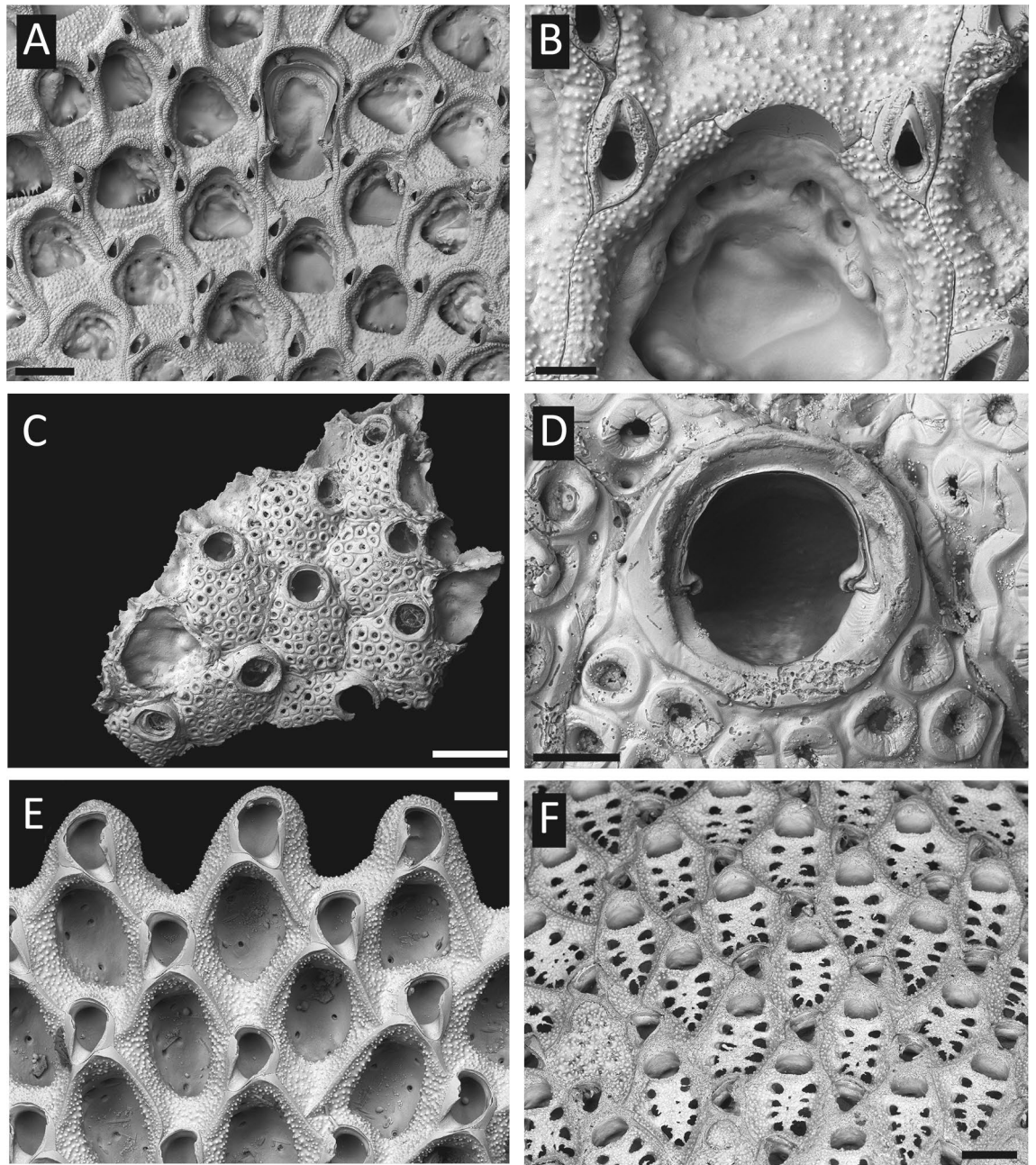
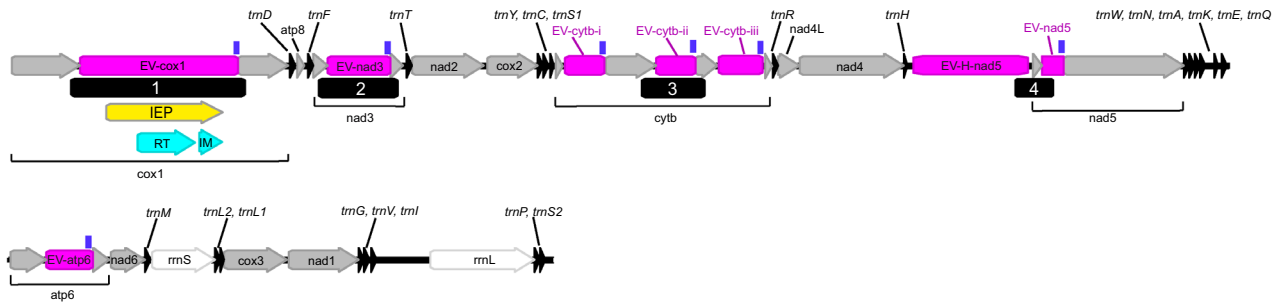


Figure 1. Scanning electron microscope images of morphological voucher specimens. *Parantropora penelope* (NHMUK 2021.2.9.1.) (A,B), *Exechonella vieirai* (NHMUK 2018.5.17.81) (C,D), *Cupuladria biporosa* (NHMUK 2018.5.17.101) (E), *Discoporella cookae* (NHMUK 2021.2.9.2.) (F). Scale bars = 200 μ m (A), 50 μ m (B), 500 μ m (C), 100 μ m (D,E), 200 μ m (F).

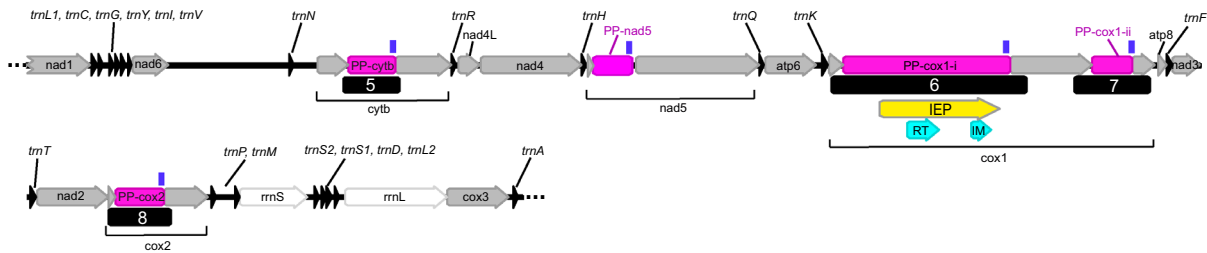
	No. paired-end reads (pre- and post-trimming)	Mitogenome contig size (bp)	Circularised (Y/N)	Mean coverage	GenBank accession
<i>Exechonella vieirai</i>	3,979,105 3,656,084	23,057	Y	51.6 \times	MW592986
<i>Parantropora penelope</i>	4,832,737 4,419,735	21,889	N	16.2 \times	MW592988
<i>Cupuladria biporosa</i>	8,038,861 7,784,830	23,200	N	132 \times	MW592987
<i>Discoporella cookae</i> Contig A	13,944,941	5,283	N	18.4 \times	MW592990
<i>Discoporella cookae</i> Contig B	13,520,261	15,602		18.2 \times	MW592989

Table 1. Number of raw Illumina reads, mitogenome contig sizes, coverage depth and GenBank accession numbers for *Exechonella vieirai*, *Parantropora penelope*, *Cupuladria biporosa* and *Discoporella cookae*.

