The longest mitochondrial protein in metazoans is encoded by the male-transmitted mitogenome of the bivalve Scrobicularia plana

Mélanie Tassé, Thierry Choquette, Annie Angers, Donald Stewart, Eric Pante, Sophie Breton

BIOLOGY LETTERS

royalsocietypublishing.org/journal/rsbl

Research

Cite this article: 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. https://doi.org/10.1098/rsbl.2022.0122

Received: 9 March 2022
Accepted: 19 May 2022

Subject Areas:
molecular biology, evolution

Keywords:
mitogenomics, DNA insertion, cytochrome c oxidase subunit 2, doubly uniparental inheritance, Bivalvia

Authors for correspondence:
Eric Pante
e-mail: eric.pante@cnrs.fr
Sophie Breton
e-mail: s.breton@umontreal.ca

†Co-senior authors.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.6013522.

THE ROYAL SOCIETY PUBLISHING

Molecular evolution

The longest mitochondrial protein in metazoans is encoded by the male-transmitted mitogenome of the bivalve *Scrobicularia plana*

Mélanie Tassé[1], Thierry Choquette[1], Annie Angers[1], Donald T. Stewart[2], Eric Pante[3,†] and Sophie Breton[1,†]

[1]Département de sciences biologiques, Université de Montréal, Montréal, QC, Canada
[2]Department of Biology, Acadia University, Wolfville, NS, Canada
[3]Littoral, Environnement et Sociétés (LIENSs), UMR 7266 CNRS–La Rochelle Université, 2 rue Olympe de Gouges, 17000 La Rochelle, France

EP, 0000-0001-7680-2112; SB, 0000-0002-8286-486X

Cytochrome c oxidase subunit II (COX2) is one of the three mitochondrially encoded proteins of the complex IV of the respiratory chain that catalyses the reduction of oxygen to water. The *cox2* gene spans about 690 base pairs in most animal species and produces a protein composed of approximately 230 amino acids. We discovered an extreme departure from this pattern in the male-transmitted mitogenome of the bivalve *Scrobicularia plana* with doubly uniparental inheritance (DUI) of mitochondrial DNA (mtDNA), which possesses an important in-frame insertion of approximately 4.8 kb in its *cox2* gene. This feature—an enlarged male *cox2* gene—is found in many species with DUI; the COX2 protein can be up to 420 amino acids long. Through RT-PCRs, immunoassays and comparative genetics, the evolution and functionality of this insertion in *S. plana* were characterized. The in-frame insertion is conserved among individuals from different populations and bears the signature of purifying selection seemingly indicating maintenance of functionality. Its transcription and translation were confirmed: this gene produces a polypeptide of 1892 amino acids, making it the largest metazoan COX2 protein known to date. We hypothesize that these extreme modifications in the COX2 protein affect the metabolism of mitochondria containing the male-transmitted mtDNA in *Scrobicularia plana*.

1. Introduction

The mitochondrial DNA (mtDNA) of Metazoa is habitually depicted as a small double-stranded circular molecule of about 13–19 kb, with remarkable uniformity in gene content: it contains 37 genes, 13 of which encode protein subunits acting in oxidative phosphorylation (OXPHOS) and ATP production [1,2]. There are, however, exceptions to this typical architecture (e.g. [2,3] and see also [4–6] for newly discovered mtDNA-derived micropeptides). Notably, bivalve molluscs exhibit considerable variation in mtDNA size (less than 14 to greater than 67 kb), and supplementary, atypical protein-coding genes have also been reported [2,7–9]. The only exception to strict maternal inheritance of mtDNA in metazoans, known as doubly uniparental inheritance or DUI, is found in this group and has been reported in over 100 bivalve species [10]. DUI is characterized by two highly divergent (genetic distance greater than 40% in some species [11]) sex-linked mtDNA lineages, one that is maternally inherited (F mtDNA) and present in oocytes and

somatic tissues of both sexes, and the other that is paternally inherited (M mtDNA) and present in sperm, but also sometimes in male soma [12–14].

An apparently convergent evolution of the male-specific version of the *cox2* gene (M*cox2*) represents another mitochondrial curiosity in bivalves [9]. The modifications reported in M*cox2* include long 3′ extensions or in-frame insertions [15–21]. For example, freshwater mussels (order Unionoida) have a fast-evolving 3′ extension of variable length (144–681 bp) in their M*cox2* gene, but still display a pattern of purifying selection in this additional sequence [15–17,19,21]. This extension is transcribed and translated, and its c-terminus tail has been localized at the mitochondrial surface of sperm mitochondria, leading to the hypothesis that it could serve as a tag to facilitate their sex-specific fate in bivalve embryos [17,22,23]. However, more work is needed to understand the functional importance of these modifications in the MCOX2 protein of bivalves with DUI.

