Caenogastropod mitogenomics

David Osca, José Templado, Rafael Zardoya*

Museo Nacional de Ciencias Naturales (MNCN-CSIC), José Gutiérrez Abascal 2,

28006, Madrid, Spain

Emails: davidosca@mncn.csic.es, templado@mncn.csic.es, rafaz@mncn.csic.es

*Corresponding author

Abstract

In order to further resolve the phylogenetic relationships within Caenogastropoda, the complete mitochondrial (mt) genomes of *Cochlostoma hidalgoi* (Cyclophoroidea), *Naticarius hebraeus* (Naticoidea), *Galeodea echinophora* (Tonnoidea), and *Columbella adansoni* (Buccionidea), and the partial mt genome of *Erosaria spurca* (Cypraeoidea) were sequenced. All newly determined mt genomes conformed to the consensus gene order of caenogastropods, except that of *C. hidalgoi*, which differed in the relative positions of the *trnD*, *trnQ*, *trnG*, *trnY*, and *trnT* genes. Phylogenetic reconstruction of the caenogastropod tree was performed using probabilistic methods and based on the deduced amino acid sequences of concatenated mt protein coding genes. The reconstructed phylogeny recovered Architaenioglossa (superfamilies Cyclophoroidea, Ampullarioidea, and Viviparoidea) as a grade. The monophyly of Sorbeoconcha (all caenogastropods but Architaenioglossa) was supported by most but not all phylogenetic analyses (excluding Vermetoidea, which has a long branch). The relative phylogenetic position of Cerithioidea with respect to Hypsogastropoda remains unresolved. The monophyly of Hypsogastropoda (without Vermetoidea) is strongly supported. Within this clade, Littorinimorpha should be considered a grade. Several superfamilies (Abyssochrysoidea, Risssoidea, Truncatelloidea, and Naticoidea) branched off successively before a siphonate clade (Stromboidea, Cypraeoidea, Tonnoidea, Neogastropoda), which is strongly supported. The relative phylogenetic position of Vermetoidea could not be determined due to long-branch attraction artifacts. The superfamily Tonnoidea was recovered within Neogastropoda, which questions the monophyly of the latter as traditionally defined. We propose to resolve the polyphyly of Muricoidea by erecting two new superfamilies Volutioidea and Babylonioidea.

Key words: mitochondrial genomes; Caenogastropoda; Architaenioglossa; Littorinimorpha; Hypsogastropoda; Neogastropoda

1. Introduction

With thousands of described species, the caenogastropods (periwinkles, whelks, cones, cowries, miters, etc.) constitute the largest and most evolutionary successful group of living gastropods (Ponder et al., 2008). These snails are highly diverse in the morphology of their typically coiled shell, exhibit a wide variety of dietary specializations (detritus feeders, suspension feeders, algal grazers, carnivores, parasites), and are adapted to very different marine, freshwater and terrestrial habitats worldwide (being particularly abundant in the tropics; Colgan et al., 2007). According to the fossil record, the extraordinary diversity of extant caenogastropods (about 140 families according to the most recent comprehensive classification by Bouchet and Rocroi (2005) and the WoRMS database; <http://www.marinespecies.org/>) was generated in successive radiations during the upper Mesozoic and lower Cenozoic (Bandel, 1993).

The clade Caenogastropoda was originally created by Cox (1960) to accommodate most Mesogastropoda and all Stenoglossa of Thiele (1929–35), but current understanding of Caenogastropoda broadly follows the morphological cladistic treatment of the overall gastropod phylogeny by Ponder and Lindberg (1997). Within Caenogastropoda, two higher groups are recognized, Architaenioglossa and Sorbeoconcha (Bouchet and Rocroi, 2005; Colgan et al., 2007; Ponder and Lindberg, 1997; Strong, 2003). The Architaenioglossa include the freshwater Ampullarioidea and Viviparoidea, and the terrestrial Cyclophoroidea. The Sorbeoconcha include the Cerithioidea, the Campaniloidea, and the Hypsogastropoda. This latter group contains the majority of the diversity of caenogastropods, and was provisionally divided into three main groups, Littorinimorpha, Ptenoglossa, and Neogastropoda (= Stenoglossa) (Bouchet and Rocroi, 2005; Ponder and Lindberg, 1997). While the monophyly of the

