Structure, Transcription, and Variability of Metazoan Mitochondrial Genome: Perspectives from an Unusual Mitochondrial Inheritance System

Fabrizio Ghiselli[1],*, Liliana Milani[1], Davide Guerra[1], Peter L. Chang[2], Sophie Breton[3], Sergey V. Nuzhdin[2], and Marco Passamonti[1]

[1]Dipartimento di Scienze Biologiche, Geologiche ed Ambientali (BiGeA), Università di Bologna, Bologna, Italy

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

[3]Département de Sciences Biologiques, Université de Montréal, Montréal, Québec, Canada

*Corresponding author: E-mail: fabrizio.ghiselli@unibo.it.

Abstract

Despite its functional conservation, the mitochondrial genome (mtDNA) presents strikingly different features among eukaryotes, such as size, rearrangements, and amount of intergenic regions. Nonadaptive processes such as random genetic drift and mutation rate play a fundamental role in shaping mtDNA: the mitochondrial bottleneck and the number of germ line replications are critical factors, and different patterns of germ line differentiation could be responsible for the mtDNA diversity observed in eukaryotes. Among metazoan, bivalve mollusc mtDNAs show unusual features, like hypervariable gene arrangements, high mutation rates, large amount of intergenic regions, and, in some species, an unique inheritance system, the doubly uniparental inheritance (DUI). The DUI system offers the possibility to study the evolutionary dynamics of mtDNAs that, despite being in the same organism, experience different genetic drift and selective pressures. We used the DUI species *Ruditapes philippinarum* to study intergenic mtDNA functions, mitochondrial transcription, and polymorphism in gonads. We observed: 1) the presence of conserved functional elements and novel open reading frames (ORFs) that could explain the evolutionary persistence of intergenic regions and may be involved in DUI-specific features; 2) that mtDNA transcription is lineage-specific and independent from the nuclear background; and 3) that male-transmitted and female-transmitted mtDNAs have a similar amount of polymorphism but of different kinds, due to different population size and selection efficiency. Our results are consistent with the hypotheses that mtDNA evolution is strongly dependent on the dynamics of germ line formation, and that the establishment of a male-transmitted mtDNA lineage can increase male fitness through selection on sperm function.

Key words: doubly uniparental inheritance, mitochondrial intergenic regions, novel mitochondrial ORFs, germ line mitochondria, mitochondrial polymorphism, CORR.

Introduction

Since the symbiosis event that originated the eukaryotic cell, mitochondria underwent a massive process of genome reductive evolution (GRE) (Andersson and Kurland 1998; Khachane et al. 2007). The protomitochondrion (most likely an alpha-proteobacterion, for details see Müller and Martin 1999; Atteia et al. 2009; Abhishek et al. 2011; Thrash et al. 2011) lost the majority of its genome in a short evolutionary time, before the split of eukaryotic lineages, about 1,200 Ma (Khachane et al. 2007). After that, mitochondria coevolved with different hosts and experienced both neutral modifications and adaptive responses that led to the diversity that we observe today in mitochondrial genomes (mtDNAs) (Embley and Martin 2006). The most radical difference is between land plants and animals: plant mtDNAs are large and rich in non-coding sequences, while animal mtDNAs are more compact and much smaller. According to the mutation pressure theory (Petrov 2001; Lynch et al. 2006, 2011; Lynch 2007) genome evolution is shaped by mutation rate and random genetic drift. Nonfunctional intergenic DNA is mutationally

hazardous because, while it cannot suffer from loss-of-function mutations, it can be the substrate for gain-of-function deleterious mutations (Lynch et al. 2006, 2011; Lynch 2007). Thus, genomes with a high mutation rate are subject to a more intense selection for GRE, but the efficiency of this selection is determined by the amount of random genetic drift (i.e., effective population size, N_e). In taxa with reduced N_e, selection against the accumulation of nonfunctional DNA is less effective, and that would be the reason for the observed genome expansion during eukaryote evolution (Lynch 2007). As random genetic drift in plants and animals is similar, the difference in mitochondrial genome size can be explained by the much lower (~100×) mutation rate in plant mtDNAs compared with animal mtDNAs (Lynch et al. 2006).

In animal mitochondria, genomic features such as mutation rate, gene content, genome architecture, compositional properties, and gene strand asymmetry are variable among taxa, reflecting their different evolutionary histories (Gissi et al. 2008). A large number of studies attempted to unveil the reasons behind the different mutation rates among animal mtDNA lineages, investigating the relationship between such rates and body mass, metabolic rate, reactive oxygen species (ROS) production, and lifespan (see Galtier et al. 2009a for an overview). The matter remains unsolved, but there is clear evidence for a leading role of DNA replication on base-substitution mutations: despite its proof-reading function, most mutations arise from DNA polymerase errors (Drake et al. 1998; Lynch et al. 2006). Following this rationale, most of the heritable mutations are accumulated during germ line proliferation, when germ cells undergo several rounds of replication, and this implies that reproduction mode and gonad physiology affect evolutionary rates, as suggested by several authors (Rand 2001; Davison 2006). For example, in bivalve molluscs gametes are formed by proliferation of germinal cells in acini (Devauchelle 1990), the gonadic units containing the germinative tissue that lines the acinus wall. The gonad develops until it becomes fully mature then, after one or more spawning events, it is depleted. At the beginning of the following reproductive season, the spent gonad undergoes a period of reconstitution, and the cycle starts again (Gosling 2003). It follows that in bivalves the number of cell divisions in germ line does not show a marked asymmetry between males and females in contrast with what happens, for example, in mammals (Davison 2006). This feature, together with the production of an extremely large number of gametes due to broadcast spawning, implies a large number of cell divisions in both germ lines, resulting in a higher mutation rate in comparison to species that show male-driven evolution (Ellegren 2007). Actually, bivalves show an extraordinary amount of nucleotide polymorphism in both mitochondrial and nuclear genomes (Saavedra and Bachere 2006), and, in sharp contrast with deuterostomes which have almost invariant mitochondrial gene order (Gissi et al. 2008, but see

Gissi et al. 2010 for an exception), bivalves present highly rearranged mtDNAs, even at the intra-genus level. The association between polymorphism and gene order variability is not surprising: it is well established that sequence evolution and genome rearrangement are positively correlated (Begun and Aquadro 1992; Shao et al. 2003; Xu et al. 2006; Koonin 2009), even if the reasons behind this are still object of a heated debate (Begun and Aquadro 1992; Charlesworth et al. 1993; Nachman 2001). What is more surprising is the association, in bivalve mtDNAs, of a high mutation rate with the presence of quite large mitochondrial genomes.

An even more interesting feature of bivalves is the presence of an unusual mitochondrial inheritance system: the doubly uniparental inheritance (DUI; Skibinski et al. 1994; Zouros et al. 1994). So far, DUI has been detected in 46 bivalve species (Theologidis et al. 2008; Breton et al. 2011b), belonging to seven families. In DUI species, two mtDNAs are present: one is transmitted through eggs (F-type, for female-inherited), the other through sperm (M-type, for male-inherited), and the divergence between conspecific M and F genomes ranges from 10% to over 50% (see Breton et al. 2007 and Zouros 2012 for reviews).

In this work, we analyzed the mtDNAs of the DUI species *Ruditapes philippinarum* (Manila clam). The complete M and F genomes of *R. philippinarum* were submitted to GenBank in 2001 by Okazaki and Ueshima (Accession Nos.: AB065374 and AB065375, respectively), but a detailed characterization has not been published so far. We Sanger-sequenced the M and F major unassigned regions (URs), identifying the control regions (CRs) as well as motifs and secondary structures at both DNA and RNA level. Then, we obtained the M-type and F-type transcriptomes by RNA-Seq on Illumina GA IIx platform and performed a single-nucleotide polymorphism (SNP) analysis. Our main objectives were to 1) identify conserved functional elements and novel open reading frames (ORFs) that could explain the evolutionary persistence of intergenic regions in this species and other bivalves with DUI, 2) test, for the first time, if the mtDNA transcription in bivalves with DUI is lineage-specific and/or independent from the nuclear background, and 3) verify whether the male-transmitted and female-transmitted mtDNAs have a similar amount of polymorphism, and investigate the type of molecular variation occurring in the two mitochondrial lineages.

On a more general level, DUI systems can help understanding the complex relationship among multiple levels of selection and complex population dynamics that underlay mitochondrial genome evolution. Our data support the hypothesis that mtDNA evolution is strongly dependent on the dynamics of germ line formation, and suggest that the establishment of a male-transmitted mtDNA lineage can be beneficial, increasing male fitness through selection on sperm function.

Materials and Methods

Proportion of URs in Mitochondrial Genomes of Metazoans

In February 2012, 2,656 complete mitochondrial genomes were downloaded from the MitoZoa database release 10 (http://mi.caspur.it/mitozoa/ [last accessed August 2, 2013], Lupi et al. 2010; D'Onorio de Meo et al. 2012), and analyzed with custom Unix and R scripts to obtain the data shown in table 1. Given the marked difference in sample size among animal groups, to improve statistical power, we included in the analysis only taxa for which more than 60 complete mitochondrial genomes were available.

Gamete Collection and DNA Extraction

Gametes were collected from seven males and eight females using the procedure described in Ghiselli et al. (2011). Sperm samples were purified using a Percoll (GE Healthcare) gradient, as in Venetis et al. (2006). Egg samples were collected and centrifuged, then seawater was replaced with absolute ethanol. Total DNA was extracted from gametes with the DNeasy (Qiagen) and the MasterPure Complete DNA and RNA Purification Kit (Epicentre).

Polymerase Chain Reactions and Sequencing

DNA extractions were used as template for the polymerase chain reactions (PCRs): sperm extractions were used to obtain male largest unassigned region (MLUR) and male unassigned region 21 (MUR21) sequences, whereas eggs extractions for female unassigned region 21 (FUR21) and female largest unassigned region (FLUR). Primers were designed with Primer 3 (Rozen and Skaletsky 2000) on the complete *R. philippinarum* M and F mitochondrial genomes present in GenBank (Accession Nos.: AB065374-5; Okazaki M and Ueshima R, unpublished data). Primer pairs and their sequences are enlisted in supplementary table S1, Supplementary Material online. PCR amplifications were performed on a 2720 Thermal Cycler (Applied Biosystems) in a 50 μL reaction volume using the GoTaq Flexi Dna Polymerase

(Promega) kit. The reaction volume was composed of 24 μL of Nuclease-free Water (Ambion Inc.), 10 μL of Green GoTaq Flexi Buffer 5× (Promega), 6 μL of MgCl 25 mM, 1 μL of dNTPs (Promega) mix 40 μM (10 μM each dNTP), 2.5 μL of each primer (10 μM) (Invitrogen SRL), 4 μL of DNA template and 0.25 μL of GoTaq DNA Polymerase (Promega) 5 U/μL. PCRs were performed with the following cycle: an initial denaturation at 95 °C for 2 min, followed by 30 cycles of denaturation at 95 °C for 30 s, annealing at 48 °C for 30 s and extension at 72 °C for 90 s, then a final extension at 72 °C for 5 min. Every PCR product was checked by agarose gel electrophoresis. PCR products were purified using the Wizard SV Gel and PCR clean-up system (Promega) kit or the GenElute PCR clean-up kit and the GenElute Gel extraction kit (Sigma-Aldrich), following the manufacturer instructions. Sequencing was performed by Macrogen Inc. (Seoul, South Korea). Sequences were checked, aligned, and assembled manually using MEGA5 (Tamura et al. 2011).

Annotation of LURs

Ruditapes philippinarum largest unassigned regions (LURs) structure was defined using blastn (http://blast.ncbi.nlm.nih.gov/Blast.cgi, last accessed August 2, 2013) and with manual alignments. Repeat units were identified with Tandem Repeats Finder (http://tandem.bu.edu/trf/trf.html, last accessed August 2, 2013) (Benson 1999) and Repeat Finder (http://www.proweb.org/proweb/Tools/selfblast.html, last accessed August 2, 2013). ORFs in MUR21 and FLUR were identified with ORF Finder (http://www.ncbi.nlm.nih.gov/gorf, last accessed August 2, 2013) using the invertebrate mitochondrial genetic code.

