AUTHOR COVER PAGE

The Complete Mitochondrial Genome of the Grooved Carpet Shell, *Ruditapes decussatus* (Bivalvia, Veneridae).

Fabrizio Ghiselli[1,*,º], Liliana Milani[1,º], Mariangela Iannello[1], Emanuele Procopio[1], Peter L. Chang[2], Sergey V. Nuzhdin[2], and Marco Passamonti[1]

[1] Department of Biological, Geological and Environmental Sciences - University of Bologna, Italy.

[2] Program in Molecular and Computational Biology, Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA.

º Equal contribution

*Corresponding Author: Fabrizio Ghiselli fabrizio.ghiselli@unibo.it

Abstract

Despite the~~a~~ large number of animal complete mitochondrial genomes ~~being~~ currently available in public databases, ~~the current~~ knowledge about mitochondrial genomics in invertebrates is ~~sketchy~~uneven. This paper reports, for the first time, the complete mitochondrial genome of the grooved carpet shell, *Ruditapes decussatus*, also known as the European clam. *R. decussatus* is morphologically and ecologically similar to the Manila clam *Ruditapes philippinarum*, which has been recently introduced for aquaculture in the very same habitats of *R. decussatus*, and that is replacing the native species. Currently the production of the European clam is almost insignificant, nonetheless it is considered a high value product, and therefore it is an economically important species, especially in Portugal, Spain and Italy.

In this work we: *i)* assembled *R. decussatus* mitochondrial genome from RNA-Seq data, and validated it by Sanger sequencing, *ii)* analyzed and characterized the *R. decussatus* mitochondrial genome, comparing its features with those of other venerid bivalves; *iii)* assessed mitochondrial genetic polymorphism, both from a nucleotide (SNPs) and a structural (CNV of tandem repeats) point of view, across 26 samples.

Despite using high-throughput approaches we did not find evidence of sex-linked mitochondrial genomes, so it seems that *R. decussatus* does not have Doubly Uniparental Inheritance of mitochondria, a phenomenon known in ~100 bivalve species. According to our analyses, *R. decussatus* is more genetically similar to species of the Genus Paphia than to the congeneric *R. philippinarum*, a finding that bolsters the ~~already~~ already-proposed need of a taxonomic revision.

We also found a quite low genetic variability across the examined samples, with ~~a very low number~~ few~~of~~ SNPs and ~~a low~~little polymorphism of the hypervariable sequences flanking the control region (Largest Unassigned Regions, LURs). Strikingly, ~~contrasting with~~although we

Commented [TC1]: use full term on first use

Commented [TC2]: explain term on first use

found ~~the l~~ow nucleotide variability along the entire mitochondrial genome, we observed a ~~pretty~~ high levels of length polymorphism in the LUR ~~length~~ due to CNV of tandem repeats, and even a LUR length heteroplasmy in two samples. It is not clear if the lack of genetic variability in the mitochondrial genome of *R. decussatus* is a cause or an effect of the ongoing replacement of *R. decussatus* with the invasive *R. philippinarum*, and more analyses, especially on nuclear sequences, are required to assess this point. In the coding regions we found ~~putative~~ some indication~~s~~ for RNA editing.

Keywords: (3-10)

complete mitochondrial genome; mitochondrial length polymorphism; mitochondrial repeats; mitochondrial RNA editing; codon usage; bivalve molluscs; European clam; comparative mitochondrial genomics.

Introduction

Despite a large number of animal complete mitochondrial genomes (mtDNAs) being available in public databases (>55,000 in GenBank), up to now sequencing has been focused mostly on vertebrates (~50,000 in GenBank), and the current knowledge about mitochondrial genomics in invertebrates—with the notable exception of few model organisms (e.g. Drosophila and *Caenorhabditis elegans*)—is uneven~~sketchy~~. To better understand invertebrate mitochondrial biology—and, most importantly, mitochondrial biology and evolution in general—it is necessary to adopt a more widespread approach in gathering and analyzing data. Failing to do so would bias our knowledge toward a few taxonomic groups, with the risk of losing a big part of the molecular and functional diversity of mitochondria. Actually, despite maintaining its core features in terms of genetic content, mtDNA in Metazoa shows a wide range of variation in some other traits such as, for example, genome architecture, abundance of unassigned regions (URs)—namely regions with no assigned product (protein, RNA)—repeat content, gene duplications, introns, UTRs, and even additional coding genes (see Breton et al., 2014 for a review) or genetic elements (e.g, small RNAs, see Pozzi et al., 2017) . All this emerging diversity is in sharp contrast with the—at this point outdated—textbook notion about mtDNAs role being limited to the production of a few subunits of the protein complexes involved in oxidative phosphorylation (OXPHOS).

This paper reports, for the first time, the complete mitochondrial genome of the grooved carpet shell, *Ruditapes decussatus* (Linnaeus, 1758). *R. decussatus*—also known as the European clam—is distributed all over the Mediterranean coasts, as well as on the Atlantic shores, from Lofoten Islands (Norway) to Mauritania, including the British Isles. *R. decussatus* lives in warm

85 coastal waters, especially in lagoons, and it is morphologically and ecologically similar to the
86 Manila clam *Ruditapes philippinarum*, which has been recently introduced for aquaculture in the
87 very same habitats of *R. decussatus. R. philippinarum*, native from the Philippines, Korea, and
88 Japan, was accidentally introduced into North America in the 1930s, and from there it was
89 purposely introduced in France (1972), UK (1980), and Ireland (1982) for aquaculture purposes
90 (Gosling, 2003). According to historical records, *R. decussatus* was one of the most important
91 species for aquaculture in Europe, but overfishing, irregular yields, recruitment failure, and
92 outbreaks of bacterial infection pushed the producers to introduce *R. philippinarum*; Italy
93 imported large quantities of *R. philippinarum* seed from UK in 1983 and 1984. Compared to the
94 European clam, the Manila clam turned out to be faster growing, more resistant to disease, to
95 have a more extended breeding period and a greater number of spawning events, and to begin
96 sexual maturation earlier (i.e. at a smaller size). The unavoidable consequence was the
97 replacement of *R. decussatus* with *R. philippinarum*, and currently the production of the
98 European clam is almost insignificant. Nonetheless the grooved carpet shell is considered a high
99 value product, and therefore it is an economically important species, especially in Portugal,
100 Spain and Italy (Gosling, 2003; Leite et al., 2013; de Sousa et al., 2014).
101 Molluscs in general, and bivalves in particular, exhibit an extraordinary degree of mtDNA
102 variability and unusual features, such as: large mitochondrial genomes (up to ~47Kb), high
103 proportion of URs, novel protein coding genes with unknown function, frequent and extensive
104 gene rearrangement, and differences in strand usage (Gissi, Iannelli & Pesole, 2008; Breton et
105 al., 2011; Ghiselli et al., 2013; Milani et al., 2014b; Plazzi, Puccio & Passamonti, 2016).
106 Moreover, mitochondrial genome size varies among bivalves because of gene duplications and
107 losses (Serb & Lydeard, 2003; Passamonti et al., 2011; Ghiselli et al., 2013), and sometimes