Caenogastropoda is generally not questioned either with morphology (Bouchet and Rocroi, 2005; Ponder and Lindberg, 1997; Strong, 2003; but see Haszprunar, 1988) or with molecules (Colgan et al., 2007; Harasewych et al., 1998; McArthur and Harasewych, 2003; Osca et al., 2014a; Zapata et al., 2014), the monophyly and phylogenetic relationships of the above mentioned main lineages within the group are under constant debate and revision (Colgan et al., 2007; Ponder et al., 2008). For instance, several morphological studies suggested that Architaenioglossa could represent a grade (Haszprunar, 1988; Ponder et al., 2008; Ponder and Lindberg, 1997) and considered Ptenoglossa non-monophyletic (Ponder and Lindberg, 1997), whereas several molecular studies fail to recover the monophyly of Littorinimorpha and Neogastropoda (Colgan et al., 2007; Cunha et al., 2009).

Morphological phylogenies of Caenogastropoda have been hampered by the difficulty in defining character homology and character states, as well as because numerous characters are prone to convergence (Ponder et al., 2008; see Sigwart and Lindberg, 2014 for a more general pattern affecting mollusc phylogenies). In addition, it has been difficult to make compatible the rich fossil data of the group with phylogenies and classifications based on extant taxa (Ponder et al., 2008). Similarly, molecular studies had their own caveats. While some early molecular phylogenies were based on relatively short sequences (Harasewych et al. 1998), others had a biased representation of the main lineages of Caenogastropoda (Aktipis and Giribet, 2010; Cunha et al., 2009; McArthur and Harasewych, 2003; Williams et al., 2014). Consequently, the reconstructed molecular phylogenies were mainly characterized by low resolution of the deeper nodes (e.g., Colgan et al., 2007) and contradicting results. The Bayesian phylogenetic tree using combined morphological and molecular data that is reported in Ponder et al. (2008) summarizes the state of the art regarding the phylogeny and

systematics of Caneogastropoda. Available data suggests that some of the currently recognized higher taxa (orders, suborders) are probably paraphyletic or even polyphyletic, and the relationships of intermediate groups (superfamilies, families) are unresolved. The monophyly of Architaenioglossa is not recovered whereas the monophyly of Sorbeoconcha is supported with a 0.94 Bayesian posterior probability (BPP; Ponder et al., 2008). Within Sorbeoconcha, the Campaniloidea are recovered as sister group of Cerithioidea and Hypsogastropoda, although with low statistical support. The monophyly of Hypsogastropoda is highly supported (0.98 BPP) and the group is divided into two low supported clades of siphonate (with siphonal canal) and asiphonate (without siphonal canal) snails, respectively (Ponder et al., 2008; Vermeij, 2007). Within the asiphonates (including e.g., Naticoidea, Littorinoidea, Vermetoidea, and Risssooidea), phylogenetic relationships are unresolved whereas within the siphonates, resolution is high, and the following lineages branch off successively: (1) Calyptraeoidae; (2) Stromboidea and Xenophoroidea; (3) Cypraeoidea and Tonnoidea; and (4) Neogastropoda (Ponder et al., 2008). The monophyly of Neogastropoda is supported by several morphological synapomorphies related with the digestive system (Kantor, 1996) but questioned by molecular analyses (Colgan et al., 2007; Cunha et al., 2009).

Complete mitochondrial (mt) genome sequences can be used to reconstruct robust phylogenies if applied at the proper taxonomic level (e.g., San Mauro et al., 2014). Moreover, mt genome arrangements may also contain important phylogenetic information (see e.g., Stöger and Schrödl, 2013 that propose the KARNI cluster of mt tRNAs as molecular synapomorphy of molluscs). When applied to resolve phylogenetic relationships among bilaterians (Bernt et al., 2013) or the main classes of molluscs (Osca et al., 2014a; Stöger and Schrödl, 2013; Schrödl and Stöger, 2014), however, mt

