

Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN  
BIODIVERSITÀ ED EVOLUZIONE

Ciclo XXVII

**Settore Concorsuale di afferenza:** 05/B1 - Zoologia e Antropologia

**Settore Scientifico disciplinare:** BIO/05 - Zoologia

CHARACTERIZATION OF MITOCHONDRIAL GENOMES IN  
BIVALVE SPECIES WITH DOUBLY UNIPARENTAL  
INHERITANCE OF MITOCHONDRIA

**Presentata da: Dott. Davide Guerra**

**Coordinatore Dottorato**

**Relatore**

**Prof. Barbara Mantovani**

**Prof. Marco Passamonti**

**Esame finale anno 2015**

*Within the shadow of the ship,  
I watched their rich attire:  
Blue, glossy green, and velvet black,  
They coiled and swam; and every track  
Was a flash of golden fire.*

*O happy living things! no tongue  
Their beauty might declare:  
A spring of love gushed from my heart,  
And I blessed them unaware:  
Sure my kind saint took pity on me  
And I blessed them unaware.*

S. T. Coleridge

THE RIME OF THE ANCIENT MARINER

<b>Thesis structure ...</b>	p. 1
<b>Abbreviations ...</b>	p. 2
<b>Chapter 1 - Introduction ...</b>	p. 3
1.1 Animal mitochondrial genomes ...	p. 3
1.1.1 Evolution and structure ...	p. 3
1.1.2 Inheritance and homoplasmy ...	p. 4
1.1.3 Mitochondrial heteroplasmy ...	p. 6
1.2 Doubly uniparental inheritance (DUI) of mitochondria ...	p. 7
1.2.1 The mechanisms of DUI ...	p. 7
1.2.2 Gene content of F and M mt genomes ...	p. 11
1.2.3 Control regions ...	p. 16
1.3 Aims of the thesis ...	p. 17
<b>Chapter 2 - Early replication dynamics of sex-linked mitochondrial DNAs in the doubly uniparental inheritance species <i>Ruditapes philippinarum</i> (Bivalvia Veneridae) ...</b>	p. 19
2.1 Introduction ...	p. 19
2.2 Materials and Methods ...	p. 23
2.2.1 Sample characteristics ...	p. 23
2.2.2 DNA extractions ...	p. 24
2.2.3 Real-Time qPCR procedure ...	p. 25
2.2.4 Statistical analyses ...	p. 26
2.3 Results ...	p. 27
2.4 Discussion ...	p. 35
2.4.1 mtDNA replication is dormant during early embryogenesis ...	p. 35

2.4.2	Mitochondrial heteroplasmy during male development ...	p. 37
2.4.3	Approaching adulthood: sex-specific mtDNA dynamics in young clams ...	p. 42
2.5	Conclusions ...	p. 44
<b>Chapter 3 - Published works summary ...</b>		
p. 46		
3.1	Description of the papers ...	p. 46
3.2	Paper 1 ...	p. 49
3.2.1	Background ...	p. 49
3.2.2	Analyses and results ...	p. 50
3.2.3	Discussion ...	p. 53
3.3	Paper 2 ...	p. 57
3.3.1	Background ...	p. 57
3.3.2	Analyses and results ...	p. 58
3.3.3	Discussion ...	p. 60
3.4	Paper 3 ...	p. 64
3.4.1	Background ...	p. 64
3.4.2	Analyses and results ...	p. 66
3.4.3	Discussion ...	p. 67
3.5	Paper 4 ...	p. 70
3.6	Authors affiliations ...	p. 74
<b>Chapter 4 - Literature cited ...</b>		
p. 75		
<b>Appendix 1 ...</b>		
p. 94		
Supplementary Table 1 ...		
p. 95		
Supplementary Table 2 ...		
p. 102		
Supplementary Table 3 ...		
p. 111		
Supplementary Table 4 ...		
p. 112		

Supplementary Table 5 ... p. 113

Supplementary Table 6 ... p. 114

Supplementary Figure ... p. 115

**Appendix 2** ... p. 116

Paper 1 ... p. 117

Paper 2 ... p. 144

Paper 3 ... p. 164

Paper 4 ... p. 174

## THESIS STRUCTURE

This PhD Thesis is subdivided in seven main parts.

A list of **Abbreviations**, containing all acronyms used throughout the text, opens this Thesis.

**Chapter 1** is a general introduction focused on the main biological aspects that will be discussed in the rest of the Thesis. The scopes of my Thesis are described at the end of this chapter.

**Chapter 2** is the description, in article form, of the main experimental project of my PhD course, which was focused on the analysis of mtDNA replication during the development of the bivalve *Ruditapes philippinarum* (Veneridae). At the time of printing, the resulting paper is submitted for publication.

**Chapter 3** summarizes the four articles I participated during my PhD course. My contributions are enlisted in the opening of the chapter.

**Chapter 4** contains cited literature.

**Appendix 1** contains the supplementary information to Chapter 2, and **Appendix 2** reports full texts of the four published papers.

## ABBREVIATIONS

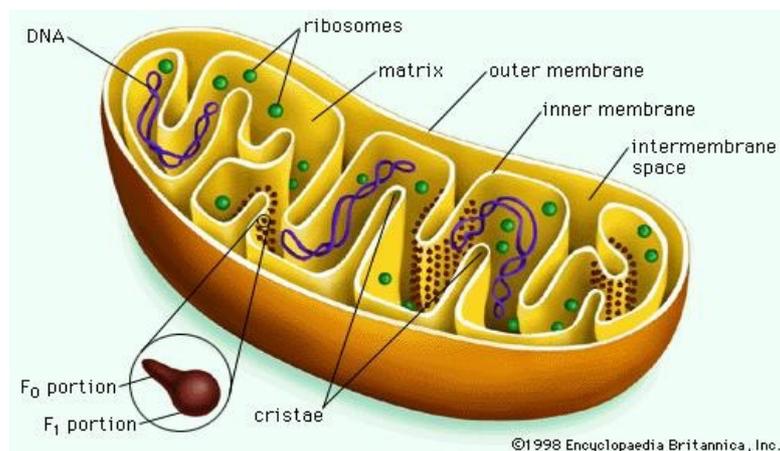
aa: amino acid	mORF: ORFan specific of DUI species
AT: adenine and thymine content	M mtDNA
ATP: adenosine triphosphate	MLUR: M mtDNA LUR
bp: base pairs	mt: mitochondrial
cDNA: complementary DNA	O <sub>H</sub> : origin of heavy strand replication
CDS: coding DNA sequence	O <sub>L</sub> : origin of light strand replication
CMS: cytoplasmic male sterility	ORF: open reading frame
CoRR: co-location for redox regulation	ORFan: ORF with unknown ontology and function
Cq: quantification cycle	OXPHOS: oxidative phosphorylation
CR: control region	qPCR: quantitative polymerase chain reaction
DNA: deoxyribonucleic acid	RNA: ribonucleic acid
DUI: doubly uniparental inheritance	ROS: reactive oxygen species
ETC: electron transport chain	rRNA: ribosomal RNA
F: prefix for “female-transmitted”	SMI: strictly maternal inheritance
FLUR: F mtDNA LUR	SNP: single nucleotide polymorphism
fORF: ORFan specific of DUI species F mtDNA	SP: signal peptide
GRE: genome reductive evolution	tRNA: transfer RNA
hpf: hours post-fertilization	UR: unassigned region
kb: kilobases	
LUR: largest unassigned region of a mt genome	
M: prefix for “male-transmitted”	

## INTRODUCTION

### 1.1 ANIMAL MITOCHONDRIAL GENOMES

#### 1.1.1 EVOLUTION AND STRUCTURE

Mitochondria are double-membraned cellular organelles that produce energy in the form of ATP through OXPHOS, and are involved in cellular processes such as cell signaling and differentiation, fertilization, ageing, and apoptosis (Scheffler 2008, Van Blerkom 2011, López-Otín *et al.* 2013, Chandel 2014). Mitochondria possess their own genome, the mt genome or mtDNA, which is organized in nucleoids located into the mitochondrion matrix (Figure 1).



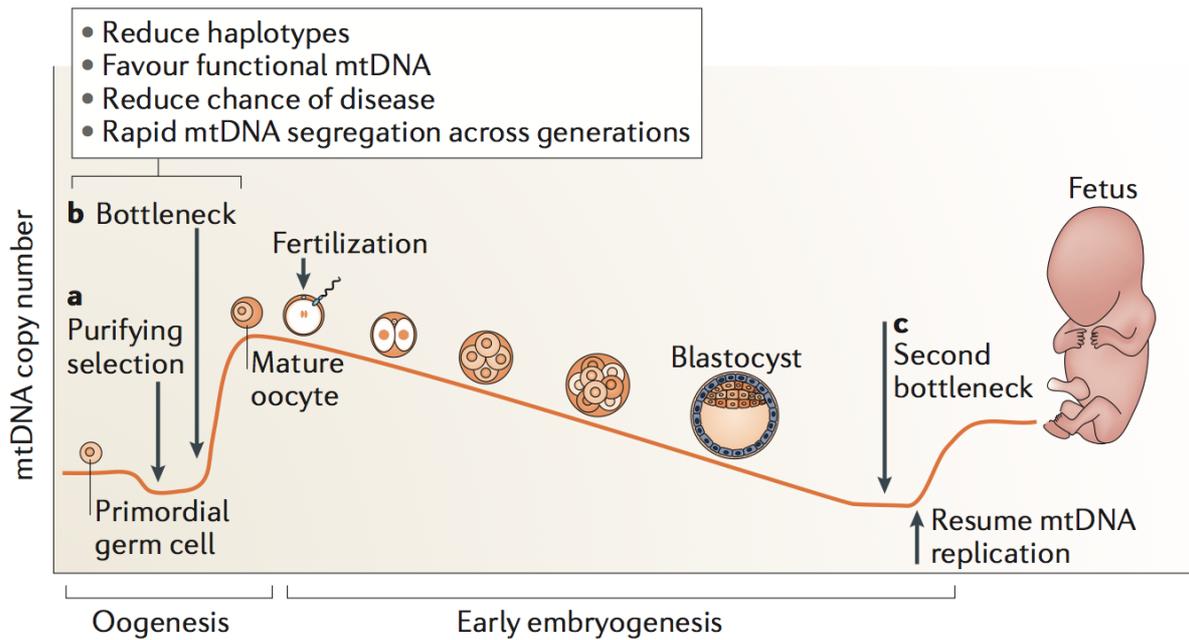
**Figure 1** Schematic structure of a mitochondrion. [Taken from [www.britannica.com](http://www.britannica.com)]

Phylogenetic evidence indicates that the mitochondrion originated from an ancient symbiosis event, where an  $\alpha$ -proteobacterium was incorporated into an archaeon (Müller and Martin 1999, Koonin and Martin 2005, Embley and Martin 2006, Martin and Koonin 2006,

Atteia *et al.* 2009, Abhishek *et al.* 2011, Thrash *et al.* 2011). The mtDNA has been subject to a process of reduction since the symbiosis event (Andersson and Kurland 1998, Timmis *et al.* 2004, Khachane *et al.* 2007), resulting into a generally small and circular genome whose genes are translated with its own genetic code, different from the nuclear one. The typical animal mtDNA is a rather simple molecule of ~16kb encoding 13 protein-coding genes (that produce subunits of complexes I, III, IV, and V of the OXPHOS system), plus 22 tRNA genes and two rRNA subunits (12S and 16S) for their transcription and translation. A non-coding sequence of variable length and nucleotide content, the CR, contains the signals to start the replication of the mtDNA molecule (Scheffler 2008). However, many exceptions to the “standard” mt genome organization are found throughout the animal kingdom (reviewed in Breton *et al.* 2014).

### **1.1.2 INHERITANCE AND HOMOPLASMY**

Mitochondria and the genome they carry are inherited maternally through oocytes in almost all animal species, in a system called SMI. This uniparental mode of transmission is achieved by a series of mechanisms that eliminate the mtDNA carried by spermatozoon mitochondria at various stages of fertilization and development (reviewed in Birky 2001): this inheritance pattern is thought to be an adaptive condition whose advantage is to avoid conflicts between maternal and paternal mtDNA lineages in a developing individual (Hoekstra 2011, Lane 2011), thus ensuring its homoplasmy. During oogenesis the mt genome is subject to a bottleneck that promotes the homoplasmy of oocytes (Figure 2). The outcome of this bottleneck is that a mature oocyte contains only one mtDNA haplotype, or a highly reduced number of variants. However, the nature of this bottleneck is unclear: proofs have been gathered that support either the reduction of mtDNA copy number followed by amplification of this reduced pool during oocyte maturation (Cree *et al.* 2008), or the



**Figure 2** mtDNA dynamics during animal embryogenesis, exemplified by mammalian development. During oogenesis, purifying selection (a) eliminates deleterious mtDNA haplotypes, then a bottleneck occurs (b) before oocyte maturation. Replication of mtDNA is not active during early embryogenesis: mtDNA copy number is thus reduced in the proliferating cells, resulting into a second bottleneck (c). In a later phase, mtDNA replication is resumed. [Taken from Mishra and Chan (2014)]

selective amplification of only certain mtDNA variants, and not of others, during oogenesis, resulting in a “virtual” bottleneck (Wai *et al.* 2008), maybe without a reduction in copy number inside germ cells (Cao *et al.* 2007, Cao *et al.* 2009). To achieve homoplasmy of mature oocytes, the bottleneck must act on “segregating units”, which can be the single mtDNA molecules or the nucleoids, which could contain multiple mtDNA copies. This has implications on the mtDNA population dynamics during oocyte maturation, as the two possibilities can produce different rates of genetic drift, but the debate on what is the real segregating unit is still ongoing (see Khrapko 2008). The transmission of only functional mtDNA molecules seems to be enhanced by purifying mechanisms that negatively select deleterious mutant mt genomes during oogenesis (Sato *et al.* 2007, Fan *et al.* 2008, Stewart *et al.* 2008). Moreover, Allen (1996) proposed that oocyte mitochondria are inactive to avoid ROS damage to their mtDNA. Furthermore, another bottleneck occurs during embryogenesis

(Figure 2): mtDNA replication is inactive during early cell divisions, causing the progressive decrease of mt genome copy number per cell in the embryo, which consequently reduces mtDNA variability in each of them (reviewed in Mishra and Chan 2014). This bottleneck enhances the homoplasmy of the tissues deriving from these cells.

### **1.1.3 MITOCHONDRIAL HETEROPLASMY**

Heteroplasmy, the presence of more than one mtDNA haplotype in an individual, is thus considered an exceptional condition that can have many causes, such as the maternal transmission of multiple haplotypes, the acquisition of *de novo* mutations during ageing, or from paternal inheritance (as in the human case reported by Schwartz and Vissing 2002). However, recent deep-sequencing studies demonstrated that humans generally carry many different mtDNA haplotypes at very low levels, even potentially harmful ones (Ye *et al.* 2014). In a heteroplasmic condition, if a harmful haplotype exceed a concentration threshold in a tissue or organ, it may cause severe pathologies (Jokinen and Battersby 2013, Wallace and Chalkia 2013, Mishra and Chan 2014). The presence of two different but functional, non-harmful, mt genomes in engineered laboratory mouse strains seemed to affect the overall viability of an individual in various ways, both from a physiological and behavioral point of view (see Sharpley *et al.* 2012 and references therein). Two functional mtDNA haplotypes may segregate in tissue- and haplotype-specific ways in artificially heteroplasmic mice, but the reasons behind these patterns remain unknown (Burgstaller *et al.* 2014).

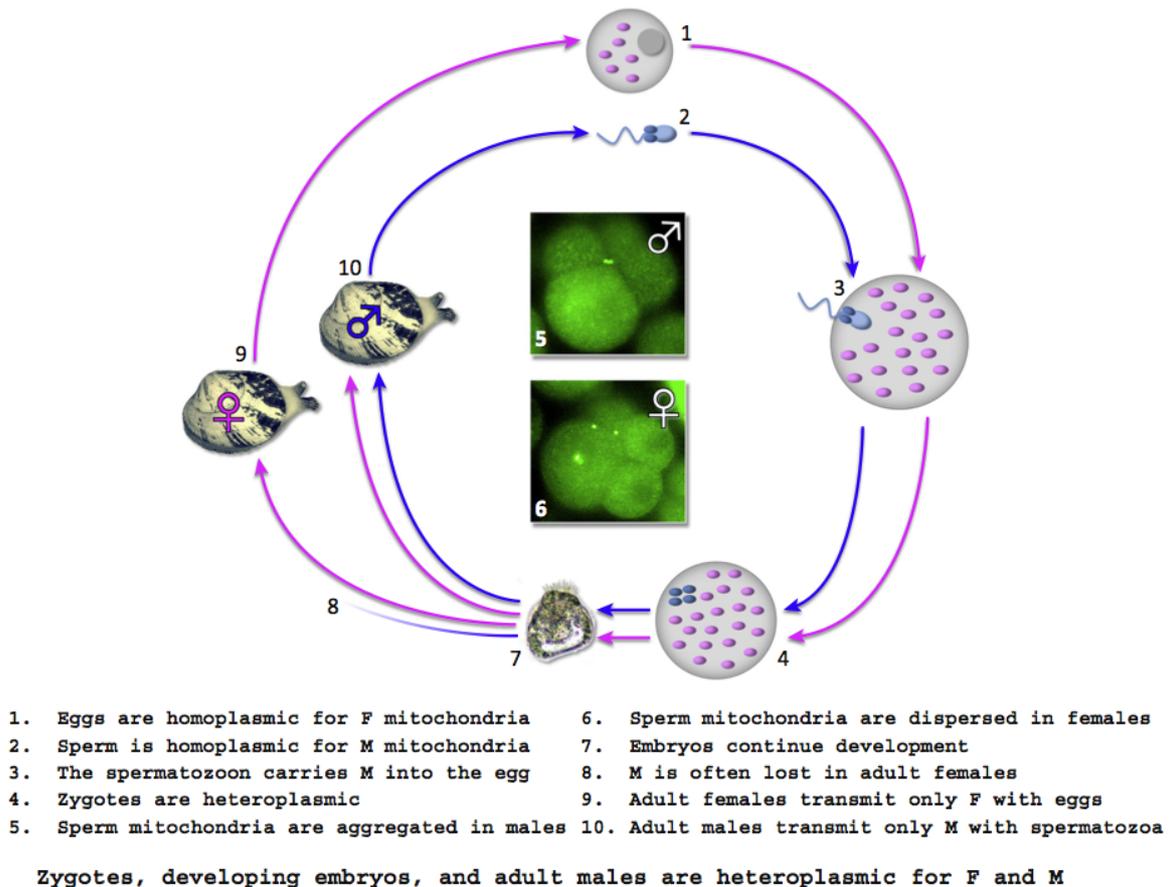
## 1.2 DOUBLY UNIPARENTAL INHERITANCE (DUI) OF MITOCHONDRIA

### 1.2.1 THE MECHANISMS OF DUI

In animals, the most striking exception to SMI and homoplasmy of mtDNA is DUI (Skibinski *et al.* 1994a, Skibinski *et al.* 1994b, Zouros *et al.* 1994a, Zouros *et al.* 1994b), a system of mitochondrial transmission found only in gonochoric bivalves belonging to eight families (Theologidis *et al.* 2008, Boyle and Etter 2013). Under DUI, two mt genomes occur in a single species: one is transmitted by eggs (called F, from female-transmitted), and the other by spermatozoa (called M, from male-transmitted) (Figure 3). At fertilization, spermatozoon mitochondria enter the egg: the zygote is thus heteroplasmic for both lines, but, depending on the sex of the individual, the fate of paternal mitochondria and their M genome is different.

The *Mytilus edulis* (Mytilidae) complex, and the clam species *Ruditapes philippinarum* (Veneridae), have been the most used models to unravel the molecular and cellular mechanisms of DUI and several models for DUI have been elaborated so far. For instance, Ghiselli *et al.* (2011) proposed the crucial stages that differentiate female and male paths of development as three checkpoints (Figure 4). During Checkpoint #1, sperm mitochondria can be scattered throughout the embryo, in a mode called “dispersed pattern” (Figure 3, panel 5), or, on the contrary, retained as a single mass in an “aggregated pattern” on a specific plane of cleavage of the developing embryo (Figure 3, panel 6) (Cao *et al.* 2004a, Obata and Komaru 2005, Cogswell *et al.* 2006, Milani *et al.* 2011, Milani *et al.* 2012). The first pattern has been linked to the female path of development, while the second (observed only in DUI species and absent from non-DUI species), has been referred to the male one. The embryo plasm, where sperm mitochondria locate, is indeed the one from which the germ line will originate, and this may explain why male gonads contain almost exclusively M mtDNA (see below) and

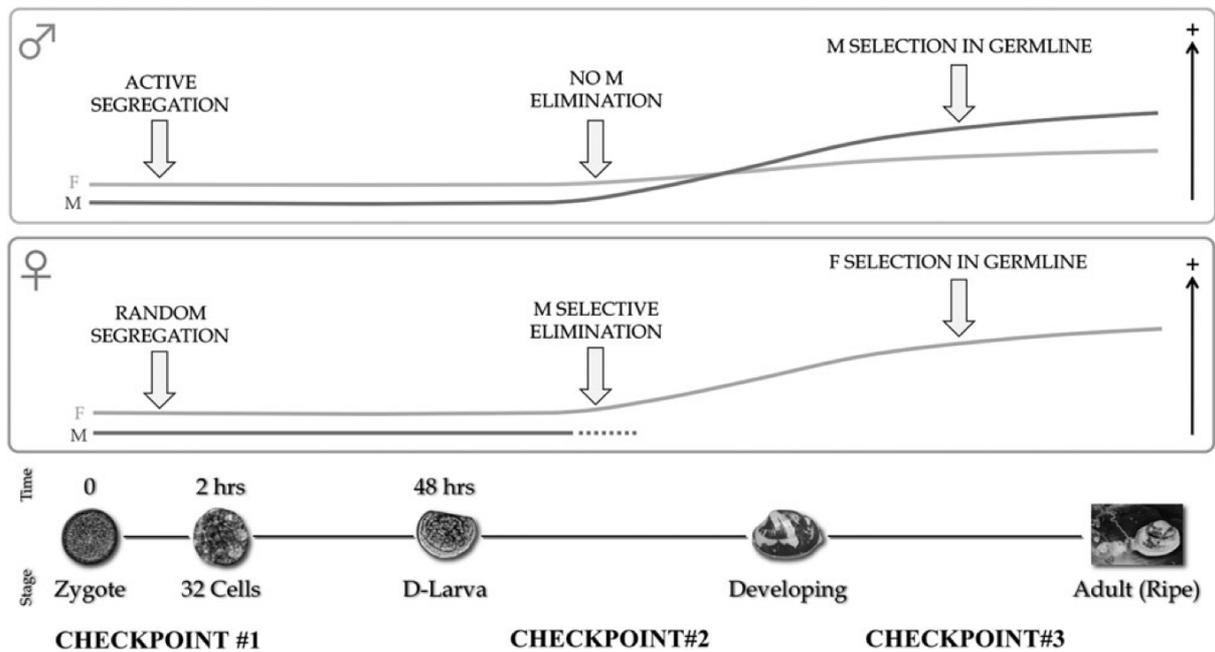
## The doubly uniparental inheritance of mitochondria



TRENDS in Genetics

**Figure 3** Functioning of DUI. Pink ellipses: egg mitochondria containing F mtDNA. Blue ellipses: sperm mitochondria carrying M mtDNA. Pink and blue arrows: F and M mtDNA routes of transmission, respectively. Frames 5 and 6: MitoTracker® Green FM staining of sperm mitochondria in early embryos of *Ruditapes philippinarum*. [Taken from Breton *et al.* (2014)]

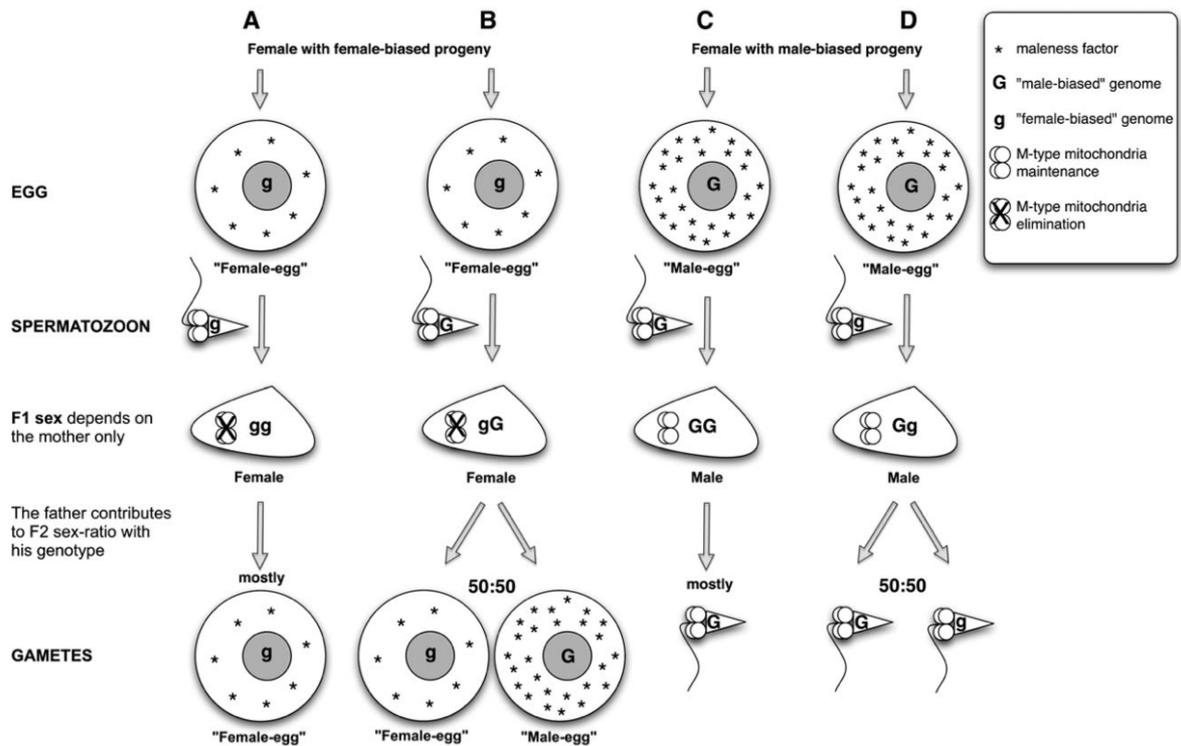
produce spermatozoa that are homoplasmic for M. In a second stage, Checkpoint #2, sperm mitochondria are eliminated in females so that they become homoplasmic for the F mt genome, as in species with SMI (Figure 3, panel 6, and Figure 4). It is still unclear however if M mtDNA is really lost or just diluted below detectable levels: the occurrence of heteroplasmic females in *Mytilus* (see references in Zouros 2013) and *Ruditapes philippinarum* (Ghiselli *et al.* 2011) indicates that females can sometimes retain M mtDNA in their somatic tissues. In the last phase, Checkpoint #3, the mtDNA that is going to be transmitted by gametes is selected in the germ line, probably by means of active selection



**Figure 4** Checkpoints during male (upper panel) and female (lower panel) development of a DUI species. F: F mtDNA, M: M mtDNA. [Taken from Ghiselli *et al.* (2011)]

mechanisms (Figure 4) (Venetis *et al.* 2006, Ghiselli *et al.* 2011, Milani *et al.* 2011, Guerra *et al.* 2014). Females select F, while males select M: eggs and spermatozoa will be therefore homoplasmic for the F and M mtDNA, respectively, thus maintaining the uniparental inheritance of the two mt lines.

M is the dominant mtDNA line in male gonads (Fisher and Skibinski 1990, Skibinski *et al.* 1994b, Zouros *et al.* 1994a, Stewart *et al.* 1995, Quesada *et al.* 1996, Beagley *et al.* 1997, Saavedra *et al.* 1997, Ghiselli *et al.* 2011), but the presence of this mt genome in soma is variable from species to species. *Mytilus* males have very low quantities of somatic M mtDNA (Garrido-Ramos *et al.* 1998, Obata *et al.* 2006, Kyriakou *et al.* 2010, Batista *et al.* 2011, Obata *et al.* 2011), while adult *Ruditapes philippinarum* males usually have high amounts of it in their tissues, often more than F (Passamonti and Scali 2001, Ghiselli *et al.* 2011). The reason for this disparity has been linked to the degree of leakage of sperm mitochondria from the early embryo aggregate, which can differ in various species (Milani *et*



**Figure 5** Model for sex determination and DUI. Transcription factors (e.g., ubiquitination genes) stored in female oocytes would activate sex–gene expression in early embryonic developmental stages, and male development would require the crossing of a critical threshold of masculinizing transcripts. The sperm genotype contributes to F2 sex bias. (A, B) A “female egg” will produce a female regardless the genotype of the spermatozoon. (A) If it is fertilized by a spermatozoon with a “female-biased” genotype (g), the F1 female will produce mostly female eggs. (B) If it is fertilized by a spermatozoon with a “male-biased” genotype (G), the F1 female will produce both egg types (50:50). (C, D) A “male egg” will produce a male regardless the genotype of the spermatozoon. (C) If it is fertilized by a spermatozoon with a “male-biased” genotype (G), the F1 male will produce sperm carrying a male-biased genotype (G). (D) If it is fertilized by a spermatozoon with a “female- biased” genotype (g), the F1 male will produce both sperm types (50:50). Some ubiquitination factors could also be involved in mitochondrial inheritance, and their differential expression could be responsible for the different fate of sperm mitochondria in the two families: degradation (A, B) or maintenance (C, D). Note that the genomic sex-determining factors (G and g) probably comprise more than one gene; recombination among these genes and environmental factors could account for the nearly continuous distribution of sex ratios among families. [Taken from Ghiselli *et al.* (2012)]

*al.* 2012): following this rationale, the aggregate might be looser in *Ruditapes philippinarum* than in *Mytilus*, thus allowing a more widespread distribution of M mtDNA outside the gonad. Females failing to eliminate sperm mitochondria also result as heteroplasmic, and the distribution of M in their soma is variable.

Another particular feature observed in DUI species is that different females produce progenies with different sex bias (female biased, male biased, or balanced), and that this bias is independent of the mating male, thus being only mother-dependent (Saavedra *et al.* 1997, Kenchington *et al.* 2002, Ghiselli *et al.* 2012). A preformation process has been proposed to explain the sex bias, in which different quantities of certain transcripts in eggs can drive sex determination: in particular, the crossing of a threshold for some of them could lead to maleness (Figure 5) (Kenchington *et al.* 2009, Ghiselli *et al.* 2012). Factors involved in male sex determination may also avoid sperm mitochondria degradation, which in turn can participate in the formation of the gonad in this sex (Breton *et al.* 2007, Passamonti and Ghiselli 2009).

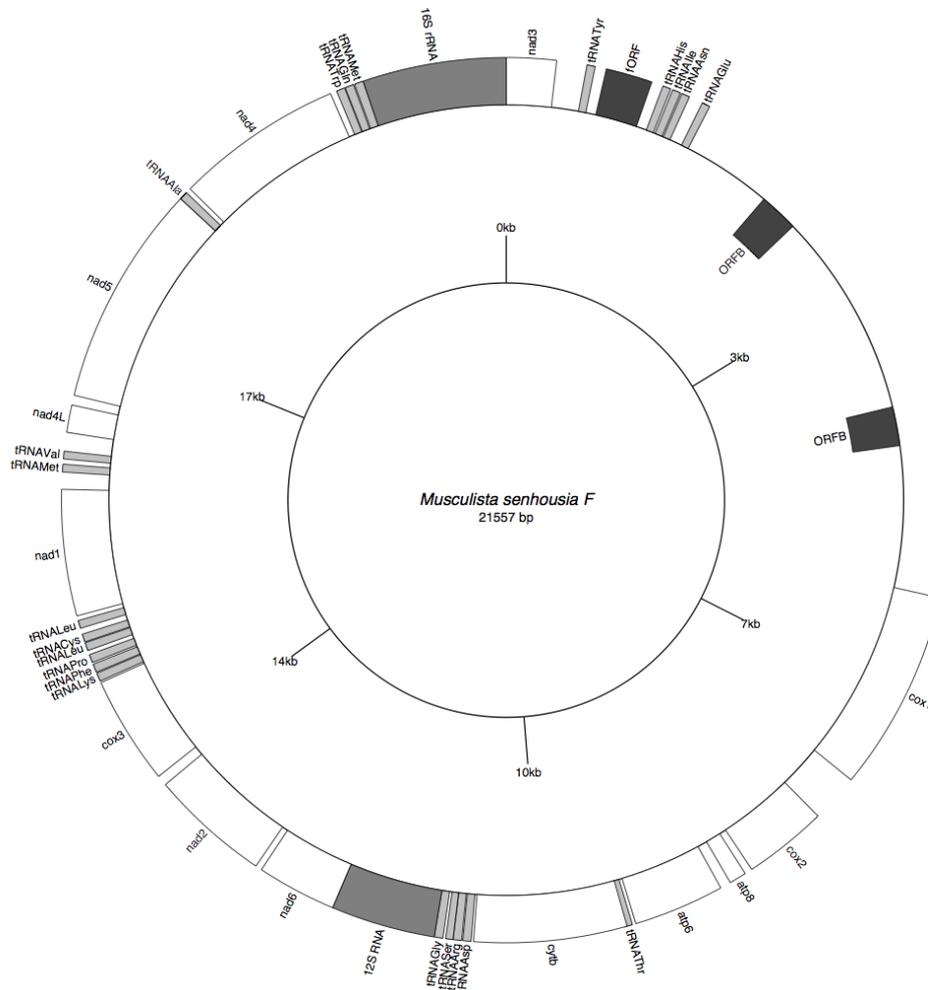
Therefore, DUI and sex determination are strictly linked, but it is still obscure if this link is associative (DUI is a side-effect of sex determination; Kenchington *et al.* 2009, Zouros 2013) or causative (the maintenance of sperm mitochondria triggers male development; Breton *et al.* 2007, Breton *et al.* 2011a, Yusa *et al.* 2013). Bivalves do not possess sexual chromosomes, and the presence of sex determining factors in the F and M mtDNA of DUI species has been proposed to support the association between DUI and sex determination (Breton *et al.* 2011a, Yusa *et al.* 2013, Milani *et al.* 2014).

### **1.2.2 GENE CONTENT OF F AND M MT GENOMES**

Genes encoded by F and M mtDNAs have a high sequence divergence, spanning from ~22% in *Mytilus edulis* and *Mytilus galloprovincialis* (Mytilidae; Zouros 2013) to ~43% in







**Figure 8** *Musculista senhousia* F mtDNA (GenBank accession number GU001953; Passamonti *et al.* 2011). A lineage-specific ORFan, fORF, is found in this mt genome. The LUR of this mtDNA is composed of two large repetitive units, and both of them contain an ORFan on the reverse strand named ORF-B: this same ORFan is found also in the M mtDNA (see Figure 9). See Milani *et al.* (2013) for a characterization of fORF and ORF-B, and Guerra *et al.* (2014) for a detailed annotation of the LUR.

relatively low transcription levels may suggest the non-functionality of this gene (Ghiselli *et al.* 2013).

Even more significantly, lineage-specific ORFans, called fORF and mORF, are found in all F and M mt genomes of DUI species (Breton *et al.* 2009, Breton *et al.* 2011a, Breton *et al.* 2011b, Milani *et al.* 2013), and their presence has been putatively linked to the functioning of DUI (for detailed discussions see Breton *et al.* 2011a, Milani *et al.* 2013, Milani *et al.* 2014, Breton *et al.* 2014). Ghiselli *et al.* (2013) provided clear evidence for the transcription of



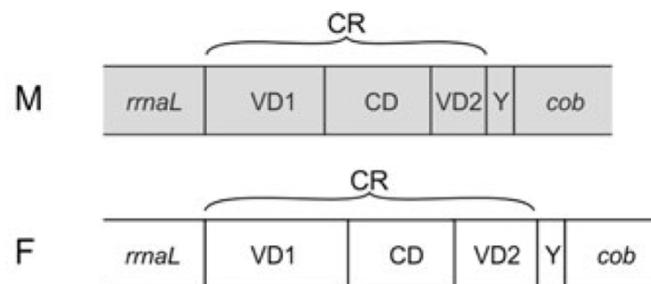
clear evidence of their transcription has been found so far (Breton *et al.* 2011b, Kyriakou *et al.* 2014), and confirmation of a protein encoded by them is still lacking.

### 1.2.3 CONTROL REGIONS

The CR of the two mt genomes can contain large shared segments that comprise motifs and secondary structures related to replication and transcription of the mtDNA molecule (Cao *et al.* 2004b, Breton *et al.* 2009, Ghiselli *et al.* 2013, Guerra *et al.* 2014). The characterization of a CR in DUI species has a special advantage: if the two mtDNAs are not too divergent, the blocks conserved between F and M are often easily alignable, and thus it is easier to focus on these regions to identify shared features which may have a common, non sex-linked, biological function in both mt genomes. Doing the same in non-DUI species requires comparisons among related species, but given the high variability of bivalve mt genomes, if the phylogenetic distance is too large, the differences among them would be too high to recognize conserved features. This rationale has been useful to characterize the CR of many DUI species. *Mytilus* spp. (Mytilidae) CR is composed of two domains, variable between F and M, flanking a conserved one (Figure 10; Cao *et al.* 2004b). A comparable situation is found in *Musculista senhousia* (Crenellinae) (see Figures 16 and 20 in Chapter 3; Passamonti *et al.* 2011, Guerra *et al.* 2014). *Mytilus* and *Musculista* CRs share similar structure and a number of other features, so the CR conformation in Mytilidae seems to be conserved even among distantly related species (Guerra *et al.* 2014).

An interesting feature of *Musculista senhousia* F mtDNA, differently from any other bivalve species, is that the CR is duplicated in tandem, with the two copies evolving in concert (Guerra *et al.* 2014). In the case of *Ruditapes philippinarum*, two long URs are present (see Figures 6 and 7), sharing three large conserved subunits arranged in a different fashion between F and M mtDNA (see Figure 17 in Chapter 3; Ghiselli *et al.* 2013). The CR

in this species was identified as the region containing the two conserved subunits that carry motifs putatively involved in replication and transcription (a third subunit is in different locations and does not contain such motifs) (Ghiselli *et al.* 2013). In freshwater mussels (Unionidae), URs are much shorter than all other DUI species. However, a region with the same relative position between F and M mtDNAs may contain the O<sub>H</sub> and thus is considered the CR (Breton *et al.* 2009).



**Figure 10** Scheme of F and M mtDNA CRs of *Mytilus edulis*, *Mytilus galloprovincialis*, and *Mytilus californianus*. *Mytilus trossulus* F mtDNA has a more complex CR organization, while in its M mt genome it is identical to the one depicted. VD1 and VD2: variable domain 1 and 2, CD: conserved domain. VD1 and VD2 are highly different between F and M mt genomes, while CD is similar in the two mtDNAs. [Modified from Zouros (2013)]

### 1.3 AIM OF THE THESIS

The main experimental project of my PhD was to analyze the replication dynamics of F and M mtDNAs during the development of the DUI species *Ruditapes philippinarum*, from early embryos to young individuals reaching the first reproductive season, a time frame never explored before. To do this, I used a multiplex Real-Time qPCR approach that allowed me to detect the variations in mt genome quantities in this large developmental interval. The aim of my project was to frame in a temporal scale some of the crucial passages that characterize the development of a DUI species, like Checkpoint #2 and #3 (described above; see Figure 4),

and the start of gonad production. In these steps, we should observe in the male sex the retaining and amplification of paternally-transmitted mtDNAs. The identification of such phases is a necessary stepping stone to future studies that will focus on these temporal windows, to understand what drives these processes at the molecular level, and to unravel the differences between female and male paths of development. The theoretical background, design, and results of this project are summarized in article form in Chapter 2. At the time of printing the paper is submitted.

During my PhD course I also characterized the large URs in the mtDNAs of the species *Ruditapes philippinarum* and *Musculista senhousia*, and identified the respective CRs. The fine characterization of these non-coding sequences allowed me to identify features shared between F and M mt genomes that may regulate their replication and transcription. The identification of such structures will give a better knowledge on how the complex bivalve mt CRs are organized and how they function. Moreover, I analyzed the presence of lineage-specific ORFans in many DUI species, to gain insights into their evolution and conservation. Finally, because of specific knowledge on mtDNA gained during my PhD course, I was asked to contribute in writing of a review paper that summarizes the current knowledge of animal mtDNAs with an unusual gene content.

My works have been published in four papers, which are summarized in Chapter 3. Full articles are attached in Appendix 2.

# EARLY REPLICATION DYNAMICS OF SEX-LINKED MITOCHONDRIAL DNAs IN THE DOUBLY UNIPARENTAL INHERITANCE SPECIES *RUDITAPES PHILIPPINARUM* (BIVALVIA VENERIDAE)

## 2.1 INTRODUCTION

In animals, the mitochondrial genome (mtDNA) is usually transmitted maternally by avoiding paternal inheritance through spermatozoa. This ‘strictly maternal inheritance’ (SMI) is partly responsible for offspring homoplasmy, namely the presence of only one mitochondrial haplotype in an individual (as opposed to heteroplasmy). Homoplasmy is also thought to be maintained by a bottleneck during oogenesis that lowers the diversity of mtDNA variants in the mature egg (reviewed in Jokinen and Battersby, 2013, and Mishra and Chan, 2014). After fertilization, during early embryogenesis, another bottleneck is thought to occur due to an absence of mtDNA replication as cells rapidly divide, resulting into a drastic decline of mtDNA copy number per cell. The amplification of the reduced pool of mtDNA molecules inside cells takes place in a later developmental phase (Mishra and Chan, 2014). The maintenance of mitochondrial homoplasmy is commonly thought to be adaptive, because it avoids potentially harmful genomic conflicts that may arise between two different mtDNAs within a cell, and it promotes co-adaptation of interacting mitochondrial and nuclear genes (Hoekstra, 2011; Lane, 2011).

Heteroplasmy in an individual may result from maternal inheritance of multiple mtDNA variants that pass the bottleneck, from acquisition of *de novo* mutations during development and aging, or from paternal leakage. Heteroplasmy resulting from the coincidental presence of functional and deleterious mtDNAs is a cause of pathology in humans, when the harmful variant surpasses a certain copy-number threshold in a tissue or organ (Jokinen and Battersby, 2013; Wallace and Chalkia, 2013; Mishra and Chan, 2014). Stochastic segregation of heteroplasmic mtDNA variants during mitotic or meiotic cell division allows highly heterogeneous tissue distribution. Added to the stochastic segregation of mtDNA variants is the tissue-specific segregation of some mtDNA haplotypes that cannot be explained without invoking a selective process (Jokinen and Battersby, 2013; Wallace and Chalkia, 2013; Mishra and Chan, 2014). However, the mechanisms by which some mtDNA variants come to predominate in certain tissues or in the germ line remain poorly understood (see for example Burgstaller *et al.*, 2014), and this may reflect a lack of suitable *in vivo* model systems in which to study these processes.

Only one exception to SMI of mitochondria is known in animals, the ‘doubly uniparental inheritance’ (DUI) of mtDNA (Skibinski *et al.*, 1994a, 1994b; Zouros *et al.*, 1994a, 1994b), a mode of mitochondrial transmission confirmed in eight families of bivalve mollusks (Theologidis *et al.*, 2008). In the DUI system, two mtDNAs, named F and M, with high sequence divergence (*Mytilus* spp. nucleotide p-distance: 22 to 39%; amino acid p-distances in *Ruditapes philippinarum* and unionoids freshwater mussels: 34% and up to ~51%, respectively; reviewed in Zouros, 2013), co-exist in the same species and are inherited through separate routes. The F is transmitted by females through eggs to both sons and daughters, whereas the M is transmitted by males through spermatozoa to sons only (Breton *et al.*, 2007; Passamonti and Ghiselli, 2009; Zouros, 2013; Breton *et al.*, 2014). Females are essentially homoplasmic for the F mtDNA, while males are heteroplasmic carrying both F

and M mtDNAs. DUI females have been shown to produce progenies with variable sex ratios, i.e. female-biased, male-biased, or balanced (females and males in similar proportions), a feature that appears to be dependent on the mother only (Saavedra *et al.*, 1997; Kenchington *et al.*, 2002; Ghiselli *et al.*, 2012). Ghiselli *et al.* (2011) proposed the main steps of the DUI mechanism as a series of three consecutive checkpoints: in Checkpoint #1, which takes place shortly after fertilization, sperm mitochondria enter the egg and are maintained as an aggregate in male embryos, while they disperse in females (observed in *Mytilus* by Cao *et al.*, 2004a, Obata and Komaru, 2005, Cogswell *et al.*, 2006; and in *R. philippinarum* by Milani *et al.*, 2011 and Milani *et al.*, 2012); in Checkpoint #2 the M mt genome disappears only from females, following the dilution and/or degradation of sperm mitochondria; in Checkpoint #3 the M mtDNA is segregated in the male germ line and the F in the female one, becoming the dominant mtDNA in the gonad and the only mt line transmitted by sperms and eggs, respectively. The loss of M mtDNA in developing females, i.e. Checkpoint #2, has been studied in *Mytilus* (Sutherland *et al.*, 1998; Sano *et al.*, 2011) and appears to take place in the first 24 hours of development. To our knowledge, the proposed Checkpoint #3, which should occur before gamete production starts, has never been characterized so far.

The DUI system represents a promising model to study mechanisms controlling transmission, segregation, replication, and expression of the mitochondrial genome. DUI males are naturally heteroplasmic for the two highly divergent F and M mtDNAs: gonads contain a high amount of M mtDNA and produce sperms carrying only the M genome, whereas levels of heteroplasmy in somatic tissues can be variable depending on species. In adult *Mytilus* males, the M mtDNA is usually at low concentrations or absent in somatic tissues (reviewed in Zouros, 2013), whereas in *R. philippinarum*, adult male soma can contain high quantities of M, even more than the F mtDNA (Passamonti and Scali, 2001; Ghiselli *et al.*, 2011). The different relative quantities of the M genome in these two models could be

explained by a different ‘degree’ of aggregation of sperm mitochondria in male embryos: a weaker aggregation, as hypothesized in *R. philippinarum*, could let more copies of M to leak through the developing embryo, reaching blastomeres that give rise to somatic tissues (Milani *et al.*, 2012). Conversely, heteroplasmy seems to happen extremely infrequently in DUI females, except in *Mytilus*, where it is sometimes observed in female tissues, probably due to the failure of sperm mitochondria degradation (Ghiselli *et al.*, 2011; Zouros, 2013), hybridization events (reviewed in Brannock *et al.*, 2013), or being even typical of some populations where DUI disruption is frequent (Brannock *et al.*, 2013). Heteroplasmic females in *R. philippinarum* seem to be uncommon so far (Ghiselli *et al.*, 2011).

Evidence indicates that DUI relies on the same molecular machine of SMI typical of animals, with some modifications that let paternal inheritance of M mtDNA happen (Breton *et al.*, 2014). The preferential segregation of M mtDNA in male germ line seems more linked to active processes than to replication rate only (Venetis *et al.*, 2006; Ghiselli *et al.*, 2011; Milani *et al.*, 2011; Guerra *et al.*, 2014), while the driving forces behind F and M distribution in somatic tissues are still unknown. Studying F and M behavior and replication dynamics during female and male developmental paths is of central importance not only to understand the mechanism of DUI, but also to shed light on the factors that govern tissue-specific segregation of heteroplasmic variants during development, a topic of great importance for example in many human health issues (Lane, 2012; Jokinen and Battersby, 2013; Wallace and Chalkia, 2013; Mishra and Chan, 2014).

As a first step towards a better understanding of the mechanisms underlying mtDNA segregation and replication in a species with DUI, the present study explored the developmental dynamics of F and M mtDNA amounts using Real-Time qPCR in two groups of samples of *R. philippinarum*, respectively representing early (embryos and larvae) and advanced (sub-adults) developmental stages. Consistent with previous studies in mammals

and other animal models (reviewed in Jokinen and Battersby, 2013, and Mishra and Chan, 2014), our results indicate no detectable F and M mtDNA replication during the first phases of development, resulting in a dramatic reduction of mtDNA copies per cell owing to an increase in cell number. This reduction is followed by tissue- and sex-specific mtDNA replication boosts, most probably related to massive cell proliferation associated with the start of gonad and/or gamete formation.

## **2.2 MATERIALS AND METHODS**

### **2.2.1 SAMPLE CHARACTERISTICS**

Two groups of samples were analyzed in this work, provided by the ‘Centro Ricerche Molluschi (CRiM)’ (Goro, Italy). The first group, named ‘embryo series’, represents the early stages of *R. philippinarum* development and consists of individuals whose age was from 2 to 86 hours post-fertilization (hpf) (a time interval that covers the developmental stages of 8-cells embryos, trochophore, D-larva and veliger), divided into six different age subgroups (2hpf, 6hpf, 12hpf, 24hpf, 48hpf, and 86hpf). Spawning, fertilization and embryo collection were performed in controlled lab conditions. We assume a balanced sex ratio of each stage in the embryo series, given that (1) the spawning involved dozens of different females expected to produce progenies with varying sex-bias, (2) the overall sex ratio of wild populations of DUI species is 1:1 (Kenchington *et al.*, 2002; Ghiselli *et al.*, 2012), and (3) we collected a high number of individuals per stage (see below). These samples were preserved in 100% ethanol at 4°C until use.

The second group, named ‘young series’, was collected in the field, and is composed of individuals from 1mm to 20mm of shell length: the largest ones (15-20mm) were those

nearest to their first reproductive season (Devauchelle, 1990). These samples were grouped into four distinct classes (named 1, 2, 3, and 4) based on their dimensions. Table 1 enlists dimension intervals of classes and the number of specimens for each class. Samples of the young series were put in 100% ethanol immediately after collection and conserved at -80°C until use.

**Table 1** Characteristics and composition of the four classes of the young series. Sex of the specimens was assigned after the qPCR experiments, based on the absence/presence of a M mtDNA signal (see text for details).

Class	Shell length interval (mm)	# of specimens	# females	# males	Sample types			
					whole animal	body	adductor muscle	mantle
1	1-3	10	5	5	10	-	-	-
2	5-6	13	8	5	5	8	7	8
3	9-15	5	3	2	-	5	-	-
4	18-20	16	8	8	-	16	6	8

### 2.2.2 DNA EXTRACTIONS

For the embryo series, DNA was extracted using pools of individuals, because extractions from single individuals gave too little yield. The overall developmental stage of specimens was checked under a light microscope before extraction. Ten pools for each of the six stages were used for the extractions, for a total of 60 pools. We sampled approximately 100 individuals per pool, that corresponds to ~1,000 individuals per stage, for a total of ~6,000 among embryos and larvae. Each pool was identified with a unique name composed by the stage and a letter (see Supplementary Table 1).

DNA extractions for the young series were performed from whole animals for the smallest specimens, or from bodies for the bigger specimens (the sample identified as ‘body’ consists of the central visceral mass of the animal, with foot, gills and digestive gland

removed). To analyze the mtDNA content variation in other tissues, we also extracted DNA from adductor muscles and mantles from classes 2 and 4 specimens (see Table 1 for sample size of these tissues). To avoid contaminations, single animals were processed separately, a different scalpel was used to dissect each tissue, all reusable tools (glass plates and tweezers) were carefully washed and rinsed between dissections, and all disposable materials (gloves and table covers) were changed every time. After dissection, tissues were kept frozen in separate tubes at -80°C until DNA extraction. A unique name composed of class number, a letter identifying the specimen, and a letter identifying the tissue was given to each juvenile sample (see Supplementary Table 2).

Extractions of embryo pools were performed using MasterPure™ Complete DNA and RNA purification kit (Epicentre) or DNeasy® Blood & Tissue kit (Qiagen) (extraction modes for each pool are specified in Supplementary Table 1), while DNA from juvenile samples was extracted using MasterPure™ kit (reducing the quantities of reagents when needed to obtain a better yield). Concentration and quality of the extractions were determined with a NanoDrop 2000 (Thermo Scientific) or a Biodrop™ DUO (Biodrop).

### **2.2.3 REAL-TIME qPCR PROCEDURE**

To quantify the amounts of F and M mtDNAs, Real-Time qPCR experiments were performed in multiplex using TaqMan® chemistry, following the approach developed and utilized by Ghiselli *et al.* (2011) and Milani *et al.* (2012). Three targets were considered: the nuclear heat shock protein 70 (*hsp70*) gene, the mitochondrial NADH dehydrogenase subunit 1 (*nad1*) and the small subunit of ribosomal RNA (*12S*) genes for the F and M mtDNA, respectively (refer to Ghiselli *et al.*, 2011, for details on the choice of target sequences). Primers were provided by Alpha DNA (Montréal, Québec, Canada) and TaqMan® probes by Sigma-Aldrich (The Woodlands, Texas, USA). See Supplementary Table 3 for the

characteristics of primers and probes. All experiments were performed in triplicates on a PikoReal™ 96 Real-Time PCR System (Thermo Scientific) with the DyNAmo ColorFlash Probe qPCR Kit (Thermo Scientific), in a 10µL total volume. The PCR cycle consisted in 2min at 50°C for uracil-N-glycosylase treatment, an initial denaturation and polymerase activation for 7min at 95°C, then 40 cycles of denaturation at 95°C for 5sec and annealing plus extension at 60°C for 30sec.

Following the method developed by Gallup and Ackermann (2008), we built a Stock1 dilution for all 60 embryo pools and three for young series samples, composed respectively of classes 1, 2, and 3 plus 4. All Stock1 dilutions were used to build four separate 1:10 dilution series (from 100ng to 0.1ng of DNA in reaction) to calculate the efficiencies of the targets in Real-Time. For all four Stock1 dilution series, primers and probes concentrations were adjusted to obtain the best possible combination of efficiencies: these concentrations (enlisted in Supplementary Table 4) were then used for the quantifications of the respective samples. Efficiency of the three targets in each Stock1 dilution series (enlisted in Supplementary Table 4) was calculated automatically by PikoReal™ Software 2.2 (Thermo Scientific). Quantification experiments on all samples were performed using 100ng of DNA in reaction, except for class 2 adductor muscles where 10ng were used because of the low quantity of DNA obtained during extraction.

#### **2.2.4 STATISTICAL ANALYSES**

Ratios of F and M mtDNAs in samples from both series, normalized to the nuclear *hsp70* quantity, were calculated using the efficiency-corrected ratio method (equation 3.5 from Pfaffl, 2004) for all reactions, while the procedure of Livak and Schmittgen (2001) for multiplex reactions was used to calculate average ratio values. Replicates in which the nuclear target failed to amplify, or showing amplification issues for the mt targets, were excluded

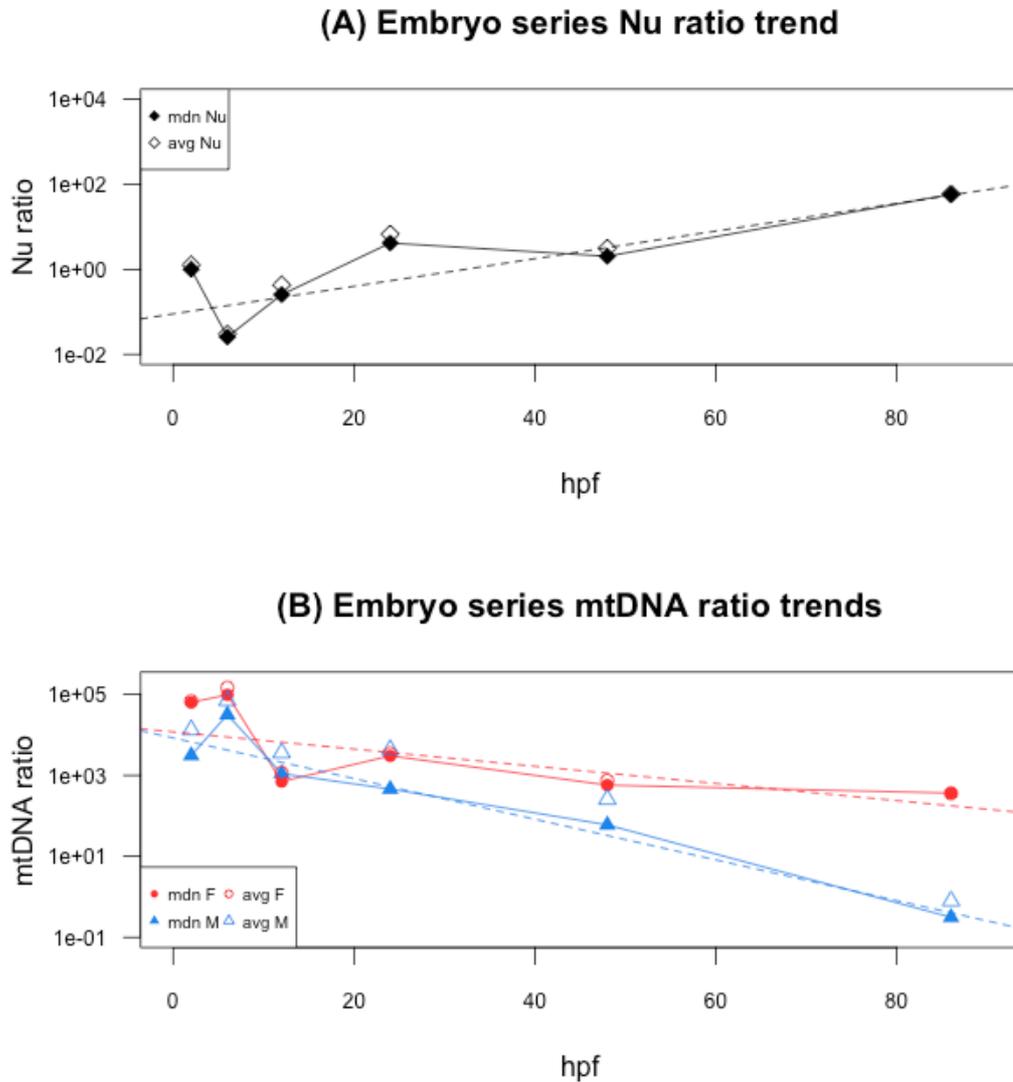
from the subsequent calculations and analyses (see Supplementary Tables 1 and 2 for the excluded reactions). We avoided normalization to a calibrator sample for the mtDNA ratios in both series because: (1) we would have needed to choose it arbitrarily for each group, and (2) it is not technically correct to directly compare ratios from different experiments normalized to different calibrators. To calculate the growth in nuclear content in the embryo series, we normalized the content of each pool to the *hsp70* median quantification cycle (Cq) of stage 2hpf, the first of the series, using equation 3.3 from Pfaffl (2004). Statistical analyses (Wilcoxon test, cluster analysis using the single linkage method, linear models on the  $\log_{10}$  transformed ratio values) and graphic elaborations were performed with R 3.1.0 (R Core Team, 2014) on RStudio 0.95 (RStudio, 2012).

## 2.3 RESULTS

A total of 399 triplex quantifications for 133 samples were performed in this study, 180 for the 60 embryo series pools and 219 for the 73 samples of the young series. All Cq values and ratio calculations for each technical replicate in the embryo and young series, as well as the excluded replicates, are enlisted in Supplementary Tables 1 and 2, respectively.

Trends in median ratios for the three targets of each stage in the embryo series are represented in Figure 11, while detailed statistics of the ratio distributions are shown in Supplementary Table 5. Given the fluctuating ratios between stages 2hpf and 12hpf, likely because of low nuclear target abundance, we cannot exclude erroneous estimations due to qPCR amplification issues in these stages. Nonetheless, the general trends of targets are clear and supported by the linear models (Figure 11): the nuclear content grows as expected from 2hpf to 86hpf (Figure 11A), while both F and M mtDNAs ratios decrease in time (Figure

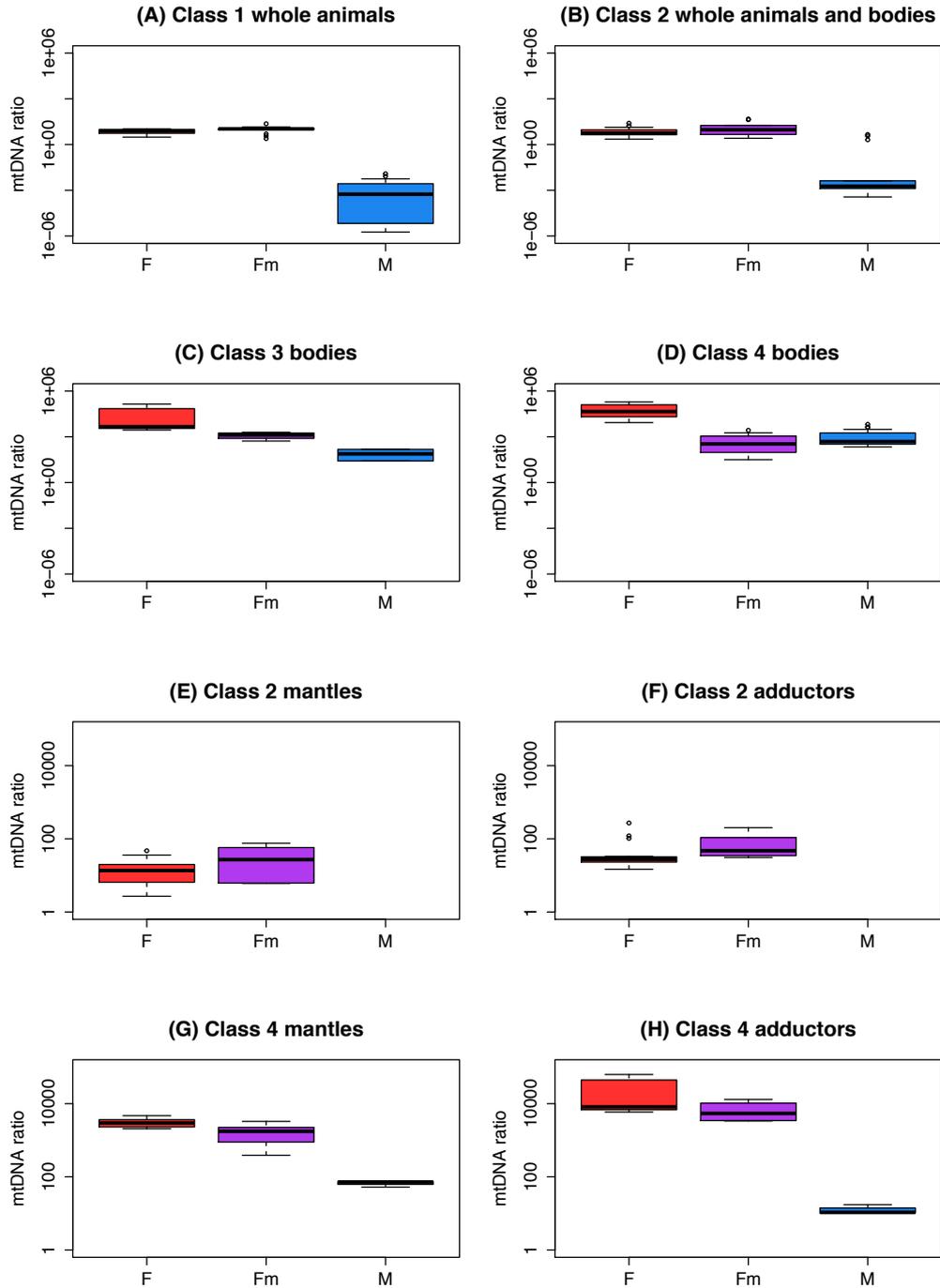
11B). The M ratio trend becomes clearly distinct from the F one after 24hpf, showing a more steep descending after this stage (Figure 11B, Supplementary Table 5).



**Figure 11** Trends of the three targets in embryos and larvae from 2hpf to 86hpf. Abbreviations: mdn, median ratio; avg, average ratio; Nu, nuclear *hsp70* ratio; F, F mtDNA ratio; M, M mtDNA ratio. Full lines, trends of median ratio; dotted lines, linear model trends calculated on all suitable technical replicates. Lines color code: black, nuclear *hsp70*; red, F mtDNA; blue, M mtDNA. Y axis are in  $\log_{10}$  scale. See Supplementary Table 5 for median absolute deviations and standard deviations of all ratios for each stage.

(A) Stage 2hpf median nuclear ratio is 1, as the median Cq of this stage has been used as the reference to calculate all nuclear ratios (see Materials and Methods). Nuclear target linear model: adjusted  $R^2 = 0.67$ , p-value  $< 2.2E-16$ .

(B) F mtDNA linear model: adjusted  $R^2 = 0.45$ , p-value  $< 2.2E-16$ . M mtDNA linear model: adjusted  $R^2 = 0.84$ , p-value  $< 2.2E-16$ .



**Figure 12** Boxplots representing the mtDNA ratio distributions in the young series samples. Abbreviations: F, F mtDNA ratio in females; Fm, F mtDNA ratio in males; M, M mtDNA ratio in males. Y axis are in log<sub>10</sub> scale. See Table 1 for sex ratio of each class and sample size of each tissue type, and Table 2 for p-value of the comparisons among targets and classes using the Wilcoxon test. Supplementary Table 6 enlists detailed statistics of the distributions.

(A-D) Ratio distributions in whole animals and bodies.