genes are fragmented as in the case of ribosomal genes in oysters (Milbury et al., 2010). The most notable feature of bivalves is the Doubly Uniparental Inheritance (DUI) of mitochondria (Skibinski, Gallagher & Beynon, 1994a,b; Zouros et al., 1994a,b). Under DUI, two different mitochondrial lineages (and their respective genomes) are transmitted to the progeny: one is inherited from the egg (female-transmitted or F-type mtDNA), the other is inherited from the spermatozoon (male-transmitted or M-type mtDNA). Following fertilization, the early embryo is heteroplasmic, but the type of mitochondria present in the adult is tightly linked to its sex. Females are commonly homoplasmic for F, while males are heteroplasmic with the following distribution of mtDNA types: the germ line is homoplasmic for the M-type (which will be transmitted via sperm to male progeny), the soma is heteroplasmic to various degrees, depending on tissue type and/or species (Ghiselli, Milani & Passamonti, 2011; Zouros, 2013). To date, the only known animals exhibiting DUI are about 100 species of bivalve molluscs (Gusman et al., 2016). This natural and evolutionarily stable heteroplasmic system can be extremely useful to investigate several aspects of mitochondrial biology (see Passamonti & Ghiselli, 2009; Breton et al., 2014; Milani & Ghiselli, 2015; Milani, Ghiselli & Passamonti, 2016). Indeed, despite many aspects of DUI are still unknown, there is evidence that DUI evolved from a strictly maternal inheritance (SMI) system (Milani & Ghiselli, 2015; Milani, Ghiselli & Passamonti, 2016), by modifications of the molecular machinery involved in mitochondrial inheritance, through as-yet-unknown specific factors (see Diz, Dudley & Skibinski, 2012; and Zouros, 2013 for proposed models). The detection of DUI is not a straightforward process, especially using PCR-based approaches: given that the divergence between F and M genomes is often comparable to the distance between mtDNAs of different classes of Vertebrates, primers may fail to amplify one of the two mtDNAs, yielding a false-negative result. Moreover, M-type mtDNA can be rare in

somatic tissues, so it may be difficult to amplify from animals sampled outside of the reproductive season, when gonads are absent (thoroughly discussed in Theologidis et al., 2008). High-throughput sequencing (HTS) approaches can overcome such problems, because a prior knowledge of the mtDNA sequence is not needed, and low-copy variants can be easily unveiled. Until now, HTS has been scarcely utilized to study mitochondrial transcriptomes and genomes (Pesole et al., 2012; Smith, 2013), even if it showed very good potential (Lubośny et al., 2017/2; see for example Yuan et al., 2016). In this work we: *i)* assembled *R. decussatus* mitochondrial genome from RNA-Seq data, and validated it by Sanger sequencing, *ii)* analyzed and characterized *R. decussatus* mitochondrial genome, comparing its features with those of other venerid bivalves; *iii)* assessed mitochondrial genetic polymorphism among the sampled animals, both from a nucleotide (SNPs) and a structural (CNV of tandem repeats) point of view.

Materials & Methods

Sampling

The 26 *Ruditapes decussatus* specimens used in this study were collected from the Northern Adriatic Sea (Goro) during the spawning season. Each individual was sexed, frozen in liquid nitrogen, and stored at -80ºC. Supplementary Table 1 shows the sample list, and details about data availability.

RNA-Seq

12 samples (6 males and 6 females) were used for RNA-Seq. Total RNA extraction and library preparation were performed following the protocol described in Mortazavi et al. (2008), with the

modifications specified in Ghiselli et al. (2012). The 12 samples were indexed, pooled and sequenced in two lanes (two technical replicates) of Illumina GA IIx, using 76bp paired-end reads.

De Novo Assembly

The mitochondrial genome of *R. decussatus* was not available in the databases, so we used the transcriptome data to generate a draft to be used as a guide for Sanger sequencing. The mtDNA *de novo* assembly followed this pipeline: *i)* the reads of the 12 individuals of *R. decussatus* were combined; *ii)* an alignment against 20 complete mitochondrial genomes of Veneridae species (retrieved from GenBank) was performed using BLASTN, with cutoff E-value ≤1E-5; *iii)* the aligned reads were assembled using the A5 pipeline (Andrew And Aaron's Awesome Assembly pipeline, Tritt et al., 2012); *iv)* the obtained contigs were joined into scaffolds using CAP3 (Huang & Madan, 1999). A5 is an assembly pipeline that reduces the entire genome assembly process by automating stages (data cleaning, error correction, assembly, and quality control). A5 can produce quality assemblies without any prior knowledge of the particular genome being assembled and without the extensive parameter tuning required by ~~the~~ other assembly algorithms. A5 can operate directly on FASTQ format data generated by illumina sequencing. CAP3 clips 5' and 3' low-quality regions of reads, and uses base quality values in computation of overlaps between reads, construction of multiple sequence alignments of reads, and generation of consensus sequences. The program also uses forward-reverse constraints to correct assembly errors and link contigs.

Commented [TC3]: Were these all pieces of individual genes, or were some polycistronic? Did you recover tRNAs?

Commented [TC4]: So, were these reads all within-gene or did some span gene boundaries?

Sanger Validation

14 *R. decussatus* samples from the same collection campaign—sexed, and stored at -80°C—were used for DNA extraction. DNA from the gonadic tissue was extracted using the Qiagen DNeasy kit. Primers for mtDNA amplification were designed based on contigs obtained from RNA-Seq matching venerid mtDNA sequences, then the "primer walking" method was used to Sanger-sequence the complete mitochondrial genome of *R. decussatus*. The primers were designed with the software Primer3 (Rozen & Skaletsky, 2000) and tested on several samples, then a female (F4) was chosen as reference sample for Sanger validation of mtDNA *de novo* assembly. In addition, we amplified the Largest Unassigned Region (LUR) of 13 females to assess its variability (see Results and Discussion). The list of the primers and their sequences are reported in Supplementary Table 2. PCR reactions were performed in a final volume of 50µl using the GoTaq Flexi DNA Polymerase Kit (Promega), on a 2720 Thermal Cycler (Applied Biosystem). The PCR reactions were set as follows: initial denaturation 95°C for 1 min, then 30 cycles of amplification (denaturation 95°C for 1 min, annealing 48- 60°C for 1 min, extension 72°C for 1 min/kb), then the final extension at 72°C for 5 min. PCR products were checked by electrophoretic run on 1% agarose gel, and then purified using the DNA Clean & Concentrator-25 kit (Zymo Research).

Sanger sequencing was performed by Macrogen Inc. (http://www.macrogen.com).

Sequences were aligned with the software MEGA 6.0 (Tamura et al., 2013), using the contigs obtained by RNA-seq as a reference.

