

RESEARCH ARTICLE

Open Access



# Variability of mitochondrial ORFans hints at possible differences in the system of doubly uniparental inheritance of mitochondria among families of freshwater mussels (Bivalvia: Unionida)

Davide Guerra<sup>1\*</sup> , Manuel Lopes-Lima<sup>2,3</sup>, Elsa Froufe<sup>3</sup>, Han Ming Gan<sup>4</sup>, Paz Ondina<sup>5</sup>, Rafaela Amaro<sup>5</sup>, Michael W. Klunzinger<sup>6,7,8</sup>, Claudia Callil<sup>9</sup>, Vincent Prié<sup>10</sup>, Arthur E. Bogan<sup>11</sup>, Donald T. Stewart<sup>12</sup> and Sophie Breton<sup>1</sup>

## Abstract

**Background:** Supernumerary ORFan genes (i.e., open reading frames without obvious homology to other genes) are present in the mitochondrial genomes of gonochoric freshwater mussels (Bivalvia: Unionida) showing doubly uniparental inheritance (DUI) of mitochondria. DUI is a system in which distinct female-transmitted and male-transmitted mitotypes coexist in a single species. In families Unionidae and Margaritiferidae, the transition from dioecy to hermaphroditism and the loss of DUI appear to be linked, and this event seems to affect the integrity of the ORFan genes. These observations led to the hypothesis that the ORFans have a role in DUI and/or sex determination. Complete mitochondrial genome sequences are however scarce for most families of freshwater mussels, therefore hindering a clear localization of DUI in the various lineages and a comprehensive understanding of the influence of the ORFans on DUI and sexual systems. Therefore, we sequenced and characterized eleven new mitogenomes from poorly sampled freshwater mussel families to gather information on the evolution and variability of the ORFan genes and their protein products.

**Results:** We obtained ten complete plus one almost complete mitogenome sequence from ten representative species (gonochoric and hermaphroditic) of families Margaritiferidae, Hyriidae, Mulleriidae, and Iridinidae. ORFan genes are present only in DUI species from Margaritiferidae and Hyriidae, while non-DUI species from Hyriidae, Iridinidae, and Mulleriidae lack them completely, independently of their sexual system. Comparisons among the proteins translated from the newly characterized ORFans and already known ones provide evidence of conserved structures, as well as family-specific features.

**Conclusions:** The ORFan proteins show a comparable organization of secondary structures among different families of freshwater mussels, which supports a conserved physiological role, but also have distinctive family-specific features. Given this latter observation and the fact that the ORFans can be either highly mutated or completely absent in species that secondarily lost DUI depending on their respective family, we hypothesize that some aspects of the connection among ORFans, sexual systems, and DUI may differ in the various lineages of unionids.

**Keywords:** Freshwater mussels, Doubly uniparental inheritance of mitochondrial DNA, mtDNA sequencing, Mitochondrial ORFan genes, Evolution of protein structures and functions, Mitochondria and sexual systems

\* Correspondence: [guerra.davide86@yahoo.com](mailto:guerra.davide86@yahoo.com)

<sup>1</sup>Département de Sciences Biologiques, Université de Montréal, Montréal, QC, Canada

Full list of author information is available at the end of the article



## Background

Many species of gonochoric bivalve molluscs from four different orders (Unionida, Mytilida, Venerida, Nuculanida) possess a peculiar mode of mitochondrial transmission called doubly uniparental inheritance, or DUI, which is particularly well-documented (~ 80 species) in freshwater mussels of the order Unionida [1, 2]. DUI basically consists of a cyclic parent-specific inheritance of two distinct mitochondrial (mt) genomes (or mtDNA), where females segregate only the so called F (“female-transmitted”) mtDNA in their eggs and males segregate only the other mitotype, named M (“male-transmitted”), in their sperm. The zygote is heteroplasmic but, depending on its sexual development in the subsequent stages, i.e. whether it will become a female or a male as an adult, an individual will transmit only one of these two types of mt genomes to the next generation [1]. DUI seems to be strictly associated with the gonochoristic sexual system of a particular species, as it was discovered that four species of unionids and one species of margaritiferid each appear to have lost independently their M mtDNA during the transition from dioecy to hermaphroditism [3]. Notice that, however, it is not known if the loss of the M type is perfectly contemporaneous with the switch to hermaphroditism. All these mentioned obligate hermaphroditic species now retain only a modified version of the F genome that is called the H genome (for “hermaphrodite”) [3]. In an attempt to understand if and how DUI and sex determination are connected, genomic studies have highlighted the following features of DUI in freshwater mussel mtDNAs: (1) a high level of sequence divergence between F and M (up to ~ 40% of difference in their nucleotide sequences, resulting into an almost 50% difference in the amino acid sequence of their encoded proteins [4, 5]); (2) the presence of a 3'-elongation of the *cox2* gene in the M mtDNA (compared to that found in most metazoans) [6]; (3) the presence in both mt genomes of “ORFan” genes, i.e., genes without obvious homology or function [7], named *F-orf* in the F mtDNA and *M-orf* in the M mtDNA, respectively [8]. An ORFan is found also in the H mtDNAs mentioned above: it is a highly mutated *F-orf* and is named *H-orf* for distinction [3]. It is worth noting that these features of DUI-positive freshwater mussel mtDNAs also appear in practically all DUI bivalves outside Unionida, although with many variations and combinations (e.g., F vs M divergence can be lower; elongated *cox2* genes can appear in the F instead of the M, or be absent from both; and mtDNA-specific ORFans can be present in only one of the two mtDNAs or duplicated in a same mt genome [1, 5, 9–12]). Additional coding sequences are sometimes found in the mtDNA of bivalves (with or without DUI) and even other molluscan species [9, 13–22]. Usually they are identifiable as duplicated standard mitochondrial protein coding genes, with a variable degree of similarity to the original sequence. In fact, the two copies can evolve in tandem but, in other

cases, the only recognizable parts consist of small segments encoding functional domains of the original mitochondrial genes. In absence of further functional studies, this leads us to hypothesize that the primary function of these genes, if indeed they retain functionality, might be related to oxidative phosphorylation.

Functional studies focusing on the ORFan genes *F-orf* and *M-orf* in DUI bivalves provided evidence they are transcribed and translated into proteins (from here on respectively indicated as *F-ORF* and *M-ORF*) located inside and outside the mitochondria [3, 10, 23–26]. Moreover, extensive bioinformatic studies on both the gene sequences and their translated proteins produced evidence for two options for the origin of the ORFans, which may be the result of either (1) the insertion of viral sequences into the host mt genome [23, 27] or (2) the duplication and subsequent modification of extant mitochondrial genes and sequences [5, 28]. Analyses aimed at understanding their origin by predicting the functions of their protein products broadly converged on similar patterns in all species considered (freshwater mussels and others). *F-ORFs* and *M-ORFs*, for example, were predicted to interact with nucleic acids and/or membranes (for signaling and/or interactions with the immune system), and *M-ORFs*, in particular, were also predicted to interact with the cytoskeleton and to have a role in the ubiquitination processes [27, 28]. The high variability of the ORFans among distantly related families and orders, however, questions their homology in all DUI bivalves. To have a better understanding of ORFans evolution and the functions of their encoded proteins, efforts should focus on a taxon in which DUI is widespread, such as the Unionida [2]. Until a few years ago, only F, M, and H mt genomes from the family Unionidae and a few F and H genomes of Margaritiferidae were available, but recently mtDNAs from families Hyriidae, Iridinidae, Mulleriidae, and the first margaritiferid M mtDNAs have been published [5, 29]. Given the hypothesized, although still untested, link between ORFans and sexual systems in freshwater mussels [3], the sequencing of these new mt genomes allowed examination of the evolution of ORFan genes, DUI, and sexual systems in a phylogenetic context [5]. It was suggested that DUI may have been present as an ancestral state before the radiation of the order Unionida, and that some ORFans have been partially or totally purged from the remaining mtDNAs of some lineages that may have lost DUI in their early stages of radiation (Iridinidae and Mulleriidae). However, the ORFans have been maintained in the other families that regularly show DUI (Hyriidae, Margaritiferidae, Unionidae), and in each of these taxa, the mt genomes, especially the M, show their own family-specific peculiarities. For example, in margaritiferid M mtDNAs the *M-orf* is duplicated (one copy,

*M-orf1*, appears to be homologous to the *M-orf* of Unionidae and Hyriidae, while the second copy, *M-orf2*, is specific to Margaritiferidae only), whereas the hyriid *M-orf* is much longer than those of Margaritiferidae and Unionidae [5].

In this study, we present eleven new mt genomes from ten species of freshwater mussels, with or without DUI and with different sexual systems (for each species, references describing DUI status and/or reproductive modes on which we relied for this study are given): *Chambardia rubens* (Lamarck, 1819) [30, 31] for Iridinidae; *Anodontites elongata* (Swainson, 1823) [32], *Fossula fossiculifera* (d'Orbigny, 1835) [33], *Lamproscapha ensiformis* (Spix and Wagner, 1827) (C. Callil personal observation), *Monocondylaea parchappii* (d'Orbigny, 1835) (C. Callil personal observation) for Mulleriidae; *Castalia ambigua* Lamarck, 1819 [34, 35], *Diplodon suavidicus* (Lea, 1856) [35], *Prisodon obliquus* Schumacher, 1817 [35], *Westralunio carteri* (Iredale, 1934) [36] for Hyriidae; and *Pseudunio auricularius* (Spengler, 1793) [37] for Margaritiferidae. First, we characterize the overall structure of the new mt genomes and highlight their unique features, some of which are described for the first time, and we build a phylogeny of freshwater mussels using these and other genomes. Then we identify new *F-orf* and *M-orf* genes and describe the new *F-ORF* and *M-ORF* proteins by comparing them to a set of already published sequences, showing how, despite having evolved different three-dimensional configurations, they share some key features. Finally, considering our findings, we discuss whether the DUI system works the same way in all DUI freshwater mussels or if there may be family-specific differences, as well as the modifications occurring in the mtDNAs after DUI is lost.

## Results

### Sequencing, assembly, and general features of the new mt genomes

We obtained a complete sequence for ten of the eleven new mt genomes; the M mtDNA of *W. carteri* showed a sequencing gap in a non-coding region between *trnV* and *trnH*. All sequences were deposited in GenBank under the accession numbers MK761136–46 (Table 1). A summary of their length and a comparison of their gene order are given in Table 2. In summary, they present the same general features recently highlighted by [5] (for families Iridinidae, Mulleriidae, Hyriidae, Margaritiferidae) and [29] (for Margaritiferidae). As is typical for DUI freshwater mussels, the *cox2* gene carried by *W. carteri* M mt genome is longer compared to its F counterpart (respectively 1329 bp and 693 bp) and to those of non-DUI species [6] (Table 3). Apart from the occasional species-specific difference in length of some non-coding regions, particularly in Hyriidae (i.e., between *atp8* and *nad4L* in *D. suavidicus*, and between *trnV* and *trnH* in *W. carteri* M mtDNA) and Iridinidae (1049 bp between *nad5* and *trnF* in *C. rubens*, compared to the 23–76 bp of the other species mtDNAs), the most notable features lie in the presence/absence of ORFans, on which we will focus.

### Phylogeny of freshwater mussel mt genomes

A Bayesian inference analysis was performed with MrBayes [38] on 12 protein coding genes and their respective protein sequences extracted from the new 11 mt genomes, from 51 additional mtDNAs of freshwater mussels available in GenBank (25 F, 22 M, and other 15 mtDNAs from non-DUI species; Additional file 1: Table S1), and from three outgroup

**Table 1** Summary of the newly sequenced species and their mt genomes

| Family           | Species                         | Country    | Latitude  | Longitude  | Sexual system  | GenBank accession numbers |          |          |
|------------------|---------------------------------|------------|-----------|------------|----------------|---------------------------|----------|----------|
|                  |                                 |            |           |            |                | non-DUI                   | F        | M        |
| Iridinidae       | <i>Chambardia rubens</i>        | Mauritania | 16.33880  | −11.97809  | dioecious      | MK761138                  |          |          |
| Mulleriidae      | <i>Anodontites elongata</i>     | Brasil     | −14.67936 | −56.23611  | dioecious      | MK761136                  |          |          |
|                  | <i>Fossula fossiculifera</i>    | Brasil     | −14.67936 | −56.23611  | dioecious      | MK761140                  |          |          |
|                  | <i>Lamproscapha ensiformis</i>  | Brasil     | −16.00573 | −55.90867  | unknown        | MK761141                  |          |          |
|                  | <i>Monocondylaea parchappii</i> | Brasil     | −14.67936 | −56.23611  | hermaphroditic | MK761142                  |          |          |
| Hyriidae         | <i>Castalia ambigua</i>         | Brasil     | −14.67936 | −56.23611  | dioecious      | MK761137                  |          |          |
|                  | <i>Diplodon suavidicus</i>      | Brasil     | −6.018083 | −60.194306 | unknown        | MK761139                  |          |          |
|                  | <i>Prisodon obliquus</i>        | Brasil     | −6.018083 | −60.194306 | unknown        | MK761143                  |          |          |
|                  | <i>Westralunio carteri</i>      | Australia  | −33.30210 | 115.81770  | dioecious      |                           | MK761145 | MK761146 |
| Margaritiferidae | <i>Pseudunio auricularius</i>   | France     | 47.01710  | 0.56310    | dioecious      |                           | MK761144 |          |

For each species for which at least one mt genome sequence was obtained in this study (either complete or largely complete), the respective family, and the provenance of the specimen(s) used for the sequencing (columns 'Country', 'Latitude', 'Longitude') are indicated, as well as the respective sexual system. For each mtDNA sequenced, its type is specified: non-DUI, mtDNA of dioecious or hermaphroditic species with no evidence of DUI presence or secondary loss; F and M, female- and male-transmitted mtDNAs of dioecious DUI species, respectively. The only available mtDNA of *Pseudunio auricularius*, a dioecious species for which no evidence of DUI has been produced yet (no M mtDNA sequence available yet), which carries an *F-orf*, has been considered as F for simplicity