The longest in-frame insertion (4827 bp) in the M*cox2* gene of a DUI bivalve has recently been reported in *Scrobicularia plana* (order Cardiida) [20], meaning that this gene could be expressed as a single polypeptide drastically larger (1892 amino acids) than the typical one (approx. 230 aa) or the enlarged one found in some male mitogenomes of other DUI species (up to 420 aa) [15–17,19–21]. However, this possibility has not been evaluated yet. Here, we used a combination of RT-PCR, immunoassays and comparative genetics to better characterize the evolution and functionality of this newly discovered M*cox2* insertion in *S. plana* and confirmed the existence of the largest mtDNA-encoded protein reported in a metazoan so far.

2. Material and methods

Adult specimens of *Scrobicularia plana* were collected in May 2013 from Concarneau (France; 47.8728° N, 3.9207° W), and June 2018 and 2019 from Fouras-les-Bains (France; 45.9838° N, 1.0931° W). Individuals were dissected by cutting adductor muscles and sampling a fragment of the mantle. Gonads were gently nicked with a scalpel and gametes were sampled with a P200 pipette and inspected under the microscope for sexing. Female and male gametes, mature gonads and somatic tissues were preserved in 95% ethanol or flash-frozen in liquid nitrogen and sent to the Université de Montréal on dry ice to be stored at −80°C. Total genomic DNA was extracted from gonads and somatic tissues with the Qiagen DNeasy Blood & Tissue Kit (QIAGEN 69506). RNA was extracted from gametes and somatic tissues with the *Quick*-RNA™ Miniprep kit (Zymo Research R1054) and the Nucleospin RNA for NucleoZOL kit (Macherey-Nagel 740406.50 and 740404.200). RNA was retrotranscribed into cDNA using the GoScript™ Reverse Transcription Mix (Promega A2790) and the Superscript IV 1-Step System (Invitrogen 15895891). The quality and quantity of DNA, RNA and cDNA were assessed by electrophoresis on a 1% agarose gel and/or with a BioDrop µLITE spectrophotometer. Proteins were extracted from gametes and somatic tissues by mechanical grinding and sonication in two volumes of RIPA buffer (Tris-HCL 50 mM, NaCl 150 mM, EDTA 1 mM, Triton X-100 1%, sodium deoxycholate 0.5%, sodium dodecyl sulfate [SDS] 0.1%, pH 7.6). Protein concentration was determined by Bradford protein assay kit (Sigma B6916).

Two regions of M*cox2* (R1 and R2) and one region of F*cox2* (F mtDNA) were PCR-amplified using DNA from 23 males and 18 females (figure 1*a*; electronic supplementary material, table S1 and datasets S1–S3; GenBank accession number OM928010-064). PCR products were purified with the QIAquick® Gel Extraction Kit (QIAGEN 28706) and sent to Génome Québec for Sanger sequencing in both directions. Nucleotide and amino acid sequences were aligned by MAFFT v. 7 with L-INS-I method [24]. MEGA-X was used to compute amino acid *p*-distances and rates of synonymous (K_S) and non-synonymous substitutions (K_A) using the Nei Gojobori method with the Jukes–Cantor correction [25]. Secondary structures of FCOX2 and MCOX2 were predicted using Protter [26]. Transmembrane alpha-helices (TMHs) were predicted using TMHMM [27], TMpred [28], Phobius [29], TOPCONS [30], InterProScan [31] and Phyre2 [32,33]. Protein three-dimensional structures were modelled using Phyre2 [33]. Function prediction was performed using BLASTp and BLASTn searches against (i) the NCBI entire non-redundant protein sequences (nr) and nucleotide sequences (nr/nt) databases, (ii) the Reference sequence (RefSeq) database, and (iii) the UniProtKB/Swiss-Prot database all with default parameters [34]. Open reading frames (ORFs) localized in the insertion with a length greater than 150 nucleotides and ATG or alternative initiation codons as start codons were found with NCBI ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/, last accessed July 2021) using the invertebrate mitochondrial genetic code. Function prediction was performed for these ORFs using BLASTp against all the databases mentioned above [34]. All matches with *e*-values less than 0.001 were kept.

To check for expression, F*cox2* and two regions of M*cox2* (R1 and R3) were PCR-amplified from gamete and somatic cDNA (two-step RT-PCR) (figure 1*a*; electronic supplementary material, table S1). Male gDNA (gamete DNA) was used as a positive control, and non-treated RNA and molecular grade water were used as negative controls. The same three regions were also amplified using RNA extracted from somatic tissues with the SuperScript™ IV One-Step RT-PCR System (Invitrogen 12594025) (electronic supplementary material, table S1). Male gDNA and molecular grade water were used as negative controls. PCR and RT-PCR products were visualized on a 1% agarose gel.