Conserved Motifs

A search for conserved sequence motifs in *R. philippinarum* mt LURs and in 9 other Veneroid mt LURs (supplementary table S2, Supplementary Material online) was performed with MEME (Multiple Em for Motifs Elicitation; http://meme.nbcr.net/meme/cgi-bin/meme.cgi, last accessed August 2, 2013) (Bailey et al. 2009). The found motifs were submitted to

Table 1
Proportion of URs in the Mitochondrial Genomes of Metazoans

Taxa	N	Median Total Length	Median URs Length	Median %cod	Median %URs	Significance
Metazoa	2,656	16,544	1,047	93.4	6.6	n.s.
Chordata	1,852	16,606	1,062	93.6	6.4	n.s.
Arthropoda	415	15,587	945	93.9	6.1	n.s.
Nematoda	66	13,972	843	94.0	6.0	n.s.
Mollusca	134	16,195	1,311	91.9	8.1	n.s.
Gastropoda	49	15,129	258	98.3	1.7	***
Bivalvia	64	16,898	1,886	88.8	11.2	***

NOTE.—N, sample number; median total length, median total length of the mitochondrial genome; median URs length, median total length of the URs; median %cod, median proportion of coding regions in the genomes; median %URs, median proportion of URs in the genomes.
Significance, Wilcoxon rank-sum test significance: ***$P < 0.001$, n.s., nonsignificant.

GOMO (Gene Ontology for Motifs; http://meme.nbcr.net/meme/cgi-bin/gomo.cgi, last accessed August 2, 2013) (Buske et al. 2010), which assigned them a list of GO terms.

AT-Skew Analysis

To find indications on the location of the H-strand and L-strand origin of replication (O_H and O_L, respectively) in *R. philippinarum* mt genomes, we calculated the AT-skew values on 4-fold redundant sites of protein-coding genes, using the formula $(A + T)/(A - T)$. See Breton et al. (2009) for a detailed discussion. To support the findings, the analysis was extended to eight other Veneridae mt genomes (supplementary table S2, Supplementary Material online).

Secondary Structures

The mfold web server (http://mfold.rna.albany.edu/?q=mfold/download-mfold, last accessed August 2, 2013) (Zuker 2003) was used for DNA secondary structure prediction. The analysis was performed with default settings except for folding temperature: we used the value of 15 °C, which is the mean water temperature in the Venice lagoon during the reproductive season. Only the structures with the lowest ΔG value and showing conservation among the analyzed samples were selected.

The RNAz web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAz.cgi, last accessed August 2, 2013) (Washietl et al. 2005) was used for RNA secondary structure prediction. A window size of 200 bp and a window step-size of 100 bp were used. According to the software manual, alignments with $P > 0.5$ are classified as functional, and a negative z score indicates a stable structure. To avoid misinterpretation, we used a strict cutoff and excluded all the structures with a $P > 0.95$ and a z score < -4. Structure names legend: DS, DNA structure; RS, RNA structure; m, M-type; f, F-type.

Transcriptome Analysis

A cDNA library from 6 male and 6 female gonads was produced following the protocol of Mortazavi et al. (2008). The library was sequenced on an Illumina GAIIx platform with 71-bp paired-end reads. The samples were barcoded, pooled, and sequenced on two lanes (two technical replicates). For detailed information about sampling, library preparation, and sequencing see Ghiselli et al. (2012). Reads were mapped to the *R. philipinarum* complete mitochondrial genomes (GenBank Accession Nos. AB065374–5) and allowed up to six mismatches per end.

SNPs

We used the Genome Analysis Toolkit (GATK, McKenna et al. 2010) for base quality score recalibration, indel realignment, duplicate removal, and performed SNP and INDEL discovery and genotyping using standard hard filtering parameters or

variant quality score recalibration (DePristo et al. 2011). SNP effects were analyzed using the snpEff software (Cingolani et al. 2012).

Statistical Analysis

All data were checked for homoscedasticity, and variance stabilizing transformations were applied, where necessary, before tests. To show statistical dispersion, nontransformed data were used in boxplots. As in most cases data were not normally distributed, for uniformity we always applied nonparametric tests and, where not specified, P values are referred to the Wilcoxon rank-sum test. Statistical analysis and graphs were produced using R. Post hoc multiple comparison tests after Kruskal–Wallis nonparametric analysis of variance (ANOVA) were performed with the kruskalmc function (Siegel and Castellan 1988) implemented in the pgirmess R package.

Results

Proportion of URs in Mitochondrial Genomes of Metazoans

The analysis of 2,656 complete mitochondrial genomes present in the MitoZoa Database allowed us to assess the proportion of URs in several groups of metazoans (table 1). gastropods and bivalves have a proportion of URs which is significantly different from all other groups ($P < 0.001$): gastropods have the most compact mitochondrial genome, whereas bivalves show the highest median percentage of URs.

Structure of R. *philippinarum* Large URs

Figure 1 resumes the main features of the major URs in M- and F-type mtDNAs. Figure 1*a* shows conserved regions, motifs, repeated units, and major secondary structures (both at DNA and RNA level). Figure 1*b* shows transcription depth, nucleotidic variability (inter-lineage *p*-distance), and the ribbons link conserved region and motifs within and between major URs. Length of conserved blocks, repeated units, and motifs are included in supplementary tables S3–S6, Supplementary Material online.

Ruditapes philippinarum F and M mitochondrial genomes contain two major URs. In the M genome, the MLUR is located between *nd4L* and *tRNA-Ile* and it is preceded by the second largest UR (between *nd2* and *nd4L*), MUR21. In the F genome, the FLUR is located between *nd2* and *nd4L* and it is followed by the second largest UR (FUR21; between *nd4L* and *tRNA-Ile*). Overall, the two major URs (MUR21 and MLUR in M, FLUR and FUR21 in F) represent about 90% of the total amount of intergenic DNA in *R. philippinarum* mtDNAs. The obtained sequences are available in GenBank (FLURs: accession nos. KC243324-31; FUR21s: accession nos. KC243332-9; MLURs: accession nos. KC243340-6; MUR21s: accession nos. KC243347-53). The largest of these four URs is MLUR (3,588–3,610 bp), while the shortest is MUR21 (959 bp). FLUR
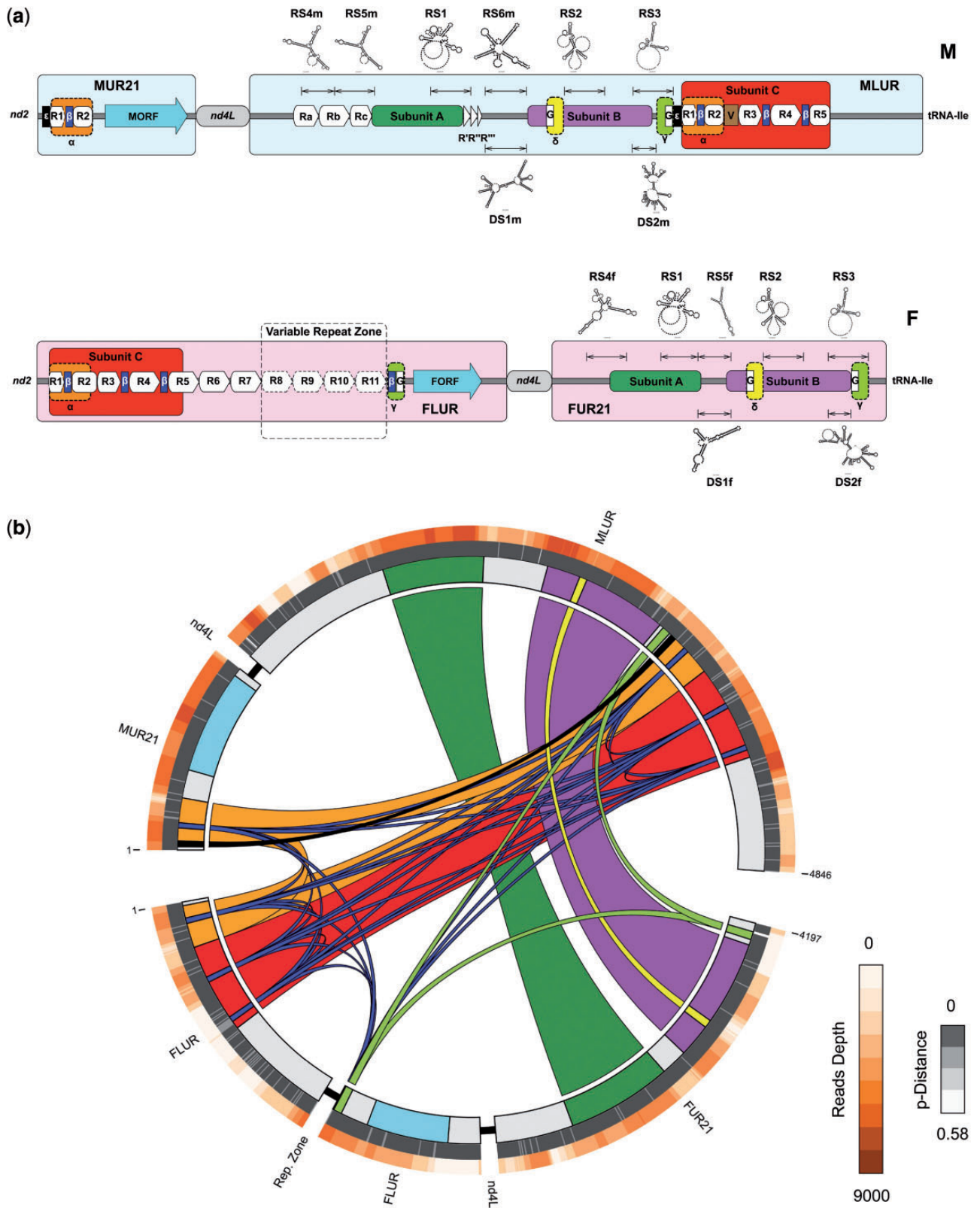
FIG. 1.—Features of *Ruditapes philippinarum* major URs. Main features of the largest URs in M- and F-type mtDNAs. (*a*) Conserved regions, motifs, repeated units and major secondary structures (both at DNA and RNA level). M, M-type mtDNA; F, F-type mtDNA; orange, motif α; turquoise, novel ORFs; dark green, subunit A; orchid, subunit B; red, subunit C; yellow, motif δ; light green, motif γ; black, motif ε; blue, motif β; G, G-homopolymer; Ra, Rb, Rc, R′,

(continued)

length is highly variable (from 2,185 to ~2,800 bp) due to a different number of repeated units (fig. 1a). FUR21 ranges from 1,767 to 1,771 bp.

Both MUR21 and FLUR contain, just upstream *nd4L*, a novel conserved lineage-specific ORF to which we will refer, from now on, as MORF and FORF, respectively (see supplementary figs. S1–S4 [Supplementary Material online] for nucleotidic and amino acid sequence alignments). MORF sequence is 519 bp long (172 aa), while FORF is 408 bp (135 aa). These sequences did not show any obvious homology with known proteins. To better understand origin and function of novel mitochondrial ORFs in DUI bivalves, an in-depth comparative analysis using multiple in silico approaches was performed in Milani et al. (2013).

Conserved Functional Motifs and Identification of Origins of Replication

We used the MEME suite and AT-skew analysis to identify molecular signatures of the origins of replication. Interestingly, two motifs, δ and γ (supplementary fig. S5a and b, Supplementary Material online), showed sequence similarity with motifs Sp1, Sp2, and Sp3 of the sea urchin *Strongylocentrotus purpuratus* mitochondrial CR (Jacobs et al. 1989; Cao et al. 2004). These motifs were found to be conserved also in the mitochondrial LURs of nine veneroid species (supplementary table S2, Supplementary Material online). Motif δ (40 bp; supplementary table S7 and fig. S5a, Supplementary Material online) corresponds to sea urchin Sp2 and the first part of Sp3 (P value = 7.32E−16), while motif γ (41 bp; supplementary table S8 and fig. S5b, Supplementary Material online) corresponds to a reversed segment of Sp1 (P value = 3.45E−10). A search with GOMO assigned to these two motifs a series of GO terms, many of which are related to transcription and DNA binding (supplementary table S9, Supplementary Material online). MEME also identified two motifs, β and ε (supplementary fig. S5c and d, Supplementary Material online), that are specific of *R. philippinarum*. β is present in both M- and F-type mtDNAs, while ε is M-type specific (fig. 1a). All the performed analyses failed to identify similarities with known motifs, therefore we are unable to assign a putative function to motifs β and ε.

AT-skew values, calculated in nine Veneridae species, are shown in supplementary table S10, Supplementary Material online. In *R. philippinarum*, F genome AT-skew values do not show any significant similarity with those of the other genomes, so comparisons cannot be made. As a general pattern, the genes with the highest values are those nearest to the LUR while the lowest-scoring genes are associated to the same three tRNAs, that is, *tRNA-His*, *tRNA-Glu*, and *tRNA-Ser*. *Paphia undulata*, *P. textile*, and *Meretrix lamarckii* mt genomes differ from this general scheme in only one of the aspects, whereas *R. philippinarum* F genome in both.