**Table 2** Gene order comparison for the newly sequenced mt genomes (Continued)

| <i>Chambardia rubens</i> | <i>Anodontites elongatus</i> | <i>Fossula fossiculifera</i> | <i>Lamproscapha ensiformis</i> | <i>Monocoondylaea parchappii</i> | <i>Pseudunio auricularius</i> | <i>Castalia ambigua</i> | <i>Diplodon suavidicus</i> | <i>Prisodon obliquus</i> | <i>Westralunio carteri</i> | <i>Westralunio carteri</i> |
|--------------------------|------------------------------|------------------------------|--------------------------------|----------------------------------|-------------------------------|-------------------------|----------------------------|--------------------------|----------------------------|----------------------------|
| 16,697 bp                | 15,255 bp                    | 15,632 bp                    | 15,583 bp                      | 15,528 bp                        | 16,195 bp                     | 17,092 bp               | 18,211 bp                  | 16,691 bp                | 16,766 bp                  | 18,038 bp                  |
| trnW                     | trnW                         | trnW                         | trnW                           | trnW                             | trnW                          | trnW                    | trnW                       | trnW                     | trnW                       | trnW                       |
| trnM                     | trnM                         | trnM                         | trnM                           | trnM                             | trnM                          | trnM                    | trnM                       | trnM                     | trnM                       | trnM                       |
| nad2                     | nad2                         | nad2                         | nad2                           | nad2                             | nad2                          | nad2                    | nad2                       | nad2                     | nad2                       | nad2                       |
| trnE                     | trnE                         | trnE                         | trnE                           | trnE                             | F-orf                         | trnE                    | trnE                       | trnE                     | trnE                       | trnE                       |
| trnS1                    | trnS1                        | trnS1                        | trnS1                          | trnS1                            | trnS1                         | trnS1                   | trnS1                      | trnS1                    | trnS1                      | trnS1                      |
| trnS2                    | trnS2                        | trnS2                        | trnS2                          | trnS2                            | trnS2                         | trnS2                   | trnS2                      | trnS2                    | trnS2                      | trnS2                      |
| trnA                     | trnA                         | trnA                         | trnA                           | trnA                             | trnA                          | trnA                    | trnA                       | trnA                     | trnA                       | trnA                       |
| trnH                     | <b>trnH</b>                  | <b>trnH</b>                  | <b>trnH</b>                    | <b>trnH</b>                      | <b>trnH</b>                   | <b>trnH</b>             | <b>trnH</b>                | <b>trnH</b>              | <b>trnH</b>                | trnC                       |
| nad3                     | <b>nad3</b>                  | <b>nad3</b>                  | <b>nad3</b>                    | <b>nad3</b>                      | <b>nad3</b>                   | <b>nad3</b>             | <b>nad3</b>                | <b>nad3</b>              | <b>nad3</b>                | <b>nad3</b>                |
| cox2                     | <b>cox2</b>                  | <b>cox2</b>                  | <b>cox2</b>                    | <b>cox2</b>                      | <b>cox2</b>                   | <b>cox2</b>             | <b>cox2</b>                | <b>cox2</b>              | <b>cox2</b>                | <b>cox2</b>                |

Under every species name is indicated the length of the mtDNA. Genes in bold are encoded in forward direction, while genes in normal font in reverse direction. Gaps between genes in the table are present only as a consequence of the alignment of identical gene blocks. Positions with the same gene in every mtDNA are indicated with an asterisk (\*) at the right side of the table

**Table 3** Cox2 gene length variability in freshwater mussels

| Family           | mtDNA          |              | Number of <i>cox2</i> analyzed | Length (bp) |         |        |       |
|------------------|----------------|--------------|--------------------------------|-------------|---------|--------|-------|
|                  | transmitted by | total number |                                | minimum     | maximum | mean   | SD    |
| Iridinidae       | eggs           | 2            | 2                              | 681         | 681     | 681.0  | 0.0   |
| Mulleriidae      | eggs           | 5            | 4                              | 681         | 681     | 681.0  | 0.0   |
| Hyriidae         | eggs           | 5            | 5                              | 681         | 693     | 683.4  | 4.8   |
|                  | sperm          | 2            | 2                              | 1329        | 1380    | 1354.5 | 25.5  |
| Margaritiferidae | eggs           | 7            | 7                              | 600         | 693     | 672.4  | 29.9  |
|                  | sperm          | 3            | 3                              | 1215        | 1224    | 1218.0 | 4.2   |
| Unionidae        | eggs           | 21           | 21                             | 681         | 684     | 681.1  | 0.6   |
|                  | sperm          | 17           | 17                             | 825         | 1272    | 1193.2 | 130.8 |

Range, mean, and standard deviation (SD) of the *cox2* gene length in egg- and sperm-transmitted mtDNAs in different freshwater mussel families. Sperm-transmitted mtDNAs are present only in families showing DUI. Uniparental transmission of mtDNA through female gametes is assumed in Iridinidae and Mulleriidae, which do not show DUI. In Mulleriidae, the partially sequenced *cox2* gene from *Anodontites trapesimalis* was excluded from this analysis; however, it does not show the typical 3' elongation of M *cox2* genes [5]

species (the bivalves *Neotrigonia margaritacea* and *Solemya velum*, plus a member of the Caudofoveata, *Chaetoderma nitidulum*). The evolutionary models calculated for the aligned and trimmed gene sequences were 'GTR + G' for *nad4L* and 'GTR + I + G' for all others. The most supported model for the trimmed protein alignment was 'Jones' (posterior probability = 1.000). The nucleotide- and amino acid-based phylogenetic reconstructions reached convergence (standard deviation of split frequencies stabilized at values < 0.01) respectively after 39,000 and 17,000 generations.

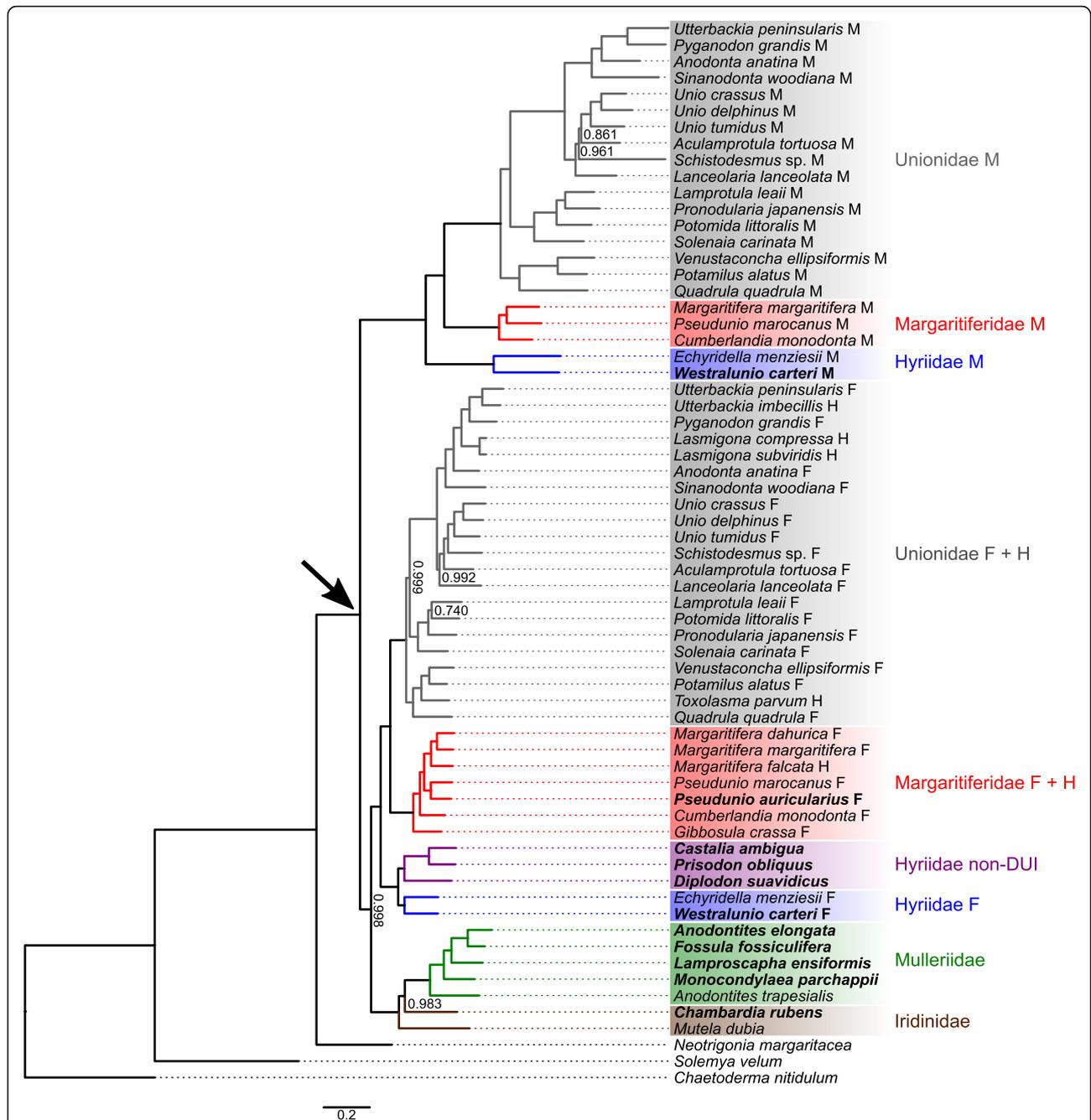
In the nucleotide-based tree (Fig. 1), freshwater mussel mtDNAs form a monophyletic group divided in two main branches: one containing all M mtDNAs and another containing all female-transmitted ones, from both DUI and non-DUI taxa. Relationships among DUI species in these two main branches are maintained in most cases, i.e. the phylogeny of F mt genomes mirrors that of M ones. In the few cases where this situation does not occur, either the nodes usually have posterior probabilities < 1.000 (e.g., see the different relative position of *Aculamprotula tortuosa* mt genomes in the Unionidae clades) or the number of F and M genomes for a taxon is different (e.g., Margaritiferidae). In both of the two main branches, family Hyriidae is sister group to Margaritiferidae and Unionidae, which always form reciprocally sister groups. For Hyriidae clades, *W. carteri* mt genomes are always sister to the respective *Echydrella menziesii* ones. In Margaritiferidae, *P. auricularius* female-transmitted mtDNA is sister to *Pseudunio maroccanus* F mtDNA. Iridinidae and Mulleriidae form a single branch sister to Hyriidae + Margaritiferidae + Unionidae female-transmitted mtDNAs, where Mulleriidae form a monophyletic clade but Iridinidae do not: *C. rubens* is sister to all Mulleriidae (node posterior probability = 0.983), while the other iridiniid *Mutela dubia* is sister to all. Inside Mulleriidae, the dioecious [32] *A. elongata* is

recovered as distantly related to the congeneric hermaphrodite [39–41] *Anodontites trapesimalis*.

The main difference of the protein-based tree (Fig. 2) with the nucleotide-based one is the position of the M mtDNAs clade: here, it is sister to a clade containing *N. margaritacea* mtDNA as sister to all female-transmitted mt genomes of freshwater mussels. Minor differences with the nucleotide tree are as follows. In the female-transmitted mtDNAs clade: for Hyriidae, *W. carteri* F mtDNA is sister to all other hyriid female-transmitted mt genomes; for Unionidae, *Lanceolaria lanceolata* F mtDNA has a different position; in the Mulleriidae + Iridinidae clade, the positions of *C. rubens* and *M. dubia* are switched, with *M. dubia* here sister to all Mulleriidae (node posterior probability = 0.739). The M mt genomes clade differs from the nucleotide tree in the following instances: Margaritiferidae M mtDNAs have different relationships; in Unionidae, *Aculamprotula tortuosa* M mtDNA and *Schistodesmus* sp. [42] M mtDNA exchange positions, and *Sinanodonta woodiana* M mtDNA and *Anodonta anatina* M mtDNA become sister groups.