To check for the translation of the full M*cox2* gene, a polyclonal antibody against the synthetic epitope CDKYKVFPHWE specific to MCOX2 was produced in rabbits and affinity-purified (MediMabs, Montréal, Canada). Protein extracts from male and female gametes and somatic tissues were supplemented with loading buffer (50 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 1% b-mercaptoethanol and 0.02% bromophenol blue) then heated for 5 min at 95°C before being loaded into a denaturing 4–20% polyacrylamide gel. Samples migrated for 18 h at 50 V in migration buffer (Tris 25 mM, glycine 192 mM and SDS 0.1%) and were transferred on nitrocellulose membranes at 1000 mA for 1 h 30 min in transfer buffer (glycine 1.5%, Tris base 0.3%, methanol 20%). Membranes were coloured with Ponceau S and then decoloured and blocked in a 1X phosphate-buffered saline Tween-20 (PBST; sodium phosphate monobasic (NaH_2PO_4) 57 mM, NaCl 154 mM, 0.05% Tween-20 and pH 8.0) with 5% BLOTTO Blocker. Membranes were incubated overnight at 4°C either with primary antibodies directed against MCOX2 (1 : 2000–1 : 20 000 in PBST) or with antibodies Anti-ATP5A (mitochondrial marker, Abcam ab14748) as positive control (1 : 2000 in PBST). Membranes were washed 3 × 15 min in PBST and then incubated for 2 h at room temperature with secondary antibodies (1 : 2000 in PBST) goat anti-rabbit-peroxidase (for membranes exposed to Anti-MCOX2 – Jackson Immuno 111-035-144) and goat anti-mouse-peroxidase (for membranes exposed to Anti-ATP5A—Jackson Immuno 115-035-003). Membranes were again washed 3 × 15 min in PBST and then washed 5 min in Tris-buffered saline (Tris-HCl 20 mM, NaCl 150 mM, pH 7.4). Finally, membranes were revealed using MBI evolution Borealis plus western blot detection (MBI BORA200ML).

3. Results

Figure 1*a* shows the regions of M*cox2* (R1–R3) and F*cox2* (F mtDNA) used to compute amino acid *p*-distances and rates