Secondary Structures

Supplementary table S11, Supplementary Material online, summarizes the principal features of DNA and RNA secondary structures, while supplementary table S12, Supplementary Material online, shows the detailed results of RNAz analysis. Four major DNA structures were identified, two in M-type and two in F-type (fig. 1). DS1m and DS2m (supplementary figs. S6 and S7, Supplementary Material online) in the MLUR, DS1f and DS2f (supplementary figs. S8 and S9, Supplementary Material online) in the FUR21. The most interesting features are 1) the terminal "b" loop of DS1m shows a TT/AA polymorphism; 2) the terminal "f" loop of DS1m shows a TGT/ACA polymorphism; 3) the "m" loop of DS2m and the "i" loop of DS2f have the same sequence (CGGTTTCAGAAG); and 4) the "l" loop of DS2m and the "h" loop of DS2f share the first 4 and the last 3 bases (*TAAG*TAAA*ACG* in the male, and *TAAG*GTY*ACG* in the female).

The analysis with RNAz identified 6 structures in the MLUR and 5 in the FUR21 (fig. 1a). Among them, three structures (RS1, RS2, and RS3, supplementary figs. S10–S12, Supplementary Material online) are conserved between M-type and F-type, three (RS4m, RS5m, and RSm6; supplementary figs. S13–S15, Supplementary Material online) are M-type specific and two (RS4f and RS5f; supplementary fig. S16 and S17, Supplementary Material online) are F-type specific.

Transcription of Mitochondrial Genomes

Overall, of the 90,233,244 sequenced reads, 9,895,466 (9.12%) mapped to mtDNA. Transcription mapping to the mitochondrial genomes is shown in the Circos diagram of figure 2, while supplementary figure S18, Supplementary Material online, shows the amount of mitochondrial reads: there is no significant difference in total amount of reads between males and females. The distribution of M-type and F-type transcripts is also shown in supplementary figure S18, Supplementary Material online: on average, 90.11% of the transcripts in male gonads are M-type. We found small traces (0.36%) of M-type transcripts in female gonads.

Fɪɢ. 1.—Continued
R″, R‴, M-type-specific repeats; R1-R11, Repeats; V, variable length spacer. (b) Circos diagram of the LURs of M- and F-type mtDNA showing transcription depth (orange gradient) and nucleotidic variability (inter-lineage p-distance, gray gradient) of the largest URs. The ribbons link conserved region and motifs within and between major URs. Nᴏᴛᴇ.—M-type above, F-type below. From the outside to the inside: transcription level (orange gradient scale 0–9000), p-distance (gray gradient scale 0–0.58), subunits and motifs with links between M-type and F-type. Orange, Motif α; turquoise, novel ORFs; dark green, subunit A; orchid, subunit B; red, subunit C; yellow, motif δ; light green, motif γ; black, motif ε; blue, motif β.
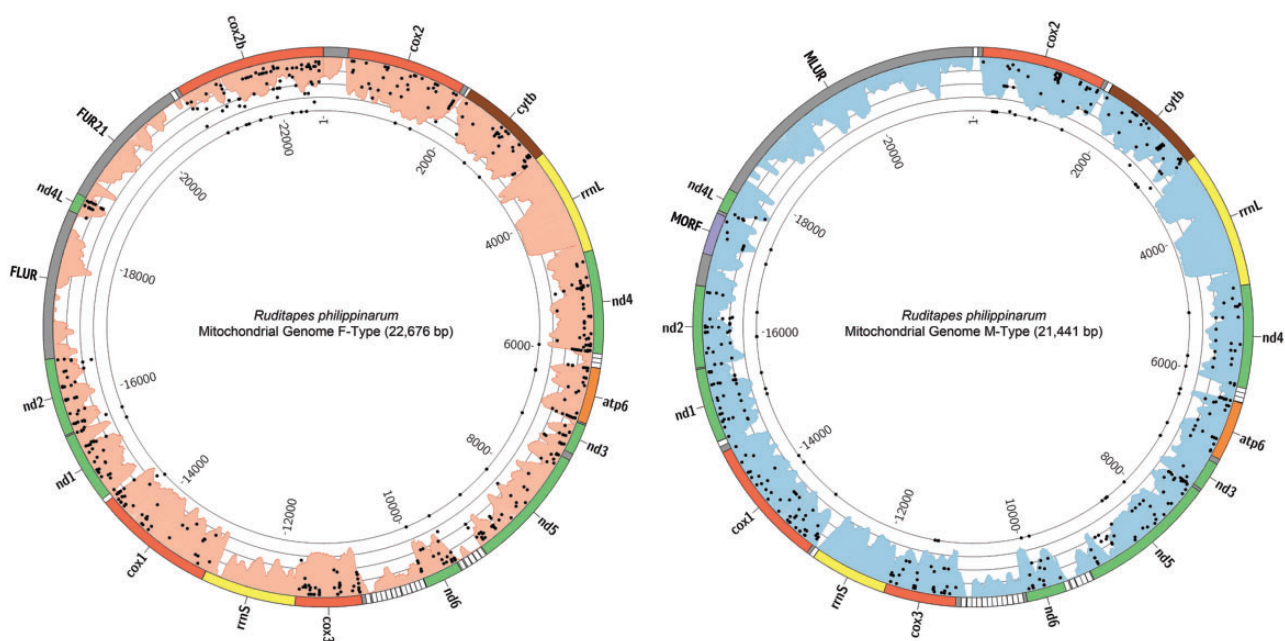
Fig. 2.—Circos diagrams of M-type and F-type transcriptomes. Transcription depth and SNPs mapped to the *Ruditapes philippinarum* mitochondrial genomes (GenBank accession nos.: AB065374 and AB065375). Genes are colored according to ETC complexes: green, complex I; brown, complex III; red, complex IV; orange, complex V. Ribosomal genes are colored in yellow, URs in gray, MORF in purple, and tRNAs in white. Histograms represent reads depth of F-type mtDNA (light red) and M-type mtDNA (light blue); black lines scale $0–4[\log_{10}^{-1}]$. Dots represent SNP position and frequency in protein coding genes; black lines scale 0–1.

The analysis of mitochondrial Coding Sequences (CDSs) revealed significant transcriptional differences: boxplots in figure 3a–c show the gene-by-gene transcription levels of M-type in males (black), F-type in females (white) and F-type in males (gray), while figure 3d compares the three transcription profiles. Wilcoxon rank-sum test was used to assess differential transcription between M-type (solid line with squares) and F-type in females (dashed line with circles). Spearman rank correlation test and Kendall tau test were used to assess the correlation between transcription of M-type (M), F-type in males (Fm), and F-type in females (F) (table 2). It is worth noting that F-type follows the same transcription profile independently from the nuclear background (i.e., the profile is the same in males and females; $\rho = 0.965$, $P < 0.001$; $\tau = 0.890$, $P < 0.001$).

Supplementary table S13, Supplementary Material online, shows the list of annotated nuclear-encoded ETC genes used in the analysis. The transcription of nuclear-encoded ETC genes is reported in figure 4a. No significant differences were found between males and females, except for genes of Complex III that show a slightly higher transcription in males ($P < 0.05$). Conversely, transcription of mitochondrially encoded ETC genes is always significantly different between M- and F-type, with the former being more transcribed for Complexes I and V, the latter for Complexes III and IV (fig. 4b).

The analysis of M-type mtDNA transcriptome showed that three mitochondrial coding genes (*nd4*, *nd5*, and *nd4L*) have a similar transcription level to MORF, and one (*nd2*) is less transcribed (supplementary table S14, Supplementary Material online; fig. 3a). On the contrary, FORF showed a very low transcription rate and its transcription level is significantly lower than all the F-mtDNA CDSs (supplementary table S15, Supplementary Material online; fig. 3b).

SNP Analysis

Table 3 reports SNP quality and coverage. In all the three mitochondrial genomes (F, Fm, and M) more than 93% of the SNPs exceeds a Phred score of 50. SNPs with Phred scores below 30 were not called. The coverage is high: only 8 SNPs (1.4% of the total) in the Fm genome and 2 SNPs (0.0048%) in the M genome have a depth less than 25. On the other side, the vast majority of the SNPs have a coverage less than 100× (97%, 92.9%, and 98.5% for F, Fm, and M genomes). Supplementary figure S19, Supplementary Material online, shows the scatter plot of the coverage against the number of SNPs (normalized to gene length). Spearman rank correlation test and Kendall tau test are not significant ($\tau = -0.18$, $P = $ ns; $\rho = -0.26$, $P = $ ns), supporting the absence of correlation between number of reads and number of called SNPs.

The kernel density plot of allele frequencies (fig. 5) evidences a different distribution between F and M mitochondrial genomes (Kolmogorov-Smirnov $P = 0.0061$): the F-type shows an excess of rare alleles (frequency < 0.125), while M-type has
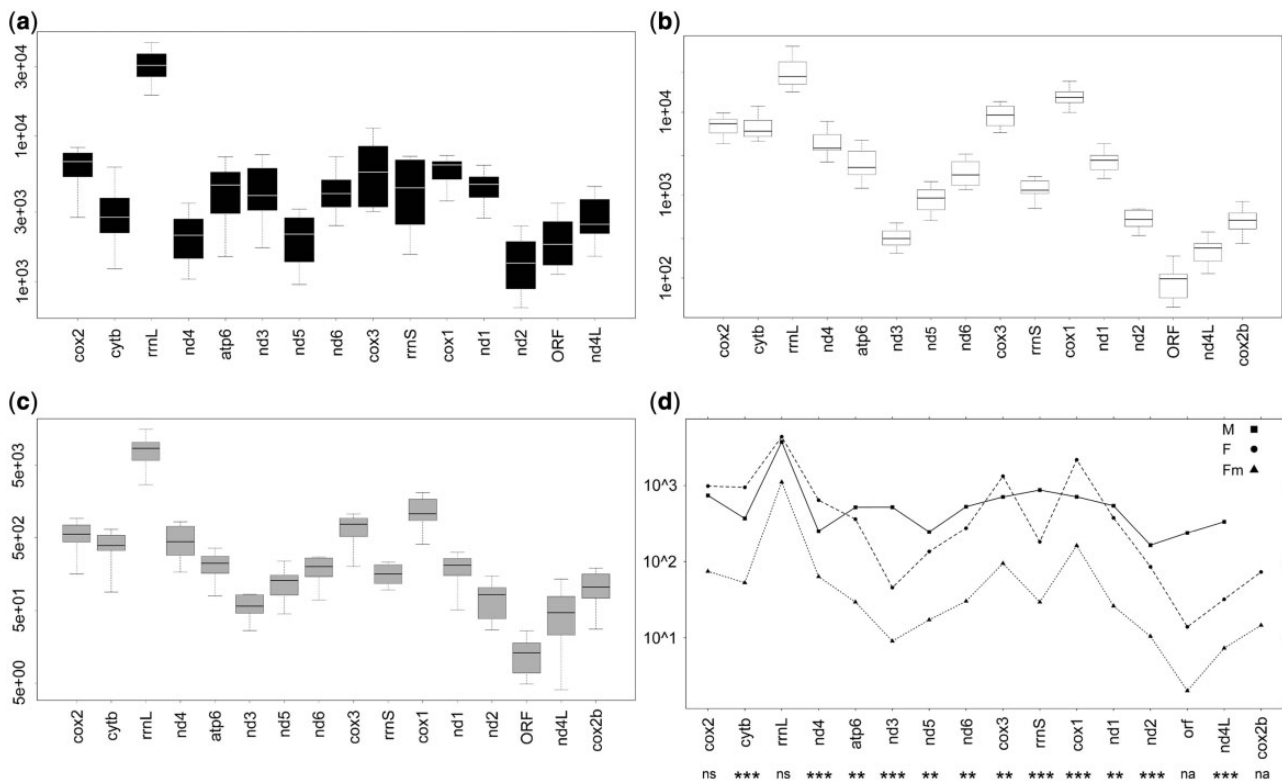
FIG. 3.—Transcription level of mitochondrial protein coding genes. (a) Transcription of M-type in males; (b) transcription of F-type in females; (c) transcription of F-type in males; and (d) transcription profiles (median values used). On the y axis is plotted the FPKM value. The lines that links the genes in (d) are virtual. Their purpose is to highlight the differences and similarities of transcription profiles. See table 3 for the correlation tests between mtDNA transcripts. In (d), the significance of Wilcoxon rank-sum test between M-type (M) and F-type in females (F) is reported below the x axis. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ns, nonsignificant; na, not applicable.