### Search and annotation of ORFan genes

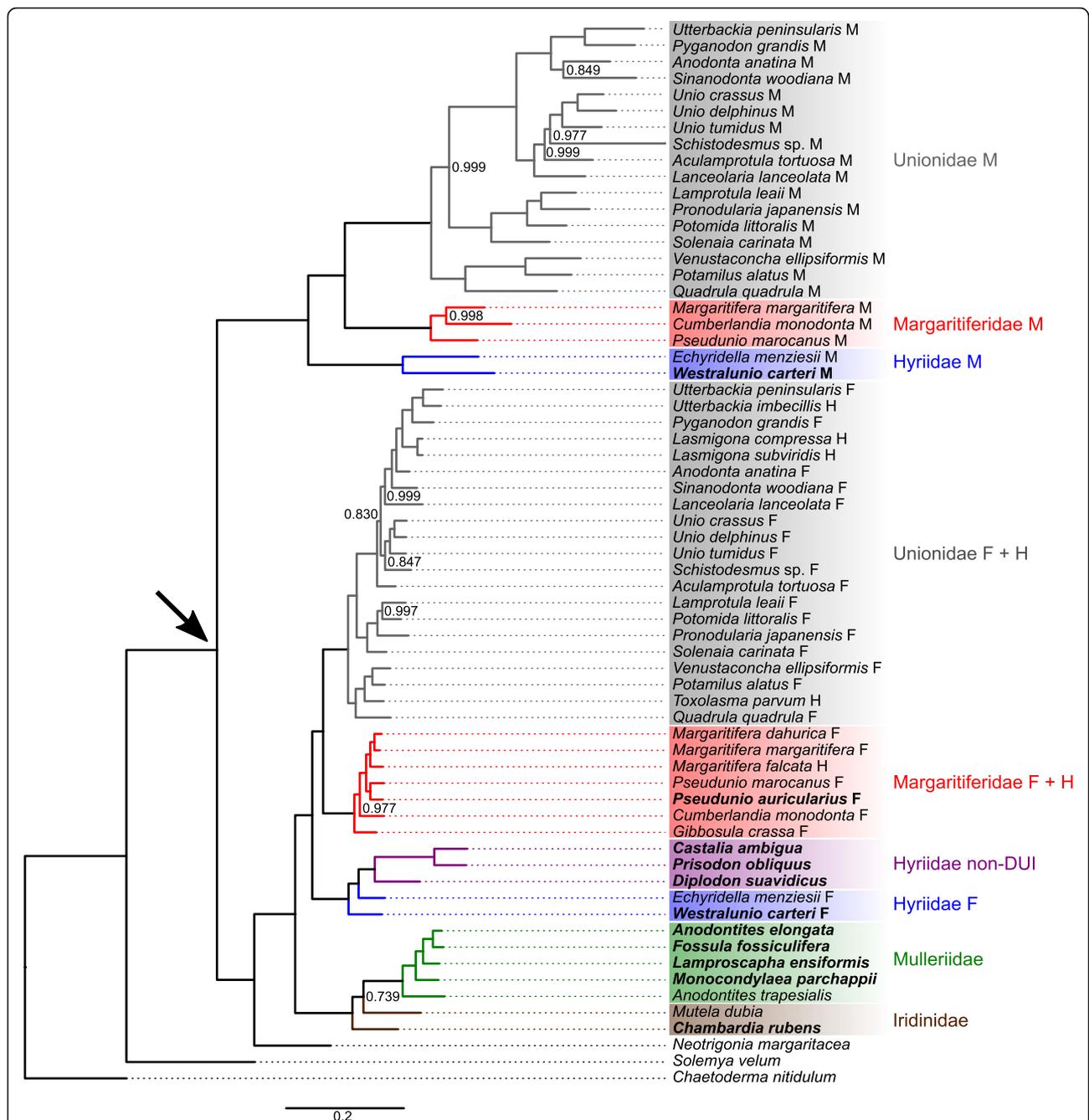
To search for new ORFan genes, a set of 35,032 open reading frames (ORFs) (Table 4) was extracted from the new mt genomes in Table 1 and the additional 51 mtDNAs of freshwater mussels from GenBank (Additional file 1: Table S1), for a total of 62 mt genomes analyzed. Using as a criterion of choice the level of similarity between the known ORFan proteins and the translated proteins of the nucleotide sequences in the ORFs set, we found with the HMMER [43] suite of programs 25 F-*orfs*, 5 H-*orfs*, and 26 M-*orfs*, three of which are M-*orf2* from Margaritiferidae species [5] and one appears to be a recent duplication specific to the unionid *S. woodiana* (we named the two copies M-*orfa* and M-*orfb*; see also [12]) (Additional file 1: Tables S2 and S3).



**Fig. 1** Bayesian inference phylogenetic tree of freshwater mussel mt genomes based on nucleotide sequences from 12 of their protein coding genes (*atp8* was excluded). All nodes have posterior probability 1.000, except where indicated. An arrow indicates the split of freshwater mussel M mtDNAs clade. Clades and groups of mtDNAs are color coded according to the family and/or type of mtDNA indicated on the right side of the figure. The new mtDNAs sequenced in this study are in bold character

The nucleotide sequences of these 56 ORFans are listed in Additional file 2. In summary, single ORFan protein sequences used as seeds mainly recognized homologous ORFan sequences, with few non-ORFan hits (as in the case for *E. menziesii* M-ORF which recognized some full, in-frame *nad4L* sequence) (Additional file 1: Table S2). The F-ORF Hidden Markov Model (HMM) profile we

used recognized, in addition to all sequences from which it was assembled, also the H-ORFs (Additional file 1: Table S3). The M-ORF HMM profiles recognized all the M-ORFs forming them, plus, with lower scores and E-values, some protein translated from ORFs overlapping *trnD*, *atp8* (in two cases the hit comprised the whole in-frame sequence of its protein), *nad6*, and *nad4L* genes



**Fig. 2** Bayesian inference phylogenetic tree of freshwater mussel mt genomes based on protein sequences translated from 12 of their protein coding genes (*atp8* was excluded). All nodes have posterior probability 1.000, except where indicated. An arrow indicates the split of freshwater mussel M mtDNAs clade. Clades and groups of mtDNAs are color coded according to the family and/or type of mtDNA indicated on the right side of the figure. The new mtDNAs sequenced in this study are in bold character

of many mt genomes (Additional file 1: Table S3). The putative proteins from the *AtraUR219* and *AtraUR2218* ORFans of *A. trapesialis*, located between *atp8* and *nad4L* of this species mt genome (Fig. 3) (which have been proposed to have originated from duplication and divergence of *atp8* and might be related to M-orfs of DUI species [5]), have no apparent homologs in other

species. However, some ORFs from other species were retrieved that overlap the same region where the two *A. trapesialis* ORFans are located and from which are hypothesized to have originated: among these hits, for example, four include either a segment of the *atp8* gene or its full in-frame sequence (Additional file 1: Table S4). In short, as shown also in Fig. 3, we found that: (1) all

**Table 4** ORFs dataset description

| Family           | Number of species | mtDNA type | Number of mtDNAs | Number of ORFs | F- <i>orf</i> | M- <i>orf</i> | H- <i>orf</i> |
|------------------|-------------------|------------|------------------|----------------|---------------|---------------|---------------|
| Iridinidae       | 2                 | non-DUI    | 2                | 1206           | 0             | 0             | 0             |
| Mulleriidae      | 5                 | non-DUI    | 5                | 2820           | 0             | 0             | 0             |
| Hyriidae         | 5                 | F          | 2                | 1063           | 2             | 0             | 0             |
|                  |                   | M          | 2                | 1099           | 0             | 2             | 0             |
|                  |                   | non-DUI    | 3                | 1668           | 0             | 0             | 0             |
|                  |                   | all        | 7                | 3830           | 2             | 2             | 0             |
| Margaritiferidae | 7                 | F          | 6                | 3273           | 6             | 0             | 0             |
|                  |                   | M          | 3                | 1657           | 0             | 6             | 0             |
|                  |                   | H          | 1                | 556            | 0             | 0             | 1             |
|                  |                   | all        | 10               | 5486           | 6             | 6             | 1             |
| Unionidae        | 21                | F          | 17               | 9698           | 17            | 0             | 0             |
|                  |                   | M          | 17               | 9826           | 0             | 18            | 0             |
|                  |                   | H          | 4                | 2166           | 0             | 0             | 4             |
|                  |                   | all        | 38               | 21,690         | 17            | 18            | 4             |
|                  |                   | all F      | 25               | 12,367         | 25            | 0             | 0             |
|                  |                   | all M      | 22               | 12,582         | 0             | 26            | 0             |
|                  |                   | all H      | 5                | 4389           | 0             | 0             | 5             |
| Total            | 40                |            | 62               | 35,032         | 25            | 26            | 5             |

For each taxonomic group of freshwater mussels, the number of species and mtDNAs considered in this study are given, as well as the number and kind of ORFans (F-, M-, or H-*orfs*) retrieved from them. mtDNA types: F and M, female- and male-transmitted mtDNAs of dioecious DUI species, respectively; H, mtDNA of stably hermaphroditic species that lost DUI sensu [3]; non-DUI, mtDNA of dioecious and hermaphroditic non-DUI species. mtDNAs of dioecious species carrying an F-*orf* but for which no evidence of DUI has been produced yet, or of dioecious DUI species that can show occasional hermaphroditism, were considered as F for simplicity

DUI species of freshwater mussels analyzed carry a F-*orf* in their F mtDNA and at least one M-*orf* in their M; (2) secondarily hermaphrodite (i.e., that switched from gonochorism to hermaphroditism) species of Unionidae and Margaritiferidae that lost DUI always possess a H-*orf* [3]; (3) species that do not show evidence of DUI (i.e., no evidence of heteroplasmy) from families Iridinidae, Mulleriidae, and Hyriidae have none of these ORFans. The only mtDNA we retrieved from *P. auricularius* presents a standard F-*orf*, and given that this species is dioecious, it is plausible it will be revealed as a DUI species, and therefore we treated this genome as F mtDNA.

#### Sequence-based analyses of ORFan protein products

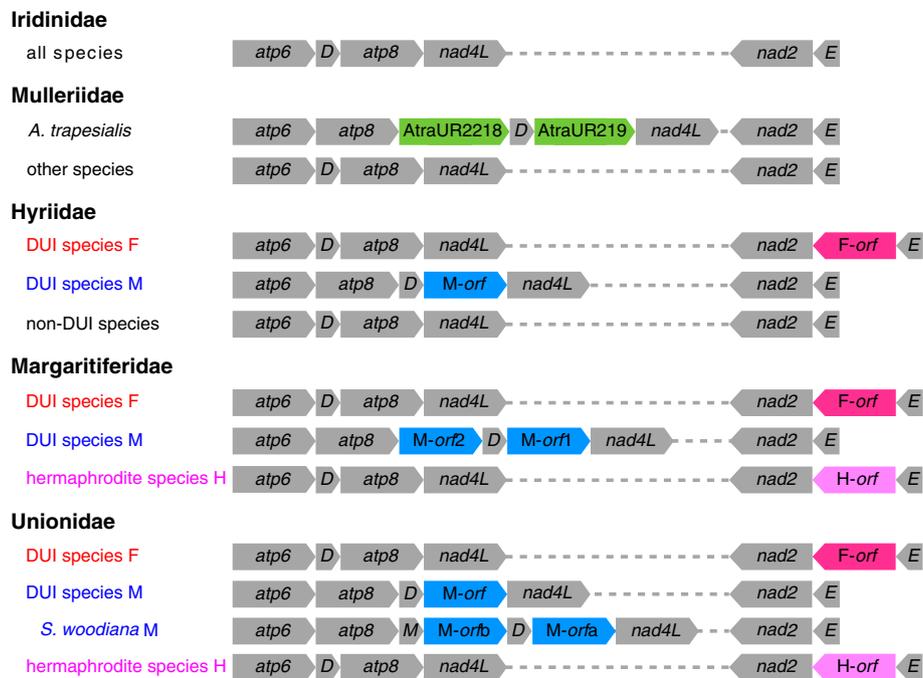
The amino acid composition of F-ORFs appears to be rather homogeneous among families, with no clear differences and, although more variable, a general trend is observable also in the M-ORFs (Fig. 4). For Margaritiferidae, in some cases, the distribution of some amino acid percentage of M-ORF2 proteins (which do not have a homolog in Hyriidae or Unionidae; Fig. 3), differ distinctly from the M-ORF1 and fall outside the range of other M-ORFs. The patterns for AtraUR219 is distinct from those of all M-ORFs but only in terms of sheer percentage of

amino acid usage, as the peaks and declines of its pattern are located in the same position as the M-ORFs. On the contrary, AtraUR2218 profile is quite different and does not follow that of the other proteins; however, this may be an effect of its extremely short length (22 aa).

Both the CLANS [44] analyses (Fig. 5) and the maximum likelihood (ML) trees (Fig. 6) tend to separate F-ORF and M-ORF protein sequences in family-specific clusters: notably, the relationships among single ORFans in the ML trees broadly resemble those among their respective mt genomes in our phylogenies based on 12 protein-coding genes (Figs. 1 and 2), especially in the case of F-ORFs. For Margaritiferidae M-ORFs, it is notable to see how M-ORF2 sequences cluster together with the M-ORF1 sequences in the CLANS analysis (Fig. 5), but form a separate branch in the ML tree (which has, however, low resolution) (Fig. 6). We also attempted to add AtraUR219 and AtraUR2218 sequences to the M-ORF alignment for the ML reconstruction, but this disrupted the clustering of M-ORFs, especially for the more numerous Unionidae (not shown).

#### Tertiary structure prediction of ORFan proteins

Currently, there are no data derived from crystallographic studies on the ORFan proteins, nor established



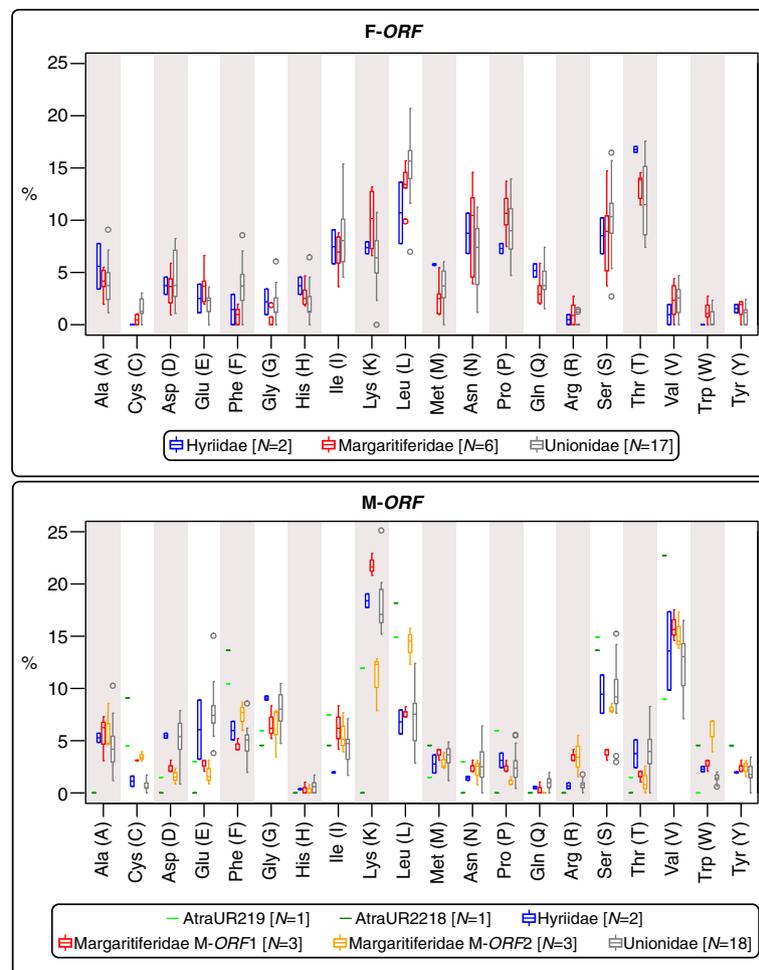
**Fig. 3** Schematic organization of the *atp6-nad4L* and *nad2-trnE* regions of the freshwater mussel (Unionida) based on mt genomes presented in this study and of already published ones. Sample size for each family: 2 Iridinidae (all non-DUI), 5 Mulleriidae (all non-DUI), 7 Hyriidae (2 F, 2 M, 3 non-DUI), 10 Margaritiferidae (6 F, 3 M, 1 H), 38 Unionidae (17 F, 17 M, 4 H). GenBank accession numbers of the mt genomes used are enlisted in Table 1 and Additional file 1: Table S1. Standard mitochondrial genes are in grey, while ORFan genes (see the main text for a complete description of these genes) are colored following this code: green, *Anodonta trapesialis* specific ORFans; blue, M-orfs; pink, F-orfs; light pink, H-orfs. Genes are pointed according to their relative direction on the mtDNAs. tRNA genes are indicated with the one-letter code of their respective amino acid. Dotted lines represent the segments between the two regions, which are not indicated for simplicity. *Sinanodonta woodiana* annotation is based on [12] and on the current study

