THE ROYAL SOCIETY
PUBLISHING

Confirmation of the first intronic sequence in the bivalvian mitochondrial genome of *Macoma balthica* (Linnaeus, 1758)

Marek Lubośny[1], Beata Śmietanka[1], Rafał Lasota[2] and Artur Burzyński[1]

[1]Department of Genetics and Marine Biotechnology, Institute of Oceanology Polish Academy of Sciences, Sopot 81-712, Poland
[2]Faculty of Oceanography and Geography, Division of Marine Ecosystems Functioning, University of Gdańsk, Gdynia 81-378, Poland

ML, 0000-0001-6370-8880; BŚ, 0000-0002-9149-9120; RL, 0000-0002-2745-3515; AB, 0000-0002-0185-197X

In 2020, the first male-type mitochondrial genome from the clam *Macoma balthica* was published. Apart from the unusual doubly uniparental inheritance of mtDNA, scientists observed a unique (over 4k bp long) extension in the middle of the *cox2* gene. We have attempted to replicate these data by NGS DNA sequencing and explore further the expression of the long *cox2* gene. In our study, we report an even longer *cox2* gene (over 5.5 kbp) with no stop codon separating conserved *cox2* domains, as well as, based on the rtPCR, a lower relative gene expression pattern of the middle part of the gene (5′ = 1; mid = 0.46; 3′ = 0.89). Lastly, we sequenced the *cox2* gene transcript proving the excision of the intronic sequence.

1. Introduction

With the growing popularity of next-generation sequencing methods, scientists are discovering more new features of mitogenomes, breaking the established view of conserved structure and compactness of animal mitogenomes. Considering only animal mitogenomes, the length of mtDNA can vary greatly from the usual 16–17 k base pairs (bp) [1]. The longest known animal mitochondrial DNA belongs to clams from the Arcidae family *Anadara sativa* and reaches a length of 48 kbp (KF667521). The shortest 10 kbp long belongs to the comb jelly *Mnemiopsis leidyi* [2] (NC_016117). Mitochondrial DNA can also differ in structure from typical circular, throughout linear in some medusozoan cnidarians, calcareous sponges, ending in multipartite, fragmented in hydrozoan cnidarians, lice and jellyfish [3]. What is more, the number of protein-coding genes can also vary. There are mitogenomes lacking some of the canonical 13 protein: *atp8* in flatworms [4,5], nematodes [6,7], *atp6* and *atp8* in comb jelly [2], *atp8* and *nad5* in Hong Kong whipping frog [8], *nad5* in tuatara [9] up to mitogenomes with additional genes: *mtMuts* in Octocorallia [10]; *dnaB* in jellyfish [11]; *atp9* in sponges [12] and protein-coding novel open reading frames (*orf's*) in bivalves [13–17]. It is worth mentioning here that in some rare cases, the 'missing' genes might have been omitted and remain unannotated due to the low-sequence homology to their counterpart genes present in the available databases at that time (e.g. GenBank). One of those situations was reported in bivalves from genus *Mytilus*, which were thought for a long time to be missing *atp8* gene in their mtDNA [18–21]. The typical 'average' metazoan mitogenome does not contain any introns, but as usual in these cases, even here scientists found exceptions. Most cases concern group I and

II introns found in non-bilaterian animals: hexacorals [22–26], demosponges [27–31], and placozoan [32,33]. Such taxonomic distribution is often used as empirical support for mutational-pressure hypothesis explaining lack of introns in fast-evolving mtDNA [34]. Up to this point, the only bilaterian animals where mitochondrial intronic sequences have been identified are several species of annelids [35–38].

Bivalves are one of the animal classes with a wide range of mitochondrial features scattered over many species that distinguish them from the typical structure of mtDNA. Some of them vary in mitogenome size, some possess additional tRNA, gene extensions, and additional protein-coding open reading frames often associated with doubly uniparental inheritance of mitochondrial DNA (DUI), which is unique for bivalves. In DUI species male individuals are heteroplasmic, possess two divergent mitogenomes (Female-type and faster evolving Male-type), while most of the female individuals are homoplasmic toward F-type mitogenome. M-type mitogenomes are localized mostly in male gonads and are typically the only mitotype passed to progeny through sperm [15,21, 39–45]. Fertilized egg cells receive F-type mtDNA from mother and M-type from father, then depending on the sex of future progeny (scenario when the embryo grows into the female individual), male mitochondria get dispersed during cell divisions resulting in quick loss of detectable signal from male-type mtDNA. In the second case, when the embryo develops into a male specimen, M-type mitochondria migrate grouped together during the first cell divisions, becoming the dominant fraction in the future gonads [46–49].

The Baltic clam *Macoma balthica* (Linnaeus 1758) is an infaunal tellinid bivalve that is commonly found in marine and estuarine soft-bottom habitats in the Northern Hemisphere. The first sequencing of mitogenome of this species was performed by Saunier *et al.* [50], who assembled and annotated six nearly complete mitochondrial genomes from five mitochondrial lineages distinguished based on *cox1* haplotypes [51]. Recently, a male-type mitogenome of *Macoma balthica* was published [52]. Its unusual feature was a very long (over 4 kbp long) cytochrome c oxidase subunit II gene with stop codon separating the 5′ part of *cox2* gene. This caught our interest, so we decided to analyse the expression pattern of this gene to check if the whole sequence is expressed as a single transcript, as well as to verify if the reported stop codon that separates *cox2* domains is true or only an assembly artefact.