Table 2

Mitochondrial Transcription Correlation Tests

Test	Genomes	Significance	Notes
Spearman	M vs. F	*	$\rho = 0.600$
Spearman	M vs. Fm	*	$\rho = 0.600$
Spearman	F vs. Fm	***	$\rho = 0.965$
Kendall	M vs. F	*	$\tau = 0.451$
Kendall	M vs. Fm	*	$\tau = 0.429$
Kendall	F vs. Fm	***	$\tau = 0.890$

NOTE.—M, M-type mtDNA; F, F-type mtDNA; Fm, F-type mtDNA in males; ρ, Spearman's rank correlation coefficient; τ, Kendall tau rank correlation coefficient.
*$P < 0.05$.
***$P < 0.001$.

a pronounced peak around 0.5. The distribution in the Fm genome (not shown) is not statistically different from that in F.

Table 4 summarizes the SNP analysis. M-type has significantly less SNPs ($P < 0.001$) in comparison with both the F-types (F and Fm), which, conversely, do not differ between them (table 4 and fig. 6a). We subdivided the SNPs according to whether they are present with a single allele or multiple alleles in an individual. We called the former type "monoallelic SNP" and the latter "polyallelic SNP". Polyallelic SNPs have

always the reference allele among their variants. Compared with polyallelic SNPs, monoallelic SNPs have a lower proportion of nonsynonymous substitutions (Ns/Tot, table 4) in all the genomes ($P = 1.904E-7$). Boxplots in figure 6b and c show the proportion of nonsynonymous changes in polyallelic (fig. 6b) and monoallelic (fig. 6c) SNPs. The SNPs were subdivided in three classes according to their effect on genes (high, moderate, and low): boxplots in figure 6d–f show the proportion of the total amount of SNPs pertaining to each class, whereas in figure 6g–i only the monoallelic SNPs are considered.

Discussion

Bivalve mtDNAs Contain a High Proportion of URs

Bivalvian mtDNAs have, on average, 1.7× the amount of URs in respect to analyzed Metazoa (11.2% vs. 6.6%, $P < 0.001$; table 1). How does noncoding DNA accumulate in mitochondrial genomes? The principal mechanisms affecting mitochondrial genome structural evolution are 1) slipped-strand mispairing, 2) errors in termination of replication, 3) recombination, and according to the duplication–random loss model,
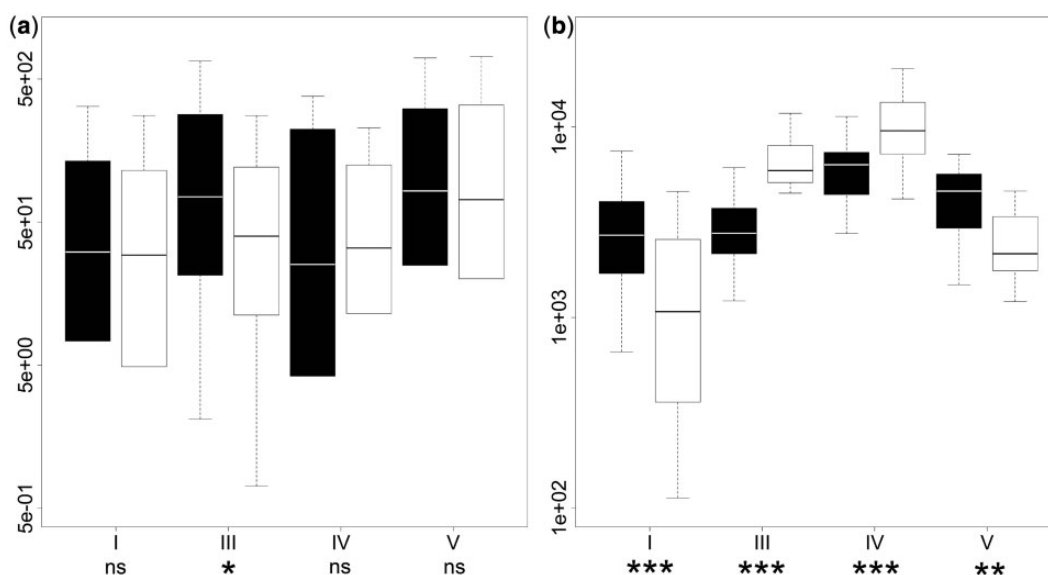
Fig. 4.—Transcription level of electron transport chain (ETC) genes. (a) Nuclear-encoded genes: black, male gonad; white, female gonad. (b) Mitocondrially encoded genes: black, M-type mtDNA; white, F-type mtDNA. I, III, IV, and V represents the ETC complexes: the analyzed genes and their accession numbers are enlisted in supplementary table S10–S13, Supplementary Material online. Complex II proteins are encoded only by nuclear genes, so they were not included in the analysis. On the y axis is plotted the FPKM value. Wilcoxon rank-sum test significance: *$P < 0.05$, **$P < 0.01$; ***$P < 0.001$.

Table 3
SNP Quality and Coverage

Genome	Phred Score						
	Min	Max	Mean	Median	30–40	40–50	>50
F	30	122,730	9,259.64	804.02	3.7%	3.1%	93.2%
Fm	30.23	126,287	12,170.75	623.44	3%	2.9%	94.1%
M	30.23	126,287	17,005.83	861.77	2.6%	3.6%	93.8%

Genome	Depth						
	Min	Max	Mean	Median	<25	25–100	100–1,000	>1,000
F	25	2,997	1,628.723	1,772	0%	3%	34.4%	62.6%
Fm	3	3,000	1,540	1,574	1.4%	5.7%	36%	56.9%
M	20	3,000	1,866	2,150	4.8E–3%	9.6E–3%	27.4%	71.1%

noncoding regions may arise from random pseudogenization of duplicated gene copies (Boore 2000).

The already mentioned high variability of gene order and the presence of duplicated genes (Ren et al. 2010; Passamonti et al. 2011; Okazaki M and Ueshima R, unpublished data) support the common occurrence of gene rearrangements in bivalve mitochondrial genomes. In particular, in bivalve species with DUI, mtDNA recombination is easily detectable, given the sequence divergence between M and F genomes (Ladoukakis et al. 2011 and references therein), and extensive rearrangements and duplications of parts of the CR have been well documented in Mytilus (Burzynski et al. 2003, 2006; Breton et al. 2006; Venetis et al. 2007; Cao et al. 2009). Recently, Ladoukakis et al. (2011) reported mitochondrial recombination between sequences with more than 20% divergence in the DUI species Mytilus galloprovincialis, showing that recombination is not restricted to sequences with low divergence. As explanation, the authors hypothesized a relaxation of the mismatch repairing system in animal mitochondria, but then, why are not mtDNA rearrangements more common in metazoans? Gissi et al. (2010) found hypervariability in ascidian mtDNA gene order, comparable only with that observed in molluscs. The only conserved feature among mitochondrial genomes of Tunicata is that all the genes are coded on the
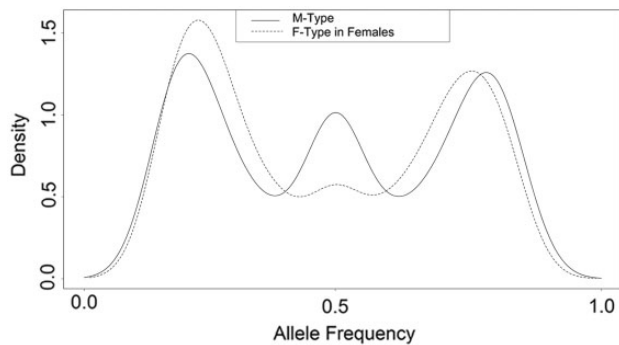
Fig. 5.—Kernel density plot of allele frequencies in mitochondrial CDSs. Probability density function of allele frequencies calculated by kernel density estimation. Solid line: M-type mtDNA; dashed line: F-type mtDNA in female gonads. The two distributions are significantly different (Kolmogorov-Smirnov $P = 0.0061$): the F-type shows an excess of rare alleles (frequency < 0.125), while M-type has a pronounced peak around 0.5. The distribution in the Fm genome (not shown) is not statistically different from that in F.

same strand, a feature that they share with all marine bivalves. Ren et al. (2010) suggested that coding on both strands could be a factor inhibiting recombination. Interestingly, among bivalves, freshwater mussels (family Unionidae) have dual-strand coding and show few mtDNA rearrangements with a proportion of URs that is much lower compared with the other species of the class (median in unionids = 7.9%, $N = 18$; median in other bivalves = 13%, $N = 46$; $P < 0.001$).

According to the Mutation Pressure theory, fast evolving organelle genomes are more exposed to a selective pressure for genome reduction. Bivalve mtDNAs seem to contradict this theory, because their hypervariability is coupled with a high percentage of intergenic DNA. But are bivalvian URs really nonfunctional? What if their retention in the genome is caused by the presence of functional sequences and/or structures?

Lineage-Specific Novel ORFs

Lineage-specific novel ORFs in DUI mtDNAs were already found in Mytilidae and Unionidae (Breton et al. 2009, 2010, 2011a, 2011b) and this is the first evidence from the family Veneridae. In the unionid, *Venustaconcha ellipsiformis* the translation of both FORF and MORF was demonstrated by Western blot (Breton et al. 2009), and the FORF protein was localized by immuno electron microscopy in both mitochondria and nucleus of the eggs (Breton et al. 2011b). A functional role of the lineage-specific mitochondrial ORFs identified in DUI bivalves was hypothesized: specifically, Breton et al. (2011b) proposed a role in germ line determination and maintenance of gonochorism. Given the tight association between the presence of M-type mtDNA and maleness, a role of DUI in sex differentiation was proposed

(reviewed in Passamonti and Ghiselli 2009; Zouros 2012), but whether this coupling is causative or associative is still matter of debate (Zouros 2012). It is worth noting that the influence of a mitochondrial ORF on germ line development is well documented in plants (Cytoplasmic Male Sterility, CMS; Chase 2007).

One might ask why do MORFs and FORFs need to be retained in the mtDNA and do not migrate in the nucleus. If these ORFs have a lineage-specific role (i.e., they are functionally linked to M- or F-type, and/or they represent some sort of tag) their migration to the nuclear genome would likely affect their function, especially considering that bivalves do not have sex chromosomes, or at least they are not morphologically distinguishable, thus they do recombine. Another possibility is that a nuclear copy of the ORFs exists and the mitochondrial copy will be lost as a result of selection (see Allen 2003, §4 g, point [iii]), even if our analyses do not indicate an accumulation of mutations in the ORFs.

Conserved Motifs and Origin of Replication

Figure 1*b* highlights the connections between subunits and motifs between the major URs of M- and F-type. We compared M and F mtDNAs to identify similarities and differences: similarities are supposed to be linked to a common physiological function (i.e., control of replication and transcription), whereas differences could be involved in the different "behavior" of the two mitochondrial lineages. Sequence alignments identified three conserved regions, subunit A, subunit B, and subunit C (fig. 1). From a functional point of view, subunits A and B and their neighboring regions seem to be the most interesting. Inside and right after subunit B are present two motifs (δ and γ, respectively), which show a strong conservation among the family Veneridae and, most importantly, with the sea urchin *S. purpuratus* (E value 5.5E−81 for δ and 9.4E−54 for γ; see Results and supplementary tables S7 and S8, Supplementary Material online, for details) whose CR has been characterized (Jacobs et al. 1989). Motifs δ and γ match elements of the sea urchin CR, which are homologous to the mammalian Conserved Sequence Blocks (CSBs) (Cantatore et al. 1989, 1990; Jacobs et al. 1989). CSBs have a fundamental role in the initiation of mtDNA replication, particularly in the formation of the R-loop, an RNA primer that is necessary for the formation of the D-loop and the start of H-strand synthesis (Scheffler 2008). The GOMO tool assigned GO terms related to transcription and DNA binding to both motifs δ and γ, further supporting their involvement in replication and transcription initiation, which are intimately linked in mitochondria (Scheffler 2008). Moreover, Cao et al. (2004) reported a match between some motifs found in the CR of the marine mussels *M. edulis* and *M. galloprovincialis* with the abovementioned elements of the sea urchin CR. All that considered, we can deduce that subunit B is close to O_H and that MLUR and FUR21 are the CRs of the M and F mitochondrial