structural similarities with known proteins in databases, that may guide bioinformatic analyses aimed at predicting the ORFan proteins folding. Therefore, for completeness, we decided to show the best results we obtained with I-Tasser [45] regardless of the C-scores assigned to the models: C-scores are usually comprised between -5 and 2, therefore higher values indicate higher confidence of the model. The predicted tertiary structure of some selected ORFan proteins may appear to be highly different at first sight, but similarities among proteins of the same kind can be recognized (Figs. 7 and 8). A common feature between the F-ORFs of Hyriidae and Unionidae is the presence of two antiparallel helices separated by a loop. This conformation is not found in the F-ORF of Margaritiferidae, in which only one small helix is predicted (preceded by a small beta strand in *Cumberlandia monodonta* and *Pseudunio marocanus* but not in *Margaritifera margaritifera*). Hyriidae M-ORF, Margaritiferidae M-ORF1 and M-ORF2, and *S. woodiana* M-ORFa all share the presence of three antiparallel helices in their N-terminus. The three helices have the same relative orientation, but the third one is in front of the first two in Hyriidae and behind them in Margaritiferidae and *S. woodiana* (and it is also much smaller in this species).

The portion beyond this third helix varies for each species, but, for example, Hyriidae M-ORFs are similar in this part of the protein and are clearly discernible from those of Margaritiferidae, and M-ORF1s and M-ORF2s of this family are again distinguishable between them. *Cumberlandia monodonta* M-ORF1 structure is less defined compared to the homologous proteins from *M. margaritifera* and *P. marocanus*, and in its M-ORF2, the third helix appears to be on the same plane as the other two. The three Unionidae M-ORFs examined have extremely divergent configurations, and no obvious similarities can be recognized among them. The two *S. woodiana* M-ORFs, most probably the product of a duplication event specific to this species [12], do not resemble one another. AtraUR219 protein is constituted by a short N-terminal beta strand, two helices crossing each other and connected by a simple loop, and a small C-terminal beta strand. AtraUR2218 protein is very short (22 aa) and it is predicted to be only a single helix.

**Three-dimensional alignments of ORFan proteins**

Summarized statistics for pairwise three-dimensional (3D) alignments of the ORFan protein models are shown in Table 5. In the pairwise interspecies 3D alignment for

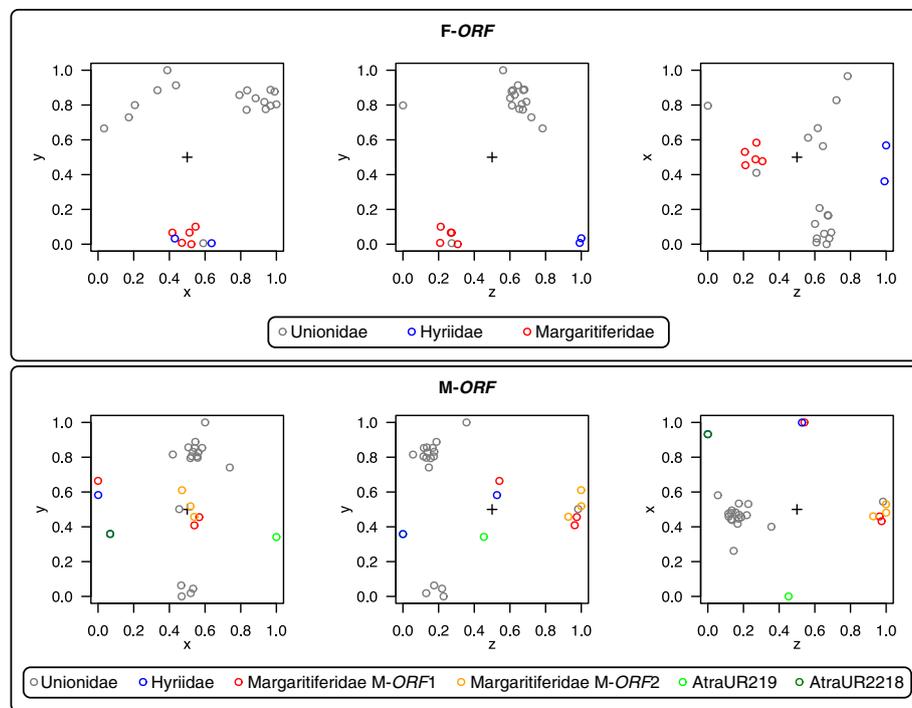


**Fig. 4** Percentage amino acid composition of the ORFan proteins considered in this study (their relative nucleotide sequences are enlisted in Additional file 2). The sample size for each boxplot is indicated inside the legends in square parentheses as 'N'. Amino acid names are indicated with the IUPAC three-letter and one-letter codes

each single freshwater mussel family, the distances (expressed as root mean square deviation of distances) between C $\alpha$  and between C $\beta$  atoms (respectively the  $\alpha$ - and  $\beta$ -carbon atoms of an amino acid) tend to be lower in the M-ORF alignments (for Unionidae, only the *Venustaconcha ellipsiformis* M-ORF vs *S. woodiana* M-ORFa comparison) compared to the F-ORF ones, while percentages of aligned amino acids and secondary structures vary depending on the family and the ORFan proteins considered. Margaritiferidae M-ORF1 and -2 pairwise alignments, both in between-species and in single-species comparisons, on average obtain relatively good (and in some cases better) values compared to the separate M-ORF1 and M-ORF2 alignments. The protein encoded by the recently duplicated *M-orfb* in *S. woodiana* appears to be rather distant in structure and sequence from the same species M-ORFa and *V. ellipsiformis* M-ORF. When aligning homologous ORFan

proteins from different families, the best overall C $\alpha$  and C $\beta$  atoms distance values and identity scores for the F-ORFs are those from the Unionidae versus Margaritiferidae comparisons. Hyriidae and Unionidae M-ORFs and Margaritiferidae M-ORF1 obtain similar, if not identical, C $\alpha$  and C $\beta$  atoms distances results when compared among them in all combinations. These values are slightly higher when considering Margaritiferidae M-ORF2 sequences versus Hyriidae and Unionidae M-ORFs and, while the sequence identity of aligned amino acids tends to be higher for M-ORF1, the identity of aligned secondary structures is higher for M-ORF2 than that of M-ORF1 in the same comparisons.

Multiple 3D alignments of ORFan proteins proved challenging for all groups we considered, i.e. all F-ORFs, all M-ORFs, and only Margaritiferidae M-ORFs (for the M-ORFs, we also tried to include the *A. trapesialis* proteins into the alignments). Only for a few combinations



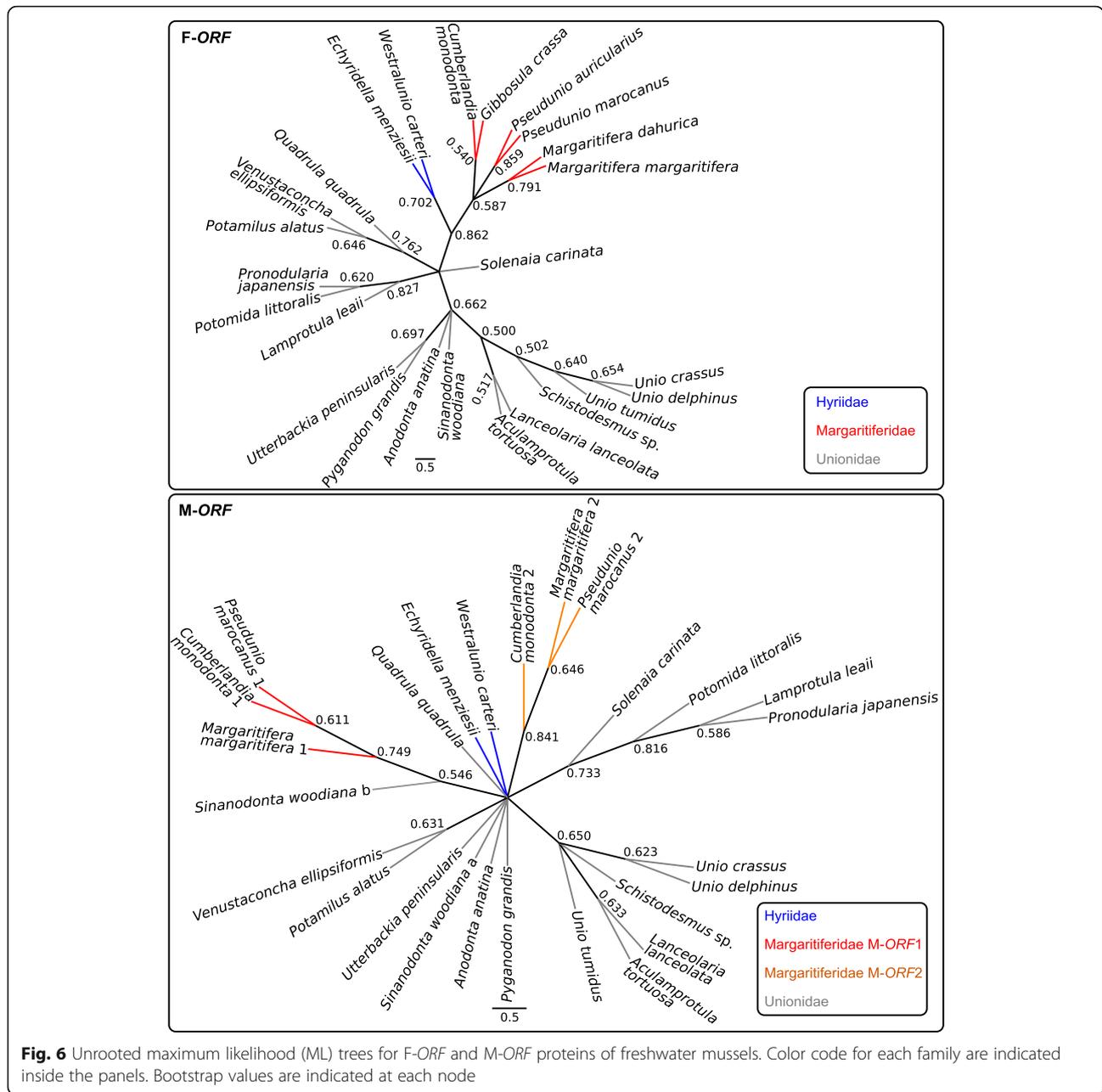
**Fig. 5** Summary of the CLANS analysis for F-ORFs and M-ORFs. Because the original CLANS output is a three-dimensional space, here are shown the three two-dimensional faces of the cube (one for each possible couple of axis: X vs Y, Z vs Y, Z vs X) obtainable by rotating the three-dimensional space of each analysis with 90° movements on one axis. The '+' inside each panel represents the center of the cube. Each dot represents a single protein sequence (color code in the legends)

of sequences the alignment could successfully produce a tree but, because of the low number of sequences available, the trees did not show any informative clear-cut clustering patterns that could split the proteins into, for example, family-specific (e.g., Hyriidae vs Margaritiferidae vs Unionidae) or kind-specific (e.g., Margaritiferidae M-ORF1 vs M-ORF2) patterns as in the sequence-based analyses.