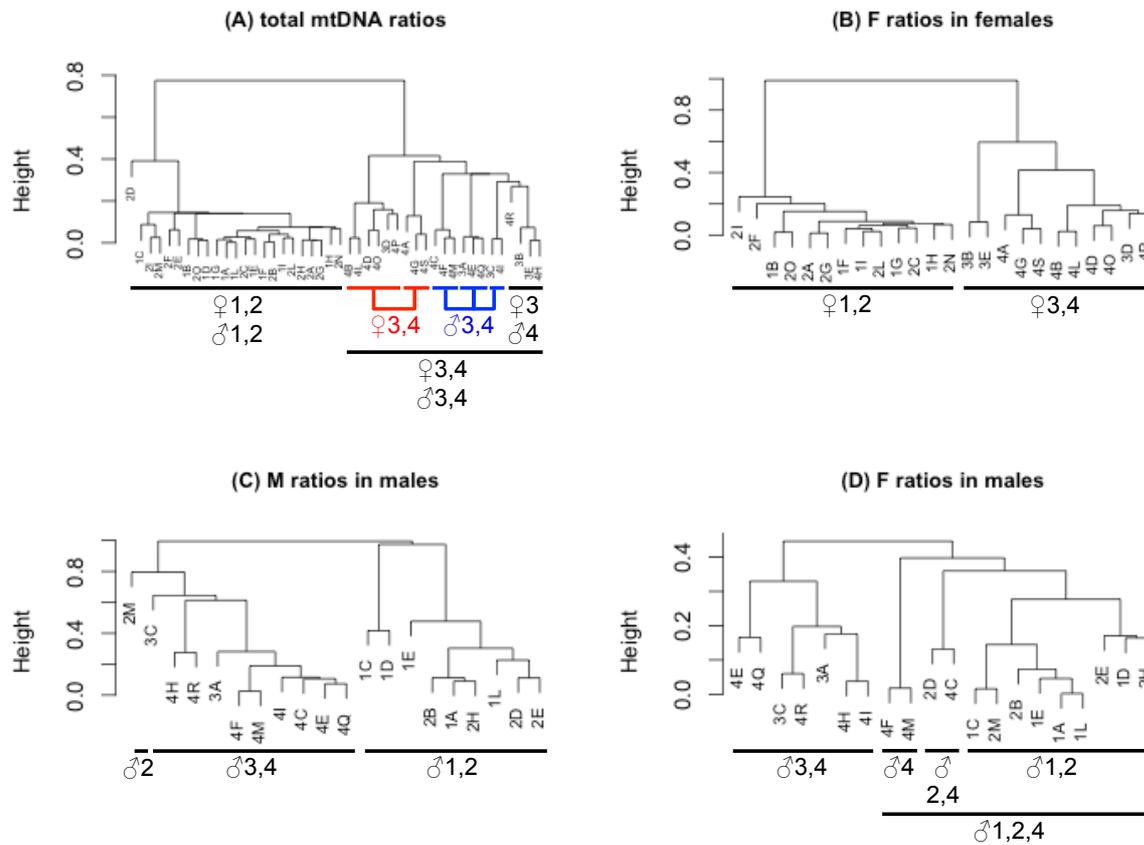
(C and G) Ratio distributions in mantles.

(F and H) Ratio distributions in adductor muscles.

Specimens of the young series have been considered ‘males’ or ‘females’ based on the presence or absence of M mtDNA signal in their bodies. This sexing method is not error-free, but given previous observations (Ghiselli *et al.*, 2011), we assume that female heteroplasmy is infrequent in *R. philippinarum*, and there is no other consistent way to sex clams at this developmental stage. On a total of 44 animals, 24 females and 20 males were determined this way (Table 1). Boxplots in Figure 12A-D show the distribution of mt ratios in both sexes for each class, while Supplementary Table 6 contains the detailed statistics of mt ratios in these samples.

Average mt ratio values of whole animal and body samples of the young series have been used for a cluster analysis: see Figure 13 for the results and Supplementary Figure for a graphic resume of the values. Cluster analysis separates the samples in two main groups based on their total mtDNA ratio (average F ratio + average M ratio), as seen in Figure 13A. One cluster contains all specimens from classes 1 and 2 in a non-ordered fashion, irrespective of the sample sex and/or class. The other cluster contains all specimens from classes 3 and 4, and can be divided in six sub-clusters, five of which are sex-specific (see Figure 13A). To better understand this clustering, we separated the samples in four groups (females from classes 1 and 2, females from classes 3 and 4, males from classes 1 and 2, and males from classes 3 and 4) and performed comparisons among them using the Wilcoxon test. There is no difference in total mt ratio between sexes in classes 1 and 2 (p-value = 0.11), but this difference is significant in classes 3 and 4 (p-value = 3.97E-05), which explains the nearly perfect distinction between females and males in the cluster containing these two classes (Figure 13A). The average ratios of F in females and M in males (Figures 13B and 13C) are each separated in two clusters: one containing specimens from the first two classes and the other those from the last two. Ratios of F mtDNA in males (Figure 13D) can be also divided

in two major clusters containing either males of classes 3 and 4, or males from classes 1, 2, and 4.



**Figure 13** Cluster analysis results for the average mtDNA ratios in whole animals and bodies of the young series (specimens from 1mm to 20mm). Sample names are composed of a number indicating their class and a letter. Symbols under the clusters indicate the sex of the specimens inside them (♀, females; ♂, males), while numbers indicate their class. See Supplementary Figure and Supplementary Table 2 for average ratio values and standard deviations of each specimen.

(A) Dendrogram of average total mtDNA ratios (average F ratio + average M ratio). Classes 1+2 are clearly separated from classes 3+4. Females and males from classes 3 and 4 tend to group in sex-specific clusters (highlighted in red and blue, respectively). A cluster contains two class 3 females (3B and 3E) plus two class 4 males (4H and 4R): the two females have the lowest total mtDNA ratios among class 3 and 4 females, while the two males have the highest among males of the same classes. There is no difference in total mtDNA ratio between females and males of classes 1 and 2, while females and males of classes 3 and 4 are significantly different (see Results section for Wilcoxon test p-value of these comparisons).

(B) Dendrogram of average F ratio in females. Specimens from classes 1 and 2 are well separated from classes 3 and 4. [continues in next page]

**Figure 13** [continues] (C) Dendrogram of average M ratio in males. Males from classes 1 and 2 cluster separately from those of classes 3 and 4; specimen 2M is an outlier of class 2 with high quantities of M mtDNA (see Supplementary Figure), therefore it is more closely related to the cluster containing classes 3 and 4.

(D) Dendrogram of average F ratio in males. Animals from classes 3 and 4 tend to cluster together, but three specimens from class 4 (4C, 4F, 4M) cluster closer to males from classes 1 and 2: these animals have low F mtDNA ratios, more similar to those of classes 1 and 2 (see Supplementary Figure).

Table 2 reports the significance of the Wilcoxon test performed on comparisons between ratios within the same class in whole animal and body samples (see also Figure 12A-D). In males, M mtDNA ratios are always significantly different from F ratios (Fm vs M comparisons), except for class 4, where the two are comparable (even if M ratio is higher than F in all but one male from this class; Supplementary Figure). F levels between females and males (F vs Fm comparisons) are slightly different in class 1, comparable in class 2, and become significantly different in classes 3 and 4, where females have more F mtDNA than males (Supplementary Figure). Finally, F levels in females are always significantly different from M levels in males in all classes (F vs M comparisons).

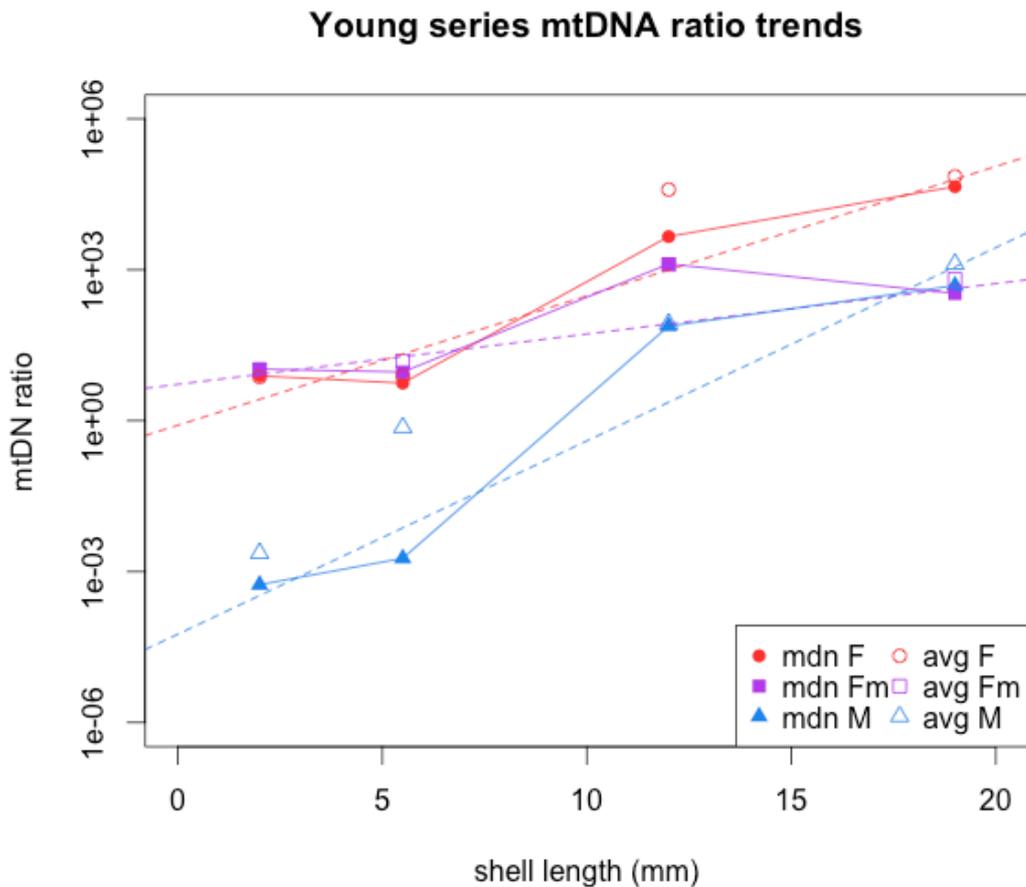
The trends of median mtDNA ratios in whole animal and body samples, and the respective linear model trends, are shown in Figure 14. Significance of ratio comparisons between classes using the Wilcoxon test is shown in Table 2. F in both sexes and M ratios in males are similar between classes 1 and 2 (1 vs 2 comparisons), but they grow significantly between classes 2 and 3 (2 vs 3 comparisons) (Figure 14). Both F and M ratios grow from class 3 to 4, with the M showing the most significant progression (3 vs 4 comparisons).

No M mtDNA was detected in class 2 adductor muscles and mantles (Figure 12E-F, Supplementary Table 6), while it was detected in only one mantle and one adductor sample from two different males of class 4 (Figure 12G-H, Supplementary Tables 2 and 6). Table 2 shows the significance of the Wilcoxon test comparisons for these tissues. In class 2 females,

<i>Tissues</i>	<i>Classes</i>	<i>Ratios</i>	<i>p-value</i>	<i>Significance</i>
whole animal + body	1	F vs Fm	0.011	*
	2	F vs Fm	0.146	ns
	3	F vs Fm	3.996E-04	***
	4	F vs Fm	6.202E-14	***
	1	Fm vs M	1.289E-08	***
	2	Fm vs M	1.728E-06	***
	3	Fm vs M	2.165E-03	**
	4	Fm vs M	0.071	ns
	1	F vs M	1.289E-08	***
	2	F vs M	4.033E-08	***
	3	F vs M	3.996E-04	***
	4	F vs M	6.202E-14	***
	1 vs 2	F vs F	0.484	ns
	2 vs 3	F vs F	5.186E-08	***
	3 vs 4	F vs F	0.020	*
	1 vs 2	Fm vs Fm	0.806	ns
	2 vs 3	Fm vs Fm	3.69E-05	***
	3 vs 4	Fm vs Fm	0.029	*
	1 vs 2	M vs M	0.057	ns
	2 vs 3	M vs M	5.16E-05	***
3 vs 4	M vs M	3.368E-06	***	
mantle	2	F vs Fm	0.494	ns
	4	F vs Fm <sup>1</sup>	1.115E-03	**
	4	Fm vs M	4.396E-03	**
	4	F vs M	4.396E-03	**
	2 vs 4	F vs F	2.312E-08	***
	2 vs 4	Fm vs Fm	1.077E-04	***
adductor	2	F vs Fm	0.029	*
	4	F vs Fm <sup>2</sup>	0.062	ns
	4	Fm vs M	9.091E-03	**
	4	F vs M	9.091E-03	**
	2 vs 4	F vs F	1.53E-06	***
	2 vs 4	Fm vs Fm	3.996E-04	***
mantle vs adductor	2	F vs F	2.37E-03	**
	2	Fm vs Fm	0.309	ns
	4	F vs F	6.804E-06	***
	4	Fm vs Fm	6.804E-06	***
	4	M vs M	0.100	ns

**Table 2** Wilcoxon test p-values of the mt ratio comparisons in the young series. Abbreviations: F, F mtDNA ratio in female samples; M, M mtDNA ratio in male samples; Fm, F mtDNA ratio in male samples. Significance of p-values: ns, not significant; \* <0.05; \*\* <0.01; \*\*\* <0.001. The Wilcoxon test has been applied to mt ratio comparisons for couples of targets in the same class, and for single targets between sexes. Comparisons involving the M ratio of class 2 adductors and mantles have not been performed, since no M mtDNA was detected in these tissues. <sup>1</sup>: this test significance remains the same (p-value = 9.288E-03, significance = \*\*) when the male mantle showing an M mtDNA signal is removed from the comparison. <sup>2</sup>: this test significance remains the same (p-value = 0.388, not significant) when the male adductor showing an M mtDNA signal is removed from the comparison.

F ratios in adductor muscle are higher than in the mantle (F vs F comparisons), while in males this difference is not significant (Fm vs Fm comparisons); in class 4, in both sexes, F ratios are higher in the adductor muscle than in the mantle (mantle vs adductor, F vs F and Fm vs Fm comparisons). No significant difference between these two tissues is found in the M ratios



**Figure 14** Trends of mtDNA ratios in whole animal and body samples from the young series (specimens from 1mm to 20mm). Abbreviations: mdn, median ratio; avg, average ratio; F, F mtDNA ratio in females; Fm, F mtDNA ratio in males; M, M mtDNA ratio in males. Full lines, trends of median ratio; dotted lines, ratio linear models. Lines color code: red, F mtDNA in females; purple, F mtDNA in males; blue, M mtDNA in males. Y axis is in  $\log_{10}$  scale. Points are positioned on the average shell length of the respective class (class 1, 2mm; class 2, 5.5mm; class 3, 12mm; class 4, 19mm; see Table 1 for length intervals of each class). Class 2 values comprise both bodies and whole animals. See Figure 12A-D for ratio distributions, and Supplementary Table 6 for detailed statistics of each ratio. F ratio linear model: adjusted  $R^2 = 0.88$ ,  $p$ -value  $< 2.2E-16$ . Fm ratio linear model: adjusted  $R^2 = 0.59$ ,  $p$ -value  $= 3.70E-13$ . M ratio linear model: adjusted  $R^2 = 0.86$ ,  $p$ -value  $< 2.2E-16$ .

for males of class 4 (mantle vs adductor, M vs M comparison), but the comparison is made only between replicates of two samples. Class 4 females have higher F mtDNA ratios than males in the mantle (Figure 12G), but in the adductor muscle (Figure 12H) the difference is not significant (F vs Fm comparisons): the significance of these two tests remains the same even when we remove the heteroplasmic male samples from comparisons, although the  $p$ -

values show a small increase (Table 2). When comparing homoplasmic versus heteroplasmic tissues, the difference in F mtDNA content was significant (mantles p-value = 2.45E-03, adductors p-value = 9.88E-04). Finally, class 4 adductors and mantles always have higher ratios than those of class 2 (F vs F and Fm vs Fm comparisons; Figure 12E-H).

## 2.4 DISCUSSION

In this work we analyzed nuclear and mitochondrial DNA replication dynamics during early development of the venerid species *R. philippinarum* (from 2hpf to 86hpf, and in young individuals from 1 to 20mm of shell length, the largest of which were approaching their first reproductive season, according to Devauchelle, 1990). The nuclear genome quantity grows as expected during embryogenesis (Figure 11A): many cellular divisions do occur and the total number of nuclei per specimen grows consequently. On the contrary, mtDNA per cell content seems to decrease during embryogenesis (Figure 11B). A rise of mtDNA ratios is observed later in development, providing an increase in F in both sexes and M in males (Figure 14).

### 2.4.1 MTDNA REPLICATION IS DORMANT DURING EARLY EMBRYOGENESIS

The observed decrease of F mtDNA ratio in our pools of embryos and larvae of about two orders of magnitude (from  $\sim 10^4$  to  $\sim 10^2$ ; see Figure 11B and Supplementary Table 5) is evidently an outcome of the denominator of the ratio, i.e. the amount of nuclear target, which increases from 2hpf to 86hpf (Figure 11A). At the same time, this indicates an undetectable, weak, or absent mtDNA replication until 86hpf (broadly corresponding to the veliger stage), which results in a significant decrease of mtDNA copies per cell. The absence of detectable replication of the maternally-transmitted F genome during *R. philippinarum* embryogenesis

(maybe until the veliger stage, ~86hpf) is consistent with previous reports of reduction in mtDNA content per cell during early development in other animal groups (see Milani *et al.*, 2012). Interestingly, in *R. philippinarum*, the paternally-transmitted mtDNA seems to behave like the F at least until 24hpf. Compared to the F ratio, the M mtDNA ratio shows a steeper drop, from  $\sim 10^4$  to  $\sim 10^{-1}$ , with a well distinguishable decrease after 24hpf (Figure 11B; see median and average M values in Supplementary Table 5). As above, this drop can be explained by the absence of detectable replication of the M genome during early embryogenesis (at least until 24hpf, D-larva stage), coupled with the loss of M mtDNA in embryos that are going to develop into females (according to the DUI routes of mtDNA transmission).

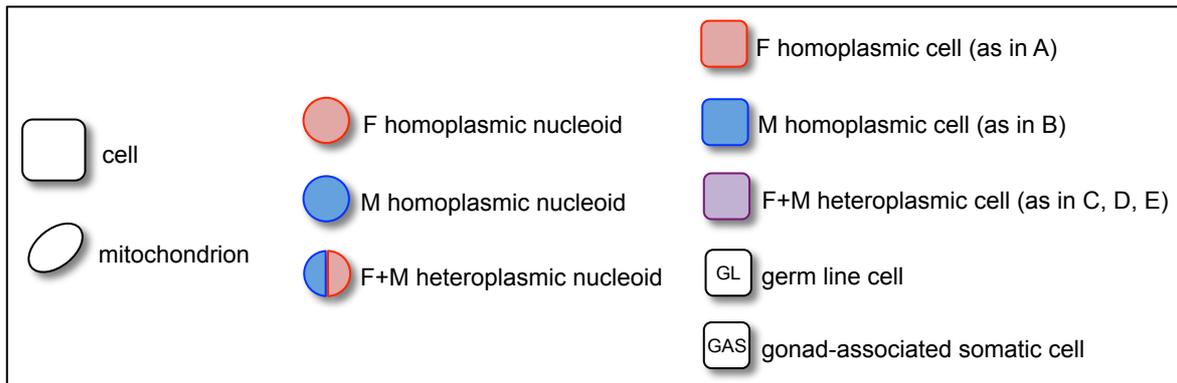
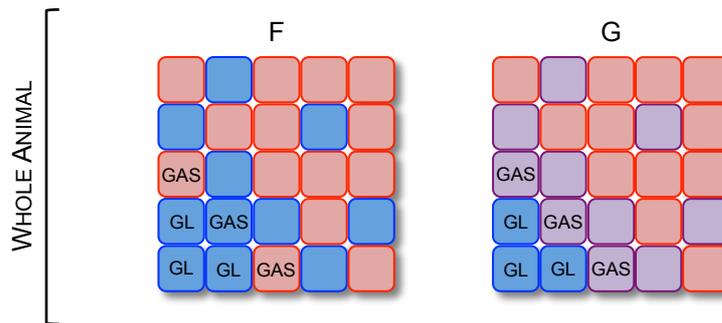
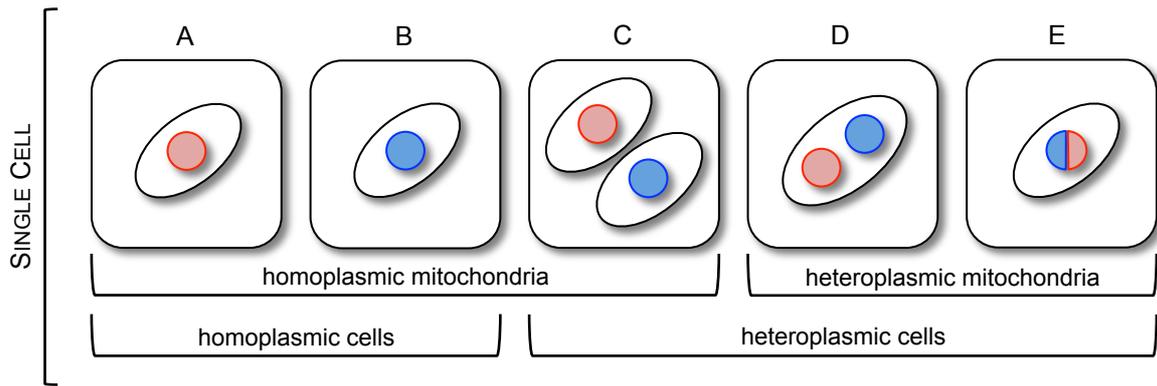
Early embryogenesis of DUI species is thought to be the time frame in which sperm mitochondria are eliminated/diluted in females and retained in male offspring, i.e. Checkpoint #2 described by Ghiselli *et al.* (2011). Several works on *Mytilus* suggest that Checkpoint #2 is not a sudden event in DUI species. In the pioneering work of Sutherland *et al.* (1998) on *Mytilus edulis*, the presence/absence of M mtDNA was tested in larvae at 18hpf, 24hpf, and 48hpf using PCR. 18hpf larvae from both female-biased and balanced progenies showed an identical M mtDNA signal, whereas starting from 24hpf the signal was absent in the female-biased progenies and in some larvae from the balanced progenies (see Saavedra *et al.*, 1997, and Kenchington *et al.*, 2002, for discussions on the topic). Sano *et al.* (2011) used Real-Time qPCR to assess the F/M mtDNA ratio variation up to 24hpf in *Mytilus galloprovincialis* larvae of female- and male-biased offsprings. In male-biased larvae the F/M ratio remained the same up to 24hpf, meaning that no sperm mitochondria elimination occurred (and that the relative quantities of F and M did not change), while in female-biased larvae this ratio increased after 3hpf, i.e. M mtDNA has started to diminish earlier in this group. From these studies it appears that in *Mytilus* Checkpoint #2 begins before 24hpf, and that after that time

M mtDNA is no more detectable in females. In *R. philippinarum*, Milani *et al.* (2012) found no changes in both F and M mtDNA copy number up to 2hpf, using a multiplex Real-Time qPCR approach on pools of embryos from different crosses, hence the Authors concluded that Checkpoint #2 happens after 2hpf in this species.

The pools of *R. philippinarum* embryos used in this study were composed of specimens coming from a mass spawning fertilization event that involved dozens of males and females, so no prior information about offspring sex bias was available. Nevertheless, considering the large number of animals per pool, the fact that they were born from many different females producing differently sex-biased progenies, and that the overall sex ratio of wild populations of DUI species does not deviate significantly from 1:1 (Kenchington *et al.*, 2002; Ghiselli *et al.*, 2012), we can assume a balanced sex ratio for each stage. Therefore, the drop of M mtDNA ratio starting from 24hpf can be ascribable to Checkpoint #2, if we assume that the M mtDNA becomes undetectable from a certain point on from half of the embryos (the females; Sutherland *et al.*, 1998), while it persists without amplification in the other half (the males). We also have to consider a stronger effect of dilution for the M mt genome in males compared to the F, given the much smaller initial quantity of M, as seen in *Mytilus* (Sutherland *et al.*, 1998; Sano *et al.*, 2011). Even if the exact timing of Checkpoint#2 could not be precisely determined given the chosen experimental approach, and considering that the M mtDNA observed after 24hpf is most likely carried only by males, the loss of M from females probably ends around 24hpf in *R. philippinarum*, a situation similar to that of *M. edulis* observed by Sutherland *et al.* (1998).

#### **2.4.2 MITOCHONDRIAL HETEROPLASMY DURING MALE DEVELOPMENT**

In animals with maternal inheritance of mitochondria, the absence of replication during early embryogenesis constitutes a bottleneck that reduces mtDNA copy number in



**Figure 15** Mitochondrial heteroplasmy in male cells (A-E) and whole animals (F and G) in DUI species. (A-B) Homoplasmic cells contain only mitochondria homoplasmic for either F or M. (C) If the embryogenesis bottleneck in developing males is not able to separate egg and sperm mitochondria in different cells, this may result in heteroplasmic cells containing differently homoplasmic mitochondria. (D) A heteroplasmic cell may contain heteroplasmic mitochondria that carry both F and M mtDNA in different nucleoids: this kind of mitochondria can arise from fusion of differently homoplasmic mitochondria as in (C). (E) Mitochondria in a heteroplasmic cell may carry heteroplasmic nucleoids that contain both F and M mtDNA molecules: in this situation, which can derive from either (C) or (D) scenarios, the two mt genomes are in strict physical contact and may recombine. (F) If the embryogenesis bottleneck is narrow enough to produce only homoplasmic cells, a male individual can be considered as a mitochondrial mosaic. [continues on next page]

**Figure 15** [continues] (G) The bottleneck may not be able to fully separate the two mt lines in different cells: this may result in the occurrence of [continues on next page]heteroplasmic cells, especially around the sperm mitochondria aggregate, where the relative concentrations of M are higher. (F and G) Germ line cells (indicated as GL) are always homoplasmic for M, as they will produce homoplasmic spermatozoa each reproductive season. Male gonad samples contain traces of F mtDNA, most likely carried by the somatic fraction of the sample (cells indicated as GAS). Sperm mitochondria that leak from the aggregate can locate themselves in somatic tissues. The heteroplasmy of these tissues can be ascribable to M-homoplasmic cells (as in F) or heteroplasmic cells (as in G). The quantity of M and the number of cells containing M mtDNA in soma are probably proportional to the degree of leakage from the sperm mitochondria aggregate.

proliferating cells, and consequently their heteroplasmy (Mishra and Chan, 2014). The mitochondrial composition in these cells will in turn influence the levels of heteroplasmy in tissues deriving from them. In DUI males, the leakage of M mtDNA from the sperm mitochondria aggregate (Milani *et al.*, 2012) and the stochastic action of the bottleneck in early embryogenesis may explain the scattered and variable presence of the M mt genome in adult somatic tissues: mitochondria leaking from the aggregate during early embryogenesis are randomly segregated in variable proportions in different blastomeres, and the quantities of M inside these cells are further exacerbated during the bottleneck, since the number of mtDNA molecules per cell diminishes at every cell division. On the other hand, cells including most of the mtDNAs of the sperm aggregate may have a higher probability of reaching homoplasmy for M. This passive process may indeed help achieving homoplasmy in germ line cells (Checkpoint #3; Ghiselli *et al.*, 2011) and avoid genetic conflicts between F and M mt genomes by physically separating the two lines. Other processes that actively select one mtDNA over the other are also thought to be involved in this checkpoint (Venetis *et al.*, 2006; Ghiselli *et al.*, 2011; Milani *et al.*, 2011), but their nature and role are still unknown.

From our results, we cannot infer the minimum number of mtDNA molecules per cell at the end of the embryogenesis bottleneck in bivalves, as this trait can be influenced by many factors such as the starting number of mitochondria and mtDNA molecules in the zygote, and

the number of cells before replication resumption. However, if the absence of mtDNA replication extends until the veliger stage, which is a complex larva with a high number of cells, the number of mt genomes per cell could be extremely low (for example, early studies in mice suggested one mtDNA molecule per cell at implantation, a stage with less cells than a bivalve veliger; Pikó and Taylor, 1987).

In a DUI male, the number of mitochondria and the amount and type of mt genomes (F or M) carried by a given cell can influence its heteroplasmic state. The heteroplasmy of a single cell can be ascribable, at the lowest levels, to the heteroplasmy of nucleoids carried by mitochondria and/or to the types of mitochondria carried by the cell (Figure 15A-E). Mitochondria can be homoplasmic by containing nucleoids composed of only one mt line (F or M), and if a cell possesses mitochondria carrying only one mtDNA type it is homoplasmic for that mt genome (Figure 15A-B): gametes, for example, fall into this category. On the other hand, a cell can be heteroplasmic in three possible ways, by carrying: (1) two kinds of homoplasmic mitochondria (like a zygote just after fertilization, or in later stages of development if the bottleneck does not separate sperm and egg mitochondria in different cells) (Figure 15C), (2) heteroplasmic mitochondria with two types of homoplasmic nucleoids (which could result from fusion of differently homoplasmic mitochondria) (Figure 15D), or (3) heteroplasmic mitochondria with heteroplasmic (F+M) nucleoids (from fusion of F and M homoplasmic nucleoids) (Figure 15E).

At the organismal level, a DUI male can be heteroplasmic in different fashions (Figure 15F-G). The two most probable possibilities are that it is either a mosaic composed of different homoplasmic cells (Figure 15F), or it is composed of both homoplasmic and heteroplasmic cells (Figure 15G), depending on the capacity of the embryogenesis bottleneck to generate homoplasmic cells. As stated above, homoplasmic cells for the M mt genome will arise with higher probability in correspondence of the sperm mitochondria aggregate during

early embryogenesis. These cells include germ line cells, which will produce M-homoplasmic spermatozoa. F mtDNA has been detected in low amounts in gonad samples of *R. philippinarum* but not in sperm (Ghiselli *et al.*, 2011). Given the constitution of *R. philippinarum* gonad samples, that is an abundance of acini containing gametes, but intimately wrapped to connective tissue and intestine, the F mtDNA is most likely ascribable to the somatic cell fraction of the sample (Figure 15F-G). Sperm mitochondria leaked from the aggregate can segregate in somatic cells, but the homoplasmy (Figure 15F) or heteroplasmy (Figure 15G) of these cells is unknown; the number of somatic cells containing M mtDNA and the amount of this mt genome, in both cases, would be proportional to the degree of leakage.

All the possibilities described above have interesting implications for a DUI species, and also for the study of mitochondrial heteroplasmy and mito-nuclear interactions. Under these scenarios, the nuclear genome of a male individual (or of an accidentally heteroplasmic female that failed to eliminate sperm mitochondria during Checkpoint #2) would have to interact in a given cell with either F or M mtDNA (Figure 15A, B, and F), or with both mt lines contemporarily (Figure 15C, D, E, and G), and still be able to maintain cell, tissue, and whole animal functionality. Analogous situations of heteroplasmy for two functional mt lines have been produced artificially in a number of studies on mice, with variable effects on the viability of the specimens (see for example Sharpley *et al.*, 2012, and references therein). In addition, the occurrence of heteroplasmic mitochondria and/or nucleoids (Figure 15D-E) can explain, and is supported by, the detection of recombinant mt genomes in *Mytilus* (Zouros, 2013). On the contrary, no clear indications of mitochondrial recombination are available for *R. philippinarum*: so far, only one F×M recombinant sequence was found in soma by Passamonti *et al.* (2003). Dedicated studies on these species can help solving the puzzle of cellular heteroplasmy in DUI species, acting as a model also for species with SMI of

mitochondria where heteroplasmy can be source of pathologies, and the scenarios described in this section could serve as working hypotheses for future investigations.

### **2.4.3 APPROACHING ADULTHOOD: SEX-SPECIFIC MTDNA DYNAMICS IN YOUNG CLAMS**

The series of sub-adult specimens of *R. philippinarum* was divided in four dimensional classes, and animals classified as females or males based on the absence or presence of an M mtDNA signal in their bodies (Table 1). As mentioned, there was no other reliable way to sex those specimens, even though accidental heteroplasmic females can be wrongly sexed as males. Female mitochondrial heteroplasmy appears to be uncommon in *R. philippinarum*: in the study of Ghiselli *et al.* (2011) only two on 33 females (6.06%) were found heteroplasmic in soma. Based on these observations, and given the results from our embryo series (i.e. M mtDNA signal in females may be lost after 24hpf), the utmost care in handling samples, and the sensitivity of our experimental approach, we are confident that categorizing as females all the animals in which bodies no trace of M mtDNA was detected is the best approximation of the real sex of the clams. Following this, the overall sex ratio of the young series resulted well balanced (24 females and 20 males), which is in line with the ratios observed in wild populations of DUI species (Kenchington *et al.*, 2002; Ghiselli *et al.*, 2012).

We observed that the total mtDNA ratio does not change significantly in whole animals and bodies (i.e. the central visceral mass after removing the foot, gills, and digestive gland) of the first two classes (1 and 2), and that in this group there is no difference between females and males, i.e. homoplasmic and heteroplasmic individuals, respectively (Figure 13A). The low levels of M present in males from classes 1 and 2 (Figure 12A-B, Supplementary Table 6, Supplementary Figure) do not affect the total mtDNA content (Figure 13A). On the contrary, specimens from classes 3 and 4 can be clearly distinguished by sex based on their total mtDNA ratio (Figure 13A), as female bodies usually have more total mtDNA than male

bodies (Figure 13A, Supplementary Figure). Our results also show that class 4 females tend to have more F than males in their mantles, but not in their adductor muscles (Figure 12G-H, Table 2, Supplementary Table 6), and that the difference remains the same even when comparing female tissues versus male homoplasmic tissues only. The reason for this difference in F content in the mantle but not in the adductor muscle between sexes remains obscure, and can be putatively ascribed to a sampling effect. It is however interesting to notice that the two heteroplasmic male tissues have significantly less F mtDNA than homoplasmic ones: again, the number of heteroplasmic tissues is low (one mantle and one adductor) and a larger sample size would be needed to confirm this observation, but some kind of conflict during mtDNA replication in these tissues that lowers the F mt genome ratio cannot be excluded *a priori*. Interestingly, in both sexes of class 4, the adductor muscle has higher F ratios than the mantle, suggesting that the former tissue has substantially higher energy needs than the latter: this is expected since the adductor, that is the muscle that opens and closes the valves, is a much more active organ than the mantle.

The disparity in total mtDNA content in bodies between sexes from classes 3 and 4 seems to be due to the rapid increase of M content in males, which starts at very low levels in classes 1 and 2, and becomes dominant over the F in most males of class 4 (Supplementary Figure). Specifically, in males from class 1 to 4, the F ratio shows a slow and rather continuous growth, whereas the M mtDNA increase is sudden, and very similar to that of the F mtDNA in females, maybe even more accelerated (Figure 14). Thus, both F mtDNA in female bodies and M mtDNA in male bodies are subject to a boost in replication between classes 2 and 3 (Figure 14, Table 2, Supplementary Table 6). Moreover, females and males of classes 1+2 and 3+4 are well distinguishable using these two ratios (Figure 13B-C).

What is the cause of this mtDNA boost? It may be due to the cellular proliferation and/or growing energetic demands of the organs located in the bodies during development,

compared to the mantle and adductor muscle, for example. On the other hand, organs in the bodies include the gonads, and the occurrence of a replication boost just before the first reproductive season, together with the observation that in males only the M mtDNA, that is going to be transmitted by sperm, is showing this increase, may hint to another non-mutually excluding explanation. We know that in DUI species the male gonad holds high quantities of M mtDNA (both in *R. philippinarum* and *Mytilus*; Ghiselli *et al.*, 2011; Zouros, 2013), and also that in *R. philippinarum* adults of both sexes, the gonad has the highest mtDNA content (Ghiselli *et al.*, 2011). Knowing that the largest analyzed animals are those closest to the first reproductive season (Devauchelle, 1990), the observed mtDNA boosts in both females and males of classes 3 and 4 are most probably the outcome of the first gonad formation.

## 2.5 CONCLUSIONS

Under DUI, females follow a pattern comparable to that of species with SMI, as they lose the paternally-inherited M mt genome in the first phases of embryogenesis. Males, on the contrary, retain this mtDNA and segregate it in the germ line. The M mtDNA in males becomes overrepresented especially in the gonad, which will produce M-homoplasmic spermatozoa. In this study, we characterized the sex-specific replication trends of F and M mtDNAs in *R. philippinarum*, and pinpointed in this species two events of pivotal importance to study how DUI can occur in place of SMI: the loss of M mtDNA from females but not in males (around 24hpf), and the start of gonad production (in young clams with a shell length >6mm, probably). Also, we hypothesized that the embryogenesis bottleneck in DUI males may be a first step towards the segregation of M mtDNA in the germ line, and discussed how this bottleneck can produce homoplasmic and/or heteroplasmic cells in this sex.

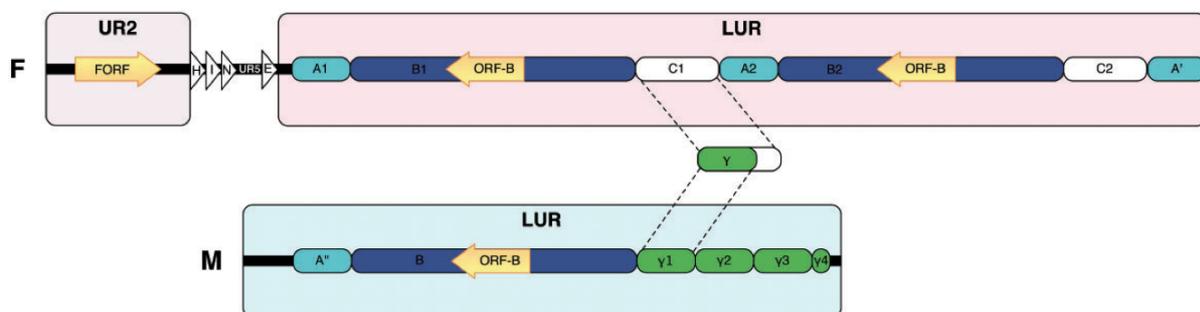
The development of *R. philippinarum* males can be used as a useful model to study the mechanisms controlling mitochondrial heteroplasmy levels and segregation, which are important topics in many fields of theoretical and applied biology. In fact, DUI has long been proposed as a privileged system to study basic biological mechanisms such as genomic conflicts, mitochondrial inheritance, and the mitochondrial bottleneck during germ line formation (Passamonti and Ghiselli, 2009). The understanding of these processes is of crucial importance, not only to comprehend the functioning of the exceptional DUI system in bivalves. The mechanisms that drive sex- and tissue-specific segregation of F and M mtDNA during development could be similar to those involved in the regulation of accidental heteroplasmy (caused by biparental inheritance or other sources) of other species including humans, where this condition is often linked to pathological phenotypes (Lane, 2012; Mishra and Chan, 2014). Since the study of such disorders requires the dedicated production of heteroplasmic lines of model organisms (see for example Sharpley *et al.*, 2012, Burgstaller *et al.*, 2014, and references therein), and bearing in mind that the main difference between DUI and these conditions is that F and M mtDNAs have evolved in parallel in the same species for millions of years, males of *R. philippinarum* can be an ideal and natural system to understand how two very different mt lines interact between each other and with the same nuclear genome during the development of an individual, and how one of the two, the M, whose levels are shown to be extremely low for a long time period, become dominant in some tissues. For this reason, studying the early gonad development in males of DUI species can be an important experimental model to understand how only a certain kind of mitochondria (those carrying M mtDNA in this case) are transmitted by gametes to the next generation.

# PUBLISHED PAPERS

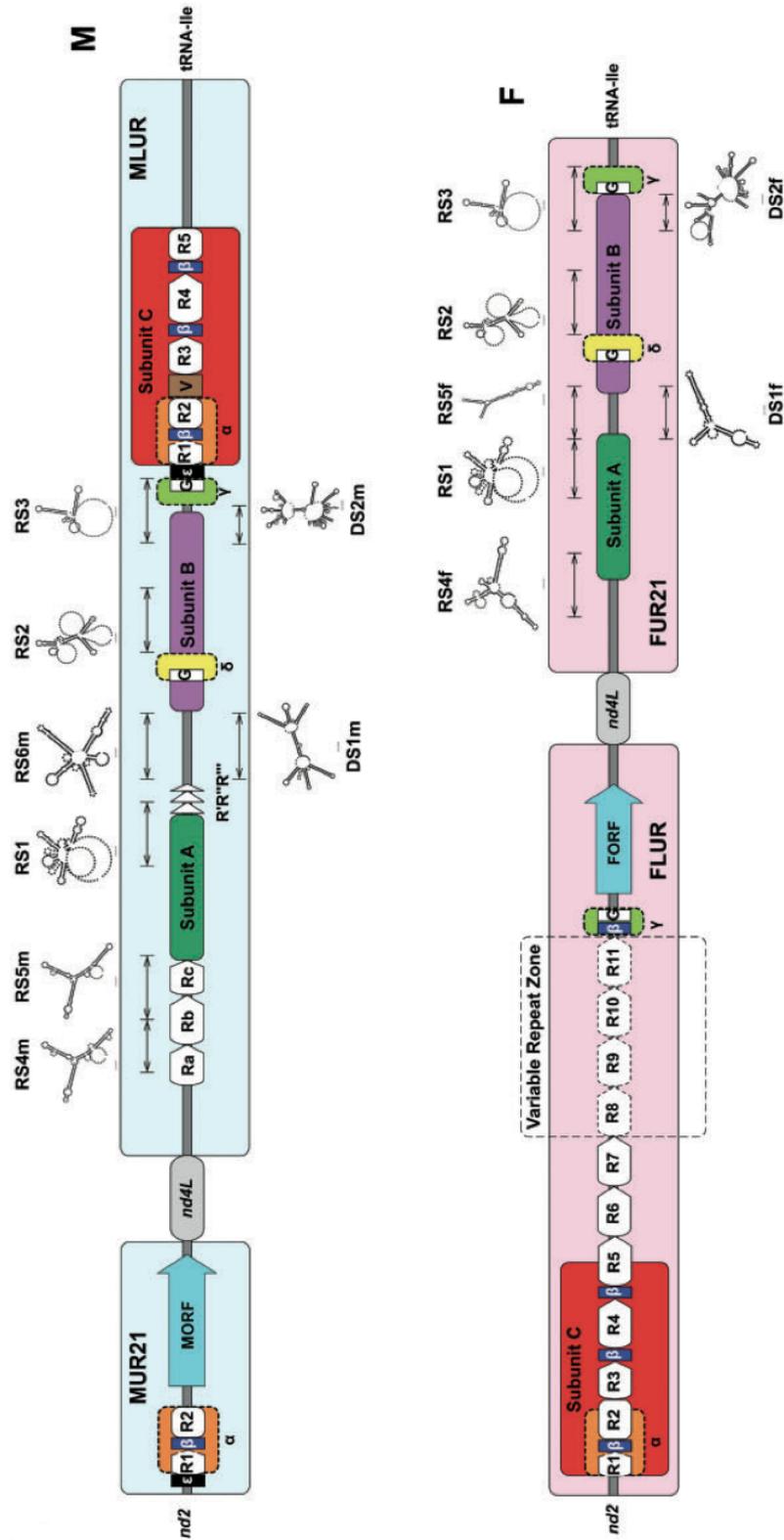
### 3.1 DESCRIPTION OF THE PAPERS

In this chapter I will outline aims and results of the four papers I contributed to during my PhD course. In **Paper 1** we performed a thorough comparative investigation of the ORFans (fORF and mORF) found in the mt genomes of bivalves with DUI, to gain insights into their origin and function. We studied the degree of conservation of these ORFans in different species and investigated the functions of the putative proteins encoded by them. My role was to search for ORFans presence and conservation in all the available mt genome sequences of known DUI species in GenBank, plus *Ruditapes philippinarum* and *Musculista senhousia* mtDNA segments specifically sequenced by me and the research group. **Paper 2** aim was to understand why bivalve mt genomes retain unusually large unassigned regions, and to understand how F and M mtDNA are transcribed in *Ruditapes philippinarum*. We described these regions and analyzed the transcriptome data from female and male gonads, also obtaining data on sequence variability of the two mt genomes. My contribution to this work was to sequence and annotate the largest unassigned regions of *Ruditapes philippinarum* mt genomes, to identify features (such as elements typical of a CR or ORFans) that could justify their retention. For **Paper 3** I contributed as first author in the characterization of the largest URs of *Musculista senhousia* F and M mt genomes. I thoroughly annotated these regions, and suggested that they contain the CR of the respective mt genomes. The CR is duplicated in the F mtDNA: by examining the conservation of the two copies through a Bayesian analysis, I provided evidence for their concerted evolution. **Paper 4** is a review that

summarizes the current knowledge on the unusual features found in many animal mt genomes, and in particular about abnormal gene content. Additional genes or ORFans not related to the OXPHOS system may greatly expand the role of mitochondria in an organism, even in processes such as sex determination or speciation. For this review, I was in charge of the section on gene duplications. The following summaries are intended to give a general overview of the papers, the topic they investigate, the methods used and the results they brought. A brief discussion contextualizing the findings closes every summary. Being Paper 4 a review, no such sections are reported in its summary. Author affiliations are enlisted at the end of this chapter. Please also read the full papers attached in **Appendix 2** to have a more detailed view of the works.



**Figure 16** Schematic structure of *Musculista senhousia* largest URs in its F and M mt genomes. UR2 contains the FORF, while Subunits B in the LURs contain ORF-B on the reverse strand. ORF-B is sometimes disrupted in the F mtDNA. See Figures 8 and 9 for the location of these regions in the respective mtDNA. [Taken from Milani *et al.* (2013)]



**Figure 17** Detailed structure of *Ruditapes philippinarum* F and M mtDNAs largest unassigned regions. MUR21 contains the mORF, while the fORF is comprised in FLUR. RS: RNA secondary structures, DS: DNA secondary structures. See Figures 6 and 7 for the location of these URs in their respective mt genomes. [Modified from Ghiselli *et al.* (2013)]

## 3.2 PAPER 1

### **A comparative analysis of mitochondrial ORFans: new clues on their origin and role in species with doubly uniparental inheritance of mitochondria**

Liliana Milani<sup>1</sup>, Fabrizio Ghiselli<sup>1</sup>, Davide Guerra<sup>1</sup>, Sophie Breton<sup>2</sup>, Marco Passamonti<sup>1</sup>

*Genome Biology and Evolution* 2013, 5 (7): 1408-1434

#### **3.2.1 BACKGROUND**

Additional ORFs are being continuously found in animal mt genomes. Some of them have been shown to derive from duplication and divergence of extant genes, as in the bivalves *Ruditapes philippinarum* (Okazaki and Ueshima, unpublished data; Figure 6 in Chapter 1), *Musculista senhousia* (Passamonti *et al.* 2011; Figure 9 in Chapter 1), and in the genus *Crassostrea* (Wu *et al.* 2012a), or in some hydroidolinan hydrozoans (Kayal and Lavrov 2008, Voigt *et al.* 2008). Additional genes found in Cnidaria have been hypothesized to derive from horizontal transfers (from linear plasmids,  $\epsilon$ -proteobacteria, or viruses; Claverie *et al.* 2009, Bilewitch and Degnan 2011, Ogata *et al.* 2011, Kayal *et al.* 2012).

In DUI species belonging to the orders Mytiloidea, Unionoidea, and Veneroidea, both F and M mtDNAs contain additional lineage-specific ORFans, respectively named fORF and mORF (Breton *et al.* 2009, Breton *et al.* 2011a, Breton *et al.* 2011b, Ghiselli *et al.* 2013): these ORFans are usually located in large URs or in the CR. In the order Mytiloidea, both fORF and mORF of *Mytilus* spp. are situated in the CR first variable domain (named VD1, see Figure 10 in Chapter 1; Cao *et al.* 2004b) of the respective mtDNAs (Breton *et al.* 2011b), and a fORF is present in UR2 of *Musculista senhousia* F mtDNA (Figure 8 in Chapter 1, Figure 16; Passamonti *et al.* 2011, Breton *et al.* 2011b). In *Ruditapes philippinarum*, fORF and mORF are located respectively in the FLUR and in MUR21 of the respective mtDNAs

(upstream *nad4L* in both mt genomes, see Figures 6 and 7 in Chapter 1, Figure 17; Ghiselli *et al.* 2013). Finally, in Unionoids, fORF is located upstream *nad2* in the F mtDNA (except for *Inversidens japonensis* and *Hyriopsis cumingii*, which have a different gene order), while mORF is placed upstream *nad4L* in the M (Breton *et al.* 2009).

These ORFans have been hypothesized to be linked to the maintenance of gonochorism in these species, as well as to be responsible of the peculiar mitochondrial transmission system that is DUI (Breton *et al.* 2009, Breton *et al.* 2011a, Breton *et al.* 2011b). The finding that the two lineage-specific ORFans are translated in the unionid *Venustaconcha ellipsiformis* (Breton *et al.* 2009), and in particular that the fORF protein product is present in mitochondria, nuclear membrane, and egg nucleoplasm of this species (Breton *et al.* 2011a), supports their direct involvement in the DUI mechanism.

For this work we sequenced and analyzed the major mt URs of the DUI Mytilid *Musculista senhousia* to search for novel ORFans, and compared them with those found in other DUI species. A functional analysis on the proteins produced by these ORFs was performed to gain insights into their function, origin, and putative role in the DUI system.

### **3.2.2 ANALYSES AND RESULTS**

We sequenced the LURs of *Musculista senhousia* F and M mtDNA plus FUR2 to confirm the presence of fORF and search for other, not previously annotated, additional ORFs, using DNA extracted from 12 egg samples and 11 sperm samples spawned from different specimens. fORF presence was confirmed in all analyzed sequences and a new ORF on the reverse strand, named ORF-B, was found in both FLUR and MLUR (Figure 16): FLUR is composed of two large tandem repeats (Passamonti *et al.* 2011) and this ORF is found in both of them. ORF-B is always conserved in males, while in some female specimen

its sequence showed mutations and deletions that altered the reading frame, modifying its length or breaking it in two overlapping ORFs.

I gathered all complete and partial CR sequences of the genus *Mytilus* from GenBank to confirm the presence of fORF and mORF in the VD1 described by Breton *et al.* (2011b). 689 sequences belonging to four species of *Mytilus* (*Mytilus edulis*, *Mytilus galloprovincialis*, *Mytilus californianus*, *Mytilus coruscus*) were analyzed: 201 of them contained full-length ORFs (197 fORFs and 17 mORFs; some sequences are recombinant and can contain more than one ORF). All cases where fORF was incomplete were just because the VD1 was not fully sequenced, thus truncating the fORF sequence: the reading frame of all partial and full fORFs was found intact. On the contrary, most of mORF showed a disrupted reading frame. The mORFs annotated by Breton *et al.* (2011b) in *Mytilus edulis*, *Mytilus galloprovincialis*, and *Mytilus trossulus*, comprised a long poly-A string in their reading frame: the different number of adenines in this string, and mutations upstream of it, were the main reasons why many mORFs were found compromised. Nonetheless, the part downstream this homopolymers is conserved in the majority of mORFs.

I then assessed the variability and conservation of all available fORF and mORF sequences from the DUI species *Musculista senhousia* (obtained in this study; also ORF-Bs were considered), the four *Mytilus* species (from GenBank), *Ruditapes philippinarum* (obtained by Ghiselli *et al.* 2013), and *Venustaconcha ellipsiformis* (from GenBank), using p-distance (Table 3). In *Musculista senhousia* fORFs are more variable than male ORF-B, but more conserved than female ORF-B. *Mytilus* fORFs are less variable than mORFs, while in *Ruditapes philippinarum* mORF is more conserved than fORF. All available *Venustaconcha ellipsiformis* fORF sequences are identical among each other. I calculated the p-distances also for the translation of the ORFs (Table 3): all values are higher than those of the respective nucleotide sequences, except for *Ruditapes philippinarum* mORF where the p-D value is zero.

**Table 3** p-Distance (p-D) and Standard Error (SE) values of novel mitochondrial ORFs in DUI bivalves. Number of ORF sequences used for each species is dependant on the number of available and suitable sequences on GenBank. p-D of *Mytilus edulis*, *Mytilus galloprovincialis*, and *Mytilus trossulus* mORFs were calculated only on the last part of the ORF immediately following the poly-A sequence (see Milani *et al.* 2013 in Appendix 2 for details). *N* = number of sequences used. <sup>a</sup> Only complete female ORF-B were considered. <sup>b</sup> Male ORF-B and complete female ORF-B were considered. <sup>c</sup> mORF sequences matching *Mytilus edulis* mORF. [Taken from Milani *et al.* (2013)]

Species	ORF	Nucleotide		Translation		N
		p-D	SE	p-D	SE	
<i>Musculista senhousia</i>	fORF	0.019	0.004	0.035	0.010	11
	Male ORF-B	0.004	0.002	0.008	0.004	12
	Female ORF-B <sup>a</sup>	0.024	0.005	0.056	0.012	8
	Overall ORF-B <sup>b</sup>	0.030	0.006	0.063	0.014	20
<i>Mytilus californianus</i>	fORF	0.005	0.003	0.014	0.008	4
	mORF1	0.015	0.009	0.031	0.021	4
	mORF2	0.011	0.007	0.033	0.022	4
<i>Mytilus edulis</i>	fORF	0.013	0.002	0.026	0.006	134
	mORF	0.017	0.004	0.039	0.012	25
<i>Mytilus galloprovincialis</i>	fORF	0.024	0.004	0.048	0.009	16
	mORF	0.029	0.008	0.062	0.021	47
	mORF (edulis-like) <sup>c</sup>	0.023	0.007	0.042	0.017	14
<i>Mytilus trossulus</i>	fORF	0.007	0.002	0.014	0.005	8
	mORF	0.025	0.007	0.046	0.016	9
<i>Ruditapes philippinarum</i>	fORF	0.009	0.003	0.011	0.006	8
	mORF	0.004	0.002	0.000	0.000	7
<i>Venustaconcha ellipsiformis</i>	fORF	0.000	0.000	0.000	0.000	3

The translations of the lineage-specific ORFans of DUI species (henceforth named FORF and MORF), *Musculista senhousia* ORF-B, an ORF found in *Paphia euglypta* (Bivalvia Veneridae) CR, and the additional ORFs in three cnidarian species, were analyzed to infer their structure and used for a function prediction analysis. A SP was found at the N-terminus of all FORFs, always coincident with a transmembrane helix; the same situation was found in MORFs, even if the SP support was lower. The function prediction for FORFs gave many hits with proteins involved in nucleic acid binding, transcription, RNA modification or methylation, membrane association, and immune response. For MORFs, the hits were related to proteins having a role in membrane association with nucleic acid binding and transcription, DNA recombination, transcription, integration of foreign elements, cytoskeleton dynamics, ubiquitination, apoptosis, and immune response. For all DUI bivalves lineage-specific

ORFans, the first four hits found by one software were always the same: two hits are involved in cell membrane/surface anchoring, and the other two to transcription and post-transcriptional processes. Finally, all analyzed ORFs showed hits with viral proteins, except for the MORFs of *Mytilus trossulus* and *Venustaconcha ellipsiformis*. See Figure 18 for location of the hits on the protein sequences.

### 3.2.3 DISCUSSION

The conservation in DUI species of most of all analyzed additional ORFans points to a functional meaning of these sequences. Exceptions are *Musculista senhousia* ORF-B, and *Mytilus edulis*, *Mytilus galloprovincialis*, and *Mytilus trossulus* mORFs. ORF-B is not lineage-specific and its sequence is compromised in many F mtDNA samples: this may indicate that this ORF is not functional in this mt genome. The mORF of the three *Mytilus* species mentioned above contain, in their “full length” version, a long string of adenines (Breton *et al.* 2011b), and since most reading frames are compromised in this segment or in the part upstream of it, this may reflect a difficulty in sequencing this long homopolymer. All other lineage-specific ORFans in DUI species are conserved, with variable patterns among species.

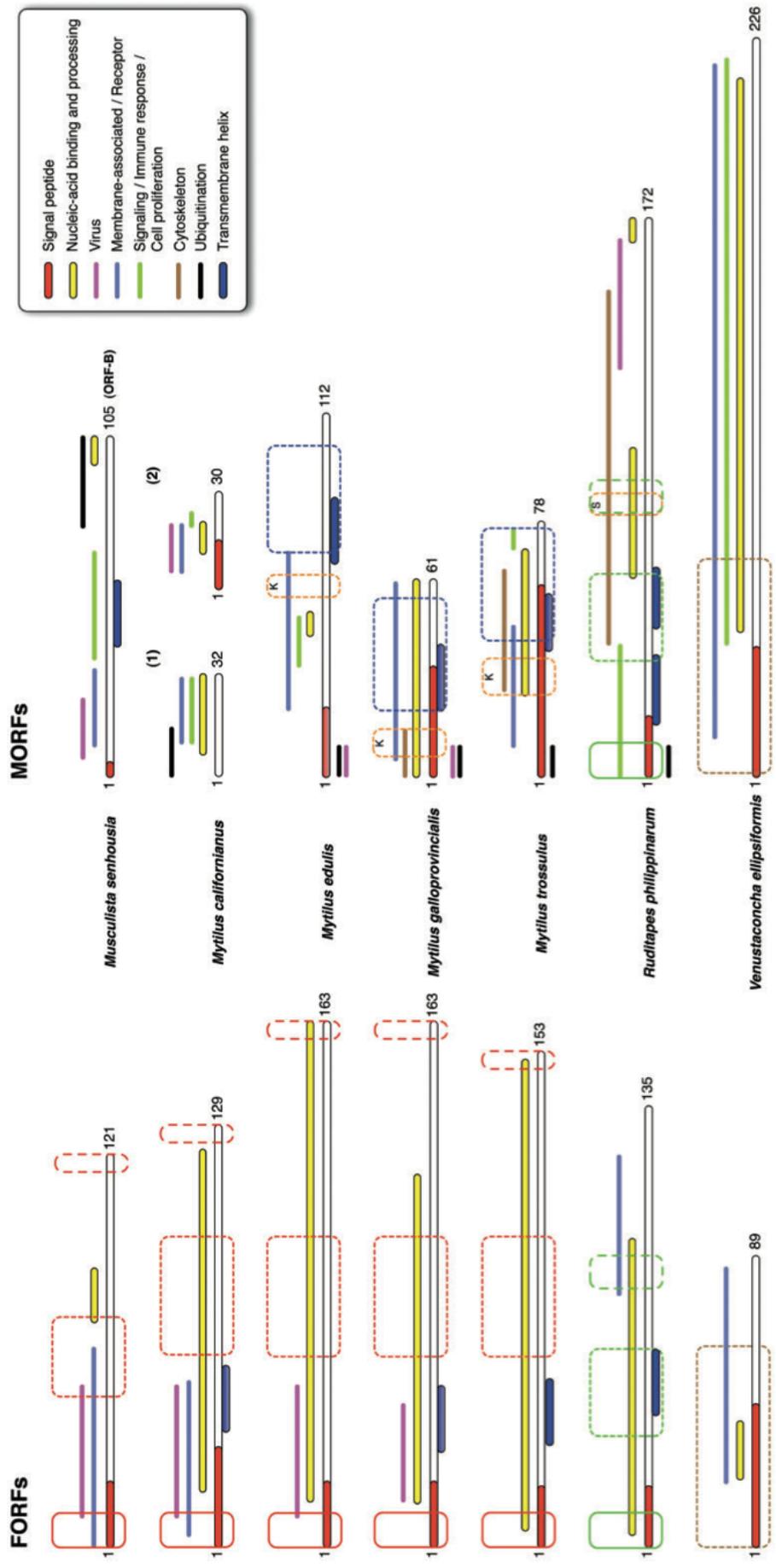
The structure of the putative proteins translated from the ORFans suggests an anchoring to membranes, in particular for MORFs of *Mytilus* and *Ruditapes philippinarum*. All proteins share similar functions, but when only considering most supported hits, FORFs are more similar among each other than with MORFs, and vice versa. Both FORFs and MORFs have hits indicating a role as signaling molecules, but most functions are typical of only one ORFan product type (FORFs: transcription regulation and immune response, cell adhesion, migration, and proliferation; MORFs: cytoskeleton organization, cell differentiation during embryonic development, nucleic acid binding and transcription regulation). Some MORFs

have similarities with DNA replication, recombination, and repair proteins, and almost all ORFan proteins show hits of ubiquitination and apoptosis regulation proteins.

Given the results, it is unlikely that the lineage-specific ORFans analyzed originated from duplication of standard mtDNA-encoded genes. The high amino acid substitution rate observed (compared to typical mt genes) could be the signature of lineage-specific adaptation (Daubin and Ochman 2004a, Daubin and Ochman 2004b, Cai and Petrov 2010, Yu and Stoltzfus 2012), and since many hits refer to viral proteins, this may indicate a viral origin of these ORFans. These viral sequences might have been endogenized by the host organism and co-opted for host cell functions (Feschotte and Gilbert 2012), and clues coming from the structural and functional analyses open many possibilities. For example, these foreign elements could originally have a role in immune response: since mitochondria have central roles in regulating immune response, the viruses from which these ORFans could come from might have targeted mitochondria to evade this process (Ohta and Nishiyama 2011); also, some structural features also suggest involvements in apoptosis control. Viral envelope proteins have been shown to cause aggregation of mitochondria (Doorbar *et al.* 1991, Galluzzi *et al.* 2008), and the many MORFs hits concerning cytoskeleton can be related to a role in the differential distribution of sperm mitochondria during male and female development. Retrograde signaling from mitochondria to nucleus has been demonstrated in plants showing CMS (Abad *et al.* 1995, Fujii and Toriyama 2008, Nizampatnam *et al.* 2009), a system of sex determination involving additional mtDNA-encoded proteins that may bind to mitochondrial membranes (Nizampatnam *et al.* 2009). The novel proteins identified in DUI bivalves can putatively tag the outer membrane of mitochondria: in particular, MORFs might mask sperm mitochondria from the degradation machinery in developing male embryos. Finally, these proteins can be located outside mitochondria (Breton *et al.* 2011a), contributing to a communication system between mitochondria and nucleus.

The co-option of the novel ORFans in bivalve mtDNAs may have influenced some aspects of their life cycle involving mitochondria (Forterre 2006, Koonin 2006). In particular, selfish viral sequences might have found a way to be transmitted through the generations, leading to DUI. DUI presence is scattered in bivalves and the findings of our work may support a multiple origin of this mechanism. On the other hand, the similarities among ORFans may suggest an origin from elements of the same kind, but their conservation only among related taxa may indicate either independent origins or that their fast evolution has cancelled all sequence similarities.

**Figure 18 (next page)** Functional domains in FORFs and MORFs of DUI species mtDNAs (position in the amino acid sequence as identified by HHpred; Söding *et al.* 2005). Sequences with similarities are boxed in the same color and with the same type of line; red: similarities among FORFs; blue: similarities among MORFs; orange: K = poly-K region, S = poly-S region. Numbers indicate aa length. [Taken from Milani *et al.* (2013)]



### 3.3 PAPER 2

#### **Structure, transcription, and variability of metazoan mitochondrial genome: perspectives from an unusual mitochondrial inheritance system**

Fabrizio Ghiselli<sup>1</sup>, Liliana Milani<sup>1</sup>, Davide Guerra<sup>1</sup>, Peter L. Chang<sup>3</sup>, Sophie Breton<sup>2</sup>, Sergey V. Nuzhdin<sup>3</sup>, Marco Passamonti<sup>1</sup>

*Genome Biology and Evolution* 2013, 5 (8): 1535-1554

#### **3.3.1 BACKGROUND**

The protomitochondrion (Müller and Martin 1999, Atteia *et al.* 2009, Abhishek *et al.* 2011, Thrash *et al.* 2011) genome underwent a massive process of GRE since the ancestral symbiosis event (Andersson and Kurland 1998, Khachane *et al.* 2007), which transformed it into the mt genomes we know today in the various eukaryote lineages, shaped by both neutral and adaptive modifications (Embley and Martin 2006). Selective pressure for GRE (thus for UR deletion) is stronger in genomes with a high mutation rate, as non-functional intergenic DNA can accumulate gain-of-function harmful mutations (Lynch *et al.* 2006, Lynch *et al.* 2011, Lynch 2007), but its efficiency is dependent on the amount of random genetic drift and effective population size.

mtDNA mutation rate is variable among animal lineages, and it has been linked to body mass, metabolic rate, ROS production, and lifespan (Galtier *et al.* 2009). Most of the variability however seems due to errors made by DNA polymerase during replication (Drake *et al.* 1998, Lynch *et al.* 2006): it follows that large part of the heritable mutations are accumulated during germ line proliferation, where cells undergo a great number of divisions. Thus, reproduction modes and gonad physiology may affect mtDNA rates of evolution (Rand 2001, Davison 2006). In bivalves, the high number of cell divisions in both female and male

gonads (which are produced *de novo* every reproductive season; Gosling 2003), and the large number of gametes produced, lead to a high mutation rate: indeed, in both nuclear and mt genomes, bivalves show a large amount of polymorphism (Saavedra and Bachere 2006). In addition, their mt genomes are very long, show an extremely variable gene order, and can be transmitted both maternally and paternally under DUI.

