



OPEN

## Mito-nuclear coevolution and phylogenetic artifacts: the case of bivalve mollusks

Alessandro Formaggioni, Federico Plazzi<sup>✉</sup> & Marco Passamonti

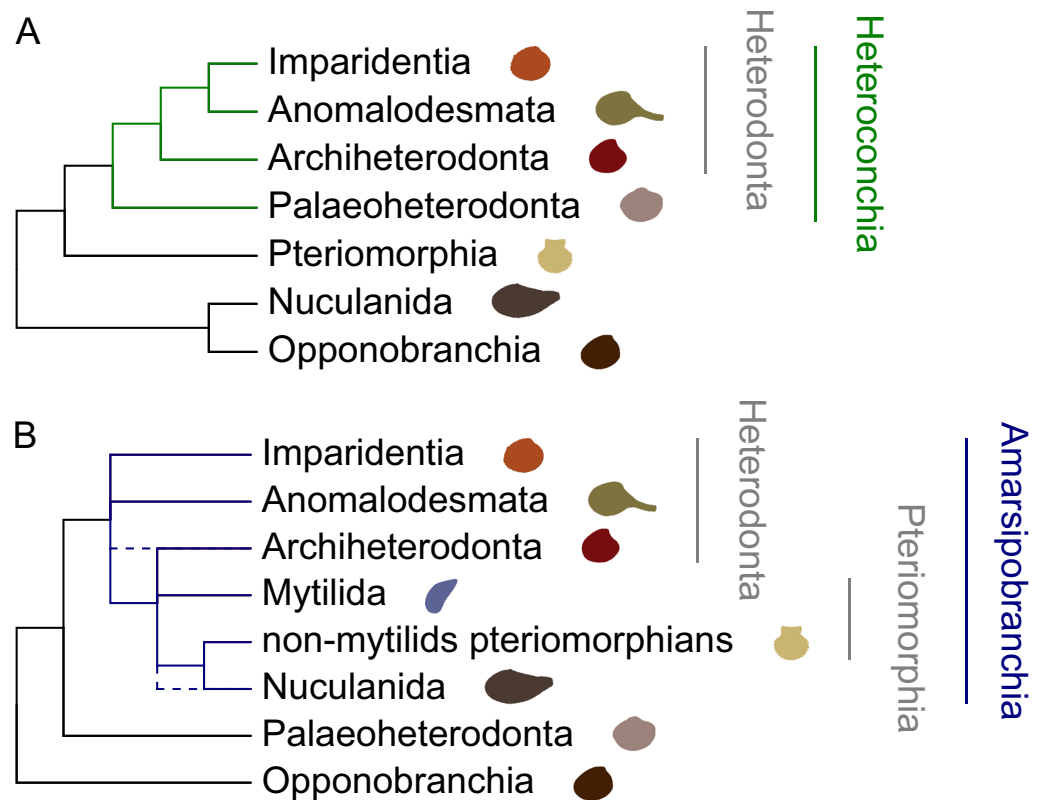
Mito-nuclear phylogenetic discordance in Bivalvia is well known. In particular, the monophyly of Amarsipobranchia (Heterodonta + Pteriomorphia), retrieved from mitochondrial markers, contrasts with the monophyly of Heteroconchia (Heterodonta + Palaeoheterodonta), retrieved from nuclear markers. However, since oxidative phosphorylation nuclear markers support the Amarsipobranchia hypothesis instead of the Heteroconchia one, interacting subunits of the mitochondrial complexes ought to share the same phylogenetic signal notwithstanding the genomic source, which is different from the signal obtained from other nuclear markers. This may be a clue of coevolution between nuclear and mitochondrial genes. In this work we inferred the phylogenetic signal from mitochondrial and nuclear oxidative phosphorylation markers exploiting different phylogenetic approaches and added two more datasets for comparison: genes of the glycolytic pathway and genes related to the biogenesis of regulative small noncoding RNAs. All trees inferred from mitochondrial and nuclear subunits of the mitochondrial complexes support the monophyly of Amarsipobranchia, regardless of the phylogenetic pipeline. However, not every single marker agrees with this topology: this is clearly visible in nuclear subunits that do not directly interact with the mitochondrial counterparts. Overall, our data support the hypothesis of a coevolution between nuclear and mitochondrial genes for the oxidative phosphorylation. Moreover, we suggest a relationship between mitochondrial topology and different nucleotide composition between clades, which could be associated to the highly variable gene arrangement in Bivalvia.

**Deep bivalve phylogeny: state-of-art.** Bivalves are an extremely diverse group with about 50,000 living species<sup>1</sup>. Deep evolutionary relationships among major clades within the molluscan class Bivalvia are only recently coming to a shared figure. The class is split into two main subgroups, Protobranchia and Autobranchia, whose origins root deep in the middle Ordovician periods<sup>2–6</sup>. Most likely, extant protobranchs resemble the Cambrian forerunners the most, for many molluscan symplesiomorphies are present, like a well-developed foot and true molluscan ctenidia devoted to gas exchange<sup>7,8</sup>; moreover, food is brought to the mouth by palp proboscides. Two sister groups are usually acknowledged within Protobranchia, Nuculida and Solemyida, which are given an ordinal status<sup>2,9–13</sup>; analyses mainly based on molecular markers proposed to exclude the protobranch superfamily Nuculanoidea from Protobranchia and to better place it within Autobranchia<sup>14–17</sup>; the name Opponobranchia was proposed for remaining protobranchs<sup>18</sup>. On the other hand, the clade Protobranchia has been recovered by most of large-scale datasets<sup>19,20</sup>, but with some exceptions<sup>20</sup>. Therefore, the monophyly of this clade still needs to be assessed.

The way of feeding is radically different in Autobranchia (= Autolamellibranchiata *sensu*<sup>18</sup>), whose common ancestor developed a feeding gill, one of the main drivers of the Ordovician bivalve radiation<sup>3</sup> and led most groups to the key ecological shift towards infaunalization<sup>4,5,21</sup>. Autobranchia is comprised by three major clades (subclasses<sup>22</sup>): Heterodonta (clams, cockles, razor clams, and their kin), Palaeoheterodonta (freshwater mussels and their kin), and Pteriomorphia (mytilids, oysters, scallops, and their kin)<sup>19,23,24</sup>. Moreover, the former subclass Anomalodesmata<sup>22,25–27</sup> has been found to be nested within Heterodonta<sup>14,15,18,20,28–31</sup>. Currently, Archiheterodonta (order Carditida) are considered sister group to other Euheterodonta, which are further split into Anomalodesmata itself and Imparidentia<sup>15,20,23,26,27,31–33</sup>.

Relationships among the main bivalve sub-lineages remained unresolved or uncertain until recently. With minor issues linked to the position of Nuculanida and Anomalodesmata, two main hypotheses have been put

Department of Biological, Geological and Environmental Sciences, University of Bologna, Via Francesco Selmi, 3, 40126 Bologna, Italy. ✉email: federico.plazzi@unibo.it



**Figure 1.** The two main alternative resolutions of the Bivalvia phylogenetic tree. **(A)** The Heteroconchia hypothesis. **(B)** The Amarsipobranchia hypothesis.

forward: the Heteroconchia hypothesis, which involves a sister group relationship between Heterodonta and Palaeoheterodonta (Fig. 1A), and the Amarsipobranchia hypothesis, which involves the sister group relationship between Heterodonta and Pteriomorphia instead (Fig. 1B).

The traditional taxonomic view and morphological analyses of Autobranchia heralded the Heteroconchia hypothesis<sup>4,11,15,26,33,34</sup>; however, a closer relationship between Heterodonta and Pteriomorphia has been suggested following palaeontological evidence<sup>35–40</sup>. The Amarsipobranchia hypothesis was also highly supported by molecular phylogenetics, using mitochondrial markers<sup>15,17,24,41–43</sup>. Contrastingly, the Heteroconchia hypothesis is always supported when nuclear markers are used (either combined with morphological data or not), as well as by means of transcriptomics<sup>15,19,20,33,44–46</sup>. This is a clear example of mito-nuclear phylogenetic discordance<sup>47</sup>.

**The OXPHOS genes and mito-nuclear coevolution.** The massive ATP production of aerobic respiration in eukaryotes is mostly made possible through the oxidative phosphorylation (OXPHOS) pathway, which takes place across the inner mitochondrial membrane. OXPHOS pathway is carried out by five enzymatic complexes (CI–V). The genes encoding for the subunits are mostly located in the nuclear genome (around 70 genes), while 13 genes are typically harbored in the mitochondrial genome (mtDNA), at least in most bilaterians. All the complexes but Complex II (CII) have cooperating subunits that are encoded by genes that are located on two different genomes, which show different mutation rate, population size and way of inheritance<sup>48</sup>.

In particular, the low recombination rate of mtDNA leads to the accumulation of slightly deleterious mutations<sup>49</sup>. This process would affect the efficiency of OXPHOS, but slightly negative mutations can be counterbalanced by compensatory mutations in the nuclear genes<sup>50</sup> or even by new nuclear subunits added to the OXPHOS complexes<sup>51</sup>. According to this model of mito-nuclear coevolution, the process is driven by the accumulation of slightly deleterious mitochondrial mutations, which affects the selective pressure on the interacting nuclear subunits. Indeed, a correlation between the amino acid substitution rate of mitochondrial genes and their interacting nuclear counterparts was shown<sup>52–54</sup>. The evolutionary rate correlation (ERC)<sup>55</sup> analysis is considered highly reliable to detect signals of mito-nuclear coevolution<sup>56</sup> and bivalves are among the clades where a positive ERC has been identified<sup>56–58</sup>.

Quite surprisingly, the Amarsipobranchia clade is also supported by nuclear genes encoding for the OXPHOS subunits<sup>57</sup>. Moreover, nuclear and mitochondrial OXPHOS genes show significant ERC and a similar dN/dS ratio<sup>57</sup> (the ratio between nonsynonymous substitution rate and the synonymous substitution rate<sup>59</sup>).

The mtDNA of bivalves has a highly variable architecture, showing features that are unique among metazoans. Gene order is not conserved inside the class and the high frequency of rearrangements prevents to infer an ancestral gene order for Autobranchia<sup>60</sup>. Among Protobranchia, in the mitochondrial sequence of *Solemya velum*

the leading strand, which is also the AC-rich one, harbors the genes *co1*, *co2*, *co3*, *nadh1*, *nadh2*, *nadh4*, *nadh4L* and *nadh5*, whereas the other strand harbors the genes *atp8*, *atp6*, *cytb*, *nadh1* and *nadh6*<sup>43</sup>. Among Bivalvia, this is likely the most ancestral gene arrangement<sup>43</sup>.