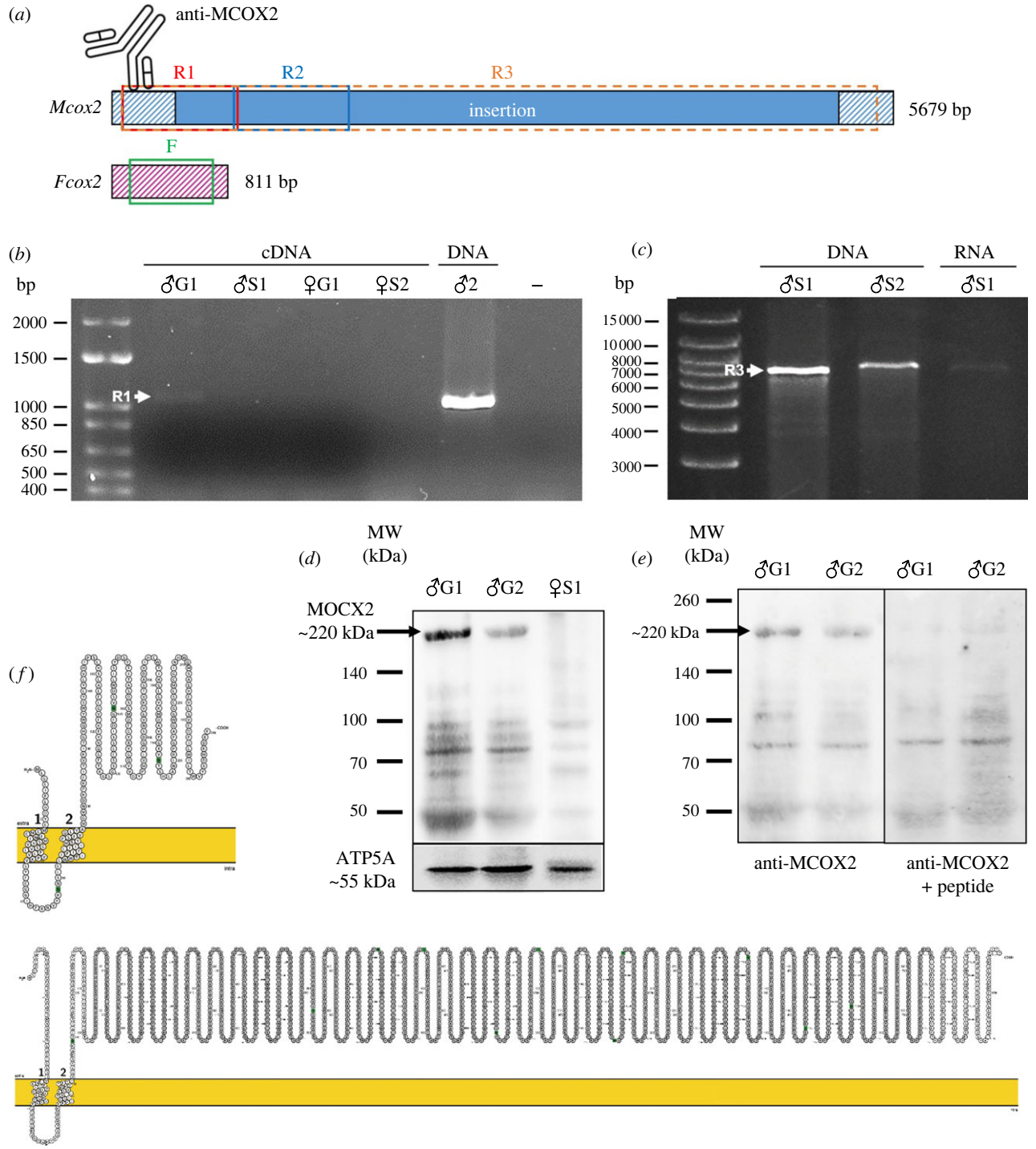

Figure 1. Transcription and translation of the M*cox2* insertion. (*a*) Representation of the M (blue) and F (pink) *cox2* genes and the regions targeted by PCR and RT-PCR (R1–3: red, blue and orange boxes, respectively; F*cox2*: green box). The M/F homologous region is hatched and the M*cox2* insertion is plain. The position of the epitope specific to MCOX2 is shown by an antibody. (*b*) R1 was amplified from male and female gametes (G) and somatic tissues (S) by two-step RT-PCR. (*c*) R3 was amplified from male somatic DNA by PCR and RNA by one-step RT-PCR. Amplified fragments are marked by white arrows. (*d*) Western blots detection of MCOX2 and ATP5A in sperm and female somatic tissues. (*e*) MCOX2 expression in sperm without (left) and with (right) prior incubation of antibodies with the synthetic antigenic peptide. (*f*) Predicted membrane topology model generated by Protter of FCOX2 (above) and MCOX2 (below). The inner mitochondrial membrane is represented in yellow. Amino acids from the MCOX2 insertion are in grey. Transmembrane regions are identified by 1 and 2.

of synonymous (K_S) and non-synonymous substitutions (K_A) (i.e. R1, R2 and F*cox2*), and to check for expression (i.e. R1 and R3). In the latter case, a fragment of approximately 950 bp was expected to be amplified by RT-PCR for the region R1 if the insertion is indeed transcribed (figure 1*a*). The expected signal was observed for sperm (figure 1*b*) and male somatic tissues (electronic supplementary material, figure S1A), a result consistent with the presence of M mtRNA in these tissues as has been previously noted in other species [13,35]. For the region R3 (figure 1*a*), a fragment of approximately 5.6 kb was expected to be amplified (with the insertion fully transcribed). PCR amplification on sperm cDNA failed to detect such a fragment (electronic supplementary material, figure S1B), but one-step RT-PCR on male somatic RNA samples showed a faint band of approximately 7 kb (figure 1*c*). It is possible that very low levels of mitochondrial RNA and retrotranscribed mtRNA were present in sperm (this added to the difficulty of retro-transcribing very long RNA molecules). The approximately 1.5 kb difference between the expected length of R3 and the

Table 1. Intraspecific amino acid divergences between and within FCOX2 and MCOX2. *Note*. The pairwise deletion option computes a distance for each pair of sequences and ignores gaps involved in the comparison. The complete deletion option does not compute any sites with gaps or missing information in the comparison. R1homo refers to the homologous portion of R1 with F*cox2*; R1ins refers to the insertion portion of R1.

			pairwise deletion		complete deletion	
	sample size (*n*)	sequence length (aa)	mean number of amino acid differences (aa)	*p*-distance	mean number of amino acid differences (aa)	*p*-distance
FCOX2 vs MCOX2	36	238	34.940 ± 5.531	50.390% ± 6.123%	32.441 ± 4.013	50.689% ± 5.836%
F*cox2*	18	176	0.737 ± 0.266	0.419% ± 0.152%	0.737 ± 0.272	0.419% ± 0.159%
R1homo	17	90	1.471 ± 0.464	1.634% ± 0.535%	1.471 ± 0.486	1.634% ± 0.532%
R1ins	17	169	4.029 ± 0.851	2.537% ± 0.526%	3.081 ± 0.752	2.233% ± 0.540%
R2	21	278	6.295 ± 1.081	2.491% ± 0.452%	5.229 ± 1.065	2.551% ± 0.530%

Table 2. Synonymous (K_S) and non-synonymous (K_A) substitution rates of F*cox2* and M*cox2* gene regions. *Note*. R1homo and R1ins refer to the homologous portion of R1 with F*cox2* and to the insertion portion of R1, respectively.