The main aims of this work were to identify the reasons why large URs can be conserved in bivalve mt genomes, and test the transcription of F and M mtDNA in a DUI species, also analyzing the amount of polymorphism of the two mtDNAs. For this, we sequenced the largest URs of *Ruditapes philippinarum* F and M mtDNAs to locate the CR and identify other features that could explain the presence of such large URs. Then, we characterized the mitochondrial transcriptome of female and male gonads of this species, and performed a SNP analysis.

**Table 4** Proportion of URs in the mitochondrial genomes of Metazoans. *N*: sample number, median total length: median total length of the mt genomes in a taxon, median URs length: median total length of mtDNA URs in a taxa, median %cod: median proportion of coding regions in the mt genomes, median %URs: median proportion of URs in the genomes. Significance: Wilcoxon rank-sum test significance (\*\*\*:  $P < 0.001$ , n.s.: nonsignificant). [Taken from Ghiselli *et al.* (2013)]

Taxa	<i>N</i>	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	***

### 3.3.2 ANALYSES AND RESULTS

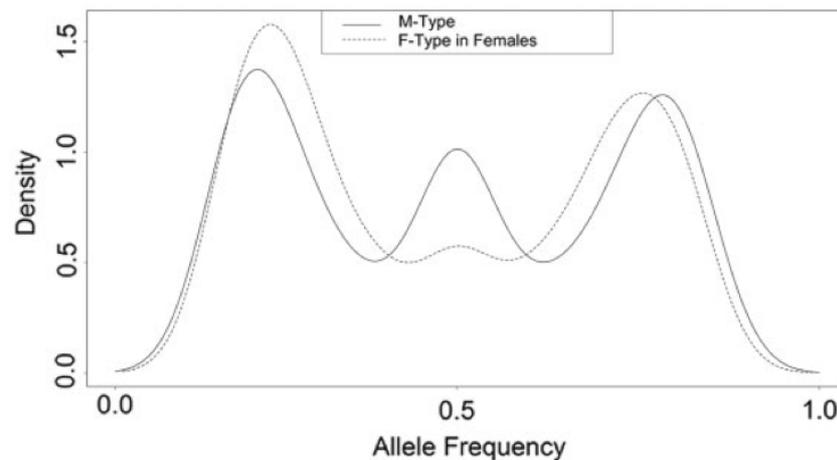
2,656 complete animal mt genomes were downloaded from the MitoZoa database to analyze the length and proportion of UR in various taxa. Bivalves are the group showing the longest mtDNAs and the highest proportion of UR (Table 4).

I used the DNA extracted from sperms and eggs spawned by 7 males and 8 females, respectively, of *Ruditapes philippinarum* as a template to amplify the mtDNA regions containing LUR and UR21 of the M and F mt genomes (the largest URs of the two mtDNAs; see Figures 6 and 7 in Chapter 1). The obtained URs were analyzed in detail to search for conserved domains and secondary structures (Figure 17). Three main subunits, named A, B, and C, were found conserved among the URs, though their position is different between F and M mtDNA. A region containing a variable number of tandem repeats was found in the FLUR. Some of the identified DNA secondary structures share similar loops in both F and M mtDNAs, and three RNA structures are conserved between the two mt genomes. In both mt genomes, in the same relative position (upstream the *nad4L* gene), an additional ORF was found conserved (fORF in the F and mORF in the M mtDNA): these ORFs have been analyzed in detail in Milani *et al.* (2013).

To locate  $O_H$  and  $O_L$  in *Ruditapes philippinarum* mtDNAs, I performed an AT-skew analysis on its complete F and M mt genomes available in GenBank, and of other Veneridae species for additional support. In general,  $O_H$  seems to be located in coincidence of the largest URs in all mt genomes, while the  $O_L$  is often associated to a conserved cluster of three tRNA genes (*tRNA-His*, *tRNA-Glu*, and *tRNA-Ser*). I also searched the large URs of all these species for motifs typical of a CR by comparing them with the CR of the sea urchin *Strongylocentrotus purpuratus*, in which mtDNA replication is well characterized: two motifs (named  $\delta$  and  $\gamma$ ; Figure 17) in *Ruditapes philippinarum* FUR21 and MLUR showed similarities with the motifs involved in the start of replication in sea urchin mtDNA.

To perform a transcriptome analysis of the mt genomes from 6 male and 6 female gonads of *Ruditapes philippinarum*, a cDNA library was sequenced with an Illumina GAIIx platform. The majority of transcripts (90.11%) in male gonads are from M mtDNA. F mt genome transcription profiles are the same in males and females, and are different from that

of the M. mORF transcription levels are comparable to those of the other M genes, while fORF levels are significantly lower than all other F genes. A SNP analysis was performed on the mt transcriptomes to assess the variability of F and M mt genomes. F and M have similar amounts of high-frequency alleles, but the F mt genome, in both sexes, has an excess of rare alleles, and less mid-frequency ones, compared to the M (Figure 19).



**Figure 19** Kernel density plot of allele frequencies in mitochondrial CDSs of *Ruditapes philippinarum* mt genomes. Solid line: M-type mtDNA in male gonads; dashed line: F-type mtDNA in female gonads. 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. [Taken from Ghiselli *et al.* (2013)]

### 3.3.3 DISCUSSION

Bivalve mtDNAs have high proportions of non-coding and intergenic sequences (Table 4), and given their high polymorphism in sequence and gene arrangement, they seem to evolve in contradiction with the GRE theory: one plausible explanation may be that these non-coding sequences harbor important functional elements, and thus their retention is favored and/or necessary. For example, from the results of our analyses in *Ruditapes philippinarum*, it is plausible that the non-coding sequences MLUR and FUR21 contain the CR of the respective mt genomes, since they contain structures and motifs (conserved between F and M mtDNAs, and among different species) putatively related to the start of

mtDNA replication. Moreover, the other two large URs of *Ruditapes philippinarum* mt genomes, MUR21 and FLUR, contain an additional ORFan whose sequence is conserved in all samples (Figures 6 and 7 in Chapter 1, Figure 17; these ORFans have been analyzed in detail in Milani *et al.* 2013). The presence of such ORFans in an organelle genome might hint to a lineage-specific function of these elements, maybe in the DUI mechanism of *Ruditapes philippinarum*, and their transfer to the nucleus would severely affect their function.

From the gonad transcriptomes, we obtained a transcription profile for M mtDNA in males, and for F in both males and females. One interesting finding is that the F mt genome transcription profile is the same in both sexes. However, the transcription levels of F mtDNA-encoded ETC genes are lower in males than in females, but those of nuclear-encoded ETC genes are the same in both sexes: this points to a mitochondrion-driven regulation of mtDNA genes expression, as stated by the CoRR hypothesis (Allen 2003). To explain the different transcription profile of M compared to the F, two scenarios can be considered. The M mtDNA might be a somewhat “nearly-selfish” element (Zouros 2013) that could be less coordinated with the nuclear-encoded subunits of ETC complexes. On the other hand, according to Allen (1996), in species with SMI, the egg mtDNA that is going to be inherited by the next generation is transcriptionally inactive to prevent ROS damage (supported by the findings of de Paula *et al.* 2013), while sperm mtDNA, which is not inherited, is active to provide energy for spermatozoa movement, and thus subject to ROS damage. In DUI species, mtDNA from both egg and sperm is inherited: it is then clear that M mtDNA does not follow the rules of Allen’s theory, and also our findings on F mtDNA do not support the transcriptional quiescence of eggs mitochondria. More studies will be required to test the theory of ageing in DUI species, but we hypothesized that sperm mitochondria may avoid damage to M mtDNA using the malate dismutase pathway, an ATP-producing metabolic process that reduces ROS production, as already observed in the DUI species *Mytilus edulis* (Müller *et al.* 2012).

The SNP analysis showed that F mtDNA (in both male and female gonads) has more polymorphic sites than M, and that the two mt genomes have different amounts of low, mid, and high frequency alleles. The difference in low frequency alleles ( $F > M$ ) can be explained by the different size of the bottleneck to which the two mt genomes are subject during gametogenesis: eggs possess more mtDNA than sperms, so the F mtDNA experiences a wider bottleneck than the M, thus allowing more rare variants to be transmitted. The different bottleneck size can also explain the maintenance of intermediate alleles in M mtDNA, but not why M has more alleles of this type compared to F. A deviation from the negative selection typical of mt genomes could explicate this finding, as already hypothesized to explain similar observations in plants showing CMS, a sex determination system involving novel mt ORFs that has shown many parallels with DUI. Analyzing the kind of SNPs in the two mt genomes, we observed that F has in general a higher amount of deleterious non-synonymous variants compared to M. The presence of such variants is explained by the different size of the bottlenecks between F and M, as above, and the maintenance of harmful mutations in the F by a stronger buffering effect. Eggs contain more mtDNA copies than spermatozoa, so the effect of deleterious mutations is masked by the presence of functional alleles, which also partially hides them from selection. In spermatozoa, on the contrary, selection against these deleterious mutations is stronger given the much lower mtDNA copy number and consequent weaker buffering, thus we observe a lower number of them.

The origin of intergenic sequences in bivalve mtDNAs can be ascribed to random processes (such as slipped-strand mispairing, errors in termination of replication and recombination; Boore 2000, Ladoukakis *et al.* 2011), which in this taxon appear to be particularly active, coupled to an elevated mutation rate and a probable low efficiency of DNA mismatch repair. Their maintenance, on the other hand, can be due to adaptive reason, as these URs may contain ORFs and/or signals necessary to the mtDNA physiology. Also,

these features might have influenced the inheritance path or the transcriptional behavior of the two mt genomes in species with DUI.

From the results of many previous PCR-based studies, M mtDNA was usually thought to evolve faster than F (Zouros 2013), but our data show a different scenario where the two mt genomes are close in terms of polymorphism amount. It has been proposed that a relaxed selection could be the reason of M fast evolution (Zouros 2013), but its role in production of gametes and sperm swimming capacity is still an important one, and a reduced selection would affect the entire population (Gemmell and Allendorf 2001, Meiklejohn *et al.* 2007): indeed, we observed more synonymous than non-synonymous mutations in M mtDNA. Hypotheses to explain M mtDNA evolution can be that it has functions in spermatogenesis and/or spermatozoa, since genes related to sex and reproduction have been shown to evolve rapidly (Ellegren and Parsch 2007, Parsch and Ellegren 2013), or maybe sperm competition (Palumbi 2009). M mtDNA of DUI species can be under selection for male functions, and this can increase the fitness of a population, and DUI itself, once established, would be an advantageous character.

### 3.4 PAPER 3

#### **The largest unassigned regions of the male- and female-transmitted mitochondrial DNAs in *Musculista senhousia* (Bivalvia Mytilidae)**

Davide Guerra<sup>1</sup>, Fabrizio Ghiselli<sup>1</sup>, Marco Passamonti<sup>1</sup>

*Gene* 2014, 536: 316-325

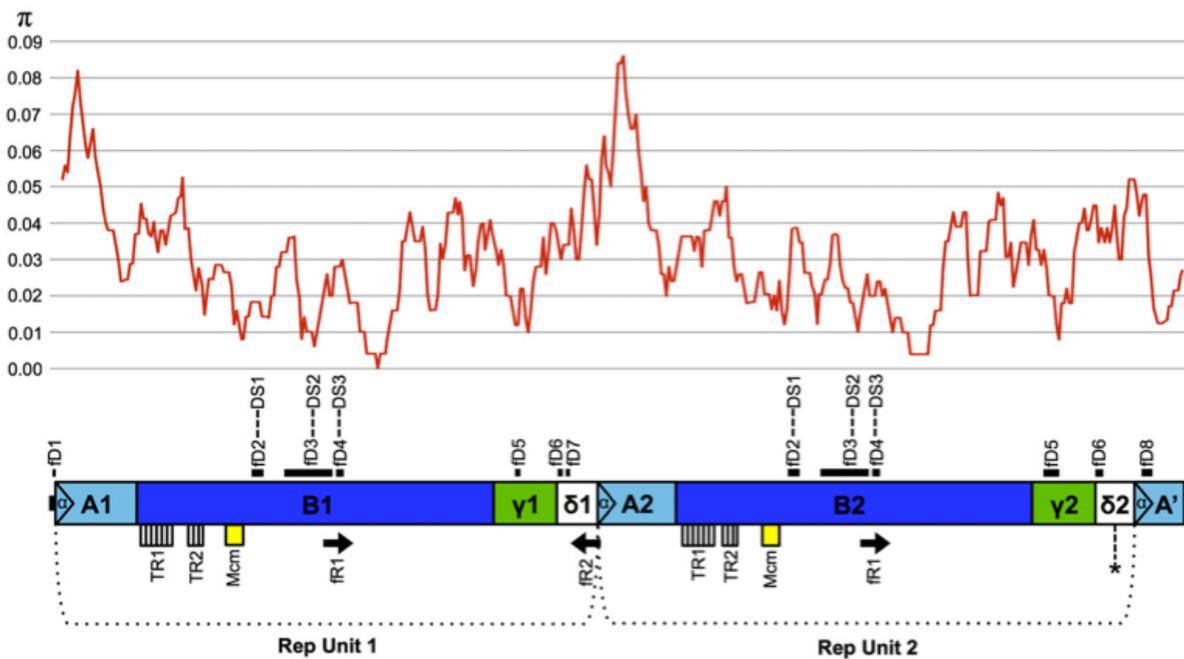
##### **3.4.1 BACKGROUND**

Mitochondrial genomes usually comprise a CR that contains features such as secondary structures and sequence motifs involved in transcription and replication of the mtDNA molecule (Scheffler 2008). The mt genome of many invertebrate and vertebrate species possesses two CRs that maintain a high sequence similarity (up to 100%), which points to a full functionality of both copies (see Schirtzinger *et al.* 2012 and references therein for examples). The concerted evolution of the two CRs is thought to be driven by gene conversion through recombination between homologous sequences (Kumazawa *et al.* 1998); also, the presence of two functional origins of replication was hypothesized to increase the replication and mutation rate of a mt genome (Kumazawa *et al.* 1996, Kumazawa *et al.* 1998).

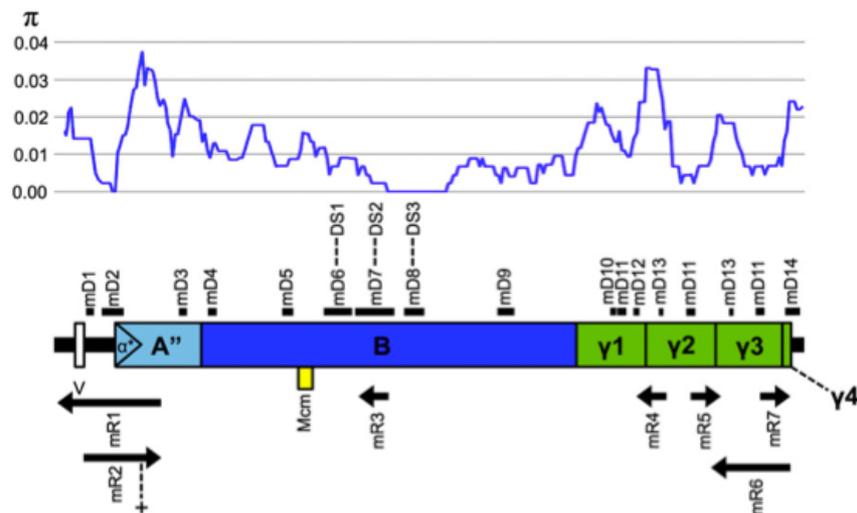
In the highly divergent F and M mtDNA of DUI species, the conservation of sequence blocks in their CRs is explained by the presence of signals necessary to replication and transcription, which could be under a more strict selective constraint than the rest of the CR or other parts of the mt genomes (Cao *et al.* 2004b). The CR parts that most differ between F and M are, on the contrary, thought to be involved in the separate transmission route of the two mt genomes (Breton *et al.* 2009).

The LURs of *Musculista senhousia* F and M mtDNAs share three main conserved blocks (named A-type Subunits, Subunits B, and Subunits  $\gamma$ ; Passamonti *et al.* 2011) (Figures

## FLUR



## MLUR



**Figure 20** *Musculista senhousia* mt LURs organization, content and variability. FLUR and MLUR are represented in scale. Graphics over the LUR schemes represent the nucleotide diversity ( $\pi$ ) levels, among FLURs and among MLURs, calculated with sliding windows on alignments of complete sequences. DNA secondary structures (fD and mD) position is indicated with black lines over the LURs. RNA secondary structures (fR and mR) location is shown below the LURs with black arrows; the orientation of the arrows specifies the structures direction. Mcm: conserved sequence motif found in Mytilids LURs. TR1 and TR2: FLUR tandem repeat series 1 and 2. V: MLUR 5' small variable region. \*: location of a 130 bp insertion. +: location of MLUR cruciform structure inside mR2. [Taken from Guerra *et al.* (2014)]

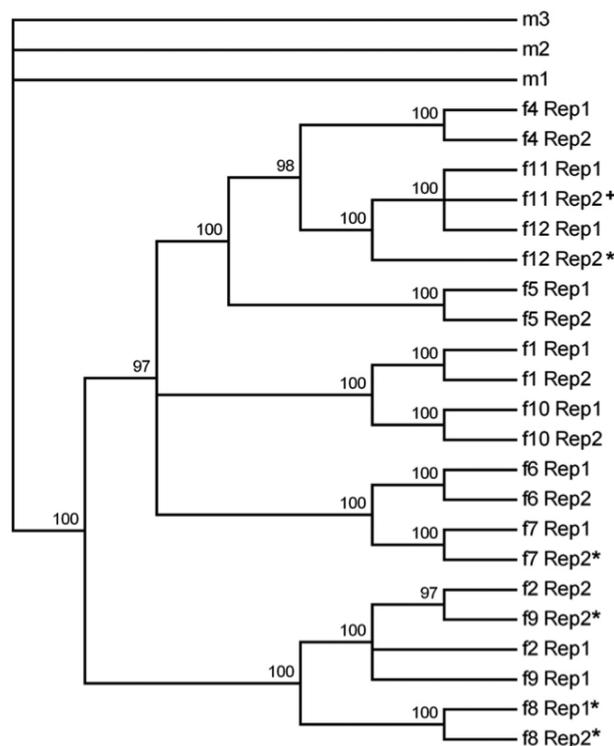
16 and 20). The FLUR is composed of two large repetitive units in tandem (named Rep Units 1 and 2), whose structure is comparable to that of the MLUR. In this work I analyzed these LURs to understand the conservation pattern of the subunits, to search for features typical of a CR by comparing them with other available Mytilidae species LURs, analyzed the AT-skew patterns of the published mt genomes of these species to locate  $O_H$  and  $O_L$ , and analyzed the variability of FLUR Rep Units to comprehend their evolution pattern. The LUR sequences used were those obtained by Milani *et al.* (2013).

### 3.4.2 ANALYSES AND RESULTS

I obtained a variability profile of the LURs using sliding windows (Figure 20), and calculated the variability of single subunits with p-distance. Among all subunits, B Subunits were the less variable, particularly in their central part; on the whole, MLUR sequences showed less polymorphisms than those of FLUR. I then searched for repeats and secondary structures in all FLUR and MLUR sequences (Figure 20): two series of tandem repeats were found private to FLUR, as well as a variable region 5' of MLUR; a larger number of DNA and RNA secondary structures were found in MLUR compared to FLUR, three of them shared between the two. A sequence motif, called M<sub>cm</sub>, was found conserved among *Musculista senhousia* LURs (Figure 20) and the CR of other five Mytilidae species (four *Mytilus* species plus *Perna viridis*). Most of the features conserved between FLUR and MLUR, and among Mytilidae, reside in the region with the lowest variability of B Subunits. Moreover, the overall structure and variability pattern of MLUR, and of FLUR Rep Units (i.e. a more conserved region flanked by two more variable ones), closely resembles that of *Mytilus* CRs (Cao *et al.* 2004b) (Figure 10 in Chapter 1, Figures 16 and 20); also, the shape and position of some secondary structures in *Musculista senhousia* is comparable to those annotated in *Mytilus* CRs. The AT-skew analysis on all protein coding genes of 11 complete

mt genomes (F and M from *Musculista senhousia* and four *Mytilus* species, plus *Perna viridis*) indicated that O<sub>H</sub> may be located in correspondence of the LUR.

The topology of the Bayesian tree built using all available FLUR Rep Units (Figure 21) showed a specimen-specific clustering of the sequences with high support, which means that Rep Units from a single female are more similar to each other than to those belonging to other specimens.



**Figure 21** Bayesian tree of *Musculista senhousia* FLUR Rep Units. Rep1 and Rep2: FLUR Rep Unit 1 and 2, respectively. m1, m2 and m3: MLUR sequences used as outgroup. \*: sequences with strings of ambiguous nucleotides. +: sequence with a 130 bp insertion. [Taken from Guerra *et al.* (2014)]

### 3.4.3 DISCUSSION

The presence in *Musculista senhousia* mt LURs of many features (Figure 20) that can function as binding sites for enzymes involved in the start of replication and/or transcription of the mtDNA (secondary structures and motifs) (Scheffler 2008) shared between F and M

type, and among different Mytilid species, and the indication that the O<sub>H</sub> could be contained in these URs, suggest that these regions most probably contain the CR.

Despite the phylogenetic distance and the differences in mt genome organization between *Musculista senhousia* and the genus *Mytilus* (they belong to two different subfamilies, Crenellinae and Mytilinae, respectively), some similarity in the organization of their CR was found. In particular, the overall organization of *Musculista senhousia* MLUR and FLUR Rep Units, and the position of some secondary structure, is comparable to that of *Mytilus*. This may indicate that in DUI mytilids the CR might have the same general structure (a conserved domain flanked by two variable ones), or at least that the central part of the CR contains all signals for the start of replication and transcription and thus is less prone to change compared to its 5' and 3' ends when the two mt genomes start to diverge.

The FLUR is composed of two Rep Units, more similar between each other inside the same individual than among different specimens (Figure 21). Given this high similarity, and that the M CR corresponds to the MLUR, whose organization is comparable to those of the Rep Units, it is conceivable that the FLUR is composed of two CRs in tandem, which evolve in concert. This is the first supported report of this kind for a bivalve species.

The higher amount of polymorphism of the FLUR compared to the MLUR mirrors the finding in protein coding genes of *Musculista senhousia* by Passamonti (2007), who analyzed individuals from the same Adriatic Sea population. In the preceding study, this observation was explained by hypothesizing a female-skewed sex ratio of the founder population, which carried more F than M variants, and can be applied also to the findings enlisted in this work. However, it has also been observed that mtDNAs with two functional CRs accumulate more mutations compared to single-CR mt genomes, maybe by means of higher replication and mutation rates (Kumazawa *et al.* 1998). The M mtDNA of DUI species was usually proposed as having a higher replication rate compared to the F, to account for its capacity to invade the

male germ line (Cogswell *et al.* 2006, Ghiselli *et al.* 2011). If really the F mtDNA replication rate is higher than that of the M in *Musculista senhousia*, this does not affect its DUI mechanism, and may indicate that a higher replication speed is not a necessary condition for the M mt genome to colonize the germ line.

### 3.5 PAPER 4

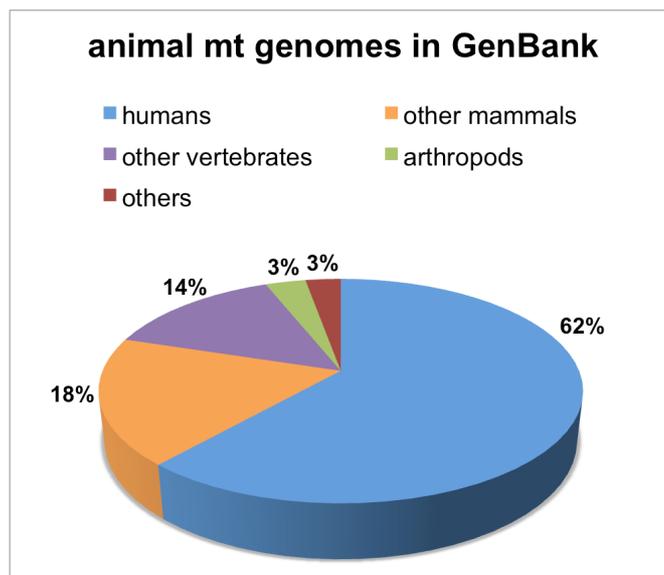
#### **A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs**

Sophie Breton<sup>2</sup>, Liliana Milani<sup>1</sup>, Fabrizio Ghiselli<sup>1</sup>, Daide Guerra<sup>1</sup>, Donald T. Stewart<sup>4</sup>, Marco Passamonti<sup>1</sup>

*Trends in Genetics* 2014, 30 (12): 555-564

A mt genome typically possesses 13 protein coding genes encoding for subunits of four ETC complexes, plus 22 tRNA and 2 rRNA genes necessary for their translation (the mt genetic code is different from the nuclear, and it is variable among taxa); moreover, a region containing signals to start replication and transcription of the molecule, the CR (Scheffler 2008), is present and its length can be highly variable. The classical image of animal mt genomes as small, circular, compact molecules with few and short intergenic spacers, if any, is most probably given by the well known architecture of *Homo sapiens* mtDNA, and to the high number of vertebrate mtDNAs among all those sequenced up to day, whose organization is generally conserved. Indeed, of all animal mtDNAs present in GenBank (~32,800 sequences in July 2014), more than 94% belong to vertebrates (of which 85% are mammalian, and 77% human), 3.2% to arthropods (the most diverse and species-rich animal taxon), and the last ~2% comprise all remaining taxa (Figure 22). It is then clear that the largest part of animal mt genome diversity is awaiting discovery.

Regarding the content of animal mtDNAs, the set of 13 protein-coding genes may be far from being a stable feature. In some cases, for example, a gene may be missing because of transfer to the nucleus (*atp6* in ctenophores; Pett *et al.* 2011, Kohn *et al.* 2012), or entirely lost: in this latter case, however, annotation issues cannot be excluded (as for *atp8* and *nad6*



**Figure 22** Proportions of animal mt genomes in GenBank (July 2014).

in various taxa; Gissi *et al.* 2008, Wang and Lavrov 2008, Pett *et al.* 2011, Bernt *et al.* 2013). Additional genes can arise from duplication of extant ones, and the random pseudogenization of the surplus copies can lead to new gene orders and arrangements (Boore 1999): indeed, pseudogenes have been observed in many taxa (Raboin *et al.* 2010, Beckenbach 2011, Schirtzinger *et al.* 2012, Wei *et al.* 2012, Kawashima *et al.* 2013). Duplicated, but not pseudogenized copies have been recognized in molluscs (Kawashima *et al.* 2013, Stöger and Schrödl 2013) and other animal groups (Fujita *et al.* 2007, Kayal *et al.* 2012, Kurabayashi and Sumida 2013). In some cases, in bivalves, such duplicated genes have mutated into coding sequences of unknown function: *f.i.* an additional and elongated copy of *cox2* is present in two species with DUI (*Ruditapes philippinarum* F mtDNA and *Musculista senhousia* M mtDNA; Okazaki and Ueshima unpublished data, Passamonti *et al.* 2011), but its functionality as *cox2* is still under study. In three oyster species the duplication and divergence of a gene may have originated novel functional genes (*nad2* in two *Crassostrea* species and *nad5* in *Pinctada maxima*; Wu *et al.* 2012a, Wu *et al.* 2012b). Animal mtDNAs may also contain genes whose function is not related to OXPHOS. “Atypical” mt genes that

fall into this category are: an endonuclease contained into group I introns inside *cox1* of many cnidarians, sponges and placozoa (Signorovitch *et al.* 2007, Gissi *et al.* 2008, Szitenberg *et al.* 2010); the cnidarian *mutS* (Pont-Kingdon *et al.* 1995) and *dnaB* (Shao *et al.* 2006); *atp9* and *tatC* in sponges (Lavrov *et al.* 2005, Lavrov *et al.* 2013, Pett and Lavrov 2013); and the *humanin* gene found inside *Homo sapiens* 16S rRNA, which appears to have many diverse putative functions (Lee *et al.* 2013, Cohen 2014). Finally, proteins produced by “typical” mtDNA-encoded genes may have functions not restricted to the OXPHOS system, such as NAD2 and COX2 in humans (Maximov *et al.* 2002, Gingrich *et al.* 2004), or the F and M mtDNA-encoded COX2 in freshwater mussels with DUI (Chakrabarti *et al.* 2006, Chakrabarti *et al.* 2007, Chapman *et al.* 2008, Chakrabarti *et al.* 2009).

Mitochondrial ORFans are an interesting category, since their origin is often unknown and the peptides they encode carry no resemblance to any known protein, and therefore their functions are largely undetermined. However, the conservation of such ORFans in the mt genomes of different species (and sometimes very distant taxa) may indicate a selective pressure for their maintenance. An extreme example of a widely conserved ORFan is *gau*, an ORFan found on the complementary strand of the *cox1* gene of eukaryote mtDNAs and  $\alpha$ -proteobacteria genomes, and evidence in human cells points to its functionality inside mitochondria (Faure *et al.* 2011). An additional ORF containing direct variant repeats (Park *et al.* 2011) and an ORFan named *ORF314* (Kayal *et al.* 2012) are found in many cnidarian species. The lineage-specific ORFans present in the mt genomes of bivalves with DUI (families Mytilidae, Unionidae, and Veneridae; Breton *et al.* 2009, Breton *et al.* 2011a, Breton *et al.* 2011b, Ghiselli *et al.* 2013, Milani *et al.* 2013) have been characterized to infer their functions. The fact that, despite their rapid evolution, some of them appear to have been conserved for a long time (up to ~200 million years in unionids; Breton *et al.* 2011a) may be proof in favor of a role in some cellular mechanism of the species in which they are found,

and the detection of transcripts and proteins encoded by these ORFans in unionids and venerids, inside and outside mitochondria (Breton *et al.* 2011a, Ghiselli *et al.* 2013, Milani *et al.* 2014), confirms that they are functional. Recent *in silico* analyses suggest that these ORFans may be involved in processes such as embryogenesis, reproduction, and spermatogenesis (Milani *et al.* 2013). Also, the co-occurrence in freshwater mussels of functional ORFans in gonochoric species with DUI, and of highly mutated ORFans in hermaphroditic species without DUI, may indicate that these coding sequences are involved in the maintenance of separate sexes: if so, DUI would be the first sex determination system involving mtDNA encoded proteins (Breton *et al.* 2011a).

From what we know, the functions of mitochondria and mt genomes are not only limited to the production of ATP via the OXPHOS system. Evidences are amassing that demonstrate their role in processes such as apoptosis, ageing, fertilization, cell signalling and differentiation (Scheffler 2008, Van Blerkom 2011, López-Otín *et al.* 2013, Chandel 2014), and the cases described above may indicate their role in other functions, such as sex determination in bivalves with DUI. The maintenance of additional genes and other unusual features in many taxa might have an adaptive meaning, and may be related to the ecological characteristics of these species. The adaptive response of mt genomes to environmental pressures, and how this response can have a role in speciation, is a topic of great interest. MtDNA variation may originate reproductive barriers and drive speciation (Dowling *et al.* 2008, Gershoni *et al.* 2009, Lane 2009, Ballard and Melvin 2010, Burton and Barreto 2012, Wallace 2013, Dowling 2014, Wolff *et al.* 2014): support for a role of mitochondria in the origin of post-zygotic incompatibilities, as stated by the Dobzansky-Müller model, has been found in many animal groups (Dowling *et al.* 2008, Gershoni *et al.* 2009, Lane 2009, Ballard and Melvin 2010, Burton and Barreto 2012). During the radiation of eukaryotes, for example, when gene transfer from mtDNA to nuclear genomes was common, relocation of

mitochondrial genes might have caused mitonuclear incompatibilities and thus driven speciation (Lynch 2007). It is feasible then that mt genome rearrangements, and in particular the rise of ORFans such as those described earlier in bivalves, might promote reproductive isolation and speciation.

### **3.6 AUTHORS AFFILIATIONS**

<sup>1</sup> Dipartimento di Scienze Biologiche, Geologiche ed Ambientali, Università di Bologna, via Selmi 3, 40126 Bologna

<sup>2</sup> Département de Sciences Biologiques, Université de Montréal, 90 Avenue Vincent d'Indy, Montréal, Québec H2V 2S9, Canada

<sup>3</sup> Program in Molecular and Computational Biology, Department of Biological Sciences, University of Southern California, Los Angeles, USA

<sup>4</sup> Department of Biology, Acadia University, 24 University Avenue, Wolfville, Nova Scotia B4P 2R6, Canada

## CHAPTER 4

### LITERATURE CITED

- Abad AR, Mehrtens BJ, Mackenzie SA (1995). Specific expression in reproductive tissues and fate of a mitochondrial sterility-associated protein in cytoplasmic male-sterile bean. *Plant Cell* 7: 271-285.
- 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.
- Andersson SG, Kurland CG (1998). Reductive evolution of resident genomes. *Trends Microbiol* 6: 263-268.
- Atteia A, Adrait A, Brugière S, Tardif M, van Lis R, *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.
- Ballard JWO, Melvin RG (2010). Linking the mitochondrial genotype to the organismal phenotype. *Mol Ecol* 19: 1523-1539.
- Batista FM, Lallias D, Taris N, Guedes-Pinto H, Beaumont AR (2011). Relative quantification of the M and F mitochondrial DNA types in the blue mussel *Mytilus edulis* by Real-Time PCR. *J Mollus Stud* 77: 24-29.

- Beagley CT, Taylor KA, Wolstenholme DR (1997). Gender-associated diverse mitochondrial DNA molecules of the mussel *Mytilus californianus*. *Curr Genet* 31: 318-324.
- Beckenbach AT (2011). Mitochondrial genome sequences of representatives of three families of scorpionflies (Order Mecoptera) and evolution in a major duplication of coding sequence. *Genome* 54: 368-376.
- Bernt M, Braband A, Schierwater B, Stadler PF (2013). Genetic aspects of mitochondrial genome evolution. *Mol Phylogenet Evol* 69: 328-338.
- Bilewitch JP, Degnan SM (2011). A unique horizontal gene transfer event has provided the octocoral mitochondrial genome with an active mismatch repair gene that has potential for an unusual self-contained function. *BMC Evol Biol* 11: 228.
- Birky CW Jr (2001). The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu Rev Genet* 35: 125-148.
- Boore JL (1999). Animal mitochondrial genomes. *Nucl Acids Res* 27: 1767-1780.
- Boore JL (2000). The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals, pp. 133-147 in *Comparative genomics* edited by Sankoff D, Nadeau JH. Springer, Amsterdam (The Netherlands).
- Boyle EE, Etter RJ (2013). Heteroplasmy in a deep-sea protobranch bivalve suggests an ancient origin of doubly uniparental inheritance of mitochondria in Bivalvia. *Mar Biol* 160 (2): 413-422.
- Brannock PM, Roberts MA, Hilbish T (2013). Ubiquitous heteroplasmy in *Mytilus* spp. resulting from disruption in doubly uniparental inheritance regulation. *Mar Ecol Prog Ser* 480: 131-143.
- Breton S, Doucet Beaupré H, 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, Doucet Beaupré H, Stewart DT, Piontkivska H, Karmakar M, *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, Stewart DT, Shepardson S, Trdan RJ, Bogan AE, *et al.* (2011a). Novel protein genes in animal mtDNA: a new sex determination system in freshwater mussels (Bivalvia: Unionoida)? *Mol Biol Evol* 28: 1645-1659.
- Breton S, Ghiselli F, Passamonti M, Milani L, Stewart DT, *et al.* (2011b). Evidence for a fourteenth mtDNA-encoded protein in the female-transmitted mtDNA of marine mussels (Bivalvia: Mytilidae). *PLoS One* 6: e19365.
- Breton S, Milani L, Ghiselli F, Guerra D, Stewart DT, *et al.* (2014). A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. *Trends Genet* 30 (12): 555-564.
- Burgstaller JP, Johnston IG, Jones NS, Albrechtová J, Kolbe T, *et al.* (2014). mtDNA Segregation in Heteroplasmic Tissues Is Common In Vivo and Modulated by Haplotype Differences and Developmental Stage. *Cell Rep* 7 (6): 2031-2041.
- Burton R, Barreto FS (2012). A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Mol Ecol* 21: 4942-4957.
- Cai JJ, Petrov DA (2010). Relaxed purifying selection and possibly high rate of adaptation in primate lineage-specific genes. *Genome Biol Evol* 2: 393-409.
- Cao L, Kenchington E, Zouros E (2004a). Differential Segregation Patterns of Sperm Mitochondria in Embryos of the Blue Mussel (*Mytilus edulis*). *Genetics* 166: 883-894.
- Cao L, Kenchington E, Zouros E, Rodakis GC (2004b). 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(2): 835-850.

- Cao L, Shitara H, Horii T, Nagao Y, Imai H, *et al.* (2007). The mitochondrial bottleneck occurs without reduction of mtDNA content in female mouse germ cells. *Nat Genet* 39: 386-390.
- Cao L, Shitara H, Sugimoto M, Hayashi J, Abe K, *et al.* (2009). New evidence confirms that the mitochondrial bottleneck is generated without reduction of mitochondrial DNA content in early primordial germ cells of mice. *PLoS Genet* 12: e1000756.
- Chakrabarti R, Walker JM, Stewart DT, Trdan RJ, Vijayaraghavan, *et al.* (2006). Presence of a unique male-specific extension of C-terminus to the cytochrome c oxidase subunit II protein coded by the male-transmitted mitochondrial genome of *Venustaconcha ellipsiformis* (Bivalvia: Unionoidea). *FEBS Lett* 580: 862-866.
- Chakrabarti R, Walker JM, Chapman EG, Shepardson SP, Trdan RJ, *et al.* (2007). Reproductive function for a C-terminus extended, male-transmitted cytochrome c oxidase subunit II protein expressed in both spermatozoa and eggs. *FEBS Lett* 581: 5213-5219.
- Chakrabarti R, Shepardson S, Karmakar M, Trdan R, Walker J, *et al.* (2009). Extra-mitochondrial localization and likely reproductive function of a female-transmitted cytochrome c oxidase subunit II protein. *Dev Growth Differ* 51: 511-519.
- Chandel NS (2014). Mitochondria as signaling organelles. *BMC Biol* 12: 34.
- Chapman EG, Piontkivska H, Walker JM, Stewart DT, Curole JP, *et al.* (2008). Extreme primary and secondary protein structure variability in the chimeric male-transmitted cytochrome c oxidase subunit II protein in freshwater mussels: Evidence for an elevated amino acid substitution rate in the face of domain-specific purifying selection. *BMC Evol Biol* 8: 165.
- Claverie JM, Grzela R, Lartigue A, Bernadac A, Nitsche A, *et al.* (2009). Mimivirus and Mimiviridae: giant viruses with an increasing number of potential hosts, including corals and sponges. *J Invertebr Pathol* 101: 172-180.

- Cogswell A, Kenchington EL, Zouros E (2006). Segregation of sperm mitochondria in two and four cell embryos of the blue mussel *Mytilus edulis*: implications for the mechanism of doubly uniparental inheritance of mitochondrial DNA. *Genome* 49: 799-807.
- Cohen P (2014). New role for the mitochondrial peptide humanin: protective agent against chemotherapy-induced side effects. *J Natl Cancer Inst* 106(3): dju006, doi: 10.1093/jnci/dju006.
- Cree LM, Samuels DC, De Sousa Lopes SC, Rajasimha HK, Wonnapijit P, *et al.* (2008). A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. *Nat Genet* 40: 249-254.
- Curole JP, Kocher TD (2002). Ancient sex-specific extension of the cytochrome c oxidase II gene in bivalves and the fidelity of doubly-uniparental inheritance. *Mol Biol Evol* 19: 1323-1328.
- Daubin V, Ochman H (2004a). Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. *Genome Res* 14: 1036-1042.
- Daubin V, Ochman H (2004b). Start-up entities in the origin of new genes. *Curr Opin Genet Dev* 14: 616-619.
- Davison A (2006). The ovotestis: an underdeveloped organ of evolution. *Bioessays* 28: 642-650.
- Devauchelle N (1990). Sexual development and maturity of *Tapes philippinarum*, pp. 48-62 in *Tapes philippinarum Biologia e Sperimentazione*. Ente Sviluppo Agricolo Veneto.
- 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.

- Doorbar J, Ely S, Sterling J, McLean C, Crawford L (1991). Specific interaction between HPV-16 E1-E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network. *Nature* 352: 824-827.
- Doucet-Beaupré H, Breton S, Chapman EG, Blier PU, Bogan AE, *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.
- Dowling DK, Friberg U, Lindell J (2008). Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends Ecol Evol* 23: 546-554.
- Dowling DK (2014). Evolutionary perspectives on the links between mitochondrial genotype and disease phenotype. *Biochim Biophys Acta* 1840: 1393-1403.
- Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998). Rates of spontaneous mutation. *Genetics* 148: 1667-1686.
- Dreyer H, Steiner G (2006). The complete sequences and gene organisation of the mitochondrial genomes of the heterodont bivalves *Acanthocardia tuberculata* and *Hiattella arctica* – and the first record for a putative Atpase *subunit 8* gene in marine bivalves. *Front Zool* 3: 13.
- 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.
- Fan W, Waymire KG, Narula N, Li P, Rocher C, *et al.* (2008). A mouse model of mitochondrial disease reveals germline selection against severe mtDNA mutations. *Science* 319: 958-962.

- Faure E, Delaye L, Tribolo S, Levasseur A, Seligmann H, *et al.* (2011). Probable presence of an ubiquitous cryptic mitochondrial gene on the antisense strand of the cytochrome oxidase I gene. *Biol Direct* 6: 56.
- Feschotte C, Gilbert C (2012). Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev Genet* 13: 283-296.
- Fisher C, Skibinski DOF (1990). Sex-biased mitochondrial heteroplasmy in the marine mussel *Mytilus*. *Proc Biol Sci* 242: 149-156.
- Forterre P (2006). The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res* 117: 5-16.
- Fujii S, Toriyama K (2008). Genome barriers between nuclei and mitochondria exemplified by cytoplasmic male sterility. *Plant Cell Physiol* 49: 1484-1494.
- Fujita MK, Boore JL, Moritz C (2007). Multiple origins and rapid evolution of duplicated mitochondrial genes in parthenogenetic geckos (*Heteronotia binoei*; Squamata, Gekkonidae). *Mol Biol Evol* 24: 2775-2786.
- Gallup JM, Ackermann MR (2008). The 'PREXCEL-Q Method' for qPCR. *Int J Biomed Sci* 4(4): 273-293.
- Galluzzi L, Brenner C, Morselli E, Touat Z, Kroemer G (2008). Viral control of mitochondrial apoptosis. *PLoS Pathog* 4: e1000018.
- Galtier N, Jobson RW, Nabholz B, Glémin S, Blier PU (2009). Mitochondrial whims: metabolic rate, longevity and the rate of molecular evolution. *Biol Lett* 5: 413-416.
- Garrido-Ramos MA, Stewart DT, Sutherland BW, Zouros E (1998). The distribution of male-transmitted and female-transmitted mitochondrial DNA types in somatic tissues of blue mussels: Implications for the operation of doubly uniparental inheritance of mitochondrial DNA. *Genome* 41: 818-824.

- Gemmell NJ, Allendorf FW (2001). Mitochondrial mutations may decrease population viability. *Trends Ecol Evol* 16: 115-117.
- Gershoni M, Templeton AR, Mishmar D (2009). Mitochondrial bioenergetics as a major motive force of speciation. *Bioessays* 31: 642-650.
- 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, Milani L, Chang PL, Hedgecock D, Davis JP, *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 (2): 771-786.
- Ghiselli F, Milani L, Guerra D, Chang PL, Breton S, *et al.* (2013). Structure, transcription and variability of metazoan mitochondrial genome. Perspectives from an unusual mitochondrial inheritance system. *Genome Biol Evol* 5(8), 1535-1554.
- Gingrich JR, Pelkey KA, Fam SR, Huang Y, Petralia RS, *et al.* (2004). Unique domain anchoring of Src to synaptic NMDA receptors via the mitochondrial protein NADH dehydrogenase subunit 2. *Proc Natl Acad Sci U.S.A.* 101: 6237-6242.
- 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.
- Gosling EM (2003). *Bivalve molluscs: biology, ecology and culture*. Blackwell Publishing Ltd, Oxford.
- Guerra D, Ghiselli F, Passamonti M (2014). The largest unassigned regions of the male- and female-transmitted mitochondrial DNAs in *Musculista senhousia* (Bivalvia Mytilidae). *Gene* 536: 316-325.

- Hoekstra RF (2011). Nucleo-cytoplasmic conflict and the evolution of gamete dimorphism, pp. 111-130 in *The evolution of anisogamy*, edited by Togashi T, Cox PA. Cambridge University Press, Cambridge.
- Jokinen R, Battersby BJ (2013). Insight into mammalian mitochondrial DNA segregation. *Ann Med* 45: 149-155.
- Kawashima Y, Nishihara H, Akasaki T, Nikaido M, Segawa S, *et al.* (2013). The complete mitochondrial genomes of deep-sea squid (*Bathyteuthis abyssicola*), bob-tail squid (*Semirossia patagonica*) and four giant cuttlefish (*Sepia apama*, *S. latimanus*, *S. lycidas* and *S. pharaonis*), and their application to the phylogenetic analysis of Decapodiformes. *Mol Phylogenet Evol* 69: 980-993.
- Kayal E, Lavrov DV (2008). The mitochondrial genome of *Hydra oligactis* (Cnidaria, Hydrozoa) sheds new light on animal mtDNA evolution and cnidarian phylogeny. *Gene* 410: 177-186.
- Kayal E, Bentlage B, Collins AG, Kayal M, Pirro S, *et al.* (2012). Evolution of linear mitochondrial genomes in medusozoan cnidarians. *Genome Biol Evol* 4: 1-12.
- Kenchington E, MacDonald B, Cao L, Tsagarakis D, Zouros E (2002). Genetics of mother-dependent sex ratio in blue mussels (*Mytilus* spp.) and implications for doubly uniparental inheritance of mitochondrial DNA. *Genetics* 161: 1579-1588.
- Kenchington EL, Hamilton L, Cogswell A, Zouros E (2009). Paternal mtDNA and maleness are co-inherited but not causally linked in mytilid mussels. *PLoS One* 4: e6976.
- 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.
- Khrapko K (2008). Two ways to make an mtDNA bottleneck. *Nat Genet* 2: 134-135.

- Kohn AB, Citarella MR, Kocot KM, Bobkova YV, Halanych KM, *et al.* (2012). Rapid evolution of the compact and unusual mitochondrial genome in the ctenophore *Pleurobrachia bachei*. *Mol Phylogenet Evol* 63: 203-207.
- Koonin EV, Martin W (2005). On the origin of genomes and cells within inorganic compartments. *Trends Genet* 21: 647-654.
- Koonin EV, Senkevich TG, Dolja VV (2006). The ancient Virus World and evolution of cells. *Biol Direct* 1: 29.
- Kumazawa Y, Ota H, Nishida M, Ozawa T (1996). Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. *Mol Biol Evol* 13: 1242-1254.
- Kumazawa Y, Ota H, Nishida M, Ozawa T (1998). The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. *Genetics* 150: 313-329.
- Kurabayashi A, Sumida M (2013). Afrobatrachian mitochondrial genomes: genome reorganization, gene rearrangement mechanisms, and evolutionary trends of duplicated and rearranged genes. *BMC Genomics* 14: 633.
- Kyriakou E, Zouros E, Rodakis GC (2010). The atypical presence of the paternal mitochondrial DNA in somatic tissues of male and female individuals of the blue mussel species *Mytilus galloprovincialis*. *BMC Research Notes* 3: 222.
- Kyriakou E, Chatzoglou E, Rodakis GC, Zouros E (2014). Does the ORF in the control region of *Mytilus* mtDNA code for a protein product? *Gene* 546: 448-450.
- 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 (2009). On the origin of bar codes. *Nature* 462: 272-274.

- Lane N (2011). Mitonuclear match: optimizing fitness and fertility over generations drives ageing within generations. *Bioessays* 33: 860-869.
- Lane N (2012). The problem with mixing mitochondria. *Cell* 151: 246-248.
- Lavrov DV, Forget L, Kelly M, Lang BF (2005). Mitochondrial genomes of two demosponges provide insights into an early stage of animal evolution. *Mol Biol Evol* 22: 1231-1239.
- Lavrov DV, Pett W, Voigt O, Wörheide G, Forget L, *et al.* (2013). Mitochondrial DNA of *Clathrina clathrus* (Calcarea, Calcinea): six linear chromosomes, fragmented rRNAs, tRNA editing, and a novel genetic code. *Mol Biol Evol* 30: 865-880.
- Lee C, Yen K, Cohen P (2013). Humanin: a harbinger of mitochondrial-derived peptides? *Trends Endocrinol Metab* 24: 222-228.
- Livak KL, Schmittgen TD (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method. *Methods* 25: 402-408.
- López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G (2013). The hallmarks of aging. *Cell* 153(6): 1194-1217.
- Lynch M (2007). *The origins of genome architecture*. Sinauer Associates, Sunderland (MA).
- 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.
- Martin W, Koonin EV (2006). Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 440: 41-45.

- Maximov V, Martynenko A, Hunsmann G, Tarantul V (2002). Mitochondrial 16S rRNA gene encodes a functional peptide, a potential drug for Alzheimer's disease and target for cancer therapy. *Med Hypotheses* 59: 670-673.
- Meiklejohn CD, Montooth KL, Rand DM (2007). Positive and negative selection on the mitochondrial genome. *Trends Genet* 23: 259-263.
- 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(11): e28194.
- Milani L, Ghiselli F, Passamonti M (2012). Sex-Linked Mitochondrial Behavior During Early Embryo Development in *Ruditapes philippinarum* (Bivalvia Veneridae) a Species With the Doubly Uniparental Inheritance (DUI) of Mitochondria. *J Exp Zool B Mol Dev Evol* 318: 182-189.
- Milani L, Ghiselli F, Guerra D, Breton S, Passamonti M (2013). A comparative analysis of mitochondrial ORFans: new clues on their origin and role in species with doubly uniparental inheritance of mitochondria. *Genome Biol Evol* 5: 1408-1434.
- Milani L, Ghiselli F, Maurizii MG, Nuzhdin SV, Passamonti M (2014). Paternally transmitted mitochondria express a new gene of potential viral origin. *Genome Biol Evol* 6: 391-405.
- Mishra P, Chan DC (2014). Mitochondrial dynamics and inheritance during cell division, development and disease. *Nat Rev Mol Cell Biol* 15: 634-646.
- 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, Mentel M, van Hellemond JJ, Henze K, Woehle C, *et al.* (2012). Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol Mol Biol Rev* 76: 444-495.

- Nizampatnam NR, Doodhi H, Kalinati Narasimhan Y, Mulpuri S, Viswanathaswamy DK (2009). Expression of sunflower cytoplasmic male sterility-associated open reading frame, orfH522 induces male sterility in transgenic tobacco plants. *Planta* 229: 987-1001.
- Obata M, Komaru A (2005). Specific location of sperm mitochondria in mussel *Mytilus galloprovincialis* zygotes stained by MitoTracker. *Dev Growth Differ* 47: 255-263.
- Obata M, Kamiya C, Kawamura K, Komaru A (2006). Sperm mitochondrial DNA transmission to both male and female offspring in the blue mussel *Mytilus galloprovincialis*. *Dev Growth Differ* 48: 253-261.
- Obata M, Sano N, Komaru A (2011). Different transcriptional ratios of male and female transmitted mitochondrial DNA and tissue-specific expression patterns in the blue mussel, *Mytilus galloprovincialis*. *Dev Growth Differ* 53: 878-886.
- Ogata H, Ray J, Toyoda K, Sandaa RA, Nagasaki K, *et al.* (2011). Two new subfamilies of DNA mismatch repair proteins (MutS) specifically abundant in the marine environment. *ISME J* 5: 1143-1151.
- Ohta A, Nishiyama Y (2011). Mitochondria and viruses. *Mitochondrion* 11: 1-12.
- Palumbi SR (2009). Speciation and the evolution of gamete recognition genes: pattern and process. *Heredity* 102: 66-76.
- Park E, Song JI, Won YJ (2011). The complete mitochondrial genome of *Calicogorgia granulosa* (Anthozoa: Octocorallia): potential gene novelty in unidentified ORFs formed by repeat expansion and segmental duplication. *Gene* 486: 81-87.
- Parsch J, Ellegren H (2013). The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet* 14: 83-87.

- Passamonti M, Scali V (2001). Gender-associated mitochondrial DNA heteroplasmy in the venerid clam *Tapes philippinarum* (Mollusca Bivalvia). *Curr Genet* 39: 117-124.
- Passamonti M, Boore JL, Scali V (2003). Molecular Evolution and Recombination in Gender-Associated Mitochondrial DNAs of the Manila Clam *Tapes philippinarum*. *Genetics* 164: 603-611.
- Passamonti M (2007). An unusual case of gender-associated mitochondrial DNA heteroplasmy: the mytilid *Musculista senhousia* (Mollusca Bivalvia). *BMC Evol Biol* 7(Suppl. 2): 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.
- Pett W, Ryan JF, Pang K, Mullikin JC, Martindale MQ, *et al.* (2011). Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi*: insight from mtDNA and the nuclear genome. *Mitochondrial DNA* 22: 130-142.
- Pett W, Lavrov DV (2013). The twin-arginine subunit C in *Oscarella*: origin, evolution, and potential functional significance. *Integr Comp Biol* 53: 495-502.
- Pfaffl MW (2004). Quantification strategies in real-time PCR, pp. 87-112 in *A-Z of quantitative PCR*, edited by Bustin SA. International University Line, La Jolla.
- Pikó L, Taylor KD (1987). Amounts of mitochondrial DNA and abundance of some mitochondrial gene transcripts in early mouse embryos. *Dev Biol* 123: 364-374.
- Pont-Kingdon GA, Okada NA, Macfarlane JL, Beagley CT, Wolstenholme DR, *et al.* (1995). A coral mitochondrial mutS gene. *Nature* 375: 109-111.

- Quesada H, Warren M, Skibinski DAG, Skibinski DOF (1996). Sex-biased heteroplasmy and mitochondrial DNA inheritance in the mussel *Mytilus galloprovincialis*. *Curr Genet* 29: 423-426.
- R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raboin MJ, Timko AF, Howe DK, Félix MA, Denver DR (2010). Evolution of *Caenorhabditis* mitochondrial genome pseudogenes and *Caenorhabditis briggsae* natural isolates. *Mol Biol Evol* 27(5): 1087-1096.
- Rand DM (2001). The units of selection on mitochondrial DNA. *Annu Rev Ecol Syst* 32: 415-448.
- RStudio (2012). *RStudio: Integrated development environment for R*. Boston, MA.
- Saavedra C, Reyero MI, Zouros E (1997). Male-dependent doubly uniparental inheritance of mitochondrial DNA and female-dependent sex ratio in the mussel *Mytilus galloprovincialis*. *Genetics* 145(4): 1073-1082.
- Saavedra C, Bachere E (2006). Bivalve genomics. *Aquaculture* 256: 1-14.
- Sano N, Obata M, Ooie Y, Komaru A (2011). Mitochondrial DNA copy number is maintained during spermatogenesis and in the development of male larvae to sustain the doubly uniparental inheritance of mitochondrial DNA system in the blue mussel *Mytilus galloprovincialis*. *Dev Growth Differ* 53: 816-821.
- Sato A, Nakada K, Shitara H, Kasahara A, Yonekawa H, *et al.* (2007). Deletion-mutant mtDNA increases in somatic tissues but decreases in female germ cells with age. *Genetics* 177: 2031-2037.
- Scheffler IE (2008). *Mitochondria*. Wiley-Liss, Hoboken, N.J.

- Schirtzinger EE, Tavares ES, Gonzales LA, Eberhard JR, Miyaki CY, *et al.* (2012). Multiple independent origins of mitochondrial control region duplications in the order Psittaciformes. *Mol Phylogenet Evol* 64(2): 342-356.
- Schwartz M, Vissing J (2002). Paternal Inheritance of Mitochondrial DNA. *N Engl J Med* 347: 576-580.
- Shao Z, Graf S, Chaga OY, Lavrov DV (2006). Mitochondrial genome of the moon jelly *Aurelia aurita* (Cnidaria, Scyphozoa): a linear DNA molecule encoding a putative DNA-dependent DNA polymerase. *Gene* 381: 92-101.
- Sharpley MS, Marciniak C, Eckel-Mahan K, McManus M, Crimi M, *et al.* (2012). Heteroplasmy of Mouse mtDNA Is Genetically Unstable and Results in Altered Behavior and Cognition. *Cell* 151: 333-343.
- Signorovitch AY, Buss LW, Dellaporta SL (2007). Comparative genomics of large mitochondria in placozoans. *PLoS Genet* 12: e13.
- Skibinski DO, Gallagher C, Beynon CM (1994a). Mitochondrial DNA inheritance. *Nature* 368: 817-818.
- Skibinski DOF, Gallagher C, Beynon CM (1994b). Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. *Genetics* 138(3): 801-809.
- Söding J, Biegert A, Lupas AN (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33: W244-W248.
- Stewart DT, Saavedra C, Stanwood RR, Ball AO, Zouros E (1995). Male and female mitochondrial lineages in the blue mussel (*Mytilus edulis*) species group. *Mol Biol Evol* 12: 735-747.
- Stewart JB, Freyer C, Elson JL, Wredenber A, Cansu Z, *et al.* (2008). Strong purifying selection in transmission of mammalian mitochondrial DNA. *PLoS Biology* 6: e10.

- Stöger I, Schrödl M (2013). Mitogenomics does not resolve deep molluscan relationships (yet?). *Mol Phylogenet Evol* 69: 376-392.
- Sutherland B, Stewart D, Kenchington ER, Zouros E (1998). The Fate of Paternal Mitochondrial DNA in Developing Female Mussels, *Mytilus edulis*: Implications for the Mechanism of Doubly Uniparental Inheritance of Mitochondrial DNA. *Genetics* 148: 341-347.
- Szitenberg A, Rot C, Ilan M, Huchon D (2010). Diversity of sponge mitochondria introns revealed by *cox 1* sequences of Tetillidae. *BMC Evol Biol* 10: 288.
- 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(4): 959-970.
- Thrash JC, Boyd A, Huggett MJ, Grote J, Carini P, *et al.* (2011). Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. *Sci Rep* 1: 13.
- Timmis JN, Ayliffe MA, Huang CY, Martin W (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* 5: 123-135.
- Van Blerkom J (2011). Mitochondrial function in the human oocyte and embryo and their role in developmental competence. *Mitochondrion* 11: 797-813.
- 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.
- Voigt O, Erpenbeck D, Wörheide G (2008). A fragmented metazoan organellar genome: the two mitochondrial chromosomes of *Hydra magnipapillata*. *BMC Genomics* 9: 350.
- Wai T, Teoli D, Shoubridge EA (2008). The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. *Nat Genet* 40: 1484-1488.

- Wallace DC (2013). Bioenergetics in human evolution and disease: implications for the origins of biological complexity and the missing genetic variation of common diseases. *Philos Trans R Soc Lond B Biol Sci* 368: 2012027.
- Wallace DC, Chalkia D (2013). Mitochondrial DNA Genetics and the Heteroplasmy Conundrum in Evolution and Disease. *Cold Spring Harb Perspect Biol* 5(11): a021220.
- Wang X, Lavrov DV (2008). Seventeen new complete mtDNA sequences reveal extensive mitochondrial genome evolution within the Desmospongiae. *PLoS ONE* 3(7): e2723.
- Wei DD, Shao R, Yuan ML, Dou W, Barker SC, *et al.* (2012). The multipartite mitochondrial genome of *Liposcelis bostrychophila*: insights into the evolution of mitochondrial genomes in bilateral animals. *PLoS ONE* 7(3): e33973.
- Wolff JN, Ladoukakis ED, Enríquez JA, Dowling DK (2014). Mitonuclear interactions: evolutionary consequences over multiple biological scales. *Philos Trans R Soc Lond B Biol Sci* 369: 20130443.
- Wu X, Li X, Li L, Xu X, Xia J, *et al.* (2012a). New features of Asian *Crassostrea* oyster mitochondrial genomes: a novel alloacceptor tRNA gene recruitment and two novel ORFs. *Gene* 507: 112-118.
- Wu X, Li X, Li L, Yu Z (2012b). A unique tRNA gene family and a novel, highly expressed ORF in the mitochondrial genome of the silver-lip pearl oyster, *Pinctada maxima* (Bivalvia: Pteriidae). *Gene* 510: 22-31.
- Ye K, Lu J, Ma F, Keinan A, Gu Z (2014). Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. *Proc Natl Acad Sci U.S.A.* doi: 10.1073/pnas.1403521111.
- Yu G, Stoltzfus A (2012). Population diversity of ORFan genes in *Escherichia coli*. *Genome Biol Evol* 4: 1176-1187.

Yusa Y, Breton S, Hoeh WR (2013). Population Genetics of Sex Determination in Mytilus Mussels: Reanalyses and a Model. *J Hered* 104(3): 380-385.

Zouros E, Oberhauser Ball A, Saavedra C, Freeman KR (1994a). An unusual type of mitochondrial DNA inheritance in the blue mussel Mytilus. *Proc Natl Acad Sci U.S.A.* 91: 7463-7467.

Zouros E, Oberhauser Ball A, Saavedra C, Freeman KR (1994b). Mitochondrial DNA inheritance. *Nature* 368: 818.

Zouros E (2013). Biparental inheritance through uniparental transmission: the doubly uniparental inheritance (DUI) of mitochondrial DNA. *Evol Biol* 40(1): 1-31.

## **APPENDIX 1**

This section includes the supplementary materials to Chapter 2.