In Palaeoheterodonta the genome organization is highly conserved, and notable rearrangements were never detected within this subclass. Most of the protein coding genes are retained on the GT-rich strand (*atp6*, *atp8*, *co1*, *co2*, *co3*, *nadh3*, *nadh4*, *nadh4L* and *nadh5*), whereas the other strand harbors *cytb*, *nadh1*, *nadh2* and *nadh6*<sup>61</sup>.

Heterodonta and Pteriomorpha show a high degree of rearrangement. Few blocks of genes are shared between different orders of the same clade, and sometimes even among the same family<sup>62</sup>. It is however worth noting that in Heterodonta and Pteriomorpha all genes are retained on one strand, which is rich in G + T. As a matter of fact, the unidirectional replication of the mitochondrial genome leads to an asymmetric nucleotide composition of the two strands, increasing the G + T content in the heavy strand<sup>63</sup>. Most metazoans harbor most of the genes on the light strand, which is rich in A + C, but mollusks show an inverted pattern, in that in these species most of the genes are located on the GT-rich strand<sup>64,65</sup>. The position of mitochondrial genes on different strands has already been reported as a source of phylogenetic artifacts<sup>64,66</sup>. Thus, there could be a relationship between the diverging phylogenetic signal of the mitochondrial markers and the location of some genes in Palaeoheterodonta compared to Heterodonta and Pteriomorpha.

In this study, we performed a phylogenetic analysis using mitochondrial (mt-OXPPOS) and nuclear OXPPOS (nu-OXPPOS) markers, exploiting different phylogenetic approaches. For the sake of comparison, we added two more datasets: genes related to the glycolytic pathway and the genes related to the biogenesis of regulative small noncoding RNAs (sncRNAs). We also analyzed different features of markers selected for phylogenies: how the phylogenetic signal is distributed along the genes, codon usage, amino acid composition and strand location of the markers. We tested possible relationships between these features and the retrieved phylogenetic signals.

Regardless of the phylogenetic method, the Amarsipobranchia are supported only by the OXPPOS markers, both nuclear and mitochondrial. This phylogenetic signal is mostly retained in the organellar markers; among nuclear genes, subunits in direct contact with the mitochondrial counterparts lend most support to this topology. Moreover, we report an unbalanced nucleotide and amino acid composition between Amarsipobranchia and the Palaeoheterodonta, with a higher guanine and thymine content in the latter clade. We suggest that this pattern might be related to a different transcriptional mechanism, which has driven the mitochondrial phylogenetic signal to support Amarsipobranchia.

## Results

**The phylogenetic analysis on the four datasets.** The datasets were comprised by 35 species, for four species two mitochondrial haplotypes were sampled (i.e., the female and male mitochondrial haplotypes; see below) (Table 1). All four datasets were incomplete, glycolysis being the most incomplete matrix (Table S1). Conversely, the mt-OXPPOS dataset was the most complete. Species showed a different range of completeness as well: *Myzuhopecten yessoensis* was the most complete species, while the outgroup *Graptacme eborea* was the least complete species (Fig. S1). After the masking step, the mt-OXPPOS dataset was the shortest but also that with the lowest number of discarded sites. The longest dataset was the glycolysis one; the sncRNAs dataset was that with the highest number of discarded sites (Table S1).

The three maximum-likelihood (ML) trees and the single Bayesian tree inferred from the mt-OXPPOS dataset were never significantly different and did not show any alternative resolution of major clades (Fig. 2a, Fig. S2 and Table S2). Protobranchia were basal, exception made for *Aequiyoldia eightsii* (Nuculanida), which clusters within Amarsipobranchia. Autobranchia were fully supported by all four trees. Then, the tree was divided into Amarsipobranchia and Palaeoheterodonta, both fully supported. The Amarsipobranchia were divided into Heterodonta and a clade comprised by *A. eightsii* and Pteriomorpha. Within this clade a polytomy between *A. eightsii*, Mytilida (*Perna viridis*, *Bathymodiolus azoricus*, *Mytilus edulis*) and the other pteriomorphians was recovered. Heterodonta were split into Imparidentia and Anomalodesmata, both fully supported.

The ML and Bayesian trees inferred from the nu-OXPPOS dataset were never significantly different and did not show any alternative resolution of major clades (Fig. 2b, Fig. S3 and Table S2). Protobranchia were basal, but monophyletic in the MrBayes tree only (Fig. 2b); according to the other trees this group was not monophyletic or not robustly supported (Fig. S3). As for the mt-OXPPOS dataset, Autobranchia were split into Palaeoheterodonta and monophyletic Amarsipobranchia. Amarsipobranchia were divided into Pteriomorpha and Heterodonta, and the latter clade was split into Anomalodesmata and Imparidentia; all these clades were fully supported. Within Pteriomorpha, Mytilida are the sister group of remaining OTUs.

The ML and Bayesian trees inferred from the sncRNAs dataset were never significantly different and did not show any alternative resolution of the main clades (Fig. 3a, Fig. S4 and Table S2). Overall, several phylogenetic relationships were not resolved and some species were placed in unexpected major clades. After the separation of *Ennucula tenuis*, there was a polytomy with 6 branches: Heteroconchia; Mytilida + Ostreida, exception made for *Pinna atropurpurea*; Pectinida; *A. eightsii* + *P. atropurpurea*; *Tegillarca granosa*; *Solemya velum* (Fig. 3a). Heteroconchia were divided into Palaeoheterodonta and Heterodonta. Heterodonta were split into Anomalodesmata and Imparidentia, even if within the latter clade the palaeoheterodont *Margaritifera margaritifera* was recovered, which does belong to freshwater mussels.

The ML and Bayesian trees inferred from the glycolysis dataset were never significantly different and did not show any alternative resolution of major clades (Fig. 3b, Fig. S5 and Table S2). A long branch led to the Bivalvia node, which further separated Pteriomorpha from other bivalves, leading to the paraphyly of Autobranchia. Namely, Protobranchia and Heteroconchia clustered into a monophyletic group that was supported by all four trees. Heteroconchia were split into Palaeoheterodonta and Heterodonta. The latter clade was divided in

Clade	Order	Family	Species
Protobranchia	Nuculida	Nuculanidae	<i>Ennucula tenuis</i>
	Solemyida	Solemyidae	<i>Solemya velum</i>
	Nuculanida	Sareptidae	<i>Aequiyoldia eightsii</i>
Pteriomorphia	Pectinida	Pectinidae	<i>Amusium pleuronectes</i>
	Pectinida	Pectinidae	<i>Mizuhopecten yessoensis</i>
	Arcida	Arcidae	<i>Tegillarca granosa</i>
	Ostreida	Ostreidae	<i>Magallana angulata</i>
	Ostreida	Ostreidae	<i>Saccostrea glomerata</i>
	Ostreida	Pinnidae	<i>Pinna atropurpurea</i>
	Ostreida	Margaritidae	<i>Pinctada margaritifera</i>
	Mytilida	Mytilidae	<i>Bathymodiolus azoricus</i>
	Mytilida	Mytilidae	<i>Mytilus edulis</i> (F and M)
Palaeoheterodonta	Unionida	Unionidae	<i>Cristaria plicata</i> (F and M)
	Unionida	Unionidae	<i>Lampsilis cardium</i>
	Unionida	Unionidae	<i>Sinohyriopsis cumingii</i> (F and M)
	Unionida	Margaritiferidae	<i>Margaritifera margaritifera</i>
	Trigoniida	Trigoniidae	<i>Neotrigonia margaritacea</i>
Anomalodesmata	Laternulidae	Pandorida	<i>Laternula elliptica</i>
	Lyonsiidae	Pandorida	<i>Lyonsia floridana</i>
Imparidentia	Venerida	Acticidae	<i>Arctica islandica</i>
	Venerida	Cyrenidae	<i>Corbicula fluminea</i>
	Venerida	Mactridae	<i>Mactra chinensis</i>
	Venerida	Veneridae	<i>Paratapes textilis</i>
	Venerida	Veneridae	<i>Ruditapes philippinarum</i> (F and M)
	Venerida	Veneridae	<i>Ruditapes decussatus</i>
	Venerida	Glossidae	<i>Glossus humanus</i>
	Myida	Myidae	<i>Mya arenaria</i>
	Sphaeriida	Sphaeriidae	<i>Sphaerium nucleus</i>
	Adapendonta	Pharidae	<i>Simonovacula constricta</i>
	Galeommatida	Galeommatidae	<i>Galeomma turtoni</i>
Outgroups	Dentaliida	Dentalidae	<i>Graptacme eborea</i>
	Octopoda	Octopodidae	<i>Octopus bimaculoides</i>
	Chitonida	Acanthochitonidae	<i>Acanthochitona crinita</i>
	Lepetellida	Haliotidae	<i>Haliotis tuberculata</i>