Commented [TC5]: season? date?

Commented [TC6]: How many regions were amplified? How much did they overlap? What is their average size?

Annotation

Open Reading Frames (ORFs) were identified with ORF finder (Wheeler et al., 2005). Alternative start codons were considered functional~~also used~~ because they are common in Bivalvia. ORFs were annotated starting from the first available start codon (ATG, ATA or ATC) downstream of the preceding gene, and ending with the first stop codon in frame (TAA or TAG). tRNA genes and their structure were identified with MITOS (Bernt et al., 2013) and ARWEN (Laslett & Canback, 2008). Secondary structures were predicted using the RNAFold Server, included in the ViennaRNA Web Services (http://rna.tbi.univie.ac.at/; Gruber et al., 2008); the folding temperature was set at 16°C which is the average annual temperature of the water from which the *R. decussatus* specimens used in this work were fished (results available on figshare: https://doi.org/10.6084/m9.figshare.4970762.v1). tRNAs and other secondary structures were drawn with the software Varna GUI (Darty, Denise & Ponty, 2009). Ribosomal small subunit (*rrnS*) and large subunit (*rrnL*) were identified with BLASTN, and annotated considering the start and the end of the adjacent genes as the boundaries of the rRNA genes. Non-genic regions were annotated as Unassigned Regions (URs). In order to identify the putative D-loop/control region (CR), we analyzed the LUR with the MEME suite (Bailey et al., 2009) to find DNA motifs using the following bivalve species as comparison: *Acanthocardia tuberculata*, *Arctica islandica*, *Coelomactra antiquata*, *Fulvia mutica*, *Hiatella arctica*, *Loripes lacteus, Lucinella divaricata, Lutraria rhynchaena, Meretrix lamarckii* (F-type), *Meretrix lamarckii* (M-type), *Meretrix lusoria, Meretrix lyrata, Meretrix meretrix, Meretrix petechialis, Moerella iridescens, Nuttallia olivacea, Paphia amabilis, Paphia euglypta, Paphia textile, Paphia undulata, Ruditapes philippinarum* (F-type), *Ruditapes philippinarum* (M-type*), Semele scabra, Sinonovacula constricta, Solecurtus divaricatus, Solen grandis, Solen strictus, Soletellina diphos*

and the sea urchin *Strongylocentrotus purpuratus* (Echinoidea, Strongylocentrotidae). The list of the species used in the phylogenetic analysis and in the comparative analyses of DNA motifs, sequence similarity, and gene order are available in Supplementary Table 3. The GOMo (Gene Ontology for Motifs; Buske et al., 2010) tool of the MEME suite was used to assign GO terms to the ~~found~~ motifs discovered.

The number of repeats in the LUR of the reference sample (F4) was calculated with tandem repeat finder (http://tandem.bu.edu/trf/trf.html), since the complete LUR sequence was available (results available on figshare: https://doi.org/10.6084/m9.figshare.4970762.v1). In the other cases, in which the LUR could not be sequenced without gaps, the number of repeats was inferred from agarose gel electrophoresis.

Other Analyses

Comparisons among venerid complete mtDNAs were performed with BLAST Ring Image Generator (BRIG, Alikhan et al., 2011) and Easyfig (Sullivan, Petty & Beatson, 2011). Descriptive statistics were obtained with MEGA v6.0 (Tamura et al., 2013), except for the codon usage table, which was obtained with the Sequence Manipulation Suite (Stothard, 2000). SNP calling was performed with the Genome Analysis Toolkit (GATK, McKenna et al., 2010), with the Sanger-sequenced mtDNA as reference. For SNP/INDEL discovery and genotyping we used standard hard filtering parameters or variant quality score recalibration (DePristo et al., 2011). The MitoPhast pipeline (Tan et al., 2015) was used to obtain the Maximum Likelihood (ML) tree, which was visualized with Evolview v2 (He et al., 2016). Briefly, MitoPhast takes as input GenBank files (.gb), extracts the coding sequences, profiles the sequences with Pfam (Finn et al., 2016) and PRINTS (Attwood et al., 2003), performs a multiple sequence alignment with Clustal

Omega (Sievers et al., 2011), removes poorly aligned regions with trimAl (Capella-Gutiérrez, Silla-Mart'inez & Gabaldón, 2009), concatenates the coding sequences, performs data partitioning and model selection, and then carries out a ML analysis using RAxML (Stamatakis, 2014). The species used in the ML analysis, and their GenBank Accession Numbers are listed in Supplementary Table 3.

Results

Sex-linked mitochondrial haplotypes were not discovered in the ~~The~~ RNA-Seq analysis of 12 *R. decussatus* mature gonads (6 males and 6 females~~) has not retrieved sex-linked mitochondrial haplotypes~~.

The mitochondrial genome contains 13 protein-coding genes, and in the reference female is 18,995 bp long (Figure 1); the gene arrangement and other details are shown in Table 1. All genes are located on the heavy strand, and in addition to the classic start codon ATG (Met), the alternative start codons ATA (Met) and ATC (Ile) are present. The most frequently used start codons are: ATA (*cox1, nd1, nd4L, cox2, cob, atp8, nd4*), and ATG (*nd2, atp6, nd3, nd5, nd6, cox3*). The stop codons found are TAG (*cox1, nd2, nd4L, cox2, cytb, nd4*) and TAA (*nd1, atp6, nd3, atp8, nd6*). The *nd4* gene has an incomplete stop codon (TA-). 22 tRNA genes were identified, including two tRNAs for leucine, tRNA-Leu1(TAG) and tRNA-Leu2(TAA), and two for serine, tRNA-Ser1(TCT) and tRNA-Ser2(TGA), both showing ~~a~~ degenerate D-arm branches. tRNA structures are shown in Supplementary Figure 1. The two rRNAs, *rrnS* and *rrnL*, were both identified: the *rrnS* is located between *cox3* and *cox1*, while *rrnL* is between *cytb* and *atp6*. Unassigned Regions (URs) were identified on the basis of unannotated spaces between different genes; we found 24 URs (Table 2).

The analysis of the nucleotide composition points out that the mitochondrial genome of this bivalve species ~~presents a~~exhibit high A+T content, total~~l~~ing 63% versus 37% ~~of~~ G+C. The minimum values of A+T are found in *cytb* (60.1%) and *nd4* (61%). The nucleotide composition of every gene is shown in Table 3. According to the analysis above, both A and T occur very frequently at the third ~~base of the~~ position of codons (64.6% on average of A+T), while the less frequent base in third position is C (12%). The most used codons are UUU (Phe), counted 269 times, and UUA (Leu) counted 210 times (6.78% and 5.29% of the total, respectively), while the less used codons are CGC (Arg) counted 6 times (0.15%), ACC (Thr) and CCG (Pro) each counted 16 times (0.4%) (Table 4). Only in 4 cases over 20 (Lys, Leu, Gln, Val), the most frequently used codon matches the correspondent mitochondrial tRNA anticodon.