sequences show important problems of rate heterogeneity among lineages, which ultimately limit their phylogenetic utility (systematic biases from strand-specific skews as well as correlation of high rates of gene rearrangements and sequence evolution were proposed as potential causes for inference artifacts; Stöger and Schrödl, 2013; Schrödl and Stöger, 2014). Similarly, the long branches produced by Patellogastropoda (Simison et al., 2006) and Heterobranchia (Grande et al., 2008) mt genomes also hamper the reconstruction of the gastropod phylogenetic tree (Stöger and Schrödl, 2013). However, within the main lineages of gastropods, due to the younger diversification and the rather stable gene arrangements, the use of mt genomes is promising and have rendered in many instances rather resolved phylogenies (Cunha et al., 2009; Grande et al., 2008; Osca et al., 2014b; Williams et al., 2014), which complement those based on nuclear multi-gene sequence data sets (Zapata et al., 2014). Thus far, only two studies utilized complete mt genomes to address phylogenetic relationships within Caenogastropoda, one focused on testing the monophyly of Neogastropoda (Cunha et al., 2009), the other on resolving the relative phylogenetic position of the Abysochrysoidea (Osca et al., 2014b).

Here, we sequenced a total of four complete, and one partial caenogastropod mt genomes (all selected to fill in important gaps in the caenogastropod tree), and compiled all those available (and non-redundant) for the group in GenBank (<http://ncbi.nlm.nih.gov>). For the first time, we reconstructed a phylogeny of Caenogastropoda including a total of 34 mt genomes that represent the main lineages within the group, with the exception of the Campaniloidea and the superfamilies included within the non-monophyletic Ptenoglossa (Table 1). The reconstructed phylogeny was used to test the monophyly of the classical taxonomic groups recognized within the megadiverse Caenogastropoda, to tackle the open questions about internal

phylogenetic relationships, and to investigate the evolution of mt gene rearrangements within the group.

2. Materials and Methods

2.1. DNA extraction, PCR amplification and sequencing

One specimen each of *Cochlostoma hidalgoi* (Cyclophoroidea), *Naticarius hebraeus* (Naticoidea), *Erosaria spurca* (Cypraeoidea), *Galeodea echinophora* (Tonnoidea), and *Columbella adansoni* (Buccinoidea) was used for this study (Table 1). All samples were stored in ethanol 100% and total genomic DNA was isolated from up to 1g of foot tissue following a standard phenol-chloroform extraction. The PCR amplification of the mt genomes followed a three-phase strategy: first, several short fragments per genome were PCR amplified and sequenced using universal primers for *coxI* (Folmer et al., 1994), *rrnL* (Palumbi et al., 1991), *rrnS* (Kocher et al., 1989), *nad4* (Shao et al., 2006), *cox3* and *cob* (Boore and Brown, 2000) and standard PCR conditions (Osca et al., 2014a); second, new primers were designed outwards the short fragments; third, the remaining mtDNA was amplified in 2-3 overlapping fragments by long PCR (see appendix for the specific strategy and long PCR primer sequences for each mt genome). Long PCRs were carried out in 50 µl reactions containing 5 µl of 10x LA Buffer II (Mg⁺² plus), 8 µl of dNTP Mixture (2.5 mM each), 1 µl of each primer, 0.5 µl of template DNA, 0.5 µl of TaKaRa LA Taq DNA polymerase (5 units/µl), and sterilized distilled water up to 50 µl. The following PCR conditions were used: an initial denaturing step at 98°C for 30 s; 45 cycles of denaturing at 98°C for 10 s, annealing at 57°C for 30 s, and extension at 68°C for 60 s per kb; and a final extension step at 68°C for 10 min. Long-PCR products were purified by ethanol precipitation. Long PCR

fragments were sequenced either following a shotgun strategy and Sanger sequencing (Osca et al., 2014a) or directly through massive parallel sequencing (Osca et al., 2014b) in an automated DNA sequencer ABI PRISM 3700 at Secugen (Madrid, Spain) or in an Illumina HiSeq2000 at Macrogen (Seoul, Korea), respectively. Shorter PCR fragments and regions of poor base calling were sequenced using nested PCRs and primer walking (see appendix).