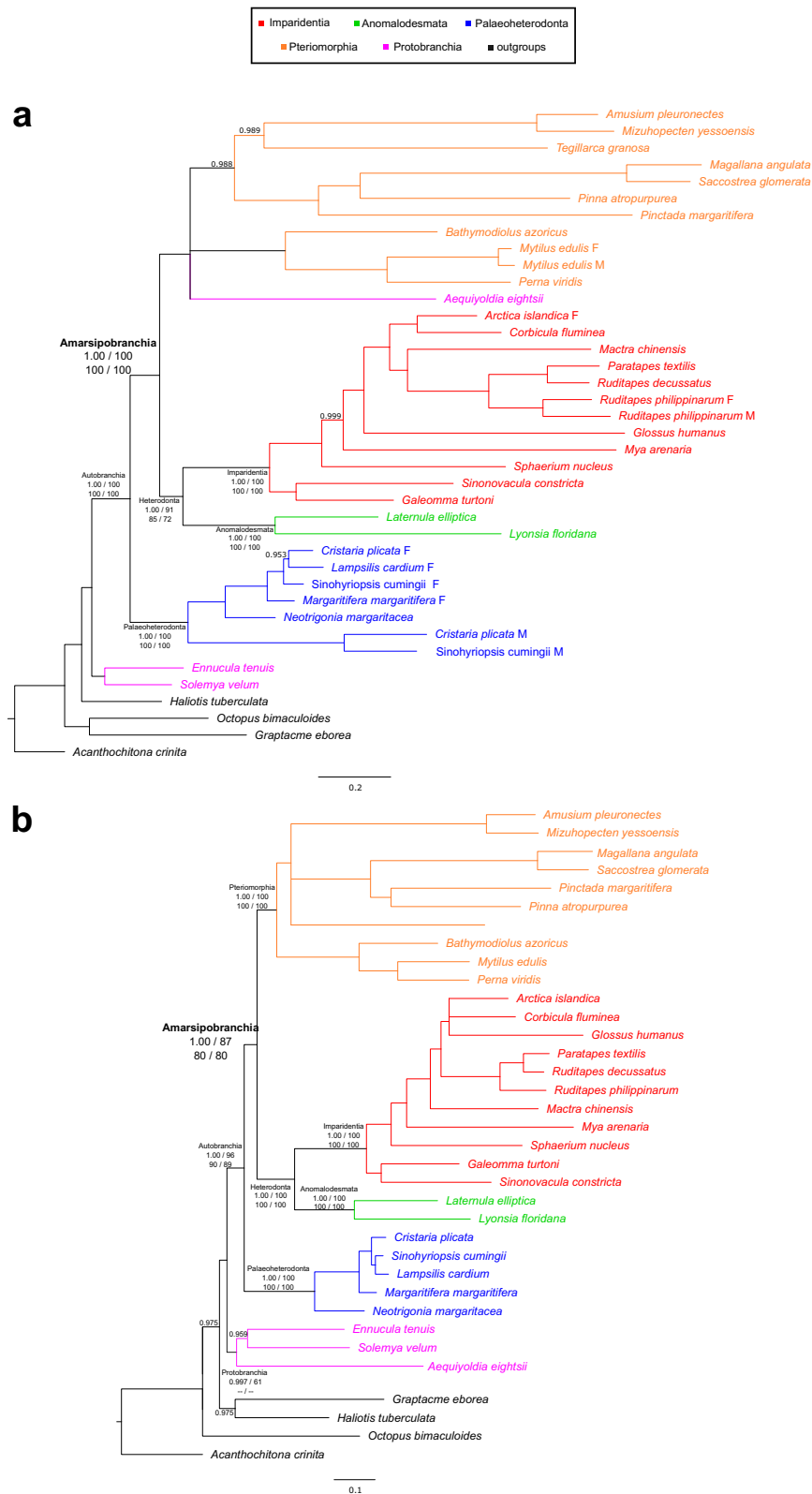
**Table 1.** List of species included in the phylogenetic analysis divided by higher classification taxa, orders and families according to Carter and colleagues<sup>118</sup> and WoRMS database<sup>119</sup>.

Anomalodesmata and Imparidentia; all these clades were fully supported. Within major clades all relationships were resolved and supported and the pteriomorphian and imparidentian species clustered in the expected orders.

Concluding, notwithstanding some issues with the major clade of Protobranchia which blurred the comparison and the substantial overlapping of all phylogenetic trees, the Amarsipobranchia hypothesis was supported in both OXPPOS datasets, while the Heteroconchia hypothesis was supported in the glycolysis and sncRNAs datasets. Henceforth, we will use mt-topology to refer to the Amarsipobranchia hypothesis and nuc-topology for the Heteroconchia hypothesis.

**Phylogenetic signal and its distribution across markers and complexes.** Markers belonging to the same dataset may support a different phylogenetic signal. Gene concordance factor (gCF), site concordance factor (sCF)<sup>67</sup> and ultrafast bootstrap approximation<sup>68</sup> (UFBoot) were calculated for the Heteroconchia and Amarsipobranchia (which represent alternative resolutions of a node). The mt-OXPPOS dataset showed high support for Amarsipobranchia according to each value (UFBoot = 100; gCF = 30.8; sCF = 48.6), and low support for the Heteroconchia (UFBoot = 0; gCF = 0; sCF = 25.5). Despite a non-zero gCF suggests more markers concordant with the nuc-topology than with the mt-topology, the nu-OXPPOS dataset similarly favors mt-topology (UFBoot = 87; gCF = 3.57; sCF = 37.2) against nuc-topology (UFBoot = 12; gCF = 5.36; sCF = 32.5). Regarding the sncRNAs and glycolysis datasets, markers are more concordant with Heteroconchia, since the UFBoot, gCF and sCF calculated for this topology are considerably higher (Table S3).

For the two OXPPOS datasets we clustered the markers according to the OXPPOS complexes; the sCF for each complex was computed; moreover, it was compared to the sitewise log-likelihood score (SLS) calculated for



**Figure 2.** Bayesian trees inferred from the two OXPPOS datasets. **(a)** The mt-OXPPOS tree inferred through MrBayes. **(b)** The nu-OXPPOS tree inferred through MrBayes. Notably, both trees support the Amarsipobranchia hypothesis. The posterior probability on each node is reported when lower than 1.00; nodes with posterior probability lower than 0.95 were collapsed. Major nodes are annotated and support values of each of the four trees inferred for the present work are shown, as follows: MrBayes posterior probability, partitioned and mixture-model IQ-TREE UFBoot values, and RAXML bootstrap value. A double dash instead of the support means that the clade is not monophyletic in that tree. Red, Imparidentia; green, Anomalodesmata; blue, Palaeoheterodonta; orange, Pteriomorpha; purple, Protobranchia; and outgroups are shown in black.

**Figure 3.** Bayesian trees inferred from the sncRNAs and glycolysis datasets. (a) The sncRNAs tree inferred through MrBayes. (b) The glycolysis tree inferred through MrBayes. Notably, both trees support the Heteroconchia hypothesis. The posterior probability on each node is reported when lower than 1.00; nodes with posterior probability lower than 0.95 were collapsed. Major nodes are annotated and support values of each of the four trees inferred for the present work are shown, as follows: MrBayes posterior probability, partitioned and mixture-model IQ-TREE UFBoot values, and RAxML bootstrap value. A double dash instead of the support means that the clade is not monophyletic in that tree. Red, Imparidentia; green, Anomalodesmata; blue, Palaeoheterodonta; orange, Pteriomorpha, purple, Protobranchia; outgroups are shown in black.

both topologies. The difference between the mt-topology sitewise log-likelihood score and the nuc-topology sitewise log-likelihood score ( $\Delta$ SLS) can tell which topology is favored by each site: sites with  $\Delta$ SLS > 0 support the mt-topology; sites with  $\Delta$ SLS < 0 support the nuc-topology. Moreover, by summing all the  $\Delta$ SLS within a complex we obtained a complexwise log-likelihood score<sup>69,70</sup> ( $\Delta$ CLS; Table 2). Since the summed  $\Delta$ CLS highly depends on the number of sites within each complex, we divided the  $\Delta$ CLS for the number of sites of each complex (average  $\Delta$ CLS). For the mitochondrial markers that belong to CI we made a distinction between those *nadh* genes that in Palaeoheterodonta are on the plus strand (CI-ps) from those *nadh* genes located on the minus strand (CI-ms), since we were willing to test if the mt-topology is mostly supported in the genes that are in different strands in Palaeoheterodonta and Amarsipobranchia (i.e. *cytb*, *nadh1,2,6*; see “Introduction”).

All the mitochondrial groups (Table 2) show a positive  $\Delta$ CLS; a positive average  $\Delta$ CLS; more sites that strongly support the mt-topology; more sites in the alignment that agree with the mt-topology. The only exception is CI-ps, where the  $\Delta$ CLS and average  $\Delta$ CLS are negative, although the other statistics follow the pattern of the other groups.

Complexes III to V of the nu-OXPHOS dataset (Table 2) support Amarsipobranchia; sites that strongly support the mt-topology (with  $\Delta$ SLS > 0.5) are more than those supporting the nuc-topology and most sites in the alignment agree with the mt-topology. Contrastingly, CI and CII do not support Amarsipobranchia. In CII there is an equal number of sites for either topology, while in the CI those with a  $\Delta$ SLS > 0.5 are more. The sCF calculated on the nuc-topology is higher in CII and almost equal in CI with respect to the sCF calculated on the mt-topology.

Overall, in all complexes  $\Delta$ CLS, average  $\Delta$ CLS and sCF variate together; statistics related to the strongly supporting sites do not always follow the same pattern, since CI shows a negative  $\Delta$ CLS but a higher number of sites with  $\Delta$ SLS > 0.5.

To test whether the mt-topology phylogenetic signal is mostly retained in the nu-OXPHOS subunits that interact with the mitochondrial subunits, we calculated the sCF referred to each marker and we split the markers into two groups: those that are in direct contact with the mitochondrial counterparts and those that are not. The sCF values of the “contact” nu-OXPHOS markers are significantly higher than the values of “non-contact” nu-OXPHOS makers (*p* value = 0.006363; Fig. 4).