## Discussion

The deep relationships among freshwater mussel families have long been debated [46] but, although mt genomes from the sixth family Etheriidae are at present still not available, the phylogenies presented here support a sister group relationship between Hyriidae and Margaritiferidae + Unionidae, and an equally strict relationship between Mulleriidae and Iridinidae, although not well resolved: comparable family-level topologies were also found by other recent studies [5, 42, 46] using different analytical methods and/or taxa. Such topology supports an early classification by [47] that splits freshwater mussels in two superfamilies, Unionoidea (Unionidae + Margaritiferidae + Hyriidae) and Etherioidea (Mulleriidae + Iridinidae + Etheriidae). In the light of our and the other mentioned results [5, 42, 46], and as properly discussed by [46], the separation in these two major taxa better reflects the

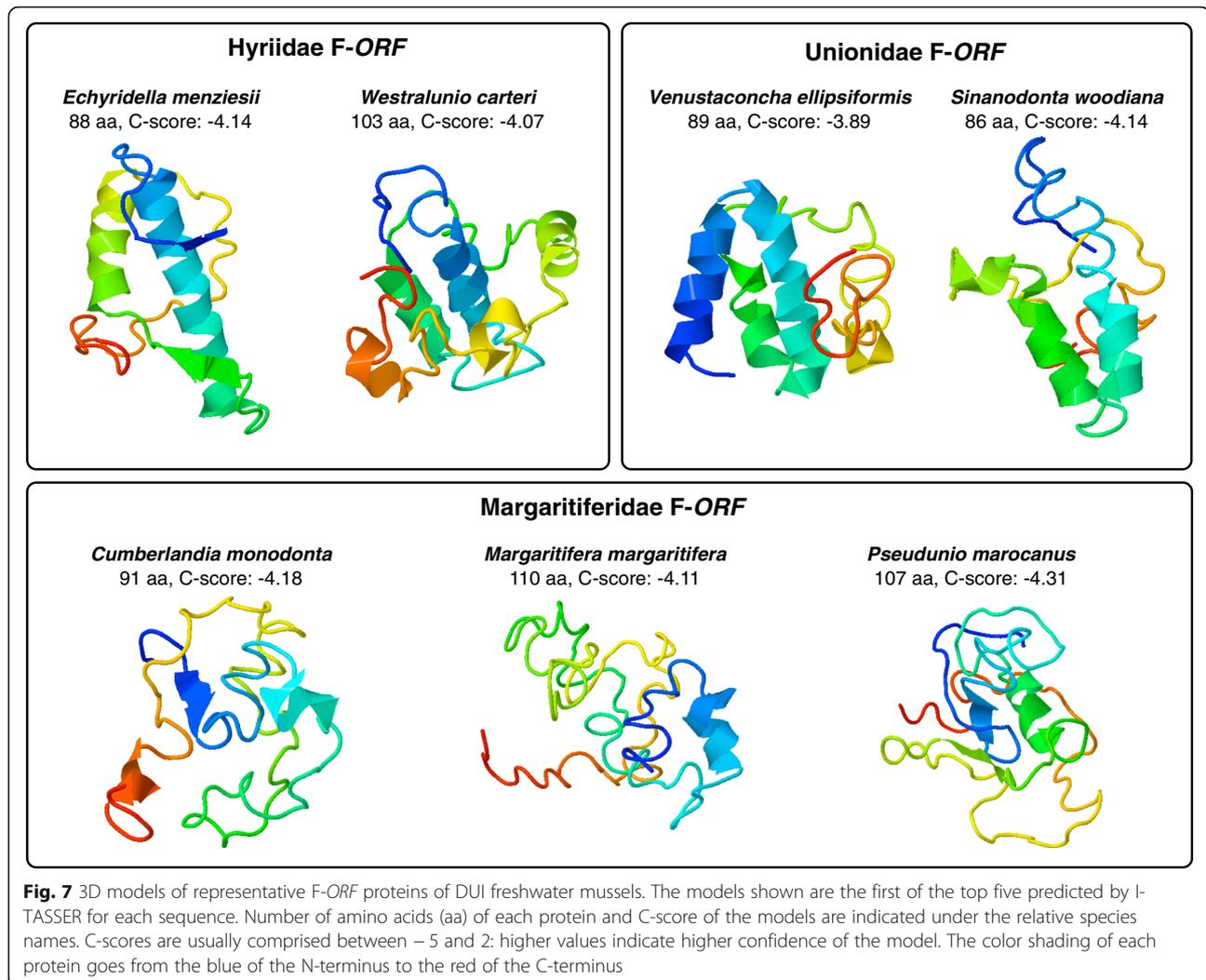
monophyly of shared characters among families than others that introduce a third superfamily, Hyrioidea, for Hyriidae only (as in [5]). It is worth noting how the female-transmitted mt genomes of Unionoidea, the superfamily where DUI is common, form a single clade sister to one containing mtDNAs of the non-DUI superfamily Etherioidea. This, however, does not imply that DUI was present only in the common ancestor of the Unionoidea. Indeed, the variable position of the M mtDNAs clade suggests the presence of DUI either in the last common ancestor of all freshwater mussels (Fig. 1) or even earlier, before the split between orders Unionida and Trigoniida, represented by the species *N. margaritacea* (Fig. 2). This is because when speciation occurs after DUI appears, F and M genomes evolve according to a “sex-associated” phylogenetic pattern [48] in two distinct clades and, inside these two clades, the relationships among mtDNAs of the various species are the same. Our phylogenies, therefore, suggest that (1) DUI was lost by Iridinidae and Mulleriidae, as well as by the South American lineage of Hyriidae (as discussed in detail below), and (2) that at least the last common ancestor of all Unionida had DUI. Comparable phylogenies were already retrieved with different methods and taxa [5, 42, 46] and, although the mt genomes sequenced in this study largely follow already described architectures [5, 29] (Fig. 3), we observed new interesting



features that help us reconstruct the evolution of freshwater mussel mt genomes and their relationship with DUI.

Starting with family Mulleriidae, the four new mtDNAs have no trace of rearrangements that are reminiscent of the ORFans *AtraUR219* and *AtraUR2218* found in *A. trapesialis* (Table 2, Fig. 3), which were hypothesized to be early versions of M-orfs [5]. The structure of Mulleriidae mt genomes, therefore, makes them much more similar to Iridinidae mtDNAs, which also lack additional ORFans between *atp8* and *nad4L* (Table 2, Fig. 3). This can be interpreted as evidence for negative selection against the rise of novel

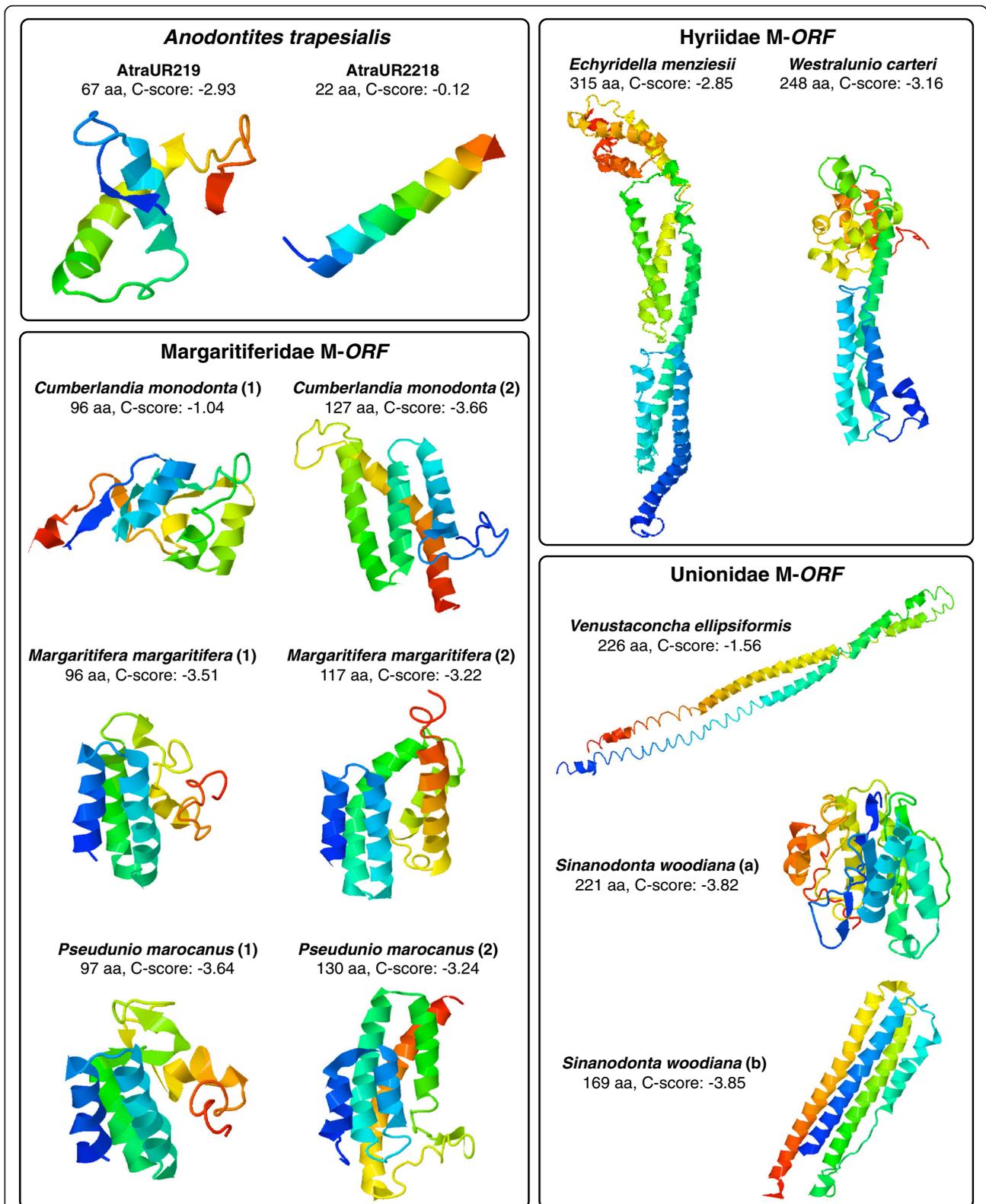
coding sequences in both families Iridinidae and Mulleriidae. It is however notable how the two mentioned ORFans are present only in *A. trapesialis*, sister species to all other Mulleriidae in our phylogenies (Figs. 1, 2 and 3). Until further studies, this might indicate an independent and relatively recent genomic rearrangement in *A. trapesialis* giving rise to its two ORFans: therefore, future works investigating ORFans evolution should consider the possibility that both *AtraUR219* and *AtraUR2218* may be relicts of a species-specific duplication event, unrelated to M-orfs of other DUI families and possibly non-functional. Also, the selection against new sequences may not be correlated to the sexual



system of these species, as both dioecious (*A. elongata*, *F. fossiculifera*, *C. rubens*) and hermaphroditic (*M. parchappi*) ones (Table 1) seem subject to it. For Iridinidae, *C. rubens* mtDNA also confirms that the non-coding region between *nad5* and *trnF* is unusually large compared to other freshwater mussel families [5]; whether it is a control region containing regulative motifs not found in other freshwater mussels will be a matter for future studies. Finally, the relationships among the examined Iridinidae and Mulleriidae species obtained from our phylogenies do not help in solving the history of their mt genome architectures. The two families, although strictly related, do not form separate monophyletic clades, and the two congeneric species *A. trapesialis* (hermaphrodite [39–41]) and *A. elongata* (dioecious [32]) are distantly related (Figs. 1 and 2); whether this situation calls for taxonomic revisions or not should be a matter for further ad hoc studies.

The new Margaritiferidae female-transmitted mt genome of *P. auricularius* has an F-*orf*, as previously

described for this taxon [3, 5, 29] (Table 2, Fig. 3), and is strictly related to F mt genomes: further studies are surely needed to fully confirm the presence of DUI in this species, but the available evidence points to this direction. The re-analysis of published Unionidae mt genomes allowed us to confirm a recent species-specific duplication of the M-*orf* in *S. woodiana* M mtDNA, as noted by [12] (Fig. 3). We are unable to say what effects (if any) this mutation may have caused to *S. woodiana* DUI system, but our results suggest that the M-*orfa* (immediately upstream of *nad4L*) is the one more similar to those of other Unionidae, while M-*orfb* (immediately upstream of *trnD*) appears different, although still recognizable as an M-*orf*. This feature of *S. woodiana* and the previously described rearrangements in *A. trapesialis* support the idea that the *atp8-nad4L* region in the mtDNAs of freshwater mussels is a hotspot for significant rearrangements and gene duplications, therefore offering additional support to the hypothesis for which the



**Fig. 8** 3D models of the proteins encoded by *Anodontites trapesimalis* ORFs and of representative M-ORF proteins of DUI freshwater mussels. The models shown are the first of the top five predicted by I-TASSER for each sequence. Number of amino acids (aa) of each protein and C-score of the models are indicated under the relative species names. C-scores are usually comprised between -5 and 2: higher values indicate higher confidence of the model. The color shading of each protein goes from the blue of the N-terminus to the red of the C-terminus

**Table 5** Summary statistics of the pairwise 3D alignments performed with MATRAS

| Families                     | Proteins compared   | N | Rdis (%) |       | RMS (Å) |      | DRMS (Å) |      | SqID (%) |       | Sec (%) |       |
|------------------------------|---|---|----------|-------|---------|------|----------|------|----------|-------|---------|-------|
|                              |   |   | mean     | SD    | mean    | SD   | mean     | SD   | mean     | SD    | mean    | SD    |
| Hyriidae                     | F-ORF   | 1 | 4.50     | n.a.  | 5.04    | n.a. | 3.38     | n.a. | 9.30     | n.a.  | 46.50   | n.a.  |
|                              | M-ORF   | 1 | 1.00     | n.a.  | 4.11    | n.a. | 2.07     | n.a. | 5.10     | n.a.  | 100.00  | n.a.  |
| Margaritiferidae             | F-ORF   | 3 | 7.50     | 4.16  | 8.98    | 4.13 | 5.11     | 0.68 | 13.20    | 7.28  | 92.43   | 7.92  |
|                              | M-ORF1  | 3 | 10.90    | 17.33 | 3.58    | 1.92 | 2.58     | 1.21 | 20.87    | 23.86 | 63.97   | 13.95 |
|                              | M-ORF2  | 3 | 9.67     | 7.11  | 5.35    | 3.84 | 3.53     | 2.29 | 14.23    | 5.66  | 83.37   | 5.70  |
|                              | all M-ORF1 x M-ORF2                                       | 9 | 7.29     | 9.22  | 3.05    | 1.13 | 2.21     | 0.81 | 17.12    | 11.84 | 71.82   | 19.55 |
|                              | single species M-ORF1 x M-ORF2                            | 3 | 12.30    | 13.03 | 3.34    | 1.78 | 2.68     | 1.30 | 22.73    | 7.35  | 73.23   | 17.50 |
| Unionidae                    | F-ORF   | 1 | 4.10     | n.a.  | 4.18    | n.a. | 2.76     | n.a. | 5.40     | n.a.  | 70.30   | n.a.  |
|                              | <i>V. ellipsiformis</i> M-ORF x <i>S. woodiana</i> M-ORFa | 1 | 0.40     | n.a.  | 2.36    | n.a. | 1.22     | n.a. | 0.00     | n.a.  | 77.30   | n.a.  |
|                              | <i>V. ellipsiformis</i> M-ORF x <i>S. woodiana</i> M-ORFb | 1 | 5.20     | n.a.  | 35.75   | n.a. | 34.40    | n.a. | 9.20     | n.a.  | 86.20   | n.a.  |
|                              | <i>S. woodiana</i> M-ORFa x M-ORFb                        | 1 | 0.10     | n.a.  | 3.75    | n.a. | 1.84     | n.a. | 0.00     | n.a.  | 52.90   | n.a.  |
| Hyriidae x Margaritiferidae  | F-ORF   | 6 | 0.08     | 0.21  | 2.79    | 1.50 | 2.33     | 0.92 | 3.97     | 6.26  | 66.83   | 43.03 |
|                              | M-ORF x M-ORF1  | 6 | 0.45     | 0.28  | 3.50    | 3.40 | 1.85     | 1.62 | 8.77     | 9.03  | 76.22   | 16.18 |
|                              | M-ORF x M-ORF2  | 6 | 2.10     | 1.47  | 4.74    | 2.24 | 2.62     | 0.60 | 8.17     | 4.72  | 84.40   | 11.79 |
| Hyriidae x Unionidae         | F-ORF   | 4 | 5.83     | 3.60  | 6.36    | 3.33 | 4.07     | 1.43 | 8.55     | 3.52  | 60.18   | 10.90 |
|                              | M-ORF   | 6 | 1.30     | 1.54  | 3.50    | 0.90 | 1.92     | 0.57 | 7.65     | 3.57  | 70.77   | 7.06  |
| Unionidae x Margaritiferidae | F-ORF   | 6 | 0.03     | 0.08  | 2.00    | 0.98 | 1.78     | 0.49 | 18.75    | 14.37 | 72.22   | 44.31 |
|                              | M-ORF x M-ORF1  | 9 | 0.33     | 0.39  | 2.50    | 1.42 | 1.90     | 0.82 | 11.01    | 9.32  | 67.89   | 27.66 |
|                              | M-ORF x M-ORF2  | 9 | 1.84     | 1.53  | 3.83    | 1.10 | 2.35     | 0.86 | 6.07     | 4.16  | 71.53   | 21.04 |