Table 4

SNP Analysis Summary

Sample	Mis	Non	Syn	Indel	Tot	Ns/Tot	Ns_Freq	Hi	Mod	Low	%Hi	%Mod	%Low	#SNPs*	SNP_Freq*	Ns/Tot*	Ns_Freq*	%Hi*	%Mod*	%Low*
M-Type																				
mRPU1_p	122	6	35	1	164	0.787	1.018E−02	7	122	35	0.043	0.744	0.213	201	1.59E−02	0.70	1.11E−02	0.040	0.662	0.298
mRPU1_m	11	0	25	1	37	0.324	9.465E−04	1	11	25	0.027	0.297	0.676	—	—	—	—	—	—	—
mRPU2_p	86	4	29	1	120	0.758	7.178E−03	5	86	29	0.042	0.717	0.242	177	1.40E−02	0.62	8.60E−03	0.039	0.576	0.384
mRPU2_m	18	0	39	0	57	0.316	1.420E−03	2	16	39	0.035	0.281	0.684	—	—	—	—	—	—	—
mRPU3_p	105	5	38	1	149	0.745	8.755E−03	6	105	38	0.040	0.705	0.255	186	1.47E−02	0.67	9.86E−03	0.038	0.634	0.328
mRPU3_m	13	0	23	1	37	0.378	1.104E−03	1	13	23	0.027	0.351	0.622	—	—	—	—	—	—	—
mRPU10_p	104	4	38	1	147	0.741	8.598E−03	5	104	38	0.034	0.707	0.259	211	1.66E−02	0.59	9.86E−03	0.038	0.554	0.408
mRPU10_m	14	1	48	1	64	0.250	1.262E−03	3	13	48	0.047	0.203	0.750	—	—	—	—	—	—	—
mRPU11_p	117	3	30	1	151	0.801	9.544E−03	6	115	30	0.040	0.762	0.199	187	1.47E−02	0.72	1.06E−02	0.037	0.679	0.283
mRPU11_m	12	0	23	1	36	0.361	1.025E−03	1	12	23	0.028	0.333	0.639	—	—	—	—	—	—	—
mRPU12_p	114	5	46	1	166	0.723	9.465E−03	6	114	46	0.036	0.687	0.277	233	1.84E−02	0.59	1.09E−02	0.043	0.549	0.408
mRPU12_m	15	1	49	2	67	0.269	1.420E−03	4	14	49	0.060	0.209	0.731	—	—	—	—	—	—	—
Median_p	109.5	4.5	36.5	1	150	0.752	9.110E−03	6	109.5	36.5	0.040	0.712	0.248	—	—	—	—	—	—	—
Median_m	13.5	0	32	1	47	0.320	1.183E−03	1.5	13	32	0.031	0.289	0.680	—	—	—	—	—	—	—
Median_t*	—	—	—	—	—	—	—	—	—	—	—	—	—	194	1.53E−02	0.64	1.02E−02	0.039	0.605	0.356
Fm-type																				
mRPU1_p	120	5	96	4	225	0.573	9.194E−03	18	111	96	0.080	0.493	0.427	333	2.37E−02	0.54	1.43E−02	0.072	0.471	0.456
mRPU1_m	50	2	56	0	108	0.481	3.706E−03	6	46	56	0.056	0.426	0.519	—	—	—	—	—	—	—
mRPU2_p	132	3	74	3	212	0.651	9.835E−03	10	128	74	0.047	0.604	0.349	299	2.13E−02	0.59	1.38E−02	0.043	0.542	0.415
mRPU2_m	36	1	50	0	87	0.425	2.637E−03	3	34	50	0.034	0.391	0.575	—	—	—	—	—	—	—
mRPU3_p	124	5	66	3	198	0.667	9.408E−03	13	119	66	0.066	0.601	0.333	287	2.05E−02	0.57	1.30E−02	0.056	0.519	0.425
mRPU3_m	32	1	56	0	89	0.371	2.352E−03	3	30	56	0.034	0.337	0.629	—	—	—	—	—	—	—
mRPU10_p	100	5	106	3	214	0.505	7.697E−03	15	93	106	0.070	0.435	0.495	317	2.26E−02	0.51	1.27E−02	0.069	0.438	0.492
mRPU10_m	50	3	50	0	103	0.515	3.777E−03	7	46	50	0.068	0.447	0.485	—	—	—	—	—	—	—
mRPU11_p	101	4	72	3	180	0.600	7.697E−03	11	97	72	0.061	0.539	0.400	263	1.87E−02	0.53	1.10E−02	0.053	0.479	0.468
mRPU11_m	31	1	51	0	83	0.386	2.281E−03	3	29	51	0.036	0.349	0.614	—	—	—	—	—	—	—
mRPU12_p	110	6	61	2	179	0.659	8.410E−03	12	106	61	0.067	0.592	0.341	263	1.87E−02	0.58	1.21E−02	0.057	0.525	0.418
mRPU12_m	34	1	49	0	84	0.417	2.494E−03	3	32	49	0.036	0.381	0.583	—	—	—	—	—	—	—
Median_p	115	5	73	3	205	0.625	8.802E−03	12.5	108.5	73	0.066	0.566	0.375	—	—	—	—	—	—	—
Median_m	35	1	50.5	0	88	0.421	2.566E−03	3	33	50.5	0.036	0.386	0.579	—	—	—	—	—	—	—
Median_t*	—	—	—	—	—	—	—	—	—	—	—	—	—	293	2.09E−02	0.56	1.29E−02	0.056	0.499	0.441
F-type																				
frPU4_p	169	4	62	5	240	0.742	1.269E−02	22	156	62	0.092	0.650	0.258	263	1.87E−02	0.71	1.48E−02	0.091	0.622	0.285
frPU4_m	10	0	13	0	23	0.435	7.127E−04	2	8	13	0.087	0.348	0.565	—	—	—	—	—	—	—
frPU5_p	160	4	43	5	212	0.797	1.204E−02	23	146	43	0.108	0.689	0.203	232	1.65E−02	0.77	1.40E−02	0.108	0.659	0.233
frPU5_m	9	0	11	0	20	0.450	6.414E−04	2	7	11	0.100	0.350	0.550	—	—	—	—	—	—	—
frPU6_p	165	4	42	4	215	0.805	1.233E−02	19	154	42	0.088	0.716	0.195	315	2.25E−02	0.68	1.70E−02	0.082	0.600	0.317
frPU6_m	41	1	58	3	100	0.420	2.993E−03	7	35	58	0.070	0.350	0.580	—	—	—	—	—	—	—
frPU7_p	189	6	60	5	260	0.769	1.425E−02	28	172	60	0.108	0.662	0.231	318	2.27E−02	0.69	1.74E−02	0.097	0.594	0.308

(continued)

Table 4 Continued

Sample	Mis	Non	Syn	Indel	Tot	Ns/Tot	Ns_Freq	Hi	Mod	Low	%Hi	%Mod	%Low	#SNPs*	SNP_Freq*	Ns/Tot*	Ns_Freq*	%Hi*	%Mod*	%Low*
fRPU7_m	19	1	38	0	58	0.345	1.425E−03	3	17	38	0.052	0.293	0.655	—	—	—	—	—	—	—
fRPU8_p	175	6	52	4	237	0.781	1.319E−02	22	163	52	0.093	0.688	0.219	292	2.08E−02	0.70	1.62E−02	0.082	0.620	0.298
fRPU8_m	20	0	35	0	55	0.364	1.425E−03	2	18	35	0.036	0.327	0.636	—	—	—	—	—	—	—
fRPU9_p	149	3	55	5	212	0.741	1.119E−02	18	139	55	0.085	0.656	0.259	266	1.90E−02	0.67	1.40E−02	0.075	0.594	0.331
fRPU9_m	21	0	33	0	54	0.389	1.497E−03	2	19	33	0.037	0.352	0.611	—	—	—	—	—	—	—
Median_p	167	4	53.5	5	226	0.775	1.251E−02	22	155	53.5	0.092	0.675	0.225	—	—	—	—	—	—	—
Median_m	19.5	0	34	0	54.5	0.404	1.425E−03	2	17.5	34	0.061	0.349	0.596	—	—	—	—	—	—	—
Median_t*	—	—	—	—	—	—	—	—	—	—	—	—	—	279	1.99E−02	0.70	1.55E−02	0.087	0.610	0.303

Note.—_p, polyallelic SNPs; _m, monoallelic SNPs; _t, total; mis, missense; non, nonsense; syn, synonymous; tot, total number of SNPs; Ns/Tot, nonsynonymous to total ratio; Ns_freq, frequency of nonsynonymous SNPs (number of nonsynonymous SNPs normalized to CDS length); Hi, Mod, Low, high, moderate, low SNP effects; %Hi, proportion of high effect SNPs; %Mod, proportion of moderate effect SNPs; %low, proportion of low effect SNPs; No. of SNPs, total number of SNPs; SNP_Freq, SNP frequency (total number of SNPs normalized to CDSs length); *, values calculated on total number of SNPs (polyallelic and monoallelic).

genomes, respectively. To further support this hypothesis, we performed a comparative AT-skew analysis on complete mt genomes of *R. philippinarum* and other eight Veneridae (supplementary table S10, Supplementary Material online). The distribution of the values points to a location of the O_H inside the LUR, corroborating the hypothesis that this region is/contains the CR. The analysis also gave us clues about the localization of O_L that seems to be strictly associated to the presence of a conserved tRNA cluster composed by *tRNA-His*, *tRNA-Glu*, and *tRNA-Ser*. O_L is thought to be associated to secondary structures: given our findings, this tRNA cluster could provide such signal and a similar situation, where a three tRNA cluster may function as O_L, is also found in unionid mtDNAs (Breton et al. 2009). Interestingly, *R. philippinarum* F mtDNA did not show a pattern comparable with other genomes. Possible explanations for this incongruence may be: 1) variable locations of the origins of replication in these species, as observed by Breton et al. (2009) in unionid bivalves; 2) recent mtDNA rearrangements (Ovchinnikov and Masta 2012). At the present time, we do not have enough information to choose between these two options.

Secondary Structures

CRs are typically rich in secondary structures (Breton et al. 2009 and references therein) both at DNA and RNA level. There is clear evidence that secondary structures play a crucial role in biological processes such as DNA replication, transcription, recombination, repair, cleavage, control of gene expression, and genome organization (Pereira et al. 2008; Brázda et al. 2011). For instance, hairpins and cruciform structures can function as recognition sites for transcription factors, and their presence has been proved or inferred in many mitochondrial CRs (Cao et al. 2004; Mizi et al. 2005; Arunkumar and Nagaraju 2006; Pereira et al. 2008; Passamonti et al. 2011). When getting direct molecular evidence is not possible, secondary structures can be predicted in silico, and there are basically two methods to do it: free energy minimization and alignments of homologous sequences. In the first case, structures are inferred by calculating the variation of Gibbs free energy (ΔG) due to the folding of the nucleic acid (Zuker 2000). The structure with the lowest free energy is the most thermodynamically stable, therefore it is supposed to be the most common state of the molecule. The downsides are that nucleic acid sequences have more than one biologically active structure, and that the thermodynamic minimum is not always the actual conformational status of the sequence in vivo. The second method can provide information about evolutionary conservation improving the structure prediction accuracy (Xu and Mathews 2011). We utilized a combination of both approaches and found structures conserved between M mtDNAs, between F mtDNAs and shared by the two lineages (supplementary table S11 and figs. S6–S17, Supplementary Material online). The most significant

FIG. 6.—Boxplots of SNP polymorphism and SNP effects in F (white), Fm (gray), and M (black) mitochondrial genomes. (a) number of SNPs normalized to coding sequence (CDS) length; (b) nonsynonymous (Ns) SNPs to total number of SNPs ratio (polyallelic SNPs only); (c) nonsynonymous (Ns) SNPs to total number of SNPs ratio (monoallelic SNPs only); (d) percentage of high-effect SNPs (polyallelic + monoallelic); (e) percentage of moderate-effect SNPs (polyallelic + monoallelic); (f) percentage of low-effect SNPs (polyallelic + monoallelic); (g) percentage of high-effect SNPs (monoallelic only); (h) percentage of moderate-effect SNPs (monoallelic only); (i) percentage of low-effect SNPs (monoallelic only). NOTE.—A Kruskal–Wallis nonparametric ANOVA was performed. Significance levels of post hoc multiple comparison tests are reported below the x axis of each plot. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ns, nonsignificant.

structures were predicted in, or close to, the conserved regions of the CRs, subunits A and B. The low intra- and interlineage variability strongly support a functional role of such sequences, and can also be explained by a modulation of mutation rates by secondary structures: paired bases in double-stranded stem regions are less prone to mutations (Hoede et al. 2006). We identified two major DNA structures per lineage: DS1m and DS2m in M-type, DS1f and DS2f in F-type (fig. 1). The structures show lineage-specific differences with regard to shape and number of substructures (stem-loops and stacks). The M-type specific structure shows an interesting polymorphism

in two loops (TT/AA and TGT/ACA; supplementary table S11 and fig. S6, Supplementary Material online), whose function, if any, is unknown. The most notable feature of DS2m/f is the presence of an invariant sequence in the loop of a substructure (DS2m-m and DS2f-i), 26–30 bp upstream motif γ. Loop sequences are more vulnerable to mutations, and a 100% conservation among all sequenced M and F mtDNAs hardly can be labeled as coincidental. Moreover, the loop of another substructure (DS2m-l and DS2f-h) shows an inter-lineage sequence conservation, although partial: the central part of the loop is indeed different (TAAA in M-type and GTY in F-type).