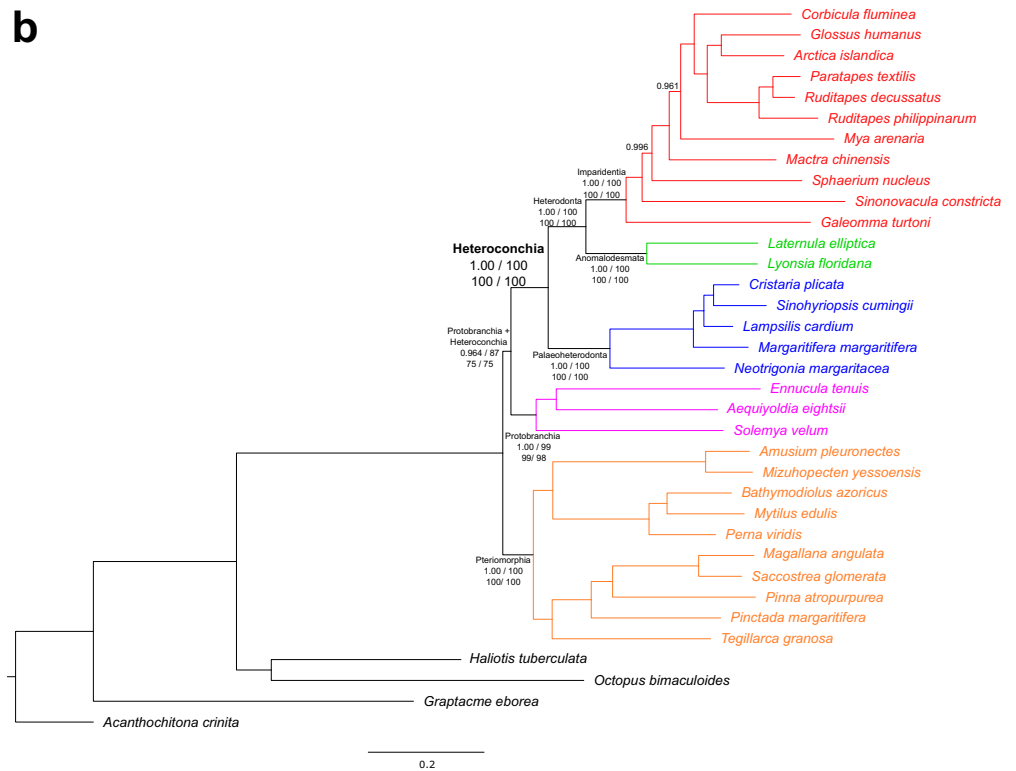
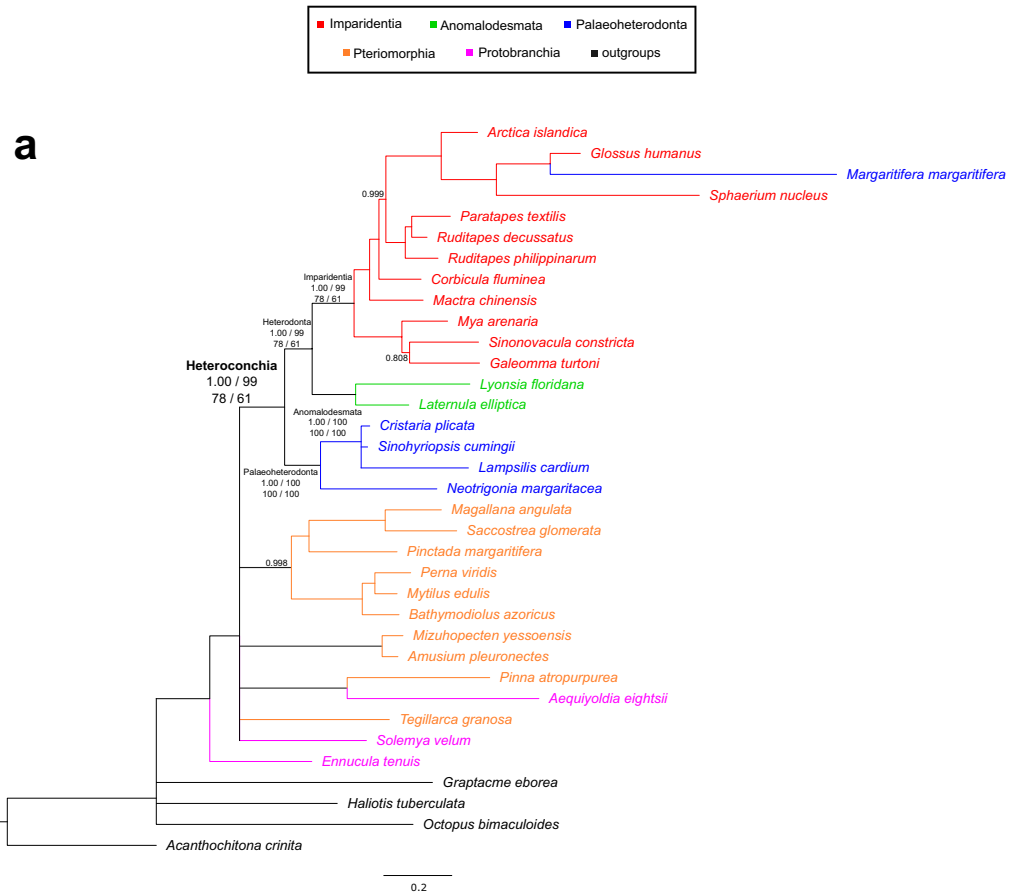
In the sncRNAs and glycolysis datasets the  $\Delta$ SLS was calculated on the whole matrix. In both datasets the average  $\Delta$ SLS is negative and there are more sites strongly supporting Heteroconchia (Table S3).

**Nucleotide composition and mitochondrial topology.** We placed attention on the nucleotide asymmetry between the two mitochondrial strands, which can be assessed calculating the AT skew and GC skew<sup>71</sup>. In Bivalvia the plus strand is richer in guanines and thymines than the minus strand<sup>64,72</sup>. Thus, we also analyzed possible dissimilarities in the guanine and thymine content (G + T content) between markers that in Pteriomorpha, Imparidentia, Anomalodesmata and Palaeoheterodonta are on the same strand (i.e., *atp6,8*, *cox1-3*, *nadh3-5*; Table S4).

For each mt-OXPHOS marker of each species we calculated the AT skew, the GT content, the frequency of codons with guanines or thymines at the first and the second position (GT-rich codons) and the GT content at the third position of four-fold degenerated codons (Fig. 5). Among Imparidentia, Anomalodesmata, Pteriomorpha and Palaeoheterodonta the markers show an AT skew < 0 and a GT content > 0.5. On average, Palaeoheterodonta show the highest values in all the statistics but the AT skew (Fig. 5a). Indeed, Palaeoheterodonta are always significantly different from the other groups, with the only exception of Anomalodesmata in GT-rich codons (Fig. 5b). On the other hand, the comparisons between Pteriomorpha and Imparidentia are never significant. Regarding the outgroups and Protobranchia values, data show a high standard deviation in most of the cases. The only exception is *A. eightsii*, whose statistics are in line with the values of Imparidentia and Pteriomorpha.

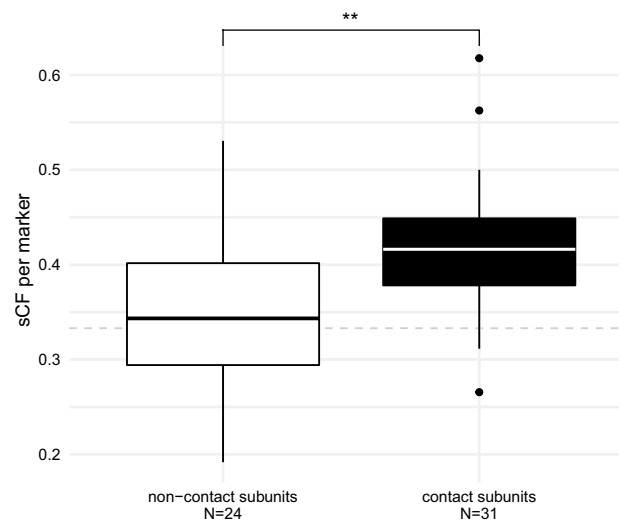
Finally, we studied if the nucleotide compositional patterns outlined in the protein coding regions were extended to the unassigned regions (URs): we downloaded the mitochondrial genomes available on NCBI of all the species that belong to Imparidentia, Anomalodesmata, Palaeoheterodonta, Pteriomorpha and Protobranchia. Then, we calculated the GT content in the URs of the genomes. The GT content of URs calculated on 92 Palaeoheterodonta entries is significantly higher than the one calculated on 77 Pteriomorpha entries, 70 Imparidentia entries and 4 Protobranchia entries. Conversely, it is not significantly higher than the one calculated on 6 Anomalodesmata entries (Fig. S6). For what concerns the other comparisons, no clade is significantly different from any other.

Overall, the nucleotide composition of Palaeoheterodonta mt-OXPHOS markers is most of the times significantly different from the one of other major clades. In particular, we detected a higher GT content. This pattern is reflected in all codon positions as well as in the URs. On the other hand, statistics are overlapping between Imparidentia and Pteriomorpha.



Group	$\Delta$ CLS	Average $\Delta$ CLS	% $\Delta$ SLs > 0.5 (%)	% $\Delta$ SLs < -0.5 (%)	mt-sCF	nuc-sCF
<b>nu-OXPPOS dataset</b>						
CV	20.1	0.0079	0.90	0.35	40.9	30.7
CIV	6.2	0.0043	1.39	1.04	42.6	31.0
CIII	8.9	0.0097	1.19	0.59	42.1	33.6
CII	-7.9	-0.0073	0.83	0.83	32.9	36.3
CI	-1.0	-0.00019	0.96	0.72	35.4	35.0
<b>mt-OXPPOS dataset</b>						
CV	5.9	0.0380	3.22	1.29	48.1	26.6
CIV	25.3	0.0272	3.11	0.75	54.6	22.8
CIII ( <i>cytb</i> )	6.3	0.0181	1.97	0.28	42.6	28.4
CI-ms	11.4	0.0110	2.14	0.77	49.7	24.8
CI-ps	-7.2	-0.0090	1.62	0.85	38.1	32.4

**Table 2.** The phylogenetic signal of nu and mt-OXPPOS markers grouped by complexes. The CI mt-markers are split into two groups: CI-ms is comprised by *nadh1,2,6* and CI-ps is comprised by *nadh3,4,4l,5*. For each group it was calculated:  $\Delta$ CLS; average  $\Delta$ CLS; percentage of sites with % $\Delta$ SLs > 0.5; percentage of sites with % $\Delta$ SLs < -0.5; sCF for the mt-topology; sCF for the nuc-topology.



**Figure 4.** The phylogenetic signal in “contact” and “non-contact” nu-OXPPOS subunits. Boxplot comparing the sCF of nu-OXPPOS markers in direct contact with mitochondrial subunits and the sCF of nu-OXPPOS markers that are not in direct contact. Dashed gray line at 0.33 marks the threshold below which the branch with the highest figure of sCF between the three possible resolutions is not the one that support the mt-topology. Significance calculated through a Student’s t-test ( $t = -2.862$ ,  $p$  value = 0.006363, d.f. = 23, 30).