2.2. Genome assembly and annotation

Short-read DNA sequences from Illumina were assembled into a single sequence corresponding to a complete mt genome using CLC Genomics Workbench 5.5 (CLC bio, Aarhus, Denmark). Shotgun clone sequences were assembled into a complete mt genome using Sequencher v. 5.0.1 (Gene Codes Co.; Ann Arbor, MI, USA). Gaps in the final circular sequences were closed using sequences obtained through primer walking. The 13 mt protein-coding genes were annotated by identifying their open reading frames and by comparing them with other reported mollusc mt genomes using the MITOS (Bernt et al., 2013) and DOGMA (Wyman et al., 2004) webserver. The transfer RNA (tRNA) genes were further identified with tRNAscan-SE 1.21 (Schattner et al., 2005) and ARWEN 1.2 (Laslett and Canbäck, 2008), which infer cloverleaf secondary structures. The ribosomal RNA (rRNA) genes were identified by sequence comparison with other reported mollusc mt genomes, and assumed to extend to the boundaries of adjacent genes (Boore et al., 2005). The gene orders of the newly sequenced mt genomes were compared to the gene arrangement of other reported caenogastropod mitochondrial genomes (Table 1), one of Vetigastropoda, *Haliotis rubra* (Maynard et al., 2005), and one of Neritimorpha, *Nerita tessellata* (Arquez et al., 2014).

2.3. Sequence alignment

The mtDNA sequences of the newly sequenced mt genomes were aligned to the orthologous sequences of 34 complete mt genomes of representatives of main caenogastropod lineages (Table 1), as well as the following outgroups: two vetigastropods, *Tegula brunnea* (Simison, 2012; unpublished) and *Lunella* aff. *cinerea* (Williams et al., 2014); two neritimorphs, *Nerita tessellata* (Arquez et al., 2014) and *Nerita melanotragus* (Castro and Colgan, 2010); six heterobranchs, *Aplysia californica* (Knudsen et al., 2006), *Roboastra europaea* (Grande et al., 2002), *Pupa strigosa* (Kurabayashi and Ueshima, 2000), *Albinaria caerulea* (Yamazaki et al., 1997), *Onchidella celtica* and *Pyramidella dolabrata* (Grande et al., 2008); and one chaetodermomorph, *Scutopus ventrolineatus* (Osca et al., 2014a). Derived amino acid sequences from all 13 mt protein-coding genes were downloaded from GenBank using MitoBank (Abascal et al., 2007). Phylogenetic analyses were conducted based on the concatenation of the deduced amino acid sequences of the 13 protein-coding genes (hereafter the mt data set).

In addition, a nuclear data set was constructed by retrieving from GenBank the partial nucleotide sequences of the nuclear genes encoding the rRNAs 18S and 28S (segments d1, d4-d5, d6, and B) and the amino acid sequences of the Histone 3 (H3) gene of representatives of the main lineages of caenogastropods (Colgan et al., 2007; Colgan et al., 2003; Colgan et al., 2000), one neritimorph, two vetigastropods, one patellogastropod, and one chaetodermomorph (Table 2). In order to minimize missing data, in some cases we had to concatenate sequences from species for which there is strong evidence for the monophyly of the higher taxon rank (Table 2). Finally, we concatenated mt and nuclear sequence data into a single data set, hereafter named the combined data set.

In order to construct the different data sets, deduced amino acid sequences of the 13 protein-coding genes (mt data set) were aligned separately using Translator X (Abascal et al., 2010). For the nuclear data set, the amino acid sequences of the H3 gene and the nucleotide sequences of the ribosomal RNA nuclear genes were aligned separately using Translator X (Abascal et al., 2010) and MAFFT (Kato et al., 2002), respectively. Ambiguously aligned positions were removed using Gblocks, v.0.19b (Castresana, 2000) under least stringent settings. Finally, the different single alignments were concatenated.

2.4. Phylogenetic analyses

Alignment format conversions were performed using the ALTER webserver (Glez-Peña et al., 2010). Phylogenetic relationships were inferred using maximum likelihood (ML; (Felsenstein, 1981) and Bayesian inference (BI; (Huelsenbeck and Ronquist, 2001). ML analyses were conducted with RAxML v7.0.4 (Stamatakis, 2006) using the rapid hill-climbing algorithm. BI analyses were conducted using MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003) and running four simultaneous Markov chains for 10 million generations, sampling every 1000 generations, and discarding the first 25% generations as burn-in to prevent sampling before reaching stationarity (as judged by plots of ML scores and low standard deviation of split frequencies). Two independent Bayesian inference runs were performed to increase the chance of adequate mixing of the Markov chains and to increase the chance of detecting failure to converge.