As far as RNA is concerned, we found three structures (RS1, RS2, RS3; fig. 1) shared by the two mitochondrial genomes. RNAz alignments show a high inter-lineage conservation and multiple compensatory base changes in the stem regions (supplementary figs. S10–S17, Supplementary Material online), which suggest the functionality of the structures. The three structures are localized on the reverse strand; this is interesting especially in the case of RS3, which is formed in the same region as DS2m/f but on the opposite strand. Although being on the complementary strand, RS3 does not have the same folding as the corresponding DNA structure (DS2m/f), but notably the substructures m, i, l, and h are present also in RNA, forming a complementary copy. This is another clue pointing to some biological function for these substructures and for the conserved sequence that they carry in their loops. Our analysis also identified 5 lineage-specific RNA secondary structures: RS4m, RS5m, and RS6m in the M-type, RS4f, and RS5f in the F-type. RS4m and RS5m are very similar between each other because are formed in a region with repeated sequences (Ra, Rb, Rc, fig. 1). RS6m occupies the same position as DS1m, it forms on the opposite strand, and shares substructures e and f with the correspondent DNA structure. In the F mtDNA, upstream subunit A, only one RNA structure is present (RS4f). Finally, RS5f is in the same position of DS1f, it forms on the same strand and is quite similar to its DNA counterpart.

The presence of secondary structures showing inter-lineage conservation and forming in proximity of motifs that have a role in transcriptional/replicational control (δ and γ) suggests that they are probably involved in the same process.

Mitochondrial Transcription

Slightly more than 9% of the total number of reads mapped to mitochondrial DNA and no significant difference between males and females was detected (supplementary fig. S18, Supplementary Material online), meaning that the amount of mitochondrial transcripts in male and female gonads is approximately the same. Mitochondrial transcripts have different sources in males and females: males are heteroplasmic so their transcripts come from both M and F mtDNAs, while the only source of mitochondrial transcripts in females is the F-type. More specifically, on average, 90.11% of the transcripts in male gonads are from M mtDNA, while the remaining 9.89% are F-type transcripts (supplementary fig. S18, Supplementary Material online). This result is expected given that, in this species, M-type is always strongly predominant in male gonads (Ghiselli et al. 2011), thus the main reason for the difference in transcription is probably the different mtDNA copy number. Our analyses showed small traces of M-type transcripts in female gonads (0.36%) which can be explained in two ways: by a small amount of cross-contamination between samples, and/or by the actual presence of M mtDNA in female gonads, which can occur sometimes (Ghiselli et al.

2011). Given the exiguous amount (and thus the nonsignificant statistical weight), these reads were treated as contamination and excluded from the analyses. Thus, three types of mitochondrial genomes (and their transcripts) were considered: 1) M-type, which is localized in male gonads and that can be inherited by male progeny through sperm; 2) Fm-type, which is the F-type present in male gonads and that is an evolutionary dead-end because is not transmitted to progeny (Ghiselli et al. 2011); 3) F-type, which is localized in female gonads and that can be inherited by both male and female progeny through eggs.

mtDNA is transcribed as a polycistronic primary transcript which is edited to form mRNAs, but this does not mean that mitochondrial genes have always the same relative expression level, since differential expression is achieved by post-transcriptional control (Lynch 2007; Scheffler 2008). We generated the RNA-Seq library selecting polyadenylated transcripts, so our analysis only includes transcripts that underwent an editing phase.

Autonomous Regulation of Mitochondrial Expression

Figure 3 shows the transcriptional differences among mitochondrial genes in M-type (black, fig. 3a), F-type (white, fig. 3b), and Fm-type (gray, fig. 3c). In figure 3d, the three transcriptional profiles are compared: with the exception of cox2 and rrnL, the transcription is always significantly different between M and F (solid line and dashed line, respectively, see P values below x axis). Fm transcription (dotted line) is obviously significantly lower in respect to both M and F, but shows an interesting feature: its transcriptional profile is almost identical to that of F (Spearman's correlation coefficient $\rho = 0.965$, $P < 0.001$; Kendall's correlation coefficient $\tau = 0.890$, $P < 0.001$; table 2). Except for a small difference in Complex III, the transcription level of nuclear-encoded ETC genes does not change between male and female gonads (fig. 4a), whereas the mitochondrially encoded ETC genes have always a significantly different transcription (fig. 4b). Taken together, these observations are consistent with the hypothesis of CO-location for Redox Regulation (CORR; Allen 2003). The aim of Allen's hypothesis is to explain the retention of genes in cytoplasmic organelles: it states that mitochondria and chloroplasts retained genes whose expression need to be under direct regulation of the redox state of their products or of electron carriers with which their products interact. This permits "direct and autonomous redox regulation of gene expression" (Allen 2003). The fact that M and Fm show different transcription profiles under the same nuclear environment (male gonad), is consistent with a regulation operated by mitochondrial components. Moreover, our data about transcription of nuclear-encoded ETC genes (fig. 4a) match a prediction of the CORR hypothesis: the nucleus would provide a fairly constant pool of transcripts producing mitochondrial precursor proteins, ready to be imported in the

mitochondrion following the "decision" of the organelle genome (Lane 2007).

Lineage-Specific Transcription and M-Type Bioenergetic Activity

To explain the observed transcriptional differences between M- and F-type mtDNAs, we propose two hypotheses. 1) According to several Authors (Zouros 2012) M genome could be a selfish or "nearly selfish" element that found a way to be inherited through sperm. Under this light, the transcription profiles shown in figure 3d could support this: the "regular" transcription in R. philippinarum gonad would be that showed by F and Fm, while M would be less coordinated with nuclear factors, therefore showing a different transcription pattern. 2) According to the mitochondrial theory of ageing (Allen 1996) there is a division of labor between female and male germ line mitochondria. The former have a repressed bioenergetic function to prevent mutagenesis caused by ROS production thus facing only mutations due to replication errors. On the other side, male germ line mitochondria are bioenergetically active (their energy is needed for spermatozoa movement), thus more prone to mutagenesis by ROS. Therefore, in gametes there is a tradeoff between motility and fidelity of mtDNA transmission, implying that mitochondria that become bioenergetically functional are genetically disabled (Allen 1996). Recently, de Paula et al. (2013) found evidence supporting the hypothesis that oocyte mitochondria are quiescent in the jellyfish Aurelia aurita and discussed the Weismann barrier in germ line mitochondria. The mitochondrial theory of ageing and de Paula et al. (2013) results support the continuity of mitochondrial germ plasm (i.e., that acquired mitochondrial mutation is not inherited; see de Paula et al. 2013, fig. 7e). It is clear that DUI represents an interesting system to test the mitochondrial theory of ageing, as it seems that M mtDNA is breaking the rule of mitochondrial germ line continuity. Our results show a significantly different transcription pattern between M and F mtDNAs (figs. 3d and 4b), but they cannot support the quiescence of oocyte mitochondria. Indeed (fig. 3d), even if seven protein coding genes showed a higher transcription in M (atp6, nd3, nd5, nd6, nd1, nd2, and nd4L), four showed a higher transcription in F (cytb, nd4, cox3, and cox1) and one showed no significant difference (cox2). In contrast, de Paula et al. (2013) found a marked difference in mitochondrial transcription between testis and ovary, even though the analysis was made on three genes (nd1, cytb, and cox1). This work cannot be conclusive about this subject, and further analyses (e.g., membrane potential, ROS content and transcription in somatic tissues) are needed to better assess the activity of the two types of mitochondria in R. philippinarum.

In DUI organisms, M mitochondria are transmitted through sperm to male progeny, thus playing both the roles of energy-transducers and genetic templates. How can they escape the ROS-induced mutagenesis affecting bioenergetically active organelles? Bivalve molluscs habitats (i.e., sediments and intertidal environments) are subject to recurring hypoxia or anoxia. Along with several marine invertebrates, M. edulis (a DUI species) has been found to have facultatively anaerobic mitochondria capable of malate dismutation, a metabolic pathway (common to most parasitic helminths) that produce ATP through degradation of carbohydrates (reviewed in Müller et al. 2012). Such pathway of facultative anaerobic metabolism in M. edulis bypasses the ETC Complexes II, III and IV, thus reducing ROS production. Interestingly, our data show that, compared with F-type mtDNA, M-type transcription is lower for Complexes III and IV and higher for complex I and V (fig. 4b). We speculate that, to reduce ROS production in male germ line, M-type mitochondria in R. philippinarum might use malate dismutation as an alternative way to produce ATP. We think that this working hypothesis deserves further investigation.

MORF Is Transcribed

Our data support the functionality of the MORF. Not only the sequence is conserved among all the analyzed males and does not show indels or stop codons, but it is also transcribed at a level which is comparable with that of the other typical mitochondrially encoded ETC genes (fig. 3a and d; supplementary table S14, Supplementary Material online). On the other hand, FORF shows a very low level of transcription, the lowest among F-type mtDNA genes (fig. 3b and d; supplementary table S15, Supplementary Material online); therefore, we are inclined to believe that it is not functional, or that it is transcribed in a different developmental stage.

The cox2 Duplication

The F genome contains a duplicaton of the cox2 gene, named cox2b (fig. 2), a feature that has been also observed in the M-type mtDNA of another DUI species, Musculista senhousia (Passamonti et al. 2011). The two copies have different length: the shortest, cox2, is 1,569 bp long (523 aa), while the longest, cox2b, is 1,971 bp long (657 aa). They have also a markedly different transcription level (fig. 3b and d), so, for all these reasons, we think that cox2b is undergoing a pseudogenization process, or that it is not functioning as a cytochrome oxidase subunit 2 anymore. In the M genome of the freshwater mussel V. ellipsiformis, the cox2 gene has a 555 bp coding extension that has been hypothesized to have a reproductive function (Breton et al. 2007 and references therein). Whether R. philippinarum cox2b underwent a neofunctionalization process acquiring a similar function will be matter of future investigations.

Amount of Polymorphism

The fast-evolving nature of bivalve mtDNA is a well known feature, but the underlying mechanisms are not. In DUI

species, the M-type mtDNA always showed a higher amount of variation in respect to the F-type mtDNA (Zouros 2012 and references therein) except in *M. senhousia* where the opposite pattern was observed, probably due to a historical effect of its introduction in the Adriatic Sea (Passamonti 2007). These observations led to the hypothesis of a faster evolution of the M-type mtDNA, confirmed by several studies in which comparisons of whole mitochondrial genomes were used (Mizi et al. 2005; Breton et al. 2006; Zbawicka et al. 2010; Doucet-Beaupré et al. 2010). Here, for the first time, we used a high-throughput approach to assess the amount and the type of polymorphism in the gonadal mitochondrial populations. Given that the vast majority of gonadal mtDNAs are localized in gametes, this analysis is useful to estimate, both quantitatively and qualitatively, the standing genetic variation of the mitochondrial population that is going to be transmitted to the progeny. The high coverage (table 3) allowed us to detect rare alleles, and the RNA-Seq protocol gave us the chance to avoid PCR-based methods: in a situation were DNA sequences are highly polymorphic, PCR primers fail to amplify mutated targets, leading to an underestimation of the actual variability (see Theologidis et al. 2008 for a detailed discussion).