N Number of pairwise comparisons considered, Rdis Normalized Sdis score (%), RMS Root mean square deviation of C $\alpha$  atoms location of aligned amino acids after optimal superimposition (Å), DRMS Root mean square deviation of distances between C $\beta$  atoms of aligned amino acids (Å), SqID % of identical amino acid pairs on the total number of aligned amino acids, Sec % of identical secondary structure residues on the total number of aligned amino acids, SD Standard deviation of the mean for a given value (n.a. when N = 1)

unionid M-*orf* may have originated from a duplication of *atp8* [5, 28].

The mtDNAs of the only dioecious species of Hyriidae in this study showing DUI, *W. carteri*, are comparable in all aspects to those of the previously sequenced DUI species *E. menziesii* [5], a strictly related species with the same Australasian distribution. In particular, both the M-*orf* and the *cox2* in the M mtDNA are confirmed to be longer in this family compared to the others (Table 3). The other three Hyriidae species, *C. ambigua*, *D. suavidicus*, *P. obliquus* (all from South America and always forming a single monophyletic clade; Table 1, Figs. 1 and 2), did not show evidence of DUI and, contrary to DUI gonochoric and DUI-less hermaphroditic unionids and margaritiferids, their F-like mtDNAs do not possess any F-*orf* or H-*orf*, and we did not find evidence of their translocation in the unassigned regions of these mt genomes (Table 2, Fig. 3). Even if the sex determination system is known only for *C. ambigua*, a gonochoristic DUI-less species, we can propose a working hypothesis stating that, in Hyriidae, losing DUI (1) may always cause the complete disappearance

of the F-*orf* in the former F mtDNA (as in *C. ambigua*, *D. suavidicus*, *P. obliquus*), and (2) may not affect the gonochoristic sexual system (as in *C. ambigua*). In this family, similarly to Mulleriidae and Iridinidae which are hypothesized to have lost DUI in the early stages of their radiation [5], it seems therefore that the relationship among DUI, presence of ORFans, and gonochorism may be somewhat different compared to Unionidae and Margaritiferidae, which retain a H-*orf* in their mt genomes after losing DUI [3]. The parallelism between Hyriidae and Mulleriidae + Iridinidae may, however, lead to another hypothesis: we can see that among all Hyriidae species studied until now, only the Australasian ones (*E. menziesii*, *W. carteri*) show DUI, while the Neotropical ones (*C. ambigua*, *D. suavidicus*, *P. obliquus*) do not (Table 1). This may hint that the last common ancestor for these two lineages had DUI, which was lost (together with the ORFan in the remaining F-like mtDNA) only in the South American lineage during its radiation. Considering the current information from all families of freshwater mussels, we can speculate that once DUI and the M mtDNA are lost by a species, the ORFan

in the remaining F mtDNA (i.e. the *F-orf*) no longer plays a role in the DUI mechanism and gradually disappears. First, it may start to accumulate mutations and degenerate (like the *H-orf* in Unionidae and Margaritiferidae [3]) and then, given enough time, it completely disappears from the mtDNA without leaving recognizable traces (as in the South American Hyriidae, which no longer carry traces of the *F-orf*; Table 2, Fig. 3).

Despite very similar amino acid compositions (Fig. 4), sequence-based analyses of ORFan proteins managed in most cases to distinguish families (Figs. 5 and 6), with *F-ORFs* giving better resolution compared to *M-ORFs* (as in the ML analysis, for example). Following this lead, we explored for the first time the total putative 3D folding of freshwater mussels ORFan proteins (as secondary structures and other features have been already thoroughly characterized [27, 28, 49]), to search for patterns that could help unravel their evolutionary history. Indeed, even if we sampled only a few representative species, the predicted 3D foldings demonstrate how in each family the *F-* and *M-ORFs* have their own peculiar shape (Figs. 7 and 8). The tertiary structure of a protein is influenced by its amino acid sequence, and as a consequence, the proteins of closely related species and/or with a common origin exhibit comparable shapes. In Margaritiferidae, for example, all *M-ORF1* and *M-ORF2* proteins fold in a similar way, and this finding could also hint at a full functionality of the *M-ORF2* protein in DUI margaritiferids, possibly with a physiological role comparable to that of *M-ORF1* and other *M-ORFs*. In contrast, in the case of Unionidae *M-ORFs* (Fig. 8), the proteins exhibit a range of very divergent foldings; this could be due to the method used, the phylogenetic distance among species (see their positions in the trees in Figs. 1 and 2), or the exclusive evolutionary history of each protein (as we discussed above for *S. woodiana* *M-ORFs*). A broader sampling from each family and a comparison among different methods of tertiary structure prediction in the future could help to better define these ambiguous foldings, as well as to improve the distance analyses that lacked resolution in the current study.

Nonetheless, despite the technical difficulties, our explorative study presents evidence of possibly conserved features among ORFan proteins of the same kind, such as the relative arrangement of certain helices in *F-ORFs* and *M-ORFs*. These structural features, together with properties already characterized (e.g., [5, 27, 28, 49]) and others yet to be discovered, will lead us to give a precise physiological role to the ORFan proteins and their respective genes (like those already hypothesized before, for example [28]). With further study, we might also be able to answer the long-standing questions about the relationships among the

ORFan genes, the sex determination system, and the peculiar mitochondrial inheritance mode of freshwater mussels and of all other bivalves showing DUI [3]. However, given the rather different length and shape of Hyriidae *M-ORFs* compared to those of Unionidae and Margaritiferidae, and the fact that Hyriidae DUI-less species are not always hermaphroditic (as *C. ambigua*) and their mtDNA does not possess any *F-* or *H-orf* (*C. ambigua*, *D. suavidicus*, *P. obliquus*), as opposed to what occurs in Margaritiferidae and Unionidae [3], we suspect that the nature of the link among ORFans and sexual system may not be exactly the same in all DUI freshwater mussels. On the other hand, it is also possible that long evolutionary times in the absence of DUI, as well as various ecological pressures [50], may shape freshwater mussel mt genomes and/or sexual systems, leading to the situations described for the first time in this study (i.e., no ORFan genes in both hermaphroditic and gonochoric species). To answer these questions, a broader sampling and in vivo studies on the ORFan proteins will be needed.

## Conclusion

In this study we produced ten entire, plus one almost complete, new mtDNA sequences of freshwater mussel species with or without DUI from still poorly sampled families (Iridinidae, Mulleriidae, Hyriidae, and Margaritiferidae). Besides being a useful basis for future sequencing efforts, much needed for this group of endangered animals [51], we provided new data on the mitochondrial ORFan genes of freshwater mussels, whose origin is still unknown and whose function and conservation are likely related to their sex determination system [3]. We observed that the rearrangements occurring in mitochondrial genomes of species and lineages that secondarily lost DUI (i.e., that lost their ancestral M mtDNA and retain only a mutated F) are not always the same, and that losing DUI is not always linked to a switch to hermaphroditism. By analyzing the 3D structures of their translated proteins, we also evidenced common characteristics and similarities among them, hinting at conserved physiological roles of *F-orf* and *M-orf* genes in all DUI lineages of freshwater mussels, as well as family-specific ones. We therefore questioned if the family-specific structures of the ORFan proteins can influence some detail of the DUI system in different manners, so that the downstream effect of losing DUI on the sexual system of a species may vary. An alternative, but not necessarily mutually exclusive, hypothesis we propose to explain the observed differences among non-DUI lineages is that time and other factors may play an important role in reshaping both the mitochondrial genome and sexual system of a species after it loses DUI.

## Methods

### Sequencing of new mt genomes

Freshwater mussel species were selected across the main families within the order Unionida and to cover distinct sexual strategies, i.e. gonochorism and hermaphroditism (Table 1). *W. carteri* specimens were taken from the wild with permits for field and laboratory studies obtained from the Western Australian Department of Environment and Conservation under Regulation 17 of the Wildlife Conservation Act 1950 (SF007049) and Department of Fisheries under Exemption from the Fisheries Resources Management Act 1994 (1724–2010-06). Sex of the specimens was determined by observing gonad tissue smears for sexual cells and/or the demibranchs for the presence of marsupia, using a dissecting microscope. Tissue samples were excised from one specimen per species, and placed in 100% ethanol for DNA extraction: for all species, a foot clip was available for DNA extraction, while for *W. carteri* an additional male gonad sample from the same specimen was also used. DNA extraction followed [52]. DUI presence or absence for every species was assumed from previous studies [30–37]. The complete mitogenomes sequencing and assemblage was accomplished using the pipeline proposed by [53]. Annotations were performed using MITOS [54] with the final tRNA genes limits being rechecked with ARWEN [55]. Finally, personal scripts were developed and applied to adjust the mtDNA protein-coding limits since MITOS seems to underestimate gene length (for details, go to <https://figshare.com/s/a756ef19cec8f65d506a>).

### Phylogenetic analyses of freshwater mussel mt genomes

The set of eleven new mt genomes (Table 1) was expanded by adding other 51 freshwater mussel mt genomes from GenBank (see Additional file 1: Table S1 for the complete list and details), for a total of 62 mtDNAs for 40 species. The mt genomes added encompass families Iridinidae, Mulleriidae, Hyriidae, Margaritiferidae, and Unionidae. For Hyriidae and Unionidae, we took the mt genomes from DUI species for which both the F and M mtDNAs were available (resulting in two mtDNAs per species), for Margaritiferidae all F and M mt genomes available, and for Margaritiferidae and Unionidae only also those from secondarily hermaphrodite species (i.e., the H mtDNAs). The final set was thus composed of 22 M, 25 F, and 15 mt genomes from non-DUI species, either secondarily hermaphrodite that lost DUI (sensu [3]) or gonochoric ones. For the purpose of the phylogenetic analyses, three outgroup species mt genomes were also used: the bivalves *Neotrigonia margaritacea* (Trigoniida) and *Solemya velum* (Solemyida), plus *Chaetoderma nitidulum* (Caudofoveata, Chaetodermitida) as outgroup to all bivalves (respective GenBank accession numbers: KU873118, JQ728447, EF211990). A

total of 65 mt genomes was therefore considered for the phylogenies.

We extracted all protein coding gene sequences, except *atp8* because of its short length and high variability, from the 65 mtDNAs and translated them with the invertebrate mitochondrial genetic code to obtain the relative protein sequences. The 12 protein sets were then aligned with M-Coffee (<http://tcoffee.crg.cat/apps/tcoffee/do:mcoffee>) [56, 57] using all multiple methods available, then from these protein alignments we retro-aligned the codons of the respective genes using the TranslatorX server (<http://translatorx.co.uk>) [58]. Both protein and codon alignments were trimmed on the Gblocks server version 0.91b ([http://molevol.cmima.csic.es/castresana/Gblocks\\_server.html](http://molevol.cmima.csic.es/castresana/Gblocks_server.html)) [59, 60] using the option for a more stringent selection. jModelTest2 [61, 62] was used to calculate, under the Bayesian Inference Criterion (BIC), the best-fit models of nucleotide substitution for the trimmed codon alignments. Finally, the two sets of trimmed alignments were concatenated respectively into a codon alignment and an amino acid alignment, with a respective length of 7914 and 2363 gapless positions. MrBayes 3.2.3 [38] was used to infer amino acid- and nucleotide-base phylogenies of the mt genomes. The analysis for both concatenated alignments consisted of two separate runs of four chains, 5,000,000 generations, sampling every 100 trees with a burn-in of 0.1%. In the nucleotide analysis, the models retrieved with jModelTest2 were specified for each gene partition, and a ‘4by4’ nucleotide substitution model was adopted for the whole alignment. In the amino acid analysis, a ‘mixed’ rate matrix was specified. Completed runs were accepted for further examination after checking that their standard deviation of split frequencies stabilized at values < 0.01 over the generations (as in [63]). jModelTest2 and MrBayes 3.2.3 were ran on the CIPRES Science Gateway (<http://www.phylo.org>) [64]. Trees were graphically edited with FigTree v.1.4.3 [65].