**Supplementary Table 1 Quantification cycles and ratio calculations of each target for all pools in the embryo series**

<i>hpf</i>	<i>Extraction</i>	<i>Sample</i>	<i>Nu Cq</i>	<i>F Cq</i>	<i>M Cq</i>	<i>Nu ratio</i>	<i>F ratio</i>	<i>M ratio</i>	<i>avg Nu ratio</i>	<i>Nu SD</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>avg M ratio</i>	<i>M SD</i>
2	E	2h_A	31.70	22.16	27.84	3.20	40794.72	637.03	1.45	1.52	112071.49	72224.29	2408.89	1587.51
		2h_A	33.74	23.09	27.94	0.66	110212.37	2887.91						
		2h_A	34.12	22.73	28.01	0.49	185207.40	3701.72						
	E	2h_B	31.58	22.37	27.13	3.51	32600.22	919.74	1.54	1.70	86990.01	55042.10	3525.93	2303.91
		2h_B	33.92	22.90	27.22	0.58	142662.03	5291.45						
		2h_B	33.99	23.80	27.60	0.55	85707.78	4366.59						
	E	2h_C	33.63	22.75	27.33	0.72	125253.33	3938.29	0.97	0.24	91809.17	29783.88	2935.14	948.30
		2h_C	32.98	22.92	27.56	1.19	68144.79	2053.39						
		2h_C	33.22	22.92	27.36	0.99	82029.40	2813.73						
E	2h_D	33.08	23.12	27.16	1.10	64952.30	2874.60	2.12	0.98	50575.22	13805.58	1809.91	924.45	
	2h_D	32.18	22.89	27.26	2.21	37422.74	1344.09							
	2h_D	31.76	21.93	26.92	3.05	49350.61	1211.03							
E	2h_E	31.75	22.08	27.06	3.08	44580.04	1097.50	2.68	0.44	44176.05	966.89	1122.95	142.24	
	2h_E	32.18	22.60	27.34	2.21	44875.41	1276.20							
	2h_E	31.90	22.32	27.39	2.74	43072.68	995.15							
E	2h_F	32.68	23.50	27.61	1.50	37585.09	1576.64	1.11	0.35	45321.74	6706.66	2167.59	514.01	
	2h_F	33.45	24.03	27.87	0.83	48894.25	2415.26							
	2h_F	33.19	23.69	27.50	1.01	49485.87	2510.87							
E	2h_G	33.25	23.19	27.53	0.97	70892.98	2579.39	0.91	0.29	69217.99	1953.13	3194.76	568.90	
	2h_G	33.00	22.97	26.85	1.17	67072.62	3303.37							
	2h_G	33.86	23.97	27.70	0.60	69688.36	3701.53							
E	2h_H	33.80	24.08	27.60	0.63	62102.44	3770.37	0.44	0.18	100588.04	38404.74	5625.98	2034.26	
	2h_H	34.85	24.09	27.73	0.28	138911.41	7801.08							
	2h_H	34.41	24.06	27.80	0.39	100750.26	5306.49							
Q	2h_I	35.64	26.68	24.72	0.15	50515.16	100969.10	0.17	0.06	64949.46	14129.57	98887.87	22380.97	
	2h_I	35.84	26.51	24.69	0.13	65580.04	120155.53							
	2h_I	35.08	25.28	24.50	0.23	78753.19	75538.98							

<i>hpf</i>	<i>Extraction</i>	<i>Sample</i>	<i>Nu Cq</i>	<i>F Cq</i>	<i>M Cq</i>	<i>Nu ratio</i>	<i>F ratio</i>	<i>M ratio</i>	<i>avg Nu ratio</i>	<i>Nu SD</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>avg M ratio</i>	<i>M SD</i>
2	Q	2h_L	32.96	25.26	24.62	1.21	15498.81	13583.04	1.34	0.13	14802.89	1333.40	11682.19	2042.48
		2h_L	32.81	25.06	24.64	1.36	15644.36	11940.81						
		2h_L	32.71	25.20	24.87	1.47	13265.52	9522.72						
6	E	6h_A	37.00	27.56	29.72	0.05	83264.33	11316.15	0.03	0.02	110610.11	57728.55	22614.15	10275.64
		6h_A	38.17	27.80	29.54	0.02	176930.33	31402.44						
		6h_A	37.94	28.96	29.61	0.03	71635.67	25123.85						
	E	6h_B	36.84	27.45	28.59	0.06	78828.82	20795.01	0.06	na	78828.82	na	20795.01	na
		6h_B*	nn	27.86	28.76	na	na	na						
		6h_B*	nn	28.27	29.02	na	na	na						
	E	6h_C	38.82	28.76	27.89	0.01	160260.50	151129.71	0.02	0.00	612601.94	206532.42	386883.50	59838.25
		6h_C*	nn	28.40	27.86	na	na	na						
		6h_C	39.32	27.72	27.80	0.01	452341.45	235753.78						
	Q	6h_D*	nn	30.60	29.49	na	na	na	0.01	na	114465.84	na	90440.04	na
		6h_D	39.43	30.05	29.41	0.01	114465.84	90440.04						
		6h_D*	nn	27.83	28.84	na	na	na						
	Q	6h_E*	13.72	30.70	29.57	na	na	na	na	na	na	na	na	na
		6h_E*	nn	31.16	29.15	na	na	na						
		6h_E*	nn	29.57	29.51	na	na	na						
	Q	6h_F*	nn	nn	31.96	na	na	na	na	na	na	na	na	na
		6h_F*	nn	nn	34.04	na	na	na						
		6h_F*	nn	nn	33.81	na	na	na						
Q	6h_G*	nn	31.65	28.41	na	na	na	na	na	na	na	na	na	
	6h_G*	nn	32.97	28.49	na	na	na							
	6h_G*	nn	34.86	28.67	na	na	na							
Q	6h_H*	nn	33.44	29.76	na	na	na	na	na	na	na	na	na	
	6h_H*	nn	33.48	29.63	na	na	na							
	6h_H*	nn	32.69	29.46	na	na	na							
Q	6h_I*	nn	31.49	28.63	na	na	na	na	na	na	na	na	na	na

<i>hpf</i>	<i>Extraction</i>	<i>Sample</i>	<i>Nu Cq</i>	<i>F Cq</i>	<i>M Cq</i>	<i>Nu ratio</i>	<i>F ratio</i>	<i>M ratio</i>	<i>avg Nu ratio</i>	<i>Nu SD</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>avg M ratio</i>	<i>M SD</i>	
6		6h_I*	nn	33.03	28.43	na	na	na							
		6h_I*	nn	28.94	28.07	na	na	na							
	Q	6h_L*	nn	28.78	29.16	na	na	na	0.03	0.00	93728.87	4994.77	31602.69	1808.17	
		6h_L		37.76	28.37	28.98	0.03	90197.04	32881.26						
		6h_L		37.89	28.41	29.26	0.03	97260.71	30324.12						
12	E	12h_B	33.14	28.36	31.40	1.05	2556.09	193.06	0.97	0.12	2960.02	510.21	210.36	20.05	
		12h_B	33.17	27.88	31.15	1.03	3533.39	232.33							
		12h_B	33.44	28.59	31.66	0.83	2790.57	205.69							
	E	12h_C	33.76	28.51	30.39	0.65	3756.94	599.70	0.47	0.21	3123.47	1107.05	808.78	327.43	
		12h_C	34.08	28.90	30.67	0.51	3768.30	640.51							
		12h_C	35.02	31.20	30.84	0.25	1845.18	1186.13							
	Q	12h_D	37.10	37.21	31.36	0.05	213.53	4224.78	0.10	0.09	479.91	379.14	4379.56	3200.77	
		12h_D	35.29	34.37	31.07	0.20	312.23	1258.98							
		12h_D	37.45	35.32	30.86	0.04	913.98	7654.91							
	Q	12h_E	34.94	32.85	30.99	0.26	617.21	1011.77	0.55	0.44	976.37	332.56	770.24	352.14	
		12h_E	34.60	31.60	30.71	0.34	1038.27	932.76							
		12h_E	33.13	29.46	30.40	1.06	1273.64	366.19							
	Q	12h_F	32.03	29.24	30.49	2.48	624.82	147.66	1.39	0.95	1290.73	676.40	277.77	113.62	
		12h_F	33.24	29.60	30.70	0.97	1270.23	328.25							
		12h_F	33.61	29.35	31.01	0.73	1977.15	357.39							
	Q	12h_G	35.68	33.55	31.85	0.15	705.30	1026.65	0.08	0.06	366.38	296.95	2805.90	1714.17	
		12h_G	37.27	37.22	32.12	0.04	241.98	2944.47							
		12h_G	37.51	38.26	31.77	0.04	151.87	4446.57							
	Q	12h_H	37.66	34.22	31.15	0.03	2140.81	7461.25	0.04	0.03	1035.04	982.79	7593.83	4013.15	
		12h_H	36.60	34.69	30.99	0.07	703.14	3648.60							
		12h_H	38.39	38.48	31.33	0.02	261.17	11671.62							
Q	12h_I	38.84	38.31	31.27	0.01	411.31	17179.88	0.01	0.00	377.36	129.76	17257.54	315.61		
	12h_I	38.88	39.26	31.28	0.01	234.00	17604.73								

<i>hpf</i>	<i>Extraction</i>	<i>Sample</i>	<i>Nu Cq</i>	<i>F Cq</i>	<i>M Cq</i>	<i>Nu ratio</i>	<i>F ratio</i>	<i>M ratio</i>	<i>avg Nu ratio</i>	<i>Nu SD</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>avg M ratio</i>	<i>M SD</i>
12		12h_I	38.75	37.93	31.18	0.01	486.77	16988.01						
	Q	12h_L	35.79	36.82	30.91	0.14	99.07	2055.06	0.21	0.08	234.92	201.46	1638.69	391.96
		12h_L	35.21	35.56	30.62	0.21	139.31	1584.13						
		12h_L	34.78	33.10	30.44	0.30	466.39	1276.86						
	Q	12h_M	34.08	32.10	31.36	0.51	507.95	409.62	0.46	0.04	699.40	168.40	482.14	67.95
		12h_M	34.31	31.61	31.35	0.43	824.66	492.47						
		12h_M	34.23	31.63	31.10	0.45	765.58	544.34						
24	E	24h_C	30.12	22.48	28.54	10.85	9848.51	119.40	8.23	2.40	8200.05	1436.76	152.51	30.64
		24h_C	30.56	23.52	28.63	7.72	7213.84	158.24						
		24h_C	30.86	23.82	28.79	6.12	7537.81	179.88						
	Q	24h_D	33.57	28.57	29.52	0.75	3124.33	909.84	0.50	0.23	3943.14	1104.25	1386.75	505.73
		24h_D	34.61	29.04	29.61	0.34	5199.02	1917.06						
		24h_D	34.4	29.41	29.92	0.40	3506.07	1333.34						
	Q	24h_E	30.57	26.2	31.36	7.66	1357.15	27.20	6.21	1.62	1564.21	203.74	42.56	14.97
		24h_E	31.27	26.83	31.05	4.46	1571.02	57.10						
		24h_E	30.78	26.04	30.89	6.51	1764.47	43.38						
	Q	24h_F	30.38	25.81	28.61	8.87	1496.03	139.49	10.16	3.17	2123.27	701.98	142.66	26.51
		24h_F	30.54	25.55	28.49	7.84	1992.25	170.61						
		24h_F	29.81	24.06	28.19	13.78	2881.52	117.88						
	Q	24h_G*	nn	nn	nn	na	na	na	na	na	na	na	na	na
		24h_G*	nn	nn	nn	na	na	na						
		24h_G*	nn	nn	39.53	na	na	na						
	Q	24h_H*	nn	nn	nn	na	na	na	na	na	na	na	na	na
		24h_H*	nn	nn	nn	na	na	na						
		24h_H*	nn	nn	nn	na	na	na						
	Q	24h_I	33.44	27.72	28.25	0.83	4811.85	1873.64	0.76	0.22	4442.15	1029.84	2326.46	977.19
		24h_I	33.29	27.4	28.26	0.94	5236.11	1657.82						
		24h_I	34.07	29.11	28.06	0.51	3278.50	3447.92						

<i>hpf</i>	<i>Extraction</i>	<i>Sample</i>	<i>Nu Cq</i>	<i>F Cq</i>	<i>M Cq</i>	<i>Nu ratio</i>	<i>F ratio</i>	<i>M ratio</i>	<i>avg Nu ratio</i>	<i>Nu SD</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>avg M ratio</i>	<i>M SD</i>
24	Q	24h_L	38.71	36.47	29.71	0.01	1177.56	42691.17	0.12	0.18	945.72	659.35	29384.22	23725.05
		24h_L	39.27	36.82	30.35	0.01	1457.84	43468.91						
		24h_L	34.66	34.29	29.61	0.32	201.76	1992.58						
	Q	24h_M	32.39	26.67	27.25	1.88	4126.08	1591.15	2.73	1.01	4313.71	163.70	1124.53	432.74
		24h_M	32.04	26.14	27.48	2.46	4387.68	1046.03						
		24h_M	31.46	25.41	27.33	3.85	4427.37	736.41						
	Q	24h_N	28.6	23.67	27.9	35.10	1444.28	55.85	25.16	8.94	2278.25	763.03	82.41	23.41
		24h_N	29.17	23.53	27.82	22.60	2449.09	91.37						
		24h_N	29.48	23.62	28.05	17.78	2941.39	100.02						
48	E	48h_A	30.78	27.69	32.44	6.51	627.85	15.89	5.23	1.93	650.02	40.52	17.67	6.83
		48h_A	30.85	27.61	32.97	6.17	696.79	11.90						
		48h_A	31.78	28.93	32.92	3.01	625.42	25.21						
	E	48h_B	31.20	29.57	31.60	4.71	267.60	37.88	2.57	1.87	338.04	61.98	64.40	26.70
		48h_B	32.90	31.09	32.27	1.27	384.19	91.28						
		48h_B	32.50	30.69	32.34	1.72	362.33	64.04						
	E	48h_C	32.46	30.33	32.04	1.78	440.15	75.41	2.91	1.03	343.33	99.47	53.00	20.24
		48h_C	31.48	30.08	32.01	3.79	241.40	36.06						
		48h_C	31.72	29.79	31.87	3.15	348.45	47.53						
	E	48h_D	31.13	28.48	31.41	4.97	501.70	40.59	5.05	0.94	719.07	272.99	48.17	10.33
		48h_D	31.36	28.40	31.56	4.16	630.06	43.99						
		48h_D	30.88	27.03	30.51	6.03	1025.45	59.94						
	E	48h_E	30.61	26.94	30.83	7.43	880.62	39.54	8.95	2.40	866.17	69.24	24.44	13.13
		48h_E	30.56	27.05	31.98	7.72	790.85	18.06						
		48h_E	30.02	26.13	31.55	11.72	927.05	15.72						
	E	48h_F	31.63	28.71	32.31	3.38	639.25	33.34	2.44	0.82	594.55	194.91	67.31	44.52
		48h_F	32.29	30.35	31.15	2.03	381.16	117.71						
		48h_F	32.37	29.34	32.54	1.91	763.22	50.88						
	E	48h_G	32.76	30.39	31.89	1.41	534.50	104.79	1.07	0.29	422.75	229.42	111.93	10.78

<i>hpf</i>	<i>Extraction</i>	<i>Sample</i>	<i>Nu Cq</i>	<i>F Cq</i>	<i>M Cq</i>	<i>Nu ratio</i>	<i>F ratio</i>	<i>M ratio</i>	<i>avg Nu ratio</i>	<i>Nu SD</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>avg M ratio</i>	<i>M SD</i>
48		48h_G	33.30	30.94	32.27	0.93	574.87	124.33						
		48h_G	33.37	33.08	32.59	0.88	158.87	106.67						
	E	48h_H	33.41	33.10	32.16	0.85	161.81	145.36	0.73	0.20	227.72	65.50	205.09	67.54
		48h_H	34.10	33.40	31.98	0.50	228.54	278.38						
		48h_H	33.44	32.19	31.77	0.83	292.81	191.54						
	Q	48h_I*	nn	nn	34.12	na	na	na	0.03	0.02	na	na	2553.90	1339.00
		48h_I*	39.02	nn	33.94	0.01	na	3500.71						
		48h_I*	37.35	nn	33.15	0.04	na	1607.08						
	Q	48h_L	33.54	28.26	31.00	0.77	3706.83	340.77	1.33	0.49	2539.24	1238.62	192.05	129.56
	48h_L	32.52	28.75	31.62	1.70	1240.09	103.69							
	48h_L	32.67	27.71	31.43	1.51	2670.81	131.68							
86	E	86h_A	27.81	24.69	37.10	64.63	414.13	0.08	67.46	4.91	325.51	80.26	0.04	0.03
		86h_A	27.81	25.18	38.51	64.63	304.70	0.03						
		86h_A	27.65	25.25	39.72	73.13	257.71	0.01						
	E	86h_B	27.71	24.97	36.34	69.82	321.68	0.12	67.87	5.92	283.43	86.83	0.17	0.14
		86h_B	27.88	25.07	34.95	61.22	344.57	0.33						
		86h_B	27.66	25.80	37.08	72.57	184.04	0.07						
	E	86h_C	28.59	25.62	37.10	35.37	422.61	0.14	51.60	15.45	386.51	37.82	0.28	0.12
		86h_C	28.06	25.28	35.27	53.27	347.19	0.31						
		86h_C	27.78	24.75	34.63	66.14	389.72	0.38						
	E	86h_D	28.09	24.74	35.83	52.05	498.31	0.22	85.00	29.58	380.97	133.73	0.23	0.07
		86h_D	27.33	25.00	34.41	93.64	235.37	0.31						
		86h_D	27.13	23.87	35.12	109.29	409.23	0.17						
	E	86h_E	26.88	23.63	34.27	132.58	392.06	0.24	106.50	24.73	468.01	95.02	0.35	0.11
		86h_E	27.20	23.85	34.10	103.54	437.42	0.34						
		86h_E	27.48	23.76	33.95	83.40	574.56	0.47						
	E	86h_F	27.73	24.67	35.82	68.75	394.21	0.17	54.01	15.46	306.18	79.48	0.14	0.03
		86h_F	28.50	26.14	37.45	37.92	284.65	0.11						

<i>hpf</i>	<i>Extraction</i>	<i>Sample</i>	<i>Nu Cq</i>	<i>F Cq</i>	<i>M Cq</i>	<i>Nu ratio</i>	<i>F ratio</i>	<i>M ratio</i>	<i>avg Nu ratio</i>	<i>Nu SD</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>avg M ratio</i>	<i>M SD</i>
86		86h_F	28.01	25.81	36.55	55.37	239.68	0.13						
	E	86h_G	27.73	25.09	34.21	68.75	303.04	0.48	55.36	12.80	274.33	120.51	0.63	0.14
		86h_G	28.04	25.12	34.07	54.10	377.89	0.67						
		86h_G	28.33	27.04	34.21	43.24	142.07	0.76						
	E	86h_H	28.75	26.92	36.08	31.26	211.87	0.31	29.26	3.88	268.96	103.98	0.38	0.13
		86h_H	28.73	26.94	36.13	31.75	206.02	0.30						
		86h_H	29.05	26.32	35.63	24.79	388.98	0.53						
	Q	86h_I	30.47	28.08	32.94	8.28	387.04	9.04	10.67	2.11	529.10	295.82	5.04	3.51
		86h_I	29.96	27.70	34.31	12.27	331.11	2.51						
		86h_I	30.05	26.27	33.88	11.45	869.15	3.56						
	Q	86h_L	28.06	24.96	33.57	53.27	424.22	0.93	64.95	23.05	371.67	79.70	0.68	0.22
		86h_L	27.36	24.76	33.54	91.50	279.96	0.55						
		86h_L	28.14	25.11	34.46	50.08	410.81	0.56						

Abbreviations: hpf, hours post-fertilisation; Cq, quantification cycle, avg: average; SD, standard deviation; Nu, nuclear *hsp70*; F, F mtDNA *nad1*; M, M mtDNA *12S*; nn, non numeric Cq (interpreted as amplification failure); na, calculation not possible; E, MasterPure™ Complete DNA and RNA purification kit (Epicentre); Q, DNeasy® Blood & Tissue kit (Qiagen).

Method of DNA extraction of each pool is specified. All quantifications have been performed in triplicates. Efficiencies of the targets are shown in Supplementary Table 4. Nuclear and mtDNA ratios were calculated using equations 3.3 and 3.5 from Pfaffl *et al.* (2004), respectively. The median Cq value of stage 2hpf used to normalize the nuclear ratios of all replicates is 33.205.

Reactions in which a target failed to amplify, marked with an asterisk (\*), were excluded from the ratio calculation (e.g., if the nuclear target failed to amplify, no mtDNA ratio could be calculated; or, if a mtDNA target failed to amplify in a replicate, that replicate was excluded from the standard deviation calculations).

**Supplementary Table 2 Quantification cycles of all targets and mtDNA ratio calculations for each sample in the young series**

<i>Class</i>	<i>Sample</i>	<i>Nu Cq</i>	<i>F Cq</i>	<i>M Cq</i>	<i>F ratio</i>	<i>M ratio</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>avg M ratio</i>	<i>M SD</i>	<i>Sex</i>
1	1Aw	31.28	24.58	27.51	9.96	1.96E-03	10.55	0.56	2.45E-03	5.45E-04	male
	1Aw	31.49	24.56	27.40	11.08	2.35E-03					
	1Aw	31.81	24.95	27.25	10.60	3.04E-03					
	1Bw	30.34	25.58	nn	3.98	na	3.49	0.46	na	na	female
	1Bw	30.60	26.13	nn	3.42	na					
	1Bw	30.62	26.38	nn	3.06	na					
	1Cw	31.45	24.84	36.97	9.48	1.82E-06	15.54	6.86	2.93E-06	1.41E-06	male
	1Cw	31.96	24.50	36.87	14.14	2.47E-06					
	1Cw	33.13	24.60	36.78	23.00	4.51E-06					
	1Dw	30.68	26.89	35.00	2.45	5.56E-06	3.70	1.26	7.13E-06	1.36E-06	male
	1Dw	31.15	26.49	34.80	3.69	8.00E-06					
	1Dw	30.74	25.50	34.58	4.97	7.82E-06					
	1Ew	31.86	24.45	25.43	13.84	1.21E-02	11.61	1.95	8.66E-03	3.26E-03	male
	1Ew	30.72	23.90	25.24	10.75	8.28E-03					
	1Ew	30.59	23.88	25.68	10.23	5.61E-03					
	1Fw	29.02	22.68	nn	8.97	na	9.04	0.68	na	na	female
	1Fw	28.92	22.72	nn	8.40	na					
	1Fw	29.78	23.22	nn	9.75	na					
	1Gw	29.64	22.91	nn	10.64	na	10.24	0.35	na	na	female
	1Gw	29.66	23.05	nn	10.03	na					
	1Gw	30.24	23.59	nn	10.04	na					
	1Hw	29.57	24.41	nn	4.95	na	5.99	1.06	na	na	female
	1Hw	29.67	24.13	nn	5.94	na					
	1Hw	29.41	23.53	nn	7.08	na					
	1lw	29.54	23.77	nn	6.68	na	7.95	1.37	na	na	female
	1lw	29.06	23.01	nn	7.77	na					
	1lw	29.04	22.60	nn	9.41	na					

<i>Class</i>	<i>Sample</i>	<i>Nu Cq</i>	<i>F Cq</i>	<i>M Cq</i>	<i>F ratio</i>	<i>M ratio</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>avg M ratio</i>	<i>M SD</i>	<i>Sex</i>
1	1Lw	29.44	22.72	27.97	10.66	6.00E-04	10.59	1.09	5.52E-04	4.66E-05	male
	1Lw	29.61	23.12	28.30	9.47	5.07E-04					
	1Lw	29.56	22.65	28.16	11.65	5.50E-04					
2	2Aw	25.39	21.99	nn	5.35	na	4.91	0.44	na	na	female
	2Aw	26.28	23.15	nn	4.46	na					
	2Aw	26.06	22.78	nn	4.91	na					
	2Bw	26.90	22.72	36.43	8.43	1.83E-03	9.11	0.62	3.06E-03	1.16E-03	male
	2Bw	27.72	23.30	35.90	9.64	4.15E-03					
	2Bw	27.52	23.17	36.12	9.26	3.21E-03					
	2Cw	27.06	22.37	nn	11.51	na	11.20	1.26	na	na	female
	2Cw	26.88	22.09	nn	12.27	na					
	2Cw	26.61	22.19	nn	9.82	na					
	2Dw	27.86	20.93	38.48	45.05	9.22E-04	44.42	1.70	8.75E-04	6.64E-05	male
	2Dw	28.03	21.19	38.82	42.50	8.28E-04					
	2Dw	27.33	20.39	nn	45.72	na					
	2Ew	25.24	23.03	35.37	2.58	1.30E-03	2.62	0.06	1.09E-03	6.51E-04	male
	2Ew	25.14	22.86	34.92	2.70	1.62E-03					
	2Ew	25.26	23.04	37.46	2.59	3.64E-04					
	2Fb	28.28	26.10	nn	2.41	na	2.32	0.08	na	na	female
	2Fb	28.42	26.32	nn	2.29	na					
	2Fb	28.50	26.42	nn	2.26	na					
	2Gb	29.44	25.89	nn	5.49	na	5.03	0.40	na	na	female
	2Gb	29.50	26.19	nn	4.73	na					
	2Gb	29.62	26.26	nn	4.87	na					
2Hb	28.45	24.80	37.43	5.93	2.50E-03	5.16	1.67	2.05E-03	3.99E-04	male	
2Hb	27.49	24.85	36.99	3.24	1.85E-03						
2Hb	28.52	24.77	38.05	6.30	1.78E-03						
2Ib	27.99	23.01	nn	13.54	na	18.49	5.90	na	na	female	

<i>Class</i>	<i>Sample</i>	<i>Nu Cq</i>	<i>F Cq</i>	<i>M Cq</i>	<i>F ratio</i>	<i>M ratio</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>avg M ratio</i>	<i>M SD</i>	<i>Sex</i>
2	2Ib	28.00	22.66	nn	16.90	na					
	2Ib	27.69	21.72	nn	25.03	na					
	2Lb	26.51	22.30	nn	8.64	na	8.33	0.30	na	na	female
	2Lb	26.70	22.60	nn	8.05	na					
	2Lb	26.76	22.61	nn	8.29	na					
	2Mb	27.29	22.47	25.40	12.42	2.05	16.04	3.15	3.42	1.27	male
	2Mb	28.50	23.09	25.63	17.50	3.67					
	2Mb	28.77	23.29	25.54	18.19	4.56					
	2Nb	29.68	25.54	nn	7.86	na	6.85	1.12	na	na	female
	2Nb	29.65	25.69	nn	7.04	na					
	2Nb	29.56	25.96	nn	5.65	na					
	2Ob	29.68	26.52	nn	4.30	na	3.62	0.73	na	na	female
	2Ob	28.81	26.34	nn	2.85	na					
	2Ob	29.16	26.26	nn	3.70	na					
	2Fm	26.11	23.62	nn	3.02	na	2.89	0.17	na	na	female
	2Fm	26.10	23.64	nn	2.96	na					
	2Fm	26.25	23.94	nn	2.70	na					
	2Gm	27.13	23.38	nn	6.45	na	7.03	0.97	na	na	female
	2Gm	27.50	23.36	nn	8.14	na					
	2Gm	27.43	23.66	nn	6.49	na					
	2Hm	26.44	22.81	nn	6.05	na	6.31	0.35	na	na	male
	2Hm	26.49	22.83	nn	6.16	na					
	2Hm	26.57	22.77	nn	6.71	na					
	2Im	26.21	22.01	nn	8.63	na	10.52	1.66	na	na	female
	2Im	26.25	21.63	nn	11.17	na					
	2Im	25.77	21.08	nn	11.75	na					
	2Lm	25.33	19.68	nn	21.36	na	20.32	0.91	na	na	female
	2Lm	25.45	19.93	nn	19.68	na					

<i>Class</i>	<i>Sample</i>	<i>Nu Cq</i>	<i>F Cq</i>	<i>M Cq</i>	<i>F ratio</i>	<i>M ratio</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>avg M ratio</i>	<i>M SD</i>	<i>Sex</i>
2	2Lm	25.50	19.96	nn	19.91	na					
	2Mm	27.12	20.13	nn	47.31	na	60.16	14.26	na	na	male
	2Mm	27.81	20.48	nn	57.66	na					
	2Mm	28.25	20.47	nn	75.51	na					
	2Nm	28.34	21.32	nn	47.24	na	39.11	7.10	na	na	female
	2Nm	27.79	21.31	nn	34.19	na					
	2Nm	28.60	22.02	nn	35.89	na					
	2Om	28.00	22.59	nn	17.64	na	16.92	1.34	na	na	female
	2Om	27.79	22.61	nn	15.37	na					
	2Om	27.66	22.25	nn	17.74	na					
	2Fa	31.77	26.48	nn	15.41	na	15.00	0.36	na	na	female
	2Fa	31.88	26.66	nn	14.73	na					
	2Fa	32.52	27.27	nn	14.85	na					
	2Ga	31.20	25.30	nn	22.64	na	25.62	4.38	na	na	female
	2Ga	31.74	25.76	nn	23.57	na					
	2Ga	32.24	25.82	nn	30.64	na					
	2Ha	33.83	27.36	nn	30.79	na	35.77	5.66	na	na	male
	2Ha	33.07	26.43	nn	34.61	na					
	2Ha	33.01	26.06	nn	41.92	na					
	2Ia	32.87	26.48	nn	29.77	na	31.42	1.65	na	na	female
	2Ia	32.95	26.47	nn	31.43	na					
	2Ia	32.10	25.56	nn	33.07	na					
	2La	30.71	24.58	nn	26.29	na	26.58	0.95	na	na	female
	2La	31.02	24.80	nn	27.64	na					
	2La	30.70	24.60	nn	25.81	na					
	2Ma	33.59	26.25	nn	52.79	na	120.70	75.06	na	na	male
	2Ma	34.57	26.04	nn	108.02	na					
	2Ma	36.01	26.43	nn	201.29	na					

<i>Class</i>	<i>Sample</i>	<i>Nu Cq</i>	<i>F Cq</i>	<i>M Cq</i>	<i>F ratio</i>	<i>M ratio</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>avg M ratio</i>	<i>M SD</i>	<i>Sex</i>
2	2Oa	35.11	26.39	nn	120.35	na	164.59	92.20	na	na	female
	2Oa	34.94	26.48	nn	102.85	na					
	2Oa	36.73	26.65	nn	270.58	na					
3	3Ab	26.30	22.53	24.17	530.03	113.85	772.35	235.51	138.91	21.82	male
	3Ab	27.33	22.79	24.94	1000.40	153.63					
	3Ab	26.85	22.58	24.41	786.61	149.25					
	3Bb	27.59	21.40	nn	2920.88	na	3420.60	994.09	na	na	female
	3Bb	28.32	21.59	nn	4565.41	na					
	3Bb	27.92	21.89	nn	2775.51	na					
	3Cb	27.59	22.04	27.41	1956.75	38.00	1775.33	177.45	30.00	6.93	male
	3Cb	27.62	22.24	28.04	1767.10	25.87					
	3Cb	27.34	22.05	27.69	1602.14	26.12					
	3Db	32.92	22.05	nn	120822.28	na	110288.65	35805.30	na	na	female
	3Db	33.22	22.19	nn	139645.55	na					
	3Db	31.94	21.70	nn	70398.12	na					
	3Eb	30.87	24.69	nn	4728.71	na	4047.57	620.56	na	na	female
	3Eb	30.50	24.54	nn	3899.68	na					
	3Eb	30.60	24.83	nn	3514.32	na					
4	4Ab	28.27	19.08	nn	21133.86	na	22431.88	2928.42	na	na	female
	4Ab	28.74	19.72	nn	20376.81	na					
	4Ab	29.06	19.74	nn	25784.97	na					
	4Bb	28.10	17.85	nn	40007.91	na	54353.08	15882.17	na	na	female
	4Bb	28.51	17.95	nn	51630.96	na					
	4Bb	29.05	18.10	nn	71420.39	na					
	4Cb	23.37	23.10	18.55	38.33	446.58	34.08	3.73	381.51	56.70	male
	4Cb	23.20	23.21	18.70	31.36	355.26					
	4Cb	23.07	22.99	18.60	32.55	342.69					
	4Db	29.52	17.12	nn	189829.21	na	151598.81	36715.44	na	na	female

<i>Class</i>	<i>Sample</i>	<i>Nu Cq</i>	<i>F Cq</i>	<i>M Cq</i>	<i>F ratio</i>	<i>M ratio</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>avg M ratio</i>	<i>M SD</i>	<i>Sex</i>
4	4Db	29.19	17.49	nn	116613.80	na					
	4Db	29.63	17.65	nn	148353.42	na					
	4Eb	24.64	21.27	19.88	322.33	505.15	278.67	43.65	544.77	70.14	male
	4Eb	25.05	22.01	20.04	278.67	625.76					
	4Eb	24.41	21.49	19.61	235.02	503.41					
	4Fb	23.03	20.57	18.77	143.51	297.63	102.98	35.22	247.58	43.83	male
	4Fb	22.70	21.10	18.87	79.76	216.04					
	4Fb	22.55	20.80	18.60	85.68	229.06					
	4Gb	28.40	20.47	nn	9791.71	na	15957.98	5566.50	na	na	female
	4Gb	29.22	20.56	nn	17469.88	na					
	4Gb	29.28	20.37	nn	20612.35	na					
	4Hb	26.19	21.22	19.15	1105.10	2692.10	1102.15	23.43	2855.60	189.10	male
	4Hb	26.43	21.49	19.37	1123.97	2811.99					
	4Hb	26.44	21.57	19.25	1077.39	3062.69					
	4Ib	25.34	21.04	21.28	640.24	351.16	1189.09	482.00	685.46	290.23	male
	4Ib	26.29	20.81	21.01	1543.49	873.00					
	4Ib	26.27	20.96	21.06	1383.56	832.21					
	4Lb	29.10	18.99	nn	42531.05	na	56927.26	20939.94	na	na	female
	4Lb	30.06	19.15	nn	80949.25	na					
	4Lb	29.52	19.34	nn	47301.47	na					
	4Mb	24.34	22.33	20.28	131.59	309.08	106.72	22.61	260.36	42.25	male
	4Mb	24.03	22.60	20.34	87.40	233.83					
	4Mb	23.92	22.23	20.18	101.18	238.16					
	4Ob	30.21	18.51	nn	135725.18	na	171083.41	30779.98	na	na	female
	4Ob	30.59	18.48	nn	185639.49	na					
	4Ob	30.56	18.39	nn	191885.57	na					
	4Pb	26.79	17.55	nn	17495.89	na	83617.10	61995.65	na	na	female
	4Pb	29.22	17.89	nn	92919.55	na					

<i>Class</i>	<i>Sample</i>	<i>Nu Cq</i>	<i>F Cq</i>	<i>M Cq</i>	<i>F ratio</i>	<i>M ratio</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>avg M ratio</i>	<i>M SD</i>	<i>Sex</i>
4	4Pb	29.64	17.75	nn	140435.85	na					
	4Qb	24.48	20.84	19.81	372.70	466.94	389.41	46.84	471.16	73.12	male
	4Qb	25.00	21.21	20.19	442.32	546.30					
	4Qb	24.54	21.00	20.12	353.22	400.25					
	4Rb	26.78	21.16	19.19	1812.27	4143.44	2001.43	594.20	5029.93	1338.45	male
	4Rb	27.40	21.31	19.22	2667.18	6569.55					
	4Rb	27.26	22.03	19.68	1524.83	4376.79					
	4Sb	27.42	19.47	nn	8569.78	na	17319.39	7788.91	na	na	female
	4Sb	28.62	19.61	nn	19891.32	na					
	4Sb	28.73	19.48	nn	23497.07	na					
	4Am	26.27	19.80	nn	2859.79	na	3392.95	553.10	na	na	female
	4Am	26.46	19.78	nn	3355.04	na					
	4Am	26.95	20.12	nn	3964.03	na					
	4Bm	25.81	19.04	nn	3222.27	na	3947.21	729.24	na	na	female
	4Bm	26.36	19.40	nn	3938.69	na					
	4Bm	26.30	19.05	nn	4680.67	na					
	4Cm	25.58	19.87	nn	1603.79	na	2015.29	631.86	na	na	male
	4Cm	25.59	19.79	nn	1699.26	na					
	4Cm	26.41	20.04	nn	2742.81	na					
	4Dm	25.26	18.96	nn	2212.36	na	2231.51	160.30	na	na	female
	4Dm	25.43	19.04	nn	2400.53	na					
	4Dm	25.65	19.54	nn	2081.65	na					
	4Em	25.67	19.58	nn	2061.88	na	1724.22	456.42	na	na	male
	4Em	25.73	19.78	nn	1905.82	na					
	4Em	24.71	19.25	nn	1204.95	na					
	4Fm	23.89	20.05	22.49	386.90	52.19	496.11	96.83	66.23	12.23	male
	4Fm	24.45	20.12	22.61	571.46	74.52					
	4Fm	24.28	20.03	22.46	529.97	71.98					

<i>Class</i>	<i>Sample</i>	<i>Nu Cq</i>	<i>F Cq</i>	<i>M Cq</i>	<i>F ratio</i>	<i>M ratio</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>avg M ratio</i>	<i>M SD</i>	<i>Sex</i>
4	4Gm	25.94	19.92	nn	2054.38	na	2507.62	500.82	na	na	female
	4Gm	26.04	19.78	nn	2423.17	na					
	4Gm	26.44	19.91	nn	3045.29	na					
	4Hm	26.31	20.58	nn	1810.32	na	2499.84	735.26	na	na	male
	4Hm	27.01	20.50	nn	3273.59	na					
	4Hm	26.65	20.54	nn	2415.60	na					
	4Ba	29.63	19.21	nn	55875.92	na	54254.34	9273.78	na	na	female
	4Ba	29.16	19.00	nn	44276.71	na					
	4Ba	29.89	19.35	nn	62610.38	na					
	4Ca	26.13	19.33	26.88	3443.46	17.28	3401.44	49.21	12.69	3.99	male
	4Ca	26.15	19.40	27.65	3347.30	10.66					
	4Ca	26.45	19.74	28.09	3413.55	10.12					
	4Da	27.60	19.45	nn	9976.30	na	8654.78	1154.27	na	na	female
	4Da	27.33	19.50	nn	7843.92	na					
	4Da	27.33	19.44	nn	8144.11	na					
	4Ea	28.69	20.38	nn	12968.74	na	11534.99	1332.24	na	na	male
	4Ea	28.48	20.34	nn	11300.95	na					
	4Ea	28.09	20.00	nn	10335.27	na					
	4Fa	26.13	18.47	nn	5899.00	na	5407.10	491.76	na	na	male
	4Fa	26.04	18.65	nn	4915.48	na					
	4Fa	26.47	19.03	1.84*	5406.84	na					
4Ga	27.83	20.58	nn	5877.25	na	6408.08	473.84	na	na	female	
4Ga	27.81	20.38	nn	6558.63	na						
4Ga	28.42	21.08	nn	6788.36	na						

Abbreviations: Cq, quantification cycle; Nu, nuclear *hsp70*; F, F mtDNA *nad1*; M, M mtDNA *12S*; avg, average; SD, standard deviation; nn, non numeric Cq (target non detectable); na, calculation not possible.

Suffixes to sample names: w, whole animal; b, body; a, adductor muscle; m, mantle.

All quantifications have been performed in triplicates. Efficiencies of the targets are provided in Supplementary Table 4. mtDNA ratios were calculated using equations 3.5 from Pfaffl *et al.* (2004). The sex of each specimen was assigned after the quantifications in whole animals (classes 1 and 2) or their respective bodies (classes 2, 3, and 4): animals in which M mtDNA signal was absent were considered as females, while those in which M was present were classified as males (see main text for details).

The M Cq in a replicate of sample 4Fa, marked with an asterisk (\*), has been considered as an erroneous read and was excluded from the calculations.

Male specimen 2M can be considered an outlier among class 2 males: compared to other males of the same class, its body has the highest M ratio compared, and its adductor muscle and mantle have the highest F ratios.

**Supplementary Table 3 Details of primers and probes used in the Real-Time qPCR experiments.**

<i>Target</i>	<i>Type</i>	<i>Name</i>	<i>Direction</i>	<i>Length (bp)</i>	<i>Sequence 5'-3'</i>	<i>5' fluorophore</i>	<i>3' quencher</i>
nuclear <i>hsp70</i>	primer	Rph_Nu FWD	forward	24	TCACTTTGTTGAGGAGTTCAAACG		
	primer	Rph_Nu REV	reverse	20	TTGCTTCGGCACTGTTAGAC		
	probe	Rph_Nu probe	reverse	24	CTCTTTGCTCGCTCACACGCCGTC	6-FAM	BHQ1
F mtDNA <i>nad1</i>	primer	Rph_F FWD	forward	24	TTAGGTCTGTTTTTCATTGGGTTTCG		
	primer	Rph_F REV	reverse	24	GCAAAATTTACCCCACCAAATTCC		
	probe	Rph_F probe	reverse	24	ACCTGCCACCAACTCTGACTCCCC	Cy5	BHQ3
M mtDNA <i>12S</i>	primer	Rph_M FWD	forward	20	TGACCCGCCTTTTCAGCTAAC		
	primer	Rph_M REV	reverse	24	TAGGAATAGTTTAACCGCGATTGC		
	probe	Rph_M probe	forward	24	CGCTTGTCATGGGCTCTGCTCCAG	HEX	BHQ1

Abbreviations: 6-FAM, 6-carboxyfluorescein; Cy5, cyanine 5; HEX, 6-carboxy-2',4',5',7'-hexachlorofluorescein; BHQ1-3, Black Hole Quencher 1-3.

**Supplementary Table 4 Primer and probe concentrations used in the quantification experiments, plus target efficiencies**

<i>Stock1</i>	<i>Probes</i> $\mu$ M	<i>Primers Nu</i> $\mu$ M	<i>Primers F</i> $\mu$ M	<i>Primers M</i> $\mu$ M	<i>Nu eff%</i>	<i>F eff%</i>	<i>M eff%</i>
Embryo series	0.30	0.10	0.05	1.00	86.17	70.72	118.99
Young series class 1	0.30	0.125	0.10	0.75	58.25	62.63	111.10
Young series class 2	0.30	0.20	0.15	1.00	81.79	85.20	85.19
Young series classes 3+4	0.25	0.60	0.30	0.50	116.56	87.06	91.15

Abbreviations: Nu, nuclear *hsp70*; F, F mtDNA *nad1*; M, M mtDNA *12S*; eff%, amplification efficiency of the target in percentage.

The enlisted reagent concentrations were used for the quantification in each of the four Stock1 dilution series to obtain the respective efficiencies, and for the respective samples in Real-Time. Each Stock1 contains all samples from the indicated group(s). Real-Time experiments to evaluate the efficiencies have been performed following the Stock1 procedure by Gallup and Ackermann (2008). Primer and probe features are enlisted in Supplementary Table 3. Probe concentration for each group of samples is the same for all targets.

**Supplementary Table 5 Ratio statistics for each stage of the embryo series.**

<i>hpf</i>	<i>N</i>	<i>mdn Nu</i>	<i>Nu MAD</i>	<i>avg Nu</i>	<i>Nu SD</i>	<i>mdn F</i>	<i>F MAD</i>	<i>avg F</i>	<i>F SD</i>	<i>mdn M</i>	<i>M MAD</i>	<i>avg M</i>	<i>M SD</i>
2	30	1	0.68	1.27	0.99	63,527.37	29,208.72	68,050.21	40,074.67	3,095.64	2,647.18	13,336.12	29,759.34
6	9	0.02	0.02	0.03	0.02	97,260.71	27,327.11	147,242.74	120,131.65	31,402.44	15,726.58	69,907.37	76,480.89
12	30	0.25	0.33	0.43	0.43	704.22	691.23	1,154.36	1,121.54	1,106.39	1,224.77	3,622.48	5,338.21
24	24	4.15	5.43	6.73	8.50	3,032.86	2,222.92	3,476.31	2,324.91	458.14	626.93	4,330.26	11,968.77
48	27	2.03	1.77	3.13	2.80	574.87	320.21	744.54	768.16	59.94	62.09	257.24	688.87
86	30	58.30	22.16	59.27	29.22	362.54	88.64	359.47	136.08	0.31	0.28	0.79	1.72

Abbreviations: hpf, hours post fertilization; N, number of replicates considered for the analyses; mdn, median; MAD, median absolute deviation; avg, average; SD, standard deviation; Nu, nuclear ratio; F, F mtDNA ratio; M, M mtDNA ratio.

Statistics are calculated on all suitable technical replicates available for each pool. 2hpf median Nu ratio is 1 because its median Cq in this stage has been used as the calibrator to normalize the Nu ratios of the other stages (see Materials and Methods).

**Supplementary Table 6 mtDNA ratio statistics for all sample types in the four classes of the young series.**

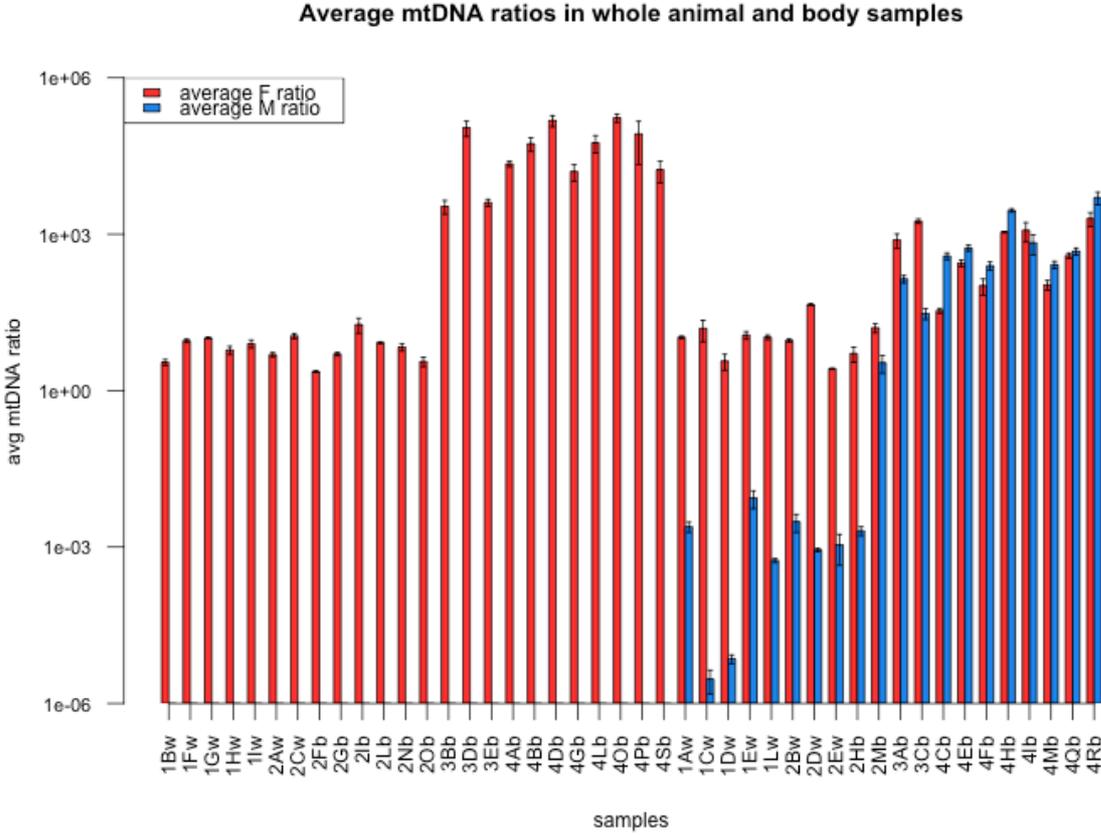
<i>Class</i>	<i>Tissue</i>	<i>mdn F ratio</i>	<i>F MAD</i>	<i>avg F ratio</i>	<i>F SD</i>	<i>mdn Fm ratio</i>	<i>Fm MAD</i>	<i>avg Fm ratio</i>	<i>Fm SD</i>	<i>mdn M ratio</i>	<i>M MAD</i>	<i>avg M ratio</i>	<i>M SD</i>
1	whole animal	7.77	2.93	7.34	2.57	10.60	1.65	10.40	4.83	5.50E-04	8.12E-04	2.33E-03	3.63E-03
2	whole animal	7.58	4.29	8.05	3.55	9.26	9.89	18.72	19.50	1.46E-03	8.65E-04	1.78E-03	1.28E-03
	body	5.57	3.85	7.44	5.86	9.36	7.08	10.60	6.37	1.02	1.52	1.71	2.04
	whole animal + body	5.57	3.85	7.59	5.31	9.26	9.74	15.33	15.77	1.84E-03	1.43E-03	1.85E-03	1.54
	adductor	27.64	7.42	52.64	67.86	47.36	21.73	78.24	66.56	0	na	0	na
	mantle	13.56	9.94	16.13	12.41	27.01	30.99	33.23	30.85	0	na	0	na
3	body	4,565.40	2,438.17	39,252.27	56,208.46	1,301.27	726.83	1,273.84	580.15	75.92	74.02	84.45	61.38
4	body	44,916.26	40,672.72	71,661.11	62,589.45	337.77	413.25	650.57	712.82	485.17	322.14	1,309.54	1,704.89
	adductor	8,144.11	2,716.40	23,105.73	23,845.21	5,406.84	2,955.24	6,781.18	3,737.75	10.66	0.80	12.69	3.99
	mantle	2,952.54	957.90	3,019.82	846.25	1,754.79	897.45	1,683.86	898.44	71.98	3.76	66.23	12.23

Abbreviations: mdn, median; MAD, median absolute deviation; avg, average; SD, standard deviation; F, F mtDNA ratio in female samples; M, M mtDNA ratio in male samples; Fm, F mtDNA ratio in male samples; na, calculation not possible.

Statistics are calculated on all suitable technical replicates available for each sample. See Table 1 for sample size of female and male tissues for each class.

No M mtDNA has been detected in class 2 adductors and mantles.

**Supplementary Figure** Barplot of average mtDNA ratios in whole animal and body samples of the young series. Error bars represent standard deviations (SD). Average ratios and relative SD have been calculated following Livak and Schmittgen (2001) (see Supplementary Table 2 for these values). Y axis is in  $\log_{10}$  scale. Animals that did not show a M mtDNA signal were considered females (from 1Bw to 4Sb), while those showing M mtDNA males (from 1Aw to 4Rb). Samples are ordered first by sex, then by class, and finally by name in alphabetical order. Suffixes to sample names indicate sample type: w, whole animal; b, body. The average ratios have been used for the cluster analysis (see results in Figure 13). F ratio values are comparable between females and males of classes 1 and 2. F ratios in classes 3 and 4 are higher in females than in males. M ratio is lower than F in all males from classes 1, 2, and 3, while it is higher than F in all males of class 4 except 4Ib. Male sample 2Mb has an unusually high M ratio value compared to the other males of class 2 (see outlier points of M ratio in Figure 12B, and Figure 13C).



## APPENDIX 2

This section includes the full texts of the following papers.

**A comparative analysis of mitochondrial ORFans: new clues on their origin and role in species with doubly uniparental inheritance of mitochondria**

Liliana Milani, Fabrizio Ghiselli,  Davide Guerra, Sophie Breton, Marco Passamonti

*Genome Biology and Evolution* 2013, 5 (7): 1408-1434

**Structure, transcription, and variability of metazoan mitochondrial genome: perspectives from an unusual mitochondrial inheritance system**

Fabrizio Ghiselli, Liliana Milani,  Davide Guerra, Peter L. Chang, Sophie Breton, Sergey V.

Nuzhdin, Marco Passamonti

*Genome Biology and Evolution* 2013, 5 (8): 1535-1554

**The largest unassigned regions of the male- and female-transmitted mitochondrial DNAs in *Musculista senhousia* (Bivalvia Mytilidae)**

Davide Guerra, Fabrizio Ghiselli, Marco Passamonti

*Gene* 2014, 536: 316-325

**A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs**

Sophie Breton, Liliana Milani, Fabrizio Ghiselli,  Davide Guerra, Donald T. Stewart, Marco Passamonti

*Trends in Genetics* 2014, 30 (12): 555-564

# A Comparative Analysis of Mitochondrial ORFans: New Clues on Their Origin and Role in Species with Doubly Uniparental Inheritance of Mitochondria

Liliana Milani<sup>1,\*</sup>, Fabrizio Ghiselli<sup>1</sup>, Davide Guerra<sup>1</sup>, Sophie Breton<sup>2</sup>, and Marco Passamonti<sup>1</sup>

<sup>1</sup>Dipartimento di Scienze Biologiche, Geologiche ed Ambientali, University of Bologna, Bologna, Italy

<sup>2</sup>Département de Sciences Biologiques, Université de Montréal, Montréal, Québec, Canada

\*Corresponding author: E-mail: liliana.milani@unibo.it.

Accepted: June 27, 2013

## Abstract

Despite numerous comparative mitochondrial genomics studies revealing that animal mitochondrial genomes are highly conserved in terms of gene content, supplementary genes are sometimes found, often arising from gene duplication. Mitochondrial ORFans (ORFs having no detectable homology and unknown function) were found in bivalve molluscs with Doubly Uniparental Inheritance (DUI) of mitochondria. In DUI animals, two mitochondrial lineages are present: one transmitted through females (F-type) and the other through males (M-type), each showing a specific and conserved ORF. The analysis of 34 mitochondrial major Unassigned Regions of *Musculista senhousia* F- and M-mtDNA allowed us to verify the presence of novel mitochondrial ORFs in this species and to compare them with ORFs from other species with ascertained DUI, with other bivalves and with animals showing new mitochondrial elements. Overall, 17 ORFans from nine species were analyzed for structure and function. Many clues suggest that the analyzed ORFans arose from endogenization of viral genes. The co-option of such novel genes by viral hosts may have determined some evolutionary aspects of host life cycle, possibly involving mitochondria. The structure similarity of DUI ORFans within evolutionary lineages may also indicate that they originated from independent events. If these novel ORFs are in some way linked to DUI establishment, a multiple origin of DUI has to be considered. These putative proteins may have a role in the maintenance of sperm mitochondria during embryo development, possibly masking them from the degradation processes that normally affect sperm mitochondria in species with strictly maternal inheritance.

**Key words:** mitochondrial ORFans, mitochondrial inheritance, Doubly Uniparental Inheritance of mitochondria, endogenous virus.

## Introduction

Comparative mitochondrial genomics revealed that animal mitochondrial DNAs (mtDNAs) are highly conserved in terms of gene content (Boore 1999; Gissi et al. 2008). These small, typically circular and intron-less molecules encode 2 ribosomal RNAs, 22 transfer RNAs, and 13 protein subunits of the mitochondrial respiratory complexes and ATP synthase. The other subunits of the electron transport chain and all the proteins involved in other mitochondrial functions, such as mtDNA replication and expression, are encoded by the nucleus (Boore 1999). However, supplementary genes are sometimes found in mtDNA. Many mechanisms are responsible for the origin of such new genes. For example, novel mitochondrial Open Reading Frames (ORFs) can arise from gene duplication.

In bivalve molluscs, a *cox2* duplication is found in the clam *Ruditapes philippinarum* (Bivalvia, Veneridae) (Okazaki M and Ueshima R, unpublished data; GenBank AB065375.1) and in the mussel *Musculista senhousia* (Bivalvia, Mytilidae) (Passamonti et al. 2011). Moreover, *nad2* duplication is at the origin of two novel ORFs in the oyster genus *Crassostrea* (Bivalvia, Ostreidae) (Wu et al. 2012). Extra elements were also found in Cnidaria mtDNA, either from duplication of extant genes or not: a duplicated *cox1* in some hydrozoan hydrozoans (Cnidaria, Hydrozoa), two novel ORFs in Medusozoa (Kayal et al. 2011), and a novel ORF in every octocoral (Cnidaria, Anthozoa) that has been screened to date (McFadden et al. 2010). One of the two medusozoan ORFs shares several conserved motifs characteristic of the

polymerase domain typical of family B-DNA polymerases (polB; Shao et al. 2006). The other ORF, named ORF314, do not resemble any other known protein. Kayal et al. (2011) attributed the origin of these two extra elements to an ancient invasion by a linear plasmid that caused the linearization of the mtDNA in Medusozoa, consistent with a previously established hypothesis for polB-like sequences found in the linear mtDNA of fungi and algae (Mouhamadou et al. 2004). The conservation of both sequence length and position suggested some level of selection pressure for their maintenance in the mtDNA of most medusozoans (Kayal et al. 2011). The octocoral extra ORF is recognized as a putatively DNA mismatch repair protein (mtMutS) (Pont-Kingdon et al. 1995; Claverie et al. 2009; Bilewitch and Degnan 2011; Ogata et al. 2011). As for medusozoan ORFs, mtMutS was supposed to be originated by horizontal gene transfer, but in this case either through an epsilonproteobacterium or a viral infection (Claverie et al. 2009; Bilewitch and Degnan 2011; Ogata et al. 2011).

Interestingly, novel mitochondrial ORFs have been also discovered in bivalve molluscs with Doubly Uniparental Inheritance (DUI) of mitochondria (Skibinski et al. 1994a, 1994b; Zouros et al. 1994a, 1994b). Specifically, in metazoans, mitochondria are commonly inherited maternally by Strictly Maternal Inheritance (SMI) (Birky 2001), whereas in DUI animals two mitochondrial lineages are present: one transmitted through females (F-type) and the other through males (M-type). In DUI bivalves, females inherit F-type mtDNA, whereas males inherit both F- and M-types (Skibinski et al. 1994a, 1994b; Zouros et al. 1994a, 1994b). In DUI bivalves (orders Mytiloidea, Unionoidea, and Veneroidea), two novel lineage-specific ORFs were found, one in the F-mtDNA (fORF) and one in the M-mtDNA (mORF) (Breton et al. 2009; Breton et al. 2011a, 2011b; Ghiselli et al. 2013). These novel ORFs have been hypothesized to be responsible for the different mode of mtDNA transmission and the maintenance of gonochorism in DUI bivalves (Breton et al. 2009, 2011a, 2011b).

In all the analyzed DUI *Mytilus* species, the novel fORF is localized in the Largest Unassigned Region (LUR) and encodes a putative protein of more than 100 amino acids (aa), suggesting its maintenance in the subfamily Mytilinae for more than 10 million years (Breton et al. 2011b). A fORF is present also in the F-mtDNA of *Musculista senhousia*, a DUI mytilid of the subfamily Crenellinae (Breton et al. 2011b). In the venerid *R. philippinarum*, the fORF is localized in the Female Largest Unassigned Region (FLUR), whereas the mORF in the Male Unassigned Region 21 (Ghiselli et al. 2013). Interestingly, the two lineage-specific ORFs found in the freshwater mussel *Venustaconcha ellipsiformis* (Bivalvia, Unionidae), the fORF (found between *tRNA-Glu* and *nad2*) and the mORF (found between *tRNA-Asp* and *nad4L*), are both translated (Breton et al. 2009), and the female-transmitted novel protein is not only present in mitochondria but also in the nuclear

membrane and in egg nucleoplasm (Breton et al. 2011a). These findings might support an involvement of these novel mitochondrial genes in some, still unknown, key biological functions in bivalve species with DUI. For instance, it has been suggested that the newly identified mtORFs in DUI bivalves might have a role in determining the fate of sperm mitochondria in fertilized eggs, maybe leading to the two distribution patterns of spermatozoon mitochondria observed in DUI early embryos: the aggregated pattern, in which these mitochondria form a cluster along the cleavage furrow in two-blastomere embryos and among blastomeres in four-cell embryos, and the dispersed pattern, in which sperm mitochondria are randomly scattered (Cao et al. 2004; Cogswell et al. 2006; Milani et al. 2011, 2012).

The analysis of 34 mitochondrial major Unassigned Regions (URs) of *M. senhousia* F- and M-mtDNA allowed us to verify the presence of novel mitochondrial ORFs in this species and to compare them with novel ORFs from other bivalve species with ascertained DUI, with other bivalves and with animals showing new mitochondrial elements. We found that many features are shared by all novel ORFs, allowing us to formulate an hypothesis on their possible shared origin.

## Materials and Methods

### Gametes Collection, DNA Extraction, PCRs, and Sequencing

*M. senhousia* specimens from Venice lagoon (Italy) were induced to spawn in sea water with oxygen peroxide, according to Morse et al. (1977). Each spawning was analyzed with a light microscope to sex specimens. Sperm and eggs were collected and then centrifuged at 3,000 × g; after that, sea water was removed and replaced with ethanol. Gametes were stored at −20 °C. Total DNA extraction from gametes of 11 females and 12 males was performed with DNeasy Tissue Kit (Qiagen) following manufacturer instructions. All polymerase chain reactions (PCRs) were executed on a 2720 Thermal Cycler (Applied Biosystems). All primers were provided by Invitrogen™ (see list of primers in [supplementary material S1, Supplementary Material](#) online).

Long PCRs, using gamete DNA extractions as template, were performed to obtain a segment containing the whole Largest Unassigned Region (LUR) (i.e., in both mtDNAs, the region between *rrnL* and *cob*); in the F-mtDNA, this region also contains the Female Unassigned Region 2 (FUR2) (see Passamonti et al. 2011 for annotation details). Primers for long-PCRs are the same used in Passamonti et al. (2011): M-mtDNA from sperm was amplified with primers M-16S103F and M-cob386R, whereas F-mtDNA from eggs with primers F-16S142F and F-cob383R ([supplementary material S1, Supplementary Material](#) online). Both segments were amplified with Herculase II Fusion Enzyme kit (Stratagene) in a 50 µl reaction volume composed of 10 µl

5× Herculase II Run Buffer, 0.5 µl of 100 mM dNTP mix, 1.25 µl of 10 µM primers, 0.5 µl of Herculase II Fusion DNA Polymerase, 5 µl of total DNA, and 31.5 µl of Nuclease-free water (Ambion Inc.). Long PCR cycles followed the same scheme for the M- and the F-mtDNA. The reactions started with an initial denaturation at 95 °C for 5 min, then 30 cycles of denaturation at 95 °C for 20 s, annealing at 48 °C for 20 s and extension at 68 °C for 10 s, then a final extension at 68 °C for 8 min.

Long PCR products were used as a template to amplify single overlapping segments of the LURs and the FUR2 with standard PCRs. Primers for standard PCRs ([supplementary material S1, Supplementary Material](#) online) were designed with Primer3 (Rozen and Skaletsky 2000) on the two complete *M. senhousia* F- and M-mtDNAs (GenBank accession nos. GU001953–4). GoTaq® Flexi Dna Polymerase (Promega) kit was used for standard PCRs. Reactions were performed in a 50 µl volume composed of 10 µl of 5× Green GoTaq Flexi Buffer, 6 µl of 25 mM MgCl<sub>2</sub>, 1 µl of 40 µM dNTP mix (10 µM each dNTP), 2.5 µl of 10 µM primers, 0.25 µl of GoTaq Dna Polymerase 5 U/µl, 4 µl of template DNA from the long PCRs, and 24 µl of Nuclease-free water (Ambion Inc.). LURs and FUR2 were amplified with the following cycle: initial denaturation at 95 °C for 2 min, 30 cycles of denaturation at 95 °C for 30 s, annealing at 48 °C for 30 s, extension at 72 °C for 90 s, and a final extension at 72 °C for 5 min.

All PCR products were purified with Wizard SV Gel and PCR clean-up System (Promega) kit, GenElute PCR clean-up kit, and GenElute Extraction kit (Sigma-Aldrich), following manufacturer instructions. Sequencing was performed at Macrogen Inc. (Seoul, South Korea). Sequences were assembled and aligned with MEGA5 (Tamura et al. 2011).

## Novel Mitochondrial ORFs

### Nucleotide Level: Sequence Conservation

We used ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf>, last accessed July 23, 2013) to assess the presence of novel ORFs in DUI species LURs present in GenBank, using the invertebrate mitochondrial genetic code. For DUI species, novel mitochondrial sex-specific ORFs were already described and confirmed in literature (*Mytilus* spp., *M. senhousia*: Breton et al. 2011b; *V. ellipsiformis*: Breton et al. 2009; *R. philippinarum*: Ghiselli et al. 2013). The obtained sequences of *M. senhousia* FUR2 and 689 annotated mt LURs of four *Mytilus* species (*Mytilus californianus*, *Myt. edulis*, *Myt. galloprovincialis*, and *Myt. trosulus*) (Bivalvia, Mytilidae) were checked to assess the conservation of the ORFs described in Passamonti et al. (2011) and Breton et al. (2011b) (last GenBank access: September 2012). The new sequences of *M. senhousia* LURs were also searched for the presence of novel ORFs (only the longest ORFs found in all sequences were considered). In the analyzed DUI species, we will refer to the ORFs present either in the F or the M

mtDNA (i.e., lineage-specific ORFs) as fORF and mORF, respectively. For comparison, ORFs were searched also in the LUR of the venerid *Paphia euglypta*, a species in which the presence of DUI has not been investigated yet (only one LUR sequence is available; table 1). Specific names are given to non-lineage-specific extra mtORFs, comprising mtORFs in non-DUI species. p-distances of novel ORFs of *M. senhousia* and other DUI species were calculated with MEGA5 (Tamura et al. 2011) using the bootstrap method on all suitable sequences available in GenBank.

### Protein Level: Structural and Functional Analysis

The above-mentioned ORFs were translated and analyzed at the amino acid level (see table 1 for the sequences in which the analyzed ORFs are included, and [supplementary material S2, Supplementary Material](#) online, for amino acid sequences). We will refer to the translations of fORFs and mORFs of DUI species as FORF and MORF, respectively.

To find Signal Peptides (SPs) we used Phobius (<http://phobius.sbc.su.se/>, last accessed July 23, 2013; Käll et al. 2004), InterProScan (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>, last accessed July 23, 2013; Zdobnov and Apweiler 2001), PrediSi (<http://www.predisi.de/>, last accessed July 23, 2013; Hiller et al. 2004), and SignalP 4.0 (<http://www.cbs.dtu.dk/services/SignalP/>, last accessed July 23, 2013; Petersen et al. 2011) softwares, while TMpred ([http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html), last accessed July 23, 2013; Hofmann and Stoffel 1993), Phobius (<http://phobius.sbc.su.se/>, last accessed July 23, 2013; Käll et al. 2004), InterProScan (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>, last accessed July 23, 2013; Zdobnov and Apweiler 2001), Prodiv-TMHMM (<http://topcons.cbr.su.se/>, last accessed July 23, 2013; Bernsel et al. 2009), and Rhythm (<http://proteininformatics.charite.de/rhythm/index.php?site=references>, last accessed July 23, 2013) were used to localize putative transmembrane helices (TM-helices). Atome 2 (<http://atome.cbs.cnrs.fr/AT2/meta.html>, last accessed July 23, 2013; Pons and Labesse 2009), I-Tasser (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>, last accessed July 23, 2013; Zhang 2008), and HHpred (<http://toolkit.tuebingen.mpg.de/hhpred>, last accessed July 23, 2013; Söding et al. 2005) were used to find similarities with known proteins and to find clues on the possible functions of the mtORFs. Alignments of the putative novel mitochondrial proteins were performed with PSI-COFFEE (<http://tcoffee.crg.cat/apps/tcoffee/do:psicoffee>, last accessed July 23, 2013; Di Tommaso et al. 2011).

Mitochondrial novel ORFs recently found in Cnidaria were included in the function analysis for comparison: two putatively active proteins, DNA polymerase beta (PolB) (*Alatina moseri*: Cnidaria, Cubozoa, Alatinidae) (Smith et al. 2011) and DNA mismatch repair protein (mtMutS) (*Incrustatus comauensis*: Cnidaria, Anthozoa, Clavulariidae) (McFadden and van Ofwegen 2013), and ORF-314 (*Pelagia noctiluca*:

**Table 1**

Sequences Used in the Analyses

Species	mt Genome	Accession Number	ORF
<i>Mollusca, Bivalvia</i> <i>Musculista senhousia</i>	F	GU001953	Mse-FORF, Mse-ORF-B
		KC243365–75	Mse-FORF
		KC243354–64	Mse-ORF-B
	M	GU001952	Mse-ORF-B
		KC243376–87	Mse-ORF-B
<i>Mytilus californianus</i>	F	AY515227	Mca-FORF
	M	AF188284	Mca-MORF1, Mca-MORF2
<i>Mytilus edulis</i>	F	AY350784	Med-FORF
	M	AY823623	Med-MORF
<i>Mytilus galloprovincialis</i>	F	AY497292	Mga-FORF
	M	HM027630	Mga-MORF
<i>Mytilus trossulus</i>	F	GU936625	Mtr-FORF
	M	AF188282	Mtr-MORF
<i>Ruditapes philippinarum</i>	F	AB065375	Rph-FORF
		KC243324–31	Rph-FORF
	M	AB065374	Rph-MORF
		KC243347–53	Rph-MORF
<i>Venustaconcha ellipsiformis</i>	F	FJ809753	Vel-FORF
	M	FJ809752	Vel-MORF
<i>Paphia euglypta</i>		GU269271	Peu-ORF
<i>Cnidaria</i> <i>Pelagia noctiluca</i> <i>Alatina moseri</i> <i>Incrustatus comauensis</i>		JN700949	Pno-ORF314
		YP_005353032.1	Amo-PolB
		AFU34533.1	Ico-mtMutS

NOTE.—Mitochondrial genome type is specified only for ascertained DUI species. ORF column is the name given to the amino acid sequence.