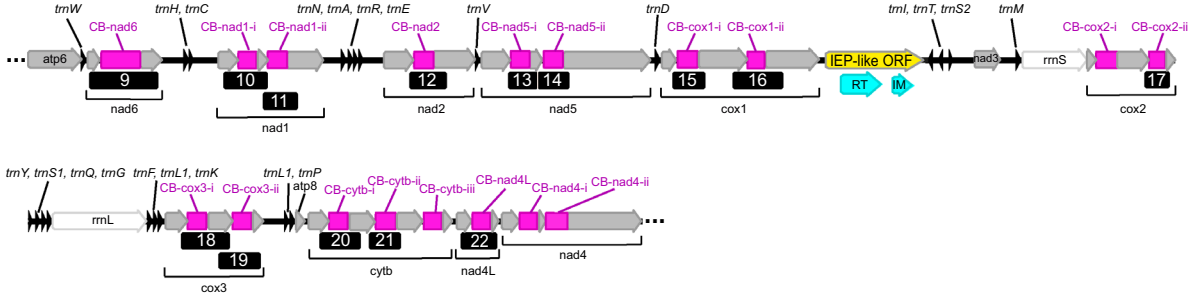
A) *Exechonella vieirai*; 23,057 bp



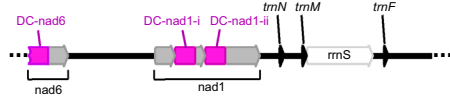
B) *Parantropora penelope*; 21,889 bp



C) *Cupuladria biporosa*; 23,200 bp



D) *Discoporella cookae* (Contig A); 5,283 bp



E) *Discoporella cookae* (Contig B); 15,602 bp

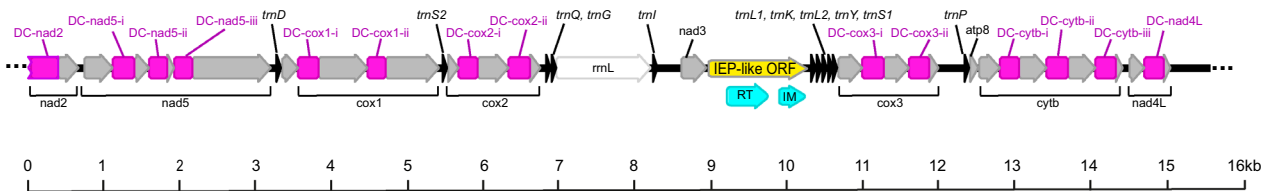


Figure 2. Mitogenome maps showing gene order and PCR product locations (black boxes 1–22). Linearised complete mitogenome of *Exechonella vieirai* (A). Partial mitogenome of *Parantropora penelope* (B). Non-circularised mitogenome of *Cupuladria biporosa* (C). Contig A of *Discoporella cookae* (D). Contig B of *Discoporella cookae* (E). Introns are labelled as pink boxes. Conserved Group II ribozyme domains V and VI, as found by Rfam, are indicated by vertical blue bars. Capital prefixes in intron labels represent genus/species initials. Whenever multiple introns were found per gene, introns were labelled with suffixes i-iii. Open reading frames (ORFs) of intron-encoded proteins (IEPs) are shown as yellow arrows. In the cases of *C. biporosa* and *D. cookae*, these ORFs are labelled as ‘IEP-like ORFs’ because they do not nest within introns. ORFs of reverse transcriptase (RT) and intron maturase (IM) are shown as turquoise arrows.

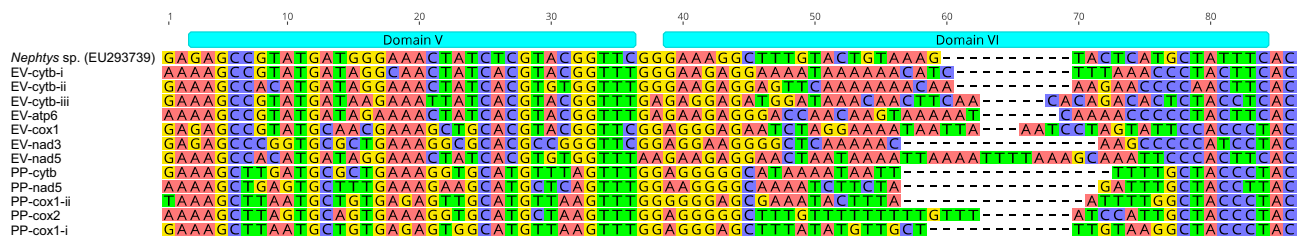


Figure 3. Group II ORF-less ribozyme domains V and VI alignments of *Exechonella vieirai* and *Parantropora penelope* with reference sequence *Nephtys* sp. (domain annotation from Fig. 1 in³⁰).

Sanger sequencing. Each of the introns examined with PCRs were confirmed to be bona fide. PCR products were of expected sizes (Supplementary Fig. S1) and flanking Sanger reads matched the existing assemblies.

The mitogenome of *E. vieirai* harboured single introns in three genes (EV-cox1, EV-nad3, EV-atp6; intron name prefixes indicate genus/species initials; Fig. 2, Supplementary Table S2). Furthermore, three introns were found in cytb (EV-cytb-i-iii) and there was one intron-like region situated between tRNA histidine and nad5 (EV-H-nad5) and one intron in nad5 (EV-nad5). The concatenation of EV-H-nad5 + EV-nad5 aligned well with other introns. Thus, it seems that this intron is interrupted by a short nad5 exon of 114 bp (Fig. 2; Supplementary Table S1). To eliminate the possibility of this being an assembly artefact, we confirmed the presence of this exon by PCR with primers annealing in flanking intron regions (Fig. 2). The PCR product was of the expected size of 607 bp (Supplementary Fig. S1) and Sanger reads matched the existing assembly. Thus, this is a bona fide result. To confirm that the short 114 bp exon is indeed part of nad5, we examined this region in the context of a wider alignment of published and unpublished bryozoan nad5 sequences. Although the 5' end of nad5 is rather variable, we conclude that the 114 bp exon does fit into the alignment lengthwise and sequence-wise (see Supplementary Fig. S2). This result, however, raises the question on how intron-splicing and translation of the exon can be performed effectively seeing that the intron may not be able to fold into a functional secondary structure. This issue ought to be revisited in future work. The mitogenome of *P. penelope* had single introns in cox2 (PP-cox2), cytb (PP-cytb) and nad5 (PP-nad5) and two introns in cox1 (PP-cox1-i-ii) (Fig. 2; Supplementary Table S2). Furthermore, both *E. vieirai* and *P. penelope* had IEPs with RVT and IM domains in cox1 introns, EV-cox1 and PP-cox1-i, respectively (Fig. 2, Supplementary Tables S1 and S2; see Supplementary Fig. S3 for Pfam results).

The largest number of introns, however, was found in the two members of the family Cupuladriidae: *C. biporosa* and *D. cookae*. Although there are substantial similarities in the intron distribution across these two species (two introns each in cox1, cox2, cox3, nad1; single introns in nad2, nad4L and nad6 [but note that the *D. cookae* fragments start on incomplete nad2 and nad6 sequences, so there may be more undiscovered introns]), *D. cookae* harboured three introns in nad5, whereas *C. biporosa* harboured only two introns in this gene. A further two introns were found in nad4 in *C. biporosa*, but due the partial nature of the *D. cookae* mitogenome, nad4 was missing from its contigs, thus, no inferences can be made about shared introns for this gene. Furthermore, both species have one intergenic region that harbours an open reading frame containing RVT and IM domains (for Pfam results, see Supplementary Fig. S3). In *C. biporosa* this is located between cox1 and tRNA isoleucine and in *D. cookae* it is found between nad3 and tRNA leucine 1 (Fig. 2; Supplementary Tables S1 and S2; see Supplementary Fig. S3 for Pfam results).

Group II introns typically exhibit a secondary structure composed of six domains¹³. An alignment to reference sequence *Nephtys* sp. in which all six domains were annotated according to Vallès et al.³⁰, revealed that all introns in *E. vieirai* and *P. penelope* had Group II intron ribozyme domains V and VI, as determined by Rfam searches (Supplementary Table S3), all of which aligned well with reference sequence *Nephtys* sp. (Fig. 3). None of the cupuladriid introns harboured conserved Group II ribozyme domains.

All but one intron of *E. vieirai* and *P. penelope*, started and ended on conserved splice motifs GUGYG and YAY^{12,13}. PP-cox1-ii intron deviated from this pattern and started on GUAUG (Supplementary Table S2). In contrast, none of the cupuladriid introns started on the conserved GUGYG motif. As a result, intron starts were determined by the end of the preceding exon, which resulted in highly variable putative start motifs (Supplementary Table S2). Similarly, the ends of introns were determined by the start of the following exons, which meant that 16 out of the 35 cupuladriid introns did not terminate on a conserved AY motif (Supplementary Table S2). The lack of these conserved motifs suggests that the cupuladriid introns might splice via alternative 5' and 3' splice sites (see¹⁴; Fig. 3I). Most Group II introns with alternative splice sites contain LAGLIDADG homing endonucleases¹⁴, but these could not be found in the cupuladriid mitogenomes using Pfam searches.

Concerning the identification of intron types in our study species, we consider the introns of *E. vieirai* and *P. penelope* to be possible Group II introns due to the presence of a) IEP with RVT-IM domains, b) conserved Group II ribozyme domains V and VI, c) Group II intron-typical start and stop motifs (except the non-standard start motif in PP-cox1-ii). All introns in these two species finish on YAY, which is a typical Group IIA end motif, versus RAY, which is typical for Group IIB introns¹³. However, the secondary structures of the two IEP-containing introns (EV-cox1, PP-cox1-i; Fig. 4, Supplementary Fig. S4) do not adhere to the conserved motifs shown in the Group IIA consensus secondary structure given on the Zimmerly lab website (<http://webapps2.ucalgary.ca/~groupii/>)^{10,49,50}. Although the secondary structure drawing of EV-cox1 (Fig. 4) shows multiple domains radiating from a central wheel, only domains V and VI could be identified reliably. Furthermore, only a few tertiary interaction features that facilitate intron-exon pairings, i.e. intron-exon binding sites IBS1-EBS1 and delta (δ) and delta prime (δ')⁵¹ could be determined reliably; the IBS2-EBS2 pairing, although indicated in Fig. 4 should