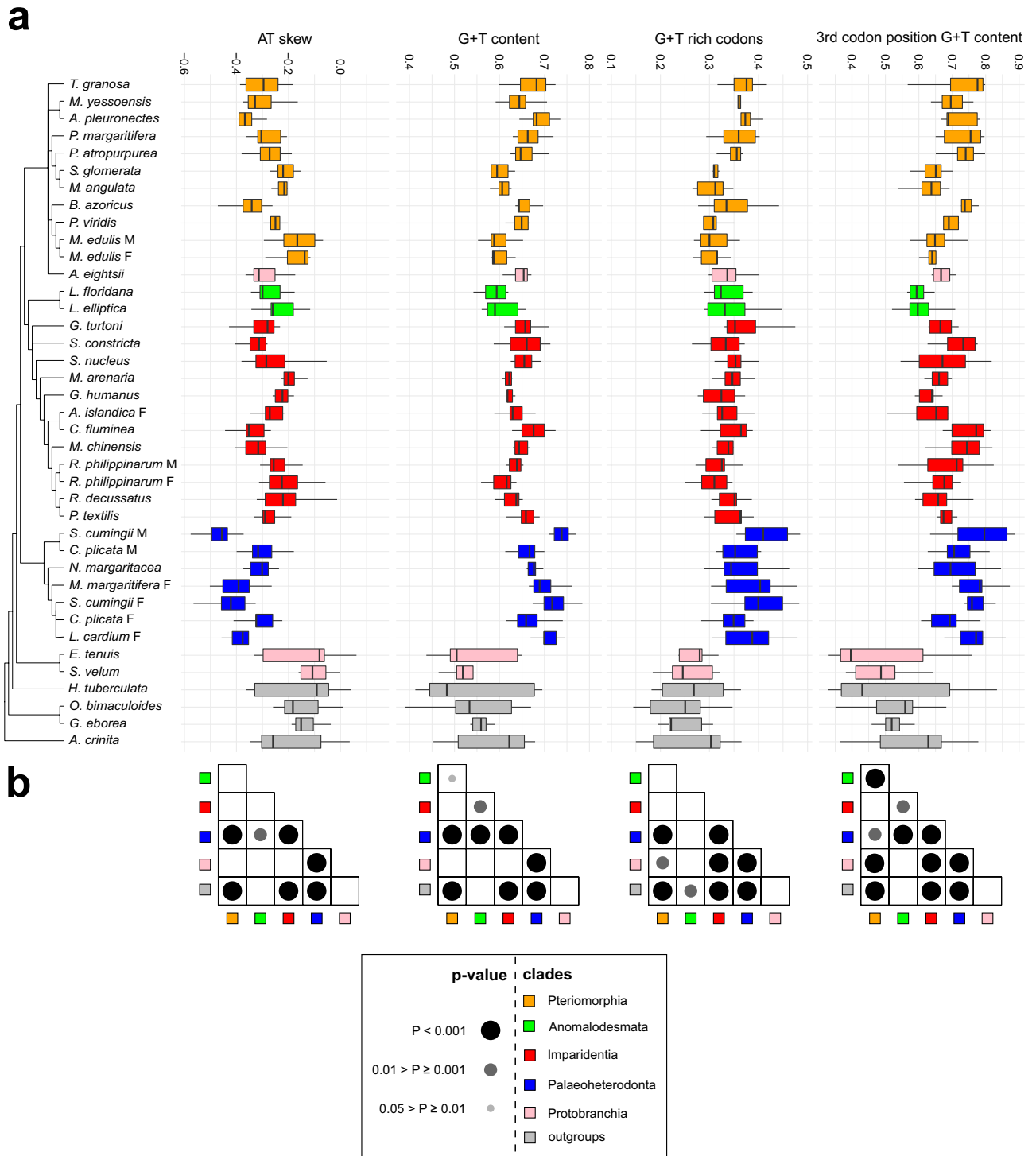
## Discussion

For all four datasets, the more recent nodes were resolved and highly supported. In Imparidentia and Pteriomorpha the OTUs were always placed in the expected orders and major clades (i.e. Pteriomorpha, Anomalodesmata, Imparidentia, Palaeoheterodonta and Protobranchia). Only few exceptions were detected, i.e. the position of *A. eightsii* in the sncRNAs and mt-OXPPOS trees and the position of *M. margaritifera* in the sncRNAs trees. The latter was likely a long branch attraction bias<sup>73</sup>, since the final branches of the OTU and its sister species were the longest in the tree (Fig. 3a). Generally speaking, we regard to these misplacements as minor phylogenetic issues in the broader figure of deep evolutionary relationships among bivalves, which do not significantly blur the topology connecting major clades.

Major clades were retrieved with higher support and with better resolution from the OXPPOS datasets with respect to the glycolysis and sncRNAs dataset. Overall, OXPPOS genes are known to be more conservative, therefore these markers might be more informative in the resolutions of cladogenetic events dating to the Ordovician, approximately 470–480 million years ago (Mya)<sup>36,38,39</sup>.

The mt-OXPPOS trees were mostly coherent with the previous mitochondrial phylogenetic analyses, exception made for the monophyly of the Heterodonta with Anomalodesmata inferred from our analysis<sup>17,24,41,42,57,64,74</sup>. Since all the nu-OXPPOS trees supported the Amarsipobranchia clade (Fig. 1b), our data confirmed that the





**Figure 5.** Nucleotide and codon composition statistics in mt-OXPHOS markers compared across OTUs. (a) OTUs are reported according to the mitochondrial consensus tree. The x-axis is divided in four boxplots with no outliers, each one reports a different statistic calculated on a set of mt-OXPHOS markers (*atp6,8, cox1-3, nadh3-5*). From left to right, plots report the AT skew, the GT content, the frequency of codons that have T or G at first and second codon position, and the GT content of the third codon position in four-fold degenerated codons, respectively. (b) Each table report the significance of pair-wise comparisons between the values reported in the plot right above grouped according to the six clades. The significance is calculated through the Dunn test with the Bonferroni correction. Black and grey dots inside the table mark the significant comparisons; as reported in the legend, the bigger and the darker the dot, the more significant the comparison.

mt and nu-OXPPOS markers share the same phylogenetic signal, which is different from that inferred from transcriptome-wide analyses or other nuclear markers<sup>14,19,33,44,45</sup>.

Among the interacting sites of coevolving proteins there are epistatic interactions, which lead the sites of both proteins to evolve at the same rate<sup>75</sup>. Bivalvia OXPPOS subunits show a positive ERC, which is the most solid clue of protein coevolution<sup>55,56</sup>. Our data enforce the hypothesis of mito-nuclear coevolution in bivalves, depicting a clear relationship between the phylogenetic signal of interacting subunits. Moreover, they provide an overview on how the phylogenetic signal of OXPPOS subunits may be biased under this type of interaction. The CF and  $\Delta$ SSL analyses suggested that OXPPOS markers did not equally support the Amarsipobranchia, yet the two dataset were largely coherent with each other: CIII-V markers from both OXPPOS datasets largely supported the mt-topology; contrastingly, CI did not show a clear pattern, and the nuclear-only Complex II favors the nuc-topology. Moreover, the nu-OXPPOS subunits in contact with the mitochondrial counterparts were significantly more concordant with the mt-topology than the subunits that are not directly in contact. Accordingly, previous analyses reported that the CII is the only complex that shows uncorrelated rates of evolution compared to the other subunits<sup>57</sup>.

Finally, the support and concordance statistics (BP, PP, UFBoot, gCF and sCF) calculated for the mt-OXPPOS dataset on the Amarsipobranchia node were always equal to or higher than those calculated for the nu-OXPPOS dataset. Thus, the mt-topology in the first dataset was more consistent: more sites and markers agreed with this topology and the signal was less susceptible to resampling.

The mito-nuclear coevolution is expected to be mainly driven by slightly deleterious mitochondrial mutations that are compensated by the nuclear genome<sup>48,50</sup>. Even if previous data did not show signal of nuclear compensation<sup>57</sup>, it is tempting to conclude that the mitochondrial genome acquired the mutations leading to the mt-topology at first, and then the phylogenetic signal has been traced by interacting sites in the nuclear markers through nuclear compensation.

Pteriomorphia and Imparidentia share some unique mitochondrial features: their gene order is highly rearranged, but all genes are on the same strand. Contrastingly, Palaeoheterodonta show a highly conserved gene order, with a set of genes on the minus strand (*nadh1,2,6* and *cytb*; see Introduction for further details). According to the nucleotide composition analyses, the Palaeoheterodonta mt-OXPPOS markers were significantly GT-richer in each codon position as well as in URs, while Pteriomorphia and Imparidentia did not show any significant difference (Fig. 5, Fig. S6). Thus, this pattern was accounted for either synonymous and non-synonymous substitutions and it was extended also to URs. Mito-nuclear coevolution largely explains why the nu and mt-OXPPOS markers support a common topology.

Mito-nuclear discordance is a quite common phenomenon and a multitude of processes can cause it<sup>76</sup>. The introgression of mitochondrial lines from a phylogenetically distant population is widely used to explain mito-nuclear discordance<sup>47,76</sup>. In some cases it has been hypothesized that a set of nuclear genes might cointrogress to avoid mito-nuclear incompatibilities<sup>77</sup>. The mito-nuclear cointrogression would explain very well our data, since the phylogenetic artifact is mostly supported by the nuc-OXPPOS markers that directly interact with the mt-OXPPOS markers, whereas almost all mt-OXPPOS markers support the mt-topology. In this case, the use of a single mitochondrial strand and other features would be apomorphies arisen along a single branch and subsequently acquired by the other branch through introgression. Having said that, other evidences of cointrogression are limited and only restricted to populations within the same genus<sup>77-79</sup>. In our case the discordance mainly resides in the resolution of deep nodes, which originated around 480 Mya<sup>5,36,80</sup>, between clades that already evolved quite different life habits<sup>5</sup>. Under this scenario, the mito-nuclear cointrogression might not be the most likely hypothesis.

Another source of mito-nuclear discordance can be found in how markers are located on the two mitochondrial strands<sup>66</sup>. In mollusks, whose mitochondrial genome is highly rearranged, the nucleotide bias is also reflected in amino acid bias<sup>64</sup>. Our results showed that the signal supporting the mt-topology (the Amarsipobranchia clade) is not only retained in the set of markers that in Palaeoheterodonta are on the minus strand. Instead, the mt-OXPPOS CIV-V markers on the plus strand clearly favor the Amarsipobranchia hypothesis (Table 2). The higher GT content in Palaeoheterodonta is consistent throughout different parts of the mitochondrial genome, from coding to unassigned regions. Therefore, the nucleotide substitutions that have led to this pattern are likely to be produced by a process that act on the whole genome. Possible candidates might be the mitochondrial transcription and replication, which are indeed notable source of deamination<sup>63,81</sup>. Moreover, mitochondrial replication constitutes the main source of mitochondrial point mutations, at least in humans<sup>82</sup>.

The position of all genes on the same strand is probably linked to the fact that the two clades do not show any significant difference in GT content. It is tempting to hypothesize that transcription involves the coding strand only for these mtDNAs and, thus, the aforementioned deamination effect may be less pronounced. Indeed, even those sncRNAs that were recently described in the imparidentian *R. philippinarum* were annotated on the same coding strand<sup>83</sup>, thus corroborating the idea that only one strand is transcribed in these clades.

The use of a single strand seems also linked to the mitochondrial architecture: among most of the metazoan taxon that share this feature it has been detected a higher mitochondrial rearrangement rate<sup>24,84,85</sup>. Likewise, Pteriomorphia and Imparidentia show highly rearranged mitochondrial genomes<sup>62</sup>. An additional clue is the behavior of *A. eightsii*: the protobranch species cluster with Pteriomorphia and shows similar nucleotide composition features (Fig. 2a, Fig. 5a). Indeed, although no mitochondrial genome has been annotated for the order Nuculanida, it is possible that *A. eightsii* mitogenome harbors all the genes on the heavy strand, since all its mitochondrial genes show AT skew < 0 and GC skew > 0 (Table S5).