Cnidaria, Scyphozoa, Discomedusae) (Kayal et al. 2011) (supplementary material S2, Supplementary Material online). Last accession to databases was in September 2012. p-distances of amino acid sequences of each novel ORFs were calculated using the bootstrap method with MEGA5 (Tamura et al. 2011). Percentage of amino acid difference of novel proteins and of all mtDNA-encoded protein genes were calculated with MEGA5 (as in Breton et al. 2011a). For the *Myt. edulis* species complex (i.e., *Myt. edulis*, *Myt. Galloprovincialis*, and *Myt. trossulus*), pairwise sequence difference was first calculated for each gene and the results were then exported to Microsoft Excel for calculations of means and standard deviations (SDs).

## Results

### Novel Mitochondrial Open Reading Frames in Bivalves

The obtained *M. senhousia* LUR (FLUR of 11 females, 4,518–4,643 bp; MLUR of 12 males, 2,812–2,854 bp) and FUR2

(11 females, 542–543 bp) sequences were deposited in GenBank (FLUR accession nos.: KC243354–64; MLUR accession nos.: KC243376–87; FUR2 accession nos.: KC243365–75). The fORF, found in FUR2 on the heavy strand (as all standard coding genes) (fig. 1), is conserved in all samples (supplementary fig. S1, Supplementary Material online): its start and stop codons are always ATC and TAA, respectively, and its length is always 366 bp (121 aa). For nucleotidic p-distance see table 2. Another ORF, ORF-B, has been identified in MLUR and FLUR in the middle of Subunits B, on the reverse strand (fig. 1). In all males, ORF-B is always 318 bp long and its start and stop codons are ATG and TAA, respectively (supplementary fig. S2, Supplementary Material online). In females, Subunit B is duplicated (fig. 1) and ORF-B is not conserved as in males. The start codon is always ATG, and the stop codons can be TAA or TAG. Subunit B can contain one complete ORF-B (342–408 bp; supplementary fig. S2, Supplementary Material online) or two overlapping ORFs, together forming an ORF-B, due to a deletion of one T in a five-T string which breaks the frame. Two females showed only the version

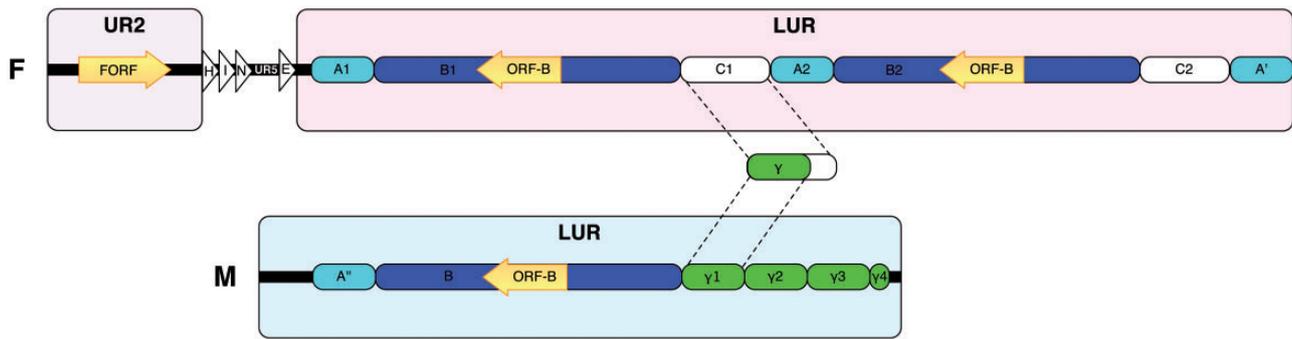


Fig. 1.—Largest Unassigned Regions (LURs). Schematic structure of female (F) and male (M) LURs of *Musculista senhousia*. Triangles indicate tRNAs.

Table 2

p-Distance (p-D) and Standard Error Values of Novel Mitochondrial ORFs in DUI Bivalves

Species	ORF	Nucleotide		Translation		N
		p-D	SE	p-D	SE	
<i>Musculista senhousia</i>	fORF	0.019	0.004	0.035	0.010	11
	Male ORF-B	0.004	0.002	0.008	0.004	12
	Female ORF-B <sup>a</sup>	0.024	0.005	0.056	0.012	8
	Overall ORF-B <sup>b</sup>	0.030	0.006	0.063	0.014	20
<i>Mytilus californianus</i>	fORF	0.005	0.003	0.014	0.008	4
	mORF1	0.015	0.009	0.031	0.021	4
	mORF2	0.011	0.007	0.033	0.022	4
<i>Mytilus edulis</i>	fORF	0.013	0.002	0.026	0.006	134
	mORF	0.017	0.004	0.039	0.012	25
<i>Mytilus galloprovincialis</i>	fORF	0.024	0.004	0.048	0.009	16
	mORF	0.029	0.008	0.062	0.021	47
	mORF (edulis-like) <sup>c</sup>	0.023	0.007	0.042	0.017	14
<i>Mytilus trossulus</i>	fORF	0.007	0.002	0.014	0.005	8
	mORF	0.025	0.007	0.046	0.016	9
<i>Ruditapes philippinarum</i>	fORF	0.009	0.003	0.011	0.006	8
	mORF	0.004	0.002	0.000	0.000	7
<i>Venustaconcha ellipsiformis</i>	fORF	0.000	0.000	0.000	0.000	3

NOTE.—Number of ORF sequences used for each species is dependant on the number of available and suitable sequences on GenBank. p-distances of *Myt. edulis*, *Myt. galloprovincialis*, and *Myt. trossulus* mORFs were calculated only on the last part of the ORF immediately following the poly-A sequence (see text for details). N = number of sequences used.

<sup>a</sup>Only complete female ORF-B were considered.

<sup>b</sup>Male ORF-B and complete female ORF-B were considered.

<sup>c</sup>mORF sequences matching *Myt. edulis* mORF.

with the two overlapping ORFs, never showing the complete ORF-B sequence (supplementary fig. S2, Supplementary Material online).

mt LURs of four *Mytilus* species (GenBank accession nos. in supplementary table S1, Supplementary Material online) were searched for the presence of the novel lineage-specific ORF described in Breton et al. (2011b): only the longest f- and mORFs were considered, as the shortest ones are often parts of them. A total of 201 *Mytilus* sequences containing complete ORFs were found (downloaded in September 2012): 197 fORFs and 17 mORFs. Many mORFs were found showing frame-disrupting mutations (supplementary table S1, Supplementary Material online). These alterations were more common in the first part of

the expected mORF in *Myt. edulis*, *Myt. galloprovincialis*, and *Myt. trossulus*, before and inside a long poly-A sequence (from 17 to 48 nucleotides), while the last part is usually conserved in comparison to the ORFs described in Breton et al. (2011b). p-distances of *Myt. edulis*, *Myt. galloprovincialis*, and *Myt. trossulus* mORFs, because of alignment issues, were calculated only on the part of the ORF following the poly-A sequence. As indicated by the p-distance analysis (table 2), *Mytilus* spp. fORFs are less variable than mORFs. In *R. philippinarum* the situation is the opposite, as mORF is more conserved than fORF. For *V. ellipsiformis* only three fORF sequences were available, but they show a remarkable conservation. An ORF was also found in the LUR of the venerid *P. euglypta*.

### Putative Novel Proteins from Bivalve Mitochondrial ORFs

Table 1 and [supplementary material S2, Supplementary Material](#) online, show sequences of the analyzed novel ORFs. A global alignment including all the analyzed amino acid sequences was not possible due to their divergence ([supplementary fig. S3, Supplementary Material](#) online), but groups with some similarities were found. Mse-ORF-B translation has practically the same amino acid sequence in the two genomes ([supplementary fig. S4, Supplementary Material](#) online). Mytilid FORFs are largely similar among each other ([fig. 2A](#)), most of all those of *Myt. edulis* complex (Med-, Mga-, and Mtr-FORFs) ([supplementary fig. S5, Supplementary Material](#) online). With the only exception of Mca-MORFs, *Mytilus* MORFs are also highly similar ([fig. 2B](#); [supplementary fig. S6, Supplementary Material](#) online), and show a characteristic string of lysines (poly-K region) of variable length (8–12 aa; translation of a poly-A nucleotide sequence), absent from MORFs of other species and from FORFs. Downstream the poly-K region, *Mytilus* MORFs show a high similarity among each other, whereas in their N-terminus they are quite variable ([supplementary fig. S6, Supplementary Material](#) online). Although *Mytilus* FORFs and MORFs appear different between each other (see for example *Myt. edulis*, [fig. 3A](#)), Rph-FORF and MORF show several shared domains ([fig. 3B](#)), and also Vel-FORF and MORF have a big domain in their N-terminal showing similarity ([supplementary fig. S7, Supplementary Material](#) online).

Shared domains among the novel putative proteins are boxed in [figure 4](#). Amino acid p-distances are reported in [table 2](#). A common feature of all ORFs amino acid sequences (with the exception of *R. philippinarum* MORF and *V. ellipsiformis* FORF) is their major p-distance value in respect to their own nucleotidic sequences: this indicates that non-synonymous mutations are more common than synonymous mutations. The variability of FORFs and MORFs was confirmed by the amino acid sequence difference analysis of all mtDNA-encoded protein genes ([fig. 5](#)). Our findings, together with previous studies (Breton et al. 2009, 2011a), showed that lineage-specific mitochondrial proteins are among the fastest evolving proteins coded by the mtDNA of the analyzed species.

A SP was found in the N-terminus of all FORFs ([table 3](#)). Among the TM-helices, the N-terminal helix coincides with the SP sequence ([table 3](#)). Besides this helix, one more TM-helix supported by at least two programs was found in Mga-FORF, in Mtr-FORF, and in Rph-FORF ([table 3](#)). A sound SP was not always found in MORFs, even if some softwares point to the same SP sequence with a low score ([table 3](#)). Also in this case, the N-terminal TM-helices coincide with the SP sequence. Other probable TM-helices detected by at least two of the softwares were found in Mse-ORF-B, Med-MORF, and in Rph-MORF ([table 3](#)).

### Novel Mitochondrial ORFs: Function Prediction

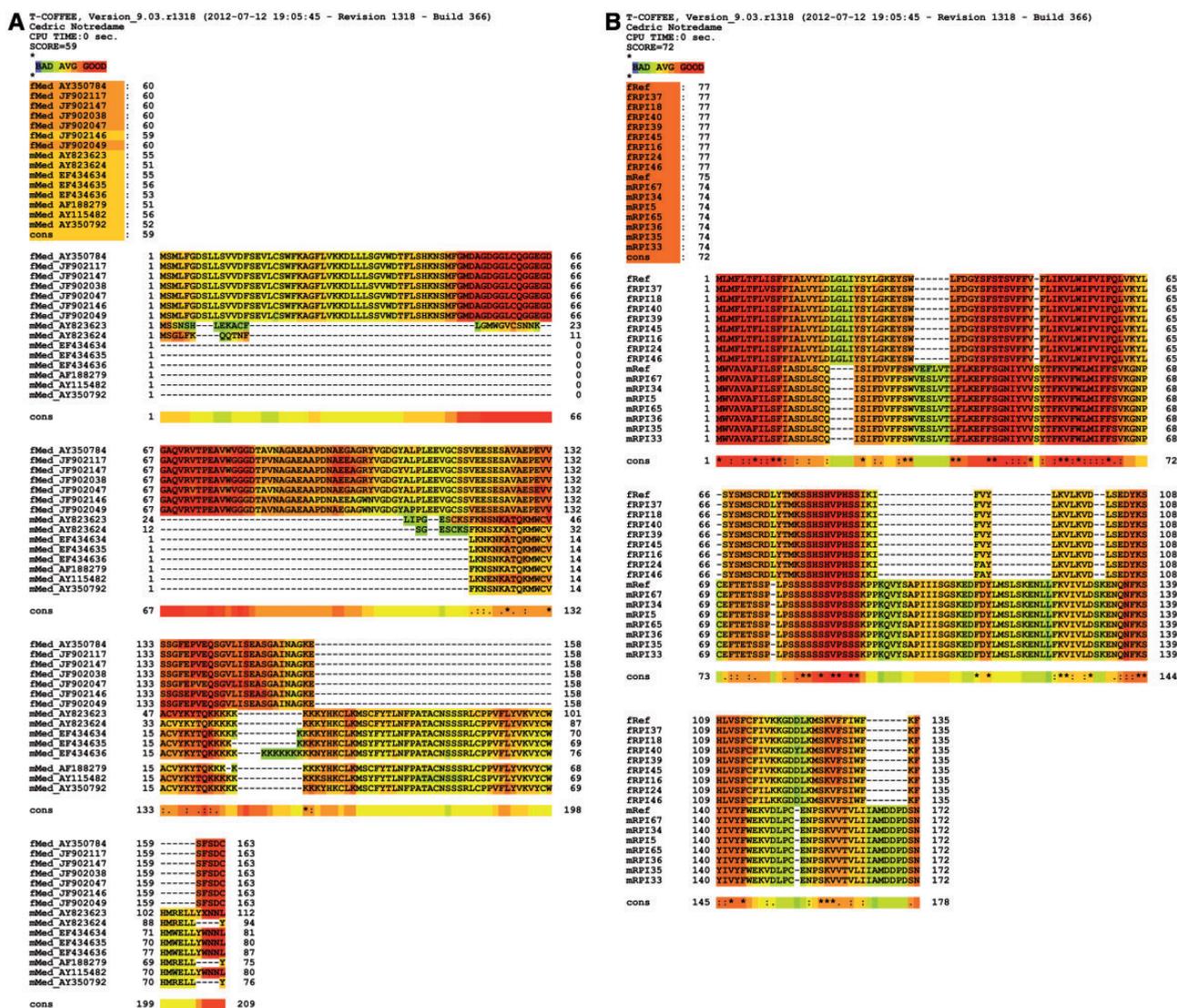
Atome 2, I-Tasser, and HHpred found domains similar, in structure or ligands, to known proteins, in both FORFs ([tables 4, 5](#) and [supplementary tables S2–S8, Supplementary Material](#) online) and MORFs ([tables 4, 5](#) and [supplementary tables S9–S16, Supplementary Material](#) online). FORF highest probability hits include proteins involved in nucleic acid binding and transcription (e.g., helicase/hydrolase, transcription factors), in some cases with specific aspects of nucleic acid processing, like RNA modification (e.g., Med-FORF and Vel-FORF), and methylation (e.g., Mtr-FORF). Other hits are proteins with a membrane association, for example involved in transport across membrane, in cell adhesion, but also receptors, most of all involved in hormone signalling. Many proteins point to a role in immune response, for example in cytokine release for immune system activation (e.g., Mca-FORF).

MORF hits with the highest probability include membrane-associated proteins with a role in nucleic acid binding and transcription, mainly related to signalling for cell differentiation and development (e.g., embryonic development). Some ORFs appear to be involved in DNA recombination and repair, in transposition regulation, and DNA integration of foreign elements (e.g., Mca-MORF1 and Rph-MORF). Moreover, several hits are proteins that regulate cytoskeleton formation and dynamics, from cell polarity regulation to cell proliferation. Other hits point to a role in ubiquitination and apoptosis with high probability (e.g., Mca-MORF1, Med-MORF, and Rph-MORF). Finally, many of the proteins have a role in immune response, for example in cytokine release (e.g., Mca-MORF2 and Med-MORF).

We found similar hits in Peu-ORF and Pno-ORF314 ([tables 4](#) and [5](#) and [supplementary tables S17](#) and [S18, Supplementary Material](#) online), connected with nucleic acid binding and transcription, with membrane association (Pno-ORF314), with signalling for cell differentiation during embryogenesis, with foreign elements (mobile genetic element and viral proteins), and with immune response regulation (Pno-ORF314).

All the hits come from different animal and plant proteins, from both unicellular and pluricellular organisms. The position of the most represented functional domains is reported in [figure 5](#) (see also [table 1](#) for acronyms). On the whole, with the only exception of Mtr-MORF and Vel-MORF, every analyzed protein showed hits referred to viral proteins ([table 5](#) and [fig. 5](#)). In some cases (Mse-FORF, Mca-FORF, Mse-ORF-B, and Rph-MORF) the similarity with viral proteins was confirmed by all the three softwares used, in other cases (Mtr-FORF, Mca-MORF, Med-MORF, and Mga-MORF) by two of the softwares, and for the remaining proteins (Med-FORF, Mga-FORF, Rph-FORF, and Vel-FORF) by one program. Moreover, the same first four hits found by HHpred are present in all the novel putative proteins analyzed ([supplementary table S19, Supplementary Material](#) online), except for Amo-PolB, which showed complete homology with base-excision repair DNA





**Fig. 3.—**(A) PSI-Coffee alignment of *Mytilus edulis* FOR and MORF (accession nos. of sequences containing the ORF are reported in the figure); (B) PSI-Coffee alignment of *Ruditapes philippinarum* FOR (accession nos. of entire FLURs: KC243324–31) and MORF (accession nos. of entire MUR21 sequences: KC243347–53).

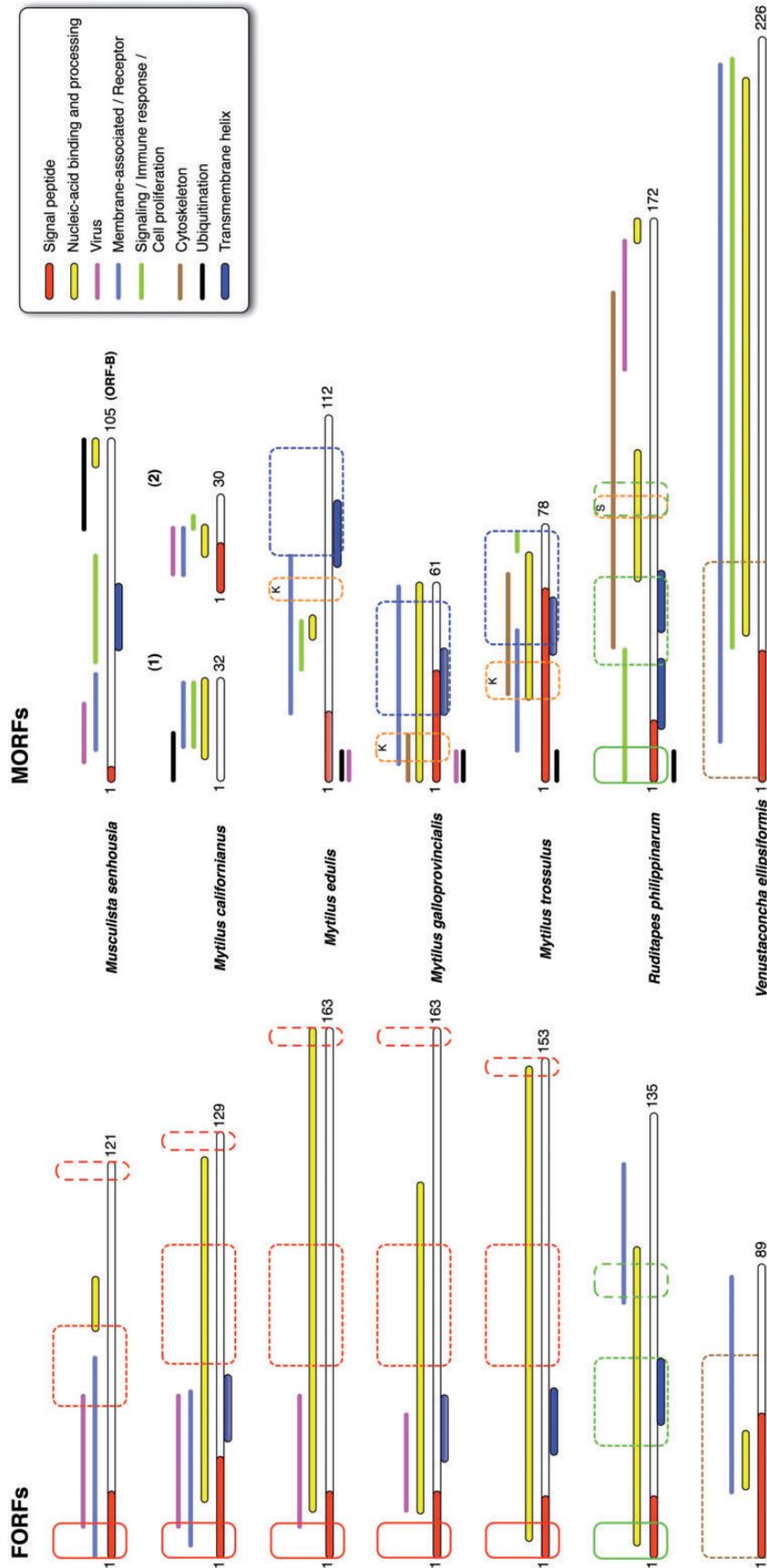
polymerases, mainly polymerase beta (HHpred probability: 100.0), and Ico-mtMutS, which showed a complete homology with a DNA mismatch repair protein (HHpred probability: 100.0), in both cases with hits from many organisms.

**Discussion**

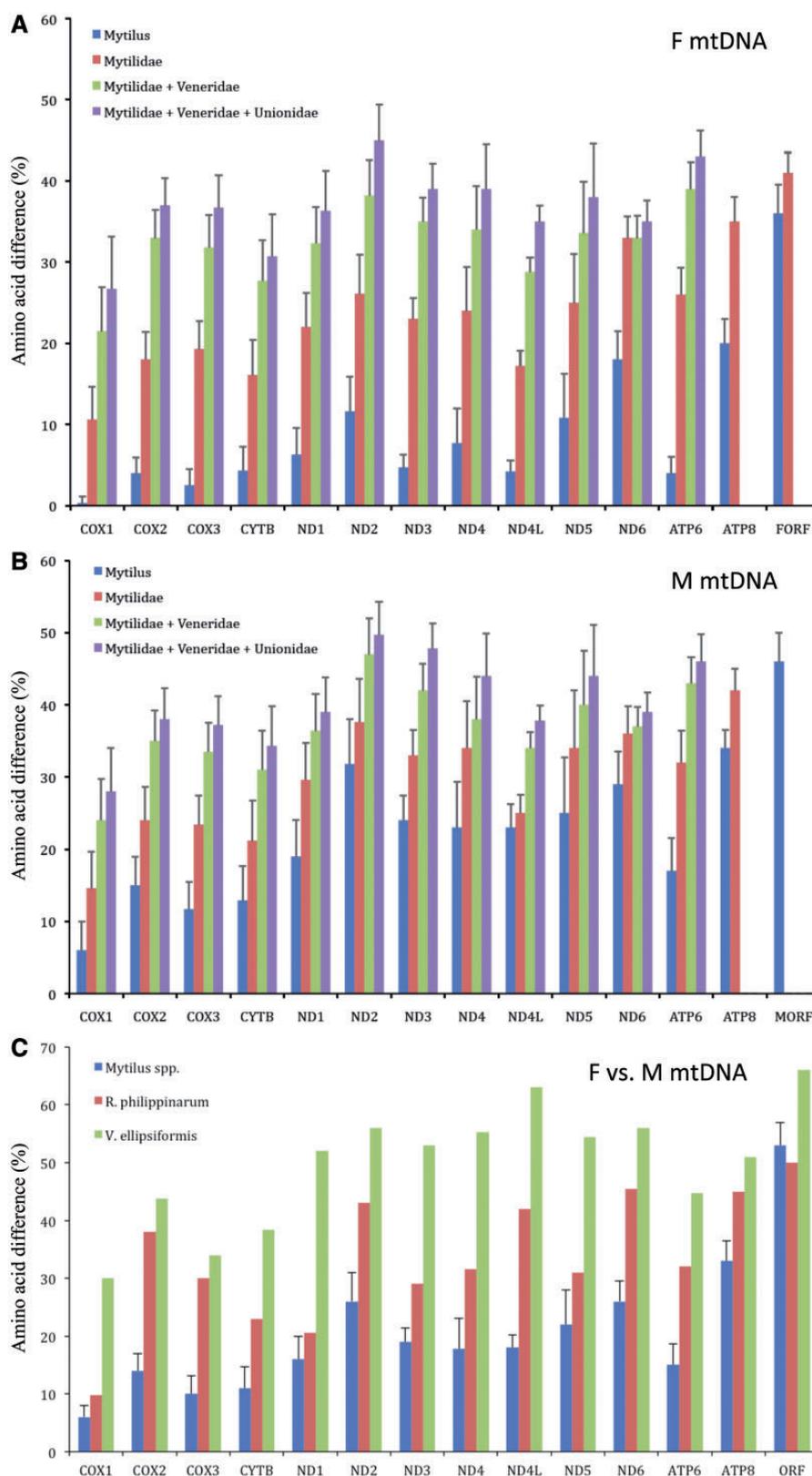
**Novel ORFs Characterization**

As mentioned, mt genomes of bivalve species with DUI have novel lineage-specific ORFs of unknown origin and function. Generally, homologous proteins, or their fragments, have similar structure because structures diverge much more slowly than their sequences (Chothia and Lesk 1986). Depending on the degree of divergence between them, homologous

proteins may also maintain similar cellular function, ligands, protein interactions partners, or enzymatic mechanisms (Todd et al. 2001). Because bivalve novel ORFs do not have known homologous (i.e., they are ORFans; Fischer and Eisenberg 1999), we performed multiple analyses of their structure, in order to infer the function. These ORFs are found in extra-genic regions, often inside the LUR. Except for *M. senhousia* ORF-B, that is found in both mt genomes (in the middle of LUR Subunit B), the other analyzed ORFs are lineage-specific. ORF-B nucleotide sequence is extremely conserved between the two mt genomes (supplementary fig. S2, Supplementary Material online), but considering that in some *M. senhousia* females the complete ORF-B is absent, ORF-B might not be functional in females.



**Fig. 4.**—Functional domains in FORFs and MORFs (position in the amino acid sequence as identified by HHpred). Sequences with similarities are boxed in the same color and with the same type of line; red: similarities among FORFs; blue: similarities among MORFs; orange: K = poly-K region; S = poly-S region (see also PSI-Coffee alignments, figs. 2 and 3 and [supplementary tables S2–S18, Supplementary Material online](#)). Numbers indicate sequence length.



**FIG. 5.**—Percentage of amino acid difference of novel proteins and of all mtDNA-encoded protein genes. Amino acid divergence (% amino acid difference) was calculated with MEGA5 for each mt protein coding gene among: (A) F mt genomes [for (i) *Mytilus* spp.; (ii) Mytilidae, i.e., *Mytilus* spp.

(continued)

Lineage-specific mitochondrial ORFs were found in all the analyzed DUI species (table 1; [supplementary material S2, Supplementary Material](#) online). In *Mytilus* male genomes, the last part of the mORFs, after the poly-A region, is the most conserved (fig. 4). A number of mORFs found in sequences annotated as *Myt. galloprovincialis* are identical to *Myt. edulis* mORF, and probably derive from hybridization that is extremely common inside the *Myt. edulis* complex: these “*edulis*-like” mORFs are more conserved than *Myt. galloprovincialis* own mORF and fORF, but are more diverse than *Myt. edulis* own mORF, from which they seem to derive (table 2). Nonetheless, *Myt. edulis* complex mORFs could be the same element, considering the extreme conservation of most of their sequence. Instead, *M. californianus* has two largely overlapping putative mORFs that do not contain a poly-A sequence like the other three species and are completely diverse from them. This is not surprising given the high divergence between *Myt. edulis* complex and *M. californianus* mitochondrial genomes (Zouros 2012).

Putative TM-helices were not found in all the analyzed proteins. In some cases the same region was identified as SP (table 3): being SP a peptide chain of hydrophobic amino acids, it can be difficult for softwares to discern it from a TM-helix (Käll et al. 2004). A clue in favour of a membrane association of MORFs comes from the poly-lysine (Med-, Mga-, and Mtr-MORF) and poly-serine (Rph-MORF) regions. Poly-lysine motif is required for membrane lipid binding (Bouaouina et al. 2012), and poly-serine domains characterize proteins anchored to bacterial outer membrane (Howard et al. 2004). Being mitochondria derived from alpha-proteobacteria (Andersson et al. 1998), we can hypothesize a similar membrane association in these organelles. Interestingly, the first four hits found with HHpred are the same for both FORFs and MORFs of DUI bivalves ([supplementary table S19, Supplementary Material](#) online), and for Peu-ORF and Pro-ORF314. Two of these hits are involved in the anchor to cell membrane/surface (LPXTG-motif cell wall anchor domain and outer membrane insertion C-terminal signal); the other two are typical of proteins involved in transcription (X-X-X-Leu-X-X-Gly heptad repeats) and in post-transcriptional processes (pentatricopeptide repeats, PPR). The detected motifs are

not long enough to claim a functional homology, but their involvement in membrane binding and in transcription is sustained also by other hits (see tables 4 and 5; [supplementary tables S2–S16, Supplementary Material](#) online).

The existence of Vel-FORF and MORF was shown by western blot analysis (Breton et al. 2009), and Vel-FORF was shown to be present in mitochondria and in the nuclear membrane (Breton et al. 2011a). Likely, these novel mitochondrial proteins have a role in different cellular compartments, thus including domains that allow them to interact with several substrates such as membranes, cytoskeleton, and nucleic acids. It is important to investigate the existence of ORF translation products in other DUI species. We are performing these kind of analyses and first data confirm the existence of Rph-MORF protein (Milani et al. in preparation). Furthermore, increasing the number of analyzed DUI species and sequences may help in explaining the evolutionary dynamics that led to the highest similarity found between FORF and MORF of some species (i.e., Rph-FORF/MORF and Vel-FORF/MORF) in comparison to other species (i.e., *Myt. edulis* complex) (see alignments and fig. 4).

The similarity region between an ORF and a known protein sometimes includes a large part of the protein, even with high probability (see for example Vel-MORF), in other cases, as said before, it is found in short amino acid sequences. In such cases we are confident we retrieved sound similarities, because the same homolog proteins from very distant taxa, from both unicellular and pluricellular organisms, are present among the hits ([supplementary tables S2–S16, Supplementary Material](#) online).

Overall, the analyzed ORFs show many common functions (see [supplementary tables S2–S16, Supplementary Material](#) online), but, when we consider only hits with the highest scores, FORFs are more similar among each other than with MORFs, and vice-versa (tables 4 and 5). FORFs appear to be involved in transcription regulation and in immune response, also linked to cell adhesion, migration, and proliferation. MORFs appear to have a main role in cytoskeleton organization (cell differentiation during embryonic development), but also capable, as FORFs, of nucleic acid binding and transcription regulation. FORFs and MORFs appear to share a role as

#### FIG. 5.—Continued

and *Musculista senhousia*; (iii) Mytilidae + the venerid *Ruditapes philippinarum*; and (iv) Mytilidae + the venerid *R. philippinarum* + the unionoid *Venustaconcha ellipsiformis*], (B) M mt genomes [for (i) *Mytilus* spp.; (ii) Mytilidae, i.e., *Mytilus* spp. and *M. senhousia*; (iii) Mytilidae + the venerid *R. philippinarum*; and (iv) Mytilidae + the venerid *R. philippinarum* + the unionoid *V. ellipsiformis*], and (C) between F and M mt genomes [for (i) *Mytilus* spp.; (ii) *R. philippinarum*; and (iii) *V. ellipsiformis*]. For the *Mytilus edulis* species complex (i.e., *Myt. edulis*, *Myt. galloprovincialis*, and *Myt. trossulus*), pairwise sequence difference was first calculated for each gene and the results were then exported to Microsoft Excel for calculations of means and SDs. For both *R. philippinarum* and *V. ellipsiformis* only, one whole F mtDNA and one whole M mtDNA are present in database and no error can be calculated. Omitted comparisons are due to the impossibility to obtain a good alignment. NOTE: F mtDNA = female mitochondrial genome; M mtDNA = male mitochondrial genome. *Mytilus* spp. = *Myt. edulis* species complex. Accession nos. mitochondrial genomes (F-type and M-type mtDNA, respectively): *Myt. edulis* NC\_006161 and AY823623; *Myt. galloprovincialis* NC\_006886 and AY363687; *Myt. trossulus* DQ198231 and DQ198225; *M. senhousia* GU001953 and GU001954; *R. philippinarum* AB065375.1 and AB065374.1; *V. ellipsiformis* FJ809753 and FJ809752.

**Table 3**  
Signal Peptide and Transmembrane-Helix Prediction in the Novel Putative Proteins

Signal Peptide							
FORF	Mse	Mca	Med	Mga	Mtr	Rph	Vel
Software							
Phobius	1–20	—	1–20*	1–20*	1–18*	1–18	—
InterProScan	1–20	1–31	—	—	1–18	1–18	1–44
PrediSi	1–28	—	1–20*	1–20*	1–18*	1–18	1–44*
SignalP 4.0	1–20	1–20*	1–20*	1–20*	1–18*	—	1–44*
MORF	Mse <sup>ORFB</sup>	Mca	Med	Mga	Mtr	Rph	Vel
Software							
Phobius	—	- (1–13)	—	1–34	—	1–18	—
InterProScan	1–5	- (1–18)	—	—	—	1–18	1–40
PrediSi	—	- (1–16)*	1–22*	1–34*	1–59	1–17	1–40
SignalP 4.0	1–6*	- (1–14)	1–21*	1–34*	1–59*	1–18	1–40*
Transmembrane Helices							
FORF	Mse	Mca	Med	Mga	Mtr	Rph	Vel
Software							
TMpred	4–23	3–25	8–29	8–29	31–52	1–18/40–59	21–42
Phobius	—	—	—	—	—	7–27/39–62	21–42
InterProScan	—	—	—	—	—	5–23/42–62	21–41
Prodiv-TMHMM	5–27	5–25/35–55	9–29	5–25/28–48	26–47	3–23/39–59	21–42
Rhythm	6–23	4–23	—	18–37	33–50	5–27/40–62	21–42
MORF	Mse <sup>ORFB</sup>	Mca	Med	Mga	Mtr	Rph	Vel
Software							
TMpred	40–61	- (-)	69–96**	19–35	41–57**	1–23/16–38/46–64	21–39
Phobius	—	- (-)	—	—	38–56	42–62	20–38
InterProScan	—	- (-)	—	20–38	38–56	5–27/41–61	21–41
Prodiv-TMHMM	41–61	- (-)	65–86	21–41	38–59	3–23/44–64	—
Rhythm	—	- (-)	—	17–34	41–57	20–37/46–64	21–38

NOTE.—Signal peptide: Only signal peptides statistically supported (Phobius posterior label probability > 0.5; PrediSi score > 0.5; SignalP score > D-cutoff 0.5; significance test not provided by InterProScan) or found at least by two softwares are shown; \*Significance < 0.5; (n) = Mca-MORF2 results. Transmembrane helices: Only transmembrane helices considered significant (TMpred score > 500; Phobius posterior label probability > 0.5; significance test not provided by the other softwares) or found by at least two softwares are shown; \*\*TMpred score < 500; values in bold indicate helices not overlapping with the predicted signal peptide; (n) = Mca-MORF2 results.

signalling molecules, more specifically involved in hormone signalling and immune response regulation. Interestingly, some MORFs show similarity with DNA replication, recombination, and repair proteins (see for example the transposition regulation and DNA binding-integration hits of Mca-MORF1). Moreover, hits of ubiquitination and apoptosis regulation proteins are found in almost all ORFs (supplementary tables S2–S16, Supplementary Material online).

### Are Novel Mitochondrial ORFans of Viral Origin?

The sequences analyzed in this article do not show homologies with any known mitochondrial protein, therefore they unlikely originated from recent duplication events, as instead happened for *nad2* in *Crassostrea* (Wu et al. 2012), for *cox2* in

*R. philippinarum* F-mtDNA (Okazaki M and Ueshima R, unpublished data), and in *M. senhousia* M-mtDNA (Passamonti et al. 2011). Another origin should be taken into account for these proteins and the observed hits to viral proteins provide a possible working hypothesis: bivalve ORFs could have arisen from different events of insertion, thus showing a narrow distribution similar to other ORFans (Yu and Stoltzfus 2012).

The analyzed ORFs show a higher amino acid substitution rate than the typical mitochondrial coding genes (fig. 5). Lineage-specific genes evolve at a faster rate than broadly distributed genes, in both bacteria and eukaryotes (Daubin and Ochman 2004a, 2004b; Yu and Stoltzfus 2012). One reason could be that lineage-specific genes participate more in lineage-specific adaptation, therefore evolving faster

**Table 4**

Function Analysis of Novel Mitochondrial ORFs

Mse-FORF	Mca-FORF
<p><b>Hormone receptor/Cell adhesion, migration, proliferation/Immune response</b></p> <p>Atome2 (highest probability):            Chemokine (13), highest score 75.17            Human tissue factor, score 70.69            Eotaxin (2), score 67.89 and 62.51            Erythrocyte binding antigen 175, score 54.15</p> <p>I-Tasser (confirmation):            Cell division protein kinase 9/Protein Tat, Z-score 0.79            RhoGAP protein, Z-score 0.90</p> <p>Glypican-1, Z-score 0.62            Small-inducible cytokine A13, Z-score 0.91            Erythrocyte binding antigen 175, Z-score 0.63</p> <p>HHpred (confirmation):            SARS receptor-binding domain-like, probability 54.78, aa 31–61            Small inducible cytokine A1 precursor, probability 27.01, aa 5–91</p> <p><b>Protein binding/transport</b></p> <p>I-Tasser (highest probability):            Exportin-5, Z-score 0.75            Cullin-5, Z-score 0.69            Nucleoporin NUP170, Z-score 0.84            BRO1 protein, TM-score &gt; 0.5            GTP-binding nuclear protein Ran, TM-score &gt; 0.5</p> <p><b>Nucleic acid binding</b></p> <p>I-Tasser (highest probability):            Telomeric repeat-binding factor (2), TM-score &gt; 0.5            ATP-dependent RNA helicase (2), TM-score &gt; 0.5</p> <p><b>Membrane association</b></p> <p>HHpred (highest probability):            More than 40 hits, highest probability 75.84, aa 1–21</p> <p><b>Transcription factor translocator</b></p> <p>HHpred (highest probability):            Glucocorticoid receptor-like (10), highest probability 64.44, aa 68–85</p>	<p><b>Transport across membrane/Receptor/Immune response</b></p> <p>Atome2 (highest probability):            Unique short US2 glycoprotein, score 82.16            Killer cell immunoglobulin-like receptor 2DL1 (2), highest score 75.39            Pertussis toxin subunit 5, score 55.44            Putative ABC type-2 transporter, score 54.92</p> <p>I-Tasser (confirmation):            Receptor-type adenylate cyclase (2), TM-score &gt; 0.5</p> <p>HHpred (confirmation):            RAB6-interacting protein 2 (2), highest probability 62.09, aa 39–99            Integral membrane protein, probability 48.70, aa 10–38            TonB Periplasmic protein TonB, probability 45.44, aa 44–75            NIPSNAP, probability 35.39, aa 12–34            Membrane protein, probability 31.83, aa 54–66            Membrane or secreted protein, probability 30.95, aa 3–51            Transport protein Sec24A (2), probability 30.88, aa 3–51            Membrane protein containing DUF1112, probability 28.02, aa 14–66</p> <p><b>Cell adhesion and migration/Hormone receptor</b></p> <p>Atome2 (highest probability):            Fibronectin, score 70.61            PfEMP1 variant 2 of strain MC, score 58.80            Human tissue factor (2), highest score 52.67</p> <p><b>Helicase activity/Replication/Immune response</b></p> <p>I-Tasser (highest probability):            Antiviral helicase SKI2, Z-score 0.68            Proliferating cell nuclear antigen PcnA, Z-score 0.85            Infectivity protein G3P, Z-score 0.62            Cyclophilin-like domain, Z-score 0.59</p> <p>HHpred (confirmation):            SKI2/RNA helicase, probability 50.91, aa 101–123            Peptidyl-prolyl isomerase G/cyclophilin G, probability 38.59, aa 17–69</p> <p><b>Cytoskeleton/Cytokine release/Immune system activation</b></p> <p>HHpred (highest probability):            Keratin (9 hits), highest probability 76.37, aa 25–95</p> <p><b>Transcription regulator</b></p> <p>HHpred (highest probability):            Sterol regulatory element binding protein (2), highest probability 71.01, aa 17–66            CG17964-PH, isoform H, probability 28.03, aa 54–122</p>
<p><b>Med-FORF</b></p> <p><b>DNA binding and replication</b></p> <p>Atome2 (highest probability):            Uncharacterized protein AF_1548, score 85.31            Exotoxin A, score 74.29            Minichromosome maintenance protein, score 63.74</p> <p>I-Tasser (highest probability):            Minichromosome maintenance protein (3), highest Z-score 1.64, TM-score &gt; 0.5            ATPase involved in replication control (3), highest Z-score 1.31, TM-score &gt; 0.5            P97 (Cell division cycle), TM-score &gt; 0.5</p> <p>HHpred (confirmation):            Zinc fingers (2), highest probability 35.41, aa 13–32 and 36–42</p>	<p><b>Mga-FORF</b></p> <p><b>DNA binding and replication</b></p> <p>Atome2 (highest probability):            Uncharacterized protein AF_1548, score 82.33            Exotoxin A, score 72.53            Minichromosome maintenance protein, score 62.01</p> <p>I-Tasser (highest probability):            Minichromosome maintenance protein (2), highest Z-score 1.64, TM-score &gt; 0.5            ATPase involved in replication control (3), highest Z-score 1.21, TM-score &gt; 0.5</p> <p>HHpred (confirmation):            Zinc fingers (2), highest probability 36.31, aa 13–32, 36–42            NPH4/transcription factor, probability 30.20, aa 31–114</p>

(continued)

Table 4 Continued

Med-FORF	Mga-FORF
<p><b>Development/Growth hormone receptor/Cell adhesion</b> Atome2: Nicotinamidase, score 59.05 Human tissue factor (2 hits), highest score 52.26 Fibronectin, score 51.93</p> <p><b>Lyase/Hydrolase activity</b> HHpred (highest probability): Cyanase C-terminal domain (2), highest probability 54.52, aa 5–16</p> <p><b>Immune response/RNA binding and processing</b> HHpred (highest probability): Cyclophilin/Peptidylprolyl isomerase (13), highest probability 46.01, aa 59–163</p> <p><b>Cell adhesion/Lipid metabolism</b> HHpred (highest probability): GYF domain (2 hits), highest probability 39.73, aa 16–28 Malonyl-CoA decarboxylase (4), highest probability 36.39, aa 14–29</p>	<p><b>Development/Growth hormone receptor/Cell adhesion</b> Atome2: Human tissue factor, score 56.15 Fibronectin, score 52.65 Nicotinamidase, score 44.75 Tudor domain-containing protein 5 (Germ line integrity), score 43.50</p> <p><b>Lyase/Hydrolase activity</b> HHpred (highest probability): Cyanase C-terminal domain (2), highest probability 55.41, aa 5–16</p> <p><b>Lipid metabolism/Cell adhesion</b> HHpred (highest probability): Malonyl-CoA decarboxylase (6), highest probability 39.89, aa 14–29 GYF domain (3 hits), highest probability 39.13, aa 16–28</p>
Mtr-FORF	Rph-FORF
<p><b>Ligase activity</b> Atome2 (highest probability): D-alanine—poly(phosphoribitol) ligase subunit 1 (3), highest score 80.77</p> <p><b>Lipid metabolism</b> Atome2 (highest probability): Acetyl-coenzyme A synthetase (3), highest score 79.06</p> <p><b>Receptor/Membrane-associated protein/Immune response</b> Atome2: Unique short US2 glycoprotein, score 65.52 Interleukin 18 binding protein/Cytokine, score 50.72</p> <p>I-Tasser (confirmation): Gramicidin synthetase 1, Z-score 2.25 D-alanine—poly(phosphoribitol) ligase subunit 1, Z-score 2.12</p> <p><b>Cytoskeleton-associated protein</b> I-Tasser (highest probability): Kinesin-like protein Nod, TM-score &gt; 0.5 Tubulin (3), TM-score &gt; 0.5 Integrin alpha-X, TM-score 0.498</p> <p>HHpred (confirmation): Actin-like ATPase domain (2), highest probability 45.12, aa 49–57</p> <p><b>Methylation (DNA, RNA, protein)</b> HHpred (highest probability): More than 20 hits (21), highest probability 61.81, aa 5–149</p> <p><b>Immune response/Viral infection cofactor</b> (large region) Cyclophilin, probability 26.70, aa 15–110</p>	<p><b>Nuclear transport</b> Atome2 (highest probability): Nuclear transport factor 2 (2), highest score 85.83 NTF2-related export protein 1, score 79.17</p> <p>I-Tasser (highest probability): Nuclear transport factor 2 (3), highest Z-score 0.72, TM-score &gt; 0.5 p15 (Export of mRNAs through nuclear pore complexes) (2), TM-score &gt; 0.5 Nuclear RNA export factor 2, TM-score &gt; 0.5 mRNA transport regulator Mtr2, TM-score &gt; 0.5 Rasputin (Similar to nuclear transport factor 2), TM-score &gt; 0.5</p> <p>HHpred (confirmation): DNA double-strand break repair transporter domain, probability 49.97, aa 77–88</p> <p><b>DNA replication/Transcription/Nucleic-acid binding</b> HHpred (highest probability): 20 hits Highest probability 86.90, with HemY family protein, aa 3–67 Zinc fingers, probability 86.88</p> <p>Atome2 (confirmation): Polymerase PB2, score 56.13 Restriction endonuclease Hpy99I, score 52.92 Cyclin (3), highest score 52.40 DNA gyrase inhibitor YacG, score 50.57</p> <p><b>Transport across membrane/Amino-acid transporter</b> HHpred: About 30 hits, highest probability 86.47, aa 2–75</p> <p><b>Receptor site</b> HHpred: Neurotoxin type G, probability 63.95, aa 77–120</p> <p><b>Membrane-associated protein/Immune response</b> HHpred: Macoilin/transmembrane protein 57 (2), probability 50.56, aa 1–113 LysM domain, probability 33.34, aa 115–123</p> <p>Atome2 (confirmation): HLA class II histocompatibility antigen, score 47.74</p>

(continued)

Table 4 Continued

**Vel-FORF****Nuclear proteins/Nuclear transport/RNA processing**

Atome2 (highest probability):

- Poly(A) polymerase, score 84.27
- Ran GTPase-activating protein 1, score 60.97
- Chimera of Histone H2B.1 and Histone H2A.Z, score 49.66

I-Tasser (confirmation):

- VP1/mRNA-capping machine (2), highest Z-score 0.82
- Poly(A) polymerase (2), highest Z-score 0.70
- ATP-dependent DNA helicase RecG-related protein, Z-score 0.71

**DNA binding/Transcription**

Atome2 (highest probability):

- Bifunctional protein GlmU, score 58.95
- Serine/threonine-protein phosphatase (2), highest score 58.24
- SAGA-associated factor 73, 21.79

HHpred (highest probability):

- ComGC (2), highest probability 94.43, aa 2–38
- CG13581-PA transcription factor, probability 39.77, aa 77–89

**Membrane-associated proteins**

Atome2 (highest probability):

- Bactericidal permeability-increasing protein, score 53.65
- Photosystem II reaction center protein I, score 31.82

HHpred (confirmation):

- More than 10 hits in the N-terminus of the sequence

**Hormone receptor/Transcription**

I-Tasser (highest probability):

- Progesterone receptor ligand-binding domain, TM-score > 0.5
- Androgen receptor ligand-binding domain, TM-score > 0.5
- AncCR, TM-score > 0.5
- Mineralocorticoid receptor (nuclear receptor), TM-score > 0.5

**Immune system/Transport across membrane**

HHpred:

- C-type LECTin family member (clec-35) (7), highest probability 78.84, aa 19–86

**Mca-MORF1****Mca-MORF2****Transposition regulation/DNA binding and integration/Transcription**

Atome2 (highest probability):

- Transposase (3), highest score 76.46
- Protein RDM1/RNA-directed DNA methylation, score 65.30
- Modification methylase TaqI, score 55.99
- Nuclear factor NF-kappa-B p100 subunit, score 55.42
- Replication termination protein, score 54.13
- DNA-binding protein RAP1, score 50.90

I-Tasser (confirmation):

- C25G10.02, chromosome I (Hydrolase/DNA duplexes separation), Z-scores > 1
- Rad50 (Hydrolase/DNA-double strand break repair), Z-scores > 1
- Replication factor c small subunit, TM-scores > 0.5
- O-sialoglycoprotein endopeptidase/protein kinase (Hydrolase), TM-scores > 0.5

HHpred (highest probability):

- “Winged helix” DNA-binding domain (2), highest probability 88.51, aa 7–25

**Protein folding**

Atome2 (highest probability):

- Huwentoxin-II, score 79.45
- Alanine racemes, score 76.95
- Heat shock 70 kDa protein 8/Chaperone (2), highest score 55.37
- BAG-family molecular chaperone regulator-1, score 47.88

**Cytokine/Immune response/Cell proliferation/Embryonic development**

Atome2 (highest probability):

- Interleukin-6 receptor subunit beta, score 71.60
- Interleukin-1 beta, score 40.82
- Erythropoietin receptor, score 54.98
- Tumor necrosis factor ligand superfamily member 13, score 50.90
- Natural killer cell activating receptor, score 46.35
- Myeloid antimicrobial peptide 27, score 41.93
- Tumor necrosis factor receptor associated protein 2, score 41.37
- T-cell immunoglobulin and mucin domain-containing protein 4, score 39.37

I-Tasser (confirmation):

- Tumor protein P73 (cell cycle control), Z-score > 1

(continued)

**Table 4** Continued

Mca-MORF1	Mga-MORF2
<p>C2H2 and C2HC zinc fingers (3), highest probability 80.42, aa 16–32            Transcription factor E2F-4, winged-helix (2), highest probability 60.60, aa 7–16</p> <p><b>Hormone signaling</b>            I-Tasser (highest probability):              Parathyroid hormone (4), Z-score &gt; 1</p> <p>HHpred (confirmation):              Kazal-type inhibitors/growth factor receptor (9), highest probability 68.60, aa 13–20</p> <p><b>Apoptosis</b>            I-Tasser (highest probability):              Apoptosis regulator BCL-2 (4), TM-scores &gt; 0.5              Apoptosis regulator BAK, TM-score &gt; 0.5</p> <p><b>Signaling/Regulation of cytoskeleton formation/Cell proliferation</b>            HHpred (highest probability):              GTPase-activator protein (47), highest probability 87.22</p> <p><b>Ubiquitination</b>            HHpred (highest probability):              UBA-like (4), highest probability 72.94, aa 1–15</p> <p><b>Membrane association</b>            HHpred (highest probability):              Tim10-like/Mitochondrial translocase (2), highest probability 68.38, aa 19–28</p> <p>Atome2, confirmation:              Photosystem I reaction center subunit IX, score 43.87</p>	<p><b>Membrane association</b>            Atome2 (highest probability):              Rieske protein, score 71.25              NADH-cytochrome b5 reductase 3, score 70.00              ATP synthase subunit alpha, score 39.91</p> <p><b>DNA replication, recombination, and repair</b>            HHpred (highest probability):              Methylated DNA-protein cysteine methyltransferase (24), highest probability 80.02, aa 13–19</p> <p>I-Tasser (confirmation):              DNA topoisomerase I, TM-score &gt; 0.5</p> <p><b>Receptor/Signaling (Immune response)</b>            HHpred (highest probability):              XII secretory phospholipase A2 precursor, probability 76.62, aa 18–24              Toxin_33/Waglerin family (acetylcholine receptor), probability 70.71, aa 11–20              Immunoglobulin domain (12), highest probability 62.63, aa 5–21              Tumor necrosis factor receptor superfamily member 17 (2), highest probability 60.38, aa 16–25</p>
Med-MORF	Mga-MORF
<p><b>Membrane association</b>            Atome2 (highest probability):              Alcohol dehydrogenase 4/Oxidoreductase, score 77.67</p> <p>I-Tasser (highest probability):              AP-2 complex subunit beta-2, Z-score 0.64</p> <p><b>Ubiquitination</b>            Atome2 (highest probability):              UPF0147 protein Ta0600/Ubiquitin-conjugating enzyme E2, 72.27</p> <p><b>Cytokine/Receptor/Immune response</b>            I-Tasser (highest probability):              Complement C5A anaphylatoxin, Z-score 0.61              Glutathione S-transferase omega-2, Z-score 0.58              Discoidin domain receptor 2, Z-score 0.74              Receptor protein-tyrosine kinase erbB-3, Z-score 0.55              Interleukin-13, Z-score 0.63              Coagulogen, Z-score 0.56</p> <p>Atome2 (confirmation):              Interleukin-12 subunit alpha, score 51.43              Tumor necrosis factor alpha-induced protein 3, score 44.65</p> <p>HHpred (highest probability):              Glutathione transferase domain/Thioredoxin (3), highest probability 67.32, aa 33–49              CG33975-PA/Glucocorticoid induced gene 1, probability 64.83, aa 20–51</p>	<p><b>Cytoskeleton dynamics/Cell proliferation and differentiation/Hormone signaling</b>            Atome2 (highest probability):              FGFR1 oncogene partner, score 88.04              HIV-1 envelope protein chimera/Chemokine receptor, score 59.63              Filamin-binding LIM protein 1, score 55.61              Sprouty-related, EVH1 domain-containing protein 1, score 34.00              Vasodilator-stimulated phosphoprotein, score 30.93              Protein enabled homolog, score 26.03              Proliferation-associated protein 2G4, score 25.29</p> <p>I-Tasser (confirmation):              Gamma filamin (2), highest Z-score 0.72</p> <p>HHpred (highest probability):              Actin, probability 87.91, aa 1–16              EP58/epidermal growth factor receptor kinase substrate 8-like protein 1, probability 71.11, aa 4–15</p> <p><b>Immune response</b>            I-Tasser (highest probability):              Glutathione S-transferase (5 hits), TM-scores &gt; 0.5</p> <p><b>Transcription factor/Nucleic-acid binding/Differentiation and development</b>            HHpred (highest probability):              Helix-loop-helix (bHLH) protein, Human Nulp1 (2), highest probability 87.71, aa 3–16</p>

(continued)

Table 4 Continued

Med-MORF	Mga-MORF
Nuclear Hormone Receptor family, probability 61.43, aa 28–69	PEP-CTERM putative exosortase interaction domain, probability 59.70, aa 1–10
<b>Transcription</b>	Sp1 transcription factor, probability 57.36, aa 5–61
HHpred (highest probability):	Josephin domain containing 3, probability 50.13, aa 6–15
Zinc finger protein 395 and 704, highest probability 57.58, aa 43–51	Kruppel-like factor (Growth-factor pathways), probability 48.23, aa 54–61
SLC2A4 regulator, 52.80, aa 43–51	
<b>Glycoprotein/Membrane association/Cell–cell connection</b>	<b>Signal transduction/Cell proliferation</b>
HHpred:	HHpred:
Protocadherin (13), highest probability 59.36, aa 53–61	Smoothened homolog (2), highest probability 81.43, aa 4–21
(poli-K region, aa 55–62)	<b>Membrane-associated protein/Hormone receptor</b>
	HHpred:
	Extracellular solute-binding protein (2), highest probability 74.34, aa 5–60
	EEV glycoprotein, probability 69.62, aa 7–36
	Lipoprotein, probability 56.70, aa 5–39
	FIG1, Factor-induced gene 1 protein (Mating/Pheromone-regulated membrane protein) (2), highest probability 51.63, aa 28–47
	<b>Glycoprotein/Membrane association/Cell–cell connection</b>
	HHpred:
	Protocadherin (26), highest probability 75.96, aa 7–14
	(poli-K region, aa 7–15)
Mtr-MORF	Rph-MORF
<b>Growth hormone receptor/Cell adhesion, migration, proliferation during embryonic development</b>	<b>Ubiquitination factors</b>
Atome2 (highest probability):	Atome2 (highest probability):
Human tissue factor (2), highest score 90.71	26S proteasome regulatory subunit rpn10, score 78.22
Skeletal dihydropyridine receptor, score 61.37	HHpred (confirmation):
Angiostatin, score 47.55	Zinc ion binding, ubiquitin interaction motif-containing protein (2), highest probability 59.68, aa 72–95
Fibronectin, score 46.00	NEDD8 ultimate buster-1/Ubiquitin-like protein, probability 41.84, aa 73–96
<b>Membrane-binding proteins</b>	<b>Membrane association</b>
Atome2 (highest probability):	Atome2 (highest probability):
Complexin (2), highest score 52.74	L-aspartate dehydrogenase/Oxidoreductase, score 71.39
HHpred (confirmation):	Transient receptor potential cation channel subfamily V member 1, score 59.26
N-acetylglucosaminyl-phosphatidylinositol de-n-acetylase, probability 76.89, aa 9–38	Unique short US2 glycoprotein, score 34.59
Membrane protein, probability 75.99, aa 26–47	
<b>Cell growth and differentiation/signaling</b>	<b>Transcription</b>
I-Tasser (highest probability):	I-Tasser (highest probability):
T-lymphoma invasion and metastasis-inducing protein 2, Z-score 0.75	Archaeal transcriptional regulator TrmB, Z-score 1.04
C3, Z-score 0.64	Atome2 (confirmation):
KEX1(DELTAP), Prohormone-processing serine carboxypeptidase, Z-score 0.74	Tumor suppressor p53-binding protein 1, score 57.01
<b>Cell differentiation</b>	HHpred (confirmation):
HHpred:	Restricted Tev Movement 2 (hormone receptor), probability 41.60, aa 61–94
Gametogenetin binding protein 2, probability 71.72, aa 3–40	Forkhead-associated phosphopeptide binding domain 1 isoform 19, probability 31.88, aa 68–101
<b>Microtubule association</b>	Exonuclease, probability 30.93, aa 69–99
HHpred:	<b>Immune resistance</b>
Kinectin 1 microtubule-dependent transport, probability 68.69, aa 26–63	HHpred (highest probability):
<b>Nucleic-acid binding/Transcription factor/DNA repair ATPase</b>	CRISPR-associated DEAD/DEAH-box helicase Csf4, probability 71.11, aa 144–165

(continued)

**Table 4** Continued

<b>Mtr-MORF</b>	<b>Rph-MORF</b>
<p>HHpred (highest probability):            Helix-loop-helix (bHLH) protein; Human Nulp1 (2), highest probability 95.13, aa 23–37            Telomeric telomere cycle, DNA-binding, protein binding, probability 68.11, aa 51–64            PHD FINGER domain, probability 62.04, aa 25–63            DNA double-strand break repair ATPase Rad50, probability 61.51, aa 42–70</p> <p><b>Signaling</b>            HHpred:            Cysteine alpha-hairpin motif, probability 65.87, aa 70–77</p> <p><b>Glycoprotein/Membrane association/Cell-cell connection</b>            HHpred:            Protocadherin, highest probability 74.29, aa 25–37 (poli-K region, aa 25–37)</p>	<p><b>Cytoskeleton organization/Cell proliferation, migration, differentiation/Immune response</b>            HHpred (highest probability):            Structural maintenance of chromosomes (3), highest probability 63.66, aa 65–140            Translation proteins SH3-like domain, 58.57, aa 61–75            RAD50 (4), highest probability 35.41, aa 163–172            Subunit of MRX complex with Mre11p and Xrs2p, probability 29.87, aa 163–172            Gelsolin (6), highest probability 46.96, aa 40–146            Villin (6), highest probability 36.65, aa 40–146            C15A11.5/Collagen family member, probability 42.97, aa 1–41            CG14217-PB, isoform B (Serine threonine kinase), probability 42.82, aa 69–91            Mitochondrial tumor suppressor 1 isoform 5, probability 38.92, aa 65–101            EGF/Laminin, probability 32.22, aa 64–99            Keratin (2), highest probability 30.68, aa 63–109            Segment polarity protein Dishevelled (Development), probability 29.40, aa 66–94            CG12047-PC, isoform C (Centrosome/spindle organization), probability 28.75, aa 65–78</p> <p>Atome2 (confirmation):            Thymosin beta-4, score 36.83            Adseverin, score 36.03            I-Tasser (confirmation):            Proliferating cellular nuclear antigen 1, Z-score 1.03            Guanine nucleotide-binding protein G(q) subunit alpha, Z-score 0.61            Chimera of Gelsolin domain 1 and C-Terminal domain of thymosin Beta-4, Z-score 0.74</p>
<b>Vel-MORF</b>	<b>Mse-ORF-B</b>
<p><b>Protein folding</b>            Atome2 (highest probability):            Chaperone protein ClpB (2), highest score 89.09</p> <p><b>Actin cytoskeleton and cell polarity regulator/Cell differentiation and adhesion/Cell cycle</b>            Atome2 (highest probability):            Myosin-7 (2), highest score 82.66            Rho-associated protein kinase 1, score 65.91            Tropomyosin alpha-1 chain, score 53.16            DNA topoisomerase 4 subunit A, score 53.11            Cell division protein ZapB, score 52.81</p> <p>I-Tasser (highest probability):            ATP-dependent helicase/nuclease subunit A, Z-score 1.19            YIUU, Z-score 0.61            Spectrin (4), highest Z-scores 1.19            Myosin-5A, Z-score 1.19            Cdc42-interacting protein 4, Z-score 0.63            Desmoplakin, TM-score 0.47</p> <p>HHpred (confirmation):            Keratin (6) (cytokine release/immune system), highest probability 94.31, aa 81–171</p>	<p><b>Cytoskeleton organization/Cell adhesion, migration, proliferation/Immune response</b>            Atome2 (highest probability):            Myomesin-1, score 90.17            Fibronectin, score 66.05            Fibrinogen-binding protein, score 32.61</p> <p><b>Hormone receptor</b>            Atome2 (highest probability):            Human tissue factor (hormone signaling/cell adhesion) (2), highest score 82.66</p> <p>HHpred (confirmation):            F11G11.10/Collagen family member, probability 41.10, aa 36–69            Alpha-actinin, probability 38.91, aa 70–85            TyrPK_CSF1-R (Cytokine/Immune response), probability 31.97, aa 95–102            Fibrinogen-binding protein/cell adhesion complex (3), highest probability 30.60, aa 82–93            PDGF Platelet-derived and vascular endothelial growth factors, probability 21.10, aa 13–28</p> <p><b>Membrane association</b>            Atome2 (highest probability):</p>

(continued)

**Table 4** Continued

Vel-MORF	Mse-ORF-B
Laminin (5) (cytokine release/immune system), highest probability 93.43, aa 41–218	Unique short US2 glycoprotein, score 77.87
<b>Membrane protein/ Receptor/Immune response</b>	I-Tasser (highest probability): mRNA export factor Mex67 (Associated to nuclear pores), Z-score 0.90
Atome2 (highest hits): C-Jun-amino-terminal kinase-interacting protein 4 Isoform 4 (Sperm surface protein), score 73.95	<b>Signaling</b> I-Tasser (highest probability): Sensor protein (3 hits), TM-scores > 0.5
HHpred (highest probability): More than 20 hits of antigens, all probabilities higher than 90, aa 12–218	<b>Nucleic acid binding/Immune response</b> HHpred (highest probability): Recombination-activating protein 2 (2), highest probability 79.31, aa 9–33
Nuclear pore complex proteins, 15 hits, all probabilities higher than 90, aa 41–220	Nucleic acid-binding proteins (4), highest probability 74.26, aa 94–105
I-Tasser (confirmation): Sensor protein (3), TM-scores > 0.5	I-Tasser (confirmation): Transcription intermediary factor 1-alpha, Z-score 0.66
Methyl-accepting chemotaxis transducer (MCPs), TM-score > 0.5	DNA polymerase sliding clamp C, Z-score 0.66
Invasin IPAD, TM-score > 0.5	
Cell invasion protein SIPD, TM-score > 0.5	
Pathogenicity island 1 effector protein, TM-score > 0.5	
Translocator protein bid, TM-score > 0.5	
Toll-like receptor 5b and variable lymphocyte receptor B.61 chimeric protein, TM-score > 0.5	
<b>Transcription factor/Nucleic-acid binding and transport</b>	
HHpred: Basic leucine zipper (bZIP) transcription factor (2), highest probability 92.14, aa 44–171	
Nucleotide binding, probability 91.76, aa 90–213	
mRNA localization machinery, probability 90.81, aa 50–171	
<b>Peu-ORF</b>	<b>Pno-ORF314</b>
<b>Cell differentiation during embryogenesis/Hormone receptor</b>	<b>Nucleic-acid binding and transcription</b>
Atome2 (highest probability): Cytoplasmic FMR1-interacting protein 1, score 61.36	Atome2 (highest probability): Small protein B, score 82.73
Tumor necrosis factor alpha/Cytokine, score 54.49	ATP-dependent RNA helicase SUPV3L1, mitochondrial, score 69.03
Atrial natriuretic peptide receptor A, score 52.42	DNA topoisomerase 4 subunit A, score 50.41
Mesoderm development candidate 2, score 44.10	I-Tasser (highest probability): Anti-sigma F factor (Prokaryote gene expression regulation) (6), highest Z-score 0.68
I-Tasser (confirmation): Mesoderm development candidate 2, Z-score 0.73	Transcriptional regulator LRPA (2), highest Z-score 0.64
Cytoplasmic FMR1-interacting protein 1, Z-score 0.92	Conserved domain protein/Transcriptional regulator, score 0.57
HHpred (confirmation): FnI-like domain (Cell adhesion/migration during embryonic development) (4), highest probability 62.25, aa 52–64	Bromodomain and PHD finger-containing protein 3; SPOIIAA, score 0.69
Jun-like transcription factor/Mitogen-activated protein kinases (Cellular responses to cytokines/Cell proliferation/differentiation), probability 50.47, 2–26	HHpred: Histone-fold (2), highest probability 58.93, aa 62–77
Resistin/Cytokine (2), highest probability 46.20, aa 49–63	CCAAT-BOX DNA binding protein subunit B, probability 50.87, aa 64–77
<b>DNA replication</b>	<b>Cell differentiation during embryogenesis</b>
I-Tasser (highest probability): Proliferating cell nuclear antigen, Z-score 0.81	Atome2 (highest probability): Mesoderm development candidate 2, score 79.76
DNA polymerase processivity factor, Z-score 0.69	<b>Membrane association</b>
Poly [ADP-ribose] polymerase 15, Z-score 0.63	I-Tasser (highest probability): Sulfate transporter, TM-score 0.608
Flap structure-specific endonuclease (DNA repair/replication), Z-score 0.70	<b>Viral protein</b>
HHpred (confirmation): Proliferating cell nuclear antigen, probability 42.24, aa 6–22	HHpred (highest probability): 8 hits, highest probability 82.37, aa 3–59

(continued)

Table 4 Continued

Peu-ORF	Pno-ORF314
<b>Immune resistance</b>	I-Tasser (highest probability):
HHpred (highest probability):	Capsid protein P27 (2), highest Z-score 0.92
CRISPR-associated DxTHG motif protein, probability 75.05, aa 4–17	<b>Protein folding</b>
<b>Nucleic-acid binding/Transcriptional regulator</b>	HHpred (highest probability):
HHpred (highest probability):	LDLR chaperone BOCA, probability 77.86, aa 2–52
More than 40 hits, highest probability 60.91, aa 34–66	<b>Immune response</b>
	HHpred:
	Immunoglobulin domain, probability 45.90, aa 79–103

NOTE.—Hits with the highest probability are reported for each of the three programs together with eventual confirmation of the same biological process from the other two softwares. Norm. Z-score > 1 = good alignment; TM-score > 0.5 = similar fold with query (Zhang 2008; Xu and Zhang 2010); (n) = number of the same hit (protein), when more than one. See also [supplementary tables S2–S16](#), [Supplementary Material](#) online.