2. Methods

Individuals of *Macoma balthica*, used for DNA analysis and NGS DNA sequencing, were collected in the Gulf of Gdansk (southern Baltic Sea, Poland) using a bottom dredge on 8 November 2017. The Gulf of Gdansk is a brackish-water basin, dominated by soft-bottom habitats, with an elevated levels of pollution, eutrophication and primary production. Water temperature ranges annually from −0.43 to 24.7°C, whereas average annual water salinity is around 7–8 PSU ([53] and references therein). Collected specimens were checked for the presence of sperm and egg cells under a microscope, sectioned and stored in the freezer in 70% ethanol until DNA extraction. Individuals for RNA samples were gathered from the beach in Sopot (Poland) after the storm in February 2021 and stored frozen at −80°C. DNA extraction from male gonads was performed according to Hoarau *et al.* [54] (the protocol was described in detail in [55]). RNA was

extracted from the entire body (half of the individual; approximately 1 cm × 1 cm shells) with the GenElute Mammalian Total RNA miniprep kit (Sigma). After RNA extraction, the remaining contaminating DNA was digested with DNase I (EurX) and purified again on new columns from the GenElute Mammalian Total RNA kit. To ensure the presence of chaotropic salts allowing for the binding of the nucleic acids to the columns silica resin, RNA extracts were mixed with isolation buffer in 1 : 1 ratio. Total DNA from single-male individuals was sent to Macrogen Inc. for high-throughput NGS sequencing (NovaSeq Illumina, TruSeq NGS library 2 × 150 bp). Complete mitochondrial sequences have been recovered with NOVOplasty [56] and validated by mapping NGS reads onto the assembled mitogenomes in CLC Genomic Workbench 9.5 (QIAGEN). For gene prediction and annotation MITOCONSTRICTOR [55,57] script with default parameters was used, running and combining results from CRITICA [58], Wise2 [59], GLIMMER [60], ARWEN [61], nhmmer [62] and Phobius [63] software (detailed parameters for dependent programs can be found in MITOCONSTRICTOR pythons scrips). The annotated mitochondrial genomes of *Macoma balthica* can be found in the GenBank database under MH285593 and MH285592 accession numbers. The *cox2* gene expression pattern was checked by mapping NGS RNA-seq reads (SRR5758183; SRR5758184) [64] onto the assembled mitogenome in CLC genomic workbench 9.5.

Based on the cytochrome oxidase subunit II sequence from M-type mitogenome, six pairs of primers spanning 5′-end, mid-section, and 3′-end of the sequence have been designed in Primer3 [65]. Only optimized primer pairs (with high efficiency approx. 100%) used in the study will be shown (electronic supplementary material, table S1). Primers and PCR reaction conditions were optimized on an approximately 6 kbp long PCR M-*cox2* amplicon containing complementary targets for all primers. Composition of the PCR reaction mix for the approximately 6 kb amplicon was as follows: 25 ng DNA, 0.4 U polymerase Phusion (New England Biolabs), 1 × Buffer GC for difficult templates, 0.4 μM for each primer (Lim-F1z and Lim-R3z) and 200 μM dNTP. Conditions for the PCR reaction were as follows: initial denaturation at 98°C for the 30 s then denaturation at 98°C for 10 s, annealing at 66°C for 30 s, elongation for 3 min at 72°C and final elongation 4 min at 72°C. Denaturation, annealing and elongation were repeated in 35 cycles. The products of amplification from three male individuals were then checked by gel electrophoresis and purified with the PCR/ DNA Clean-up kit (EurX). Next, products were digested with the restriction enzymes (BamHI) or sent for Sanger sequencing (Genomed SA). The standard curve for rtPCR was repeated nine times in total and showed a very good correlation (five 10 × dilution points) between 10^3 and 10^7 target gene copies per reaction (determination coefficient $R^2 > 0.99$) (electronic supplementary material, table S2). Parameters for all rtPCR reactions were as follows: reverse transcription at 50°C for 20 min, initial denaturation at 95°C for 10 min, followed by the 35 cycles of denaturation (at 95°C for 30 s), annealing (at 58°C for 30 s) and extension (at 72°C for 30 s). The final concentration of reagents in a single 20 μl rtPCR reaction was as follows: 1 x SG Buffer (SG OneStep RT-PCR kit; EurX), 0.4 μM of primers, 1 μl/reaction SG Enzyme Mix and approximately 20 ng of RNA template. In addition to figure 2, the rtPCR results were put in a tabularized form in electronic supplementary material, tables S3 and S4.