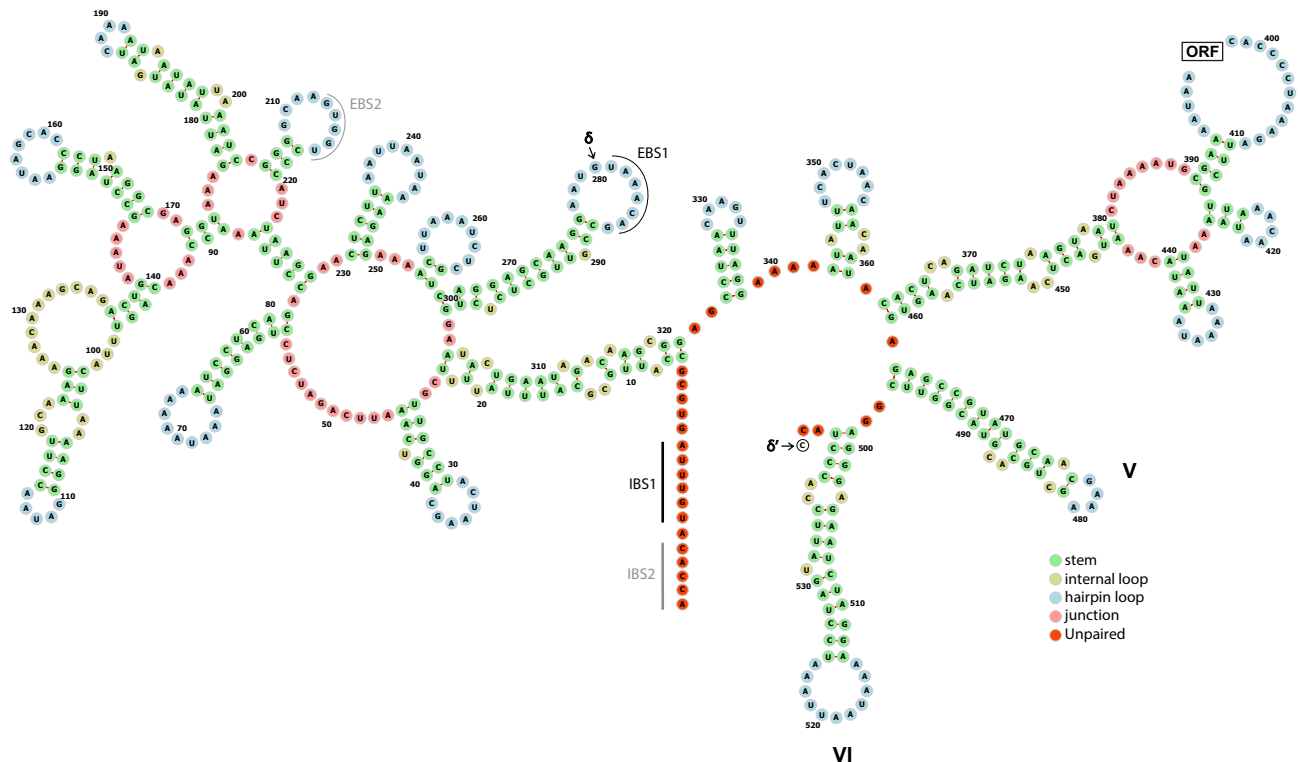


Figure 4. Putative secondary structure drawing of EV-cox1 intron (*Exechonella vieirai*). The intron-encoded protein open-reading frame (ORF) was excised prior to folding. Domains V and VI and intron-binding site IBS1 and corresponding exon-binding site EBS1 are indicated. Less certain IBS2 and EBS2 sites are indicated in grey. Tertiary interaction sites delta (δ) and delta prime (δ') are labelled. Nucleotides are coloured according to their positions in the secondary structure.

be treated with caution. Regarding the secondary structure of intron PP-cox1-i (Supplementary Fig. S4), only Group II domains V and VI could be reliably identified. Thus, the lack of unequivocal evidence prevents us from assigning a Group II subcategory to these introns.

Concerning *C. biporosa* and *D. cookae*, the identity of the introns cannot be ascertained because of the absence of identifiable ribozyme domains and Group II intron start/stop motifs. However, the presence of RVT-IM domains in both species indicates that they might be atypical Group II introns. Furthermore, none of the reconstructed secondary structures resemble the characteristic P1-P10 stems of Group I introns¹⁵ (Supplementary Figs S5, S6). Moreover, none of the Pfam searches of any of the six reading frames of the cupuladriid introns provided any hits with homing endonucleases, which are characteristic of Group I introns. Thus, we conclude that they are unlikely Group I introns. We also considered the possibility of them being Group III introns, which have been described from euglenoid chloroplast genomes⁵². Group III introns are somewhat degenerate versions of Group II introns. They are typically short (91–119 bp), AU-rich with a base bias of U > A > G > C, have degenerate Group II intron-like boundaries (5': NUNNG; 3': ANNUNNN), and lack any consistent secondary structure^{52–54}, although they have been shown to have a structure resembling Group II intron domain VI⁵⁵. Regarding the cupuladriid introns, their lengths, although shorter than introns in the non-cupuladriids, exceed the typical Group III intron size. Intron sizes range from 221–519 bp (average 272 bp) and 240–397 bp (average 273 bp) in *C. biporosa* and *D. cookae*, respectively (Supplementary Tables S1, S2). Although AU rich, the most frequent nucleotide in cupuladriid introns is adenine. In *C. biporosa* the average frequencies of adenine and uracil are 53.6% and 22.6%, respectively. In *D. cookae* the average frequencies of adenine and uracil are 49.9% and 31.9%, respectively (Supplementary Table S1). Concerning intron boundaries, only one of the 18 introns in *C. biporosa* had a degenerate Group III intron 5' end motif (NUNNG). This was found toward the 5' end of intron CB-cox3-ii (AACUAAG). Similarly, only two of the 15 introns in *D. cookae* for which the 5' ends were available, had the degenerate Group III intron motif towards their 5' ends: DC-nad5-i (UGACUUUUG) and DC-nad4L (GUAUG); nucleotides not emboldened are currently considered parts of the introns as they are the nucleotides immediately following the preceding exons. The degeneracy of the 3' end motif means that it is present too frequently to make any meaningful inferences. Lastly, the secondary structure drawings frequently show a stem towards the 3' end (Supplementary Figs. S5, S6) which may be a domain VI-like structure. However, at this stage, we consider the evidence for/against Group III introns too ambiguous to make any informed conclusions.

Putative intron origin(s). In order to examine the origin(s) of bryozoan mitochondrial introns in an evolutionary context, a phylogeny of RVT-IM domains with representatives from across the tree of life was constructed. The following description of phylogenetic interrelationships of RVT-IM domains (Fig. 5; for ML tree