(Cai and Petrov 2010). Similarly, the lineage-specific novel mtORFs may experience such a kind of evolutionary pressure, maybe for features related to sexual differentiation.

A large amount of pathways toward new gene origin through the domestication of parasitic genome sequences has been documented (Kaessmann 2010). In addition to their infectious properties, which enable them to spread horizontally between individuals and across species, many viruses can also become part of the genetic material of their host, a process that is called endogenization: endogenous viruses have integrated into the germ line of their host, allowing for vertical transmission and fixation in the host population (Boeke and Stoye 1997; Belshaw et al. 2004; Feschotte and Gilbert 2012). Viruses are able to integrate both in eukaryote and prokaryote genomes: for example, ORFans present in bacterial genomes are hypothesized to have been acquired through horizontal transfer from viruses (Daubin and Ochman 2004a, 2004b). Quite remarkably, the initiator protein DnaC in bacteria and the mitochondrial DNA replication and transcription apparatus have been recently documented to have a viral origin (Forterre 2010 and references therein). In the light of what reported above about endogenization in prokaryotes, a viral origin of novel mitochondrial genes is not unconceivable.

Novel ORFs were recently found also in the linear mitochondrial genome of Medusozoa. Using the same approach as for bivalve novel ORFs, we found a complete homology of Amo-PolB with the polymerase beta of several organisms and of Ico-mtMutS with a DNA mismatch repair protein (thus confirming the results obtained by Smith et al. 2011 and McFadden and van Ofwegen 2013, respectively). In both cases, the function of the novel mitochondrial proteins is supported. Instead, even if the product of ORF314 was proposed to act in concert with PolB in the maintenance of chromosome ends, it did not show a sound similarity with any other protein in database (Kayal et al. 2011). Interestingly, we found that it shares many predicted functions with the novel mitochondrial ORFs of bivalves ([supplementary table S18](#), [Supplementary Material](#) online). In fact, almost all the analyzed bivalve

ORFs, together with Pno-ORF314, show hits pointing to immune response and viral proteins (tables 4 and 5). Viruses can manipulate the host cell molecular machinery to counteract antiviral defences and to control the expression of their own genes, moreover viral sequences can be co-opted for host cell functions (Feschotte and Gilbert 2012), contributing to host genome evolution. For example, a viral gene has been co-opted to serve an important function in the physiology of mammals: syncytin is the envelope gene of a human endogenous defective retrovirus and is important in human placental morphogenesis and probably in the immune tolerance of the developing embryo (Mi et al. 2000). Interestingly, recent data attest that some genes involved in mammal placental development derive from domestication of multiple retrovirus-derived genes (Nakagawa et al. 2013). Similarly, we think that virus-derived novel mitochondrial proteins may have acquired new functions in the host. All the analyzed ORFs show an involvement in transcription regulation, like many virus-derived sequences that have been incorporated into the regulatory system of mammalian genes (Britten and Davidson 1969; Feschotte 2008; Cohen et al. 2009).

### Role in Immune Response and Apoptosis

Microbial invasion generally causes an immune reaction (Galluzzi et al. 2008). Mitochondria play a central role in primary host defence mechanisms against viral infections, and a number of viral proteins interact with mitochondria to regulate cellular responses (Ohta and Nishiyama 2011). Once viruses infect their hosts, they activate signalling pathways leading to the production of specific molecules (i.e., chemokines and cytokines) (Bryant and Fitzgerald 2009; Takeuchi and Akira 2009), and viruses have developed strategies to evade host immune responses: because signalling from recognition receptors converges in mitochondria, it is plausible that viruses would target mitochondrial processes to evade immune responses (Ohta and Nishiyama 2011). A clue in favor of an interaction between novel mitochondrial ORFs and immune system comes from the many hits pointing to

**Table 5**

Hits to Viral Proteins Found in Novel Mitochondrial ORFs

DUI sp.	Hits	Position
<b>FORF</b>		
Mse	Protein Tat [Atome2; score 54.94] ( <b>Nuclear transcriptional activator of viral gene expression/Cell division</b> )	n.a.
	Protein Tat [I-Tasser; norm. Z-score 0.79]	n.a.
	Protein Tat [HHpred; probability 25.94]	62–73
	SARS receptor-binding domain-like [HHpred; 54.78]	31–61
	Hepatitis E virus ORF-2 ( <b>Capsid protein/Pro-apoptotic gene expression activation/Host-cell cytoplasm</b> ) [HHpred; 23.74]	61–69
	Fijivirus P9-2 protein ( <b>Unknown function</b> ) [HHpred; probability 23.19]	8–50
Mca	Unique short US2 glycoprotein ( <b>Viral protein/Transport across membrane/Immune recognition masking</b> ) [Atome2; score 82.16]	n.a.
	Pre-neck appendage protein (Bacteriophage) (5 hits) [Atome2; score 57.87–51.81]	n.a.
	Antiviral helicase SKI2 [I-Tasser; norm. Z-score 0.68]	n.a.
	Infectivity protein G3P ( <b>Viral protein</b> ) [I-Tasser; norm. Z-score 0.62]	n.a.
	Cyclophilin-like domain ( <b>Viral infection cofactor/RNA and protein processing</b> ) [I-Tasser; norm. Z-score 0.59]	n.a.
	Phage small terminase subunit ( <b>DNA binding/Endonuclease activity/Viral capsid assembly</b> ) [HHpred; probability 44.52]	8–45
Med	Retrovirus capsid dimerization domain-like (2) [HHpred; probability 35.34, 29.28]	14–43
Mga	Retrovirus capsid dimerization domain-like (2) [HHpred; probability 35.47, 30.09]	14–43
Mtr	Unique short US2 glycoprotein ( <b>Viral protein/Transport across membrane/Immune recognition masking</b> ) [Atome2; score 65.52]	n.a.
	Positive stranded ssRNA viruses [HHpred; probability 28.66]	16–54
Rph	Polymerase PB2 ( <b>Polymerase; Viral RNA replication</b> ) [Atome2; score 56.13]	n.a.
Vel	VP1, the protein that forms the mRNA-capping machine ( <b>Viral protein</b> ) (2) [I-Tasser; norm. Z-score 0.82, 0.70]	n.a.
	Fibrinogen ( <b>Viral protein</b> ) [I-Tasser; norm. Z-score 0.64]	n.a.
<b>MORF</b>		
Mca <sup>ORF1</sup>	Early 35 kDa protein ( <b>Apoptosis-preventing protein/Protease inhibitor/Response to the viral infection</b> ) [Atome2; score 47.39]	n.a.
	Phosphatidylinositol 3-kinase regulatory subunit alpha ( <b>Host-virus interaction/Signaling/Transferase</b> ) [Atome2; score 44.26]	n.a.
	V-bcl-2 ( <b>Viral protein/Apoptosis</b> ) [I-Tasser; TM-score > 0.5]	n.a.
Mca <sup>ORF2</sup>	Circulin A ( <b>Cyclic peptide/Virus cytopathic effects and replication inhibitor</b> ) [I-Tasser; norm. Z-score > 1]	n.a.
	First immunoglobulin (Ig) domain of nectin-3 ( <b>Poliovirus receptor related protein 3/Cell adhesion</b> ) [HHpred; probability 62.63]	12–21
	Coxsackie virus and adenovirus receptor (Glycoprotein A33; CTX-related type I transmembrane protein) [HHpred; probability 51.10]	5–21
	Coxsackie virus and adenovirus receptor (Car), domain 1 [ <i>Homo sapiens</i> , TaxId: 9606] [HHpred; probability 49.70]	12–21
	Hepatitis A virus cellular receptor 1 [ <i>Mus musculus</i> ] [HHpred; probability 45.53]	12–25
Med	Replicase polyprotein 1ab ( <b>Viral protein/RNA, DNA duplex-unwinding activities/ATPase/Deubiquitination</b> ) [Atome2; score 58.58]	n.a.
	Macro domain of Non-structural protein 3 ( <b>Viral protein/RNA binding protein</b> ) [I-Tasser; norm. Z-score 0.70]	n.a.
Mga	HIV-1 envelope protein chimera ( <b>Viral envelope glycoprotein/Chemokine receptor</b> ) [Atome2; score 59.63]	n.a.
	Proliferation-associated protein 2G4 ( <b>Viral Translation/Growth regulation/Androgen receptor/Transcriptional regulation</b> ) [Atome2; score 25.29]	n.a.
	Viral protein [I-Tasser; norm. Z-score 0.72]	n.a.
Mtr	—	—
Rph	Unique short US2 glycoprotein ( <b>Viral protein/Transport across membrane/Immune recognition masking</b> ) [Atome2; score 34.59]	n.a.
	Viral protein/Signaling protein [I-Tasser; norm. Z-score 0.57]	n.a.
	CRISPR-associated DEAD/DEAH-box helicase Csf4 ( <b>Phage genomic sequence insertion/Resistance against mobile genetic elements: viruses, transposable elements, conjugative plasmids</b> ) [HHpred; probability 71.11]	144–165
	d.172.1 gp120 core (56502) SCOP seed sequence: d1g9mg_ ( <b>Viral envelope receptor</b> ) [HHpred; probability 34.78]	125–157
Vel	—	—

(continued)

**Table 5** Continued

DUI sp.	Hits	Position
Mse <sup>ORFB</sup>	Unique short US2 glycoprotein ( <b>Viral protein/Transport across membrane/Immune recognition masking</b> ) [Atome2; score 77.87]	n.a.
	Gag-Pol polyprotein ( <b>Capsid protein/Host nucleus</b> ) [Atome2; score 54.53]	n.a.
	Glycosyltransferase (Mannosyltransferase) ( <b>Capsid viral protein/Transferase</b> ) [I-Tasser; norm. Z-score 0.90]	n.a.
	VAC_I5L (dsDNA viruses, no RNA stage; Poxviridae) ( <b>Membrane-associated protein</b> ) [HHpred; probability 31.24]	6–24
<b>Other sp.</b>		
Peu	Terminase small subunit ( <b>Viral protein</b> ) [Atome2; score 56.08]	n.a.
	CAG38821 ( <b>Viral protein</b> ) [I-Tasser; norm. Z-score 0.77]	n.a.
	Terminase small subunit ( <b>Viral protein</b> ) [I-Tasser; norm. Z-score 0.84]	n.a.
	DNA polymerase processivity factor ( <b>DNA binding/Transferase/Viral protein</b> ) [I-Tasser; norm. Z-score 0.69]	n.a.
	CRISPR-associated D <sub>x</sub> THG motif protein ( <b>Phage genomic sequence insertion/Resistance against mobile genetic elements: viruses, transposable elements, conjugative plasmids</b> ) [HHpred; probability 75.05]	4–17
Pno-ORF314	Capsid protein P27 ( <b>Viral protein</b> ) (2) [I-Tasser; norm. Z-score 0.92, 0.86]	n.a.
	Retrovirus capsid protein, N-terminal core domain ( <b>Viral replication</b> ) [HHpred; probability 82.37]	21–50
	RSV capsid protein {Rous sarcoma virus [TaxId: 11886]} [HHpred; probability 80.17]	21–59
	JSRV capsid, capsid protein P27; zinc-finger, metal-binding {Jaagsiekte sheep retrovirus} ( <b>Viral protein</b> ) [HHpred; probability 78.55]	21–59
	Capsid protein P27; retrovirus, N-terminal core domain {Mason-pfizer monkey virus} ( <b>Viral protein</b> ) [HHpred; probability 74.21]	21–59
	GAG polyprotein capsid protein P27; retrovirus, immature GAG{Rous sarcoma virus} ( <b>Viral protein</b> ) [HHpred; probability 48.94]	21–50
	Capsid protein P27; viral protein, retrovirus, GAG; 7.00 A {Mason-pfizer monkey virus} [HHpred; probability 44.98]	22–59
	Capsid protein; two independent domains helical bundles, virus/viral protein {Rous sarcoma virus} [HHpred; probability 43.53]	21–47
	Tat binding protein 1 (TBP-1)-interacting protein (TBP-1) ( <b>Eukaryotic protein/Modulates the inhibitory action of human TBP-1 on HIV-Tat-mediated transactivation</b> ) [HHpred; probability 38.93]	3–50

NOTE.—Norm. Z-score > 1 = good alignment; TM-score > 0.5 = similar fold with query (Zhang 2008; Xu and Zhang 2010); (n) = number of the same hit (protein); position: amino acid position in the query sequence; n.a. = non applicable.

receptors and signaling molecules involved in immune response (antigens and cytokines above all). Some of these hits are present in both FORFs (Mse-FORF, Mca-FORF, Mtr-FORF, Vel-FORF; [supplementary tables S2, S3, S6, and S8, Supplementary Material](#) online) and MORFs (Mca-MORF2, Med-MORF, Mga-MORF, Rph-MORF, Vel-MORF; [supplementary tables S11–S13, S15, and S16, Supplementary Material](#) online), as in other analyzed ORFs (Mse-ORF-B, Peu-ORF; [supplementary tables S9 and S17, Supplementary Material](#) online). In Vel-MORF, the homology region almost coincides with the whole sequence (table 4 and [supplementary table S16, Supplementary Material](#) online).

Proteins reported in literature as acting in bivalve immune response (Gestal et al. 2008, and references therein) have homology with the analyzed mitochondrial ORFs, as for example, tumor necrosis factors (see hits found in Vel-FORF, Mca-MORF2, Med-MORF, Peu-ORF; [supplementary tables S8, S11, S12, and S17, Supplementary Material](#) online), interleukins (a group of cytokines; hits found in Mtr-FORF, Mca-MORF2, Med-MORF; [supplementary tables S6, S11, and S12, Supplementary Material](#) online), transforming growth factor (Kruppel-like factor; hits found in Mse-FORF, Mga-MORF; [supplementary tables S2 and S13, Supplementary Material](#) online) and platelet-derived growth factor (hit found in Mse-ORF-B;

[supplementary table S9, Supplementary Material](#) online). All the reported findings strongly support a link between these mitochondrial novel proteins and the immune response of bivalves.

Microbial invasion also has a role in apoptosis regulation (Galluzzi et al. 2008): viruses have acquired the capacity to control host cell apoptosis and inflammatory responses, thus evading immune reactions (Galluzzi et al. 2008). Mitochondria have a central role also in apoptosis and, for this reason, a number of viral proteins are targeted to mitochondria to regulate this mechanism. Interestingly, hits of structural analogues with apoptotic factors were found with high probability in Mca-MORF1 (apoptosis regulator BCL-2, four hits with TM-scores > 0.5, and apoptosis regulator BAK, TM-score > 0.5) (table 4). It is known that several viral polypeptides are homologues of host-derived apoptosis-regulatory proteins, such as members of the BCL-2 family (Galluzzi et al. 2008), some of which assemble on the mitochondrial membrane (Wei et al. 2001; Kuwana et al. 2002; Nutt et al. 2002).

Viral BCL-2 homologues (vBCL-2) do not show significant sequence similarity with their host counterparts, but exhibit high structural resemblance (White et al. 1991; Cuconati and White 2002). This seems exactly the case of Mca-MORF1, in which the similarity with both BCL-2 and BAK proteins was

detected in the structure, not in the sequence (supplementary table S10, Supplementary Material online). Interestingly, viral proteins with a three-dimensional folding similar to BCL-2 are glycoprotein always showing a transmembrane domain flanked by positively charged amino acids (typically lysines) and followed by an hydrophilic tail (Wang et al. 2002; Douglas et al. 2007; Kvensakul et al. 2007). This domain is required for both the mitochondrial outer membrane targeting and the anti-apoptotic function (Douglas et al. 2007; Kvensakul et al. 2007). Interestingly, all these characters are shared by *Mytilus* MORFs and Rph-MORF (the latter with serines instead of lysines). Moreover, in some FORFs (Med-FORF, Mga-FORF, and Peu-ORF; supplementary tables S4, S5, and S17, Supplementary Material online), N-terminal homeodomain (PHD)-like regions were found. Recently, several PHD-containing viral proteins have been identified to promote immune evasion by down-regulating proteins that govern immune recognition by functioning as E3 ubiquitin ligases (Coscoy and Ganem 2003). Other hits specifically related to E3 ubiquitin ligases were found (Mse-FORF, Rph-FORF, Vel-FORF, Mse-ORF-B, Mca-MORF2; supplementary tables S2, S7–S9, and S11, Supplementary Material online). For all above-mentioned, we propose that the novel ORFs here analyzed may have originated from viral elements with a function in immune response and apoptosis control.

#### Interaction with Cytoskeleton: Mitochondrial Segregation

MORFs, together with viral hits, show many hits related to cytoskeleton/cytoskeleton-binding proteins. For example, among viral hits we obtained capsid proteins and Transactivator of transcription (Tat) proteins, a regulatory protein that enhances the efficiency of viral transcription and alters microtubule dynamics, promoting proteasomal degradation and a mitochondrion-dependent apoptotic pathway (Chen et al. 2002; Aprea et al. 2006; Egelé et al. 2008). Envelope proteins generally induce a perinuclear clustering of mitochondria by altering cytoskeleton conformation, interacting for example with keratins and microtubules, thus promoting the aggregation of these organelles (Doorbar et al. 1991; Galluzzi et al. 2008). Taking into account that mitochondria appear to respond to some viral infection by migrating with viral tegument proteins (Ohta and Nishiyama 2011), we suggest that these novel ORFs might have a role in the aggregation and localization of mitochondria, producing the aggregated and dispersed patterns of distribution of spermatozoon mitochondria observed in early DUI embryos. Many other hits are connected with cytoskeleton, such as microtubule-binding proteins, actin-binding proteins, cytoskeleton proteins themselves, and proteins with a role in cytoskeleton organization (table 4). Interestingly, several endosymbiotic pathogens can use proteins expressed on their surface to ensure their survival and/or alter host processes. These surface proteins can cause cytoskeleton remodeling, as best demonstrated in *Listeria*

*monocytogenes*: this endosymbiont induces actin to assemble on its surface, propelling it through the cytoplasm and allowing its transport between host cells, bypassing host defense mechanisms (Iretton and Cossart 1997, and references therein). It is possible that MORFs bind some cytoskeleton elements, and, if they were membrane-associated proteins, they could be responsible for spermatozoon mitochondria positioning in DUI embryos.

#### Targeting and Export of Mitochondrial Novel Proteins

It is well established that the nucleus regulates organelle gene expression through anterograde regulation (Woodson and Chory 2008 and references therein). On the other hand, several studies have recently demonstrated that signals from organelles regulate nuclear gene expression by retrograde signaling (Butow and Narayan 2004). It appears likely that, given the complex cross-talk between the nucleus and mitochondria, not only chemical messengers but also exported proteins may participate in transducing signals from mitochondrion to nucleus.

A deeply studied example is the retrograde signaling that characterizes plants with Cytoplasmic Male Sterility (CMS) (Abad et al. 1995; Fujii and Toriyama 2008; Nizampatnam et al. 2009). CMS is known to be associated with the expression of novel mitochondrial ORFs and the accumulation of these novel proteins at proper spatial or temporal development stages induces male sterility (Fujii and Toriyama 2008). Moreover, some of these proteins contain a hydrophobic N terminus, commonly found in membrane-bound proteins (Abad et al. 1995 and references therein) so that it was hypothesized that they are mitochondrial membrane-bound proteins that might lead to disruption of the mitochondrial membrane integrity in the anther tissues, leading to pollen death (Nizampatnam et al. 2009, and references therein). The possibility of binding membranes is a feature in common with the here studied novel bivalve ORFs. In fact, many hits of the novel bivalve mitochondrial ORFs we analyzed were identified as proteins with a function on the cytoplasmic side of mitochondrial outer membrane (table 4). For example, bivalve mitochondrial novel proteins may tag the surface of mitochondria: MORFs may have a role in the maintenance of sperm mitochondria aggregation in the first stages of development, possibly masking them from the degradation that normally affects mitochondria carried from sperm in species with the more usual maternal inheritance of mitochondria. This could be possible thanks to the features that novel ORFs share with anti-apoptotic factors. Maybe, a similar mechanism involving novel ORF integration in the mitochondrial genome of females makes FORFs responsible for the inheritance of F-type mitochondria in DUI species, but, in this case, no evident difference from a SMI mechanism for mitochondrial transmission could be seen.

The presence of mitochondrial proteins in diverse cellular extramitochondrial sites, such as endoplasmic reticulum and nucleus, supports the existence of specific export mechanisms by which certain proteins exit mitochondria (Soltys and Gupta 2000). Mitochondria are derived from bacteria from which they probably inherited protein exit pathways used to elude host defense mechanism before the endosymbiont became an essential organism. Some of these protein exit mechanisms might have been retained and/or modified in mitochondria, allowing certain mitochondrial proteins to have additional functions in other subcellular compartments (Soltys and Gupta 2000). For example, besides the export of mitochondrial ribosomes in the cytoplasm, some mitochondrially encoded proteins are present on the cell surface as histocompatibility antigens, and are therefore exported from mitochondria (Soltys and Gupta 2000, and references therein). These peptides derive from partial sequences of mitochondrial genes (e.g., N-terminus of NADH dehydrogenase subunit 1, in mouse and humans; internal region of ATPase 6, in rat) probably by proteolysis of parent molecules inside mitochondria or in the cytoplasm, before being transported to the cell surface (Soltys and Gupta 2000). More than one mechanism by which mitochondrial matrix macromolecules are exported may exist but the processes are not fully clear yet. For example, the presence versus the detachment by peptidase of part of the protein sequence (for example an N-terminal SP) was proposed to be the cause of the re-targeting of mitochondrial proteins, and the use of protein import machinery, the leakage from breaks in the mitochondrial membranes during fission and/or fusion, membrane fusion with other organelles (e.g., endoplasmic reticulum and nucleus), the existence of protein transporters, the autotransport through lipids (as observed for heat shock proteins), and vesicle-mediated export involving vesicle budding (as in gram-negative bacteria) are other proposed mechanisms (Soltys and Gupta 2000). In our case, given the presence of a SP in many of the analyzed ORFs, this N-terminal sequence may be used to target the proteins to sites outside mitochondria. It is possible that proteins with post-transcriptional cleavage of the SP remain attached at the mitochondrial outer membrane, whereas peptide complete with the SP may be targeted elsewhere in the cell.

### The Origin of Mitochondrial Novel ORFs and Implications for DUI Evolution

As mentioned, many clues point to a viral origin of novel mitochondrial ORFs, even if the probability of the hits is sometimes low and the regions of similarity of short length (table 5). As in the case of ORFans, this can be due to the extreme limited sampling of viral sequences (Daubin and Ochman 2004a, 2004b; Lerat et al. 2005). Suttle (2005) estimated that the virus population size in the ocean alone is  $\sim 4 \times 10^{30}$ , with a phage diversity of  $\sim 10^8$  (Rohwer 2003). For this reason, a significant fraction of the ORFs without

detectable viral homologs may have arisen from not yet sequenced or extinct viruses (Yin and Fischer 2006). Moreover, many ORFans may remain without viral homologs if they have experienced rapid evolution after the integration in the new genome, diverging to the extent that no homology to viral proteins is detectable (Charlebois et al. 2003; Domazet-Lošo and Tautz 2003; Daubin and Ochman 2004a; Siew and Fischer 2004; Yin and Fischer 2006).

The co-option of such novel genes by viral hosts may have determined some evolutionary aspects of host life cycle, possibly involving mitochondria (Forterre 2006; Koonin 2006), and, as supposed for ORFans (Hendrix et al. 2000; Juhala et al. 2000), bivalve mtORFs might now be involved in key cellular functions. The study of novel mitochondrial proteins expression during the bivalve life cycle could help in understanding their function and their possible interaction with nuclear genomes.

We can hypothesize that viral selfish elements may have colonized the mitochondrial genome in male bivalves promoting its segregation into primordial germ cells, thus allowing the transmission to next generations and leading to DUI achievement. If this is true, the insertion event and the appearance of DUI might be causally linked, and some implications on the origin and evolution of DUI become evident. DUI presents a scattered distribution in bivalves, and two main hypotheses have been proposed so far to account for this: 1) an unique ancient origin and subsequent reversion to standard maternal inheritance in some lineages, or 2) multiple independent origins during bivalve evolution. If these novel ORFs are in some way linked to DUI establishment, a multiple origin of DUI should not be discarded, even if it is in contrast to the mostly accepted evolutionary scenario of a single origin of DUI (Zouros 2012). The overall function similarity among all analyzed ORFs supports their origin from elements of the same kind, but the impossibility to obtain a comprehensive good alignment and their conservation only among close relative species may indicate that either they originated from independent events or their fast evolution wiped out sequence similarities. Both hypotheses cannot be definitely accepted or discarded.

Finally, the general mechanism proposed above for the transmission of selfish elements would imply that bivalves are in some way prone to viral integration in the mitochondrial genome and therefore in DUI establishment, and maybe that other animals can have experienced such kind of mitochondrial transmission modification but no evidence has been found so far.

### Supplementary Material

Supplementary materials S1 and S2, tables S1–S19, and figures S1–S7 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org>).

## Acknowledgments

The authors thank Eleonora Sparnanzoni for her precious help in lab work. This work was supported by the Italian Ministero dell'Università e della Ricerca Scientifica funding (PRIN09) and by the Donazione Canziani bequest.

## Literature Cited

- Abad AR, Mehrrens BJ, Mackenzie SA. 1995. Specific expression in reproductive tissues and fate of a mitochondrial sterility-associated protein in cytoplasmic male-sterile bean. *Plant Cell* 7:271–285.
- Andersson SG, et al. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396:133–140.
- Apra S, et al. 2006. Tubulin-mediated binding of human immunodeficiency virus-1 Tat to the cytoskeleton causes proteasomal-dependent degradation of microtubule-associated protein 2 and neuronal damage. *J Neurosci*. 26:4054–4062.
- Belshaw R, et al. 2004. Long-term reinfection of the human genome by endogenous retroviruses. *Proc Natl Acad Sci U S A*. 101:4894–4899.
- Bernsel A, Viklund H, Hennerdal A, Elofsson A. 2009. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res*. 37: W465–W468.
- Bilewicz JP, Degnan SM. 2011. A unique horizontal gene transfer event has provided the octocoral mitochondrial genome with an active mismatch repair gene that has potential for an unusual self-contained function. *BMC Evol Biol*. 11:228.
- Birky CW Jr. 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu Rev Genet*. 35: 125–148.
- Boeke JD, Stoye JP. 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In: Coffin JM, Hughes SH, Varmus H, editors. *Retroviruses*. Plainview (NY): Cold Spring Harbor Laboratory Press. p. 343–435.
- Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Res*. 27: 1767–1780.
- Bouaouina M, et al. 2012. A conserved lipid-binding loop in the kindlin FERM F1 domain is required for kindlin-mediated  $\alpha$ IIb $\beta$ 3 integrin co-activation. *J Biol Chem*. 287:6979–6990.
- Breton S, et al. 2009. Comparative mitochondrial genomics of freshwater mussels (*Bivalvia*: Unionoidea) 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. Novel protein genes in animal mtDNA: a new sex determination system in freshwater mussels (*Bivalvia*: Unionoidea)? *Mol Biol Evol*. 28:1645–1659.
- Breton S, et al. 2011b. Evidence for a fourteenth mtDNA-encoded protein in the female-transmitted mtDNA of marine mussels (*Bivalvia*: Mytilidae). *PLoS One* 6:e19365.
- Britten RJ, Davidson EH. 1969. Gene regulation for higher cells: a theory. *Science* 165:349–357.
- Bryant C, Fitzgerald KA. 2009. Molecular mechanisms involved in inflammasome activation. *Trends Cell Biol*. 19:455–464.
- Butow RA, Narayan GA. 2004. Mitochondrial signaling: the retrograde response. *Mol Cell*. 14:1–15.
- Cai JJ, Petrov DA. 2010. Relaxed purifying selection and possibly high rate of adaptation in primate lineage-specific genes. *Genome Biol Evol*. 2: 393–409.
- Cao L, Kenchington E, Zouros E. 2004. Differential segregation patterns of sperm mitochondria in embryos of the blue mussel (*Mytilus edulis*). *Genetics* 166:883–894.
- Charlebois RL, Clarke GD, Beiko RG, St Jean A. 2003. Characterization of species-specific genes using a flexible, web-based querying system. *FEMS Microbiol Lett*. 225:213–220.
- Chen D, Wang M, Zhou S, Zhou Q. 2002. HIV-1 Tat targets microtubules to induce apoptosis, a process promoted by the pro-apoptotic Bcl-2 relative Bim. *EMBO J*. 21:6801–6810.
- Chothia C, Lesk AM. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO J*. 5:823–826.
- Claverie JM, et al. 2009. Mimivirus and Mimiviridae: giant viruses with an increasing number of potential hosts, including corals and sponges. *J Invertebr Pathol*. 101:172–180.
- Cogswell AT, Kenchington EL, Zouros E. 2006. Segregation of sperm mitochondria in two- and four-cell embryos of the blue mussel *Mytilus edulis*: implications for the mechanism of doubly uniparental inheritance of mitochondrial DNA. *Genome* 49:799–807.
- Cohen CJ, Lock WM, Mager DL. 2009. Endogenous retroviral LTRs as promoters for human genes: a critical assessment. *Gene* 448: 105–114.
- Coscoy L, Ganem D. 2003. PHD domains and E3 ubiquitin ligases: viruses make the connection. *Trends Cell Biol*. 13:7–12.
- Cuconati A, White E. 2002. Viral homologs of BCL-2: role of apoptosis in the regulation of virus infection. *Genes Dev*. 16:2465–2478.
- Daubin V, Ochman H. 2004a. Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. *Genome Res*. 14:1036–1042.
- Daubin V, Ochman H. 2004b. Start-up entities in the origin of new genes. *Curr Opin Genet Dev*. 14:616–619.
- Di Tommaso P, et al. 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res*. 39: W13–W17.
- Domazet-Loso T, Tautz D. 2003. An evolutionary analysis of orphan genes in *Drosophila*. *Genome Res*. 13:2213–2219.
- Doorbar J, Ely S, Sterling J, McLean C, Crawford L. 1991. Specific interaction between HPV-16 E1-E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network. *Nature* 352:824–827.
- Douglas AE, Corbett KD, Berger JM, McFadden G, Handel TM. 2007. Structure of M11L: a myxoma virus structural homolog of the apoptosis inhibitor, Bcl-2. *Protein Sci*. 16:695–703.
- Egelé C, et al. 2008. Modulation of microtubule assembly by the HIV-1 Tat protein is strongly dependent on zinc binding to Tat. *Retrovirology* 5:62.
- Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet*. 9:397–405.
- Feschotte C, Gilbert C. 2012. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev Genet*. 13:283–296.
- Fischer D, Eisenberg D. 1999. Finding families for genomic ORFans. *Bioinformatics* 15:759–762.
- Forterre P. 2006. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res*. 117:5–16.
- Forterre P. 2010. The universal tree of life and the Last Universal Cellular Ancestor: revolution and counterrevolutions. In: Caetano-Anollés G, editor. *Evolutionary genomics and systems biology*. Hoboken (NJ): John Wiley and Sons, Inc. p. 58–62.
- Fujii S, Toriyama K. 2008. Genome barriers between nuclei and mitochondria exemplified by cytoplasmic male sterility. *Plant Cell Physiol*. 49: 1484–1494.
- Galluzzi L, Brenner C, Morselli E, Touat Z, Kroemer G. 2008. Viral control of mitochondrial apoptosis. *PLoS Pathog*. 4:e1000018.
- Gestal C, et al. 2008. Study of diseases and the immune system of bivalves using molecular biology and genomics. *Rev Fish Sci*. 16:133–156.
- Ghiselli F, et al. 2013. Structure, transcription and variability of metazoan mitochondrial genome. Perspectives from an unusual mitochondrial inheritance system. *Genome Biol Evol*. Advance Access published July 23, 2013, doi:10.1093/gbe/evt112.
- 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.

- Hendrix RW, Lawrence JG, Hatfull GF, Casjens S. 2000. The origins and ongoing evolution of viruses. *Trends Microbiol.* 8:504–508.
- Hiller K, Grote A, Scheer M, Münch R, Jahn D. 2004. PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res.* 32:W375–W379.
- Hofmann K, Stoffel W. 1993. TMbase—a database of membrane spanning proteins segments. *Biol Chem Hoppe-Seyler.* 374:166.
- Howard MB, Ekborg NA, Taylor LE, Hutcheson SW, Weiner RM. 2004. Identification and analysis of polyserine linker domains in prokaryotic proteins with emphasis on the marine bacterium *Microbulbifer degradans*. *Protein Sci.* 13:1422–1425.
- Ireton K, Cossart P. 1997. Host-pathogen interactions during entry and actin-based movement of *Listeria monocytogenes*. *Annu Rev Genet.* 31:113–138.
- Juhala RJ, et al. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J Mol Biol.* 299:27–51.
- Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 20:1313–1326.
- Käll L, Krogh A, Sonnhammer ELL. 2004. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol.* 338:1027–1036.
- Kayal E, et al. 2011. Evolution of linear mitochondrial genomes in medusozoan cnidarians. *Genome Biol Evol.* 4:1–12.
- Koonin EV, Senkevich TG, Dolja VV. 2006. The ancient Virus World and evolution of cells. *Biol Direct.* 1:29.
- Kuwana T, et al. 2002. Bid, Bax, and lipids cooperate to form supramolecular openings in the outer mitochondrial membrane. *Cell* 111:331–342.
- Kvansakul M, et al. 2007. A structural viral mimic of pro-survival Bcl-2: a pivotal role for sequestering proapoptotic Bax and Bak. *Mol Cell* 25:933–942.
- Lerat E, Daubin V, Ochman H, Moran NA. 2005. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* 3:e130.
- McFadden CS, Sánchez JA, France SC. 2010. Molecular phylogenetic insights into the evolution of Octocorallia: a review. *Integr Comp Biol.* 50:389–410.
- McFadden CS, van Ofwegen LP. 2013. A second, cryptic species of the soft coral genus *Incrustatus* (Anthozoa: Octocorallia: Clavulariidae) from Tierra del Fuego, Argentina, revealed by DNA barcoding. *Helgol Mar Res.* 67:137–147.
- Mi S, et al. 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403:785–789.
- Milani L, Ghiselli F, Maurizi MG, Passamonti M. 2011. Doubly uniparental inheritance of mitochondria as a model system for studying germ line formation. *PLoS One* 6:e28194.
- Milani L, Ghiselli F, Passamonti M. 2012. Sex-linked mitochondrial behavior during early embryo development in *Ruditapes philippinarum* (Bivalvia Veneridae) a species with the Doubly Uniparental Inheritance (DUI) of mitochondria. *J Exp Zool B Mol Dev Evol.* 318:182–189.
- Morse DE, Duncan H, Hooker N, Morse A. 1977. Hydrogen peroxide induces spawning in molluscs, with activation of prostaglandin endoperoxide synthetase. *Science* 196:298–300.
- Mouhamadou B, Barroso G, Labarère J. 2004. Molecular evolution of a mitochondrial *polB* gene, encoding a family B DNA polymerase, towards the elimination from *Agrocybe* mitochondrial genomes. *Mol Genet Genomics.* 272:257–263.
- Nakagawa S, et al. 2013. Dynamic evolution of endogenous retrovirus-derived genes expressed in bovine conceptuses during the period of placentation. *Genome Biol Evol.* 5:296–306.
- Nizampatnam NR, Harinath D, Yamini KN, Sujatha M, Dinesh Kumar V. 2009. Expression of sunflower cytoplasmic male sterility-associated open reading frame, *orfH522* induces male sterility in transgenic tobacco plants. *Planta* 229:987–1001.
- Nutt LK, et al. 2002. Bax and Bak promote apoptosis by modulating endoplasmic reticular and mitochondrial Ca<sup>2+</sup> stores. *J Biol Chem.* 277:9219–9225.
- Ogata H, et al. 2011. Two new subfamilies of DNA mismatch repair proteins (MutS) specifically abundant in the marine environment. *Int J Soc Microbiol Ecol J.* 5:1143–1151.
- Ohta A, Nishiyama Y. 2011. Mitochondria and viruses. *Mitochondrion* 11:1–12.
- 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.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 8:785–786.
- Pons J-L, Labesse G. 2009. @TOME-2: a new pipeline for comparative modeling of protein–ligand complexes. *Nucleic Acids Res.* 37:W485–W491.
- Pont-Kingdon G, et al. 1995. A coral mitochondrial MutS gene. *Nature* 375:109–111.
- Rohwer F. 2003. Global phage diversity. *Cell* 113:141.
- Rozen S, Skaletsky HJ. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. *Bioinformatics methods and protocols: methods in molecular biology.* Totowa (NJ): Humana Press. p. 365–386.
- Shao Z, Graf S, Chaga OY, Lavrov DV. 2006. Mitochondrial genome of the moon jelly *Aurelia aurita* (Cnidaria, Scyphozoa): a linear DNA molecule encoding a putative DNA-dependent DNA polymerase. *Gene* 381:92–101.
- Siew N, Fischer D. 2004. Structural biology sheds light on the puzzle of genomic ORFans. *J Mol Biol.* 342:369–373.
- Skibinski DO, Gallagher C, Beynon CM. 1994a. Mitochondrial DNA inheritance. *Nature* 368:817–818.
- Skibinski DO, Gallagher C, Beynon CM. 1994b. Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. *Genetics* 138:801–809.
- Smith DR, et al. 2011. First complete mitochondrial genome sequence from a box jellyfish reveals a highly fragmented linear architecture and insights into telomere evolution. *Genome Biol Evol.* 4:52–58.
- Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33:W244–W248.
- Soltys BJ, Gupta RS. 2000. Mitochondrial proteins at unexpected cellular locations: export of proteins from mitochondria from an evolutionary perspective. *Int Rev Cytol.* 194:133–196.
- Suttle CA. 2005. Viruses in the sea. *Nature* 437:356–361.
- Takeuchi O, Akira S. 2009. Innate immunity to virus infection. *Immunol Rev.* 227:75–86.
- Tamura K, et al. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.
- Todd AE, Orenge CA, Thornton JM. 2001. Evolution of function in protein superfamilies, from a structural perspective. *J Mol Biol.* 307:1113–1143.
- Wang HW, Sharp TV, Koumi A, Koentges G, Boshoff C. 2002. Characterization of an anti-apoptotic glycoprotein encoded by Kaposi's sarcoma-associated herpesvirus which resembles a spliced variant of human survivin. *EMBO J.* 21:2602–2615.
- Wei MC, et al. 2001. Proapoptotic BAX and BAK: a requisite gateway to mitochondrial dysfunction and death. *Science* 292:727–730.

- White E, Cipriani R, Sabbatini P, Denton A. 1991. Adenovirus E1B 19-kilodalton protein overcomes the cytotoxicity of E1A proteins. *J Virol.* 65:2968–2978.
- Woodson JD, Chory J. 2008. Coordination of gene expression between organellar and nuclear genomes. *Nat Rev Genet.* 9:383–395.
- Wu X, et al. 2012. New features of Asian *Crassostrea* oyster mitochondrial genomes: a novel alloacceptor tRNA gene recruitment and two novel ORFs. *Gene* 507:112–118.
- Xu J, Zhang Y. 2010. How significant is a protein structure similarity with TM-score=0.5?. *Bioinformatics* 26:889–895.
- Yin Y, Fischer D. 2006. On the origin of microbial ORFans: quantifying the strength of the evidence for viral lateral transfer. *BMC Evol Biol.* 6:63.
- Yu G, Stoltzfus A. 2012. Population diversity of ORFan genes in *Escherichia coli*. *Genome Biol Evol.* 4:1176–1187.
- Zdobnov EM, Apweiler R. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847–848.
- Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:40.
- Zouros E. 2012. Biparental inheritance through uniparental transmission: the doubly uniparental inheritance (DUI) of mitochondrial DNA. *Evol Biol.* 40:1–31.
- Zouros E, Oberhauser Ball A, Saavedra C, Freeman KR. 1994a. Mitochondrial DNA inheritance. *Nature* 368:818.
- Zouros E, Oberhauser Ball A, Saavedra C, Freeman KR. 1994b. An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Proc Natl Acad Sci U S A.* 91:7463–7467.

**Associate editor:** Bill Martin

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

Fabrizio Ghiselli<sup>1,\*</sup>, Liliana Milani<sup>1</sup>, Davide Guerra<sup>1</sup>, Peter L. Chang<sup>2</sup>, Sophie Breton<sup>3</sup>, Sergey V. Nuzhdin<sup>2</sup>, and Marco Passamonti<sup>1</sup>

<sup>1</sup>Dipartimento di Scienze Biologiche, Geologiche ed Ambientali (BiGeA), Università di Bologna, Bologna, Italy

<sup>2</sup>Program in Molecular and Computational Biology, Department of Biological Sciences, University of Southern California, Los Angeles

<sup>3</sup>Département de Sciences Biologiques, Université de Montréal, Montréal, Québec, Canada

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

Accepted: July 18, 2013

**Data deposition:** DNA sequences have been deposited at GenBank under the Accession KC243324-31, KC243332-9, KC243340-6, and KC243347-53.

## 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-proteobacterium, 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

© The Author(s) 2013. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

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.caspr.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  $\mu$ L reaction volume using the GoTaq Flexi Dna Polymerase

(Promega) kit. The reaction volume was composed of 24  $\mu$ L of Nuclease-free Water (Ambion Inc.), 10  $\mu$ L of Green GoTaq Flexi Buffer 5 $\times$  (Promega), 6  $\mu$ L of MgCl<sub>2</sub> 25 mM, 1  $\mu$ L of dNTPs (Promega) mix 40  $\mu$ M (10  $\mu$ M each dNTP), 2.5  $\mu$ L of each primer (10  $\mu$ M) (Invitrogen SRL), 4  $\mu$ L of DNA template and 0.25  $\mu$ L of GoTaq DNA Polymerase (Promega) 5 U/ $\mu$ L. PCRs were performed with the following cycle: an initial denaturation at 95  $^{\circ}$ C for 2 min, followed by 30 cycles of denaturation at 95  $^{\circ}$ C for 30 s, annealing at 48  $^{\circ}$ C for 30 s and extension at 72  $^{\circ}$ C for 90 s, then a final extension at 72  $^{\circ}$ 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	<i>N</i>	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  $\Delta G$  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. philippinarum* 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

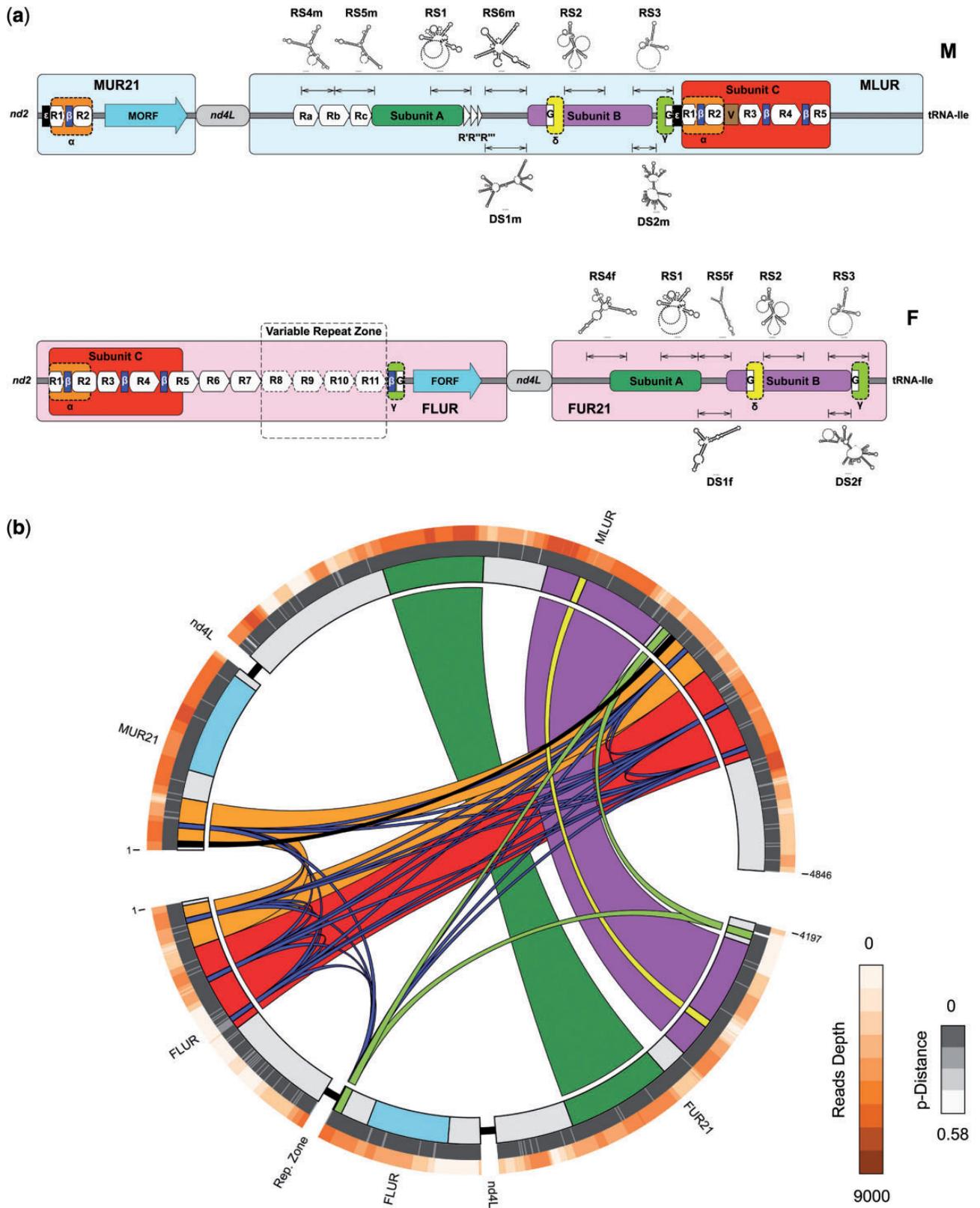
### Proportion of URs in Mitochondrial Genomes of Metazoans

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 1a shows conserved regions, motifs, repeated units, and major secondary structures (both at DNA and RNA level). Figure 1b shows transcription depth, nucleotide 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  $\alpha$ ; turquoise, novel ORFs; dark green, subunit A; orchid, subunit B; red, subunit C; yellow, motif  $\delta$ ; light green, motif  $\gamma$ ; black, motif  $\epsilon$ ; blue, motif  $\beta$ ; G, G-homopolymer; Ra, Rb, Rc, R', 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,  $\delta$  and  $\gamma$  ([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  $\delta$  (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  $\gamma$  (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,  $\beta$  and  $\epsilon$  ([supplementary fig. S5c and d](#), Supplementary Material online), that are specific of *R. philippinarum*.  $\beta$  is present in both M- and F-type mtDNAs, while  $\epsilon$  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  $\beta$  and  $\epsilon$ .

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 (TAAGTAAAACG in the male, and TAAGGTYACG in the female).

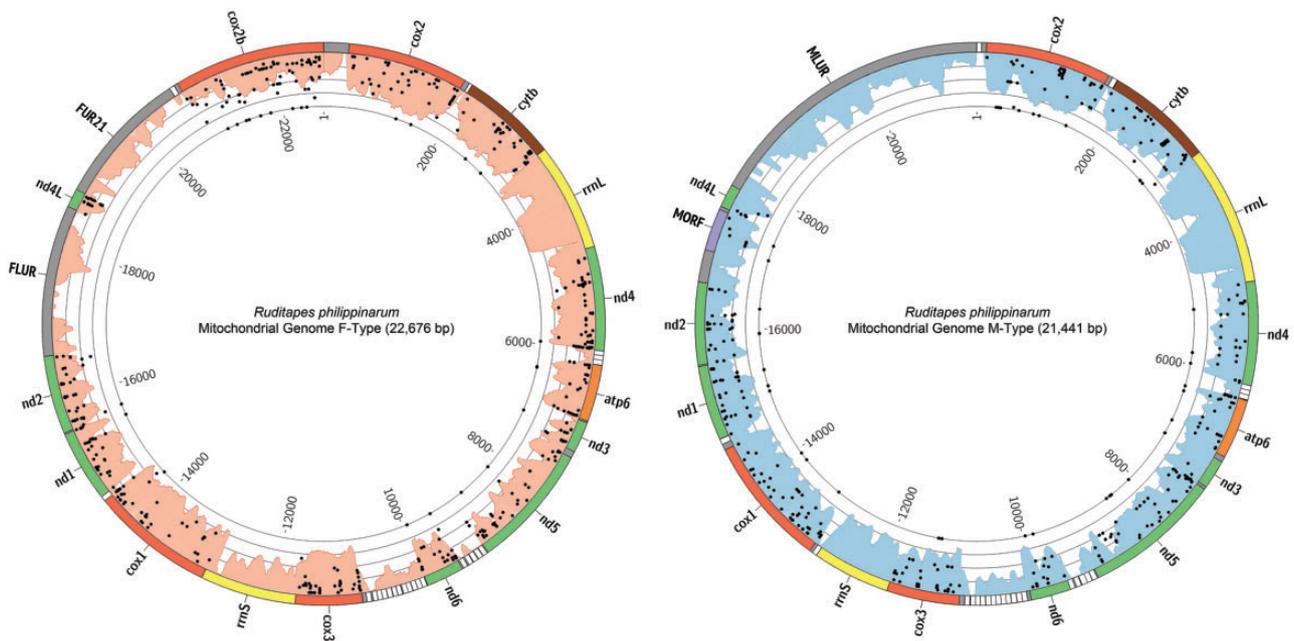
The analysis with RNAz identified 6 structures in the MLUR and 5 in the FUR21 (fig. 1a). Among them, three structures (RS1, RS2, and RS3, [supplementary figs. S10–S12](#), Supplementary Material online) are conserved between M-type and F-type, three (RS4m, RS5m, and RSm6; [supplementary figs. S13–S15](#),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

**Fig. 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. NOTE.—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  $\alpha$ ; turquoise, novel ORFs; dark green, subunit A; orchid, subunit B; red, subunit C; yellow, motif  $\delta$ ; light green, motif  $\gamma$ ; black, motif  $\epsilon$ ; blue, motif  $\beta$ .



**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).

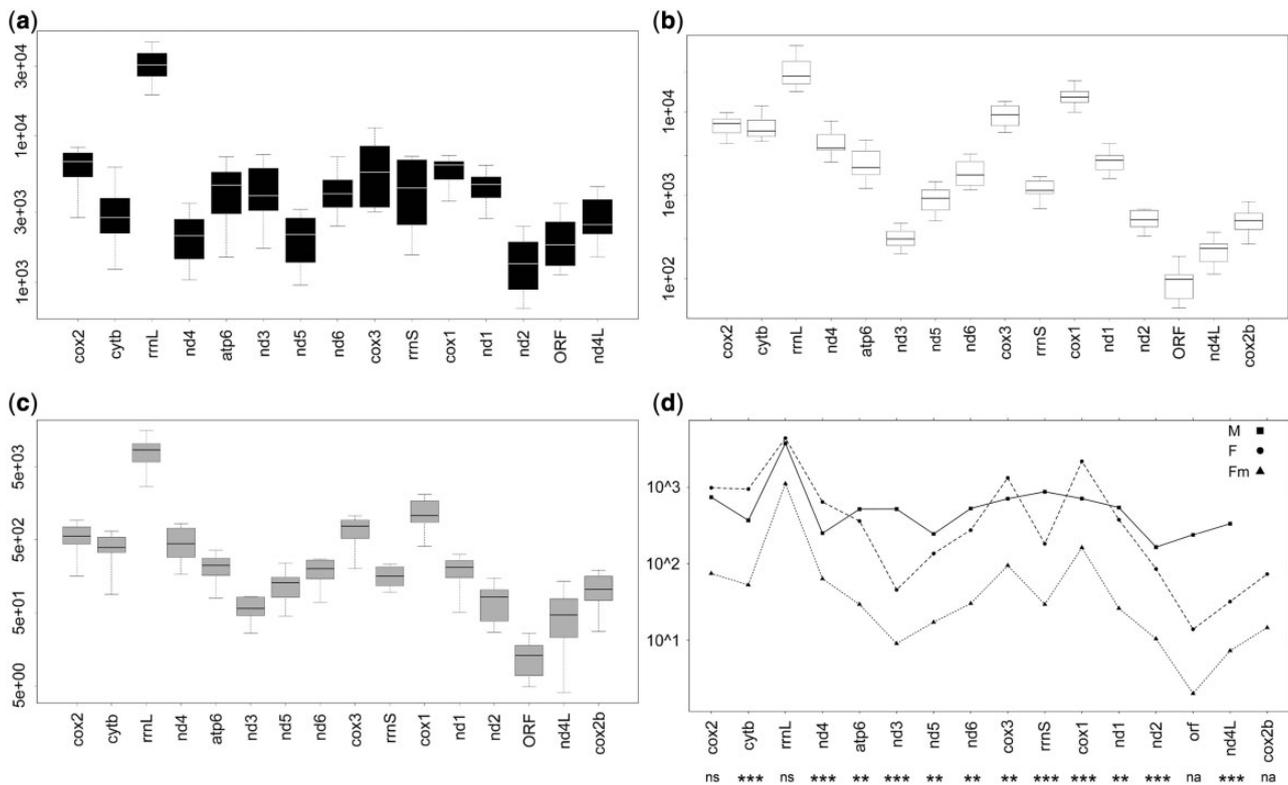
The analysis of M-type mtDNA transcriptome showed that three mitochondrial coding genes (*nd4*, *nd5*, and *nd4L*) have a

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  $100\times$  (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 ( $\tau = -0.18$ ,  $P = \text{ns}$ ;  $\rho = -0.26$ ,  $P = \text{ns}$ ), 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;  $\rho$ , Spearman's rank correlation coefficient;  $\tau$ , 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.

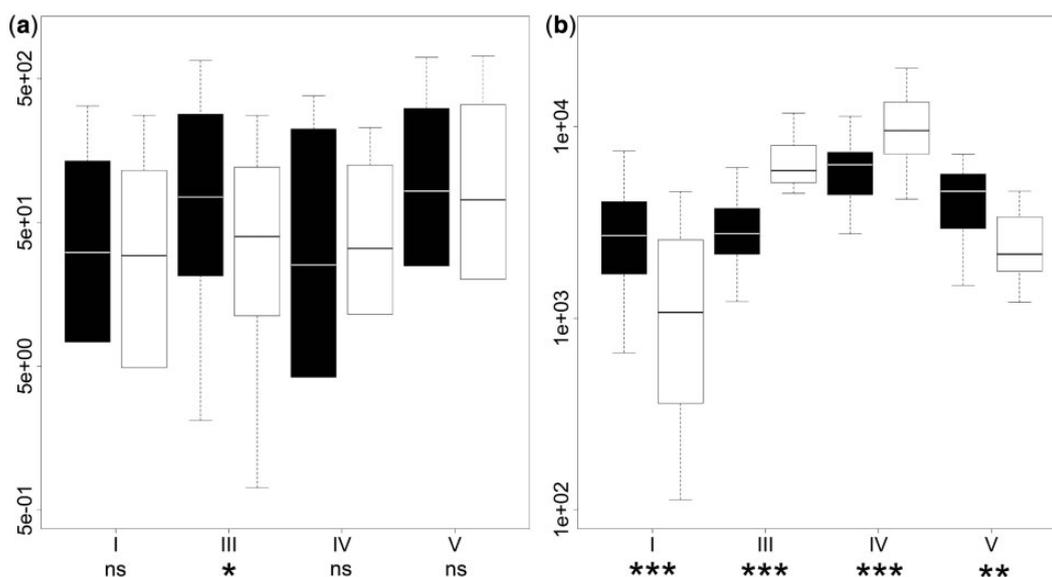
Table 4 summarizes the SNP analysis. M-type has significantly less SNPs ( $P < 0.001$ ) in comparison with both the F-types (F and Fm), which, conversely, do not differ between them (table 4 and fig. 6a). We subdivided the SNPs according to whether they are present with a single allele or multiple alleles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

Bivalvan mtDNAs have, on average,  $1.7\times$  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) Mitochondrially 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%

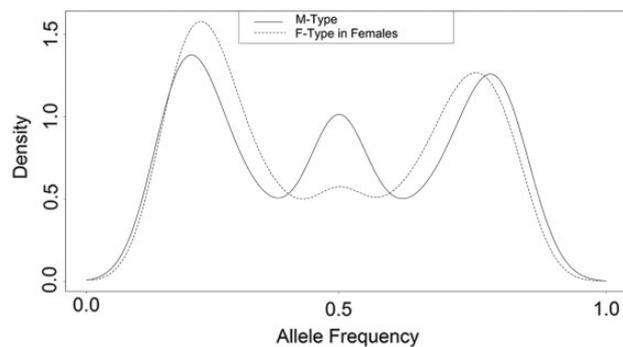
  

Genome	Depth							
	Min	Max	Mean	Median	<25	25–100	100–1,000	>1,000
F	25	2,997	1,628.723	1,772	0%	3%	34.4%	62.6%
Fm	3	3,000	1,540	1,574	1.4%	5.7%	36%	56.9%
M	20	3,000	1,866	2,150	4.8E–3%	9.6E–3%	27.4%	71.1%

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

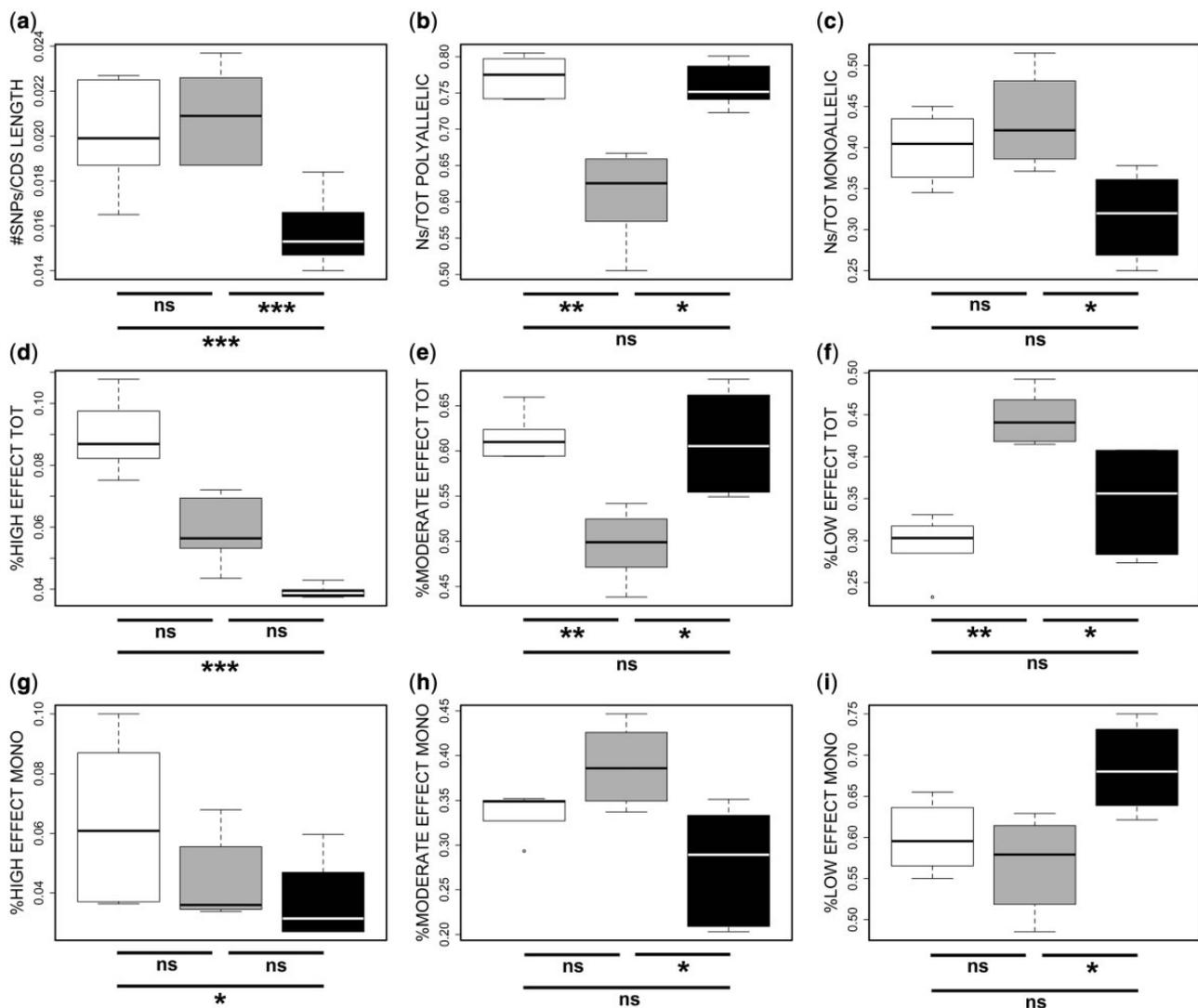
Figure 1b highlights the connections between subunits and motifs between the major URs of M- and F-type. We compared M and F mtDNAs to identify similarities and differences: similarities are supposed to be linked to a common physiological function (i.e., control of replication and transcription), whereas 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 ( $\delta$  and  $\gamma$ , 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  $\delta$  and  $9.4E-54$  for  $\gamma$ ; see Results and supplementary tables S7 and S8, Supplementary Material online, for details) whose CR has been characterized (Jacobs et al. 1989). Motifs  $\delta$  and  $\gamma$  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  $\delta$  and  $\gamma$ , 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 above-mentioned 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*
mRP1_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
mRP1_m	11	0	25	1	37	0.324	9.465E-04	1	11	25	0.027	0.297	0.676	—	—	—	—	—	—	—
mRP2_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
mRP2_m	18	0	39	0	57	0.316	1.420E-03	2	16	39	0.035	0.281	0.684	—	—	—	—	—	—	—
mRP3_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
mRP3_m	13	0	23	1	37	0.378	1.104E-03	1	13	23	0.027	0.351	0.622	—	—	—	—	—	—	—
mRP10_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
mRP10_m	14	1	48	1	64	0.250	1.262E-03	3	13	48	0.047	0.203	0.750	—	—	—	—	—	—	—
mRP11_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
mRP11_m	12	0	23	1	36	0.361	1.025E-03	1	12	23	0.028	0.333	0.639	—	—	—	—	—	—	—
mRP12_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
mRP12_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
mRP1_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
mRP1_m	50	2	56	0	108	0.481	3.706E-03	6	46	56	0.056	0.426	0.519	—	—	—	—	—	—	—
mRP2_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
mRP2_m	36	1	50	0	87	0.425	2.637E-03	3	34	50	0.034	0.391	0.575	—	—	—	—	—	—	—
mRP3_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
mRP3_m	32	1	56	0	89	0.371	2.352E-03	3	30	56	0.034	0.337	0.629	—	—	—	—	—	—	—
mRP10_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
mRP10_m	50	3	50	0	103	0.515	3.777E-03	7	46	50	0.068	0.447	0.485	—	—	—	—	—	—	—
mRP11_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
mRP11_m	31	1	51	0	83	0.386	2.281E-03	3	29	51	0.036	0.349	0.614	—	—	—	—	—	—	—
mRP12_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
mRP12_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
FRP4_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
FRP4_m	10	0	13	0	23	0.435	7.127E-04	2	8	13	0.087	0.348	0.565	—	—	—	—	—	—	—
FRP5_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
FRP5_m	9	0	11	0	20	0.450	6.414E-04	2	7	11	0.100	0.350	0.550	—	—	—	—	—	—	—
FRP6_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
FRP6_m	41	1	58	0	100	0.420	2.993E-03	7	35	58	0.070	0.350	0.580	—	—	—	—	—	—	—
FRP7_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)





**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  $\gamma$ . 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 ( $\delta$  and  $\gamma$ ) 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 *rnl*, 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 duplication 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 where 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 terms 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 ( $\sim 10\times$  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 alleles. 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 coexist 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 $\times$ ) in M, from 62.5% to 42.1% (1.48 $\times$ ) in Fm and from 77.5% to 40.4% (1.91 $\times$ ) 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 $\times$ , 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 bivalvan 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

The authors thank Prof. William Martin and three anonymous reviewers for their insightful comments and suggestions. This work was supported by the National Institutes of Health grant R01 GM098741 to S.V.N., the Italian Ministry for University and Research grant MIUR PRIN09 2009NWXMX002 to

M.P., and the Canziani Bequest fund (University of Bologna) grant A.31.CANZELSEW to M.P.