Figure 5 shows the distribution of allele frequencies in M and F. F-type mtDNAs show an U-shaped distribution, with a low proportion of intermediate-frequency alleles and a high proportion of rare alleles. The abundance of low frequency variants causes a shift of high frequency alleles towards a slightly lower frequency class. The distribution in M-type is significantly different (Kolmogorov-Smirnov test, $P = 0.0061$), with a lower proportion of rare alleles and a much higher proportion of mid-frequency alleles. The different amount of low-frequency alleles can be explained in term of bottleneck size. During its inheritance route, mitochondrial population is subject to a dramatic reduction followed by a massive expansion (see Ghiselli et al. 2011 and Milani et al. 2011 for discussions about mitochondrial bottleneck in DUI animals). After a population shrinkage, rare alleles are quickly eliminated while intermediate and high-frequency alleles are preserved (Maruyama and Fuerst 1984, 1985). Because of the higher number of mtDNAs in eggs compared with sperm (~10× in this species; see Ghiselli et al. 2011), F-type mtDNAs experience a wider bottleneck, therefore the larger population size is compatible with a higher amount of low frequency allele. The above-mentioned rationale also explains the persistence of intermediate-frequency alleles in M despite the narrower bottleneck since intermediate-frequency alleles are less likely to be eliminated by drift and more likely to be fixed by selection (Olson-Manning et al. 2012). Although population size effects can account for the loss of rare variants they cannot justify the difference in mid-frequency alleles between M- and F-type, whose explanation might be found in a different action of natural selection. It is well known that mitochondrial genomes evolve mainly under purifying selection

(Rand 2001; Meiklejohn et al. 2007; Galtier et al. 2009b), nonetheless, deviations from the negative selection regime have been reported in gynodioecious plants (Galtier et al. 2009b and references therein). Gynodioecy is a form of sexual dimorphism in which females and hermaphrodites co-exist in the same population (Couvet et al. 1998); in this system, gender is determined by epistatic interactions between mitochondrial and nuclear loci, a mechanism known as Cytoplasmic Male Sterility (CMS, see Chase 2007 for a review). In CMS mitochondrial ORFs produce chimeric proteins which cause pollen sterility, and this process can be counteracted by one or more nuclear restorer-of-fertility genes. The ongoing conflict between CMS mitotypes and nuclear restorers leads to long-term balancing selection, as observed in several CMS species (Gouyon et al. 1991; Couvet et al. 1998; Houliston and Olson 2006). Even if at speculation level, the DUI system presents some intriguing resemblances with CMS, and the distribution pattern of allele frequency in M-type mtDNA is an additional similarity that deserves to be further investigated.

Figure 6a reports the total number of SNPs: F and Fm show a higher number of SNPs (with no significant difference between them) in respect to M ($P < 0.001$). Taken together with the allele frequency data, this piece of information indicates a different kind of polymorphism between egg-transmitted (F and Fm) and sperm-transmitted (M) mtDNAs. F and Fm show more variable sites and rare alleles, while M shows a lower number of variable sites but with a higher proportion of alleles with intermediate frequency. This means that F and Fm variability has been underestimated until now: a large part of the polymorphism has been hidden, given the difficulties in amplifying and sequencing rare alleles with PCR-based methods.

Type of Polymorphism

There is a large number of mitochondria in every cell, and each mitochondrion has multiple copies of mtDNA. In such conditions, it is difficult to understand how much a deleterious mutation affects the biological function of an organelle (see Rand 2001 for a review on multi-level selection on mtDNA). The high ploidy of mtDNA in a cell implies that functional copies of the genes can buffer the malfunctioning or nonfunctioning copies, practically slowing down the action of natural selection on deleterious alleles. Selection acts on mitochondria through the autophagy process, which eliminates damaged and old organelles (mitophagy, see Youle and Van Der Bliek 2012). If natural selection is partially blinded by the buffering effect of multiple copy numbers, we expect a high amount of nonsynonymous polymorphism to exist in mitochondrial populations. This is actually what we observed: the median ratio of nonsynonymous to total number of SNPs is 0.64 in M, 0.56 in Fm, and 0.70 in F (Ns/Tot*, table 4). An even more clear indication of the buffering process comes from the comparison of nonsynonymous polymorphism between polyallelic and

monoallelic SNPs. We defined polyallelic those SNPs which are present with multiple alleles within an individual, and monoallelic those which have a single allele. Monoallelic SNPs have always a lower proportion of nonsynonymous changes ($P = 1.904E−7$): the percentage drops from 75.2% to 32% (2.35×) in M, from 62.5% to 42.1% (1.48×) in Fm and from 77.5% to 40.4% (1.91×) in F (Ns/Tot, table 4). Deleterious monoallelic SNPs cannot be buffered by alternative functional alleles, so the probability of their persistence in the population is lower. Reinforcing this concept, we observed that polyallelic SNPs had always the functional allele among their variants. Figure 6b and c show that the drop of nonsynonymous polymorphism between polyallelic and monoallelic SNPs is different in the three mtDNAs. M and F have a higher amount of nonsynonymous polyallelic SNPs, in comparison with Fm ($P < 0.05$ and $P < 0.01$, respectively), but their nonsynonymous polymorphism is more strongly reduced in monoallelic SNPs. Interestingly, the reduction is higher in the M-type (2.35×, table 4 and figure 6b and c).

To better understand the type of sequence variation in our mitochondrial populations, we analyzed the SNP effects, which were subdivided in three classes by the snpEff software (Cingolani et al. 2012). The high effect class includes nonsynonymous mutations that likely can provoke a loss of function (start lost, frameshift, nonsense, stop lost, and rare amino acid). Medium effect SNPs are also nonsynonymous substitutions, but not as disruptive as those in the previous class: they cause alterations that probably entail a lower functionality of the protein, but that can be tolerated (codon change, codon insertion, and codon deletion). In some cases, the functionality could also be improved, but the occurrence of advantageous mutations is obviously rare. Finally, low effect SNPs is substantially synonymous changes (synonymous start, nonsynonymous start, start gained, synonymous coding, and synonymous stop). The percentage of high, moderate, and low effect SNPs in the three genomes always follows the same pattern: moderate effect substitutions are the most common (%Mod*, table 4), low effects have an intermediate proportion (%Low*, table 4) and finally, as expected, high effect SNPs are the rarest (%Hi*, table 4). The abundance of moderate effects in respect to low effects is more marked in F and M (figure 6e and f; table 4), compared with Fm. This could be the result of the larger number of replications of germ line mtDNAs (Fm is not inherited so it does not undergo the same rounds of replication of the other mtDNAs): nonsynonymous mutations are more frequent and, if buffered by functional copies, their effect is small and they are not purged by selection. High-effect substitutions are more dangerous, therefore more subject to selection and for this reason are the rarest class.

We performed the same analysis also on monoallelic SNPs: due to the lack of buffering effect, selection is more effective on nonsynonymous mutations and this is reflected by the percentages of high, moderate and low effect substitutions.

Indeed, in monoallelic SNPs the most common class is the low-effect followed by moderate and high (% Low, % Mod, and % Hi; table 4), that is, synonymous substitutions are the most common. Compared with polyallelic SNPs, both high- and moderate-effect classes drop their percentages in monoallelic SNPs (fig. 6g and h; table 4).

Overall, our data are consistent with a lower amount of deleterious polymorphism in M-type in comparison with F (fig. 6d and g), that can be explained by a different efficiency of selection on gametes. The presence of hundreds of F mtDNAs in eggs entails a strong buffering effect on deleterious mutations which are complemented by wild-type alleles. R. philippinarum spermatozoa carry only four mitochondria (Milani et al. 2011), corresponding to a few dozen mtDNAs (Ghiselli et al. 2011), thus the buffering effect is much weaker, and deleterious mutations are more exposed to selection. After spawning, M-type mitochondria are subject to an intense selection since only the most viable spermatozoa can fertilize an egg and produce a healthy embryo. This leads to an unusual situation in which a smaller population size results in a more efficient selection.

Conclusions

The high amount of URs in the fast-evolving mtDNAs of bivalves seems at first to elude the evolutionary pressure towards a reduction of the genome size. The main causes for the origin of extragenic sequences in mtDNAs are slipped-strand mispairing, errors in termination of replication (Boore 2000) and recombination (Ladoukakis 2011 and references therein). The extraordinary variability in gene arrangement and the presence of gene duplications suggest that such mechanisms are particularly active in bivalves, and the elevated mutation rate plus the low efficiency of the DNA mismatch repair system could be the underlying reasons. Although the origin of intergenic DNA is due to completely stochastic processes, its persistence is probably adaptive: the presence of sequences, motifs, secondary structures with a regulatory role, and transcribed ORFs with a still unknown function, can prevent the loss through GRE. Under this view, the redundancy generated by duplications and/or acquisition of extra sequences allowed the evolution (gain-of-function) of mitochondrially encoded factors possibly interacting with the extramitochondrial environment and the nucleus by means of retrograde signaling. In DUI bivalves, these factors could be responsible for the unusual inheritance system and for the different transcriptional behavior of the two organelle lineages.

It has been established that, among the fast-evolving bivalian mtDNAs, M-type of DUI species is the fastest. From our data, it is clear that M and F are actually pretty close as far as the amount of polymorphism goes. This means that the higher evolutionary rate of M is not caused by the higher polymorphism in germ line mitochondria. If M existence is

just the effect of the acquired ability to invade male germ line, M would have to carry out its biological functions only when F cannot do it (i.e., in male gonad). Following this rationale, some Authors proposed that M faster evolutionary rate could be explained by a relaxed selection due to the reduced biological role (Zouros 2012). Even if this is true, we argue that the remaining function of M-type is an extremely important one (i.e., the contribution to gamete production and functionality), and a relaxed selection would affect gamete fitness. Indeed, even a modest reduction of energy production by mitochondria is known to reduce male fertility, and a decrease in male fitness reduces the viability of the population (Gemmell and Allendorf 2001; Meiklejohn et al. 2007). From our data on SNP effects, M has the lowest proportion of nonsynonymous polymorphism (fig. 6), particularly in the high-effect class and in monoallelic SNPs, and this is not in agreement with a relaxed selection. An alternative scenario would be that M has a function in sperm and/or spermatogenesis: sex and reproduction-related genes evolve rapidly (Ellegren and Parsch 2007; Parsch and Ellegren 2013) and a co-evolution between nuclear and mitochondrial factors involved in spermatogenesis could be the engine of M-type mtDNA fast evolution. Another hypothesis to explain M evolutionary rate is sperm competition, a particularly strong phenomenon in broadcast spawning animals (Palumbi 2009). DUI is the only known biological system in which a mtDNA can be under selection for male functions. In species with strict maternal inheritance of mitochondria, deleterious mutations that affect only males are not subject to natural selection (Gemmell and Allendorf 2001; Gemmell et al. 2004), so mtDNA mutations can reduce male fertility without effects on females. On the contrary, in DUI species natural selection can work on M mtDNA and this could increase male fitness and be beneficial for the entire species. From this point of view, the high proportion of intermediate-frequency alleles in M can be seen as a good predictor of its evolutionary potential: rare alleles do not contribute to the immediate response to selection, but intermediate-frequency alleles do (Allendorf 1986). Under this light, even if DUI arose for nonadaptive reasons, its maintenance would be selectively advantageous.

Supplementary Material

Supplementary figures S1–S19 and tables S1–S15 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

Literature Cited

Abhishek A, Bavishi A, Bavishi A, Choudhary M. 2011. Bacterial genome chimaerism and the origin of mitochondria. Can J Microbiol. 57: 49–46.

Allen JF. 1996. Separate sexes and the mitochondrial theory of ageing. J Theor Biol. 180:135–140.

Allen JF. 2003. The function of genomes in bioenergetic organelles. Philos Trans R Soc Lond B Biol Sci. 358:19–37.

Allendorf FW. 1986. Genetic drift and the loss of alleles versus heterozygosity. Zool Biol. 5:181–190.

Andersson SG, Kurland CG. 1998. Reductive evolution of resident genomes. Trends Microbiol. 6:263–268.

Arunkumar KP, Nagaraju J. 2006. Unusually long palindromes are abundant in mitochondrial control regions of insects and nematodes. PLoS One 1:e110.

Atteia A, et al. 2009. A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the alpha-proteobacterial mitochondrial ancestor. Mol Biol Evol. 26:1533–1548.

Bailey TL, et al. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37:202–208.

Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature 356:519–520.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27:573–580.

Boore JL. 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. In: Sankoff D, Nadeau JH, editors. Comparative genomics. Amsterdam (The Netherlands): Springer. p. 133–147.

Brázda V, Laister RC, Jagelská EB, Arrowsmith C. 2011. Cruciform structures are a common DNA feature important for regulating biological processes. BMC Mol Biol. 12:33.

Breton S, Beaupré HD, Stewart DT, Hoeh WR, Blier PU. 2007. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? Trends Genet. 23:465–474.

Breton S, Burger G, Stewart DT, Blier PU. 2006. Comparative analysis of gender-associated complete mitochondrial genomes in marine mussels (*Mytilus spp.*). Genetics 172:1107–1119.

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

Breton S, et al. 2009. Comparative mitochondrial genomics of freshwater mussels (Bivalvia: Unionoida) with doubly uniparental inheritance of mtDNA: gender-specific open reading frames and putative origins of replication. Genetics 183:1575–1589.