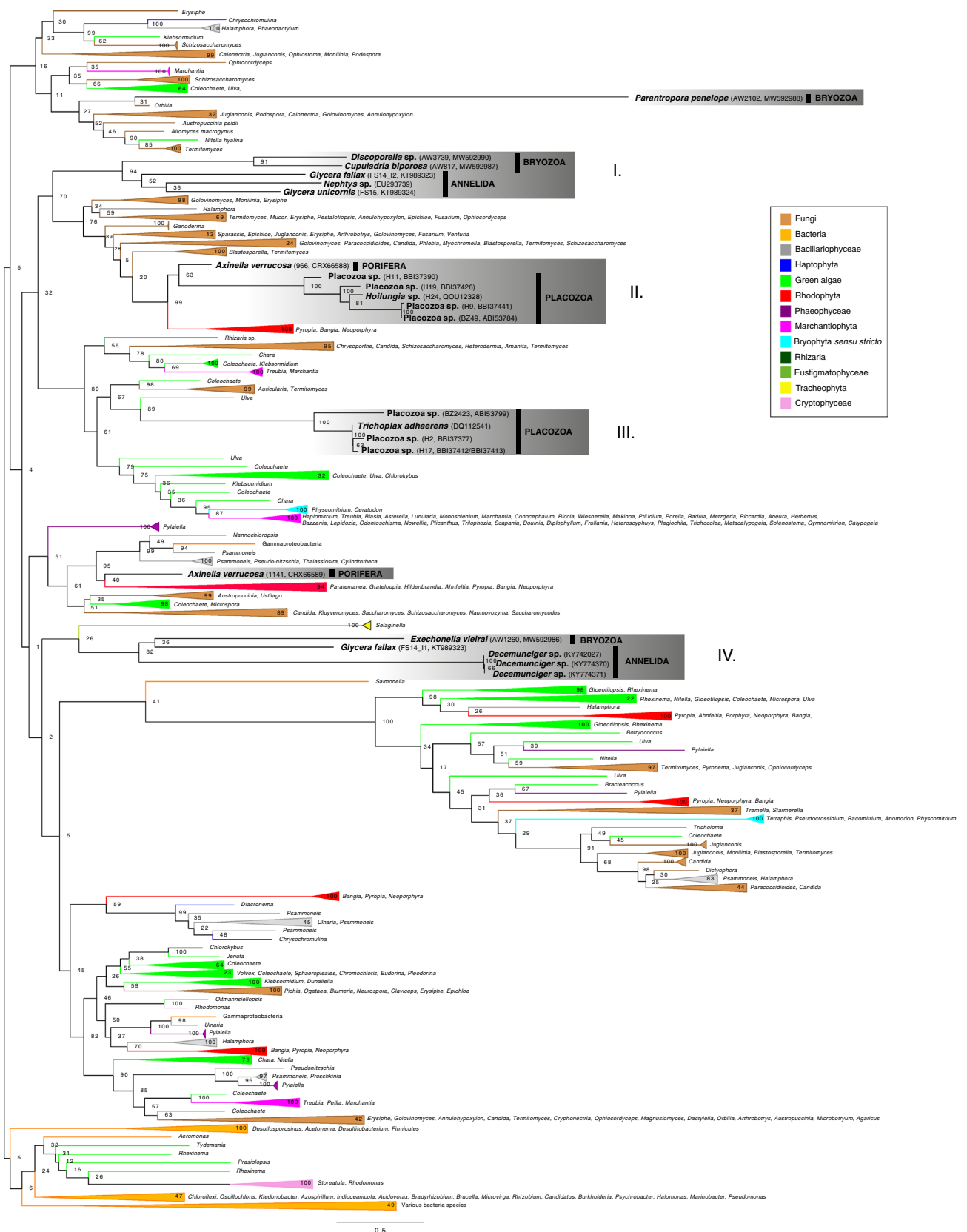


Figure 5. Maximum likelihood analysis of concatenated Group II reverse transcriptase and intron maturase open reading frames of *Exechonella vieirai*, *Parantropora penelope*, *Discoporella cookae* and *Cupuladria biporosa* and other metazoans (Porifera, Polychaeta, Placozoa), all emboldened and highlighted by shaded boxes, together with data from across the tree of life. Higher taxonomic affinities are colour coded (see insert). Ambiguously aligned positions had been excluded using Gblocks v.91b. Analysis was done using RAXML v.8.2.12 under the LG + G4 + F model with fast bootstrap analysis (1000 replicates). Values at nodes indicate maximum likelihood bootstrap values. Branch length scale bar indicates number of substitutions per site. *Clades I–IV*, as referred to in the text, are labelled.

with full terminal names, see Supplementary Fig. S7) focusses on the placement of the four bryozoan target taxa as well as other metazoans for which Group II IEPs with RVT-IM domains are known, i.e. Placozoa, Porifera and Annelida (polychaetes). Monophyletic groups of metazoan ORFs are indicated as *Clades I–IV*.

Clades I and II. The common origin of RVT-IM domains in the two cupuladriids *Cupuladria biporosa* and *Discoporella cookae* is strongly supported (91% bootstrap support [bs]). They formed strongly supported *Clade I* together with RVT-IM domains from polychaete taxa *Glycera unicornis* (FS15 isolate;³¹), *Glycera fallax* (FS14 isolate, I2 copy;³¹) and *Nephtys* sp. (94% bs); this close association of annelid terminals and the non-monophyly of the two tandemly repeated *Glycera fallax* FS14 intron copies I1 and I2 (see *Clade IV* description below for copy I1 position) was also found by Richter et al.³¹. This clade was sister to a paraphyletic assemblage of fungal sequences that also included the diatom *Halamphora calidilacuna*, and a strongly supported clade (99% bs) composed of RVT-IM domains of rhodophyte genera *Pyropia*, *Bangia* and *Neoporphyra* and the sponge *Axinella verrucosa* (copy 966²⁰), which formed the moderately supported sister (63% bs) to a clade of five placozoan sequences (Placozoa sp. haplotypes H9, H11, H19; *Hoilungia* sp. haplotype H24²⁶; Placozoa sp. strain BZ49²⁴). As in Signorovitch et al.²⁴ and Huchon et al.²⁰, an independent origin of *Placozoa* sp. strains BZ49 (here together with strains H9, H11, H19 and H24) and BZ2423 + *Trichoplax adhaerens* (here together with strain H2 and H17; see *Clade III* description) was found in our analysis. Furthermore, this close association of the abovementioned rhodophyte genera *Bangia* and *Pyropia*, placozoans and sponges is in agreement with findings by Huchon et al.²⁰.

Clade III. A second clade of placozoan sequences composed of *Placozoa* sp. haplotypes H2, H17²⁶, *Placozoa* sp. strain BZ2423²⁴ and *Trichoplax adhaerens*²³ nested in a moderately supported clade (67% bs) with charophyte *Coleochaete*, chlorophyte *Ulva* and fungi *Auricularia* and *Termitomyces*. In the wider context of a well-supported node (80% bs) that included members of Rhizaria, Fungi, Chlorophyta, Charophyta, Marchantiophyta and Bryophyta as well as placozoan *Clade III*, our results, although using a more elaborate taxon sampling, broadly agrees with results obtained by Huchon et al.²⁰. All genera recovered as close relatives of *Clade III* in Huchon et al.²⁰, i.e. *Marchantia*, *Treubia*, *Klebsormidium*, *Chlorokybus* and *Schizosaccharomyces* were also found to be part of this clade. A close association (although not strongly supported) of *Trichoplax adhaerens* with *Marchantia*, *Chlorokybus* and *Schizosaccharomyces* was also found by Vallès et al.³⁰. A novel finding of the present study, as a result of a broader taxon sampling, is the strongly supported (89%) sister-group relationship of placozoan *Clade III* with *Ulva compressa*.

Clade IV. *Exechonella vieirai* was placed with strong nodal support (82% bs) in a clade with polychaete taxa *Glycera fallax* (FS14 isolate, I1 copy³¹) and *Decemunciger* sp. The sister to *Clade IV* was formed by the tracheophyte *Selaginella*, but support for this was weak (26% bs). Similarly, neither Richter et al.³¹ nor Bernardino et al.³² were able to determine a supported position for *Glycera fallax* (FS14 isolate, I1 copy), although it formed a weakly supported clade with RVT-IM domains from *Marchantia* and land plants (*Arabidopsis*, *Solanum*, *Zea*, *Triticum*, *Vicia*, *Glycine*, *Oenothera*) in both of their analyses. Considering that both used the alignment by Zimmerly et al.⁵⁶ as a starting point (versus our de novo alignment which, amongst other more recently generated sequences, included *Selaginella* [XP_024524942, XP_024518366; published in 2018]), this is to be expected. In contrast to the present analysis, in Bernardino et al.³² the RVT-IM copies of *Decemunciger* sp. formed a strongly supported clade with *Nephtys* sp. (91% bs), nesting in a weakly supported clade with *Glycera unicornis* (FS15 isolate³¹) and *Glycera fallax* (FS14 isolate, I2 copy³¹). This difference in topology might be explained by the high sequence divergence in *Decemunciger* sp. copies, as evidenced by the long branch leading to the clade of the three *Decemunciger* sp. copies in Bernardino et al.³² and in Fig. 5 in the present study. This high sequence divergence makes an unambiguous alignment difficult and, combined with a different taxon sampling, including the closely related *Exechonella vieirai* copy, might have led to this difference in topology.

In addition to *Clades I–IV*, there were two terminals that did not group with any other metazoan RVT-IM copies. The first, *Axinella verrucosa* (copy 1141²⁰), formed a weakly supported clade (40% bs) with rhodophyte genera *Hildenbrandia*, *Ahnfeltia*, *Pyropia*, *Bangia*, *Neoporphyra* (as *Porphyra haitanensis* in Huchon et al.²⁰), all of which also formed the sister group in Huchon et al.²⁰. In addition, in the present analysis, this rhodophyte clade also included *Paralemanea*, *Grateloupia*. This *A. verrucosa* + Rhodophyta clade formed the sister group with strong support (95% bs) to a strongly supported clade (99% bs) composed of eustigmatophycean *Nannochloropsis*, a gamma-proteobacterium, and diatom genera *Psammoneis*, *Pseudo-nitschia*, *Thalassiosira* and *Cylindrotheca*; in Huchon et al.²⁰ this corresponding sister group was formed of *Thalassiosira* and *Chattonella*. Thus, there is broad agreement with our analysis and that by Huchon et al.²⁰, except that our broader taxon sampling added additional genera. We also observed, judging by the associated taxa in our analysis and that by Huchon et al.²⁰, the two intron copies of *Axinella verrucosa* (copy 966 = GenBank accession CRX66588; copy 1141 = GenBank accession CRX66589) came out in switched positions in ours and their topologies. We suspect that this could either be due to a mislabelling of Fig. 7 in Huchon et al.²⁰ or a mislabelling of their GenBank accession records. The position of the second terminal, *Parantropora penelope*, could not be unambiguously resolved as it nested on a very long branch in a poorly supported clade (16% bs) together with representatives of Fungi, Haptophyta, Bacillariophyceae, Chlorophyta, Charophyta and Marchantiophyta.