The best partition schemes and best-fit models of substitution for the three data sets were identified using Partition Finder and Partition Finder Protein (Lanfear et al., 2012) with the Akaike information criterion (AIC; Akaike, 1973). For the mt data set

the partitions tested were: all genes combined; all genes separated; all genes separated except *nad4/nad4L* and *atp8/atp6*; and genes grouped by subunits i.e., *cox*, *atp*, *nad* and *cob*. For the nuclear data set the three partitions tested were: all fragments combined, all fragments separated, and fragments grouped by gen type i.e., 18S/28S versus H3. The results of Partition Finder are shown in the Appendix. Support for internal branches was evaluated by non-parametric bootstrapping (Felsenstein, 1985) with 1000 replicates (ML) and by Bayesian posterior probabilities (BI). Maximal, high, moderate, and poor support of a node was defined for ML and BI as 100%, 1; >70%, >0.95; between 50%-70%, 0.90-0.95; and below 50%, 0.90, respectively.

3. Results and discussion

3.1. Mitochondrial genome organization and structural features

The complete mt genomes of *C. hidalgoi*, *N. hebraeus*, *C. adansoni*, and *G. echinophora*, and the partial mt genome of *E. spurca* were sequenced and assembled (see Table 1 for genome lengths and the appendix for the annotations). Like most animal mt genomes (Boore and Brown, 1994), they encode for 13 protein-coding genes, 22 tRNAs, and 2 rRNAs (Fig 1). The major strand of all the newly sequenced mt genomes encodes for all protein-coding and rRNA genes (Fig. 1). This strand also encodes for *trnD*, *trnV*, *trnL(uaa)*, *trnL(uag)*, *trnP*, *trnS(uga)*, *trnH*, *trnF*, the cluster KARNI (*trnK*, *trnA*, *trnR*, *trnN*, *trnI*), and *trnS(gcu)*. The minor stand encodes for the cluster MYCWQGE (*trnM*, *trnY*, *trnC*, *trnW*, *trnQ*, *trnG*, *trnE*) (Fig. 1). The *trnT* can be found encoded in any of the strands depending on the species (Fig. 1). The genes *nad4/ nad4L* overlapped in all mt genomes (unknown in *E. spurca*). In *C. hidalgoi*, *nad6* and *cob* may overlap or alternatively, *nad6* could end with an incomplete T-- stop codon (but this would require specific cleavage and posterior polyadenylation; see

below). Almost all protein-coding genes start their open reading frame with ATG (see appendix). However, *nad1* starts with ATA in *C. hidalgoi*, *G. echinophora*, *C. adansoni*, and *E. spurca*. In addition, *nad4* and *cox3* start with GTG and ATA, respectively, in *N. hebraeus*. The start codons for *nad4l*, *nad4*, *nad5* and *cox3* could not be determined in *E. spurca*. The most common stop codon is TAA (see appendix). The stop codon TAG is used in *atp8*, *nad1*, *nad4*, and *cox3* of *C. hidalgoi*, *nad6* and *nad4l* of *N. hebraeus*, *nad1*, *cob*, and *nad4l* of *G. echinophora*, *cox1* and *nad4l* of *C. adansoni*, and *nad1* and *nad2* of *E. spurca* (the stop codons for *cob*, *nad4l*, *nad4*, and *nad5* could not be determined in this species). The genes *atp6*, *cob*, and *nad5* of *C. hidalgoi*, *cox2* of *G. echinophora*, and *nad1*, *nad4*, and *cox3* of *C. adansoni* end with a single T whereas *nad3* of *N. hebraeus* ends with TA. These incomplete stop codons presumably become functional by subsequent polyadenylation of the transcribed messenger RNAs (Lopez Sanchez et al., 2011). The largest noncoding region was found in all mt genomes upstream to the 5' end of the *cox3* gene (34-249 bp), in a position that has been postulated as candidate to contain the control region in other gastropod mt genomes (Kurabayashi and Ueshima, 2000). This region failed to be amplified in *E. spurca*. A second intergenic region was found downstream the 3' end of the *atp6* gene (7-41 bp).