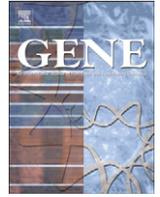
## Literature Cited

- Abhishek 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, Glémin 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, Downton 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 Blik 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



# The largest unassigned regions of the male- and female-transmitted mitochondrial DNAs in *Musculista senhousia* (Bivalvia Mytilidae)

Davide Guerra\*, Fabrizio Ghiselli, Marco Passamonti

Department of Biological Geological and Environmental Sciences, University of Bologna, via Selmi 3, 40126 Bologna, Italy

## ARTICLE INFO

### Article history:

Accepted 1 December 2013

Available online 14 December 2013

### Keywords:

Doubly uniparental inheritance

Mitochondrial genomes

Control region

Concerted evolution

Mytilids

*Musculista senhousia*

## ABSTRACT

*Musculista senhousia* is a marine mussel with doubly uniparental inheritance (DUI) of mitochondria. In this study we analyzed the largest unassigned region (LUR) of its female- and male-transmitted mitochondrial genomes, described their fine characteristics and searched for shared features. Our results suggest that both LURs contain the control region of their respective mitochondrial genomes. The female-transmitted control region is duplicated in tandem, with the two copies evolving in concert. This makes the F-mtDNA of *M. senhousia* the first Bivalve mitochondrial genome with this feature. We also compared *M. senhousia* control regions to that of other Mytilidae, and demonstrated that signals for basic mtDNA functions are retained over evolutionary times even among the fast-evolving mitochondrial genomes of DUI species. Finally, we discussed how similarities between female and male LURs may be explained in the context of DUI evolution and if the duplicated female control region might have influenced the DUI system in this species.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

Animal mitochondrial (mt) genomes are typically composed of a single, circular, double-stranded DNA molecule (mtDNA). Apart from their standard set of genes, these genomes comprise a non-coding region, the control region (CR), characterized by the presence of secondary structures, repeated sequences and conserved motifs involved in the replication and transcription of mtDNA (Scheffler, 2008). Usually, this is the only non-coding region of substantial length that is present in the mt genome, but other minor unassigned regions or spacers may be present between genes.

Several animal mt genomes, however, possess more than one CR that, in many species, maintain a high sequence similarity, a fact that points to full functionality of both copies: this unusual feature has been found across many diverse Vertebrate and non-Vertebrate taxa such as birds, snakes, lizards, fishes, crustaceans and insects (see Schirtzinger et al., 2012 and references therein for a list of studies on this topic). Inside single individuals of these species, duplicated CR sequences show an extreme conservation (up to 100%) (Schirtzinger

et al., 2012): the soundest mechanism for their concerted evolution is gene conversion by recombination between homologous sequences (Kumazawa et al., 1998). The presence of redundant CRs is in contrast to the common assumption that metazoan mt genomes tend to maintain a compact organization (Gissi et al., 2008). However, having two origins of heavy strand replication may cause an increase in its replication rate (maybe given by a more efficient initiation of replication; Kumazawa et al., 1996), an advantage that could overcome the downsides of a longer mt genome. Thus, if after a CR duplication the replication mechanism is not severely affected and both CRs are maintained functional (i.e. not subject to degeneration and loss), mtDNAs with multiple CRs might replace the single-CR versions and become fixed if they have obtained a functional advantage from the duplication (Kumazawa et al., 1996). It has been proposed that the occurrence of duplicated CRs in a mt genome might have an influence on its rate of evolution: Kumazawa et al. (1998), in a study on a snake mtDNA with duplicated CRs, suggested that a replication mechanism involving two CRs may be less accurate, allowing the mtDNA to accumulate more mutations. Nevertheless, the high variability they observed in snake mtDNAs might be the consequence of an increased replication rate.

Many Bivalve species possess a peculiar mode of mitochondrial transmission called doubly uniparental inheritance (DUI) (Breton et al., 2007; Passamonti and Ghiselli, 2009; Zouros, 2013). In species with DUI two types of mtDNA are present, one inherited from the mother through eggs (named F, from female-transmitted) and one inherited from the father through spermatozoa (named M, male-transmitted). Adult females are homoplasmic for the F line, while males have both lines in their tissues (the relative quantities of F and M vary between tissues and between species; Ghiselli et al., 2011; Obata et al., 2011; Sano et al., 2007). In male gonads, the M-type becomes dominant during

**Abbreviations:** A, adenine content; AIC, Akaike information criterion; AT, adenine and thymine content; *atp8*, gene for ATP synthase subunit 8; BIC, Bayesian information criterion; BF, Bayes factor; CR, control region; *cox1*,  $-2$  and  $-3$ , genes for cytochrome c oxidase subunits 1, 2 and 3; *cytb*, gene for cytochrome b; DUI, doubly uniparental inheritance; F, female-transmitted; GO, gene ontology; LUR, largest unassigned region; M, male-transmitted; mt, mitochondrial; *nad1*,  $-3$ ,  $-4L$  and  $-5$ , genes for NADH dehydrogenase subunits 1, 3, 4L and 5; ORF, open reading frame; *oriH* and *oriL*, origins of heavy and light strand replication; T, thymine content; tRNA, transfer RNA gene; *tRNA-Gln*, tRNA gene for glutamine; *tRNA-Glu*, tRNA gene for glutamic acid;  $\Delta$ AIC, difference between AIC values.

\* Corresponding author. Tel.: +39 051 2094172; fax: +39 051 2094286.

E-mail address: [davide.guerra7@unibo.it](mailto:davide.guerra7@unibo.it) (D. Guerra).

germ line formation: this has been explained by an active segregation mechanism during development (Zouros, 2013) and/or by a faster replication rate of the M-mtDNA (Ghiselli et al., 2011 and references therein). The length, sequence and organization of the mtDNA may differ between F and M, leading to strikingly high nucleotide and aminoacidic divergences among them (e.g. up to 43% and 50%, respectively, in Unionoids; Doucet-Beaupré et al., 2010); additionally, even the gene content may vary between them, as duplicated genes and lineage-specific novel open reading frames (ORFs) are often found in these mt genomes (Breton et al., 2009, 2011a, 2011b; Milani et al., 2013).

The CRs of the two mtDNAs usually differ in sequence and length but, despite the general great divergence of their respective mt genomes, they may maintain regions with relatively high sequence conservation. To explain the conservation of these blocks between F and M mtDNA, it has been suggested that they contain basic signals involved in replication and transcription of the molecule, like secondary structures and motifs, and thus are under more strict selective constraints than the rest of the CR (see for example Cao et al., 2004). On the other hand, the most variable parts between F and M CRs are thought to be involved in the maintenance of the separate transmission routes of the respective genomes (Breton et al., 2009 and references therein).

*Musculista senhousia* is a DUI marine mussel (Mytilidae Crenellinae), whose complete F and M mt genome sequences have been recently published (Passamonti et al., 2011). Beside a novel ORF in the F mt genome (Breton et al., 2011a; Milani et al., 2013) and a duplication of the *cox2* gene in the M, the gene content of the two mtDNAs is the same. The largest unassigned region (LUR) of the two mt genomes is placed in the same position in both lines, and it is called FLUR in the F mtDNA and MLUR in the M. The two LUR types have different lengths (FLUR: 4522 bp; MLUR: 2847 bp) but share three conserved subunits (A-type Subunits, B Subunits and  $\gamma$  Subunits), organized in a similar way in both mtDNAs. Moreover, FLUR is almost entirely composed of two large repetitive units, named Rep Units 1 and 2, whose structure resemble that of MLUR (Passamonti et al., 2011).

In this study we analyzed the LURs obtained from eleven female and twelve male Italian specimens of *M. senhousia* to better characterize them and understand which of the shared subunits are the most and least conserved in the F and M mt LURs. We also searched for molecular signatures allowing us to confirm that the analyzed LURs actually contain the CR of their respective mtDNA, by comparing *M. senhousia* and other Mytilid mt LURs and genomes. Furthermore, we analyzed the variability of FLUR Rep Units and compared it to that of duplicated CRs in other animal mt genomes. Finally, we discussed how the differences and analogies between *M. senhousia* FLUR and MLUR might have evolved and how they might have affected their respective mt genomes evolution in relation to DUI.

## 2. Materials and methods

For this study we analyzed eleven FLUR sequences (accession numbers: KC243354–KC243364) and twelve MLUR sequences (accession numbers: KC243376–KC243387) of *M. senhousia* deposited in GenBank. The same set of sequences has been also analyzed by Milani et al. (2013) with different methods to search for and compare novel ORFs in a wider phylogenetic context of DUI. MEGA5 (Tamura et al., 2011) was used for sequence alignments and to calculate p-distances and standard errors (S.E.) of LURs and conserved blocks with the bootstrap method (1000 bootstrap replications, transitions + transversions, pairwise deletion). Length variations of conserved blocks were calculated manually from the alignments. Tandem Repeats Finder (<http://tandem.bu.edu/trf/trf.html>) (Benson, 1999) was used to identify tandem repeats in the LURs subunits. DnaSP 5.10 (Librado and Rozas, 2009) was used to calculate nucleotide diversity ( $\pi$ ) with the sliding windows method (window length: 100 bp, step: 10 bp) on completely sequenced LURs alignments. DNA secondary structures predictions were performed with Mfold (<http://mfold.rna.albany.edu/?q=mfold>) (Zuker, 2003) using a folding

temperature of 16 °C (i.e. the annual average sea water temperature in the locality where *M. senhousia* specimens were collected). RNAz (<http://rna.tbi.univie.ac.at/cgi-bin/RNAz.cgi>) (Gruber et al., 2007) was used on completely sequenced LURs alignments to find conserved RNA secondary structures, using windows of 100, 200, 300 and 400 bp. Only structures with  $p > 0.95$  and a mean z-score of  $\leq -4.00$  were considered. When two or more windows with a valid z-score overlapped on the same reading direction, only the one with the lowest z-score was considered. All DNA and RNA secondary structures found were compared to those hypothesized in *Mytilus edulis* and *Mytilus galloprovincialis* CRs by Cao et al. (2004). Graphical representations of secondary structures were made with Varna 3.8 (Darty et al., 2009). MEME 4.8.1 (Bailey et al., 2009) was used to find conserved sequence motifs among the mt LURs of *M. senhousia* and those of other three DUI and non-DUI Mytilidae species (see Table 1 for the LURs used). The most supported motifs found were then submitted to all available databases of GOMO (<http://meme.nbcr.net/meme/cgi-bin/gomo.cgi>) (Buske et al., 2010) to assign them a list of GO terms, which may suggest a function for the motifs. Logo representations were generated with the online version of WebLogo (<http://weblogo.berkeley.edu>) (Crooks et al., 2004).

To find hints on the locations of heavy strand and light strand origins of replication (*oriH* and *oriL*, respectively), we calculated the AT-skew values on four-fold redundant sites of protein-coding genes in the *M. senhousia* complete mt genomes available in GenBank, using the formula  $(A + T) / (A - T)$  and following the rationale discussed in Breton et al. (2009). For a better comparison, the analysis was extended to other nine Mytilidae complete mt genomes from other five DUI and non-DUI species (see Table 1 for the mt genomes used). Given its high variability and short length, *atp8* gene was excluded from the analyses.

To evaluate the degree of similarity between FLUR Rep Units 1 and 2 among individuals, a Bayesian analysis was performed. Using both Rep Units of each FLUR plus three MLUR sequences (composed of only the parts alignable with FLUR Rep Units, i.e. Subunit A', Subunit B and  $\gamma$ 1) as outgroup, for a total of 25 sequences, we performed two analyses: a single-partition and a multi-partitions one. For the multi-partition the total alignment was subdivided in four parts (A + A', B,  $\gamma$  and  $\delta$ ). Using the MrMTgui 1.0 (Nuin, 2007) interface, the best evolutionary model for the whole single-partition and for each subunit in the multi-partitions was estimated with PAUP\* 4.0b10 (Swofford, 2002) and ModelTest 3.7 (Posada and Crandall, 1998); the Bayesian Information Criterion (BIC) was used as the model decision criterion. GapCoder (Young and Healy, 2003) was used to code alignment indels. Bayesian trees of the two datasets were built with MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003), using  $10^7$  generations; nodes with posterior probability values  $\leq 95\%$  were collapsed. To choose the best model, the results of the two analyses were compared using the Akaike Information Criterion (AIC; Akaike, 1973) and the

**Table 1**

List of mt genomes used in this study. All species belong to the family Mytilidae. Only one mt genome is available for *Perna viridis*, as no report of DUI has been made for this species.

Subfamily	Species	mt genome	Accession number	Motif search (LUR)	AT-skew analysis (whole mt genome)
Crenellinae	<i>Musculista senhousia</i>	F	GU001953	x	x
		M	GU001954	x	x
Mytilinae	<i>Mytilus californianus</i>	F	GQ527172	x	x
		M	GQ527173	x	x
	<i>Mytilus edulis</i>	F	AY484747	x	x
		M	AY823623	x	x
	<i>Mytilus galloprovincialis</i>	F	AY497292		x
		M	AY363687		x
	<i>Mytilus trossulus</i>	F	GU936625		x
		M	GU936626		x
<i>Perna viridis</i>	–	JQ970425	x	x	

Bayes Factor (BF; Kass and Raftery, 1995) following the procedure described in Plazzi and Passamonti (2010).

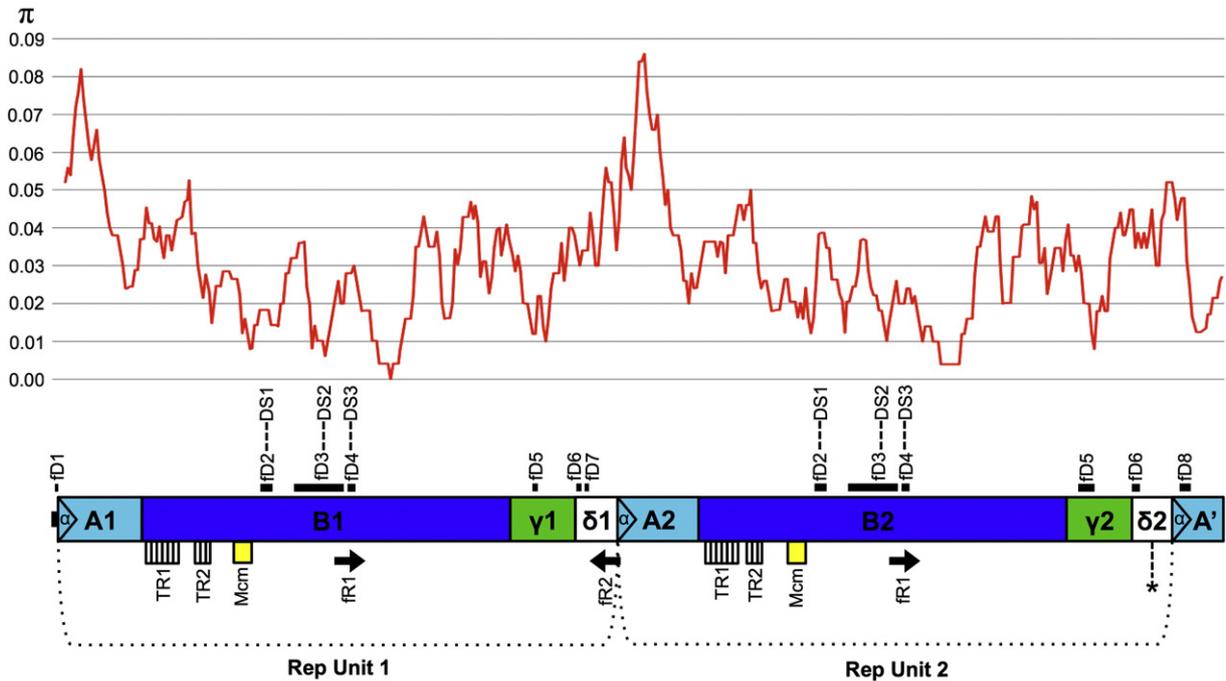
### 3. Results

Two series of small, previously unidentified tandem repeats (named TR1 and TR2) were found in FLUR B Subunits (Fig. 1). TR1 was found at the beginning of B Subunits and it is always composed of six units of 20–21 bp: the third and fifth units differ slightly in sequence from the other four (Figs. 2a–b). TR2 was found 66–67 bp downstream and it is composed of three units: the first two are 22–24 bp long, while the third is truncated to a length of 14 bp (Fig. 2c). These two series of tandem

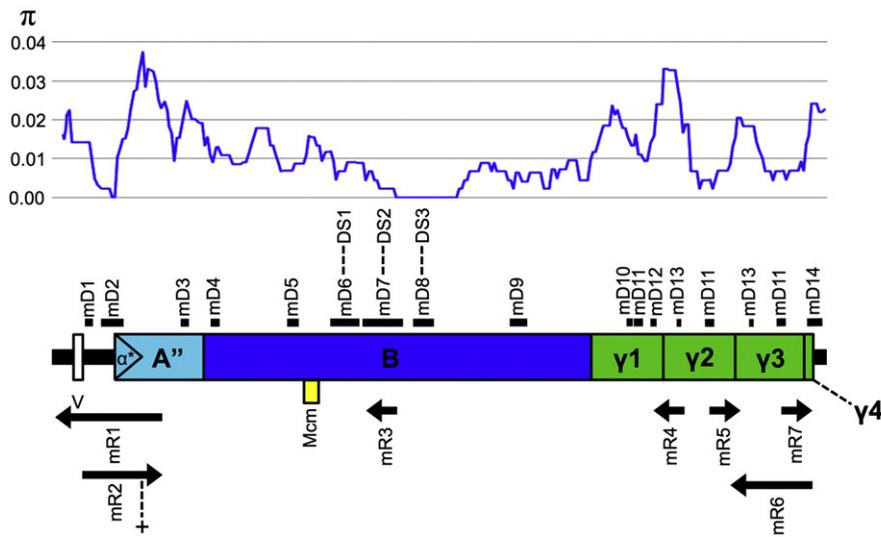
repeats are not recognizable in MLUR Subunit B: even if the sequences in the same relative positions of female B Subunits are alignable (i.e. there is no gap in the MLUR in those positions), in male Subunit B these regions are highly polymorphic, thus the repeated units are not identifiable.

A small variable region in the 5' spacer of the MLUR (named V, see Fig. 1), with a maximum length of 53 bp, can contain 5–6 copies of a 4 bp microsatellite (GTAG) or may be totally missing, as in samples m2 and m7 MLUR (accession numbers: KC243377 and KC243382, respectively). The LURs subunits show in general little length variation, in most cases due to the different extent of homopolymers. A notable exception to this general feature is a 130 bp insertion in Subunit  $\delta 2$  of

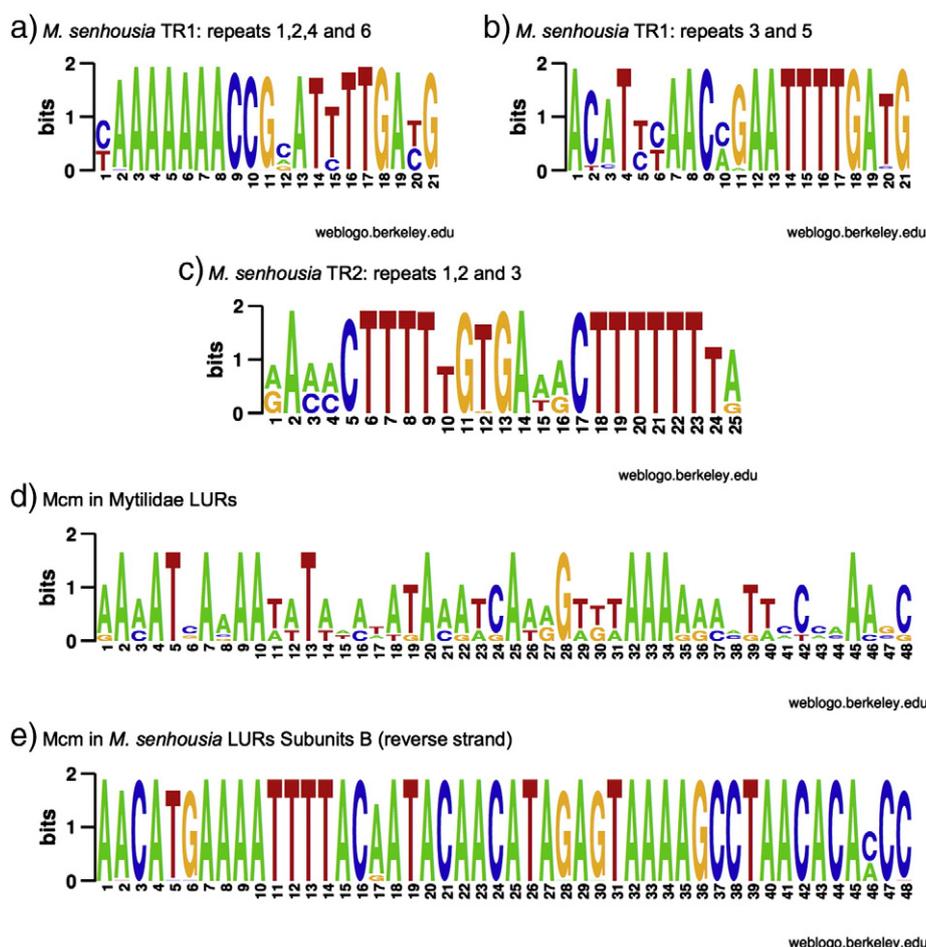
## FLUR



## MLUR



**Fig. 1.** *M. senhousia* mt LURs organization, content and variability. FLUR and MLUR are represented in scale. Graphics over the LUR schemes represent the nucleotide diversity ( $\pi$ ) levels, among FLURs and among MLURs, calculated with sliding windows on alignments of complete sequences. DNA secondary structures' (fD and mD) position is indicated with black lines over the LURs. RNA secondary structures' (fR and mR) location is shown below the LURs with black arrows; the orientation of the arrows specifies the structures' direction. Mcm: conserved sequence motif found with MEME in Mytilid LURs. TR1 and TR2: FLUR tandem repeat series 1 and 2. V: MLUR 5' small variable region. \*: location of f11 FLUR 130 bp insertion. +: location of MLUR cruciform structure inside mR2.



**Fig. 2.** Consensus logos resulting from the alignments of: (a) TR1 repeats 1, 2, 4 and 6; (b) TR1 repeats 3 and 5; (c) TR2 repeats; (d) motif Mcm in all considered Mytilidae LURs; and (e) Mcm motif in *M. senhousia* LURs.

f11 FLUR (accession number: KC243363; see Fig. 1 for the insertion location). The LURs total length is scarcely affected by the single blocks length variations: FLUR length ranges from 4518 bp to 4643 bp, while MLUR from 2812 bp to 2854 bp. For details on LURs and single subunits length variability see Table S1 in Supplementary Tables.

M subunits are generally more conserved than F counterparts and, as a consequence, MLUR is more conserved than FLUR (p-distances  $\pm$  S.E.:  $0.011 \pm 0.001$  and  $0.028 \pm 0.002$ , respectively). Inside FLUR, no great difference is found between Rep Units 1 and 2 (p-distances  $\pm$  S.E.: Rep Unit 1,  $0.030 \pm 0.002$ ; Rep Unit 2,  $0.032 \pm 0.002$ ; overall,  $0.030 \pm 0.002$ ). Among the largest conserved blocks shared between MLUR and FLUR, the less variable ones are B Subunits (p-distances  $\pm$  S.E.: M,  $0.007 \pm 0.001$ ; F,  $0.027 \pm 0.002$ ) and  $\gamma$  Subunits (p-distances  $\pm$  S.E.: M,  $0.015 \pm 0.003$ ; F,  $0.026 \pm 0.005$ ). Even between FLUR and MLUR, these subunits are the most conserved (p-distances  $\pm$  S.E.: B Subunits,  $0.061 \pm 0.004$ ;  $\gamma$  Subunits,  $0.159 \pm 0.014$ ). The small motifs  $\alpha$  at the beginning of A-type Subunits have even lower values (overall p-distance  $\pm$  S.E.:  $0.060 \pm 0.022$ ), but their length (37–47 bp) is not comparable to that of the larger subunits. For detailed p-distance values of all other subunits see Tables S2 and S3 in Supplementary Tables.

Graphical representation of DNAsp sliding windows analysis is given in Fig. 1. In FLUR,  $\pi$  has its maximum in correspondence of Subunits A1 and A2 (0.082 and 0.086, respectively). Moving towards the center of the B Subunits,  $\pi$  decreases to a minimum of 0.000 in B1 and of 0.004 in B2. From this point on, towards the  $\delta$  Subunits, the diversity rises again but has a relative minimum inside  $\gamma$  Subunits (0.010 in  $\gamma 1$  and 0.008 in  $\gamma 2$ ). In the MLUR a region of low diversity (minimum value of 0.000) between V and Subunit A' precedes the highest peak of  $\pi$ , 0.037. Proceeding towards the center of Subunit B  $\pi$  decreases, reaching

a large region of zero diversity. Variability remains low in the last part of the MLUR, with relative high peaks at the beginning of each Subunit  $\gamma$ .

A sequence motif was found by MEME in every Mytilidae LUR considered, hence it is here called Mcm (for "Mytilidae conserved motif"; Figs. 2d and e). Its length is 48 bp and it has an E-value of  $9.7E-23$ ; moreover, this motif is found in the same relative position in the LUR of both mt genomes belonging to the same DUI species. In *M. senhousia* Mcm is found inside B Subunits (Fig. 1), and in *Mytilus* inside the LURs conserved domain. In species belonging to the subfamily Mytilinae, i.e. *Mytilus* spp. and *Perna viridis*, this motif is found on the forward strand while in *M. senhousia* (subfamily Crenellinae) it is on the reverse strand. See Table 2 for the sequences forming the motif and their respective p-values. GOMO assigned GO terms related to mitochondria or to genome regulation only in the databases "*Drosophila melanogaster*" and "*Arabidopsis thaliana*". These results include the GO terms "mitochondrion", "transcription factor activity" and "regulation of transcription, DNA dependent". See Table S4 in the Supplementary Tables for every database top specific predictions.

For the DNA secondary structure characterization, we chose to consider only those structures that had the same conformation in every F and M LUR. When a structure did not respect this criterion, i.e. was present only in some specimens, it was discarded from the analysis, since no discussion could obviously be made on non-conserved features. See Fig. 1 for DNA secondary structures' location in the LURs and Supplementary Materials 1 for all structures' detailed length, shape and variable sites. In FLUR eight different secondary structures (hence called fD for female DNA structures) were identified. Some of them are found in both Rep Units, thus the total number of structures identified in the FLUR is thirteen. One structure (fD1) is located in the

**Table 2**  
Sequences found by MEME in all Mytilidae LURs forming the motif Mcm. Abbreviations: Medu = *Mytilus edulis*, Mcal = *Mytilus californianus*, Msen = *Musculista senhousia*, Pvir = *Perna viridis*.

Sequence name	Strand	p-Value	Sites
Medu MLUR	+	6.28E–23	AAAATCAGAATATATATATAAAATCAAGGTTAAAAAAATCCCAAAGC
Medu FLUR	+	7.79E–23	AAAATCAAAATATATATATAAAATCAAGGTTAAAAAAATCCCAAAGC
Mcal FLUR	+	2.76E–22	AAAATCAAAAAATAAATATAAAATCAAGGTTAAAAAAATCCCAAAGC
Mcal MLUR	+	2.47E–21	AAAATCACAATATATATAAAATCAAGGTTAAAAAAATCCCAAAGC
Msen MLUR	–	3.86E–17	AACATGAAAATTTTACGATACAACATAGAGTAAAAAGCCTAACACAACC
Msen FLUR	–	8.03E–17	AACATGAAAATTTTACAATACAACATAGAGTAAAAAGCCTAACACACCC
Pvir LUR	+	2.36E–14	GAAATAAAAAAATAGAAATGAAGTGAAGGTAAAAAGAACCTGTGGAAAG

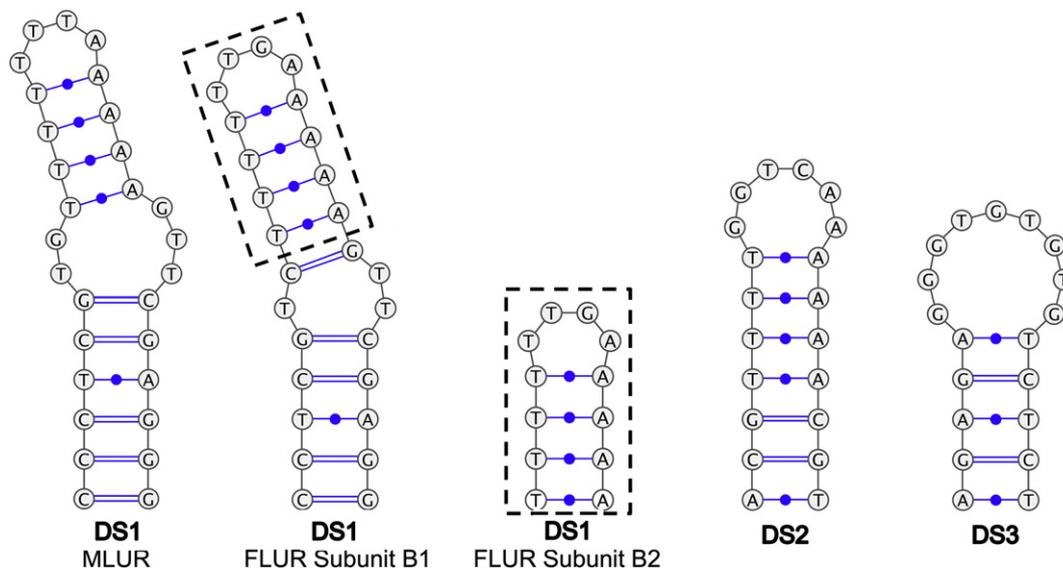
5' spacer, three (fD2 ÷ 4) in both B Subunits, one (fD5) in both  $\gamma$  Subunits, one (fD6) in both  $\delta$  Subunits, one (fD7) only in Subunit  $\delta$ 2 and one (fD8) in A'. The female structures' length ranges from 6 bp (fD1) to 180–181 bp (fD3). The MLUR contains thirteen different structures (called mD for male DNA structures): some of them are repeated in  $\gamma$  Subunits, so the total number of structures in MLUR is seventeen. Two structures (mD1 and mD2) are located in the 5' spacer, one (mD3) in A', six (mD4 ÷ 9) in Subunit B, three (mD10 ÷ 12) in Subunit  $\gamma$ 1, two (mD11 and mD13) in Subunits  $\gamma$ 2 and  $\gamma$ 3 and one (mD14) between Subunit  $\gamma$ 4 and the 3' spacer. In MLUR, the shortest structures are 8 bp long (mD12 and mD13), while the longest is 143 bp (mD7). Three short stem-and-loops are shared between FLUR and MLUR (called DS for DNA shared structures) (Fig. 3) and are all located in B Subunits. DS1 corresponds to fD2 and is part of mD6; because of mutations, in FLUR Subunit B2 only the last segment of DS1 is conserved between individuals. Mutations in DS1 are found only in the loop and in the middle bulge. DS2 is part of fD3 and mD7 and its structure shows no mutations. DS3 is comprised in mD8 and corresponds to fD4; like DS2, this hairpin has no mutations.

FLUR RNA structures are named fR and MLUR structures mR. See Fig. 1 for RNA structures location in LURs, Table S5 in Supplementary Tables for statistics of structures and Supplementary Materials 2 for their shape and details. In FLUR only two RNA structures reached the cut-off values: fR1 in B subunits and fR2 between Subunits  $\delta$ 1 and A2. Seven supported structures have been found inside the MLUR: mR1 and mR2 between the 5' spacer and A', mR3 in Subunit B, mR4 to mR7 between and inside  $\gamma$  Subunits. No RNA structures are shared between FLUR and MLUR.

Inside MLUR, three regions can fold into a similar shape both at DNA and RNA levels, on the same or on the reverse strand (see annotations in Supplementary Materials 1 and Supplementary Materials 2). mD2 can

be found, reversed and with a different bulge, inside mR1 and also inside mR2 with the exact same form and orientation. A large stem-and-loop of mD7 is very similar to mR3: apart from the opposite orientation, the extensions of the loop and bulges are the sole differences between the two structures. Lastly, mD11 is found on the reverse strand with the same form inside mR6. All *M. senhousia* structures were compared to those hypothesized in *M. edulis* and *M. galloprovincialis* CRs (Cao et al., 2004). Even though the overall shape of the structures is different, some loop and/or bulge sequence in *M. senhousia* structures somewhat matches parts of *Mytilus* ones, in a few cases even completely (annotated in Supplementary Materials 1 and Supplementary Materials 2). A further parallelism between *M. senhousia* and *Mytilus* structures regards their relative location inside the LURs. In *Mytilus*, cruciform structures are found in the first domain of the CR and a large stem-and-loop is located in the middle of the central domain (Cao et al., 2004). A comparable situation is found in *M. senhousia*, where in MLUR Subunit A', a part of mR2 forms a large cruciform structure (Fig. 4) and, in both FLUR and MLUR, the largest DNA structures (i.e. fD3 and mD7) are found in the middle of B Subunits (Fig. 1); moreover, these latter DNA structures are associated to the large RNA stem-and-loops of fR1 and mR3 (Figs. 1 and 4).

AT-skew values of all analyzed mt genomes are enlisted in Table S6 in Supplementary Tables. In *M. senhousia* the AT-skew values distribution follows a similar pattern in both genomes. In the M genome most of the genes with the highest absolute values (i.e. from *cox1* to *nad1*) are located downstream the LUR, while genes with the lowest absolute values (from *nad4L* to *nad3*) are located upstream the LUR. In the F genome AT-skew distribution pattern is similar, but it is less distinct than that of the M genome. Also in *Mytilus* F and M genomes, the distribution of these values is comparable. *Cytb*, which is immediately downstream the CR, has the highest absolute skew value in the F genomes



**Fig. 3.** DNA secondary structures shared by *M. senhousia* FLUR and MLUR.

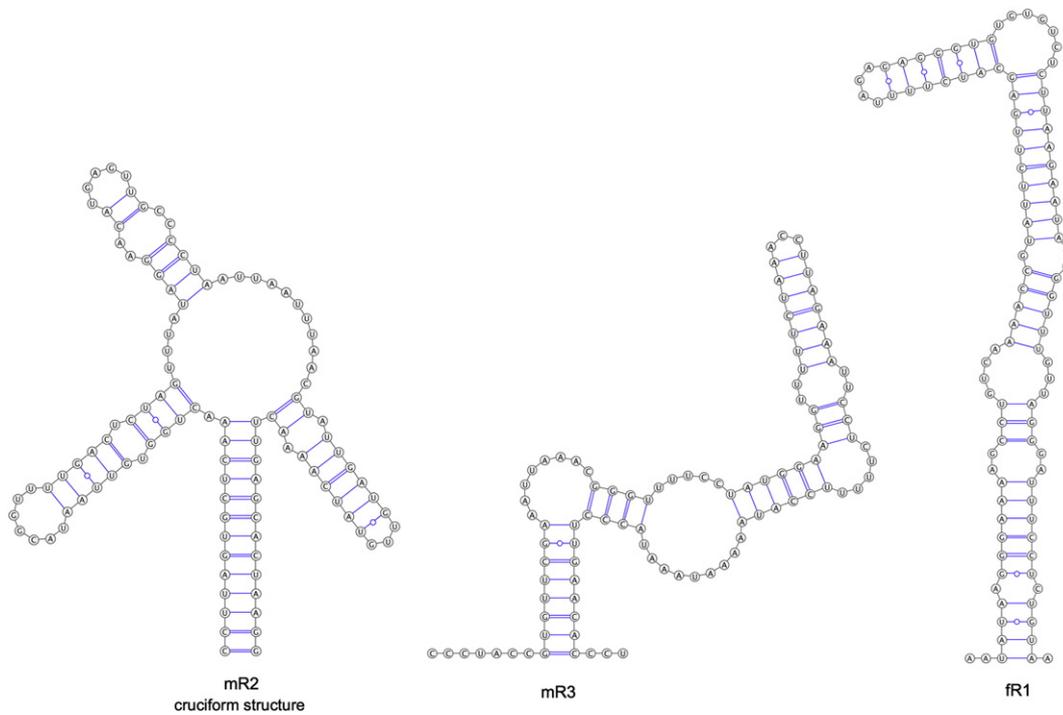


Fig. 4. *M. senhousia* LURs secondary structures whose shape and relative position in the LURs recalls those of *Mytilus* CRs found by Cao et al. (2004).

(except for *M. galloprovincialis*, where it has the second highest value) and it has always the second highest value in the M genomes. In both F and M genomes, the genes with the lowest values are located in the region between *nad3* and *nad5*, broadly on the opposite side of the CR. In *P. viridis*, the genes showing the highest values are *cox2* and *cytb*, which are downstream the LUR; *cox1*, with the third highest value, is upstream the LUR. Two of the genes with the lowest values, *nad5* and *cox3*, are found upstream the LUR alongside the rRNA genes, which are roughly on the opposite side of *cytb* and *cox2*.

The evolutionary model chosen for the single-partition analysis was GTR + G, while those of the multi-partitions were the following: TVM + G for A + A', GTR + G for B and HKY for both  $\gamma$  and  $\delta$ . Both AIC and BF tests favored the multi-partition analysis tree compared to that of the single-partition ( $\Delta\text{AIC} = 85.9$ ; BF = -135.9). The topologies of the resulting trees were the same: Fig. 5 shows the multi-partitions analysis tree. Unresolved nodes ( $\leq 95\%$ ) were collapsed: these polytomies are most probably due to the presence of long strings of ambiguous nucleotides (as in f7 Rep Unit 2, f8 Rep Units 1 and 2, f9 Rep Unit 2 and f12 Rep Unit 2; respective accession numbers: KC243359, KC243360, KC243361, KC243364) or large insertions (as in f11 Rep Unit 2; accession number: KC243363).

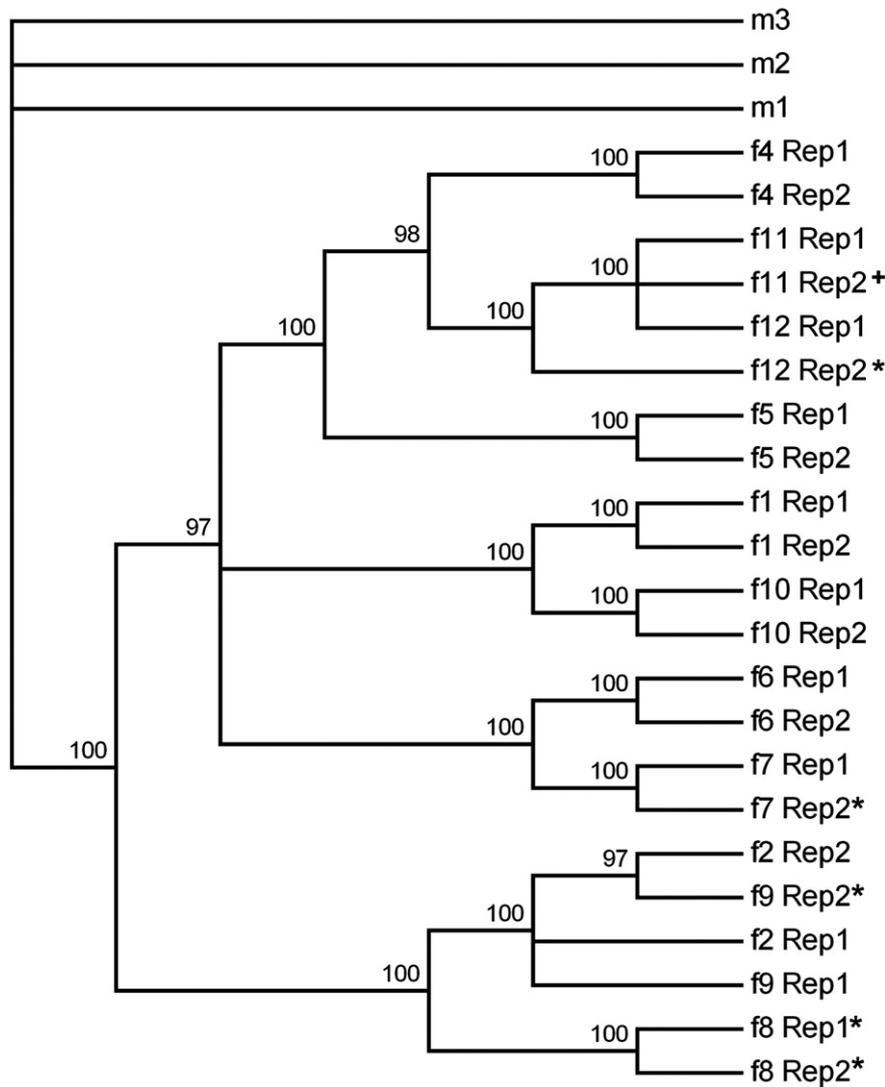
## 4. Discussion

### 4.1. FLUR and MLUR contain the CR of *M. senhousia* mt genomes

In both sex-linked mt genomes of *M. senhousia* a large unassigned region is present, named FLUR for F-mtDNA and MLUR for M-mtDNA (Passamonti et al., 2011). The two LURs have different lengths and share three main subunits, the A-type Subunits (A and A' in the FLUR, A'' in the MLUR), B Subunits and  $\gamma$  Subunits. A-type Subunits have a small motif, named  $\alpha$ , in their 5' end. Another subunit type, named  $\delta$ , is typical of the FLUR and it is not found in the MLUR. In the FLUR, the block composed of Subunits A, B,  $\gamma$  and  $\delta$  is repeated in tandem two times: the two repetitive blocks are named Rep Units 1 and 2. In the MLUR, Subunit  $\gamma$  is repeated in tandem four times, but the fourth copy is truncated to 17 bp. Both LURs start with a spacer in 5' with no similarity between them, and the MLUR has a small, unannotated spacer

also at the 3' end. In both LUR types, the most conserved regions in terms of nucleotide sequence are Subunits B and  $\gamma$ , both among FLURs and MLURs and between LURs (Table S2 in Supplementary Tables). These subunits have a high folding potential, both at DNA and RNA levels. Indeed, although the shape of the LURs two major DNA structures (fD3 and mD7) is different (Supplementary Materials 1), their relative location is the same (Fig. 1); moreover, some stem-and-loops, the DS (Fig. 3), are found with the same form in both F and M B Subunits. In the MLUR, three regions may fold into a similar shape both at DNA and at RNA levels (mD2, mR1 and mR2; Supplementary Materials 1 and 2): this last finding can be simply due to the fact that softwares for nucleic acid folding use similar algorithms but, given that only these structures were found at both DNA and RNA, it may also be taken as additional support for the existence of these structures.

Also, the analogies between the organization of *M. senhousia* LURs secondary structures and that of *M. edulis* and *M. galloprovincialis* F- and M-mtDNA CRs are noteworthy. In all three species, part of the central-most domain, i.e. the conserved domain of *Mytilus* (Cao et al., 2004) and Subunit B of *M. senhousia*, may fold into a large structure: a stem-and-loop in *Mytilus* (Cao et al., 2004) and, in *M. senhousia*, fD3 and mD7 at DNA level and fR1 and mR3 at RNA level (Figs. 1, 4, Supplementary Materials 1 and 2). Additionally, the region upstream the most conserved domains, variable between the F and M lines both in *Mytilus* (the first variable domain) and *M. senhousia* (Subunits A and A''), may form cruciform structures. In *Mytilus*, the first variable domain may form one such structure in the F line and two in the M (Cao et al., 2004), while in *M. senhousia* only MLUR Subunit A'' has a supported cruciform structure (found inside mR2; Figs. 1, 4, Supplementary Materials 2). The correspondence in position and shape of some secondary structures between *M. senhousia* and *Mytilus* may point to a conserved function of their mt genomes LUR. Another intriguing finding is the sequence similarity of some exposed parts of the structures found in *M. senhousia* and *Mytilus* (Supplementary Materials 1 and 2). Loops and bulges of secondary structures, being single-stranded, are more prone to mutations in long evolutionary times than stems (or double-stranded regions in general), which are under the constraint of legitimate base-pairing: the sequence conservation of single-stranded DNA in distantly related species may indicate some kind of selective constraint acting on these regions (Hoede et al., 2006).



**Fig. 5.** Bayesian tree of *M. senhousia* FLUR Rep Units. Rep1 and Rep2: FLUR Rep Unit 1 and 2, respectively. m1, m2 and m3: MLUR sequences used as outgroup. \*: sequences with strings of ambiguous nucleotides. +: sequence with a 130 bp insertion.

*M. senhousia* B Subunits also contain a motif, called Mcm (yellow box in Fig. 1), that is found conserved in the mt LURs of other Mytilid species (Table 2, Fig. 2a). This motif seems to be related to transcription regulation and is located before the DS, the conserved stem-and-loops contained in the largest DNA structures (Figs. 1, 3); in the FLUR this motif is also preceded by two series of small tandem repeats, TR1 and TR2 (Fig. 1). Finally, the central part of B Subunits, where all of the abovementioned conserved features are found, is the less variable region of both LUR types (see  $\pi$  trends in Fig. 1). Taken together, our analyses suggest that this region may have some importance in the mtDNAs of *M. senhousia*, regardless of it being the F or the M, since its sequence and structural conservation are high in FLUR and MLUR and between them. Other subunits do not show comparable conserved features between the two LUR types.

The AT-skew analysis we performed on *M. senhousia* and other Mytilid mt genomes evidences a similar distribution of values in all of them: genes with the highest absolute values are located immediately downstream the LUR. Following the AT-skew rationale (Breton et al., 2009), this indicates that they spend more time in a single-strand state longer than the other genes during mtDNA duplication: if we consider the strand-displacement replication model (Clayton, 2003), this could mean that these genes are near the *oriH*. The lowest-scoring

genes, on the other hand, are often located relatively far from the LUR, which means they are placed away from the *oriH* and, probably, located near the *oriL*.

Considering all these findings, we are quite confident that FLUR and MLUR contain the CR of *M. senhousia* F and M mt genomes, respectively. Indeed CRs, other than usually being the largest non-coding regions of a mt genome, contain the *oriH* and secondary structures, repeated units and sequence motifs that may function as binding sites for replication and/or transcription of the mtDNA (Scheffler, 2008). The parallelism between the organization of *M. senhousia* LURs and *Mytilus* CRs is remarkable, not only because of comparable secondary structures, but also for the order of conserved blocks. In *M. senhousia* M CR and in both FLUR Rep Units, the more conserved B Subunits are preceded at 5' by the more variable A-type Subunits; the 3' ends of M CR and FLUR Rep Units are also different, as they are composed in the first case by three  $\gamma$  Subunits and in the latter by one Subunit  $\gamma$  and a Subunit  $\delta$  (the latter specific of the FLUR and not found in the M CR). This situation is strikingly similar to that of *Mytilus* CRs, where the central domain, conserved between F and M CRs, is flanked by two highly variable domains (Cao et al., 2004). This suggests that, at least in Mytilids, the central part of the CR may contain the key signals involved in mtDNA replication and transcription and that, in the case of DUI

species, the 5' and 3' ends were those more subject to change when the F and M mtDNA lines began to diverge.

#### 4.2. *M. senhousia* FLUR contains two tandemly duplicated CRs

Many Metazoan mt genomes possess duplicated or multiple CRs and, in many cases, the duplication is located relatively far from the CR in the original position. This feature is also found in some Bivalve mtDNA. In *Paphia amabilis* (Bivalvia Veneridae) (accession number: JF969276; Xu et al., 2012) two large non-coding regions were found, one of which is clearly a degenerated copy of the other; interestingly, the copy in the original position is the degenerated one, while the one in the novel position maintains the strongest similarities with other Paphia mt LURs. *Lucinella divaricata* (Bivalvia Lucinidae) mt genome (accession number: EF043342; Dreyer, Steiner and Satler, unpublished) has two large non-coding regions of almost the same length and with a high sequence similarity. These regions were not annotated by the Authors and it is also not possible to know whether this is a common feature of *L. divaricata* mtDNAs or not. Until new evidence, this latter case could be similar to that of an M mt genome belonging to the DUI species *M. galloprovincialis* (accession number: AY363687) sequenced by Mizi et al. (2005). This genome was found to have two CRs, one of them in a novel position, although this has to be considered an exception for this species because all other complete M mt genomes only have one CR. The F mt genome of another DUI species, *Mytilus trossulus*, possesses two CRs separated by *tRNA-Gln* (Breton et al., 2006; Cao et al., 2009). Comparisons with mtDNAs of the same species and of other *Mytilus* by Cao et al. (2009) resulted on the observation that the first CR derives from a *M. trossulus* M mt genome, while the original F CR is the one placed 3' of the *tRNA-Gln*: the duplication in this species is clearly the result of a recombination event between the M and an ancestral (maybe now extinct) F mt genome. The Authors proposed a partial functionality of both CRs in this mtDNA (Cao et al., 2009). For the first three mentioned mt genomes (*P. amabilis*, *L. divaricata* and the exceptional *M. galloprovincialis* M-mtDNA), the duplication could have been produced either by a duplication of a segment containing the CR with subsequent loss of its flanking parts, or by a recombination between mt genomes of the same type. In *M. senhousia* FLUR the duplication of the Rep Units is in tandem. This kind of duplication is generally explained with slippage errors during mtDNA replication (Boore, 2000; Boore and Brown, 1998): in this particular case, the duplication has reached fixation. Both FLUR Rep Units have an organization comparable to the whole M CR (Fig. 1) and contain the same features found in the M CR: thus, given the low variability between Rep Units 1 and 2 (Table S2 in Supplementary Tables), we propose that none of the two Units is degenerating and that both of them are functional CRs.

Moreover, the Bayesian analysis showed that F CRs belonging to the same FLUR (i.e. from the same individual) cluster together in a "specimen-specific" fashion (Fig. 5). This clustering means that CRs from a single FLUR are more similar to each other than to those from other FLURs and, therefore, that CRs in the same FLUR evolve in concert. Concerted evolution of duplicated mt CR sequences inside a single individual has already been observed in animal mt genomes (Eberhard et al., 2001; Morris-Pocock et al., 2010; Ogoh and Ohmiya, 2007; Schirtzinger et al., 2012; Tatarenkov and Avise, 2007; Verkuil et al., 2010). The gene conversion model commonly used to explain this phenomenon (Kumazawa et al., 1998) may easily apply also to *M. senhousia* F CRs. When in a single-stranded state, the two CR sequences may pair in a non-homologous way (e.g. CR 1 with CR 2) and mismatch correction mechanisms may homogenize the mutations between the two strands, thus making the CR sequences very similar. However, we cannot tell from our results if this homogenization happens only between complementary strands of the same molecule or also between strands of different F-mtDNA molecules inside the same mitochondrion. If the latter scenario is true, then the mitochondrial bottleneck occurring during

germline formation may also increase sequence homogenization of the F CRs.

Our current report of a CR duplication in the F-mtDNA of *M. senhousia* is unique because: (1) the comparison of multiple sequences from different individuals allowed us to consider this duplication as a stable feature, at least in the population we analyzed and maybe in the whole species; (2) the two copies of the CR are in tandem, a situation not found in other animal mt genomes with CR duplication; (3) the duplication in this DUI species may not be the result of a recombination between an F and an M mt genome (like in *M. trossulus* F-mtDNA) but rather the result of an intra-genome duplication; (4) our analyses suggest that both copies are most probably functional. *M. senhousia* F mt genome thus adds to the list of Metazoan mtDNAs with duplicated CRs evolving in concert, the first supported report of this kind for a Bivalve species.

#### 4.3. CRs organization and evolution in relation to DUI

F and M CRs conserved blocks are organized in a comparable way, with the A-type Subunits showing more differences between the two LURs than Subunits B and  $\gamma$ . The duplications of the CR in the FLUR and of the  $\gamma$  Subunits in the M CR are probably derived states, but we cannot indubitably tell which of the two conditions is closest to the ancestral LUR state. Moreover, we do not know the ancestral position of *tRNA-Glu*, which is different between F and M mt genomes (Passamonti et al., 2011). The movement of this tRNA from a position to another may have influenced the duplications of F CRs or of M CR  $\gamma$  Subunits (Gissi et al., 2008; Stanton et al., 1994), but we cannot infer its basal position from the currently available complete mt genomes of *Mytilus* and *P. viridis*, since they have very different gene orders.

Accepting a single origin of DUI for all bivalves, the analogies between *M. senhousia* F and M CRs could be explained by a masculinization/route-reversal event which may have been triggered by an inter-genomic recombination. In a route-reversal event, a F mt genome becomes sperm transmitted and substitutes the "old" M, initially resetting to almost zero the sequence divergence between the two lines (Zouros, 2013): the oldest the reversal, the major are the differences between F and M mt genomes. A recent route-reversal, for example, has been observed in *M. galloprovincialis* by Venetis et al. (2007), who obtained the complete sequence of a newly-masculinized mtDNA, the so called C genome (accession number: DQ399833). This genome is composed of the coding sequences of a F-mtDNA and a mosaic CR composed of F and M domains: this allowed the Authors to hypothesize that the acquisition of M CR sequences by a F mt genome caused its invasion of the M route of transmission. This phenomenon may also have happened in the past during the radiation of the *M. edulis* complex, because the F and M mt genomes of the three species *M. edulis*, *M. galloprovincialis* and *M. trossulus* cluster together in a gender-joining fashion (Zouros, 2013): indeed, the major differences between *Mytilus* F and M mt genomes lie in the CR and not in the coding parts of the mtDNA. However, a *Mytilus*-like masculinization event alone cannot account at the same time for the similarities between the content of *M. senhousia* F and M CRs (compared to *Mytilus* CRs), the absence of M-specific sequences in the MLUR (apart from the small 5' and 3' spacers, all of its subunits can be found in the FLUR) and the presence of lineage-specific coding sequences in the two mt lines (Breton et al., 2011a; Milani et al., 2013; Passamonti et al., 2011). Instead, if we consider independent origins of DUI, the similarities between F and M CRs subunits of *M. senhousia* can be explained by their common origin from the CR of the ancestral maternally transmitted mtDNA, while their different organization, the lineage-specific ORF in the F and the duplicated *cox2* in the M may have been acquired after the gain of DUI.

The partial sequences of *M. senhousia* and of other three DUI species (*Brachidontes exustus*, *Geukensia demissa* and *Mytella charruana*) cluster in the Mytilid phylogeny in a taxon-joining pattern, i.e. the F and M sequences of a species cluster separately from those of the others (Alves

et al., 2012; Zouros, 2013). This further complicates the reconstruction of the affinities between DUI Mytilids, as we have to consider either multiple route-reversals or multiple independent origins of DUI for each of these species (or higher taxa) to explain their clustering pattern. However, to obtain a clearer view of the evolution of DUI in Mytilids (and in Bivalves in general) more complete mt genomes are needed, as the analyses of partial sequences surely overlook important features that could help resolve deep relationships between taxa (Doucet-Beaupré et al., 2010).

#### 4.4. Variability pattern of CRs

About the influence of the CR duplication in *M. senhousia* F-mtDNA evolution, in this study we found in the LURs the same variation pattern found in coding genes by Passamonti (2007): overall, the FLUR and the single F CRs are indeed more variable than the M CR.

The results of Passamonti (2007) were unusual, because in all other studied DUI taxa the M mt genome was shown to carry more mutations than the F, suggesting its faster evolution rate (Zouros, 2013) (but see Ghiselli et al., 2013 for a discussion on the use of PCR-based methods in evaluating the variability of DUI mt genomes). In the case of *M. senhousia*, the higher variability of F protein-coding genes was explained by the Author with the probable female-biased composition (i.e. the presence of more F than M haplotypes) of the founder population that first invaded the region of the Adriatic Sea from where the specimens used were collected. This founder-effect explanation may easily account also for the variability pattern of the LURs (the LUR sequences of this study were retrieved from individuals of the same Italian population), but it has also been proposed that the presence of two CRs in an mt genome may enhance its replication and mutation rates (Kumazawa et al., 1998).

In DUI species it has been proposed that the M-mtDNA, not the F, is the faster replicating one, to account for its capacity of invading the male germline (Cogswell et al., 2006; Ghiselli et al., 2011). If we accept the hypothesis that the higher variability of *M. senhousia* F-mtDNA is the outcome of a higher replication rate, still M-type succeeds in invading the germline in males, so that a higher duplication rate is not necessary *per se* to a successful invasion of the male germline by M-mtDNA. If the faster replication rate of the F-mtDNA is proved true, then the duplication of the CR in this mt genome has influenced only the evolution of F line, as the fate of M mitochondria and, in general, the DUI system of *M. senhousia* are unaffected by this event. Nonetheless, until new sequences of *M. senhousia* specimens from the species original geographical range (Asian Pacific coasts) are available, the F-mtDNA higher polymorphism can still be explained as an artifact due to sampling of a non-representative, relatively recent population.

## 5. Conclusions

The detailed characterization of *M. senhousia* mtDNA LURs in our study brings new information about the organization of Mytilids CRs. The presence of shared features in two distinct subfamilies, Crenellinae and Mytilinae, demonstrates that signals involved in the basic functions of the mtDNA are retained over evolutionary times even among the fast-evolving mt genomes of DUI species. Using the support of multiple sequences, we also provide evidence for the presence of two functional CRs in the FLUR, making *M. senhousia* F-mtDNA the first confirmed Bivalve mt genome with this particular feature. The variability patterns of F and M CR sequences can be explained as a sampling artifact, but more *M. senhousia* mt sequences from its original area are required to confirm or refute the observed polymorphism, which contrasts with the current knowledge on the topic. Moreover, we discussed on how *M. senhousia* mt CRs affinities can be explained in a broad DUI context: from this latter point emerges the pressing need for more complete mt genome sequences from as many DUI species as possible, to help our comprehension of DUI origin and evolution in Bivalves.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.gene.2013.12.005>.

## Conflict of interest statement

The Authors declare that no conflict of interest exists.

## Acknowledgments

We would like to thank Federico Plazzi for his precious help in the Bayesian analysis. This work was supported by the University and Research Italian Ministry (MIUR PRIN09, grant number 2009NWXMX002) and the “Canziani Bequest” fund (University of Bologna, grant number A.31.CANZELSEW).

## References

- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrox, B.N., Caski, F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, p. 267.
- Alves, F.A., Beasley, C.R., Hoeh, W.R., da Rocha, R.M., Simone, L.R., Tagliaro, C.H., 2012. Detection of mitochondrial DNA heteroplasmy suggests a doubly uniparental inheritance pattern in the mussel *Mytella charruana*. Rev. Bras. Biociênc. 10 (2), 176.
- Bailey, T.L., et al., 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37 (Suppl. 2), W202–W208.
- Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580.
- Boore, J.L., 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. Comparative Genomics. Springer, Netherlands 133–147.
- Boore, J.L., Brown, W.M., 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr. Opin. Genet. Dev. 8, 668–674.
- Breton, S., Burger, G., Stewart, D.T., Blier, P.U., 2006. Comparative analysis of gender-associated complete mitochondrial genomes in marine mussels (*Mytilus* spp.). Genetics 172 (2), 1107–1119.
- Breton, S., Doucet-Beaupré, H., Stewart, D.T., Hoeh, W.R., Blier, P.U., 2007. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? Trends Genet. 23 (9), 465–474.
- 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., Ghiselli, F., Passamonti, M., Milani, L., Stewart, D.T., Hoeh, W.R., 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.
- Buske, F.A., Bodén, M., Bauer, D.C., Bailey, T.L., 2010. Assigning roles to DNA regulatory motifs using comparative genomics. Bioinformatics 26 (7), 860–866.
- Cao, L., Kenchington, E.L., Zouros, E., Rodakis, G.C., 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 (3), 1045–1056.
- Clayton, D., 2003. Mitochondrial DNA replication: what we know. IUBMB Life 55 (4–5), 213–217.
- Cogswell, A., Kenchington, E.L., Zouros, E., 2006. Segregation of sperm mitochondria in two and four cell embryos of the blue mussel *Mytilus edulis*: implications for the mechanism of doubly uniparental inheritance of mitochondrial DNA. Genome 49, 799–807.
- Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190.
- Darty, K., Denise, A., Ponty, Y., 2009. VARNA: interactive drawing and editing of the RNA secondary structure. Bioinformatics 25 (15), 1974.
- 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 (1), 50.
- Eberhard, J.R., Wright, T.F., Bermingham, E., 2001. Duplication and concerted evolution of the mitochondrial control region in the parrot genus *Amazona*. Mol. Biol. Evol. 18, 1330–1342.
- 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., 2013. Structure, transcription and variability of metazoan mitochondrial genome: perspectives from an unusual inheritance system. Genome Biol. Evol. 5 (8), 1535–1554.
- Cissi, C., Iannelli, F., Pesole, G., 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101, 301–320.