	K_A	K_S	K_A/K_S
F*cox2*	0.001829 ± 0.000676	0.020877 ± 0.005271	0.087608 ± 0.054499
R1homo	0.007939 ± 0.002405	0.022851 ± 0.009657	0.347425 ± 0.252072
R1ins	0.012346 ± 0.002857	0.040441 ± 0.009866	0.305284 ± 0.145123
R2	0.011443 ± 0.002030	0.042749 ± 0.007393	0.267679 ± 0.093779

visualized length might be due to insertions/deletions (indels) found within the M*cox2* insertion (see below). Strong bands of expected size were obtained for F*cox2* (electronic supplementary material, figure S1C, D).

In support to RT-PCR results, western blots revealed a band of approximately 220 kDa only in sperm (figure 1*d*). This is slightly heavier than the predicted mass of the protein (204 kDa) (ExPASy server; [36]) but could be attributed to the presence of indels as observed in our sequences, and/or to the presence of TMHs or post-translational modifications such as glycosylation or ubiquitination [37]. To verify the specificity of the anti-MCOX2 antibody, the synthetic antigenic peptide was added to the primary antibody solution at a 10 X concentration before the incubation to competitively chelate every antigenic site of the primary antibody and to show the attenuation of the specific MCOX2 band (figure 1*e*). This result confirmed that the in-frame insertion in the M*cox2* gene is translated.

FCOX2 and MCOX2 amino acid sequences differed by 50% (table 1). Genetic distances were low within sexes (less than 2.6%), the lowest being among FCOX2 and the highest among MCOX2 insertions (table 1). K_A/K_S ratios were all indicative of purifying selection (less than 1 for all regions; table 2).

In metazoans, the N-terminal domain of the cytochrome c oxidase subunit II (COX2) protein typically contains two TMHs, which are followed by a 'heme-patch' region and two Cu_a-binding centres [20,38]. The insertion in the protein MCOX2 is localized between the 'heme-patch' region and the first Cu_a-binding centre [20]. Secondary structure and TMH predictions suggested that MCOX2 possesses the two conserved TMH located near the N-terminus end (figure 1*f*; electronic supplementary material, table S2). The three-dimensional model of MCOX2 predicted by Phyre2 had a high precision value (99.6%) but a small coverage (6%) including only a part of the F/M homologous C-terminal region (electronic supplementary material, figure S2).

BLASTn searches using only the M*cox2* insertion revealed hits with part of European starfish *Asterias rubens* chromosome 14 (3% coverage, approx. 74% similarity, *e*-value 2×10^{-12}) and with a translation initiation factor IF-2 in sorghum *Sorghum bicolor* (2% coverage, approx. 69% similarity, *e*-value 7×10^{-4}) (electronic supplementary material, table S3). BLASTp searches using only the MCOX2 insertion revealed hits (most of bacterial origin) with a small query coverage (less than or equal to 5%) and percentage identity of less than 52% (electronic supplementary material, table S3). A conserved domain was identified, i.e. PHA03247 large tegument protein UL36 (*e*-value 3.06×10^{-4}). Also, out of the 27 ORFs greater than 150 nt found in the insertion, there were three with significant similarity to at least one sequence (electronic supplementary material, table S3). One ORF, covering almost the entire insertion, had identical hits to the one from the complete insertion. Two other ORFs with a length of approximately 100 aa and localized on the reverse strand returned hits referring to a myosin light chain kinase (61% coverage, approx. 48% similarity, *e*-value 5×10^{-4}) and a phosphoesterase (53% coverage, approx. 45% similarity, *e*-value 7×10^{-6}) (electronic supplementary material, table S3).

4. Discussion

Herein, we report the longest mitochondrially encoded protein in the animal kingdom (1892 amino acids). The important in-frame insertion of approximately 4.8 kb previously reported in the male *cox2* gene of the DUI bivalve

Scrobicularia plana [20] results in a significant enlargement of the protein: (i) the insertion does not contain any premature stop codon, it is conserved among individuals from different populations (over 300 km apart) and it bears the signature of purifying selection, which can be interpreted as a way to maintain a functional protein and (ii) the insertion is transcribed and a protein of approximately 220 kDa was detected in sperm by western blot analysis, which corresponds to the expected size of the MCOX2 peptide with the insertion. We believe that our observations are not the result of nuclear mtDNA segments (NUMTs; [39]) because (i) the complete male mitogenome of *S. plana* has been assembled with MITObim [20], a program expected to perform well even for species with high number of NUMTs [40], (ii) NUMTs greater than 4 kb are usually rare in animal genomes, and they are mostly non-functional pseudo-genes and rarely expressed [41–43], (iii) we were able to amplify large portions of M*cox2* both at the DNA and RNA levels, and (iv) the M*cox2* gene cannot be completely translated using the universal genetic code (in particular, there is no possible ORF > 50 aa using the universal genetic code for the region containing the epitope CDKYKVFPHWE specific to MCOX2 used to generate our antibody).