The nucleic acid sequence of the cytochrome oxidase c subunit II transcript was recovered in the following steps. First, the reverse transcription reactions, with optional steps for GC-rich sequences with secondary structures, were carried out according to the protocol supplied by the manufacturer (Maxima Reverse Transcriptase, Thermo Scientific). The reaction solution contained both Random Hexamer primer, Oligo(dT)$_{18}$ primer and Lim-R3z primer. The reaction thermal profile was as follows:
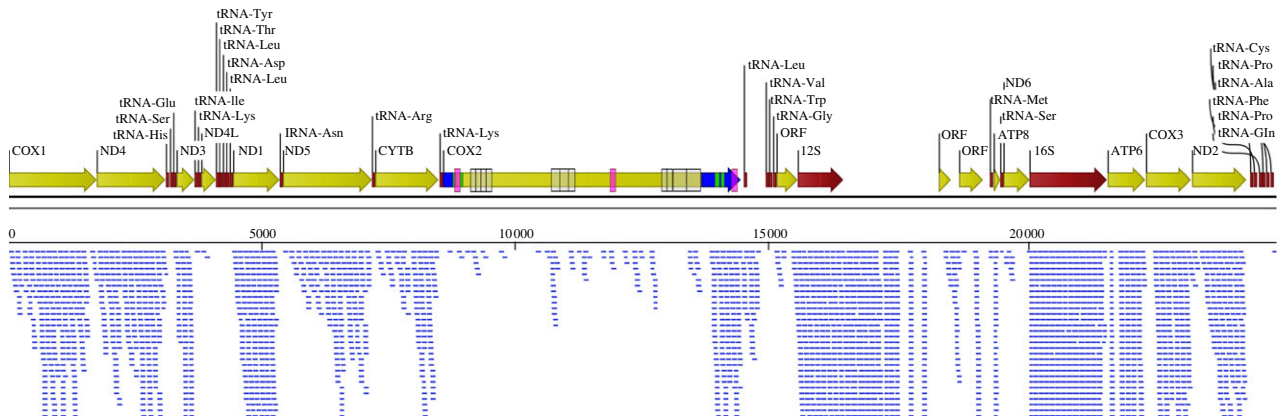
Figure 1. Visualization of the NGS RNA-seq reads coverage (SRR5758183; SRR5758184) over *Macoma balthica* M-type mitogenome. The blue lines in the lower part of the picture represent the coverage of NGS RNA reads, the yellow and red arrows represent mitochondrial genes, the blue arrows represent fragments of the *cox2* RNA transcript, the green boxes represent predicted transmembrane domains, the purple boxes represent sequences amplified during the rtPCR analysis and the transparent boxes represent repetitive sequences.

10 min at 25°C, 15 min at 50°C, 10 min at 65°C, 10 min at 50°C and finally 5 min at 85°C. The first PCR reaction on cDNA template was carried out with similar conditions as for rtPCR standard curve controls (protocol above with Phusion polymerase). The only differences were the set of primers used (Lim-F1z and Lim-R3w), annealing temperature of 60°C and 30 s extension time. The PCR product was then diluted $100\times$ times and another nested PCR with Lim-F1z and Lim_CR22 primers was performed (Lim-F1w primer also worked but was later used in verification correct PCR product). PCR was carried out in a higher volume of 50 µl containing the following concentrations of reagents: 2 µl PCR template, $1\times$ TaqNova buffer, 2 mM $MgCl_2$, 200 µM dNTP, 0.1 µM primers and 1 U polymerase TaqNova (Blirt). Cycling conditions were as follows: initial denaturation at 95°C, followed by 35 cycles of denaturation at 94°C for the 30 s, annealing at 52°C for 20 s, extension at 72°C for 20 s and final extension at 72°C lasted 1 min. After electrophoresis a 500–600 bp long band was cut out from TAE gel, purified with Agarose-Out DNA purification kit (EurX) and sent for Sanger sequencing (Genomed SA). Sequence analysis of mitogenomes and transcript was performed in CLC Genomic Workbench 9.5 (Qiagen) and MEGA7 [66].

3. Results

Based on the NGS DNA sequencing reads, we have assembled two new (F and M types) mitogenomes of *Macoma balthica* clam. During the analysis, we observed a few discrepancies between our and the published (MN528029) M-type mitogenome. The divergence is minimal (3.3% *p*-distance) and is mostly caused by the sequence and length differences in *cox2* gene and non-coding control region. The overall length of our *cox2* gene (5.8 kbp) is different, and there is no stop codon between both domains of the *cox2* gene. What is also interesting is that it seems almost impossible to map the RNA NGS reads (SRR5758183 and SRR5758184 [64]) onto the assembled *cox2* gene (figure 1). Only a few reads map onto the 'inserted' sequence fragment between 5′ and 3′ *cox2* domains. This may suggest that the assembled *cox2* with insertion is indeed an artefact created during assembly or some sort of NUMTs (nuclear mitochondrial DNA sequences) [67,68]. Another explanation why there is a discrepancy in transcript coverage with sequencing reads might be explained by a multiplicity of local secondary RNA structures or excision of the inserted fragment (intron) after transcription. To tackle this question, we took

two steps back from NGS technology, designed primers located in three distinct parts of the cytochrome c oxidase II subunit and performed PCR and rtPCR reactions.

In the beginning, we checked the presence and length of *cox2*-insertion. We performed PCR spanning from the 5′ *cox2* domain to the 3′ *cox2* domain and performed digestion with the restriction enzyme BamHI (only one restriction site within the target gene). The digestion pattern correlated with the mitogenome sequence assembled in this study (two approx. 2833–2834 bp DNA fragments migrating together). In case, if the sequence reported by Capt *et al.* [52] was correct (MN528029) the expected restriction pattern would result in the 2905 bp and 1870 bp products (electronic supplementary material, figure S1). Next, we have sequenced (Sanger sequencing) the 5′-end of the PCR product looking for stop codons in the *cox2* open reading frame. No stop codon was present in the gene sequence (checked on three different individuals). The frameshift in the published sequence might have possibly been caused by the homopolymer of four out of five guanine nucleotides (electronic supplementary material, figure S2).