Gastropod mitochondrial genomes show relatively high rates of gene rearrangement between major lineages (Grande et al., 2008). However, within each major lineage, genome organizations are relatively stable (with rearrangements normally restricted to tRNA genes), and a consensus gene order can be established. The gene orders of the newly sequenced mt genomes were found to conform to the consensus gene order shared by most caenogastropod mt genomes (Fig. 1). This caenogastropod consensus gene order is very similar to the ones reported for

vetigastropods (the closest within gastropods to the ancestral mollusc gene order; see (Osca et al., 2014a) and neritimorphs. It mainly differs from the latter two in the relative position of the MYCWQGE tRNA cluster and the inversion of a fragment that includes the following genes: *trnF*, *nad5*, *trnH*, *nad4*, *nad4L*, *trnT*, *trnS*, *cob*, *nad6*, *trnP*, *nad1*, *trnL(uaa)*, *trnL(uag)*, *rrnL*, *trnV*, and *rrnS*. In addition, the relative positions of the *trnD* and *trnN* are also different in vetigastropods (Fig. 1). Within caenogastropods, minor differences with respect to the consensus gene order are found in *C. hidalgoi*, which differs in the relative positions of *trnD*, *trnQ*, *trnG*, *trnY*, and *trnT*; in *Cipangopaludina cathayensis* (Yang et al., 2014) and *Pomacea canaliculata* (Zhou et al., 2014), which differ in the relative position of the *trnT*; in *Ifremeria nautilei* (Osca et al., 2014b), which differs in the relative position of the *trnW*; in *Oxymeris dimidiata* (Dotson and Beard, 2001), which differs in the relative position of the *trnV*; and in *Fusiturris similis* (Cunha et al., 2009), which differs in the relative position of the *trnS* (Fig. 1). However, most conspicuous gene order rearrangements are found in vermetoids (Rawlings et al., 2010) and cerithioids (Zeng et al., 2014). In the former, many genes have changed their relative position, and only the region between *rrnL* and *nad4* is relatively stable (Fig. 1). The cluster YMCWQGE is maintained (although the relative position of the genes *trnM* and *trnY* have changed with respect to the consensus gene order) but the cluster KARNI is disintegrated (Fig. 1). The instability of vermetoid mt genomes is high, and the different vermetoid species thus far sequenced show rearrangements between them (Rawlings et al., 2010). The only available mt genome of a cerithioid species, *Semisulcospira libertina* (Zeng et al., 2014), shows numerous rearrangements, and even the cluster MYCWQGE is fragmented (Fig. 1).

3.2 Phylogenetic analyses

316 The molecular phylogeny of Caenogastropoda was reconstructed based on the mt,
317 nuclear, and combined data sets using ML and BI (Figs. 2-4). Several species
318 representing other gastropod lineages (Neritimorpha, Vetigastropoda, Heterobranchia),
319 and a chaetodermomorph, *S. ventrolineatus*, were used as outgroup. After removing
320 ambiguously aligned positions, the final mt amino acid matrix contained 3,251 sites.
321 The ML (-lnL= 106317.61) and BI (-lnL= 134169.60 for run1; -lnL= 134169.68 for
322 run2) analyses based on the mt dataset arrived similar tree topologies. All nodes in the
323 BI tree were highly supported whereas several internal nodes in the ML were poorly
324 resolved. Most of the differences between the two trees affected the nodes with low
325 levels of support in the ML phylogeny (Fig. 2). The monophyly of Caenogastropoda
326 received maximal support. The relative phylogenetic position of vermetoids as sister
327 group of the remaining caenogastropods is at odds with morphological evidence, and
328 could be explained due to a long branch attraction (LBA) artifact. As explained above,
329 the vermetoid mt genomes have high rates of gene rearrangement (Rawlings et al.,
330 2010), which are generally associated with high substitution rates (evidenced by the
331 long branches in the reconstructed tree). Hence, their relative position in the mt
332 phylogeny, although highly supported and recovered in previous molecular
333 phylogenetic analyses (Zou et al., 2011), is considered spurious (see Williams et al.,
334 2014) for an even more biased position of vermetoids outside Caenogastropoda based
335 on complete mt genomes). The morphologically, highly derived vermetoids remain as a
336 rather problematic group of uncertain phylogenetic placement. The worm snails were
337 traditionally allied with Cerithioidea (e.g., Houbriek, 1988), but subsequent studies of
338 sperm ultrastructure rejected this relationship (Healy, 1988). Although their close
339 affinities to cerithioideans still persist among some paleontologists (Bandel, 2006;
340 Bandel and Kiel, 2000) and morphologists (Simone, 2011), at present, vermetoids are

commonly positioned within the paraphyletic Littorinimorpha (Bouchet and Rocroi, 2005; Ponder et al., 2008; Strong, 2003), their exact placement remaining elusive.