Our results show that Group II IEPs with RVT-IM domains were acquired independently numerous times amongst metazoans. Regarding our target bryozoan species, RVT-IM domains containing IEPs were likely acquired independently in *Exechonella vieirai* and *Parantropora penelope*. Furthermore, we infer a separate but possibly shared origin in the two cupuladriid taxa *Cupuladria biporosa* and *Discoporella cookae* (Fig. 5). Although it is conceivable that RVT-IM containing IEPs were acquired from multiple source organisms in a common ancestor of cheilostome bryozoans and were purged differentially during their evolution leading to today's

distribution of copies, the more parsimonious solution is likely the independent acquisition of introns. This question lends itself to be examined further using ancestral character estimation in the context of well sampled phylogenies which are being produced⁽⁵⁷⁾; Jenkins & Graham et al., in preparation). Moreover, the monophyly of bryozoan and annelid RVT-IM domains (*Clade I* and *Clade IV*; Fig. 5) implies that these copies had a common evolutionary origin. Much uncertainty remains regarding the insertion and propagation mechanisms of metazoan mt introns. Vertical transmission followed by independent losses has been proposed as mechanism in cnidarians and sponges¹⁷, whereas others have proposed a mixture of both horizontal gene transfer (HGT) and vertical transmission^{18,19,58}. However, much evidence points to intron insertion via HGT from microbial or algal donors. In sponges, putative intron donors include fungi^{16,21,22}, rhodophytes, diatoms and raphidophytes²⁰. Additionally, placozoans have been proposed as possible intron donors in sponges^{20,21}. Our analysis confirmed a close association of RVT-IM domains in sponges and placozoans (*Clade II*), as well as a close relationship of both with red algae (Fig. 5). Furthermore, *Clades I* and *II* formed a larger clade with fungal representatives, thus our findings corroborate previous speculations about possible origins. However, pinpointing the exact sources remains difficult, especially in the case of RVT-IM domains in *Axinella verrucosa* copy 1141, whose closest relative ranged from rhodophytes, diatoms, bacteria and Eustigmatophyceae. Still, the fact that we recovered four clades, each composed of multiple species and, in the cases of *Clades I, II* and *IV*, multiple phyla, indicates that introns in each of those clades likely originated from one type of marine organism, suggesting that certain intron donors are particularly successful in penetrating metazoan tissues.

A commonality of the metazoan taxa harbouring Group II introns is their ability to bud and regenerate. The idea that budding and regeneration ability may favour intron transmission via somatic cells/tissues was already formulated by Szitenberg et al.¹⁸ in the context of sponges. In the case of the regenerative annelid *Nephtys* sp.,⁵⁹ Vallès et al.³⁰ hypothesised that introns may have entered the germ line following HGT from possible bacterial donors via undifferentiated cells. Evidence from bryozoans in the present paper further supports the idea that organisms with regenerative abilities are easy targets for intron donors. The two cupuladriid species were found to possess an unprecedented large number of introns (18 in *Cupuladria biporosa*; 17 in *Discoporella cookae*) which is consistent with their particular reproductive life history strategy. The cupuladriid family are all free-living bryozoans that rely heavily on clonal propagation by fragmentation and regeneration of their disc-shaped colonies (e.g.⁴⁶). This mode of reproduction often results in zooids being split open⁶⁰ which could provide intron donors easy access to undifferentiated somatic cells. *Discoporella cookae* has undergone rates of clonal propagation exceeding 95% for at least 8 million years⁶¹ and clonal propagation in free-living bryozoans extends into the Cretaceous⁶². More generally, bryozoans are a rich source of bioactive compounds, many of which are likely produced by microbial symbionts^{63,64}. Although the associated microbial species are mostly unknown, detailed studies on some bryozoan species have shown their colonies to harbour symbiotic bacteria on the surfaces of rhizoids⁶⁵, intercellularly in the pallial sinus of their larvae and across larval surfaces^{65,66}, as well as in tissue strands (funicular cords) that connect colony modules to one another^{65,67}. Furthermore, bryozoans are often associated with microbial films⁶⁸ and have been found to be infected by fungal species⁶⁹. These close relationships between bryozoan hosts and microbes may have facilitated the acquisition of introns in this group.

Within species intron relationships. Phylogenetic analysis of IEP-less Group II introns from *E. vieirai* and *P. penelope* showed that the five copies of the latter formed a monophyly with strong nodal support (Fig. 6). Regarding the *E. vieirai* copies, two of them formed a clade with the *P. penelope* copies. However, their relationships were unresolved due to the low nodal support for one of the nodes (0.69 posterior probability, 48% bs). Thus, there is only evidence for a species-specific common ancestry of IEP-less Group II introns in *P. penelope*. This suggests that these introns may have either been inserted multiple times from the same type of source organism or that the introns self-propagated within the mitogenomes post initial insertion. Conversely, phylogenetic analyses of the IEP-less introns of the two cupuladriid taxa was completely unresolved and showed no grouping by species (Supplementary Fig. S8). Whether this is an indication of multiple insertion events from different sources or of a high post-insertion mutation rate cannot be inferred. In any case, intraspecific sequence divergence (uncorrected p-distances) was very high in all species and ranged from 38 to 60% in *E. vieirai* and 43–53% in *P. penelope* and from 39 to 59% in *C. biporosa* and 41–58% in *D. cookae* (Supplementary Tables S5, S6).

Unusual intron characteristics. There are several features in the bryozoan mitogenomes investigated here that distinguish them from other bilaterian and metazoan intron-harboring mitogenomes. In the context of bilaterians, our observed intron frequency in the four species of Bryozoa is unprecedented: eight in *E. vieirai* (if considering EV-H-nad5 and EV-nad5 a separate introns) and five, 18 and 17 in the incomplete mitogenomes of *P. penelope*, *C. biporosa* and *D. cookae*, respectively. This is the largest number of introns ever recorded in bilaterians (*Nephtys* sp.—one intron³⁰; *Glycera fallax*—two introns, *Glycera unicornis*—one intron³¹; *Decemunciger* sp.—three introns³²; *Cucullaea labiate*—one intron^{34,35}). Furthermore, RVT-IM domains have only ever been found in cox1 in metazoans (Placozoa^{23,24,26}, Porifera²⁰, Polychaeta^{30,32}). Although this was also the case in our *E. vieirai* and *P. penelope* mitogenomes, open-reading frames with RVT-IM domains were found in intergenic regions in *C. biporosa* and *D. cookae*. Thus, this is the first time that these domains have been found residing outside of cox1 introns in metazoans.

As foreseen by Richter et al.³¹, increased genome sequencing has revealed more Group II introns within the Bilateria, with more likely to be uncovered in the future. Nevertheless, for now at least, the frequency of mitochondrial introns in bryozoans is exceptional when compared to other bilaterians. This provides an unparalleled opportunity for bryozoans to perhaps become not only a model for studying introns but also bilaterian mitogenome architecture overall, challenging our understanding of their function and evolution. Initial observations

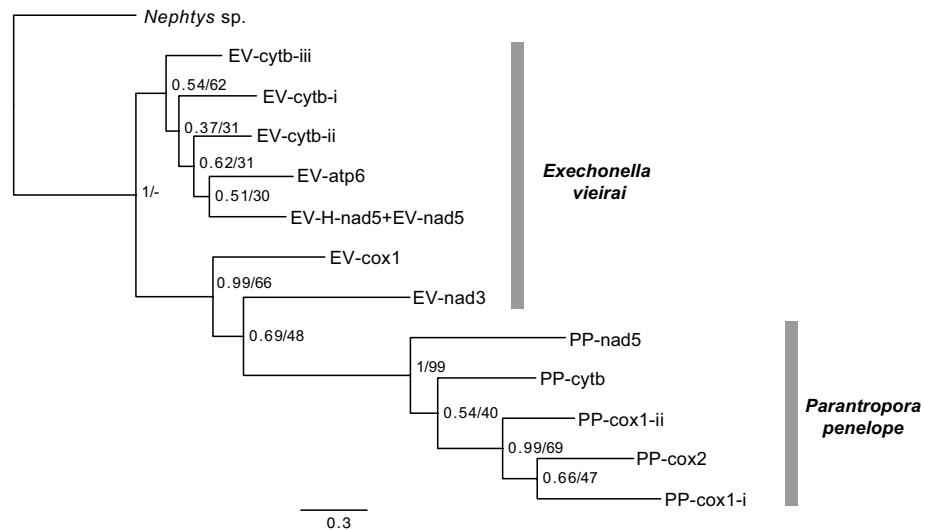


Figure 6. Bayesian analysis of *Exechonella vieirai* and *Parantropora penelope* intron sequences constructed using MrBayes v.3.2.6; 5,000,000 generations; 2,500,000 generations burn-in. Introns EV-H-nad5 and EV-nad5 had been concatenated. Ambiguously aligned positions had been excluded using Gblocks v.91b. Posterior probabilities and maximum likelihood bootstrap values (1000 replicates) as estimated using RAxML v.8.2.12 are given at the nodes. Analyses were carried out under the HKY + G model of nucleotide evolution. Reverse transcriptase and intron maturase open reading frames had been removed from EV-cox1 and PP-cox1-i introns. Intron names correspond to those shown in Fig. 2. Branch length scale bar indicates number of substitutions per site.

of intron absence versus presence in a broad phylogenetic context point towards a random acquisition process, rather than one guided by shared common ancestry (unpublished data). However, future work that dissects the process of intron gain and loss amongst closely related species ought to be explored. Furthermore, seeing that HGT from microbial donors is a likely mechanism for intron integration, studying the intraspecific distribution and variability of introns could provide interesting insights into their heritability and persistence.