- Gruber, A.R., Neuböck, R., Hofacker, I.L., Washietl, S., 2007. The RNAz web server: prediction of thermodynamically stable and evolutionarily conserved RNA structures. *Nucleic Acids Res.* 35 (Suppl. 2), W335–W338.
- Hoede, C., Denamur, E., Tenaillon, O., 2006. Selection acts on DNA secondary structures to decrease transcriptional mutagenesis. *PLoS Genet.* 2 (11), e176.
- Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17 (8), 754–755.
- Kass, R.E., Raftery, A.E., 1995. Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795.
- Kumazawa, Y., Ota, H., Nishida, M., Ozawa, T., 1996. Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. *Mol. Biol. Evol.* 13 (9), 1242–1254.
- Kumazawa, Y., Ota, H., Nishida, M., Ozawa, T., 1998. The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. *Genetics* 150, 313–329.
- Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25 (11), 1451–1452.
- Milani, L., Ghiselli, F., Guerra, D., Breton, S., Passamonti, M., 2013. A comparative analysis of mitochondrial ORFs: new clues on their origin and role in species with doubly uniparental inheritance of mitochondria. *Genome Biol. Evol.* 5 (7), 1408–1434.
- Mizi, A., Zouros, E., Moschonas, N., Rodakis, G.C., 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.
- Morris-Pocock, J.A., Taylor, S.A., Birt, T.P., Friesen, V.L., 2010. Concerted evolution of duplicated mitochondrial control regions in three related seabird species. *BMC Evol. Biol.* 10, 14.
- Nuin, P., 2007. MrMTgui: Cross-platform Interface for ModelTest and MrModeltest. Available: <http://www.genedrift.org/mtgui.php>.
- Obata, M., Sano, N., Komaru, A., 2011. Different transcriptional ratios of male and female transmitted mitochondrial DNA and tissue-specific expression patterns in the blue mussel, *Mytilus galloprovincialis*. *Dev. Growth Differ.* 53, 878–886.
- Ogoh, K., Ohmiya, Y., 2007. Concerted evolution of duplicated control regions within an ostracod mitochondrial genome. *Mol. Biol. Evol.* 24, 74–78.
- Passamonti, M., 2007. An unusual case of gender-associated mitochondrial DNA heteroplasmy: the mytilid *Musculista senhousia* (Mollusca Bivalvia). *BMC Evol. Biol.* 7 (Suppl. 2), 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 (2), 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 (1), 442.
- Plazzi, F., Passamonti, M., 2010. Towards a molecular phylogeny of mollusks: bivalves' early evolution as revealed by mitochondrial genes. *Mol. Phylogenet. Evol.* 57, 641–657.
- Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference using mixed models. *Bioinformatics* 19, 1572–1574.
- Sano, N., Obata, M., Komaru, A., 2007. Quantitation of the male and female types of mitochondrial DNA in a blue mussel, *Mytilus galloprovincialis*, using real-time polymerase chain reaction assay. *Dev. Growth Differ.* 49, 67–72.
- Scheffler, I.E., 2008. *Mitochondria*. Wiley-Liss, Hoboken, NJ.
- Schirtzinger, E.E., et al., 2012. Multiple independent origins of mitochondrial control region duplications in the order Psittaciformes. *Mol. Phylogenet. Evol.* 64 (2), 342–356.
- Stanton, D.J., Daehler, L.L., Moritz, C.C., Brown, W.M., 1994. Sequences with the potential to form stem-and-loop structures are associated with coding-region duplications in animal mitochondrial DNA. *Genetics* 137, 233–241.
- Swofford, D.L., 2002. PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods) Version 4.0b10. Sinauer Associates, Sunderland.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28 (10), 2731–2739.
- Tatarenkov, A., Avise, J.C., 2007. Rapid concerted evolution in animal mitochondrial DNA. *Proc. R. Soc. B* 264, 1795–1798.
- Venetis, C., Theologidis, I., Zouros, E., Rodakis, G.C., 2007. A mitochondrial genome with a reversed transmission route in the Mediterranean mussel *Mytilus galloprovincialis*. *Gene* 406, 79–90.
- Verkuil, Y.I., Piersma, T., Baker, A.J., 2010. A novel mitochondrial gene order in shorebirds (Scolopacidae, Charadriiformes). *Mol. Phylogenet. Evol.* 57, 411–416.
- Xu, X., Wu, X., Yu, Z., 2012. Comparative studies of the complete mitochondrial genomes of four *Paphia* clams and reconsideration of subgenus *Neotapes* (Bivalvia: Veneridae). *Gene* 494 (1), 17–23.
- Young, N.D., Healy, J., 2003. GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinform.* 4, 6.
- Zouros, E., 2013. Biparental inheritance through uniparental transmission: the doubly uniparental inheritance (DUI) of mitochondrial DNA. *Evol. Biol.* 40 (1), 1–31.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31 (13), 3406–3415.

# A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs

Sophie Breton<sup>1\*</sup>, Liliana Milani<sup>2\*</sup>, Fabrizio Ghiselli<sup>2</sup>, Davide Guerra<sup>2</sup>, Donald T. Stewart<sup>3</sup>, and Marco Passamonti<sup>2</sup>

<sup>1</sup> Département de Sciences Biologiques, Université de Montréal, 90 Avenue Vincent d'Indy, Montréal, Québec H2V 2S9, Canada

<sup>2</sup> Dipartimento di Scienze Biologiche, Geologiche ed Ambientali, University of Bologna, Via Selmi 3, 40126 Bologna, Italy

<sup>3</sup> Department of Biology, Acadia University, 24 University Avenue, Wolfville, Nova Scotia B4P 2R6, Canada

**Recent data from mitochondrial genomics and proteomics research demonstrate the existence of several atypical mitochondrial protein-coding genes (other than the standard set of 13) and the involvement of mtDNA-encoded proteins in functions other than energy production in several animal species including humans. These results are of considerable importance for evolutionary and cellular biology because they indicate that animal mtDNAs have a larger functional repertoire than previously believed. This review summarizes recent studies on animal species with a non-standard mitochondrial functional repertoire and discusses how these genetic novelties represent promising candidates for studying the role of the mitochondrial genome in speciation.**

## Beyond the powerhouse

Mitochondria have traditionally been viewed as bioenergetic organelles; their main function as described in most textbooks is ATP production, the universal currency of biological energy. These so-called ‘powerhouses’ of eukaryotic cells possess their own genome (mitochondrial DNA or mtDNA). In animals, mtDNAs are typically small (~16 kb), circular, maternally inherited molecules with an almost invariant gene content [13 genes coding for core subunits of the oxidative phosphorylation (OXPHOS) system and 24 structural RNAs for their translation; see [Glossary](#) and [Box 1](#)] and no introns [1,2]. However, mitochondria are important for more than only ATP production. There is strong evidence for their involvement in cell signaling and differentiation, fertilization, aging, and apoptosis [3–6]. Cases of animal species that deviate from the typical mitogenomic profile are starting to accumulate. For example, introns have been found in the mtDNAs of the simplest free-living animal *Trichoplax* (Placozoa), in sponges, cnidarians, and a polychaete worm [7–10]; atypical mitochondrial

genome architectures (e.g., linear or multicircle mtDNAs) have been reported in sponges, cnidarians, isopods, nematodes, and sucking lice [11–15]; the known size range for mitochondrial genomes is expanding (e.g., the largest mtDNA (46 985 bp) has been found in the bivalve mollusk *Scapharca broughtonii* [16] and the smallest (10 326 bp) in the ctenophore *Mnemiopsis leidyi* [17]); and the gene content of animal mtDNAs is being recognized as highly variable, mainly due to differences in the number of tRNA genes [2].

Despite the functional diversity of mitochondria and genome architecture variation, animal mtDNAs appear to be remarkably conserved, encoding a narrow set of functions, namely OXPHOS and translation. Yet, most models of mtDNA evolution consider neutral processes or negative selection as the major forces affecting the present pattern of mitochondrial sequence divergence among species [18,19]. Given their key role in energy metabolism, it is not surprising that several disease-causing mutations have been identified in mitochondrial genes [20–22]. Unexpectedly, it was demonstrated that the same mtDNA variant can be deleterious in a given environment and/or nuclear background but adaptive and under positive selection in another (e.g., [20]). This raises the possibility that the mitochondrial genetic system may be an important driver of speciation and, indeed, a role for mtDNA in the processes of reproductive isolation and speciation has been suggested by several authors (reviewed in [20–28]).

In contrast to the traditional view that animal mtDNA is a passive bystander of adaptive evolution owing to a limited functional repertoire, we propose that typical and atypical mitochondrial protein-coding genes may exhibit novel non-OXPHOS-related adaptations and/or extramitochondrial functions that support an even greater role in the process of speciation. Surprisingly, these findings have received limited attention in the literature even though they challenge the paradigm that mtDNA-encoded proteins are limited to a role in energy metabolism. We highlight here how atypical protein-coding genes and ORFans [open reading frames (ORFs) with unknown ontology and function] in animal mtDNAs expand the functional repertoire of mitochondria and their role in evolutionary processes.

Corresponding author: Breton, S. ([s.breton@umontreal.ca](mailto:s.breton@umontreal.ca)).

Keywords: mitochondrial genome; animals; ORFan genes; speciation.

\*These authors contributed equally to this work.

0168-9525/

Crown Copyright © 2014 Published by Elsevier Ltd. All rights reserved. <http://dx.doi.org/10.1016/j.tig.2014.09.002>

## Glossary

**Bilateria:** animals with bilateral symmetry; in other words, organisms with definite front and rear, and left and right, body surfaces.

**Cytoplasmic male sterility (CMS):** a condition under which a plant fails to produce functional pollen owing to the expression of novel or chimeric mtDNA-encoded proteins. Thus, in plant populations that are polymorphic for normal and CMS-specifying mtDNAs, two types of individuals are found: those with hermaphroditic and those with male-sterile 'functionally female' flowers. Male fertility can be restored by restorer of fertility (Rf) genes located in the nucleus. Many of these genes are members of the pentatricopeptide repeat (PPR) protein family, proteins that participate both in endonucleolytic cleavage and degradation of specific mRNAs.

**Direct variant repeat (DVR):** a repeated sequence, composed of a direct repeat (DR), plus an adjacent non-repetitive spacer sequence.

**Dobzhansky–Muller model:** a model of evolution based on the accumulation of genetic incompatibilities between populations (i.e., post-zygotic isolation). According to this model, post-zygotic isolation is a multi-gene trait that evolves by the accumulation of new mutations, eventually leading to fixation of different alleles in separate populations. Once hybrids are formed, their heterozygotic composition (for such loci) significantly reduces or disrupts hybrid fitness, thus leading to post-zygotic reproductive isolation. The model also predicts that post-zygotic reproductive isolation cannot be a trait encoded by a single locus.

**Domestication:** a process under which new sequences, obtained by horizontal gene transfer and endogenization, may acquire novel functions in the host genome.

**Effective population size:** referred to as  $N_e$ , the effective number of individuals in a population contributing to the offspring of the next generation. In natural populations, because not every potential parent actually mates,  $N_e$  equals the size of an idealized random-mating population that would have the same evolutionary behavior of the population under study. Its value depends on mode of inheritance, inbreeding level, sex ratio, variance of reproductive success, population size changes, and any genetic or geographic structuring of the effective population. Natural populations with low values of  $N_e$  are more affected by random genetic drift.

**Endogenization:** the process of inclusion of genetic material from a donor organism into a recipient genome through horizontal gene transfer. Several cases of endogenized viral sequences (mostly from retroviruses) are known in many eukaryotic genomes. Endogenized sequences may either become non-functional through pseudogenization, or be a source of new genes and functions in the recipient genome.

**Gene duplications:** one or more duplications of a gene or gene regions in a given genome. The extent of a duplication is largely dependent on the process that produces it: unequal crossing-over, replication slippage, or retrotransposition may involve one or a few genes, whereas aneuploidy and polyploidy involve single chromosomes or whole genomes, respectively. Duplicated sequences may become either non-functional through pseudogenization, or a source of new genes. Compared to endogenization, gene duplications are commonly recognized as a major source of new gene functions.

**Genetic drift:** the stochastic fluctuation of gene variants (alleles) in a given population. These variations in the presence of alleles are measured as changes in allele frequencies: the alleles in the offspring are a sample of the alleles in the parents that survive and reproduce every generation. The effect of genetic drift is larger in populations with low effective population size; that is, the less individuals are able to reproduce, the larger the fluctuations of gene frequencies in a given population. In the absence of selection, it is the main cause of variation of gene frequencies; in other words, it is the only cause of evolutionary changes in a population under neutral evolution.

**Gonochoresis:** a reproductive system in which male and female sexes are separated in different individuals (as opposed to hermaphroditism).

**Group I introns:** introns that catalyze their own excision from RNA precursors. Group I Introns have an elaborate folding pattern of large stems and loops due to nine specific paired regions (P1–P9). Group I introns can also have ORFs in their loop regions.

**Horizontal gene transfer:** the process of transferring genes (or, more generally, genetic material) from a donor to a recipient organism. Horizontal gene transfer may take place between individuals of the same species or between highly divergent organisms belonging to different domains of life. Horizontal gene transfer is distinct from 'genealogical or vertical transmission' of genetic material between parent and progeny. Horizontal gene transfer is common among viruses and prokaryotes, but has also been demonstrated in eukaryotes, especially through the process of endogenization of viral sequences. Horizontal gene transfer is considered to be a mechanism that permits the acquisition of evolutionary novelties through domestication.

**Mitochondrial control region:** a segment of the mitochondrial genome that is thought to contain the origin of replication of one strand, and the origin of transcription of both strands. The control region is normally located in the major non-coding portion of the genome. The control region frequently contains both hyper-variable domains and conserved sequence blocks. Some parts of the control region may fold into hairpin loops that likely act as signals for polymerases.

**Mitonuclear incompatibilities:** genetic incompatibilities between nuclear and mitochondrial genomes. The opportunity for mitonuclear incompatibilities arises because mitochondrial gene products must interact with nuclear gene products for respiration to occur (the core genes for OXPHOS are encoded by mitochondria while the remaining ones are encoded by the nucleus). Hence, both groups of genes must co-evolve. Other incompatibilities are conceivable because mitochondrial genes may code for functions other than respiration and ATP production as discussed herein.

**Neutral process:** the process of accumulation of mutations (and therefore new alleles in a given population) that do not affect the fitness of individuals. Because neutral alleles are not subject to natural selection, their frequency in a population is governed by the processes of mutation and random genetic drift.

**Oxidative phosphorylation (OXPHOS):** the metabolic process that mitochondria perform to produce ATP by oxidation of organic molecules (mostly glucose) to  $H_2O$  and  $CO_2$ , when oxygen ( $O_2$ ) is available.

**Parthenogenetic:** an organism that reproduces asexually. In animals, parthenogenetic virgin females produce new individuals from unfertilized eggs with no genetic contribution from a male. Parthenogenetic females may produce only females (telytochy), males (arrhenotochy), or both sexes (deuterotochy). Parthenogenesis can be accidental, facultative, or obligatory. New individuals may be haploid or diploid, either by suppressing meiosis and/or its effects (apomixis), or through a plethora of cytological mechanisms restoring the diploid constitution of eggs (automixis). Apomictic parthenogenesis produces genetic clones of the mother, while automictic parthenogenesis does not prevent genetic recombination, and the newborns are not clones. In animals, instances of parthenogenesis may also be coupled with hybridization and/or polyploidy.

**Pseudogenization:** the loss of function for a given gene/sequence. Pseudogenes are not expressed and are, therefore, non-functional relatives of functional genes. Gene duplication is a major contributor to pseudogenization.

**Reproductive isolation:** any mechanism severely limiting or preventing gene exchange between individuals/populations/species. Mechanisms of reproductive isolation either discourage mating or zygote formation (prezygotic mechanisms), or cause premature death, sterility, inviability, or low fitness of hybrids and their descendents (postzygotic mechanisms). The appearance of reproductive isolation is considered pivotal to the process of speciation. Reproductive isolation mechanisms are known to evolve under various circumstances and many different processes. The Dobzhansky–Muller model accounts for the evolution of post-zygotic genetic incompatibilities between genomes in a hybrid.

## Unusual mitochondrial gene content

During the past few years many animal mtDNAs have been sequenced, revealing numerous deviations from the usual content of 13 protein-coding genes. In bacteria, from which mitochondria originated, gene loss and acquisition are key mechanisms by which species adapt to novel environments and by which populations diverge and form distinct species. This raises the question of whether variation of mitochondrial gene content could also have adaptive explanations and/or pave the way for some evolutionary innovation in animals. The weight of evidence now firmly argues that atypical mitochondrial gene content may result in atypical mitochondrial functional properties; as a consequence, the range of functions of the mitochondrial genome is still far from being completely understood and a re-evaluation of the mtDNA genetic repertoire in animals is warranted. Below, we survey recent studies on animals with non-standard content of mitochondrial protein-coding genes (loss or duplication of 'typical' genes or additional 'atypical' genes with known function). We also discuss 'typical' protein genes involved in functions other than energy production, and ORFans.

### Loss of 'typical' protein-coding genes

Although the loss of typical protein-coding genes does not indicate an expansion of the functional repertoire of animal mtDNAs, these observations illustrate the flexibility of these genomes. For example, *atp6* is absent in the small 11 kb mtDNAs of the ctenophores *M. leidyi* and *Pleurobrachia bachei*, but it has been relocated to the

### Box 1. Mitochondrial OXPHOS system and mitonuclear coevolution

In animals, the mitochondrial OXPHOS system responsible for the production of ATP by oxidative phosphorylation is composed of five multi-subunit complexes of linked proteins that comprise the electron transport system (ETS). The ETS is the result of the assembly of 13 subunits encoded by mitochondrial DNA (mtDNA) and ~73 subunits encoded by nuclear DNA (nuDNA) [20]. Specifically, the 13 mtDNA-encoded proteins correspond to seven subunits of the NADH: ubiquinone oxidoreductase complex (complex I; ND1 to ND6, and ND4L), one subunit of the ubiquinone: cytochrome *c* oxidoreductase complex (complex III; CYTB), three subunits of the cytochrome *c* oxidase complex (complex IV; COX1 to COX3), and two subunits of ATP synthase (complex V; *ATP6* and *ATP8*) [1]. Other subunits of ETS complexes I, III, IV, and V, all the components of complex II (succinate: ubiquinone oxidoreductase), the proteins and enzymes of the mitochondrial matrix, as well as the factors involved in other mitochondrial functions (e.g., substrates and ion transport, and mtDNA replication), are all nuclear-encoded in animals [22]. Apart from the mitochondrion-encoded tRNA and rRNA genes, all other factors associated with mtDNA transcription and translation are also encoded by the nuclear genome [22].

As a consequence of mitochondria possessing their own DNA, energy production is strongly dependent on the tight cooperation and coincident coevolution of two separate genomes [20–27]. Breakdown of this process leads to suboptimal function or disruption of the ETS with severe fitness consequences (e.g., [20,21,24,25,27]). Mitonuclear coevolution is complicated by the distinct evolutionary processes that affect these two genomes. mtDNA experiences a faster basal mutation rate because of cell cycle-independent replication, a poor DNA mismatch-repair system, and the mutational damage caused by reactive oxygen species, a byproduct of OXPHOS. Moreover, the non-Mendelian, strictly maternal inheritance system reduces the effective population size of mtDNA to about 0.25 that for nuclear genes, making it more subject to genetic drift. The result is that mtDNA evolves, on average, more quickly than nuDNA (e.g., 40-fold faster in primates, see [28]). The processes involved in mitonuclear coevolution are still unknown: most likely both co-adaptation and compensatory evolution of nuclear subunits are involved [20–27]. Given that genetic variation, selection, and drift differ among populations, mitonuclear coevolution is predicted to be population-specific, and this can promote reproductive isolation and eventually speciation [20–27].

nuclear genome and has acquired introns and a mitochondrial targeting pre-sequence [17,29]. The absence of *atp6* has also been reported in several chaetognath mtDNAs [30], although it is not clear whether this is due to gene transfer to the nucleus or merely reflects a dispensable mitochondrial protein because no nuclear *atp6* has been identified yet in these species.

Among the other complete animal mtDNAs that have been studied so far, *atp8* seems to have been lost independently in many taxonomically diverse species (i.e., ctenophores, chaetognaths, sponges, nematodes, Platyhelminthes, and bivalves [2,17,29–32]), and *nad6* is apparently missing from the mtDNA of the arthropod *Metaseiulus occidentalis* [33]. However, because ATP8 and NAD6 are among the smallest and least-conserved mtDNA-encoded proteins (only 6–10 of the 50–65 amino acid residues of ATP8 are well conserved in animals), some putative ‘losses’ may be artifacts due to technical issues of genome annotation [2,7,17,32]. For example, in two recent re-annotations of bivalve mtDNAs, *atp8* candidates were identified that had not been previously recognized [34,35]. Nevertheless, loss of protein-coding genes may be viable owing to either relocation to the nuclear genome or neofunctionalization of another polypeptide. Real losses of mtDNA-encoded proteins could also result from structural differences in OXPHOS complexes among taxa [2]. Comparative analyses of mitochondrial proteomes are necessary to more accurately define the composition of OXPHOS complexes in animals.

#### Duplication of ‘typical’ protein-coding genes

The evolution of animal mtDNAs has been characterized by dynamic gene duplication resulting from duplications of segments of varying lengths that contain one or more genes and/or the mitochondrial control region (e.g., [36]). In many cases the duplicated genes are subject to pseudogenization, as in some birds (*nad6* [36]), nematodes (*nad5* [37]), mollusks (*cox3* [38]), and insects (e.g., *nad4*, *nad5*, and *cox3* in the seed crawler *Liposcelis bostrychophila* [39]; *nad2*, *cox1*, and *cox2* in scorpionflies [40]). The tandem duplication-random loss (TDRL) model has been proposed

to explain the presence of remnant pseudogenes and gene rearrangements in mtDNAs [1,2].

In contrast to pseudogenization, many mollusks possess duplicated mitochondrial genes that are identical or nearly identical in nucleotide sequence. For example, some cephalopod mtDNAs are characterized by duplicated segments containing five protein genes (*cox1*, *cox2*, *cox3*, *atp6*, and *atp8*) that evolve in concert, thus maintaining high sequence similarity [38]. The mtDNA of the aplousobranch *Chaetoderma nitidulum* has two copies of *cox2* and *nad2* [35]. *Cox2* is duplicated in tandem in the mitochondrial genomes of two bivalves with doubly uniparental inheritance (DUI) of mtDNA [41–45] (Box 2): *Ruditapes philippinarum* female-transmitted mtDNA (Okazaki and Ueshima, unpublished data) and *Musculista senhousia* male-transmitted mtDNA [44]. The additional copy of this latter gene, called *cox2b*, is longer than the original version and its functionality as *cox2* is still under study [44,45].

Other than mollusks, conservation of duplicated mitochondrial protein-coding genes has been reported in cnidarians (some Hydrozoa species have identical copies of *cox1* at both ends of their linear mtDNAs [12]), amphibians (*nad2* and *nad5* are duplicated and evolve in concert in afrobatrachian mtDNAs [46]), and reptiles (distinct lineages of the parthenogenetic hybrid gecko *Heteronotia binoei* complex possess different gene duplications that are conserved, possibly because of mitonuclear incompatibilities in these hybrids [47]). Interestingly, there are also potential examples of gene novelties after mitochondrial gene duplications: a 14th and highly transcribed gene was hypothesized to have resulted from an ancient duplication of *nad5* in the pearl oyster *Pinctada maxima* [48], whereas duplication of *nad2* followed by rapid divergence may have given rise to two novel genes in two species of *Crassostrea* oysters [49].

#### Additional ‘atypical’ protein-coding genes with known function

Mitochondrial protein-coding genes with non-OXPHOS functions were first reported in the octocoral *Sarcophyton glaucum* [50], where a homolog of *mutS*, a component of the

### Box 2. Doubly uniparental inheritance (DUI): a natural heteroplasmic system for mitochondria

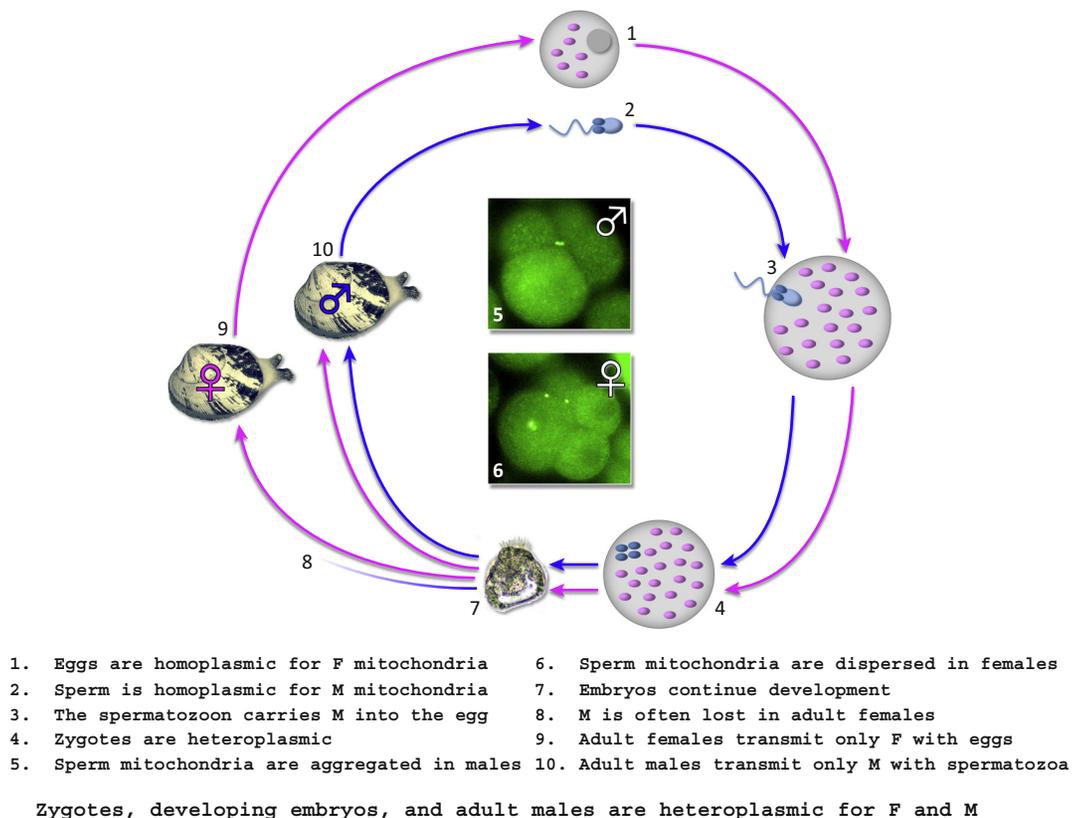
Mitochondrial heteroplasmy seems to be unfavorable, but it is increasingly observed in animals (reviewed in [95,96]). To be transmitted to future generations and influence evolution, however, heteroplasmy must persist via the germline [96]. Such systems are restricted to a few rare cases [96]. Given their rarity, natural mitochondrial heteroplasmic systems are extremely useful for investigating mitochondrial biology. One major heteroplasmic system is known as DUI (reviewed in [41–43]) (Figure 1). The only known animals exhibiting DUI are bivalve mollusks, including marine mussels, freshwater mussels, some clams, and a few others.

Under the system of DUI, two different mitochondrial lineages (and their respective genomes) are transmitted: one (the female-transmitted or F-type) from mother to daughters through eggs (Figure 1, 1), the other (the male-transmitted or M-type) from father to sons through spermatozoa (Figure 1, 2, 3). While females are usually homoplasmic for F, males are heteroplasmic for both M and F (M vs F DNA divergence  $\geq 20\%$ ), with the following typical tissue distribution: the germline is homoplasmic for the M-type (which will be transmitted via the sperm through sons), while the soma is heteroplasmic to various degrees, depending on tissue type and/or species.

In species with DUI, a peculiar aggregated pattern of sperm mitochondria has been observed in developing embryos. In male embryos, sperm mitochondria are actively carried to primordial germ cells (Figure 1, 5). The same pattern is not present in female embryos. In embryos destined to become female, the sperm mitochondria become randomly dispersed shortly after fertilization (Figure 1, 6). The molecular and regulatory machinery responsible for this peculiar pattern is still unknown, although a few models have been proposed so far. There is both direct and phylogenetic evidence that DUI can revert to maternal inheritance, with no apparent harm to mitochondrial functions. Thus, DUI presumably relies on the same molecular machinery that normally allows metazoan mitochondria to be transmitted to progeny, accordingly modified by some specific factors. At least one of these factors appears to be located in the mitochondrial genome (see the sections on mtORFans).

The nature of the processes behind mitochondrial inheritance is largely unknown, but some help is likely to come from the study of DUI. Moreover, DUI is a good experimental system to address questions about germline formation and the role of mitochondria in the process of cellular differentiation, about genomic conflicts between organelles and nucleus, and about the co-evolution of mitochondrial and nuclear genomes.

#### The doubly uniparental inheritance of mitochondria



TRENDS in Genetics

**Figure 1.** DUI. Transmission routes of the female-transmitted or F (pink) and male-transmitted or M (blue) mitochondria and their genomes in the DUI system of bivalve mollusks. 5 and 6: MitoTracker® Green FM staining of sperm mitochondria in early embryos of *Ruditapes philippinarum* (adapted from [97]).

bacterial MutSLH mismatch repair pathway, was hypothesized to originate by horizontal gene transfer either through an epsilonproteobacterium or a viral infection [51,52]. Soon after, extra protein genes were detected in a variety of animal mtDNAs: a putative homing endonuclease was

found in mitochondrial group I introns present in the *cox1* gene of several cnidarians, sponges, and placozoans [2,53,54]; *atp9*, a gene for subunit 9 of mitochondrial  $F_0$ -ATP synthase, that is usually located in the nuclear genomes of animals, was found in the mtDNAs of all major sponge

lineages [14,55]; *dnaB*, which belongs to the B-type DNA polymerase gene family, was first identified in the linear mtDNA of the medusozoan cnidarian *Aurelia aurita* [56] and was recently hypothesized to be responsible for the linearity of the mtDNAs of several medusozoan species [12,57]; and *tatC*, a subunit of the twin-arginine translocase complex, was found in sponges (Porifera: Oscarellidae) [58].

Finally, the most notable case of an additional ‘atypical’ mtDNA-encoded protein gene with known function has been found in humans (reviewed in [59,60]). The *humanin* gene is a 75 bp ORF in the mitochondrial 16S rRNA that encodes a 24 amino acid peptide that acts as a neuroprotector, an antiapoptotic agent, and a cytoprotector through retrograde signaling from the mitochondria [59,60]. It is still not clear, however, whether *humanin* is translated in the mitochondrion or the cytoplasm because it is biologically effective when synthesized using either mitochondrial or cytoplasmic genetic codes [59,60].

#### ‘Typical’ protein genes with additional functions

The functional repertoire of animal mtDNAs is further expanded by the fact that the 13 core mtDNA-encoded peptides are not exclusively associated with OXPHOS. For example, NAD2 has been shown to help regulate receptor activity in human brain synapses [61]. Similarly, COX2 is involved in Raf-1 protein kinase signaling and interacts with cyclin G1 [62].

‘Typical’ mtDNA-encoded proteins with extramitochondrial localization have also been reported in freshwater bivalves with DUI. In the male-transmitted mitochondria of these species, a unique ~185 codon extension of *cox2* (*Mcox2e*) is expressed as multiple transmembrane helices and is localized to both inner and outer mitochondrial membranes [63–65]; MCOX2e may tag sperm-derived mitochondria to facilitate their differential segregation in male and female embryos [64]. Consistent with this, seasonal variation in expression profiles suggests that MCOX2e functions in male reproduction [64]. Apparently, the COX2 protein encoded by the female-transmitted mtDNA (FCOX2) is also localized outside mitochondria; mainly in the plasma membrane, vitelline matrix, and vitelline envelope of mature eggs [66]. Although the above studies hypothesize a role for ‘typical’ mtDNA-encoded proteins beyond metabolic processes, unraveling the nature of their other functions (if any) will require in-depth biochemical studies.

#### ORFans: ‘atypical’ genes with unknown function

Mitochondrial ORFans (mtORFans) represent a noteworthy category of modifications to the standard mitochondrial repertoire that highlight the breadth of resourcefulness of this organelle. Examples include a putatively functional ORF containing repeat expansions of direct variant repeat (DVR) in-frame with the coding sequence found in the mtDNA of the octocoral *Calicogorgia granulosa* (Cnidaria) [67], as well as one extra element, *ORF314*, found in several medusozoan mtDNAs (Cnidaria) [12]. The products of these mtORFans do not resemble any known proteins, but the conservation of their DNA sequences (and position in the mtDNA for *ORF314*) suggests some level of selective pressure for their maintenance [12,67].

Recently, a positionally conserved ORF named *gau* (gene antisense ubiquitous) was found on the complementary strand of the *cox1* genes of eukaryotic mitochondria (protist, plant, fungal, and animal) and alphaproteobacteria [68]. This putative gene has also been identified in sense-oriented ESTs with poly(A) tails, and immunohistochemical experiments using an anti-GAU monoclonal antibody clearly showed a mitochondrion-specific signal in human cells, providing evidence for mitochondrial genes encoded by an overlapping genetic code [68].

To our knowledge, other than *gau*, the only animal mtORFans that have been functionally characterized are those found in bivalve mollusks with DUI (one mtORFan specific to the female-transmitted mtDNA (*fORF*) and one specific to the male-transmitted mtDNA (*mORF*) have been found in species belonging to the families Unionidae [69,70], Mytilidae [71], and Veneridae [45,72]). Their origin in bivalves may be extremely ancient, as in the case of mytilids and unionids, in which *fORF* has been maintained for at least 13 and 200 million years, respectively [70,71]. Even if the *fORF* and *mORF* are rapidly evolving [70,72], their long-term conservation in bivalves as well as their transcriptional level suggest that they are functional. Transcriptomic data of the venerid *R. philippinarum* have shown that although *fORF* is weakly transcribed compared with the 13 ‘typical’ OXPHOS genes, *mORF* is transcribed at the same level [45]. These findings are consistent with *mORF* functionality, which is also supported by the detection of its protein product (RPHM21) in male gonads (male germ cells are specifically stained with anti-RPHM21 antibodies, whereas no staining has been detected in eggs, consistent with the absence of male-transmitted mtDNA in the egg samples tested so far [73,74]). Interestingly, RPHM21 is localized not only in mitochondria but also in the nucleus of spermatogenic cells, with more intense labeling in mature spermatozoa [73]. It has also been immunodetected in early embryos, with deep staining around the animal–vegetal axis, the area in which germline determinants are localized [75]. These findings, together with *in silico* analyses (see below), are consistent with a role for *R. philippinarum* *mORF* in spermatogenesis, reproduction, and/or embryo development [72,73].

In the unionid *Venustaconcha ellipsiformis*, the translation products of both *mORF* and *fORF* genes have been verified by western blotting, and immunological analyses have localized the FORF protein in mitochondria, on the nuclear membrane, and in the nucleoplasm of eggs [70]. Notably, there is a strict correlation in unionids between DUI and the maintenance of separate male and female sexes, whereas closely related hermaphroditic species have lost DUI and have extensive mutations in the *fORF* gene of their female-type mtDNAs [70]. Thus, the *fORF* and *mORF* genes in unionids have been hypothesized to be involved in the maintenance of gonochorism [68]. If true, DUI would represent the first animal sex-determination system involving mtDNA-encoded proteins, explaining its long-term persistence in bivalves (heteromorphic sex chromosomes are absent in this taxa). However, the precise mechanisms underlying the connection between DUI and sex determination remain unknown.

### Box 3. Non-coding regions in animal mitochondrial DNA

Despite the similar patterns of nuclear genome evolution between animals and land plants, the mitochondrial genomes in these groups have evolved in strikingly different ways: plant mtDNAs are large and rich in non-coding sequences, whereas in animals they are small and compact. Lynch [91] proposed a much higher mutation rate ( $u$ ) in animal mtDNAs (~100-fold that in land plant organelles) as the main reason for the observed differences in mitochondrial genome architecture. Nonfunctional DNA is mutationally hazardous because it can originate gain-of-function deleterious mutations, so fast-evolving organelle genomes are more subject to selective pressure for genome reduction, even if the efficiency of selection depends on the effective population size ( $N_e$ ), which determines the strength of random genetic drift. Thus, drift and mutation rate are the major players shaping genome architecture and, because they are variable, they can be responsible for the mtDNA structural diversity among taxa; moreover, a high  $N_e u$  represents a barrier to non-coding DNA colonization [91]. Although animal mtDNA is more compact relative to plants, the widespread idea of its tight organization with almost no intergenic regions is the result of the skewed phylogenetic distribution of sequenced mitochondrial genomes. At the time of writing

(July 2014) more than 32 800 complete mtDNAs are available on GenBank: of these, over 94% belong to vertebrates (of which 85% are mammalian and 77% are human), and 3.2% to arthropods. These two taxa show a compact mtDNA, with the median amount of unassigned regions (URs) being ~6.4% and ~6.1% respectively (see [45]). There is growing evidence that this condition is not homogeneous among all animals: for example bivalve mollusks show a significantly higher proportion of URs (11.2%, see [45]). Interestingly, bivalves are among the most polymorphic animal species, so their molecular hypervariability coupled with a high percentage of intergenic DNA seems to contradict the hypothesis that the higher the mutation rate, the stronger the selection for genome reduction. On average,  $N_e$  in invertebrates is two orders of magnitude higher than in vertebrates ( $10^6$  vs  $10^4$ ). Accordingly, one would expect stronger selection for genome reduction. One possible explanation of this paradox is that the retention of the URs in bivalve mitochondrial genomes is caused by the presence of functional sequences and/or structures [45]. This could apply to other taxa (with some candidate sequences discussed in this review), and we suggest that it will be fruitful for future studies to focus on mitochondrial URs in all animal groups.

It is tempting to draw comparisons between DUI and cytoplasmic male sterility (CMS) in plants [76], in which sex determination involves both mitochondrial and nuclear genes. Indeed, several intriguing parallels exist between DUI and CMS: (i) both systems are associated with (or are caused by, in the case of CMS) novel or chimeric mtDNA-encoded proteins, (ii) as for plant mtDNAs [76–78], bivalve mtDNAs display dramatic variation in size (15–47 kb) (also Box 3), gene arrangement, and gene number [2,16,70], and (iii) mtDNA recombination, a mechanism responsible for the origin of chimeric CMS genes [76], has been observed in several bivalves with DUI [43]. It is worth noting that CMS has also been induced experimentally in *Drosophila* [79]. Indeed, one naturally occurring *D. melanogaster* mtDNA has been shown to cause male sterility in an evolutionary novel nuclear background, whereas in its co-evolved background fertility remained unaffected, suggesting the existence of a co-evolved restorer mutation in the natal nuclear background [79]. These results, together with numerous examples of mtDNA mutations reducing male fertility in nature (reviewed in [80]), suggest that CMS could be a more common phenomenon in animals than was previously believed, but the presence of counter-adapted nuclear mutations would prevent its detection. However, the specific role of CMS genes in causing male sterility is not yet clearly established. Most mitochondrial CMS genes in plants are unrelated in sequence [81], and the genetic mechanisms causing CMS are diverse and may vary even within a species. A common pattern observed in CMS systems is a disrupted interaction of the nuclear and the mitochondrial genomes facilitated by rearrangements in the latter [76–78]. The characterization of the genetic mechanisms of CMS systems and the corresponding fertility restoration are likely to provide new insights with respect to the complexity of mitonuclear interactions. Certainly, at least in plants, ‘atypical’ mitochondrial genes have a role in fundamental processes of organism life cycle, such as reproduction and germline development [76–78]. As for CMS systems, it is conceivable that DUI mechanisms of action are somewhat different among the evolutionarily distant DUI taxa (especially if DUI evolved

multiple times [72]). Several nuclear factors might also contribute to sex determination [82], and DUI-related factors could intervene at different levels of the sex-determination cascade leading to different effects in different taxa.

Given the absence of homologous proteins in databases, the molecular origin and function of mtORFans in bivalves with DUI have been hypothesized from *in silico* analyses [72,73]. The results indicate that FORF might be a DNA-binding protein, possibly involved in the regulation of mtDNA replication and/or transcription [70], while MORF may prevent the recognition of male-transmitted mitochondria by the degradation machinery, allowing their survival in male zygotes via a mechanism similar to that of modulators of immune recognition (MIRs), viral proteins involved in the immune-recognition pathway, to which MORF proteins showed structural similarities [73]. Indeed, many clues suggest that the analyzed bivalve mtORFans arose from endogenization of viral genes. With their infectious properties, viruses can spread horizontally, but they can also become incorporated into the genetic material of their host, their integration into the germline allowing vertical transmission and fixation in the host population [83]. Many pathways leading to new genes through the domestication of parasitic genome sequences have been documented [84]. Remarkably, the mtDNA replication and transcription apparatus has been found to have viral origins [85]. Mitochondria play a central role in primary host defense mechanisms against viral infections, and several viral proteins interact with mitochondria to regulate cellular responses [86]. Because signaling from recognition receptors converges in mitochondria, it is reasonable that viruses would target mitochondrial processes to evade immune responses [86].

Obviously, bivalve mtORFans represent candidate genes that could be responsible for or participate in the DUI mechanism (and potentially sex determination). Their viral origin could explain the acquired capability of male-transmitted mitochondria to avoid degradation and invade the germline. The process could involve a mechanism similar to that of MIRs, including the detachment of specific tags associated with sperm mitochondria that

normally lead to their degradation by the ubiquitination machinery. Each of these hypotheses needs to be evaluated, and the ultimate cause(s) of deviation from the strictly maternal inheritance of mtDNA in bivalves remains an open question.

### Involvement of the mitochondrial genome in speciation processes

The evolutionary factors that shape modern-day animal mtDNA are not yet clearly understood. The widespread idea of its tight organization and limited coding capacity is most certainly due to the skewed phylogenetic distribution of sequenced mitochondrial genomes (94% are highly conserved vertebrate mtDNAs; **Box 3**). To uncover mitochondrial novelties we must expand taxonomic sampling in a comprehensive manner. For instance, could mitochondrial features observed in the mtDNAs of sponges, cnidarians, placozoans, or bivalve mollusks have adaptive explanations, such as the occupation of different environmental niches and/or the adoption of various feeding habits or modes of locomotion? Could these exceptional mtDNAs reflect differences in breeding systems such as hermaphroditism or other intrinsic factors? It has been previously shown that one major transition in animal evolution, the origin of the Bilateria, is associated with multiple changes in the mitochondrial genetic code, losses of tRNA genes, and an accelerated rate of sequence evolution [87]. However, it is still unknown whether these changes co-occurred with the morphological transitions or evolved independently multiple times [87]. Thorough studies of exceptions to the mainstream mitochondrial tendencies are crucial to obtain insights into the processes that underlie animal mitochondrial genome diversity. One fundamental issue in the field is to what extent mtDNA can respond adaptively to environmental pressures and what role this may play in speciation. Empirical examples of speciation involving nuclear genes exist for a range of animals and plants (reviewed in [88]), and there is growing evidence that mtDNA variation can also be involved in the origin of reproductive barriers and speciation events [20–27]. The most widespread model to explain the evolution of post-zygotic isolation through accumulation of interpopulation genetic incompatibilities is the Dobzhansky–Muller model. Examples supporting a role of mitochondria in Dobzhansky–Muller incompatibilities and in speciation events in fishes, humans, mice, birds, insects, and other invertebrates have recently been reviewed [23–27]. For instance, Gershoni *et al.* [25] analyzed four main aspects of mitochondrial involvement in the generation of reproductive barriers: (i) a functional role of mitochondrial genes in germline formation/viability, (ii) the involvement of mitochondrial bioenergetics in development/differentiation, (iii) OXPHOS performance in reproductive potency, and (iv) the influence of mitochondrial bioenergetics on fitness. Following up on groundbreaking work on the copepod *Tigriopus californicus* (see [27]), which has become a model system for the study of mitonuclear coevolution, several studies clearly attribute the observed breakdown of hybrid fitness to mitonuclear interactions. Indeed, QTL mapping studies of *T. californicus* showed that the strongest genetic incompatibilities were mitonuclear [89]. Among

vertebrates, mitonuclear incompatibility explains the disruption of ATP synthase function in hybrid eels, contributing to reproductive isolation between European and American eels [90]. Mitonuclear interactions may have contributed to the explosive radiation of eukaryotes: reproductive incompatibility following the relocation of an organelle gene could have prompted speciation, especially when gene transfer from primordial mitochondria to the nucleus was common [91]. Thus, it is plausible that mtDNA rearrangements/variability (such as those discussed in this review) promote reproductive isolation and speciation. In particular, functional mtORFans are ideal candidates for the identification of positive selection at the sequence level and for testing the hypothesized involvement of mitochondria and their genomes in the establishment of reproductive barriers. Indeed, ORFan genes usually evolve at a much faster rate than other genes, and they have been hypothesized to be involved in the evolution of lineage-specific adaptive traits [92]. They are also rapidly lost, again reflecting lineage-specific functional requirements [93]. Similarly, the above-discussed bivalve mtORFans, as well as other mtORFans and ‘atypical’ or multifunctional genes observed in animal mtDNAs, could be the result of adaptive evolution and thus play an important role in speciation.

### Concluding remarks

The literature reviewed here clearly illustrates that animal mtDNAs, which have historically been regarded as functionally limited, possess an astonishing functional diversity that has yet to be fully characterized (**Box 4**). Processes such as gene duplication, gene transfer, and gene multifunctionality create opportunities for genetic novelties that could be relevant for lineage-specific adaptations and concomitant speciation. Employment of recent high-throughput techniques in ‘omics’ and biochemical assays is likely to uncover more of these ‘atypical’ mitochondrial genes. The discovery of these genes opens new avenues for the identification, characterization, and functional analyses of the many ORFs found in a large proportion of animal mtDNA regions that have been defined as ‘noncoding’. Several studies have recently demonstrated the potential for short nuclear ORFs (30–60 bp in length) to encode biologically active peptides that have regulatory roles in eukaryotic cells (reviewed in [94]). It is therefore almost certainly the case that mtORFans with key biological functions have been overlooked in animals. Finally, the hypothesis that genes encoded by mtDNA are involved in speciation processes has been circulating for a while and this is gaining new experimental support. Notwithstanding the plurality of emerging features and roles, our knowledge of animal mtDNAs is still incomplete and strongly biased towards vertebrates, which show highly conservative and compact mitochondrial genomes. In fact, by defining their genomes as ‘typical’, we overlook the real big range of mtDNA variability, as well as its ‘hidden’ functions and role in eukaryotic evolution. Indeed, we hope this review leads readers to reconsider common generalizations about animal mtDNA and highlight the emerging new features of this genome, as well as its relevance as an evolutionary trigger. Nonetheless, the available data are

#### Box 4. Outstanding questions

In addition to producing ATP, mitochondria have a central role in the cell cycle (in particular cell growth and death) as well as in cell signaling, fertilization, development, differentiation, aging, apoptosis, and even sex differentiation. Research on mitochondria and their genomes has progressed along two avenues: one focuses on their role in organismal functions, such as aging and human disorders, the other uses mtDNA as a tool for phylogenetics and population studies. To appreciate the full range of animal mitochondrial functions and behaviors, however, there is a compelling need to go beyond human and mammalian systems or other model organisms (e.g., *Drosophila*). Atypical mitochondrial systems, such as DUI in bivalves (Box 2), provide the opportunity to study mitochondria from a unique vantage point. They are uniquely suited to efficiently and cost-effectively address several fundamental questions in mitochondrial research such as:

- What are the causes and molecular mechanisms ensuring the effective transmission of functional and undamaged mitochondria to progeny?
- What are the selective pressures that maintain the near universality of uniparental inheritance?
- What are the causes, evolutionary advantages, and molecular mechanisms behind paternal mtDNA elimination in animals?
- What are the selective processes acting on mitochondria as they pass through the germline during gamete formation?
- What is the nature of the possible interactions (e.g., molecular, metabolic, or both) between nuclear and mitochondrial genomes, and how do they affect differentiation, development, and sex determination?
- If mtDNA and nuclear-encoded proteins jointly contribute to novel, non-OXPHOS related mechanisms, what is the nature of selection on these mechanisms?
- Is the male-transmitted mtDNA present in sperm capable of carrying out all mitochondrial functions?
- Does mitochondrial homoplasmy, which is the norm for most eukaryotes, provide a functional advantage? If so, what compensatory mechanisms are necessitated by heteroplasmy?
- Is recombination between mtDNAs a possible way for generating mitotypes with new functions, as is the case for nuclear DNA recombination?

still scattered, and additional research in this area will doubtless provide new insights and shed light on the origin and functions of novel mitochondrial genes.

#### Acknowledgments

This review was supported by the Natural Sciences and Engineering Research Council of Canada (S.B. and D.T.S.), the University of Montreal (S.B.), Italian Ministry for University and Research “Futuro In Ricerca 2013” grant (F.G.), and ‘Canziani Bequest’ and ‘Fondazione del Monte’ funding (M.P.).

#### References

- 1 Boore, J.L. (1999) Animal mitochondrial genomes. *Nucleic Acids Res.* 27, 1767–1780
- 2 Gissi, C. *et al.* (2008) Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity* 101, 301–320
- 3 Scheffler, I. (2008) *Mitochondria*, Wiley-Liss
- 4 Van Blerkom, J. (2011) Mitochondrial function in the human oocyte and embryo and their role in developmental competence. *Mitochondrion* 11, 797–813
- 5 López-Otín, C. *et al.* (2013) The hallmarks of aging. *Cell* 153, 1999–2127
- 6 Chandel, N.S. (2014) Mitochondria as signaling organelles. *BMC Biol.* 12, 34
- 7 Wang, X. and Lavrov, D.V. (2008) Seventeen new complete mtDNA sequences reveal extensive mitochondrial genome evolution within the Desmospongiae. *PLoS ONE* 3, e2723
- 8 Emblem, A. *et al.* (2011) Mitogenome rearrangement in the cold-water scleractinian coral *Lophelia pertusa* (Cnidaria, Anthozoa) involves a long-term evolving group I intron. *Mol. Phylogenet. Evol.* 61, 495–503
- 9 Burger, G. *et al.* (2009) Group I intron trans-splicing and mRNA editing in the mitochondria of placozoan animals. *Trends Genet.* 25, 381–386
- 10 Vallès, Y. *et al.* (2008) Group II introns break new boundaries: presence in a bilaterian’s genome. *PLoS ONE* 3, e1488
- 11 Voigt, O. *et al.* (2008) A fragmented metazoan organellar genome: the two mitochondrial chromosomes of *Hydra magnipapillata*. *BMC Genomics* 9, 350
- 12 Kayal, E. *et al.* (2012) Evolution of linear mitochondrial genomes in medusozoan cnidarians. *Genome Biol. Evol.* 4, 1
- 13 Doublet, V. *et al.* (2013) Inverted repeats and genome architecture conversions of terrestrial isopods mitochondrial DNA. *J. Mol. Evol.* 77, 107–118
- 14 Lavrov, D.V. *et al.* (2013) Mitochondrial DNA of *Clathrina clathrus* (Calcarea, Calcinea): six linear chromosomes, fragmented rRNAs, tRNA editing, and a novel genetic code. *Mol. Biol. Evol.* 30, 865–880
- 15 Jiang, H. *et al.* (2013) Substantial variation in the extent of mitochondrial genome fragmentation among blood-sucking lice of mammals. *Genome Biol. Evol.* 5, 1298–1308
- 16 Liu, Y.G. *et al.* (2013) Complete mitochondrial DNA sequence of the ark shell *Scapharca broughtonii*: an ultra-large metazoan mitochondrial genome. *Comp. Biochem. Physiol. Part D: Genomics Proteomics* 8, 72–81
- 17 Pett, W. *et al.* (2011) Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi*: insight from mtDNA and the nuclear genome. *Mitochondrial DNA* 22, 130–142
- 18 Subramanian, S. (2009) Temporal trails of natural selection in human mitogenomes. *Mol. Biol. Evol.* 26, 715–717
- 19 Soares, P. *et al.* (2013) Evaluating purifying selection in the mitochondrial DNA of various mammalian species. *PLoS ONE* 8, e58993
- 20 Wallace, D.C. (2013) Bioenergetics in human evolution and disease: implications for the origins of biological complexity and the missing genetic variation of common diseases. *Philos. Trans. R. Soc. Lond. B* 368, 2012027
- 21 Dowling, D.K. (2014) Evolutionary perspectives on the links between mitochondrial genotype and disease phenotype. *Biochim. Biophys. Acta* 1840, 1393–1403
- 22 Wolff, J.N. *et al.* (2014) Mitonuclear interactions: evolutionary consequences over multiple biological scales. *Philos. Trans. R. Soc. Lond. B* 369, 20130443
- 23 Dowling, D.K. (2008) Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends Evol. Ecol.* 23, 546–554
- 24 Ballard, J.W.O. and Melvin, R.G. (2010) Linking the mitochondrial genotype to the organismal phenotype. *Mol. Ecol.* 19, 1523–1539
- 25 Gershoni, M. *et al.* (2009) Mitochondrial bioenergetics as a major motive force of speciation. *Bioessays* 31, 642–650
- 26 Lane, N. (2009) On the origin of bar codes. *Nature* 462, 272–274
- 27 Burton, R. and Barreto, F.S. (2012) A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? *Mol. Ecol.* 21, 4942–4957
- 28 Osada, N. *et al.* (2012) Mitochondrial-nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome c oxidase complex. *Mol. Biol. Evol.* 29, 337–346
- 29 Kohn, A.B. *et al.* (2012) Rapid evolution of the compact and unusual mitochondrial genome in the ctenophore *Pleurobrachia bachei*. *Mol. Phylogenet. Evol.* 63, 203–207
- 30 Miyamoto, H. *et al.* (2010) Complete mitochondrial genome sequences of the three pelagic chaetognaths *Sagitta nagae*, *Sagitta decipiens* and *Sagitta enflata*. *Comp. Biochem. Physiol. Part D: Genomics Proteomics* 5, 65–72
- 31 Haen, K.M. *et al.* (2007) Glass sponges and bilaterian animals share derived mitochondrial genomic features: a common ancestry or parallel evolution? *Mol. Biol. Evol.* 24, 1518–1527
- 32 Bernt, M. *et al.* (2013) Genetic aspects of mitochondrial genome evolution. *Mol. Phylogenet. Evol.* 69, 328–338
- 33 Dermauw, W. *et al.* (2010) Mitochondrial genome analysis of the predatory mite *Phytoseiulus persimilis* and a revisit of the *Metaseiulus occidentalis* mitochondrial genome. *Genome* 53, 285–301
- 34 Breton, S. *et al.* (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
- 35 Stöger, I. and Schrödl, M. (2013) Mitogenomics does not resolve deep molluscan relationships (yet?). *Mol. Phylogenet. Evol.* 69, 376–392

- 36 Schirtzinger, E.E. *et al.* (2012) Multiple independent origins of mitochondrial control region duplications in the order Psittaciformes. *Mol. Phylogenet. Evol.* 64, 342–356
- 37 Raboin, M.J. *et al.* (2010) Evolution of *Caenorhabditis* mitochondrial genome pseudogenes and *Caenorhabditis briggsae* natural isolates. *Mol. Biol. Evol.* 27, 1087–1096
- 38 Kawashima, Y. *et al.* (2013) The complete mitochondrial genomes of deep-sea squid (*Bathyteuthis abyssicola*), bob-tail squid (*Semirossia patagonica*) and four giant cuttlefish (*Sepia apama*, *S. latimanus*, *S. lycidas* and *S. pharaonis*), and their application to the phylogenetic analysis of Decapodiformes. *Mol. Phylogenet. Evol.* 69, 980–993
- 39 Wei, D.D. *et al.* (2012) The multipartite mitochondrial genome of *Liposcelis bostrychophila*: insights into the evolution of mitochondrial genomes in bilateral animals. *PLoS ONE* 7, e33973
- 40 Beckenbach, A.T. (2011) Mitochondrial genome sequences of representatives of three families of scorpionflies (Order Mecoptera) and evolution in a major duplication of coding sequence. *Genome* 54, 368–376
- 41 Breton, S. *et al.* (2007) The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? *Trends Genet.* 23, 465–474
- 42 Passamonti, M. and Ghiselli, F. (2009) Doubly uniparental inheritance: two mitochondrial genomes, on precious model for organelle DNA inheritance and evolution. *DNA Cell Biol.* 28, 79–89
- 43 Zouros, E. (2013) Biparental inheritance through uniparental transmission: the Doubly Uniparental Inheritance (DUI) of mitochondrial DNA. *Evol. Biol.* 40, 1–31
- 44 Passamonti, M. *et al.* (2011) Mitochondrial genomes and doubly uniparental inheritance: new insights from *Musculista senhousia* sex-linked mitochondrial DNAs (Bivalvia: Mytilidae). *BMC Genomics* 12, 442
- 45 Ghiselli, F. *et al.* (2013) Structure, transcription, and variability of metazoan mitochondrial genome: perspectives from an unusual mitochondrial inheritance system. *Genome Biol. Evol.* 5, 1535–1554
- 46 Kurabayashi, A. and Sumida, M. (2013) Afrobatrachian mitochondrial genomes: genome reorganization, gene rearrangement mechanisms, and evolutionary trends of duplicated and rearranged genes. *BMC Genomics* 14, 633
- 47 Fujita, M.K. *et al.* (2007) Multiple origins and rapid evolution of duplicated mitochondrial genes in parthenogenetic geckos (*Heteronotia binoei*; Squamata, Gekkonidae). *Mol. Biol. Evol.* 24, 2775–2786
- 48 Wu, X. *et al.* (2012) A unique tRNA gene family and a novel, highly expressed ORF in the mitochondrial genome of the silver-lip pearl oyster, *Pinctada maxima* (Bivalvia: Pteriidae). *Gene* 510, 22–31
- 49 Wu, X. *et al.* (2012) New features of Asian *Crassostrea* oyster mitochondrial genomes: a novel alloacceptor tRNA gene recruitment and two novel ORFs. *Gene* 507, 112–118
- 50 Pont-Kingdon, G.A. *et al.* (1995) A coral mitochondrial mutS gene. *Nature* 375, 109–111
- 51 Bilewicz, J.P. and Degnan, S.M. (2011) A unique horizontal gene transfer event has provided the octocoral mitochondrial genome with an active mismatch repair gene that has potential for an unusual self-contained function. *BMC Evol. Biol.* 11, 228
- 52 Ogata, H. *et al.* (2011) Two new subfamilies of DNA mismatch repair proteins (MutS) specifically abundant in the marine environment. *ISME J.* 5, 1143–1151
- 53 Signorovitch, A.Y. *et al.* (2007) Comparative genomics of large mitochondria in placozoans. *PLoS Genet.* 12, e13
- 54 Szitenberg, A. *et al.* (2010) Diversity of sponge mitochondria introns revealed by *cox 1* sequences of Tetillidae. *BMC Evol. Biol.* 10, 288
- 55 Lavrov, D.V. *et al.* (2005) Mitochondrial genomes of two demosponges provide insights into an early stage of animal evolution. *Mol. Biol. Evol.* 22, 1231–1239
- 56 Shao, Z. *et al.* (2006) Mitochondrial genome of the moon jelly *Aurelia aurita* (Cnidaria, Scyphozoa): a linear DNA molecule encoding a putative DNA-dependent DNA polymerase. *Gene* 381, 92–101
- 57 Zou, H. *et al.* (2012) Mitochondrial genome of the freshwater jellyfish *Craspedacusta sowerbyi* and phylogenetics of Medusozoa. *PLoS ONE* 7, e51465
- 58 Pett, W. and Lavrov, D.V. (2013) The twin-arginine subunit C in Oscarella: origin, evolution, and potential functional significance. *Integr. Comp. Biol.* 53, 495–502
- 59 Lee, C. *et al.* (2013) Humanin: a harbinger of mitochondrial-derived peptides? *Trends Endocrinol. Metab.* 24, 222–228
- 60 Cohen, P. (2014) New role for the mitochondrial peptide humanin: protective agent against chemotherapy-induced side effects. *J. Natl. Cancer Inst.* 106, <http://dx.doi.org/10.1093/jnci/dju006>
- 61 Gingrich, J.R. *et al.* (2004) Unique domain anchoring of Src to synaptic NMDA receptors via the mitochondrial protein NADH dehydrogenase subunit 2. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6237–6242
- 62 Maximov, V. *et al.* (2002) Mitochondrial 16S rRNA gene encodes a functional peptide, a potential drug for Alzheimer's disease and target for cancer therapy. *Med. Hypotheses* 59, 670–673
- 63 Chakrabarti, R. *et al.* (2006) Presence of a unique male-specific extension of C-terminus to the cytochrome *c* oxidase subunit II protein coded by the male-transmitted mitochondrial genome of *Venustaconcha ellipsiformis* (Bivalvia: Unionoidea). *FEBS Lett.* 580, 862–866
- 64 Chakrabarti, R. *et al.* (2007) Reproductive function for a C-terminus extended, male-transmitted cytochrome *c* oxidase subunit II protein expressed in both spermatozoa and eggs. *FEBS Lett.* 581, 5213–5219
- 65 Chapman, E.G. *et al.* (2008) Extreme primary and secondary protein structure variability in the chimeric male-transmitted cytochrome *c* oxidase subunit II protein in freshwater mussels: Evidence for an elevated amino acid substitution rate in the face of domain-specific purifying selection. *BMC Evol. Biol.* 8, 165
- 66 Chakrabarti, R. *et al.* (2009) Extra-mitochondrial localization and likely reproductive function of a female-transmitted cytochrome *c* oxidase subunit II protein. *Dev. Growth Differ.* 51, 511–519
- 67 Park, E. *et al.* (2011) The complete mitochondrial genome of *Calicogorgia granulosa* (Anthozoa: Octocorallia): potential gene novelty in unidentified ORFs formed by repeat expansion and segmental duplication. *Gene* 486, 81–87
- 68 Faure, E. *et al.* (2011) Probable presence of an ubiquitous cryptic mitochondrial gene on the antisense strand of the cytochrome oxidase I gene. *Biol. Direct* 6, 56
- 69 Breton, S. *et al.* (2009) Comparative mitochondrial genomics of freshwater mussels (Bivalvia: Unionoidea) with doubly uniparental inheritance of mtDNA: gender-specific open reading frames (ORFs) and putative origins of replication. *Genetics* 183, 1575–1589
- 70 Breton, S. *et al.* (2011) Novel protein genes in animal mtDNA: a new sex determination system in freshwater mussels (Bivalvia: Unionoidea)? *Mol. Biol. Evol.* 28, 1645–1659
- 71 Breton, S. *et al.* (2011) Evidence for a fourteenth mtDNA-encoded protein in the female-transmitted mtDNA of marine mussels (Bivalvia: Mytilidae). *PLoS ONE* 6, 19365
- 72 Milani, L. *et al.* (2013) A comparative analysis of mitochondrial ORFans: new clues on their origin and role in species with doubly uniparental inheritance of mitochondria. *Genome Biol. Evol.* 5, 1408–1434
- 73 Milani, L. *et al.* (2014) Paternally transmitted mitochondria express a new gene of potential viral origin. *Genome Biol. Evol.* 6, 391–405
- 74 Ghiselli, F. *et al.* (2011) Strict sex-specific mtDNA segregation in the germline of the DUI species *Venerupis philippinarum* (Bivalvia Veneridae). *Mol. Biol. Evol.* 28, 949–961
- 75 Milani, L. *et al.* (2011) Doubly uniparental inheritance of mitochondria as a model system for studying germ line formation. *PLoS ONE* 6, e28194
- 76 Chase, C.D. (2007) Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. *Trends Genet.* 23, 81–90
- 77 Rieseberg, L.H. and Blackman, B.K. (2010) Speciation genes in plants. *Ann. Bot.* 106, 439–455
- 78 Horn, R. *et al.* (2014) Mitochondrion role in molecular basis of cytoplasmic male sterility. *Mitochondrion* <http://dx.doi.org/10.1016/j.mito.2014.04.004>
- 79 Clancy, D.J. *et al.* (2011) Cytoplasmic male sterility in *Drosophila melanogaster* associated with a mitochondrial CYTB variant. *Heredity* 107, 374–376
- 80 Beekman, M. *et al.* (2014) The costs of being male: are there sex-specific effects of uniparental mitochondrial inheritance? *Philos. Trans. R. Soc. Lond. B* 369, 20130440
- 81 Islam, Md.S. *et al.* (2014) Genetics and biology of cytoplasmic male sterility and its applications in forage and turf grass breeding. *Plant Breed.* 133, 299–312
- 82 Milani, L. *et al.* (2013) Nuclear genes with sex bias in *Ruditapes philippinarum* (Bivalvia, Veneridae): mitochondrial inheritance and sex determination in DUI species. *J. Exp. Zool.* 320, 442–454

- 83 Feschotte, C. and Gilbert, C. (2012) Endogenous viruses: insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* 13, 283–296
- 84 Kaessmann, H. (2010) Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 20, 1313–1326
- 85 Forterre, P. (2010) The universal tree of life and the last universal cellular ancestor: revolution and counterrevolutions. In *Evolutionary Genomics and Systems Biology* (Caetano-Anollés, G., ed.), pp. 43–62, John Wiley and Sons
- 86 Ohta, A. and Nishiyama, Y. (2011) Mitochondria and viruses. *Mitochondrion* 11, 1–12
- 87 Lavrov, D. (2007) Key transitions in animal evolution: a mitochondrial DNA perspective. *Integr. Comp. Biol.* 47, 734–743
- 88 Keller, I. and Seehausen, O. (2012) Thermal adaptation and ecological speciation. *Mol. Ecol.* 21, 782–799
- 89 Foley, B.R. *et al.* (2013) Postzygotic isolation involves strong mitochondrial and sex-specific effects in *Tigriopus californicus*, a species lacking heteromorphic sex chromosomes. *Heredity* 111, 391–401
- 90 Gagnaire, P.A. *et al.* (2012) Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American Eels. *Mol. Biol. Evol.* 29, 2909–2919
- 91 Lynch, M. (2007) *The Origins of Genome Architecture*, Sinauer Associates
- 92 Tautz, D. and Domazet-Loso, T. (2011) The evolutionary origin of orphan genes. *Nat. Rev. Genet.* 12, 692–702
- 93 Palmieri, N. *et al.* (2014) The life cycle of *Drosophila* orphan genes. *Elife* 3, e01311
- 94 Andrews, S.J. and Rothnagel, J.A. (2014) Emerging evidence for functional peptides encoded by short open reading frames. *Nat. Rev. Genet.* 15, 193–204
- 95 Lane, N. (2012) The problem with mixing mitochondria. *Cell* 151, 246–248
- 96 White, J.D. *et al.* (2008) Revealing the hidden complexities of mtDNA inheritance. *Mol. Ecol.* 17, 4925–4942
- 97 Milani, L. *et al.* (2012) Sex-linked mitochondrial behavior during early embryo development in *Ruditapes philippinarum* (Bivalvia Veneridae) a species with the doubly uniparental inheritance (DUI) of mitochondria. *J. Exp. Zool.* 318, 182–189