The monophyly of Architaenioglossa was not recovered regardless of the method of inference (Fig. 2). While the two Ampullarioidea were grouped together, the Cyclophoroidea was grouped with the Viviparoidea, against previous analyses strongly grouping Ampullarioidea and Cyclophoroidea (Ponder et al., 2008). In the ML tree, the Cyclophoroidea+ Viviparoidea were recovered as sister group of Ampullarioidea plus Sorbeoconcha (Cerithioidea + Hypsogastropoda, but without vermetoids) with high support. Both, the sister group relationship of Ampullarioidea to Sorbeoconcha (without vermetoids) and the monophyly of the latter were poorly supported (Fig. 2). In the BI tree, the Ampullarioidea were the first lineage branching off (in contradiction with the ML analysis). Next, a clade including the Cyclophoroidea+ Viviparoidea and the Cerithioidea was recovered as sister group of the Hypsogastropoda rejecting the monophyly of Sorbeoconcha (regardless of the odd position of vermetoids) (Fig. 2). Architaenioglossa has long been recognized as an assemblage of terrestrial and freshwater lineages forming the earliest two or three branches in the caenogastropod tree (Ponder and Lindberg, 1997). Its potential monophyly has long been questioned due to the absence of well-defined synapomorphies other than the non-marine habitat, and in fact the group is now considered a grade (Ponder et al., 2008). Both morphological (Simone, 2011) and molecular (Colgan et al., 2007; Harasewych et al., 1998; this work) data confirm the non-monophyly of Architaenioglossa, but the relationships of the three superfamilies with each other and with Campaniloidea and Cerithioidea remain controversial. Moreover, in some studies the architaenioglossan superfamilies are recovered as sister group of Campaniloidea (Colgan et al., 2007) or Cerithioidea (this work, but note that we could not include the Campaniloidea in our

phylogenetic analyses) also rejecting the monophyly of Sorbeoconcha. In contrast, a Bayesian inference based on a combined morphological and molecular data set recovered Sorbeoconcha as monophyletic with Campaniloidea as sister group of Cerithioidea + Hypsogastropoda (Ponder et al., 2008). In a recent phylogenetic analysis based on transcriptome data, the Ampullarioidea (represented by *Pomacea*) were placed as sister group to monophyletic Sorbeoconcha, and within this latter group, Cerithioidea was recovered as sister group to Hypsogastropoda (Zapata et al., 2014).

The Hypsogastropoda were monophyletic (without taking into account vermetoids) and highly supported in both of our analyses as has often been the case in other molecular phylogenetic analyses (Colgan et al., 2007; Harasewych et al., 1998; McArthur and Harasewych, 2003; Ponder et al., 2008; Zapata et al., 2014). Internal phylogenetic relationships within Hypsogastropoda are difficult to resolve likely due to an early radiation event in the evolutionary history of the group (Colgan et al., 2007). The traditional view of three provisional subgroups i.e., Littorinimorpha, Ptenoglossa, and Neogastropoda (Bouchet and Rocroi, 2005) is contrasted by the most recent division of hypsogastropodan superfamilies into the asiphonate and siphonate clades (Ponder et al., 2008).

Within Hypsogastropoda, our results show the Abyssochrysoidea as sister group of the remaining taxa with poor and maximal support in the ML and BI analyses, respectively (Fig. 2). Previous phylogenetic analyses nested this superfamily within Littorinimorpha (Osca et al., 2014b). A study based on morphology placed Abyssochrysoidea as sister group of Littorinidae but without statistical support (Ponder et al., 2008) whereas a molecular phylogeny based on partial mt and nuclear loci recovered Abyssochrysoidea in a clade with Cingulopsoidea and Triphoroidea also without statistical support (Takano and Kano, 2014). A previous study based on

complete mt genomes but with less number of taxa recovered with high statistical support the Abysochrysoidea right after the branching off of Rissooidea and as sister group of Tonnoidea + Neogastropoda (Osca et al., 2014b).