### Annotation of F- and M-*orf* genes

To locate the F- and M-*orf* genes in the DUI genomes in which they were not annotated, and at the same time validate previous annotations, we first used the EMBOSS tool getorf [66] to extract all possible ORFs  $\geq 30$  nucleotides long (i.e., coding at least 10 codons) under the invertebrate mitochondrial genetic code from the 62 freshwater mussel mtDNAs dataset described above, and then translated them into the corresponding proteins. This set of protein sequences was first searched with the HMMER tool jackhmmer [43] (10 iterations) using as seeds the F-ORF of *V. ellipsiformis* and the M-ORFs of *E. menziesii*, *C. monodonta* (M-ORF1) and *V. ellipsiformis*, in separate runs. The proteins retrieved from each run were then aligned with PSI-Coffee [56, 67] (<http://>

[tcoffee.crg.cat/apps/tcoffee/do:psicoffee](http://tcoffee.crg.cat/apps/tcoffee/do:psicoffee); all pairwise methods selected), and the alignments used to build HMM profiles with hmmbuild [43] (options: -fast -sym-frac 0 -fragthresh 0 -wnone -enone; see also [28]). These profiles were used to search again the whole set of proteins with hmmsearch [43] (--max option active to allow maximum sensitivity) to confirm the presence of F-orfs and M-orfs previously found with jackhmmer, retrieve the known ones not recognized by jackhmmer, and search for new homologs of these genes. Finally, phmmer [43] (--max option active) was used to search the protein set for homologs of the two proteins putatively encoded by the two ORFans *AtraUR219* and *AtraUR2218* in *A. trapesialis* (Mulleriidae) mtDNA, hypothesized to be related to M-orf genes [5].

### Sequence-based analyses of proteins

MEGA7 [68] was used to calculate the amino acid composition of F-ORFs, M-ORFs, and *AtraUR219* and *AtraUR2218* of *A. trapesialis*. To visualize the relationships among the already annotated and newly discovered ORFan proteins based on pairwise similarity, we ran CLANS [44]. Specifically, the CLANS program was conducted online on the MPI Bioinformatic Toolkit website [69] (<https://toolkit.tuebingen.mpg.de/#/tools/clans>) using the BLOSUM45 scoring matrix and BLAST HSP's E-values up to 1E-4. The output was then run locally on the CLANS application for  $\geq 10,000,000$  rounds to obtain reliable 3D clustering data. The alignments of the F-ORFs and M-ORFs used to build their HMM models, plus an alignment of the two *A. trapesialis* ORFan proteins and the M-ORFs (constructed with PSI-Coffee as mentioned above), were used to build ML trees in MEGA7 [68], using 1000 bootstraps, the 'mtREV' model of substitution, uniform rates among sites, and a gap partial deletion of 95%. The trees were built unrooted because given the uncertain origin of the ORFan genes it is not possible to choose a reliable outgroup sequence.

### 3D structure-based analyses of proteins

We used I-TASSER [45] online (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) to obtain the 3D models of the F- and M-ORF of some representative DUI species (Hyriidae: *E. menziesii*, *W. carteri*; Margaritiferidae: *C. monodonta*, *M. margaritifera*, *P. maroccanus*; Unionidae: *V. ellipsiformis*, *S. woodiana*) plus *AtraUR219* and *AtraUR2218* of *A. trapesialis* (Mulleriidae). The most supported models (i.e., the ones with the best C-score) were then used as input for MATRAS [70] (<http://strcomp.protein.osaka-u.ac.jp/matras/>) to perform pairwise and multiple 3D alignments of the proteins. The multiple alignments aimed at obtaining trees based on DRMS (root mean square deviation of C $\alpha$  atoms, measured in Å) distances among them, using as a minimal

set the ORFan proteins from *E. menziesii*, *C. monodonta*, and *V. ellipsiformis* (which have been thoroughly characterized in past studies [3, 5, 27, 28]) and adding as much proteins as MATRAS would allow from the other species. When an I-TASSER model made MATRAS fail in producing a tree, we refined it with ModRefiner [71] (<https://zhanglab.ccmb.med.umich.edu/ModRefiner/>) and repeated the 3D alignment. If the refining did not succeed in improving the results, the protein was removed from the analysis.

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12862-019-1554-5>.

**Additional file 1: Table S1.** Summary of the additional 51 freshwater mussel mt genomes used in this study that were retrieved from GenBank. **Table S2.** Summary of the jackhmmer analyses (F-ORF and M-ORF sequences input vs protein set). **Table S3.** Summary of the hmmsearch analyses (F-ORF and M-ORF HMM profiles input vs protein set). **Table S4.** Summary of the phmmer analyses (*A. trapesialis* ORFan proteins input vs protein set).

**Additional file 2.** List of mitochondrial ORFan gene sequences of freshwater mussels whose translated protein products were used in this study.

### Abbreviations

3D: Three-dimensional; aa: Amino acid; *atp8*: ATP synthase F0 subunit 8 gene; *AtraUR219*, *AtraUR2218*: ORFan genes found in *Anodontites trapesialis* (Mulleriidae) mtDNA hypothesized to be related to the M-orf in the M mtDNA of DUI freshwater mussels; *AtraUR219*, *AtraUR2218*: Protein products of *AtraUR219* and *AtraUR2218*, respectively; BIC: Bayesian Information Criterion; *cox2*: Cytochrome c oxidase subunit 2 gene; C $\beta$ :  $\beta$ -carbon atom of an amino acid, located right before the  $\alpha$ -carbon; C-terminus: End of a protein chain; C $\alpha$ :  $\alpha$ -carbon atom of an amino acid, located right before the carbonyl carbon; DUI: Doubly uniparental inheritance (of mitochondria); F-orf: ORFan gene typically found in the F mtDNA of freshwater mussels showing DUI; F-ORF: Protein product of a F-orf gene; HMM: Hidden Markov Model; H-orf: ORFan gene typically found in the mtDNA of strictly hermaphroditic Unionidae and Margaritiferidae species that secondarily lost DUI, it is a mutated version of a F-orf; H-ORF: Protein product of a H-orf gene; ML: Maximum likelihood; M-orf: ORFan gene typically found in the M mtDNA of freshwater mussels showing DUI, usually found in single-copy but present in two copies in Margaritiferidae (named 1 and 2) and in *Sinanodonta woodiana* (Unionidae) (named a and b); M-ORF: Protein product of a M-orf gene; mt: Mitochondrial; mtDNA: Mitochondrial DNA; *nad2*, -4 L, -5, 6: NADH dehydrogenase subunits 2, 4 L, 5, 6 genes, respectively; N-terminus: Start of a protein chain; ORF: Open reading frame; ORFan: ORF with no recognizable homology or similarity to known genes; tRNA: Transfer RNA; *trnD*, -E, -F, -H, -V: Genes encoding tRNAs respectively for aspartic acid, glutamic acid, phenylalanine, histidine, valine

### Acknowledgements

From the Western Australian Museum, we thank Lisa Kirkendale and Corey Whisson, Curator and Mollusca Collection Manager, respectively for arranging voucher lodgements and tissue loans of *Westralunio carteri*. We thank Alan J. Lymbery and Murdoch University for providing laboratory access during sample preparations. We thank the reviewers for their suggestions that improved the article.

### Authors' contributions

DG, MLL, SB designed the study and wrote the manuscript. PO, RA, MWK, CC, VP collected and provided samples. EF and HMG performed DNA extraction, sequencing and assembly of whole mitogenomes. MLL annotated the newly obtained mtDNAs. DG performed the analyses. EF, HMG, PO, RA, MWK, CC, VP, AEB, DTS participated in revising the manuscript drafts during the writing process. All authors read and approved the final version of the manuscript.

### Funding

This research was developed under ConBionics: the missing approach for the Conservation of freshwater Bivalves Project N° NORTE-01-0145-FEDER-030286, co-financed by COMPETE 2020, Portugal 2020 and the European Union through the ERDF, by FCT through national funds, and by Life+ Margal Ulla 09NAT/ES/000514. FCT also supported MLL (SFRH/BD/115728/2016). The research programs of SB and DTS are funded by the Natural Sciences and Engineering Research Council, Discovery Grants RGPIN/435656–2013 and RGPIN/217175–2013, respectively.

### Availability of data and materials

The mtDNA sequences obtained in this study have been submitted to GenBank under the accession numbers MK761136–46.

### Ethics approval and consent to participate

Permits for collection of *Westralunio carteri* were obtained from the Western Australian Department of Environment and Conservation (SF007049) and Department of Fisheries (1724–2010-06).

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Département de Sciences Biologiques, Université de Montréal, Montréal, QC, Canada. <sup>2</sup>CIBIO/InBIO - Research Center in Biodiversity and Genetic Resources, University of Porto, Campus Agrário de Vairão, Vairão, Portugal. <sup>3</sup>CLIMAR - Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Matosinhos, Portugal. <sup>4</sup>Deakin Genomics Centre, School of Life and Environmental Sciences, Deakin University, Geelong, Victoria, Australia. <sup>5</sup>Departamento de Zooloxia, Xenética e Antropoloxia Física, Facultade de Veterinaria, Universidade de Santiago de Compostela, Campus de Lugo, Lugo, Spain. <sup>6</sup>BWG Environmental, Brisbane, QLD, Australia. <sup>7</sup>Mollusca, Department of Aquatic Zoology, Western Australian Museum, Welshpool, WA, Australia. <sup>8</sup>School of Veterinary and Biological Sciences, Murdoch University, Perth, WA, Australia. <sup>9</sup>ECOBiv - Ecology and Conservation of Bivalves Research Group, Department of Biology and Zoology, Federal University of Mato Grosso, Cuiabá, MT, Brazil. <sup>10</sup>Institut Systématique Evolution Biodiversité ISYEB - Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France. <sup>11</sup>North Carolina Museum of Natural Sciences, Raleigh, NC, USA. <sup>12</sup>Department of Biology, Acadia University, Wolfville, NS, Canada.

Received: 3 February 2019 Accepted: 9 December 2019

Published online: 19 December 2019

### References

- Zouros E. Biparental inheritance through Uniparental transmission: the doubly uniparental inheritance (DUI) of mitochondrial DNA. *Evol Biol*. 2013; 40:1–31. <https://doi.org/10.1007/s11692-012-9195-2>.
- Gusman A, Lecomte S, Stewart DT, Passamonti M, Breton S. Pursuing the quest for better understanding the taxonomic distribution of the system of doubly uniparental inheritance of mtDNA. *PeerJ*. 2016;4:e2760. <https://doi.org/10.7717/peerj.2760>.
- Breton S, Stewart DT, Shephardson S, Trdan RJ, Bogan AE, Chapman EG, Ruminas AJ, Piontkivska H, Hoeh WR. Novel protein genes in animal mtDNA: a new sex determination system in freshwater mussels (Bivalvia: Unionoida)? *Mol Biol Evol*. 2011;28:1645–59. <https://doi.org/10.1093/molbev/msq345>.
- Doucet-Beaupré H, Breton S, Chapman EG, Blier PU, Bogan AE, Stewart DT, Hoeh WR. Mitochondrial phylogenomics of the Bivalvia (Mollusca): searching for the origin and mitogenomic correlates of doubly uniparental inheritance of mtDNA. *BMC Evol Biol*. 2010;10:50. <https://doi.org/10.1186/1471-2148-10-50>.
- Guerra D, Plazzi F, Stewart DT, Bogan AE, Hoeh WR, Breton S. Evolution of sex-dependent mtDNA transmission in freshwater mussels (Bivalvia: Unionoida). *Sci Rep*. 2017;7:1551. <https://doi.org/10.1038/s41598-017-01708-1>.
- Curole JP, Kocher TD. Ancient sex-specific extension of the cytochrome c oxidase II gene in bivalves and the fidelity of doubly uniparental inheritance. *Mol Biol Evol*. 2002;19:1323–8. <https://doi.org/10.1093/oxfordjournals.molbev.a004193>.
- Fischer D, Eisenberg D. Finding families for genomic ORFans. *Bioinformatics*. 1999;15:759–62. <https://doi.org/10.1093/bioinformatics/15.9.759>.
- Breton S, Beaupré HD, Stewart DT, Piontkivska H, Karmakar M, Bogan AE, Blier PU, Hoeh WR. 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*. 2009;183:1575–89. <https://doi.org/10.1534/genetics.109.110700>.
- Passamonti M, Ricci A, Milani L, Ghiselli F. Mitochondrial genomes and doubly Uniparental inheritance: new insights from *Musculista senhousia* sex-linked mitochondrial DNAs (Bivalvia Mytilidae). *BMC Genomics*. 2011;12:442. <https://doi.org/10.1186/1471-2164-12-442>.
- Ghiselli F, Milani L, Guerra D, Chang PL, Breton S, Nuzhdin SV, Passamonti M. Structure, transcription, and variability of metazoan mitochondrial genome: perspectives from an unusual mitochondrial inheritance system. *Genome Biol Evol*. 2013;5:1535–54. <https://doi.org/10.1093/gbe/evt112>.
- Bettinazzi S, Plazzi F, Passamonti M. The complete female- and male-transmitted mitochondrial genome of *Meretrix lamarckii*. *PLoS One*. 2016; 11(4):e0153631. <https://doi.org/10.1371/journal.pone.0153631>.
- Burzynski A, Soroka M. Complete paternally inherited mitogenomes of two freshwater mussels *Unio pictorum* and *Sinanodonta woodiana* (Bivalvia: Unionidae). *PeerJ*. 2018;6:e5573. <https://doi.org/10.7717/peerj.5573>.
- Serb JM, Lideard C. Complete mtDNA sequence of the north American freshwater mussel, *Lampsilis ornata* (Unionidae): an examination of the evolution and phylogenetic utility of mitochondrial genome organization in Bivalvia (Mollusca). *Mol Biol Evol*. 2003;20(11):1854–66. <https://doi.org/10.1093/molbev/msg218>.
- Breton S, Doucet Beaupré H, Stewart DT, Hoeh WR, Blier PU. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? *Trends Genet*. 2007;23(9):465–74. <https://doi.org/10.1016/j.tig.2007.05.011>.
- Wu X, Li X, Li L, Xu X, Xia J, Yu Z. New features of Asian *Crassostrea* oyster mitochondrial genomes: a novel alloacceptor tRNA gene recruitment and two novel ORFs. *Gene*. 2012;507:112–8. <https://doi.org/10.1016/j.gene.2012.07.032>.
- Wu X, Li X, Li L, Yu Z. 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*. 2012;510:22–31. <https://doi.org/10.1016/j.gene.2012.08.037>.
- Kawashima Y, Nishihara H, Akasaki T, Nikaido M, Tsuchiya K, Segawa S, Okada N. The complete mitochondrial genomes of deep-sea squid (*Bathyteuthis abyssicola*), bob-tail squid (*Semiosquilla 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*. 2013;69: 980–93. <https://doi.org/10.1016/j.jmpev.2013.06.007>.
- Stöger I, Schrödl M. Mitogenomics does not resolve deep molluscan relationships (yet?). *Mol Phylogenet Evol*. 2013;69:376–92. <https://doi.org/10.1016/j.jmpev.2012.11.017>.
- Uribe JE, Colgan D, Castro LR, Kano Y, Zardoya R. Phylogenetic relationships among superfamilies of Neritimorpha (Mollusca: Gastropoda). *Mol Phylogenet Evol*. 2016;104:21–31. <https://doi.org/10.1016/j.jmpev.2016.07.021>.
- Strugnell JM, Hall NE, Vecchione M, Fuchs D, Allcock AL. Whole mitochondrial genome of the Ram's horn squid shines light on the phylogenetic position of the monotypic order Spirulida (Haeckel, 1896). *Mol Phylogenet Evol*. 2017;109: 296–301. <https://doi.org/10.1016/j.jmpev.2017.01.011>.
- Williams ST, Foster PG, Hughes C, Harper EM, Taylor JD, Littlewood DTJ, Dyal P, Hopkins KP, Briscoe AG. Curious bivalves: systematic utility and unusual properties of anomalodesmatan mitochondrial genomes. *Mol Phylogenet Evol*. 2017;110:60–72. <https://doi.org/10.1016/j.jmpev.2017.03.004>.
- Zhan X, Zhang S, Gu Z, Wang A. Complete mitochondrial genomes of two pearl oyster species (Bivalvia: Pteriomorpha) reveal novel gene arrangements. *J Shellfish Res*. 2018;37(5):1039–50. <https://doi.org/10.2983/035.037.0515>.
- Milani L, Ghiselli F, Maurizii MG, Nuzhdin SV, Passamonti M. Paternally transmitted mitochondria express a new gene of potential viral origin. *Genome Biol Evol*. 2014;6:391–405. <https://doi.org/10.1093/gbe/evu021>.
- Milani L, Ghiselli F, Pecci A, Maurizii MG, Passamonti M. The expression of a novel Mitochondrially-encoded gene in gonadic precursors may drive paternal inheritance of mitochondria. *PLoS One*. 2015;10(9):e0137468. <https://doi.org/10.1371/journal.pone.0137468>.
- Minoiu I, Burzynski A, Breton S. Analysis of the coding potential of the ORF in the control region of the female-transmitted *Mytilus* mtDNA. *Gene*. 2016; 576:586–8. <https://doi.org/10.1016/j.gene.2015.09.065>.