The UR11 is the Largest Unassigned Region (LUR) and is located between *atp8* and *nd5* (Figures 1 and 2A). The LUR of the female used for whole mtDNA Sanger sequencing (i.e. the reference female, F4) is 2,110 bp long, and includes 6.5 repeated sequences—each repeat having a length of 54 bp—localized in the 3' region of the LUR, just upstream the *atp8* gene (Figure 2A). DNA secondary structure analysis predicted 3 stem-loop structures in such region (Figure 2B and supplementary files on figshare: https://doi.org/10.6084/m9.figshare.4970762.v1), with a change in Gibbs free energy (ΔG) of -71.38 Kcal/mol. We also amplified and sequenced the LUR of 13 more females. We were not able to completely sequence LURs longer than 2,110 bp, because of the known difficulties in Sanger sequencing of regions including multiple repeats. The sequence alignment of the 13 LURs is available for download from figshare (https://doi.org/10.6084/m9.figshare.4970762.v1). LUR lengths, inferred from gel electrophoresis, are reported in Table 5, and they ranges from 2,000 to 5,000 bp. Two females (F3 and F17) showed length heteroplasmy of the LUR. The portion of the genome occupied by

URs varies between 14.11% and 29.38%, depending on LUR length. The analysis with MEME (output shown in Supplementary Figures 3 and 4) unveiled two motifs (Figure 2C) that show a strong conservation within the Veneridae family, and with *S. purpuratus*. The sea urchin was included in the analysis because Cao et al. (2004) reported a match between some motifs found in the CR of the marine mussels *Mytilus edulis* and *Mytilus galloprovincialis* with regulatory elements of the sea urchin CR. Accordingly, the search with GOMo assigned a series of GO terms related to transcription to the two motifs (Table 6).

Table 7 shows the statistics associated with the SNP calling performed with GATK on the 12 samples used for RNA-Seq, with the Sanger-sequenced mtDNA as reference. Overall, 257 SNPs were called, of which 145 (56.4%) are located in CDS. Interestingly, most of the SNPs (103 over 145, that is, 71% of the SNPs in CDS) were called because of private alleles of one single male specimen (mRDI01). If we exclude the SNPs associated with this male, the genetic polymorphism drops to 42 SNPs over 14,920 bp of coding mtDNA (GATK output in VCF format and a detailed list of SNPs in tabular format is available on figshare: https://doi.org/10.6084/m9.figshare.4970762.v1). 18 SNPs are indels, 6 of which are located in 4 different coding genes: one each in *cox1*, *cytb*, and *nd5*, plus 3 in *cox3* (see Table 8). A file showing the ORF generated by the different variants of *cox3*, and alignments between them is available on figshare (https://doi.org/10.6084/m9.figshare.4970762.v1).

Figure 3 shows the *R. decussatus* mtDNA map (external gray circle), and the BLASTN identity (colored inner circles) with complete mtDNAs of other 10 venerid species (see list in Supplementary Table 3). Figure 4 shows the ML tree obtained with the MitoPhast pipeline; the complete input and output of this analysis is available on figshare (https://doi.org/10.6084/m9.figshare.4970762.v1). Figure 5 shows the variation in gene order

Commented [TC7]: Do you mean SNPs here, or any kind of polymorphism?

Commented [TC8]: use full term on first usage

Formatted: Highlight

Commented [TC9]: indels aren't SNPs

between *R. decussatus* and *P. euglypta* (Figure 5A), *M. lamarckii* F-type (Figure 5B), *R. philippinarum* F-type (Figure 5C), and among all the 4 species (Figure 5D).

Discussion

The size of the fully Sanger-sequenced mitochondrial genome of *R. decussatus* (reference female F4) is of 18,995 bp, and it includes 13 protein-coding genes, 22 tRNAs and 2 rRNAs. Our data support the presence of the *atp8* gene in the mtDNA of *R. decussatus; atp8* has been reported as missing in several bivalve species, however more accurate searches often led to the identification of the gene, so, in most cases, the alleged lack of *atp8* is likely ascribable to annotation inaccuracies due to the extreme variability and the small size of the gene (Breton, Stewart & Hoeh, 2010; Breton et al., 2014).

The mitochondrial genome of *R. decussatus* shows a high content of A-T (63%), a common feature in bivalve mtDNAs; moreover, T is the most ~~present~~ common nucleotide at the third codon base (64.6%). The most commom ~~used~~ codon is UUU (Phe), which is also the most commonly used in bivalves, as well as in other invertebrates (Passamonti et al., 2011).

As shown in Table 4, in 16 cases out of~~ver~~ 20, the most frequently used codon does not correspond to the anticodon of the inferred~~relative~~ tRNA ~~produced by mtDNA~~. In other words, there is not a correspondence between the most representative codons and anticodons of the 22 mitochondrial tRNAs. According to the "wobble hypothesis"—first proposed by Francis Crick (1966)— the conformation of the tRNA anticodon loop enables some flexibility at the first base of the anticodon, so a Watson-Crick type of base pairing in the third position of the codon is not strictly necessary. This allows an amino acid to be correctly incorporated by ribosomes even if the tRNA is not ~~rigidly~~ fully complementary to the codon; according to Crick, this explains the degeneracy of the genetic code. This feature is particularly interesting in the light of the debate about~~on~~ natural selection acting at synonymous sites: since the early 1980s, evidence of ~~about~~ a correlation between synonymous codon usage and tRNA abundances started accumulating.

Commented [TC10]: Do you mean that the most abundant anticodons do not correspond to the most abundant codons?

Commented [TC11]: citations?

According to ~~several works~~these authors, synonymous codon usage is biased to match skews in tRNA abundance, as a result of ~~the~~ selective pressure ~~for a~~ maximizing~~ed~~ protein synthesis rates (reviewed in Chamary, Parmley & Hurst, 2006). Following this rationale, the results here reported and data from other metazoans (see Passamonti et al., 2011 and references therein) would ~~entail~~ suggest that in some mitochondrial genomes ~~the~~ translation efficiency is not maximized, and this observation deserves further investigation.