To our knowledge, the only other animal species with an important in-frame insertion (greater than 3 kb) in its *cox2* gene is the hymenopteran *Campsomeris* [44]. However, contrary to the situation in *S. plana*, *cox2* in *Campsomeris* is split into two genes encoding two complementary polypeptides, COX2A and COX2B, which apparently form a heterodimer [44]. This insertion between *cox2a* and *cox2b* possesses several antiparallel overlapping ORFs, one of which encodes a nuclease potentially involved in *cox2* fission [44]. Such ORFs were not found in *S. plana*, but a conserved domain was identified in the insertion, i.e. PHA03247 large tegument protein UL36, which is conserved among herpesviruses and plays many roles in viral replication [45]. Interestingly, this result is in line with a previous hypothesis that viral selfish elements may have colonized the male mitogenome in bivalves promoting its segregation into primordial germ cells and allowing its transmission to next generations [8,46]. It is also worth mentioning that herpesviruses were reported in many bivalve species [47,48].

As suggested by Curole & Kocher [16], it is tempting to speculate that extensions and insertions in M*cox2* of bivalves with DUI are a result of the new male-specific selective environment. According to previous [15–17,20] and present analyses, the two TMHs, the 'heme-patch' region and the two Cu_a-binding centres are conserved in MCOX2 of DUI bivalves, suggesting that its role in OXPHOS is conserved. In the case of *S. plana*, secondary structure and TMH predictions suggested that the insertion is located in the mitochondrial intramembrane space, but additional putative TMHs predicted by some programs (see electronic supplementary material, table S2) could allow the protein to be exposed at the mitochondrial surface. This was reported for the MCOX2 C-terminal extension of freshwater mussels, and it is the reason why the extension has been hypothesized to act as a tag for paternal mitochondria [17,23]. However, this remains to be demonstrated.

We suggest that the apparently convergent evolution of the M*cox2* gene might play a part in the reorganization of sperm energy metabolism in bivalves with DUI [49–51]. Specifically, eggs and soma of DUI species, usually homoplasmic for the F mtDNA, express a common mitochondrial 'F-phenotype', whereas sperm and their M-type mitochondria express an 'M-phenotype', which is characterized by low OXPHOS rates and an almost null spare capacity of the cytochrome c oxidase complex [49,51]. This sperm-specific, potentially male-inducing mitochondrial phenotype in DUI species [52] could be due, at least in part, to modifications in the M*cox2* gene. Physiological comparisons of several DUI taxa with and without a functional M*cox2* extension or insertion will provide an opportunity to test this hypothesis. Such studies will also help to better understand if these important modifications in the M*cox2* gene of DUI species may represent a male-specific energetic adaptation to enhance fertilization success.

Data accessibility. All data are available as electronic supplementary material [53].

Authors' contributions. M.T.: methodology, writing—original draft and writing—review and editing; T.C.: methodology and writing—review and editing; A.A.: methodology, resources, writing—review and editing; D.T.S.: resources and writing—review and editing; E.P.: conceptualization, data curation, funding acquisition, methodology, supervision, writing—original draft and writing—review and editing; S.B.: conceptualization, funding acquisition, methodology, supervision, writing—original draft and writing—review and editing.

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

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

Funding. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) (RGPIN-2019–04076 to S.B.) and the Agence Nationale de la Recherche (ANR) (ANR-18-CE02-0004-01 to E.P.). S.B. holds the Canada Research Chair (Tier 2) in Mitochondrial Evolutionary Biology.

References

1. Boore JL. 1999 Animal mitochondrial genomes. *Nucl. Acids Res.* **27**, 1767–1780. (doi:10.1093/nar/27.8.1767)
2. Breton S, Milani L, Ghiselli F, Guerra D, Stewart DT, Passamonti M. 2014 A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. *Trends Genet.* **30**, 555–564. (doi:10.1016/j.tig.2014.09.002)
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. Miller B, Kim SJ, Kumagai H, Mehta HH, Xiang W, Liu J, Yen K, Cohen P. 2020 Peptides derived from small mitochondrial open reading frames: genomic, biological, and therapeutic implications. *Exp. Cell Res.* **393**, 112056. (doi:10.1016/j.yexcr.2020.112056)
5. Breton S, Ghiselli F, Milani L. 2021 Mitochondrial short-term plastic responses and long-term evolutionary dynamics in animal species. *Genome Biol. Evol.* **13**, evab084. (doi:10.1093/gbe/evab084)
6. Emser SV, Schaschl H, Millesi E, Steinborn R. 2021 Extension of mitogenome enrichment based on single long-range PCR: mtDNAs and putative mitochondrial-derived peptides of five rodent hibernators. *Front. Genet.* **12**, 685806. (doi:10.3389/fgene.2021.685806)
7. Breton S, Stewart DT, Shepardson S, Trdan RJ, Bogan AE, Chapman EG, Ruminas AJ, Piontkivska H, Hoeh WR. 2011 Novel protein genes in animal mtDNA: a new sex determination system in freshwater mussels (Bivalvia: Unionoida)? *Mol. Biol. Evol.* **28**, 1645–1659. (doi:10.1093/molbev/msq345)