If the hypothesis holds true, it is reasonable to consider the different transcriptional patterns among Bivalvia as the most likely source that has led the mitochondrial genome to support a different phylogenetic signal, namely a biased one. Since we demonstrated that also non-synonymous mutations have shaped the GT content pattern,

the modifications of the amino acid sequences could have altered the epistatic interactions between nuclear and mitochondrial OXPPOS subunits, leading the OXPPOS markers to support the same phylogenetic artifact.

## Conclusions

The results obtained from the phylogenetic analysis of Piccinini and colleagues<sup>57</sup> has been confirmed by our work, since markers of both OXPPOS datasets support the same biased topology, regardless of the phylogenetic pipeline used. Moreover, we depicted how the coevolution process affected the phylogenetic signal in different set of OXPPOS markers, concluding that the artifactual topology is mainly supported by the OXPPOS subunits that interact more directly.

Considering that the phylogenetic signal is more stable and stronger in the mt-OXPPOS markers, we suggest that the biased topology arose for these markers at first, then it has been acquired also by the nu-OXPPOS markers through the coevolution of interacting subunits. This model agrees with the pattern of evolution hypothesized for the mito-nuclear coevolution. That is, the mito-nuclear coevolution is mainly driven by slightly deleterious mitochondrial mutations that are compensated by the nuclear genome<sup>48,50</sup>.

Our data suggest a relationship between the mt-topology supporting Amarsipobranchia and the gene rearrangements in the Bivalvia mitochondrial genome. The clades that harbor all the mitochondrial genes on a single strand and show a similar nucleotide composition (Pteriomorpha, Heterodonta, and possibly *A. eightsii*) are grouped together in a monophyletic clade. On the other side, Palaeoheterodonta show a peculiar nucleotide composition, which is not only due to the genes located on the minus strand. Indeed, genes such *cox1-3*, *atp6,8*, *nadh3-5*, even if they are located on the plus strand, show a higher GT content compared to the Amarsipobranchia ones. Overall, the difference in GT content between OTUs may be a source of possible phylogenetic artifacts. Further analyses will be focused on understanding how the nuclear subunits compensated differently during the evolution of Palaeoheterodonta, Pteriomorpha and Heterodonta.

Finally, according to the data, the reliability of the Amarsipobranchia clade should be reconsidered. At the state of the art, although many mitochondrial phylogenies confirmed the Amarsipobranchia clade<sup>24,57,64,74</sup>, no phylogeny supports Amarsipobranchia when based on nuclear markers (exception made for the nu-OXPPOS markers<sup>57</sup>; Fig. 2b, Fig. S2). On the other side, the Heteroconchia clade has been retrieved by genome-wide, transcriptomic, and morphological analyses<sup>19,33,44,45</sup>. If the evolutionary scenario depicted in our discussion is correct, then the taxon Amarsipobranchia cannot be supported anymore and has to be considered a phylogenetic artifact: the Heteroconchia clade should be regarded as a more reliable hypothesis instead.

## Materials and methods

**The datasets.** Our phylogenetic analyses were performed on four datasets: mt and nu-OXPPOS genes, glycolytic pathway genes, and genes related to the biogenesis of sncRNAs. All markers were retrieved from the transcriptomes used by Piccinini and colleagues<sup>57</sup>: the transcriptomes of 35 molluscan species were assembled (Table 1). When available, the mt-OXPPOS markers from both sexes were retrieved for those species that show mitochondrial Doubly Uniparental Inheritance (DUI<sup>86–88</sup>).

Information about the assembly of transcriptomes is detailed in the aforementioned paper<sup>57</sup>. Briefly, the annotation of transcripts was performed using BLASTx<sup>89</sup> against a user-defined database and HMMER<sup>90</sup> against the Pfam database 30.0<sup>91</sup>; the user-defined database contains sequences of all genes of that dataset available for Bivalvia on NCBI.

Clam homologs for the first three datasets were extracted following the gene lists available in the Kyoto Encyclopedia of Genes and Genomes (KEGG<sup>92–95</sup>; <https://www.kegg.jp>), which provides a curated database of enzymes involved in specific biochemical pathways: namely, the Oxidative phosphorylation pathway (KEGG entry: map00190) and the Glycolysis/Gluconeogenesis pathway (KEGG entry: map00010). Regarding genes for the fourth dataset, i.e. genes related to the biogenesis of sncRNAs, we identified a set of genes shared across Metazoa<sup>96,97</sup>. Entries available on NCBI and UniProt<sup>98</sup> were included in the database (Table S6). Annotation was performed using BLASTp<sup>89</sup>.

Paralogs were recurrent among the markers associated to glycolysis. Therefore, we devised a method to conservatively distinguish paralogs from orthologs. We inferred the ML tree from each single marker putting orthologs together, which was obtained using IQ-TREE1.7<sup>99</sup> with mixture model as model of evolution, 1000 UFBoot<sup>68</sup> replicates, and constraining the Bivalvia clade. Through the analysis of topologies, more than one group of clear monophyletic orthologs were detected in some cases, namely in the markers with KEGG ID K00002, K00128, K00129, K00149, K00627, K00844, K01596, K01623, K01689, K01785, K01895, K03103, K08074, and K13953 (Table S7). In these cases, groups of orthologs were split and considered as single markers. Aiming to ensure that the phylogenetic signal supported by the glycolysis matrix after this scrutiny was coherent, we retained two different datasets associated to glycolysis genes: a larger dataset with all markers obtained in this way (total-glyco) and a dataset with markers that showed no evidence of paralogs (partial-glyco). All subsequent analyses were carried out independently for both datasets; since differences in results were negligible, we are confident that we identified paralogs correctly, thus in the results we mean the total-glyco dataset only when referring to the “glycolysis” dataset.

**Phylogenetic reconstruction.** We performed the phylogenetic analysis using amino acid sequences, since we were more interested in deep relationships and nucleic sequences are bound to saturate along long branches. First, we aligned sets of homologous markers with PSI-Coffee<sup>100</sup>. Then, to remove the uninformative or misleading sites for the analysis, we used and combined the results of five different masking algorithms<sup>24</sup>: BMGE<sup>101</sup>, Aliscore<sup>102</sup>, Gblocks<sup>103</sup>, ZORRO<sup>104</sup> and Noisy<sup>105</sup>. This step was performed by masking\_package 1.1,

downloaded from GitHub and available at [https://github.com/mozoo/masking\\_package](https://github.com/mozoo/masking_package). To include the indels in the phylogenetic reconstruction we ran GapCoder<sup>106</sup> on every alignment.

To assign the best-fitting evolutionary model to each marker of the matrix we used PartitionFinderProtein<sup>107</sup>. All markers belonging to the same dataset were concatenated together. For each dataset we obtained four trees. (1) One tree was obtained through IQ-TREE 1.7 with the dataset partitioned according to the PartitionFinder results. (2) One tree was obtained through IQ-TREE 1.7 with the mixture model as model of evolution<sup>99</sup>. (3) One tree was obtained through RAxML version 8.2.11<sup>108</sup> with the dataset partitioned according to the PartitionFinder results, using the CAT model instead of the Gamma model<sup>109</sup>. 1000 bootstrap replicates were executed for each run, to test the robustness of the nodes, and the UFboot approximation was chosen for IQ-TREE. (4) The fourth tree is based on the Bayesian inference, obtained through MrBayes<sup>110</sup> with the dataset partitioned according to the PartitionFinder results. Number of generations was set to 10,000,000; the convergence between runs were manually checked to set the burn-in value. To set this value, we looked at the standard deviation of average split frequency over generations; moreover, we took the Potential Scale Reduction Factor (PSRF<sup>111</sup>) into consideration. In each analysis the monophyly of Bivalvia was constrained and in the Bayesian analysis the outgroup was set to be the polyplacophoran *Acanthochitona crinita* (Table 1).

**Analyses on topologies and markers.** At the end of phylogenetic analysis, four trees were obtained for each dataset through four different pipelines, as described above. To test whether the trees obtained from the same dataset are significantly different or not, we performed the Shimodaira-Hasegawa test (SH-test<sup>112</sup>), exploiting the RAxML option “-f H”.

The support of each site for the Amarsipobranchia hypothesis (“mt-topology”) and the Heteroconchia hypothesis (“nuc-topology”) was calculated through the  $\Delta$ SLS. Sites with  $\Delta$ SLS > 0.5 or  $\Delta$ SLS < -0.5 were retained as sites with strong support for either hypothesis<sup>69,70</sup>. To calculate the sitewise log-likelihood we exploited the RAxML option “-f g” providing the RAxML ML tree when the sitewise log-likelihood was calculated on the mt-topology. A tree with the nuc-topology was obtained by running the phylogenetic analysis with the same settings, but constraining the Heteroconchia clade (as suggested by Shen and colleagues<sup>70</sup>), and the resulting ML tree was used to calculate the nuc-topology sitewise log-likelihood.

The sCF and the gCF were calculated through IQ-TREE 1.7 (again with 1000 UF-bootstrap replications) with the option “-cf-verbose” to study phylogenetic signal between and within partitions<sup>67</sup>. Each dataset was partitioned into single markers in order to calculate the sCF per marker and the gCF. Then, the matrices were partitioned according to complexes to obtain sCF per complex. The nu-OXPPOS subunits in direct contact with the mitochondrial counterparts were defined according to the list of Piccinini and colleagues<sup>57</sup>.

Custom-tailored python and R<sup>113</sup> scripts were used to analyze and plot the nucleotide and amino acid composition, using Biopython<sup>114</sup> and ggplot2. Since mitochondrial URs are missing from transcriptomes, their nucleotide composition was calculated for a list of NCBI indexes obtained through the alMight database<sup>65</sup>: a single entry was selected for each species in the database belonging to Palaeoheterodonta, Imparidentia, Anomalodesmata, Pteriomorpha or Protobranchia.

For DUI species the mtDNA of both sexes was selected. The guanine and thymine content in URs was obtained through a customized version of the HERMES tool<sup>115</sup>. The significance of the comparisons was calculated through the Kruskal and Wallis test<sup>116</sup>, followed by the Dunn’s test<sup>117</sup> with Bonferroni’s correction.

## Data availability

The data underlying this article are available in the GenBank Nucleotide Database at <https://www.ncbi.nlm.nih.gov/> and in the SwissProt database at <https://www.uniprot.org/>, and can be accessed with the accession numbers provided in the article and in the supplementary materials.