With the results of the PCR, we have designed and optimized rtPCR reactions for the 5′, mid and 3′ regions trying to prove or disprove the integrity of long *cox2* transcripts. An almost equal number of transcripts for each of the three locations or a decreasing slope from 5′ through 'mid' to 3′ would suggest that this transcript is indeed unusually long, discrepancies in the number of mid-transcript fragments would suggest possible intron excision or (if the PCR showed shorter length) problematic assembly artefacts. Reverse transcription polymerase chain reactions on average resulted in the following relative expression pattern 5′-end = 1, mid = 0.46 and 3′-end = 0.89 (figure 2). This strong decrease in expression between the mid and both 5′ and 3′ ends supported the possibility of mid-sequence excision in an intron-like manner. Finally, to prove the presence of the intron and identify the excision sites of the transcript once and for all, we have attempted to sequence reverse-transcribed *cox2* transcript. The complete *cox2* gene and transcript sequence can be found in the electronic supplementary material, data. After alignment and incorporation of the sequenced transcript fragment into the whole gene, the acquired cytochrome c subunit II gene transcript was 1005 bp long, 234 bp from the 5′-end, and 771 bp from the 3′-end. What is interesting is that the intron splicing site is located
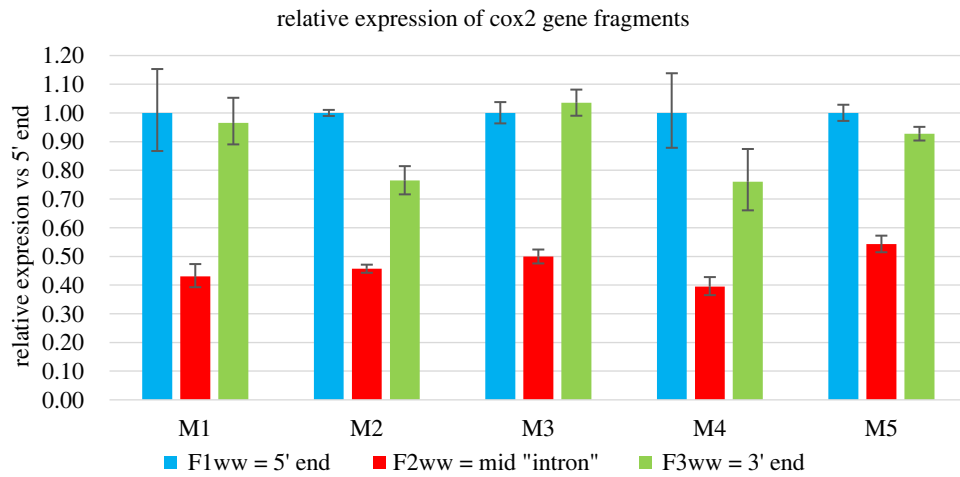
Figure 2. The relative expression of cox2 gene fragments. Data are presented in relation to the 5′-end of transcript ergo, relative expression for this gene to itself is 1.

before one of the quite conserved transmembrane domains present in F-*cox2* (*cox2* gene from female-type mitogenome) as well as in *dna*-M-*cox2* (gene from male-type mitogenome containing intronic sequence). In *rna*-M-*cox2* (*cox2* RNA transcript without the intron sequence), this missing transmembrane domain is then complemented at the 3′-end of the transcript but has low-sequence homology (however keeps structural homology) to the canonical *cox2* sequence. Transcript sequencing also revealed that primers used to amplify the 5′-end of the *cox2* gene were indeed amplifying the 5′-edge of the intron.

Bioinformatic analysis of the *dna*-M-*cox2* and RNA transcript revealed the presence of a canonical intron splicing site at the 5′-end of the sequence (5′-GT) and non-canonical but still possible GG at the 3′-end of the intron (canonical AG-3′). Furthermore, the *cox2* transcript nucleotides close to the intron excision site (5′-ATCATCTTT-ACT-splicing.site and splicing.site-GG-AAAGGTGGT-3′) can create a stable 9 bp long double-stranded structure possibly facilitating connection of the 5′ and 3′ transcripts after intron excision. Unfortunately due to the length (the sequence was too long for most of the available RNA structure prediction tools) and lack of similarity to the introns described in the literature, it is difficult to reliably predict the structure of the cox2 intron, identify the reverse transcriptase coding open reading frame and assign it to any of the existing intron groups. The entire intron is in-frame so in theory any fragment of the sequence could be coding the 'nested' protein limited by flanking secondary RNA structures, but every BLAST search performed by us with nucleotide or translated intron sequence reported no positive hits to sequences in the GenBank database (except itself, the M-type mitogenome from *Macoma balthica*).

4. Discussion

Our results have some implications for the understanding of doubly uniparental inheritance in bivalves. Supernumerary open reading frames and gene extensions are often hypothesized as a key element of the molecular mechanism that stands behind DUI. On the one hand, the intronic *cox2* extension in *Macoma balthica* may not be a part of this system (there is still *m-orf* in the mitogenome), but on the other hand, we do not know if *m-cox2* may exist in more than one isoform. For example, in theory, the short one could be used most of the time in the mitochondrial energy cycle, and the long one could possibly be used during embryo development as a

mitochondrial tagging mechanism. This is highly supported by a recent study on bivalve *Scrobicularia plana* done by Tassé *et al.* [69]. This clam represents a parallel situation to the one observed here with *M. balthica*, a long *cox2* gene with insertions between domains. The authors were able to detect signal from the long RNA transcript (longer than predicted), short targeted approximately 1 kb long transcript fragments containing homologous with typical *cox2* domains and inserted intron-like fragments, and big (Mw. approx. 220 kDa) protein detected with the Western blot technique. Unfortunately, the signal for a smaller 32–37 kDa version of the m-cox2 protein was not detected (personal communication). Furthermore, mitochondrial genes are expressed as polycistronic mRNA, so the presence of transcripts containing and lacking introns after excision is expected (figure 2) [69]. Of course, final proof for the existence of two m-cox2 protein isoforms in *M. balthica* cannot be reliably answered without proper proteomic studies.