Breton S, et al. 2011a. Evidence for a fourteenth mtDNA-encoded protein in the female-transmitted mtDNA of marine mussels (Bivalvia: Mytilidae). PLoS One 6:e19365.

Breton S, et al. 2011b. Novel protein genes in animal mtDNA: a new sex determination system in freshwater mussels (Bivalvia: Unionoida)? Mol Biol Evol. 28:1645–1659.

Burzynski A, Zbawicka M, Skibinski DO, Wenne R. 2003. Evidence for recombination of mtDNA in the marine mussel *Mytilus trossulus* from the Baltic. Mol Biol Evol. 20:388–392.

Burzynski A, Zbawicka M, Skibinski DO, Wenne R. 2006. Doubly uniparental inheritance is associated with high polymorphism for rearranged and recombinant control region haplotypes in Baltic *Mytilus trossulus*. Genetics 174:1081–1094.

Buske FA, Boden M, Bauer DC, Bailey TL. 2010. Assigning roles to DNA regulatory motifs using comparative genomics. Bioinformatics 26: 860–866.

Cantatore P, Roberti M, Polosa PL, Mustich A, Gadaleta MN. 1990. Mapping and characterization of *Paracentrotus lividus* mitochondrial transcripts: multiple and overlapping transcription units. Curr Genet. 17:235–245.

Cantatore P, Roberti M, Rainaldi G, Gadaleta MN, Saccone C. 1989. The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of *Paracentrotus lividus*. J Biol Chem. 264:10965–10975.

Cao L, Kenchington E, Zouros E, Rodakis GC. 2004. Evidence that the large noncoding sequence is the main control region of maternally and paternally transmitted mitochondrial genomes of the marine mussel (*Mytilus* spp.). Genetics 167:835–850.

Cao L, et al. 2009. The control region of maternally and paternally inherited mitochondrial genomes of three species of the sea mussel genus *Mytilus*. Genetics 181:1045–1056.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289–1303.

Chase CD. 2007. Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. Trends Genet. 23:81–90.

Cingolani P, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6:1–13.

Couvet D, Ronce O, Gliddon C. 1998. The maintenance of nucleocytoplasmic polymorphism in a metapopulation: the case of gynodioecy. Am Nat. 152:59–70.

Davison A. 2006. The ovotestis: an underdeveloped organ of evolution. Bioessays 28:642–650.

de Paula WB, Lucas CH, Agip AN, Vizcay-Barrena G, Allen JF. 2013. Energy, ageing, fidelity and sex: oocyte mitochondrial DNA as a protected genetic template. Philos Trans R Soc Lond B Biol Sci. 368: 20120263.

DePristo M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 43: 491–498.

Devauchelle N. 1990. Sviluppo sessuale e maturità di *Tapes philippinarum* (Sexual development and maturity of Tapes philippinarum). In: *Tapes philippinarum*, biologia e sperimentazione. Venezia, Italy: Ente Sviluppo Agricolo Veneto (ESAV). p. 47–62.

D'Onorio de Meo P, et al. 2012. MitoZoa 2.0: a database resource and search tools for comparative and evolutionary analyses of mitochondrial genomes in Metazoa. Nucleic Acids Res. 40: D1168–D1172.

Doucet-Beaupré H, et al. 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.

Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. Genetics 148:1667–1686.

Ellegren H. 2007. Characteristics, causes and evolutionary consequences of male-biased mutation. Proc Biol Sci. 274:1–10.

Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. Nat Rev Genet. 8:689–698.

Embley TM, Martin W. 2006. Eukaryotic evolution, changes and challenges. Nature 440:623–630.

Galtier N, Jobson RW, Nabholz B, Glémin S, Blier PU. 2009a. Mitochondrial whims: metabolic rate, longevity and the rate of molecular evolution. Biol Lett. 5:413–416.

Galtier N, Nabholz B, Glemin S, Hurst GD. 2009b. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol Ecol. 18:4541–4550.

Gemmell NJ, Allendorf FW. 2001. Mitochondrial mutations may decrease population viability. Trends Ecol Evol. 16:115–117.

Gemmell NJ, Metcalf VJ, Allendorf FW. 2004. Mother's curse: the effect of mtDNA on individual fitness and population viability. Trends Ecol Evol. 19:238–244.

Ghiselli F, Milani L, Passamonti M. 2011. Strict sex-specific mtDNA segregation in the germ line of the DUI species *Venerupis philippinarum* (Bivalvia: Veneridae). Mol Biol Evol. 28:949–961.

Ghiselli F, et al. 2012. De Novo assembly of the Manila clam *Ruditapes philippinarum* transcriptome provides new insights into expression bias, mitochondrial doubly uniparental inheritance and sex determination. Mol Biol Evol. 29:771–786.

Gissi C, Iannelli F, Pesole G. 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101:301–320.

Gissi C, et al. 2010. Hypervariability of ascidian mitochondrial gene order: exposing the myth of deuterostome organelle genome stability. Mol Biol Evol. 27:211–215.

Gosling EM. 2003. Bivalve molluscs: biology, ecology and culture. Oxford: Blackwell Publishing Ltd.

Gouyon PH, Vichot F, Vandamme JMM. 1991. Nuclear-cytoplasmic male-sterility—single-point equilibria versus limit-cycles. Am Nat. 137: 498–514.

Hoede C, Denamur E, Tenaillon O. 2006. Selection acts on DNA secondary structures to decrease transcriptional mutagenesis. PLoS Genet. 2: e176.

Houliston GJ, Olson MS. 2006. Nonneutral evolution of organelle genes in *Silene vulgaris*. Genetics 174:1983–1994.

Jacobs HT, Herbert ER, Rankine J. 1989. Sea urchin egg mitochondrial DNA contains a short displacement loop (D-loop) in the replication origin region. Nucleic Acids Res. 17:8949–8965.

Khachane AN, Timmis KN, Martins dos Santos VA. 2007. Dynamics of reductive genome evolution in mitochondria and obligate intracellular microbes. Mol Biol Evol. 24:449–456.

Koonin EV. 2009. Evolution of genome architecture. Int J Biochem Cell Biol. 41:298–306.

Ladoukakis ED, Theologidis I, Rodakis GC, Zouros E. 2011. Homologous recombination between highly diverged mitochondrial sequences: examples from maternally and paternally transmitted genomes. Mol Biol Evol. 28:1847–1859.

Lane N. 2007. Mitochondria: key to complexity. In: Martin WF, Müller M, editors. Origin of mitochondria and hydrogenosomes. Germany: Springer. p. 13–38.

Lupi R, et al. 2010. MitoZoa: a curated mitochondrial genome database of metazoans for comparative genomics studies. Mitochondrion 10: 192–199.

Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates.

Lynch M, Bobay LM, Catania F, Gout JF, Rho M. 2011. The repatterning of eukaryotic genomes by random genetic drift. Annu Rev Genom Hum Genet. 12:347–366.

Lynch M, Koskella B, Schaack S. 2006. Mutation pressure and the evolution of organelle genomic architecture. Science 311:1727–1730.

Maruyama T, Fuerst PA. 1984. Population bottlenecks and nonequilibrium models in population genetics. I. Allele numbers when populations evolve from zero variability. Genetics 108:745–763.

Maruyama T, Fuerst PA. 1985. Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. Genetics 111:675–689.

McKenna A, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

Meiklejohn CD, Montooth KL, Rand DM. 2007. Positive and negative selection on the mitochondrial genome. Trends Genet. 23:259–263.

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., doi: 10.1093/gbe/evt101. Advance Access publication July 3, 2013.

Milani L, Ghiselli F, Maurizii MG, Passamonti M. 2011. Doubly uniparental inheritance of mitochondria as a model system for studying germ line formation. PLoS One 6:e28194.

Mizi A, Zouros E, Moschonas N, Rodakis GC. 2005. The complete maternal and paternal mitochondrial genomes of the Mediterranean mussel *Mytilus galloprovincialis*: implications for the doubly uniparental inheritance mode of mtDNA. Mol Biol Evol. 22:952–967.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 5:621–628.

Müller M, Martin W. 1999. The genome of *Rickettsia prowazekii* and some thoughts on the origin of mitochondria and hydrogenosomes. Bioessays 21:377–381.

Müller M, et al. 2012. Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. Microbiol Mol Biol Rev. 76:444–495.

Nachman MW. 2001. Single nucleotide polymorphisms and recombination rate in humans. Trends Genet. 17:481–485.

Olson-Manning CF, Wagner MR, Mitchell-Olds T. 2012. Adaptive evolution: evaluating empirical support for theoretical predictions. Nat Rev Genet. 13:867–877.

Ovchinnikov S, Masta SE. 2012. Pseudoscorpion mitochondria show rearranged genes and genome-wide reductions of RNA gene sizes and inferred structures, yet typical nucleotide composition bias. BMC Evol Biol. 12:31.

Palumbi SR. 2009. Speciation and the evolution of gamete recognition genes: pattern and process. Heredity 102:66–76.

Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. Nat Rev Genet. 14:83–87.

Passamonti M. 2007. An unusual case of gender-associated mitochondrial DNA heteroplasmy: the mytilid *Musculista senhousia* (Mollusca Bivalvia). BMC Evol Biol. 7(2 Suppl):S7.

Passamonti M, Ghiselli F. 2009. Doubly uniparental inheritance: two mitochondrial genomes, one precious model for organelle DNA inheritance and evolution. DNA Cell Biol. 28:79–89.

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

Pereira F, et al. 2008. Evidence for variable selective pressures at a large secondary structure of the human mitochondrial DNA control region. Mol Biol Evol. 25:2759–2770.

Petrov DA. 2001. Evolution of genome size: new approaches to an old problem. Trends Genet. 17:23–28.

Rand DM. 2001. The units of selection on mitochondrial DNA. Annu Rev Ecol Syst. 32:415–448.

Ren J, Liu X, Jiang F, Guo X, Liu B. 2010. Unusual conservation of mitochondrial gene order in Crassostrea oysters: evidence for recent speciation in Asia. BMC Evol Biol. 10:394.

Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 132:365–86.

Saavedra C, Bachere E. 2006. Bivalve genomics. Aquaculture 256:1–14.

Scheffler IE. 2008. Mitochondria. 2nd ed. Hoboken (NJ): Wiley-Liss.

Shao R, Dowton M, Murrell A, Barker SC. 2003. Rates of gene rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of insects. Mol Biol Evol. 20:1612–1619.

Siegel S, Castellan NJ. 1988. Non parametric statistics for the behavioural sciences. 2nd ed. New York: MacGraw Hill.

Skibinski DO, Gallagher C, Beynon CM. 1994. Mitochondrial DNA inheritance. Nature 368:817–818.

Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731–9.

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 62:959–970.

Thrash JC, et al. 2011. Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. Sci Rep. 1:13.

Venetis C, Theologidis I, Zouros E, Rodakis GC. 2006. No evidence for presence of maternal mitochondrial DNA in the sperm of *Mytilus galloprovincialis* males. Proc Biol Sci. 273:2483–2489.

Venetis C, Theologidis I, Zouros E, Rodakis GC. 2007. A mitochondrial genome with a reversed transmission route in the Mediterranean mussel *Mytilus galloprovincialis*. Gene 406:79–90.

Washietl S, Hofacker IL, Stadler PF. 2005. Fast and reliable prediction of noncoding RNAs. Proc Natl Acad Sci U S A. 102:2454–2459.

Xu W, Jameson D, Tang B, Higgs PG. 2006. The relationship between the rate of molecular evolution and the rate of genome rearrangement in animal mitochondrial genomes. J Mol Evol. 63:375–392.

Xu Z, Mathews DH. 2011. Multilign: an algorithm to predict secondary structures conserved in multiple RNA sequences. Bioinformatics 27:626.

Youle RJ, van der Bliek AM. 2012. Mitochondrial fission, fusion, and stress. Science 337:1062–1065.

Zbawicka M, Burzynski A, Skibinski D, Wenne R. 2010. Scottish *Mytilus trossulus* mussels retain ancestral mitochondrial DNA: complete sequences of male and female mtDNA genomes. Gene 456:45–53.

Zouros E. 2012. Biparental inheritance through uniparental transmission: the doubly uniparental inheritance (DUI) of mitochondrial DNA. Evol Biol. 40:1–31.

Zouros E, Ball AO, Saavedra C, Freeman KR. 1994. Mitochondrial DNA inheritance. Nature 368:818.

Zuker M. 2000. Calculating nucleic acid secondary structure. Curr Opin Struct Biol. 10:303–310.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31:3406–3415.

Associate editor: Bill Martin