8. Milani L, Ghiselli F, Guerra D, Breton S, Passamonti M. 2013 A comparative analysis of mitochondrial ORFans: new clues on their origin and role in species with doubly uniparental inheritance of mitochondria. *Genome Biol. Evol.* **5**, 1408–1434. (doi:10.1093/gbe/evt101)
9. 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)
10. 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)
11. Doucet-Beaupré H, Breton S, Chapman EG, Blier PU, Bogan AE, Stewart DT, Hoeh WR. 2010 Mitochondrial phylogenomics of the Bivalvia (Mollusca): searching for the origin and mitogenomic correlates of doubly uniparental inheritance of mtDNA. *BMC Evol. Biol.* **10**, 50. (doi:10.1186/1471-2148-10-50)
12. Mioduchowska M, Kaczmarczyk A, Zając K, Zając T, Sell J. 2016 Gender-associated mitochondrial DNA heteroplasmy in somatic tissues of the endangered freshwater mussel *Unio crassus* (Bivalvia: Unionidae): implications for sex identification and phylogeographical studies. *J. Exp. Zool. A Ecol. Genet. Physiol.* **325**, 610–625. (doi:10.1002/jez.2055)
13. Breton S, Bouvet K, Auclair G, Ghazal S, Sietman BE, Johnson N, Bettinazzi S, Stewart DT, Guerra D. 2017 The extremely divergent maternally- and paternally-transmitted mitochondrial genomes are co-expressed in somatic tissues of two freshwater mussel species with doubly uniparental inheritance of mtDNA. *PLoS ONE* **12**, e0183529. (doi:10.1371/journal.pone.0183529)
14. Iannello M, Bettinazzi S, Breton S, Ghiselli F, Milani L. 2021 A naturally heteroplasmic clam provides clues about the effects of genetic bottleneck on paternal mtDNA. *Genome Biol. Evol.* **13**, evab022. (doi:10.1093/gbe/evab022)
15. Curole JP, Kocher TD. 2002 Ancient sex-specific extension of the cytochrome c oxidase II gene in bivalves and the fidelity of doubly-uniparental inheritance. *Mol. Biol. Evol.* **19**, 1323–1328. (doi:10.1093/oxfordjournals.molbev.a004193)
16. Curole JP, Kocher TD. 2005 Evolution of a unique mitotype-specific protein-coding extension of the cytochrome c oxidase II gene in freshwater mussels (Bivalvia: Unionoida). *J. Mol. Evol.* **61**, 381–389. (doi:10.1007/s00239-004-0192-7)
17. Chapman EG, Piontkivska H, Walker JM, Stewart DT, Curole JP, Hoeh WR. 2008 Extreme primary and secondary protein structure variability in the chimeric male-transmitted cytochrome coxidase subunit II protein in freshwater mussels: evidence for an elevated amino acid substitution rate in the face of domain-specific purifying selection. *BMC Evol. Biol.* **8**, 165. (doi:10.1186/1471-2148-8-165)
18. 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)
19. 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)
20. 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)
21. Soroka M. 2020 Doubly uniparental inheritance of mitochondrial DNA in freshwater mussels: history and status of the European species. *J. Zool. Syst. Evol. Res.* **58**, 598–614. (doi:10.1111/jzs.12381)
22. Chakrabarti R, Walker JM, Stewart DT, Trdan RJ, Vijayaraghavan S, Curole JP, Hoeh WR. 2006 Presence of a unique male-specific extension of C-terminus to the cytochrome c oxidase subunit II protein coded by the male-transmitted mitochondrial genome of *Venustaconcha ellipsiformis* (Bivalvia: Unionoidea). *FEBS Lett.* **580**, 862–866. (doi:10.1016/j.febslet.2005.12.104)
23. 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)
24. Katoh K, Kuma K, Toh H, Miyata T. 2005 MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucl. Acids Res.* **33**, 511–518. (doi:10.1093/nar/gki198)
25. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018 MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549. (doi:10.1093/molbev/msy096)
26. Omasits U, Ahrens CH, Müller S, Wollscheid B. 2014 Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics* **30**, 884–886. (doi:10.1093/bioinformatics/btt607)
27. Krogh A, Larsson B, Heijne G, Sonnhammer ELL. 2001 Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580. (doi:10.1006/jmbi.2000.4315)
28. Hofmann K, Stoffel W. 1993 TMbase – a database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **374**, 166.
29. Käll L, Krogh A, Sonnhammer ELL. 2004 A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036. (doi:10.1016/j.jmb.2004.03.016)
30. Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. 2015 The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucl. Acids Res.* **43**, W401–W407. (doi:10.1093/nar/gkv485)
31. Zdobnov EM, Apweiler R. 2001 InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848. (doi:10.1093/bioinformatics/17.9.847)
32. Nugent T, Jones DT. 2009 Transmembrane protein topology prediction using support vector machines. *BMC Bioinf.* **10**, 159. (doi:10.1186/1471-2105-10-159)
33. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015 The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858. (doi:10.1038/nprot.2015.053)
34. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402. (doi:10.1093/nar/25.17.3389)
35. Zouros E. 2013 Biparental inheritance through uniparental transmission: the doubly uniparental inheritance (DUI) of mitochondrial DNA. *Evol. Biol.* **40**, 1–31. (doi:10.1007/s11692-012-9195-2)
36. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. 2005 Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (ed. JM Walker), pp. 571–607. Totowa, NJ: Humana Press.
37. 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)
38. Rawson PD, Burton RS. 2006 Molecular evolution at the cytochrome oxidase subunit 2 gene among divergent populations of the intertidal copepod, *Tigriopus californicus*. *J. Mol. Evol.* **62**, 753–764. (doi:10.1007/s00239-005-0074-7)
39. Richly E, Leister D. 2004 NUMTs in sequenced eukaryotic genomes. *Mol. Biol. Evol.* **21**, 1081–1084. (doi:10.1093/molbev/msh110)
40. Hahn C, Bachmann L, Chevreux B. 2013 Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucl. Acids Res.* **41**, e129. (doi:10.1093/nar/gkt371)
41. Song S, Jiang F, Yuan J, Guo W, Miao Y. 2021 Exceptionally high cumulative percentage of NUMTs originating from linear mitochondrial DNA molecules in the *Hydra magnipapillata* genome. *BMC Genomics* **14**, 447. (doi:10.1186/1471-2164-14-447)
42. Ding L, Sang H, Sun C. 2021 Genus-wide characterization of nuclear mitochondrial DNAs in bumblebee (Hymenoptera: Apidae) genomes. *Insects* **12**, 963. (doi:10.3390/insects12110963)
43. Marshall C, Parson W. 2021 Interpreting NUMTs in forensic genetics: seeing the forest for the trees. *Forensic Sci. Int. Genet.* **53**, 102497. (doi:10.1016/j.fsigen.2021.102497)
44. 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)