There is a hypothesis explaining why the set of protein-coding genes retained in metazoan mitogenomes is limited to the genes encoding core respiratory chain subunits, stipulating that these membrane proteins are difficult to properly transport across membranes [70,71] and hence the selection favoured their *in situ* synthesis [72]. Indeed, the density of transmembrane domains within proteins encoded by the mitogenome is very high. Nevertheless, short insertions within these proteins are apparently tolerated by selection and do occur sometimes. In bivalves, such insertions are present for example in the *atp6* of giant clams [73]. The protein with a relatively small number of transmembrane domains is cox2, typically having only two such domains in a relatively long protein. Therefore, one could expect that insertions within this gene should be tolerated better. In fact there are cases in which the singular cox2 subunit was split into two separate proteins in alveolate and chlorophycean mitogenomes [74], with subsequent migration of some of the resulting genes to the nucleus. It has been also shown that fission of *cox2* occurred in an insect mitogenome of the wasp *Campsomeris plumipes* [75]. In the context of DUI bivalves, cox2 has been postulated to be involved in tagging of M type mitochondria in Unionidea [76]. In this case, the protein has a large extension with multiple additional transmembrane domains, so this interpretation seems plausible. However, there are also other, apparently independently derived, modified forms of this gene in DUI bivalves. A 100 amino acid long insertion is present in cox2 of the M-type *Meretrix lamarckii* [77], but in this case,

no additional transmembrane domains were predicted in the protein. In *Arcuatula senhousia*, the gene is duplicated in M mito-genome, and one of the copies does have an additional transmembrane domain at the C-terminus [78]. In the case of *Geukensia demissa*, the M-type cox2 is singular, but also extended and features one additional transmembrane domain [79].

In conclusion, we have confirmed the first clear case of an intronic sequence in the mitochondrial genome of bivalves and one of the first in mitogenomes of bilaterian animals. This raises the question of how this discovery fits and resonates with the hypothesis that faster evolving mitogenomes do not possess introns, especially because here we have a case of even faster evolving bivalvian M-type mitogenomes.

References

1. Boore JL. 1999 Animal mitochondrial genomes. *Nucleic Acids Res.* **27**, 1767–1780. (doi:10.1093/nar/27.8.1767)

2. Pett W, Ryan JF, Pang K, Mullikin JC, Martindale MQ, Baxevanis AD, Lavrov DV. 2011 Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi*: insights from mtDNA and the nuclear genome. *Mitochond. DNA* **22**, 130–142. (doi:10.3109/19401736.2011.624611)

3. Lavrov DV, Pett W. 2016 Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in nonbilaterian lineages. *Genome Biol. Evol.* **8**, 2896–2913. (doi:10.1093/gbe/evw195)

4. Nickisch-Rosenegk M, Brown WM, Boore JL. 2001 Complete sequence of the mitochondrial genome of the tapeworm *Hymenolepis diminuta*: gene arrangements indicate that platyhelminths are eutrochozoans. *Mol. Biol. Evol.* **18**, 721–730. (doi:10.1093/oxfordjournals.molbev.a003854)

5. Egger B, Bachmann L, Fromm B. 2017 *Atp8* is in the ground pattern of flatworm mitochondrial genomes. *BMC Genomics* **18**, 414. (doi:10.1186/s12864-017-3807-2)

6. Hyman BC. 1988 Nematode mitochondrial DNA: anomalies and applications. *J. Nematol.* **20**, 523–531.

7. Lavrov DV, Brown WM. 2001 *Trichinella spiralis* mtDNA: a nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAs and has a gene arrangement relatable to those of coelomate Metazoans. *Genetics* **157**, 621–637. (doi:10.1093/genetics/157.2.621)

8. Zhang P, Zhou H, Liang D, Liu YF, Chen YQ, Qu LH. 2005 The complete mitochondrial genome of a tree frog, *Polypedates megacephalus* (Amphibia: Anura: Rhacophoridae), and a novel gene organization in living amphibians. *Gene* **346**, 133–143. (doi:10.1016/j.gene.2004.10.012)

9. Rest JS, Ast JC, Austin CC, Waddell PJ, Tibbetts EA, Hay JM, Mindell DP. 2003 Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. *Mol. Phylogenet. Evol.* **29**, 289–297. (doi:10.1016/S1055-7903(03)00108-8)

10. Bilewitch JP, Degnan SM. 2011 A unique horizontal gene transfer event has provided the octocoral mitochondrial genome with an active mismatch repair gene that has potential for an unusual self-contained function. *BMC Evol. Biol.* **11**, 228. (doi:10.1186/1471-2148-11-228)

11. Smith DR, Kayal E, Yanagihara AA, Collins AG, Pirro S, Keeling PJ. 2012 First complete mitochondrial genome sequence from a box jellyfish reveals a highly fragmented linear architecture and insights into telomere evolution. *Genome Biol. Evol.* **4**, 52–58. (doi:10.1093/gbe/evr127)

12. Wang X, Lavrov DV. 2008 Seventeen new complete mtDNA sequences reveal extensive mitochondrial genome evolution within the Demospongiae. *PLoS ONE* **3**, e2723. (doi:10.1371/journal.pone.0002723)

13. Breton S, Beaupré HD, Stewart DT, Piontkivska H, Karmakar M, Bogan AE, Blier PU, Hoeh WR. 2009 Comparative mitochondrial genomics of freshwater mussels (Bivalvia: Unionoida) with doubly uniparental inheritance of mtDNA: gender-specific open reading frames and putative origins of replication. *Genetics* **183**, 1575–1589. (doi:10.1534/genetics.109.110700)

14. Milani L, Ghiselli F, Maurizii MG, Nuzhdin SV, Passamonti M. 2014 Paternally transmitted mitochondria express a new gene of potential viral origin. *Genome Biol. Evol.* **6**, 391–405. (doi:10.1093/gbe/evu021)