Methods

Collection information. *Parantropora penelope* (specimen ID AW2102; Fig. 1A,B) was collected from Heron Island, Queensland, Australia by A.W. and J.S.P. in January 2018 (Great Barrier Reef Marine Parks Permit G17/40024.1). *Exechonella vieirai* (specimen ID AW1260; Fig. 1C,D) was collected from the intertidal zone of Praia de Pituba, Salvador, Bahia, Brazil in July 2017 (ICMBIO/SISBIO Permit 47108-1). *Cupuladria biporosa* (specimen ID AW817; Fig. 1E) was obtained from the Golfo de Mosquitos, Caribbean Panama by S.E.C. in August 2010 (Autoridad de los recursos Acuáticos de Panamá collecting permit #DGOMI-P|CFC-N°024). *Discoporella cookae* (specimen ID AW3739; Fig. 1F) was collected around San José Island, Las Perlas archipelago, Pacific Panamá by A.O. in February 2012 (Autoridad de los recursos Acuáticos de Panamá collecting permit #DGOMI-P|CFC-N°02-A). All specimens were preserved in 95–100% ethanol. Corresponding specimens were deposited as morphological vouchers at the Natural History Museum, UK (NHMUK) collection (see Fig. 1 legend for NHMUK accession numbers). Vouchers were imaged by scanning electron microscopy using a LEO 1455-VP instrument at NHMUK.

Illumina sequencing, assembly and annotation. Total genomic DNA (gDNA) was extracted using the DNeasy Blood & Tissue Kit (Qiagen) following manufacturer's instructions. Double stranded (ds) DNA concentration was quantified using a Qubit™ fluorometer using either the Qubit™ dsDNA BR (Broad Range) or dsDNA HS (High Sensitivity) assay kits. Dual-indexed libraries were prepared using the TruSeq DNA Nano Library Prep Kit (Illumina, Inc., San Diego, California) for *C. biporosa*, and the Nextera DNA Flex Library Prep Kit (now Illumina DNA Prep; Illumina, Inc., San Diego, California) for *E. vieirai*, *P. penelope* and *D. cookae*. Sequencing was performed by Novogene (HK) Company Limited on the Illumina HiSeq 4000 platform (San Diego, California) using 2 × 150 bp paired-end sequencing.

Paired-end reads were trimmed using Trimmomatic Version 0.39⁷⁰ and assembled de novo using SPAdes 3.13.0⁷¹ with k-mer sizes of 33, 55, 77, 99 and 127. Contigs of putative mitogenomes were identified by conducting blast searches⁷² against a local custom database of reference bryozoan mitogenomes in Geneious 11.1.4 (<https://www.geneious.com>). Candidate mt contigs were further verified by blastn searches against the NCBI database (<https://blast.ncbi.nlm.nih.gov>).

Mitogenome contigs were annotated using MITOS (<http://mitos.bioinf.uni-leipzig.de/>)⁷³. PCG boundaries (start/stop codons) were refined manually in Geneious following alignment with curated sets of reference bryozoan sequences using TranslatorX (<http://translatorx.co.uk/>)⁷⁴. Mitogenomes were circularised where possible using the Repeat Finder plug-in in Geneious. Assembly quality and read coverage were assessed by reference

mapping of trimmed reads to mitogenome contigs under strict settings in Geneious. The settings used were as follows: allow gaps = maximum 1% per read, maximum gap size = 3, minimum overlap = 50, minimum overlap identity = 95%, word length = 18, index word length = 13, ignore words repeated more than 12 times, maximum mismatches per read 2%, maximum ambiguity = 4.

Intron identification and validation. Introns were identified through the observation of fragmented exons. Exons were knitted together in light of alignments with non-intron containing reference data. In cases where exons had not been identified by MITOS, nucleotide data in between exons were translated into amino acids using EMBOSS Transeq (https://www.ebi.ac.uk/Tools/st/emboss_transeq) and searched using blastp (<https://blast.ncbi.nlm.nih.gov>) to identify additional exons. In cases where blastp failed to identify matches, the amino acid translation was scanned for conserved motifs as gleaned from reference alignments.

In order to identify intron-encoded RVT-IM domains, intron sequences were translated into amino acids using EMBOSS Transeq from all six reading frames. Translated sequences were subjected to searches using Pfam 33.1 (<http://pfam.xfam.org/>)⁷⁵ and blastp. No RVT-IM domains were initially found for *Cupuladria* and *Discoporella*. However, blastx searches of the complete mt fragments against the ucalgary.ca database (<http://webapps2.ucalgary.ca/~groupii/cgi-bin/main/blastusr.php>)⁵⁰ identified intergenic regions that were subsequently confirmed as RVT-IM domains by blastp and Pfam searches. Significant as well as insignificant Pfam results were considered. Conserved Group II ribozyme domains were identified using Rfam (<https://rfam.xfam.org/>)⁷⁶. Intron boundaries were refined by searching for conserved 5' (GUGYG) and 3' terminals ([Y]AY)^{12,13}.

The *bona fide* presence of 22 of the 48 identified introns was validated by PCR and Sanger sequencing, using specific exon–intron primers (Supplementary Table S4). PCR amplification used puRE Taq Ready-to-go PCR beads (Amersham Biosciences) with a total reaction volume of 25 µl, using 1 µl of 10 µM of each primer and 3 µl of template gDNA. PCR cycling conditions were as follows: initial denaturation at 94 °C for 3 min, followed by 35 cycles at 94 °C for 30 s, T_{ann} for 30 s, 72 °C for 1 min (3 min if product > 1000 bp), with a final extension step at 72 °C for 10 min (see Supplementary Table S4 for primer pair-specific T_{ann}). Two PCRs (EV-cox1 and PP-cox1-i) failed and were repeated with Takara Long-range PCR kit (Takara Bio Inc.). Total reaction volume was 50 µl, using 0.5 µl enzyme, 5 µl buffer, 8 µl dNTPS, 2 µl of 10 µM of each primer and 4 µl gDNA. PCR cycling conditions were as follows: initial denaturation at 94 °C for 2 min, followed by 40 cycles at 94 °C for 20 s, 51 °C for 30 s, 68 °C for 3 min, with a final extension at 68 °C for 10 min. Sequencing using the PCR primers was performed using an Applied Biosystems 3730 DNA Analyser, using BigDye version 3.1. Sanger reads were edited in Geneious. Sequence identity was verified by mapping to the reference contigs in Geneious.

Secondary structure drawings. Following the excision of IEP ORFs, intron sequences of *E. vieirai* and *P. penelope* were subjected to secondary structure folding at 20 °C using mfold (www.unafold.org)^{77–79}. The consensus Group II secondary structures on the Zimmerly website (<http://webapps2.ucalgary.ca/~groupii/>)^{10,49,50} were taken as guides for subsequent manual secondary structure manipulation using forna (<http://rna.tbi.univie.ac.at/forna/>)⁸⁰ in conjunction with iterative folding of parts of the sequences in mfold. Regarding intron sequences of *C. biporosa* and *D. cookae*, secondary structure folding was conducted at 20 °C using mfold, but no manual manipulations were performed.

Intron phylogenies. Phylogenetic analyses of Group II intron RVT-IM domains were carried out as follows: Genbank searches of the protein database were conducted using terms ‘reverse transcriptase AND mitochondrion’ (2322 hits; accessed 9th March 2021) and ‘maturase AND mitochondrion’ (5187 hits; accessed 9th March 2021). Furthermore, RVT-IM domains of the following metazoans were added: *Trichoplax adhaerens* (DQ112541), Placozoa sp. (ABI53784, strain BZ49; ABI53799, strain BZ2423; BBI37377, haplotype H2; BBI37412/BBI37413, haplotype H17; BBI37390, haplotype H24), *Axinella verrucosa* (CRX66588, intron 966; CRX66589, intron 1141), *Nephtys* sp. (EU293739), *Glycera unicornis* (KT989324, isolate FS15), *Glycera fallax* (KT989323, isolate FS14, introns I1 and I2) and *Decemunciger* sp. (KY742027, KY774370, KY774371); this is the first time that RVT-IM domains of placozoan haplotypes H2, H9, H11, H17, H19 and of the previously unsampled genus *Hoilungia* (H24)²⁶, have been put into a ‘tree-of-life’ phylogenetic context. These ORFs were trimmed following Pfam analyses. No Pfam match for RVT-IM could be obtained for *Endomyzostoma* sp. In order to maximise our chances of identifying metazoan Group II intron origins, we also included blast matches of the abovementioned accessions and the bryozoans RVT-IM domains in our analyses. Duplicate accessions were removed in Mesquite v.3.51⁸¹. Amino acid sequence duplicates were removed using the web server for FASTA tools unique sequences (https://www.ncbi.nlm.nih.gov/CBBresearch/Spouge/html_ncbi/html_fasta/uniqueseq.cgi).