26. Breton S, Bouvet K, Auclair G, Ghazal S, Sietman BE, Johnson N, Bettinazzi S, Stewart DT, Guerra D. The extremely divergent maternally- and paternally-transmitted mitochondrial genomes are co-expressed in somatic tissues of two freshwater mussel species with doubly uniparental inheritance of mtDNA. *PLoS One*. 2017;12(8):e0183529. <https://doi.org/10.1371/journal.pone.0183529>.
27. Milani L, Ghiselli F, Guerra D, Breton S, Passamonti M. A comparative analysis of mitochondrial ORFans: new clues on their origin and role in species with doubly uniparental inheritance of mitochondria. *Genome Biol Evol*. 2013;5:1408–34. <https://doi.org/10.1093/gbe/evt101>.
28. Mitchell A, Guerra D, Stewart D, Breton S. *In silico* analyses of mitochondrial ORFans in freshwater mussels (Bivalvia: Unionoida) provide a framework for future studies of their origin and function. *BMC Genomics*. 2016;17:597. <https://doi.org/10.1186/s12864-016-2986-6>.
29. Lopes-Lima M, Fonseca MM, Aldridge DC, Bogan AE, Gan HM, Ghamizi M, Sousa R, Teixeira A, Varandas S, Zanatta D, Zieritz A, Froufe E. The first Margaritiferidae male (M-type) mitogenome: mitochondrial gene order as a potential character for determining higher-order phylogeny within Unionoida (Bivalvia). *J Molluscan Stud*. 2017;83:249–52. <https://doi.org/10.1093/mollus/eyx009>.
30. Walker JM, Curole JP, Wade DE, Chapman EG, Bogan AE, Watters GT, Hoeh WR. Taxonomic distribution and phylogenetic utility of gender associated mitochondrial genomes in the Unionoida (Bivalvia). *Malacologia*. 2006;48:265–82.
31. Walker JM, Bogan AE, Garo K, Soliman GN, Hoeh WR. Hermaphroditism in the Iridinidae (Bivalvia: Etherioidea). *J Molluscan Stud*. 2006;72(2):216217. <https://doi.org/10.1093/mollus/eyj072>.
32. Simone LRL. Anatomy and systematics of *Anodontites elongatus* (Swainson) from Amazon and Parana Basins, Brazil (Mollusca, Bivalvia, Unionoida, Mycetopodidae). *Rev Bras Zool*. 1997;14(4):877–88. <https://doi.org/10.1590/S0101-81751997000400011>.
33. Avelar WEP. Functional anatomy of *Fossula fossiculifera* (D'Orbigny, 1843) (Bivalvia: Mycetopodidae). *Am Malacol Bull*. 1993;10(2):129–38.
34. do Vale RS, Beasley CR, Tagliaro CH. Seasonal variation in the reproductive cycle of a Neotropical freshwater mussel (Hyriidae). *Am Malacol Bull*. 2004;18(1):71–8.
35. da Cruz Santos-Neto G, Beasley CR, Schneider H, Pimpão DM, Hoeh WR, de Simone LRL, Tagliaro CH. Genetic relationships among freshwater mussel species from fifteen Amazonian rivers and inferences on the evolution of the Hyriidae (Mollusca: Bivalvia: Unionoida). *Mol Phylogenet Evol*. 2016;100:148–59. <https://doi.org/10.1016/j.jmpev.2016.04.013>.
36. Klunzinger MW, Beatty SJ, Morgan DL, Lymbery AJ, Haag WR. Age and growth in the Australian freshwater mussel, *Westralunia carteri*, with an evaluation of the fluorochrome calcein for validating the assumption of annulus formation. *Freshwater Sci*. 2014;33(4):1127–35. <https://doi.org/10.1086/677815>.
37. Henley WF. Evaluation of diet, gametogenesis, and hermaphroditism in freshwater mussels (Bivalvia: Unionoidae). In: ETDs: Virginia Tech Electronic Theses and Dissertations, Doctoral Dissertations; 2002. <https://vtechworks.lib.vt.edu/handle/10919/27932>.
38. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61:539–42. <https://doi.org/10.1093/sysbio/sys029>.
39. Callil TC, Mansur MCD. Gametogênese e dinâmica da reprodução de *Anodontites trapesialis* (Lamarck) (Unionoida, Mycetopodidae) no lago Baía do Poço, planície de inundação do rio Cuiabá, Mato Grosso, Brasil. *Rev Bras Zool*. 2007;24(3):825–40. <https://doi.org/10.1590/S0101-81752007000300033>.
40. Callil CT, Krinski D, Silva FA. Variations on the larval incubation of *Anodontites trapesialis* (Unionoida, Mycetopodidae): synergistic effect of the environmental factors and host availability. *Braz J Biol*. 2012;72(3):1–8. <https://doi.org/10.1590/S1519-69842012000300017>.
41. Callil CT, Leite MCS, Mateus LAF, Jones JW. Influence of the flood pulse on reproduction and growth of *Anodontites trapesialis* (Lamarck, 1819) (Bivalvia: Mycetopodidae) in the Pantanal wetland, Brazil. *Hydrobiologia*. 2018;810(1):433–48. <https://doi.org/10.1007/s10750-017-3097-3>.
42. Huang XC, Sua JH, Ouyang JX, Ouyang S, Zhou CH, Wu XP. Towards a global phylogeny of freshwater mussels (Bivalvia: Unionoida): species delimitation of Chinese taxa, mitochondrial phylogenomics, and diversification patterns. *Mol Phylogenet Evol*. 2019;130:45–59. <https://doi.org/10.1016/j.jmpev.2018.09.019>.
43. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39:W29–37. <https://doi.org/10.1093/nar/gkr367>.
44. Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics*. 2004;20(18):3702–4. <https://doi.org/10.1093/bioinformatics/bth444>.
45. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. 2010;5:725–38. <https://doi.org/10.1038/nprot.2010.5>.
46. Pfeiffer JM, Breinholt JW, Page LM. Unioverse: a phylogenomic resource for reconstructing the evolution of freshwater mussels (Bivalvia, Unionoida). *Mol Phylogenet Evol*. 2019;137:114–26. <https://doi.org/10.1016/j.jmpev.2019.02.016>.
47. Parodiz JJ, Bonetto AA. Taxonomy and zoogeographic relationships of the south American naiades (Pelecypoda: Unionacea and Mutelacea). *Malacologia*. 1963;1:179–214.
48. Theologidis I, Fodelianakis S, Gaspar MB, Zouros E. Doubly uniparental inheritance (DUI) of mitochondrial DNA in *Donax trunculus* (Bivalvia: Donacidae) and the problem of its sporadic detection in Bivalvia. *Evolution*. 2008;62:959–70. <https://doi.org/10.1111/j.1558-5646.2008.00329.x>.
49. Chase EE, Robicheau BM, Veinot S, Breton S, Stewart DT. The complete mitochondrial genome of the hermaphroditic freshwater mussel *Anodonta cygnea* (Bivalvia: Unionoidae): *in silico* analyses of sex-specific ORFs across order Unionoida. *BMC Genomics*. 2018;19:221. <https://doi.org/10.1186/s12864-018-4583-3>.
50. Breton S, Capt C, Guerra D, Stewart D. Sex-determining mechanisms in bivalves. In: Leonard JL, editor. *Transitions between sexual systems*. Springer: Cham; 2018. p. 165–92. [https://doi.org/10.1007/978-3-319-94139-4\\_6](https://doi.org/10.1007/978-3-319-94139-4_6).
51. Lopes-Lima M, Burlakova LE, Karatayev AY, Mehler K, Seddon M, Sousa R. Conservation of freshwater bivalves at the global scale: diversity, threats and research needs. *Hydrobiologia*. 2018;810:1–14. <https://doi.org/10.1007/s10750-017-3486-7>.
52. Fonseca MM, Lopes-Lima M, Eackles MS, King TL, Froufe E. The female and male mitochondrial genomes of *Unio delphinus* and the phylogeny of freshwater mussels (Bivalvia: Unionoida). *Mitochondrial DNA B*. 2016;1(1):954–7. <https://doi.org/10.1080/23802359.2016.1241677>.
53. Gan HM, Schultz MB, Austin CM. Integrated shotgun sequencing and bioinformatics pipeline allows ultra-fast mitogenome recovery and confirms substantial gene rearrangements in Australian freshwater crayfishes. *BMC Evol Biol*. 2014;14:19. <https://doi.org/10.1186/1471-2148-14-19>.
54. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. MITOS: improved *de novo* metazoan mitochondrial genome annotation. *Mol Phylogenet Evol*. 2013;69:313–9. <https://doi.org/10.1016/j.jmpev.2012.08.023>.
55. Laslett D, Canbäck B. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics*. 2008;24:172–5. <https://doi.org/10.1093/bioinformatics/btm573>.
56. Di Tommaso P, Moretti S, Xenarios I, Orobítz M, Montanyola A, Chang J-M, Taly J-F, Notredame C. T-coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res*. 2011;39:W13–7. <https://doi.org/10.1093/nar/gkr245>.
57. Wallace IM, O'Sullivan O, Higgins DG, Notredame C. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res*. 2006;34(6):1692–9. <https://doi.org/10.1093/nar/gkl091>.
58. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res*. 2010;38:W7–W13. <https://doi.org/10.1093/nar/gkq291>.
59. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 2000;17:540–52. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>.
60. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol*. 2007;56:564–77. <https://doi.org/10.1080/10635150701472164>.
61. Guindon S, Gascuel O. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Syst Biol*. 2003;52:696–704. <https://doi.org/10.1080/10635150390235520>.
62. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 2012;9(8):772. <https://doi.org/10.1038/nmeth.2109>.
63. Guerra D, Bouvet K, Breton S. Mitochondrial gene order evolution in Mollusca: inference of the ancestral state from the mtDNA of *Chaetopleura apiculata* (Polyplacophora, Chaetopleuridae). *Mol Phylogenet Evol*. 2018;120:233–9. <https://doi.org/10.1016/j.jmpev.2017.12.013>.

64. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans; 2010. p. 1–8.
65. Rambaut A. FigTree v1.4.3. 2016. <http://tree.bio.ed.ac.uk/software/figtree/>. Accessed 23 Sept 2019.
66. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 2000;16:276–7. [https://doi.org/10.1016/S0168-9525\(00\)02024-2](https://doi.org/10.1016/S0168-9525(00)02024-2).
67. Chang J-M, Di Tommaso P, Taly J-F, Notredame C. Accurate multiple sequence alignment of transmembrane proteins with PSI-coffee. *BMC Bioinformatics.* 2012;13(Suppl 4):S1. <https://doi.org/10.1186/1471-2105-13-S4-S1>.
68. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33(7):1870–4. <https://doi.org/10.1093/molbev/msw054>.
69. Alva V, Nam S-Z, Söding J, Lupas AN. The MPI bioinformatics toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res.* 2016;44(W1):W410–5. <https://doi.org/10.1093/nar/gkw348>.
70. Kawabata T. MATRAS: a program for protein 3D structure comparison. *Nucleic Acids Res.* 2003;31(13):3367–9. <https://doi.org/10.1093/nar/gkg581>.
71. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J.* 2011;101(10):2525–34. <https://doi.org/10.1016/j.bpj.2011.10.024>.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