15. Guerra D, Plazzi F, Stewart DT, Bogan AE, Hoeh WR, Breton S. 2017 Evolution of sex-dependent mtDNA transmission in freshwater mussels (Bivalvia: Unionida). *Sci. Rep.* **7**, 1551. (doi:10.1038/s41598-017-01708-1)

16. Śmietanka B, Lubośny M, Przyłucka A, Gérard K, Burzyński A. 2018 Mitogenomics of *Perumytilus purpuratus* (Bivalvia: Mytilidae) and its implications for doubly uniparental inheritance of mitochondria. *PeerJ* **6**, e5593. (doi:10.7717/peerj.5593)

17. Ouimet P, Kienzle L, Lubosny M, Burzyński A, Angers A, Breton S. 2020 The ORF in the control region of the female-transmitted *Mytilus* mtDNA codes for a protein. *Gene* **725**, 144161. (doi:10.1016/j.gene.2019.144161)

18. Lubośny M, Przyłucka A, Śmietanka B, Breton S, Burzyński A. 2018 Actively transcribed and expressed *atp8* gene in *Mytilus edulis* mussels. *PeerJ* **6**, e4897. (doi:10.7717/peerj.4897)

19. Smietanka B, Burzyński A, Wenne R. 2010 Comparative genomics of marine mussels (*Mytilus* spp.) gender associated mtDNA: rapidly evolving *atp8*. *J. Mol. Evol.* **71**, 385–400. (doi:10.1007/s00239-010-9393-4)

20. Breton S, Stewart DT, Hoeh WR. 2010 Characterization of a mitochondrial ORF from the gender-associated mtDNAs of *Mytilus* spp. (Bivalvia: Mytilidae): identification of the 'missing' ATPase 8 gene. *Mar. Genomics* **3**, 11–18. (doi:10.1016/j.margen.2010.01.001)

21. Ghiselli F, Gomes-dos-Santos A, Adema CM, Lopes-Lima M, Sharbrough J, Boore JL. 2021 Molluscan mitochondrial genomes break the rules. *Phil. Trans. R. Soc. B* **376**, 20200159. (doi:10.1098/rstb.2020.0159)

22. Beagley CT, Okada NA, Wolstenholme DR. 1996 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. (doi:10.1073/pnas.93.11.5619)

23. Goddard MR, Leigh J, Roger AJ, Pemberton AJ. 2006 Invasion and persistence of a selfish gene in the Cnidaria. *PLoS ONE* **1**, e3. (doi:10.1371/journal.pone.0000003)

24. Medina M, Collins AG, Takaoka TL, Kuehl JV, Boore JL. 2006 Naked corals: skeleton loss in Scleractinia. *Proc. Natl Acad. Sci. USA* **103**, 9096–9100. (doi:10.1073/pnas.0602444103)

25. Fukami H, Chen CA, Chiou CY, Knowlton N. 2007 Novel group I introns encoding a putative homing endonuclease in the mitochondrial *cox1* gene of scleractinian corals. *J. Mol. Evol.* **64**, 591–600. (doi:10.1007/s00239-006-0279-4)

26. Emblem Å, Okkenhaug S, Weiss ES, Denver DR, Karlsen BO, Moum T, Johansen SD. 2014 Sea anemones possess dynamic mitogenome structures. *Mol. Phylogenet. Evol.* **75**, 184–193. (doi:10.1016/j.ympev.2014.02.016)

27. Rot C, Goldfarb I, Ilan M, Huchon D. 2006 Putative cross-kingdom horizontal gene transfer in sponge (Porifera) mitochondria. *BMC Evol. Biol.* **6**, 71. (doi:10.1186/1471-2148-6-71)

28. Gazave E, Lapébie P, Renard E, Vacelet J, Rocher C, Ereskovsky AV, Lavrov DV, Borchiellini C. 2010 Molecular phylogeny restores the supra-generic subdivision of homoscleromorph sponges (Porifera, Homoscleromorpha). *PLoS ONE* **5**, e14290. (doi:10.1371/journal.pone.0014290)

29. Szitenberg A, Rot C, Ilan M, Huchon D. 2010 Diversity of sponge mitochondrial introns revealed by cox1 sequences of Tetillidae. *BMC Evol. Biol.* **10**, 288. (doi:10.1186/1471-2148-10-288)

30. Erpenbeck D, Aryasari R, Hooper JNA, Wörheide G. 2015 A mitochondrial intron in a verongid sponge. *J. Mol. Evol.* **80**, 13–17. (doi:10.1007/s00239-014-9653-9)

31. Huchon D, Szitenberg A, Shefer S, Ilan M, Feldstein T. 2015 Mitochondrial group I and group II introns in the sponge orders Agelasida and Axinellida. *BMC Evol. Biol.* **15**, 278. (doi:10.1186/s12862-015-0556-1)

32. Signorovitch AY, Buss LW, Dellaporta SL. 2007 Comparative genomics of large mitochondria in placozoans. *PLoS Genet.* **3**, e13. (doi:10.1371/journal.pgen.0030013)

33. Burger G, Yan Y, Javadi P, Lang BF. 2009 Group I-intron trans-splicing and mRNA editing in the mitochondria of placozoan animals. *Trends Genet.* **25**, 381–386. (doi:10.1016/j.tig.2009.07.003)

34. Lynch M, Koskella B, Schaack S. 2006 Mutation pressure and the evolution of organelle genomic architecture. *Science* **311**, 1727–1730. (doi:10.1126/science.1118884)

35. Vallès Y, Halanych KM, Boore JL. 2008 Group II introns break new boundaries: presence in a bilaterian's genome. *PLoS ONE* **3**, e1488. (doi:10.1371/journal.pone.0001488)