Received: 27 April 2022; Accepted: 17 June 2022

Published online: 30 June 2022

## References

- Gosling, E. M. *Bivalve Molluscs: Biology, Ecology and Culture* (2003).
- Morton, B. The evolutionary history of the Bivalvia. In *Origin and Evolutionary Radiation of the Mollusca* (ed. Taylor, J. D.) 337–359 (Oxford University Press, 1996).
- Cope, J. C. W. & Babin, C. Diversification of bivalves in the Ordovician. *Geobios* **32**, 175–185 (1999).
- Cope, J. C. W. Diversification and biogeography of bivalves during the Ordovician period. In *Palaeobiogeography and Biodiversity Change: The Ordovician and Mesozoic-Cenozoic Radiations* (eds Crame, J. A. & Owen, A. W.) 25–52 (Geological Society of London, 2002).
- Fang, Z. J. An introduction to Ordovician bivalves of southern China, with a discussion of the early evolution of the Bivalvia. *Geol. J.* **41**, 303–328 (2006).
- Tsubaki, R., Kameda, Y. & Kato, M. Pattern and process of diversification in an ecologically diverse epifaunal bivalve group Pterioidea (Pteriomorpha, Bivalvia). *Mol. Phylogenet. Evol.* **58**, 97–104 (2011).
- Yonge, M. The protobranchiate mollusca; A functional interpretation of their structure and evolution. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **230**, 79–147 (1939).
- Stasek, C. R. Synopsis and discussion of the association of ctenidia and labial palps in the bivalve Mollusca. *Veliger* **6**, 91–97 (1963).
- Starobogatov, Y. I. Morphological basis for phylogeny and classification of Bivalvia. *Ruthenica* **2**, 1–25 (1992).
- von Salvini-Plawen, L. & Steiner, G. Synapomorphies and plesiomorphies in higher classification of Mollusca. In *Origin and Evolutionary Radiation of the Mollusca* (ed. Taylor, J. D.) 29–51 (Oxford University Press, 1996).
- Waller, T. R. Origin of the molluscan class Bivalvia and a phylogeny of major groups. In *Bivalves: An Eon of Evolution* (eds Johnston, P. A. & Haggart, J. W.) 1–45 (University of Calgary Press, 1998).

12. Steiner, G. & Hammer, S. Molecular phylogeny of the Bivalvia inferred from 18S rDNA sequences with particular reference to the Pteriomorpha. *Geol. Soc. Lond. Spec. Publ.* **177**, 11–29 (2000).
13. Passamanek, Y. J., Schander, C. & Halanych, K. M. Investigation of molluscan phylogeny using large-subunit and small-subunit nuclear rRNA sequences. *Mol. Phylogenet. Evol.* **32**, 25–38 (2004).
14. Giribet, G. & Wheeler, W. On bivalve phylogeny: A high-level analysis of the Bivalvia (Mollusca) based on combined morphology and DNA sequence data. *Invertebr. Biol.* **121**, 271–324 (2002).
15. Giribet, G. & Distel, D. L. Bivalve phylogeny and molecular data. In *Molecular Systematics and Phylogeography of Mollusks* (eds Lydeard, C. & Lindberg, D.) 45–90 (Smithsonian Institution Press, 2003).
16. Bieler, R. & Mikkelsen, P. M. Bivalvia—A look at the branches. *Zool. J. Linn. Soc.* **148**, 223–235 (2006).
17. Plazzi, F. & Passamonti, M. Towards a molecular phylogeny of Mollusks: Bivalves' early evolution as revealed by mitochondrial genes. *Mol. Phylogenet. Evol.* **57**, 641–657 (2010).
18. Giribet, G. Bivalvia. In *Phylogeny and Evolution of the Mollusca* (eds Ponder, W. F. & Lindberg, D. R.) 105–142 (University of California Press, 2008).
19. González, V. L. *et al.* A phylogenetic backbone for Bivalvia: An RNA-seq approach. *Proc. R. Soc. B Biol. Sci.* **282**, 20142332 (2015).
20. Lemer, S., Bieler, R. & Giribet, G. Resolving the relationships of clams and cockles: Dense transcriptome sampling drastically improves the bivalve tree of life. *Proc. R. Soc. B Biol. Sci.* **286**, 20182684 (2019).
21. Plazzi, F., Puccio, G. & Passamonti, M. Burrowers from the past: Mitochondrial signatures of ordovician bivalve infaunalization. *Genome Biol. Evol.* **9**, 956–967 (2017).
22. Newell, N. D. Classification of the Bivalvia. *Am. Mus. Novit.* **2206**, 1–25 (1965).
23. Combosch, D. J. *et al.* A family-level tree of life for bivalves based on a Sanger-sequencing approach. *Mol. Phylogenet. Evol.* **107**, 191–208 (2017).
24. Plazzi, F., Puccio, G. & Passamonti, M. Comparative large-scale mitogenomics evidences clade-specific evolutionary trends in mitochondrial DNAs of Bivalvia. *Genome Biol. Evol.* **8**, 2544–2564 (2016).
25. Myra, K. A. *Marine Molluscan Genera of Western North America: An Illustrated Key* (Stanford University Press, 1963).
26. Carter, J., Altaba, C., Anderson, L. & Yancei, T. A synoptical classification of the Bivalvia (mollusca). *Paleontol. Contrib.* <https://doi.org/10.17161/PC.1808.8287> (2011).
27. Morton, B. & Machado, F. M. Chapter one—Predatory marine bivalves: A review. *Adv. Mar. Biol.* **84**, 1–98 (2019).
28. Harper, E. M., Hide, E. A. & Morton, B. Relationships between the extant Anomalodesmata: A cladistic test. In *The Evolutionary Biology of the Bivalvia* (eds Haper, E. M. *et al.*) 129–143 (The Geological Society of London, 2000).
29. Dreyer, H., Steiner, G. & Harper, E. M. Molecular phylogeny of Anomalodesmata (Mollusca: Bivalvia) inferred from 18S rRNA sequences. *Zool. J. Linn. Soc.* **139**, 229–246 (2003).
30. Harper, E. M., Dreyer, H. & Steiner, G. Reconstructing the Anomalodesmata (Mollusca: Bivalvia): Morphology and molecules. *Zool. J. Linn. Soc.* **148**, 395–420 (2006).
31. Taylor, J. D., Williams, S. T., Glover, E. A. & Dyal, P. A molecular phylogeny of heterodont bivalves (Mollusca: Bivalvia: Heterodonta): New analyses of 18S and 28S rRNA genes. *Zool. Scr.* **36**, 587–606 (2007).
32. Taylor, J. D., Williams, S. T. & Glover, E. A. Evolutionary relationships of the bivalve family Thyasiridae (Mollusca: Bivalvia), monophyly and superfamily status. *J. Mar. Biol. Assoc. U.K.* **87**, 565–574 (2007).
33. Bieler, R. *et al.* Investigating the Bivalve tree of life—An exemplar-based approach combining molecular and novel morphological characters. *Invertebr. Syst.* **28**, 32 (2014).
34. Waller, T. R. The evolution of ligament systems in the Bivalvia. In *The Bivalvia* (ed. Morton, B.) 49–71 (Hong Kong University Press, 1990).
35. Morris, N. J. A new Lower Ordovician bivalve family, the Thoraliidae (? Nuculoida), interpreted as actinodont deposit feeders. *Bull. Br. Mus. Nat. Hist. (Geol.)* **34**, 265–272 (1980).
36. Cope, J. C. W. The early evolution of the Bivalvia. In *Origin and Evolutionary Radiation of the Mollusca* (ed. Taylor, J. D.) 361–370 (Oxford University Press, 1996).
37. Sánchez, T. M. & Babin, C. Distribution paléogéographique des mollusques bivalves durant l'Ordovicien. *Géodiversitas* **25**, 243–259 (2003).
38. Sánchez, T. M. Taxonomic position and phylogenetic relationships of the bivalve Goniophorina Isberg and related genera from the early Ordovician of northwestern Argentina. *Ameghiniana* **43**, 113–122 (2006).
39. Fang, Z. & Sanchez, T. M. Treatise Online no. 43: Part N, Revised, Volume 1, Chapter 16: Origin and early evolution of the Bivalvia. In *Treatise Online* (2012).
40. Cope, J. C. W. & Kříž, J. The Lower Palaeozoic palaeobiogeography of Bivalvia. In *Geological Society, London, Memoirs Vol. 38* (eds Harper, D. A. T. & Servais, T.) 221–241 (Geological Society of London, 2013).
41. Doucet-Beaupré, H. *et al.* Mitochondrial phylogenomics of the Bivalvia (Mollusca): searching for the origin and mitogenomic correlates of doubly uniparental inheritance of mtDNA. *BMC Evol. Biol.* **10**, 50 (2010).
42. Plazzi, F., Ceregato, A., Taviani, M. & Passamonti, M. A molecular phylogeny of bivalve mollusks: Ancient radiations and divergences as revealed by mitochondrial genes. *PLoS ONE* **6**, e27147 (2011).
43. Plazzi, F., Ribani, A. & Passamonti, M. The complete mitochondrial genome of *Solemya velum* (Mollusca: Bivalvia) and its relationships with Conchifera. *BMC Genom.* **14**, 409 (2013).
44. Kocot, K. M. *et al.* Phylogenomics reveals deep molluscan relationships. *Nature* **477**, 452–456 (2011).
45. Smith, S. A. *et al.* Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* **480**, 364–367 (2011).
46. Sharma, P. P. *et al.* Phylogenetic analysis of four nuclear protein-encoding genes largely corroborates the traditional classification of Bivalvia (Mollusca). *Mol. Phylogenet. Evol.* **65**, 64–74 (2012).
47. Toews, D. P. L. & Brelsford, A. The biogeography of mitochondrial and nuclear discordance in animals. *Mol. Ecol.* **21**, 3907–3930 (2012).
48. Sloan, D. B. *et al.* Cytonuclear integration and co-evolution. *Nat. Rev. Genet.* **19**, 635–648 (2018).
49. Lynch, M. Mutation accumulation in transfer RNAs: Molecular evidence for Muller's ratchet in mitochondrial genomes. *Mol. Biol. Evol.* **13**, 209–220 (1996).
50. Osada, N. & Akashi, H. Mitochondrial–nuclear interactions and accelerated compensatory evolution: Evidence from the primate cytochrome c oxidase complex. *Mol. Biol. Evol.* **29**, 337–346 (2012).
51. van der Sluis, E. O. *et al.* Parallel structural evolution of mitochondrial ribosomes and OXPHOS complexes. *Genome Biol. Evol.* **7**, 1235–1251 (2015).
52. Havird, J. C., Whitehill, N. S., Snow, C. D. & Sloan, D. B. Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. *Evolution (N Y)* **69**, 3069–3081 (2015).
53. Weng, M.-L., Ruhlman, T. A. & Jansen, R. K. Plastid–nuclear interaction and accelerated coevolution in plastid ribosomal genes in geraniaceae. *Genome Biol. Evol.* **8**, 1824–1838 (2016).
54. Rockenbach, K. *et al.* Positive selection in rapidly evolving plastid–nuclear enzyme complexes. *Genetics* **204**, 1507–1522 (2016).
55. Wolfe, N. W. & Clark, N. L. ERC analysis: Web-based inference of gene function via evolutionary rate covariation. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btv454> (2015).