The mtDNA of *R. decussatus* has a high proportion of URs mostly depending on the length of the LUR (Table 5); on average, bivalve mtDNAs have 1.7x the amount of URs in respect to other analyzed Metazoa (Ghiselli et al., 2013), and it is still unclear whether there is an accumulation of non-functional sequences in bivalve mtDNA due to genetic drift, or if such URs are maintained by natural selection because they contain—so far unknown—functional elements (see for example Milani et al., 2013, 2014b; Breton et al., 2014; Pozzi et al., 2017). The LUR of *R. decussatus* most likely includes the mitochondrial CR, as indicated by the presence of two motifs (Figure 2C, Supplementary Figures 3 and 4) similar to two regulatory elements identified in the sea urchin CR. These two motifs are the same identified in previous analyses on the clam *R. philippinarum* and the mussel *Musculista senhousia* (Ghiselli et al., 2013; Guerra, Ghiselli & Passamonti, 2014) so they are conserved across distant bivalve taxa, and the GO terms associated with such motifs are related to transcriptional control (Table 6). An interesting feature of *R. decussatus* LUR is its variable length (Table 2), most likely due to different repeat content. As a matter of fact, the very same repeats were present in every sequenced LUR, and our data strongly suggest that LUR length variation is actually due to repeat CNV (see supplementary files on figshare: https://doi.org/10.6084/m9.figshare.4970762.v1), as observed in other bivalve species (see for example Ghiselli et al., 2013; Guerra, Ghiselli & Passamonti, 2014). Tandem repeats

have been also reported in the mitochondrial genomes of the bivalves *Acanthocardia tuberculata (Dreyer & Steiner, 2006)*, *Placopecten magellanicus (La Roche et al., 1990)*, *Moerella iridescens*, *Sanguinolaria olivacea*, *Semele scaba*, *Sinonovacula constricta*, *Solecurtus divaricatus* (Yuan et al., 2012), *Ruditapes philippinarum (Ghiselli et al., 2013)*, and *Musculista senhousia* (Guerra, Ghiselli & Passamonti, 2014). These repeats are believed to arise from duplications caused by replication slippage (Buroker et al., 1990; Hayasaka, Ishida & Horai, 1991; Broughton & Dowling, 1994). The tandem repeats found at the 3' end of *R. decussatus* LUR are predicted to form a secondary structure (see Figure 2B, and supplementary files on figshare) composed by multiple stem-loops, which obviously increase in number with the increment of the number of tandem repeats. The effect, if any, of tandem repeats in mtDNA is unknown: since the repeats are almost always localized in proximity of the CR, they might interact with regulatory elements—or even contain some—influencing replication and/or transcription initiation, and such interactions might also be altered by the formation of secondary structures (Passamonti et al., 2011; Ghiselli et al., 2013; Guerra, Ghiselli & Passamonti, 2014). We assessed the genetic variability of *R. decussatus* mtDNA using two different approaches: by SNP calling in CDS (RNA-Seq data on 12 individuals), and by analysis of the LUR (Sanger sequencing of 14 individuals). The CR and its flanking regions are known to be hypervariable, so they are commonly used to assess polymorphism at low taxonomic levels. Our data strongly support a very low genetic variability: the number of SNPs in CDS is 145, of which 103 are private of a single individual (mRDI01)—thus reducing the number to 42—while the number of variable sites in the analyzed LURs is 98 over 3,095 aligned positions. Considering the known variability of mtDNA in bivalves (Gissi, Iannelli & Pesole, 2008; Ghiselli et al., 2013; Breton et al., 2014; Plazzi, Puccio & Passamonti, 2016), this is a surprising result. Even more if we

390 compare the results of the present work to a methodologically identical analysis performed on 12
391 *R. philippinarum* samples from the Pacific coast of USA, performed by Ghiselli et al. (2013): in
392 that work, GATK yielded 194 SNPs in the M-type mtDNA and 293 in the F-type. Strikingly, the
393 12 *R. philippinarum* samples analyzed were actually two families (6 siblings + 6 siblings). This
394 means that randomly sampled individuals of *R. decussatus* used in this work showed a much
395 lower mtDNA variability than *R. philippinarum* siblings. A previous analysis on the *cox1* gene of
396 *R. decussatus* reported a nucleotide diversity (π) of 0.15 for a population from the Northern
397 Adriatic Sea (Cordero, Peña & Saavedra, 2014). Another analysis on the same gene of *R.*
398 *philippinarum* from the same range resulted in a π=0.25 (Cordero et al., 2017), so *R. decussatus*
399 has a lower nucleotide diversity at the *cox1* locus. The difference between the variability in
400 mtDNA of *R. decussatus* that -we are reporting here and that of *R. philippinarum* reported in
401 Ghiselli et al. (2013) -appears to be more marked. It is known that the genetic variability of *R.*
402 *philippinarum* in the Adriatic Sea is lower than in populations from its native range in Asia
403 (Cordero et al., 2017), probably because of the bottlenecks that this species had to go through
404 during the multiple colonization events. The introduction in North America from Asia happened
405 first (in the 1930s), and from there the Manila clam was introduced in Northern Europe (in the
406 1970s and 1980s), and lastly into the Adriatic Sea (1983 and 1984), and it is plausible that the
407 genetic diversity decreased at each introduction event. Accordingly, Cordero et al. (2017)
408 observed that *R. philippinarum* genetic variability in Europe is lower compared to that of the
409 Pacific coast of the USA, so the samples analyzed in Ghiselli et al. (2013) could have been more
410 polymorphic than those analyzed in Cordero et al. (2014), thus explaining the more pronounced
411 differences in genetic variability between the Manila clam and the European clam discussed
412 above. In any case, all the available data point to a lower genetic diversity of *R. decussatus*

mtDNA, and it would be interesting to know whether it is a cause or an effect of the ongoing replacement of *R. decussatus* with the invasive *R. philippinarum*. It will also be important to investigate genetic variability of the nuclear genes, especially after Cordero et al. (2014) reported contrasting levels of differentiation between mitochondrial and nuclear markers.

With respect to SNP effects, we found 6 indels in CDS, 2 of which do not cause frameshift, but a simple insertion/deletion of 1 amino acid (SNP_1698, and SNP_17619, see Table 8). Of the remaining 4, SNP_6364 and SNP_10449 consist of a deletion and an insertion of a single T in two homopolymeric sequences (CTTTTTTT and CTTTTTT, respectively), raising the possibility~~suspect~~ of a sequencing error. In any case, the two SNPs yield a shorter CDS (*cytb* and *nd5*, respectively), and are present at relatively low frequencies in the specimens carrying them, except for SNP_6364 which has a frequency of 80% in fRDI04. The *cox3* gene shows 3 SNPs: the first one, SNP_17619, does not cause a frameshift, and results in the deletion of 1 alanine residue, and its frequency in mRDI01 is 97%. The second one, SNP_17621, consists of a deletion of a G with~~in~~ respect to the reference sequence, which is the Sanger-sequenced mtDNA of sample F4; all the individuals analyzed with RNA-Seq carry this deletion except for mRDI01 which, at that position, has the same sequence of the reference mtDNA (reference-like allele frequency in mRDI01 = 99%). The third indel, SNP_17624, consists of an insertion of two nucleotides, and its frequency in mRDI01 is 99%. So, basically, for *cox3* we have three types of sequences: *i)* the Sanger-sequenced reference, which yields a 966 bp (321 aa) ORF; *ii)* a sequence found in 11/12 of samples analyzed with RNA-Seq (except mRDI01) that carries a single-nucleotide deletion (SNP_17621), and yields a 963 bp (320 aa) ORF; *iii)* a sequence, private of mRDI01, which is obtained by combining SNP_17624 and SNP_17621 (both 99% of frequency, so most likely co-occurring), which produces a 963 bp (320 aa) ORF. Interestingly,

Commented [TC12]: Where is this indel located? If very near the end of the gene, may not have much effect.