45. Chadha P, Sarfo A, Zhang D, Abraham T, Carmichael J, Han J, Wills JW. 2017 Domain interaction studies of herpes simplex virus 1 tegument protein UL16 reveal its interaction with mitochondria. *J. Virol.* **91**, e01995-16. (doi:10.1128/JVI.01995-16)
46. Milani L, Ghiselli F, Passamonti M. 2016 Mitochondrial selfish elements and the evolution of biological novelties. *Curr. Zool.* **62**, 687–697. (doi:10.1093/cz/zow044)
47. Renault T, Novoa B. 2004 Viruses infecting bivalve molluscs. *Aquat. Living Resour.* **17**, 397–409. (doi:10.1051/alr:2004049)
48. Davison AJ, Trus BL, Cheng N, Steven AC, Watson MS, Cunningham C, Deuff RML, Renault T. 2005 A novel class of herpesvirus with bivalve hosts. *J. Gen. Virol.* **86**, 41–53. (doi:10.1099/vir.0.80382-0)
49. Bettinazzi S, Rodríguez E, Milani L, Blier PU, Breton S. 2019 Metabolic remodelling associated with mtDNA: insights into the adaptive value of doubly uniparental inheritance of mitochondria. *Proc. R. Soc. B* **286**, 20182708. (doi:10.1098/rspb.2018.2708)
50. Bettinazzi S, Nadarajah S, Dalpé A, Milani L, Blier PU, Breton S. 2020 Linking paternally inherited mtDNA variants and sperm performance. *Proc. R. Soc. B* **375**, 20190177. (doi:10.1098/rstb.2019.0177)
51. Bettinazzi S, Milani L, Blier PU, Breton S. 2021 Bioenergetic consequences of sex-specific mitochondrial DNA evolution. *Proc. R. Soc. B* **288**, 20211585. (doi:10.1098/rspb.2021.1585)
52. Breton S, Stewart DT, Brémaud J, Havird JC, Smith CH, Hoeh WR. 2022 Did doubly uniparental inheritance (DUI) of mtDNA originate as a cytoplasmic male sterility (CMS) system? *Bioessays* **44**, 2100283. (doi:10.1002/bies.202100283)
53. 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*. Figshare. (doi:10.6084/m9.figshare.c.6013522)