56. Yan, Z., Ye, G. & Werren, J. H. Evolutionary rate correlation between mitochondrial-encoded and mitochondria-associated nuclear-encoded proteins in insects. *Mol. Biol. Evol.* **36**, 1022–1036 (2019).
57. Piccinini, G. *et al.* Mitonuclear coevolution, but not nuclear compensation, drives evolution of OXPHOS complexes in bivalves. *Mol. Biol. Evol.* **38**, 2597–2614 (2021).
58. Forsythe, E. S., Williams, A. M. & Sloan, D. B. Genome-wide signatures of plastid-nuclear coevolution point to repeated perturbations of plastid proteostasis systems across angiosperms. *Plant Cell* **33**, 980–997 (2021).
59. Nielsen, R. Molecular signatures of natural selection. *Annu. Rev. Genet.* **39**, 197–218 (2005).
60. Ghiselli, F. *et al.* Molluscan mitochondrial genomes break the rules. *Philos. Trans. R. Soc. B Biol. Sci.* **376**, 20200159 (2021).
61. Guerra, D. *et al.* Evolution of sex-dependent mtDNA transmission in freshwater mussels (Bivalvia: Unionida). *Sci. Rep.* **7**, 1551 (2017).
62. Ren, J., Liu, X., Jiang, F., Guo, X. & Liu, B. Unusual conservation of mitochondrial gene order in Crassostreaoysters: Evidence for recent speciation in Asia. *BMC Evol. Biol.* **10**, 394 (2010).
63. Saccone, C., de Giorgi, C., Gissi, C., Pesole, G. & Reyes, A. Evolutionary genomics in Metazoa: The mitochondrial DNA as a model system. *Gene* **238**, 195–209 (1999).
64. Sun, S., Li, Q., Kong, L. & Yu, H. Multiple reversals of strand asymmetry in molluscs mitochondrial genomes, and consequences for phylogenetic inferences. *Mol. Phylogenet. Evol.* **118**, 222–231 (2018).
65. Formaggioni, A., Luchetti, A. & Plazzi, F. Mitochondrial genomic landscape: A portrait of the mitochondrial genome 40 years after the first complete sequence. *Life* **11**, 663 (2021).
66. Hassanin, A., Léger, N. & Deutsch, J. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. *Syst. Biol.* **54**, 277–298 (2005).
67. Minh, B. Q., Hahn, M. W. & Lanfear, R. New methods to calculate concordance factors for phylogenomic datasets. *Mol. Biol. Evol.* **37**, 2727–2733 (2020).
68. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
69. Castoe, T. A. *et al.* Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc. Natl. Acad. Sci.* **106**, 8986–8991 (2009).
70. Shen, X.-X., Hittinger, C. T. & Rokas, A. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat. Ecol. Evol.* **1**, 0126 (2017).
71. Reyes, A., Gissi, C., Pesole, G. & Saccone, C. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Mol. Biol. Evol.* **15**, 957–966 (1998).
72. Yu, H. & Li, Q. Mutation and selection on the wobble nucleotide in tRNA anticodons in marine bivalve mitochondrial genomes. *PLoS ONE* **6**, e16147 (2011).
73. Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401 (1978).
74. Stöger, I. & Schrödl, M. Mitogenomics does not resolve deep molluscan relationships (yet?). *Mol. Phylogenet. Evol.* **69**, 376–392 (2013).
75. Avila-Herrera, A. & Pollard, K. S. Coevolutionary analyses require phylogenetically deep alignments and better null models to accurately detect inter-protein contacts within and between species. *BMC Bioinform.* **16**, 268 (2015).
76. Funk, D. J. & Omland, K. E. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Syst.* **34**, 397–423 (2003).
77. Sloan, D. B., Havird, J. C. & Sharbrough, J. The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Mol. Ecol.* **26**, 2212–2236 (2017).
78. Beck, E. A., Thompson, A. C., Sharbrough, J., Brud, E. & Llopart, A. Gene flow between *Drosophila yakuba* and *Drosophila santomea* in subunit V of cytochrome c oxidase: A potential case of cytonuclear cointegration. *Evolution (NY)* **69**, 1973–1986 (2015).
79. Morales, H. E. *et al.* Concordant divergence of mitogenomes and a mitonuclear gene cluster in bird lineages inhabiting different climates. *Nat. Ecol. Evol.* **2**, 1258–1267 (2018).
80. Sánchez, T. M. The early bivalve radiation in the Ordovician Gondwanan basins of Argentina. *Alcheringa Aust. J. Palaeontol.* **32**, 223–246 (2008).
81. Lawless, C., Greaves, L., Reeve, A. K., Turnbull, D. M. & Vincent, A. E. The rise and rise of mitochondrial DNA mutations. *Open Biol.* **10**, 200061 (2020).
82. Zheng, W., Khrapko, K., Collier, H. A., Thilly, W. G. & Copeland, W. C. Origins of human mitochondrial point mutations as DNA polymerase  $\gamma$ -mediated errors. *Mutat. Res. Fundam. Mol. Mech. Mutagen.* **599**, 11–20 (2006).
83. Pozzi, A., Plazzi, F., Milani, L., Ghiselli, F. & Passamonti, M. SmithRNAs: Could mitochondria “bend” nuclear regulation?. *Mol. Biol. Evol.* **34**, 1960–1973 (2017).
84. Gissi, C., Iannelli, F. & Pesole, G. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity (Edinb)* **101**, 301–320 (2008).
85. Malkócs, T. *et al.* Complex mitogenomic rearrangements within the Pectinidae (Mollusca: Bivalvia). *BMC Evol. Biol.* **22**, 29 (2022).
86. Breton, S., Beaupré, H. D., Stewart, D. T., Hoeh, W. R. & Blier, P. U. The unusual system of doubly uniparental inheritance of mtDNA: Isn't one enough?. *Trends Genet.* **23**, 465–474 (2007).
87. Passamonti, M. & Plazzi, F. Doubly uniparental inheritance and beyond: The contribution of the Manila clam *Ruditapes philippinarum*. *J. Zool. Syst. Evol. Res.* **58**, 529–540 (2020).
88. Zouros, E. & Rodakis, G. C. Doubly uniparental inheritance of mtDNA: An unappreciated defiance of a general rule. *Adv. Anat. Embryol. Cell Biol.* **231**, 25–49 (2019).
89. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
90. Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41**, e121–e121 (2013).
91. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
92. Kanehisa, M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
93. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
94. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).
95. Kanehisa, M., Sato, Y. & Kawashima, M. <sc>KEGG</sc> mapping tools for uncovering hidden features in biological data. *Protein Sci.* **31**, 47–53 (2022).
96. Lewis, S. H., Salmela, H. & Obbard, D. J. Duplication and diversification of Dipteran Argonaute genes, and the evolutionary divergence of Piwi and Aubergine. *Genome Biol. Evol.* **8**, 507–518 (2016).
97. Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509–524 (2014).
98. Bateman, A. *et al.* UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
99. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

100. Chang, J.-M., di Tommaso, P., Taly, J.-F. & Notredame, C. Accurate multiple sequence alignment of transmembrane proteins with PSI-Coffee. *BMC Bioinform.* **13**, S1 (2012).
101. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).
102. Kück, P. *et al.* Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Front. Zool.* **7**, 10 (2010).
103. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
104. Wu, M., Chatterji, S. & Eisen, J. A. Accounting for alignment uncertainty in phylogenomics. *PLoS ONE* **7**, e30288 (2012).
105. Dress, A. W. *et al.* Noisy: Identification of problematic columns in multiple sequence alignments. *Algorithms Mol. Biol.* **3**, 7 (2008).
106. Young, N. D. & Healy, J. GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinform.* **4**, 6 (2003).
107. Lanfear, R., Calcott, B., Ho, S. Y. W. & Guindon, S. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **29**, 1695–1701 (2012).
108. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
109. Stamatakis, A. Phylogenetic models of rate heterogeneity: A high performance computing perspective. In *Proceedings 20th IEEE International Parallel & Distributed Processing Symposium* 8 pp. (IEEE, 2006). <https://doi.org/10.1109/IPDPS.2006.1639535>.
110. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
111. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* <https://doi.org/10.1214/ss/1177011136> (1992).
112. Shimodaira, H. & Hasegawa, M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**, 1114–1116 (1999).
113. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2021).
114. Cock, P. J. A. *et al.* Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
115. Plazzi, F., Puccio, G. & Passamonti, M. HERMES: An improved method to test mitochondrial genome molecular synapomorphies among clades. *Mitochondrion* **58**, 285–295 (2021).
116. Kruskal, W. H. & Wallis, W. A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **47**, 583 (1952).
117. Dinno, A. *dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums* (2017).
118. Carter, J. G. *et al.* A synoptical classification of the Bivalvia (Mollusca). *Paleontol. Contrib.* <https://doi.org/10.17161/PC.1808.8287> (2011).
119. WoRMS Editorial Board. *World Register of Marine Species* (2022).

## Acknowledgements

We are grateful to Dr Marco Calderone for retrieving published sequences of proteins associated to sncRNA biogenesis in public databases. We are also deeply indebted to Dr Giovanni Piccinini and Dr Mariangela Iannello who prompted us to investigate deeper in bivalve phylogenetic incongruities: the present work would not have been possible without their insightful discussions. The original manuscript was also improved thanks to two anonymous reviewers. This research was funded by the “Canziani Bequest” fund.

## Author contributions

Conceptualization, F.P. and M.P.; methodology, F.P. and A.F.; investigation, A.F.; formal analysis, A.F.; supervision, F.P. and M.P.; writing—original draft, A.F.; writing—review and editing, F.P. and M.P. All authors read and reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-15076-y>.

**Correspondence** and requests for materials should be addressed to F.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022