Commented [TC13]: So, it is a 3bp indel?

Commented [TC14]: A figure could clarify the presentation of this important inference

436 the ORFs obtained from the sequences described in *ii)* and *iii)*, are almost identical, namely the
437 sequence obtained by RNA-seq in 11/12 samples and the sequence obtained by RNA-Seq in
438 mRDI01 are basically the same, and differ from the Sanger-sequenced reference, yielding an
439 amino acid sequence that differs in the last 35 residues (all data available in supplementary files
440 on figshare: https://doi.org/10.6084/m9.figshare.4970762.v1). Given this consistent difference
441 between the sequence obtained by Sanger-sequencing of DNA, and those obtained by RNA-Seq,
442 it is tempting to speculate that this difference might be caused by RNA editing, a mechanism
443 observed in mtDNA of some animals (Lavrov & Pett, 2016), and recently reported to be ~~widely~~
444 ~~used~~common in cephalopods (Liscovitch-Brauer et al., 2017). Post-transcriptional modifications
445 (thus including RNA-editing) are still poorly understood mechanisms, but they appear to be
446 responsible for most of the mitochondrial gene expression regulation (Scheibye-Alsing et al.,
447 2007; Scheffler, 2008; Milani et al., 2014a).
448 Interestingly, in contrast with a low nucleotide variability along the entire mitochondrial genome,
449 we observed a pretty high polymorphism in LUR length due to CNV of tandem repeats, and even
450 a LUR length heteroplasmy: two females yielded two electrophoretic bands each (~2,100 and
451 ~3,500 bp in F3; ~2,500 and ~3,500 bp in F17; see Table 5).
452 A possible explanation is that the diversity (CNV) detected in the LURs could be recent: the
453 accumulation of nucleotide variation at different sites along the mitochondrial genome needs
454 time, while the kind structural variability we observed can be achieved in few generations (or
455 even one) considering that replication slippage is common in repeat-rich regions.
456 Despite *R. decussatus* and *R. philippinarum* being morphologically similar and being ascribed to
457 the same genus, the results here reported clearly show that they are quite different both for
458 mtDNA sequence (Figures 3 and 4) and mtDNA gene arrangement (Figure 5). This is an unusual

finding, even among bivalves, which are known to be fast-evolving for thesecis characters. This may point to the fact that these two species are less related than previously thought. Actually, this is not the first clue that *R. decussatus* and *R. philippinarum* are quite different genetically, as allozyme electrophoresis (Passamonti, Mantovani & Scali, 1997, 1999) and satellite DNA content (Passamonti, Mantovani & Scali, 1998) pointed out. More in--depth analyses are therefore needed to correctly trace the phylogenetic relationships of these two Ruditapes species, which may eventually end up in two different Genera. As shown in Figures 3, 4 and 5, the Genus Paphia is the most similar to *R. decussatus*. Finally, even using a HTS approach we could not find evidence for sex-specific mtDNAs, so our data appear to confirm that *R. decussatus* does not have DUI. Among Veneridae, only 3 species have been found with DUI, so far: *Cyclina sinensis*, *R. philippinarum*, and *Meretrix lamarckii* (Plazzi, Cassano & Passamonti, 2015; Gusman et al., 2016). The status of Paphia is still unknown, and in future works it would be interesting to investigate this Genus as well as other Heterodonta to understand better the distribution of DUI in this derived group of bivalves.

References

Alikhan NF., Petty NK., Ben Zakour NL., Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC genomics* 12:402.

Attwood TK., Bradley P., Flower DR., Gaulton A., Maudling N., Mitchell AL., Moulton G., Nordle A., Paine K., Taylor P., Uddin A., Zygouri C. 2003. PRINTS and its automatic supplement, prePRINTS. *Nucleic acids research* 31:400–402.

Bailey TL., Boden M., Buske FA., Frith M., Grant CE., Clementi L., Ren J., Li WW., Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic acids research* 37:W202–8.

Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf M., Stadler PF. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular phylogenetics and evolution* 69:313–319.

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 in genetics: TIG* 30:555–564.

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. *Marine genomics* 3:11–18.

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)? *Molecular biology and evolution* 28:1645–1659.

Broughton RE., Dowling TE. 1994. Length variation in mitochondrial DNA of the minnow

496 *Cyprinella spiloptera*. *Genetics* 138:179–190.
497 Buroker NE., Brown JR., Gilbert TA., O'Hara PJ., Beckenbach AT., Thomas WK., Smith MJ.
498 1990. Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation
499 model. *Genetics* 124:157–163.
500 Buske FA., Bodén M., Bauer DC., Bailey TL. 2010. Assigning roles to DNA regulatory motifs
501 using comparative genomics. *Bioinformatics* 26:860–866.
502 Cao L., Kenchington E., Zouros E., Rodakis GC. 2004. Evidence that the large noncoding
503 sequence is the main control region of maternally and paternally transmitted mitochondrial
504 genomes of the marine mussel (*Mytilus spp*.). *Genetics* 167:835–850.
505 Capella-Gutiérrez S., Silla-Martinez JM., Gabaldón T. 2009. trimAl: a tool for automated
506 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
507 Chamary JV., Parmley JL., Hurst LD. 2006. Hearing silence: non-neutral evolution at
508 synonymous sites in mammals. *Nature reviews. Genetics* 7:98–108.
509 Cordero D., Delgado M., Liu B., Ruesink J., Saavedra C. 2017. Population genetics of the
510 Manila clam (*Ruditapes philippinarum*) introduced in North America and Europe. *Scientific*
511 *reports* 7:39745.
512 Cordero D., Peña JB., Saavedra C. 2014. Phylogeographic analysis of introns and mitochondrial
513 DNA in the clam *Ruditapes decussatus* uncovers the effects of Pleistocene glaciations and
514 endogenous barriers to gene flow. *Molecular phylogenetics and evolution* 71:274–287.
515 Crick FHC. 1966. Codon—anticodon pairing: The wobble hypothesis. *Journal of molecular*
516 *biology* 19:548–555.
517 Darty K., Denise A., Ponty Y. 2009. VARNA: Interactive drawing and editing of the RNA
518 secondary structure. *Bioinformatics* 25:1974–1975.