Initial alignments were constructed in MAFFT v.7.453⁸² using the --auto setting. Alignments were examined by eye in Geneious and obvious outliers were removed. Initial neighbour-joining trees, produced in PAUP* v.4.0a⁸³, identified a large clade composed of plant RVT-IM sequences. As none of the focus sequences were nesting in this clade, it was removed from further analyses. The final alignment was carried out using the MAFFT settings --amino --bl 30 --genafpair --maxiterate 100. Ambiguously aligned positions were excluded using the stand-alone version of Gblocks v. 0.91b^{84,85} using the following settings: minimum number of sequences for a conserved position = lowest; minimum number of sequences or a flank position = lowest; maximum number of contiguous non-conserved positions = 10; minimum length of a block = 5; allowed gap positions = with half. The final alignment consisted of 1059 terminals and 3947 positions, of which 367 remained included. ModelTest-NG v.0.1.6⁸⁶ was used to determine the best model of amino acid substitution. Maximum likelihood (ML) phylogenetic analysis with 1000 fast bootstrap replicates was carried out using RAxML v.8.2.12⁸⁷ under the LG + G4 + F model. Uncorrected p-distances were calculated using PAUP* v.4.0a⁸³.

To assess common ancestry of introns within species, phylogenetic analyses of (a) Group II IEP-less introns of *Exechonella* and *Parantropora*, and (b) IEP-less introns of *Cupuladria* and *Discoporella* were carried out. The outgroup for analysis (a) was the well-annotated Group II intron *Nephtys* sp. (EU293739³⁰) from which the IEP ORF had been excised according to their secondary structure drawing (Fig. 1 in³⁰); no suitable outgroup was available for analysis (b). Alignments were constructed in MAFFT with settings --genafpair --maxiterate 1000. Ambiguously aligned positions were excluded using the stand-alone version of Gblocks as outlined above. ModelTest-NG was used to determine the best model of nucleotide substitution. ML with 1000 fast bootstrap replicates and BI analyses were conducted in RAXML and MrBayes v.3.2.6⁸⁸ under the HKY + G model. Two parallel MrBayes runs were performed for 5 million generations. The burn-in was defined as the point at which the average standard deviation of split frequencies was < 0.01.

Data availability

For alignments, trees and analyses commands see TreeBASE: <http://purl.org/phylo/treebase/phylovs/study/TB2:S29524>.

Received: 11 January 2022; Accepted: 7 June 2022

Published online: 28 June 2022

References

- Gray, M. W. Mitochondrial evolution. *Cold Spring Harb. Perspect. Biol.* **4**, a011403 (2012).
- Gray, M. W. *et al.* Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res.* **26**, 865–878 (1998).
- Zardoya, R. Recent advances in understanding mitochondrial genome diversity. *F1000* **9**, 270 (2020).
- Lynch, M., Koskella, B. & Schaack, S. Mutation pressure and the evolution of organelle genomic architecture. *Science* **311**, 1727–1730 (2006).
- Lynch, M. Streamlining and simplification of microbial genome architecture. *Annu. Rev. Microbiol.* **60**, 327–349 (2006).
- Nielsen, H. & Johansen, S. D. Group I introns moving in new directions. *RNA Biol.* **6**, 375–383 (2009).
- Hausner, G., Hafez, M. & Edgell, D. R. Bacterial group I introns: Mobile RNA catalysts. *Mob. DNA* **5**, 8 (2014).
- Haugen, P., Simon, D. M. & Bhattacharya, D. The natural history of group I introns. *Trends Genet.* **21**, 111–119 (2005).
- Nawrocki, E. P., Jones, T. A. & Eddy, S. R. Group I introns are widespread in archaea. *Nucleic Acids Res.* **46**, 7970–7976 (2018).
- Simon, D. M. *et al.* Group II introns in Eubacteria and Archaea: ORF-less introns and new varieties. *RNA* **14**, 1704–1713 (2008).
- Novikova, O. & Belfort, M. Mobile group II introns as ancestral eukaryotic elements. *Trends Genet.* **33**, 773–783 (2017).
- Lambowitz, A. M. & Zimmerly, S. Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* **3**, a003616 (2011).
- Michel, F., Umesono, K. & Ozeki, H. Comparative and functional anatomy of group II catalytic introns – A review. *Gene* **82**, 5–30 (1989).
- Zimmerly, S. & Semper, C. Evolution of group II introns. *Mob. DNA* **6**, 7 (2015).
- Michel, F. & Westhof, E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**, 585–610 (1990).
- Rot, C., Goldfarb, I., Ilan, M. & Huchon, D. Putative cross-kingdom horizontal gene transfer in sponge (Porifera) mitochondria. *BMC Evol. Biol.* **6**, 71 (2006).
- Wang, X. & Lavrov, D. V. Seventeen new complete mtDNA sequences reveal extensive mitochondrial genome evolution within the Demospongiae. *PLoS ONE* **3**, e2723 (2008).
- Szitenberg, A., Rot, C., Ilan, M. & Huchon, D. Diversity of sponge mitochondrial introns revealed by cox1 sequences of Tetillidae. *BMC Evol. Biol.* **10**, 288 (2010).
- Erpenbeck, D., Aryasari, R., Hooper, J. N. A. & Wörheide, G. A mitochondrial intron in a verongid sponge. *J. Mol. Evol.* **80**, 13–17 (2015).
- Huchon, D., Szitenberg, A., Shefer, S., Ilan, M. & Feldstein, T. Mitochondrial group I and group II introns in the sponge orders Agelasida and Axinellida. *BMC Evol. Biol.* **15**, 278 (2015).
- Kelly, M. & Cárdenas, P. An unprecedented new genus and family of Tetractinellida (Porifera, Demospongiae) from New Zealand's Colville Ridge, with a new type of mitochondrial group I intron. *Zoo. J. Linn. Soc.-Lond.* **177**, 335–352 (2016).
- Cranston, A., Taboada, S., Koutsouveli, V., Schuster, A. & Riesgo, A. A population specific mitochondrial intron from the sponge *Phakellia robusta* in the North-East Atlantic. *Deep-Sea Res. Pt. I* **172**, 103534 (2021).
- Dellaporta, S. L. *et al.* Mitochondrial genome of *Trichoplax adherens* supports Placozoa as the basal lower metazoan phylum. *Proc. Natl. Acad. Sci. USA* **103**, 8751–8756 (2006).
- Signorovitch, A. Y., Buss, L. W. & Dellaporta, S. L. Comparative genomics of large mitochondria in placozoans. *PLoS Genet.* **3**, e13 (2007).
- Burger, G., Yan, Y., Javadi, P. & Lang, B. F. Group I-intron trans-splicing and mRNA editing in the mitochondria of placozoan animals. *Trends Genet.* **25**, 381–386 (2009).
- Miyazawa, H. *et al.* Mitochondrial genome evolution of placozoans: gene rearrangements and repeat expansions. *Genome Biol. Evol.* **13**, evaa213 (2021).
- Beagley, C. T., Okada, N. A. & Wolstenholme, D. R. Two mitochondrial group I introns in a metazoan, the sea anemone *Metridium senile*: One intron contains genes for subunits 1 and 3 of NADH dehydrogenase. *Proc. Natl. Acad. Sci. USA* **93**, 5619–5623 (1996).
- Van Oppen, M. J. H. *et al.* The mitochondrial genome of *Acropora tenuis* (Cnidaria; Scleractinia) contains a large group I intron and a candidate control region. *J. Mol. Evol.* **55**, 1–13 (2000).
- Goddard, M. R., Leigh, J., Roger, A. J. & Pemberton, A. J. Invasion and persistence of a selfish gene in the Cnidaria. *PLoS ONE* **1**, e3 (2006).
- Vallès, Y., Halanynch, K. M. & Boore, J. L. Group II introns break new boundaries: Presence in a bilaterian's genome. *PLoS ONE* **3**, e1488 (2008).
- Richter, S., Schwarz, F., Hering, L., Böggemann, M. & Bleidorn, C. The utility of genome skimming for phylogenomic analyses as demonstrated for glycerid relationships (Annelida, Glyceridae). *Genome Biol. Evol.* **7**, 3443–3462 (2015).
- Bernardino, A. F., Li, Y., Smith, C. R. & Halanynch, K. M. Multiple introns in a deep-sea annelid (*Decemunciger*: Ampharetidae) mitochondrial genome. *Sci. Rep.* **7**, 4295 (2017).
- Zhong, M. Applicability of mitochondrial genome data to annelid phylogeny and the evolution of group II introns. *PhD Thesis*. (Auburn University, 2009).
- Feng, Y., Li, Q., Yu, H. & Kong, L. Complete mitochondrial genome sequence of *Cucullaea labiata* (Arcoida: Cucullaeidae) and phylogenetic implications. *Genes Genom.* **39**, 867–875 (2017).