36. Richter S, Schwarz F, Hering L, Böggemann M, Bleidorn C. 2015 The utility of genome skimming for phylogenomic analyses as demonstrated for glycerid relationships (Annelida, Glyceridae). *Genome Biol. Evol.* **7**, 3443–3462. (doi:10.1093/gbe/evv224)

37. Bernardino AF, Li Y, Smith CR, Halanych KM. 2017 Multiple introns in a deep-sea annelid (Decemunciger: Ampharetidae) mitochondrial genome. *Sci. Rep.* **7**, 4295. (doi:10.1038/s41598-017-04094-w)

38. Kobayashi G, Itoh H, Kojima S. 2022 Mitogenome of a stink worm (Annelida: Travisiidae) includes degenerate group II intron that is also found in five congeneric species. *Sci. Rep.* **12**, 4449. (doi:10.1038/s41598-022-08103-5)

39. Skibinski DO, Gallagher C, Beynon CM. 1994 Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. *Genetics* **138**, 801–809. (doi:10.1093/genetics/138.3.801)

40. Zouros E, Oberhauser Ball A, Saavedra C, Freeman KR. 1994 An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Proc. Natl Acad. Sci. USA* **91**, 7463–7467. (doi:10.1073/pnas.91.16.7463)

41. Passamonti M, Plazzi F. 2020 Doubly uniparental inheritance and beyond: the contribution of the Manila clam *Ruditapes philippinarum*. *J. Zool. Syst. Evol. Res.* **58**, 529–540. (doi:10.1111/jzs.12371)

42. Plazzi F, Passamonti M. 2019 Footprints of unconventional mitochondrial inheritance in bivalve phylogeny: signatures of positive selection on clades with doubly uniparental inheritance. *J. Zool. Syst. Evol. Res.* **57**, 258–271. (doi:10.1111/jzs.12253)

43. Zouros E, Rodakis GC. 2019 Doubly uniparental inheritance of mtDNA: an unappreciated defiance of a general rule. *Adv. Anat. Embryol. Cell Biol.* **231**, 25–49. (doi:10.1007/102_2018_4)

44. Gusman A, Lecomte S, Stewart DT, Passamonti M, Breton S. 2016 Pursuing the quest for better understanding the taxonomic distribution of the system of doubly uniparental inheritance of mtDNA. *PeerJ* **4**, e2760. (doi:10.7717/peerj.2760)

45. Ghiselli F, Milani L, Guerra D, Chang PL, Breton S, Nuzhdin SV, Passamonti M. 2013 Structure, transcription, and variability of metazoan mitochondrial genome: perspectives from an unusual mitochondrial inheritance system. *Genome Biol. Evol.* **5**, 1535–1554. (doi:10.1093/gbe/evt112)

46. Cao L, Kenchington E, Zouros E. 2004 Differential segregation patterns of sperm mitochondria in embryos of the blue mussel (*Mytilus edulis*). *Genetics* **166**, 883–894. (doi:10.1534/genetics.166.2.883)

47. Obata M, Kamiya C, Kawamura K, Komaru A. 2006 Sperm mitochondrial DNA transmission to both male and female offspring in the blue mussel *Mytilus galloprovincialis*. *Dev. Growth Differ.* **48**, 253–261. (doi:10.1111/j.1440-169X.2006.00863.x)

48. Obata M, Sano N, Kawamura K, Komaru A. 2007 Inheritance of two M type mitochondrial DNA from sperm and unfertilized eggs to offspring in *Mytilus galloprovincialis*. *Dev. Growth Differ.* **49**, 335–344. (doi:10.1111/j.1440-169X.2007.00930.x)

49. Kenchington EL, Hamilton L, Cogswell A, Zouros E. 2009 Paternal mtDNA and maleness are co-inherited but not causally linked in mytilid mussels. *PLoS ONE* **4**, e6976. (doi:10.1371/journal.pone.0006976)

50. Saunier A, Garcia P, Becquet V, Marsaud N, Escudié F, Pante E. 2014 Mitochondrial genomes of the Baltic clam *Macoma balthica* (Bivalvia: Tellinidae): setting the stage for studying mito-nuclear incompatibilities. *BMC Evol. Biol.* **14**, 259. (doi:10.1186/s12862-014-0259-z)

51. Becquet V, Simon-Bouhet B, Pante E, Hummel H, Garcia P. 2012 Glacial refugium versus range limit: conservation genetics of *Macoma balthica*, a key species in the Bay of Biscay (France). *J. Exp. Mar. Biol. Ecol.* **432–433**, 73–82. (doi:10.1016/j.jembe.2012.07.008)

52. Capt C, Bouvet K, Guerra D, Robicheau BM, Stewart DT, Pante E, Breton S. 2020 Unorthodox features in two venerid bivalves with doubly uniparental inheritance of mitochondria. *Sci. Rep.* **10**, 1087. (doi:10.1038/s41598-020-57975-y)

53. Kruk-Dowgiałło L, Szaniawska A. 2008 Gulf of Gdańsk and Puck Bay. In *Ecology of baltic coastal waters* (ed. U Schiewer), pp. 139–165. Berlin, Germany: Springer.