519 DePristo MA., Banks E., Poplin R., Garimella KV., Maguire JR., Hartl C., Philippakis AA., del
520 Angel G., Rivas MA., Hanna M., McKenna A., Fennell TJ., Kernytsky AM., Sivachenko
521 AY., Cibulskis K., Gabriel SB., Altshuler D., Daly MJ. 2011. A framework for variation
522 discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*
523 43:491–498.
524 Diz AP., Dudley E., Skibinski DOF. 2012. Identification and characterisation of highly
525 expressed proteins in sperm cells of the marine mussel *Mytilus edulis*. *Proteomics*. DOI:
526 10.1002/pmic.201100500.
527 Dreyer H., Steiner G. 2006. The complete sequences and gene organisation of the mitochondrial
528 genomes of the heterodont bivalves *Acanthocardia tuberculata* and *Hiatella arctica*—and
529 the first record for a putative Atpase subunit 8 gene in marine bivalves. *Frontiers in zoology*
530 3:13.
531 Finn RD., Coggill P., Eberhardt RY., Eddy SR., Mistry J., Mitchell AL., Potter SC., Punta M.,
532 Qureshi M., Sangrador-Vegas A., Salazar GA., Tate J., Bateman A. 2016. The Pfam protein
533 families database: towards a more sustainable future. *Nucleic acids research* 44:D279–85.
534 Ghiselli F., Milani L., Chang PL., Hedgecock D., Davis JP., Nuzhdin SV., Passamonti M. 2012.
535 De Novo assembly of the Manila clam *Ruditapes philippinarum* transcriptome provides
536 new insights into expression bias, mitochondrial doubly uniparental inheritance and sex
537 determination. *Molecular biology and evolution* 29:771–786.
538 Ghiselli F., Milani L., Guerra D., Chang PL., Breton S., Nuzhdin SV., Passamonti M. 2013.
539 Structure, transcription, and variability of metazoan mitochondrial genome: perspectives
540 from an unusual mitochondrial inheritance system. *Genome biology and evolution* 5:1535–
541 1554.

542 Ghiselli F., Milani L., Passamonti M. 2011. Strict sex-specific mtDNA segregation in the germ
543 line of the DUI species *Venerupis philippinarum* (Bivalvia: Veneridae). *Molecular biology*
544 *and evolution* 28:949–961.
545 Gissi C., Iannelli F., Pesole G. 2008. Evolution of the mitochondrial genome of Metazoa as
546 exemplified by comparison of congeneric species. *Heredity* 101:301–320.
547 Gosling EM. 2003. *Bivalve molluscs*. Wiley Online Library.
548 Gruber AR., Lorenz R., Bernhart SH., Neuböck R., Hofacker IL. 2008. The Vienna RNA
549 websuite. *Nucleic acids research* 36:W70–4.
550 Guerra D., Ghiselli F., Passamonti M. 2014. The largest unassigned regions of the male- and
551 female-transmitted mitochondrial DNAs in *Musculista senhousia* (Bivalvia Mytilidae).
552 *Gene* 536:316–325.
553 Gusman A., Lecomte S., Stewart DT., Passamonti M., Breton S. 2016. Pursuing the quest for
554 better understanding the taxonomic distribution of the system of doubly uniparental
555 inheritance of mtDNA. *PeerJ* 4:e2760.
556 Hayasaka K., Ishida T., Horai S. 1991. Heteroplasmy and polymorphism in the major noncoding
557 region of mitochondrial DNA in Japanese monkeys: association with tandemly repeated
558 sequences. *Molecular biology and evolution* 8:399–415.
559 He Z., Zhang H., Gao S., Lercher MJ., Chen W-H., Hu S. 2016. Evolview v2: an online
560 visualization and management tool for customized and annotated phylogenetic trees.
561 *Nucleic acids research* 44:W236–41.
562 Huang X., Madan A. 1999. CAP3: A DNA sequence assembly program. *Genome research*
563 9:868–877.
564 La Roche J., Snyder M., Cook DI., Fuller K., Zouros E. 1990. Molecular characterization of a

565 repeat element causing large-scale size variation in the mitochondrial DNA of the sea
566 scallop *Placopecten magellanicus*. *Molecular biology and evolution* 7:45–64.
567 Laslett D., Canback B. 2008. ARWEN: a program to detect tRNA genes in metazoan
568 mitochondrial nucleotide sequences. *Bioinformatics* 24:172–175.
569 Lavrov DV., Pett W. 2016. Animal Mitochondrial DNA as We Do Not Know It: mt-Genome
570 Organization and Evolution in Nonbilaterian Lineages. *Genome biology and evolution*
571 8:2896–2913.
572 Leite RB., Milan M., Coppe A., Bortoluzzi S., dos Anjos A., Reinhardt R., Saavedra C.,
573 Patarnello T., Cancela ML., Bargelloni L. 2013. mRNA-Seq and microarray development
574 for the Grooved Carpet shell clam, *Ruditapes decussatus*: a functional approach to unravel
575 host-parasite interaction. *BMC genomics* 14:741.
576 Liscovitch-Brauer N., Alon S., Porath HT., Elstein B., Unger R., Ziv T., Admon A., Levanon
577 EY., Rosenthal JJC., Eisenberg E. 2017. Trade-off between Transcriptome Plasticity and
578 Genome Evolution in Cephalopods. *Cell* 169:191–202.e11.
579 Lubośny M., Przyłucka A., Sańko TJ., Śmietanka B., Rosenfeld S., Burzyński A. 2017/2. Next
580 generation sequencing of gonadal transcriptome suggests standard maternal inheritance of
581 mitochondrial DNA in *Eurhomalea rufa* (Veneridae). *Marine genomics* 31:21–23.
582 McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K.,
583 Altshuler D., Gabriel S., Daly M., DePristo MA. 2010. The Genome Analysis Toolkit: a
584 MapReduce framework for analyzing next-generation DNA sequencing data. *Genome*
585 *research* 20:1297–1303.
586 Milani L., Ghiselli F. 2015. Mitochondrial activity in gametes and transmission of viable
587 mtDNA. *Biology direct* 10:22.

588 Milani L., Ghiselli F., Guerra D., Breton S., Passamonti M. 2013. A comparative analysis of
589 mitochondrial ORFans: new clues on their origin and role in species with doubly
590 uniparental inheritance of mitochondria. *Genome biology and evolution* 5:1408–1434.
591 Milani L., Ghiselli F., Iannello M., Passamonti M. 2014a. Evidence for somatic transcription of
592 male-transmitted mitochondrial genome in the DUI species *Ruditapes philippinarum*
593 (Bivalvia: Veneridae). *Current genetics* 60:163–173.
594 Milani L., Ghiselli F., Maurizii MG., Nuzhdin SV., Passamonti M. 2014b. Paternally transmitted
595 mitochondria express a new gene of potential viral origin. *Genome biology and evolution*
596 6:391–405.
597 Milani L., Ghiselli F., Passamonti M. 2016. Mitochondrial selfish elements and the evolution of
598 biological novelties. *Current zoology* 62:687–697.
599 Milbury CA., Lee JC., Cannone JJ., Gaffney PM., Gutell RR. 2010. Fragmentation of the large
600 subunit ribosomal RNA gene in oyster mitochondrial genomes. *BMC genomics* 11:485.
601 Mortazavi A., Williams BA., McCue K., Schaeffer L., Wold B. 2008. Mapping and quantifying
602 mammalian transcriptomes by RNA-Seq. *Nature methods* 5:621–628.
603 Passamonti M., Ghiselli F. 2009. Doubly uniparental inheritance: two mitochondrial genomes,
604 one precious model for organelle DNA inheritance and evolution. *DNA and cell biology*
605 28:79–89.
606 Passamonti M., Mantovani B., Scali V. 1997. Allozymic characterization and genetic
607 relationships among four species of Tapetinae (Bivalvia, Veneridae). *Italian journal of*
608 *zoology* 64:117–124.
609 Passamonti M., Mantovani B., Scali V. 1998. Characterization of a highly repeated DNA family
610 in tapetinae species (mollusca bivalvia: veneridae). *Zoological science* 15:599–605.