35. Kong, L. *et al.* Mitogenomics reveals phylogenetic relationships of Arcoidea (Mollusca, Bivalvia) and multiple independent expansions and contractions in mitochondrial genome size. *Mol. Phylogenet. Evol.* **150**, 106857 (2020).
36. Zhang, Z. *et al.* Fossil evidence unveils an early Cambrian origin for Bryozoa. *Nature* **599**, 251–255 (2021).
37. Ryland, J. S. *Bryozoa* (Hutchinson University Library, 1970).
38. Schwaha, T. *Handbook of Zoology: Phylum Bryozoa* (ed. Schwaha, T.) (De Gryter, 2020).
39. Hageman, S. J., Bock, P. E., Bone, Y. & McGowran, B. Bryozoan growth habits: Classification and analysis. *J. Paleontol.* **72**, 418–436 (1998).
40. Wood, A. C. L., Probert, P. K., Rowden, A. A. & Smith, A. M. Complex habitat generated by marine bryozoans: a review of its distribution, structure, diversity, threats and conservation. *Aquatic Conserv. Mar. Freshw. Ecosyst.* **22**, 547–563 (2012).
41. Bock, P. E. & Gordon, D. P. Phylum Bryozoa Ehrenberg, 1831. *Zootaxa* **3703**, 67–74 (2013).
42. Waeschenbach, A., Taylor, P. D. & Littlewood, D. T. J. A molecular phylogeny of bryozoans. *Mol. Phylogenet. Evol.* **62**, 718–735 (2012).
43. Bushnell, J. H. & Rao, K. S. Dormant or quiescent stages and structures among the Ectoprocta: physical and chemical factors affecting viability and germination of statoblasts. *Trans. Am. Microsc. Soc.* **93**, 524–543 (1974).
44. Lidgard, S. Ontogeny in animal colonies: A persistent trend in the bryozoan fossil record. *Science* **232**, 230–232 (1986).
45. Taylor, P. D., Di Martino, E. & Martha, S. O. Colony growth strategies, dormancy and repair in some Late Cretaceous encrusting bryozoans: Insights into the ecology of the Chalk seabed. *Palaeobiol. Palaeoenvir.* **99**, 425–446 (2019).
46. O’Dea, A., Herrera-Cubilla, A., Fortunato, H. & Jackson, B. C. Life history variation in cupuladriid bryozoans from either side of the Isthmus of Panama. *Mar. Ecol. Prog. Ser.* **280**, 145–161 (2004).
47. Ojala, D., Montoya, J. & Attardi, G. tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**, 470–474 (1981).
48. Dubot, C. *et al.* GUG is an efficient initiation codon to translate the human mitochondrial ATP6 gene. *Biochem. Biophys. Res. Co.* **313**, 687–693 (2004).
49. Dai, L., Toor, N., Olson, R., Keeping, A. & Zimmerly, S. Database for mobile group II introns. *Nucleic Acids Res.* **31**, 424–426 (2003).
50. Candales, M. A. *et al.* Database for bacterial group II introns. *Nucleic Acids Res.* **40**, D187–190 (2012).
51. McNeil, B. A., Semper, C. & Zimmerly, S. Group II introns: Versatile ribozymes and retroelements. *Wiley Interdiscip. Rev. RNA* **7**, 341–355 (2016).
52. Christopher, D. A. & Hallick, R. B. *Euglena gracilis* chloroplast ribosomal protein operon: a new chloroplast gene for ribosomal protein L5 and description of a novel organelle intron category designated group III. *Nucleic Acids Res.* **17**, 7591–7608 (1989).
53. Copertino, D. W. & Hallick, R. B. Group II and group III introns of twintrons: Potential relationships with nuclear pre-mRNA introns. *Trends Biochem. Sci.* **18**, 467–471 (1993).
54. Hong, L. & Hallick, R. B. A group III intron is formed from domains of two individual group II introns. *Gene Dev.* **8**, 1589–1599 (1994).
55. Copertino, D. W., Hall, E. T., Van Hook, F. W., Jenkins, K. P. & Hallick, R. B. A group III twintron encoding a maturase-like gene exercises through lariat intermediates. *Nucleic Acids Res.* **22**, 1029–1036 (1994).
56. Zimmerly, S., Hausner, G. & Wu, X.-C. Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Res.* **29**, 1238–1250 (2001).
57. Orr R. J. S. *et al.* Paleozoic origins of cheilostome bryozoans and their parental care inferred by a new genome-skimmed phylogeny. *Sci. Adv.* **8**, eabm7452 (2022).
58. Fukami, H., Chen, C. A., Chiou, C.-Y. & Knowlton, N. Novel group I introns encoding a putative homing endonuclease in the mitochondrial *cox1* gene of scleractinian corals. *J. Mol. Evol.* **64**, 591–600 (2007).
59. Clark, M. E. Later stages of regeneration in the polychaete. *Nephtys*. *J. Morph.* **124**, 483–510 (1968).
60. O’Dea, A., Jackson, J. B. C., Taylor, P. D. & Rodriguez, F. Modes of reproduction in recent and fossil cupuladriid bryozoans. *Palaeontology* **51**, 847–864 (2008).
61. O’Dea, A. & Jackson, J. B. C. Environmental change drove macroevolution in cupuladriid bryozoans. *Proc. R. Soc. Lond. B Biol.* **276**, 3629–3634 (2009).
62. O’Dea, A., Håkansson, E., Taylor, P. D. & Okamura, B. Environmental change prior to the K-T boundary inferred from temporal variation in the morphology of cheilostome bryozoans. *Palaeogeogr. Palaeoclimatol.* **308**, 502–512 (2011).
63. Figueroa, B. & Avila, C. The phylum Bryozoa as promising source of anticancer drugs. *Mar. Drugs* **17**, 477 (2019).
64. Ciavatta, M. L. *et al.* The phylum Bryozoa: From biology to biomedical potential. *Mar. Drugs* **18**, 200 (2020).
65. Sharp, K. H., Davidson, S. K. & Haygood, M. G. Localization of ‘*Candidatus Endobugula sertula*’ and the bryostatins throughout the life cycle of the bryozoan *Bugula neritina*. *ISME J.* **1**, 693–702 (2007).
66. Woollacott, R. M. Association of bacteria with bryozoan larvae. *Mar. Biol.* **65**, 155–158 (1981).
67. Karagodina, N. P., Vishnyakov, A. E., Kotenko, O. N., Maltseva, A. L. & Ostrovsky, A. N. Ultrastructural evidence for nutritional relationships between a marine colonial invertebrate (Bryozoa) and its bacterial symbionts. *Symbiosis* **75**, 155–164 (2018).
68. Scholz, J. Epibiontic microorganisms as a local control factor of bryozoan distribution and bryozoans ‘micro-reefs’. *Beitr. Paläontol.* **20**, 75–87 (1995).
69. Sterflinger, K., Hain, M., Scholz, J. & Wasson, K. Fungal infections of a colonial marine invertebrate: Diversity and morphological consequences. *Facies* **45**, 31–38 (2001).
70. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
71. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
72. Altschul, S. F. *et al.* Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
73. Bernt, M. *et al.* MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* **69**, 313–319 (2013).
74. Abascal, F., Zardoya, R. & Telford, M. J. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38**, W7–13 (2010).
75. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
76. Kalvari, I. *et al.* Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* **46**, D335–D342 (2018).
77. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
78. SantaLucia, J. Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA* **95**, 1460–1465 (1998).
79. Peyret, N. *Prediction of Nucleic Acid Hybridization: Parameters and Algorithms*. PhD Dissertation (Wayne State University, 2000).
80. Kerpedjiev, P., Hammer, S. & Hofacker, I. L. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics* **31**, 3377–3379 (2015).
81. Maddison, W. P. & Maddison, D. R. *Mesquite: A Modular System for Evolutionary Analysis*. Version 3.61. <http://www.mesquiteproject.org> (2019).
82. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
83. Swofford, D. L. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. (Sinauer Associates, 2003).

84. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
85. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
86. Darriba, D. *et al.* ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* **37**, 291–294 (2019).
87. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
88. Ronquist, F. *et al.* MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

Acknowledgements

This work was funded by the Leverhulme Trust (Research Project Award RPG-2016-429). We thank the staff at Heron Island Research Station (University of Queensland) for fieldwork support and Kirrily Moore at the Tasmanian Museum and Art Gallery for specimen transfer. We thank the staff at the Museu de História Natural (Universidade Federal da Bahia) for exporting the Brazilian samples to the Natural History Museum London. Permission to collect on Heron Island and in Brazil was given by the Great Barrier Reef Marine Parks and the Sistema de Autorização e Informação em Biodiversidade of the Instituto Chico Mendes de Conservação da Biodiversidade, respectively. Permits to collect in Panama were given by the Autoridad de los recursos Acuáticos de Panamá. Aaron O’Dea was supported by the *Sistema Nacional de Investigación* (SNI) de SENACYT. Leandro M. Vieira was supported by CNPq (Proc. 311523/2021-8) and FAPESP (Proc. 19/17721-9). Thanks goes to Mary Spencer Jones for handling and accessioning of the voucher specimens. We thank Paul D. Taylor and Silviu O. Martha for producing SEM images and the NHM Sequencing Facility for producing Sanger data. We also thank three anonymous reviewers for their insightful and helpful comments.

Author contributions

H.L.J. carried out DNA extractions, assembled Illumina reads, annotated mitogenomes, interpreted the data, and wrote the manuscript. R.G. assembled Illumina reads, annotated mitogenomes and produced Fig. 1. J.S.P., L.M.V., A.C.S.A., A.O., S.E.C. conducted the fieldwork to acquire specimens and commented on manuscript drafts. A.H. prepared the Illumina libraries. A.W. conceived the project, conducted the fieldwork to acquire specimens, annotated mitogenomes, carried out phylogenetic analyses, interpreted the data and wrote the manuscript. All authors have reviewed and approved the submitted manuscript version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-14477-3>.

Correspondence and requests for materials should be addressed to H.L.J. or A.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© Crown 2022