54. Hoarau G, Rijnsdorp AD, Veer HWVD, Stam WT, Olsen JL. 2002 Population structure of plaice (*Pleuronectes platessa* L.) in northern Europe: microsatellites revealed large-scale spatial and temporal homogeneity. *Mol. Ecol.* **11**, 1165–1176. (doi:10.1046/j.1365-294X.2002.01515.x)

55. Lubośny M, Przyłucka A, Śmietanka B, Burzyński A. 2020 *Semimytilus algosus*: first known hermaphroditic mussel with doubly uniparental inheritance of mitochondrial DNA. *Sci. Rep.* **10**, 11256. (doi:10.1038/s41598-020-67976-6)

56. Dierckxsens N, Mardulyn P, Smits G. 2017 NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **45**, e18. (doi:10.1093/nar/gkw955)

57. Lubośny M, Śmietanka B, Arculeo M, Burzyński A. 2022 No evidence of DUI in the Mediterranean alien species *Brachidontes pharaonis* (P. Fisher, 1870) despite mitochondrial heteroplasmy. *Sci. Rep.* **12**, 8569. (doi:10.1038/s41598-022-12606-6)

58. Badger JH, Olsen GJ. 1999 CRITICA: coding region identification tool invoking comparative analysis. *Mol. Biol. Evol.* **16**, 512–524. (doi:10.1093/oxfordjournals.molbev.a026133)

59. Birney E, Clamp M, Durbin R. 2004 GeneWise and Genomewise. *Genome Res.* **14**, 988–995. (doi:10.1101/gr.1865504)

60. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999 Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27**, 4636–4641. (doi:10.1093/nar/27.23.4636)

61. Laslett D, Canbäck B. 2008 ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **24**, 172–175. (doi:10.1093/bioinformatics/btm573)

62. Wheeler TJ, Eddy SR. 2013 nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**, 2487–2489. (doi:10.1093/bioinformatics/btt403)

63. Käll L, Krogh A, Sonnhammer ELL. 2007 Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* **35**, W429–W432. (doi:10.1093/nar/gkm256)

64. Yurchenko AA, Katolikova N, Polev D, Shcherbakova I, Strelkov P. 2018 Transcriptome of the bivalve *Limecola balthica* L. from Western Pacific: a new resource for studies of European populations. *Mar. Genomics* **40**, 58–63. (doi:10.1016/j.margen.2018.03.007)

65. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012 Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115. (doi:10.1093/nar/gks596)

66. Kumar S, Stecher G, Tamura K. 2016 MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874. (doi:10.1093/molbev/msw054)

67. Sorenson MD, Quinn TW. 1998 Numts: a challenge for avian systematics and population biology. *Auk* **115**, 214–221. (doi:10.2307/4089130)

68. Marshall C, Parson W. 2021 Interpreting NUMTs in forensic genetics: seeing the forest for the trees. *Forensic Sci. Int.* **53**, 102497. (doi:10.1016/j.fsigen.2021.102497)

69. Tassé M, Choquette T, Angers A, Stewart DT, Pante E, Breton S. 2022 The longest mitochondrial protein

in metazoans is encoded by the male-transmitted mitogenome of the bivalve *Scrobicularia plana*. *Biol. Lett.* **18**, 20220122. (doi:10.1098/rsbl.2022.0122)

70. Popot JL, de Vitry C. 1990 On the microassembly of integral membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* **19**, 369–403. (doi:10.1146/annurev.bb.19.060190.002101)

71. von Heijne G. 1986 Why mitochondria need a genome. *FEBS Lett.* **198**, 1–4. (doi:10.1016/0014-5793(86)81172-3)

72. Björkholm P, Harish A, Hagström E, Ernst AM, Andersson SGE. 2015 Mitochondrial genomes are retained by selective constraints on protein targeting. *Proc. Natl Acad. Sci. USA* **112**, 10 154–10 161. (doi:10.1073/pnas.1421372112)

73. Gan HM, Gan HY, Tan MH, Penny SS, Willan RC, Austin CM. 2016 The complete mitogenome of the giant clam *Tridacna squamosa* (Heterodonta: Bivalvia: Tridacnidae). *Mitochond. DNA A DNA Mapp. Seq. Anal.* **27**, 3220–3221. (doi:10.3109/19401736.2015.1007355)

74. Waller RF, Keeling PJ. 2006 Alveolate and chlorophycean mitochondrial *cox2* genes split twice independently. *Gene* **383**, 33–37. (doi:10.1016/j.gene.2006.07.003)

75. Szafranski P. 2017 Evolutionarily recent, insertional fission of mitochondrial *cox2* into complementary genes in bilaterian Metazoa. *BMC Genomics* **18**, 269. (doi:10.1186/s12864-017-3626-5)

76. Chakrabarti R *et al.* 2007 Reproductive function for a C-terminus extended, male-transmitted cytochrome c oxidase subunit II protein expressed in both spermatozoa and eggs. *FEBS Lett.* **581**, 5213–5219. (doi:10.1016/j.febslet.2007.10.006)

77. Bettinazzi S, Plazzi F, Passamonti M. 2016 The complete female- and male-transmitted mitochondrial genome of *Meretrix lamarckii*. *PLoS ONE* **11**, e0153631. (doi:10.1371/journal.pone.0153631)

78. Passamonti M, Ricci A, Milani L, Ghiselli F. 2011 Mitochondrial genomes and doubly uniparental inheritance: new insights from *Musculista senhousia* sex-linked mitochondrial DNAs (Bivalvia Mytilidae). *BMC Genomics* **12**, 442. (doi:10.1186/1471-2164-12-442)

79. Lubośny M, Śmietanka B, Przyłucka A, Burzyński A. 2020 Highly divergent mitogenomes of *Geukensia demissa* (Bivalvia, Mytilidae) with extreme AT content. *J. Zool. Syst. Evol. Res.* **58**, 571–580. (doi:10.1111/jzs.12354)

80. Lubośny M, Śmietanka B, Lasota R, Burzyński A. 2022 Confirmation of the first intronic sequence in the bivalvian mitochondrial genome of *Macoma balthica* (Linnaeus, 1758). Figshare. (doi:10.6084/m9.figshare.c.6203661)