611 Passamonti M., Mantovani B., Scali V. 1999. Allozymic analysis of some Mediterranean
612 Veneridae (Mollusca: Bivalvia): preliminary notes on taxonomy and systematics of the
613 family. *Journal of the Marine Biological Association of the United Kingdom. Marine*
614 *Biological Association of the United Kingdom* 79:899–906.
615 Passamonti M., Ricci A., Milani L., Ghiselli F. 2011. Mitochondrial genomes and Doubly
616 Uniparental Inheritance: new insights from *Musculista senhousia* sex-linked mitochondrial
617 DNAs (Bivalvia Mytilidae). *BMC genomics* 12:442.
618 Pesole G., Allen JF., Lane N., Martin W., Rand DM., Schatz G., Saccone C. 2012. The neglected
619 genome. *EMBO reports* 13:473–474.
620 Plazzi F., Cassano A., Passamonti M. 2015. The quest for Doubly Uniparental Inheritance in
621 heterodont bivalves and its detection in *Meretrix lamarckii* (Veneridae: Meretricinae).
622 *Journal of zoological systematics and evolutionary research* 53:87–94.
623 Plazzi F., Puccio G., Passamonti M. 2016. Comparative Large-Scale Mitogenomics Evidences
624 Clade-Specific Evolutionary Trends in Mitochondrial DNAs of Bivalvia. *Genome biology*
625 *and evolution* 8:2544–2564.
626 Pozzi A., Plazzi F., Milani L., Ghiselli F., Passamonti M. 2017. SmithRNAs: could mitochondria
627 "bend" nuclear regulation? *Molecular biology and evolution*. DOI:
628 10.1093/molbev/msx140.
629 Rozen S., Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist
630 programmers. *Methods in molecular biology* 132:365–386.
631 Scheffler IE. 2008. *Mitochondria*. Hoboken, N.J.: Wiley.
632 Scheibye-Alsing K., Cirera S., Gilchrist MJ., Fredholm M., Gorodkin J. 2007. EST analysis on
633 pig mitochondria reveal novel expression differences between developmental and adult

634 tissues. *BMC genomics* 8:367.
635 Serb JM., Lydeard C. 2003. Complete mtDNA sequence of the North American freshwater
636 mussel, *Lampsilis ornata* (Unionidae): an examination of the evolution and phylogenetic
637 utility of mitochondrial genome organization in Bivalvia (Mollusca). *Molecular biology and*
638 *evolution* 20:1854–1866.
639 Sievers F., Wilm A., Dineen D., Gibson TJ., Karplus K., Li W., Lopez R., McWilliam H.,
640 Remmert M., Söding J., Thompson JD., Higgins DG. 2011. Fast, scalable generation of
641 high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems*
642 *biology* 7:539.
643 Skibinski DO., Gallagher C., Beynon CM. 1994a. Mitochondrial DNA inheritance. *Nature*
644 368:817–818.
645 Skibinski DO., Gallagher C., Beynon CM. 1994b. Sex-limited mitochondrial DNA transmission
646 in the marine mussel *Mytilus edulis*. *Genetics* 138:801–809.
647 Smith DR. 2013. RNA-Seq data: a goldmine for organelle research. *Briefings in functional*
648 *genomics* 12:454–456.
649 de Sousa JT., Milan M., Bargelloni L., Pauletto M., Matias D., Joaquim S., Matias AM., Quillien
650 V., Leitão A., Huvet A. 2014. A microarray-based analysis of gametogenesis in two
651 Portuguese populations of the European clam *Ruditapes decussatus*. *PloS one* 9:e92202.
652 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
653 large phylogenies. *Bioinformatics* 30:1312–1313.
654 Stothard P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and
655 formatting protein and DNA sequences. *BioTechniques* 28:1102, 1104.
656 Sullivan MJ., Petty NK., Beatson SA. 2011. Easyfig: a genome comparison visualizer.

Bioinformatics 27:1009–1010.

Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* 30:2725–2729.

Tan MH., Gan HM., Schultz MB., Austin CM. 2015. MitoPhAST, a new automated mitogenomic phylogeny tool in the post-genomic era with a case study of 89 decapod mitogenomes including eight new freshwater crayfish mitogenomes. *Molecular phylogenetics and evolution*. DOI: 10.1016/j.ympev.2015.02.009.

Theologidis I., Fodelianakis S., Gaspar MB., Zouros E. 2008. Doubly uniparental inheritance (DUI) of mitochondrial DNA in *Donax trunculus* (Bivalvia: Donacidae) and the problem of its sporadic detection in Bivalvia. *Evolution; international journal of organic evolution* 62:959–970.

Tritt A., Eisen JA., Facciotti MT., Darling AE. 2012. An integrated pipeline for de novo assembly of microbial genomes. *PloS one* 7:e42304.

Wheeler DL., Church DM., Federhen S., Lash AE., Madden TL., Pontius JU., Schuler GD., Schriml LM., Sequeira E., Tatusova TA., Wagner L. 2005. Database resources of the National Center for Biotechnology. *Nucleic Acid Research* 33:D39–D45.

Yuan Y., Li Q., Yu H., Kong L. 2012. The complete mitochondrial genomes of six heterodont bivalves (Tellinoidea and Solenoidea): variable gene arrangements and phylogenetic implications. *PloS one* 7:e32353.

Yuan S., Xia Y., Zheng Y., Zeng X. 2016. Next-generation sequencing of mixed genomic DNA allows efficient assembly of rearranged mitochondrial genomes in *Amolops chunganensis* and *Quasipaa boulengeri*. *PeerJ* 4:e2786.

680 Zouros E. 2013. Biparental Inheritance Through Uniparental Transmission: The Doubly
681 Uniparental Inheritance (DUI) of Mitochondrial DNA. *Evolutionary biology* 40:1–31.
682 Zouros E., Ball AO., Saavedra C., Freeman KR. 1994a. Mitochondrial DNA inheritance. *Nature*
683 368:818.
684 Zouros E., Oberhauser Ball A., Saavedra C., Freeman KR. 1994b. An unusual type of
685 mitochondrial DNA inheritance in the blue mussel Mytilus. *Proceedings of the National*
686 *Academy of Sciences of the United States of America* 91:7463–7467.