In the next node of our reconstructed tree, the clade Rissooidea + Truncatelloidea was maximally supported in both ML and BI, and recovered as sister group of Naticoidea + the siphonate clade, although highly supported only in BI. Altogether, the recovered trees failed to support the monophyly of Littorinimorpha (Fig. 2) reinforcing this view of caenogastropod relationships (Colgan et al., 2007; Osca et al., 2014b; Ponder et al., 2008; Simone, 2011). No mt genome of any member of the Ptenoglossa has been sequenced thus far, and therefore this group could not be included in the present study. The Ptenoglossa, currently confined to Eulimoidea, Epitonioidea (= Janthinoidea) and Triphoroidea, was found to be paraphyletic (Ponder et al., 2008; Ponder and Lindberg, 1997; Takano and Kano, 2014) and treated as an informal group in the working classification by (Bouchet and Rocroi, 2005). The monophyly of the asiphonate group (Ponder et al., 2008) was also rejected.

The siphonate clade (Ponder et al., 2008) was highly supported in both ML and BI analyses. Members of this clade are characterized by having evolved an anterior siphon, often long and mobile, that greatly increases the osphradial surface, further enhancing the detection at a distance of stimuli (location of predators, food and mates), and allowing discerning the direction from which they are emanating (Lindberg and Ponder, 2001). The ML tree has Stromboidea and Cypraeoidea together with moderate support, and as sister group of the remaining members of the siphonate clade. In the BI tree, Stromboidea is the first lineage branching off, and the Cypraeoidea is recovered as sister group of the remaining members of the siphonate clade with moderate support (Fig. 2). The limpet shaped Calyptraeidea and the tent-shaped Xenophoroidea, which

416 have been previously associated with the siphonate clade (Colgan et al., 2007; Ponder et
417 al., 2008), were not included in the present study since no complete mt genomes are
418 available for these superfamilies.

419 The monophyly of Neogastropoda is not recovered regardless of the method of
420 inference, unless the littorinimorph Tonnoidea are considered part of the group (Fig. 2).
421 In the ML analysis, the neogastropod Cancellarioidea are recovered as sister group of
422 Tonnoidea and the remaining Neogastropoda, although with low bootstrap support (Fig.
423 2). While morphological analyses (Ponder et al., 2008; Simone, 2011; Strong, 2003)
424 recover Neogastropoda as a monophyletic group, molecular phylogenies (e.g., Colgan et
425 al., 2007; Cunha et al., 2009; Hayashi, 2005; Williams et al., 2014) normally include
426 tonnoideans within neogastropods (see Zou et al. (2011) for an exception). The early
427 divergence of Cancellarioidea within Neogastropoda was already suggested in some
428 previous morphological (Kantor, 1996) and molecular (Oliverio and Modica, 2010)
429 studies but its position basal to tonnoideans was only noted in previous studies also
430 based on entire mitochondrial genomes (Cunha et al., 2009; Williams et al., 2014).
431 Therefore, there is an open conflict between morphology and molecules on the
432 monophyly of Neogastropoda that awaits the inclusion of more hypsgastropodan
433 lineages in future molecular phylogenies and the analysis of larger sequence data sets.

434 Current classifications of Neogastropoda generally recognize up to six
435 superfamilies: Buccinoidea, Muricoidea, Olivoidea, Pseudolivoidea, Conoidea, and
436 Cancellarioidea (Bouchet and Rocroi, 2005; WoRMS). Nevertheless, the phylogenetic
437 relationships among neogastropod families remain rather unstable and the monophyly
438 of some of them questionable. In our BI tree, the Cancellarioidea and one neogastropod
439 Muricoidea (*Cymbium olla*, Volutidae) are recovered as sister group of Tonnoidea and
440 the remaining Neogastropoda with maximal support (Fig. 2). Within the remaining

Neogastropoda, the following lineages branch off successively: Buccinoidea, Muricoidea (except families Volutidae and Babyloniidae), Olivoidea, Babyloniidae and Conoidea (Fig. 2). The relative positions of the families Babyloniidae and Volutidae in the phylogenies (strongly supported for both in BI; only for Volutidae in ML) reject the monophyly of Muricoidea (Williams et al., 2014), and support the erection of two new superfamilies, Babylonioidea and Volutioidea, respectively. While Babyloniidae is a rather small family with only two genera (*Acinopsis* and *Babylonia*), the family Volutidae is highly diverse with about 40 recognized genera, and thus more data on other species of Volutidae are needed to further delineate this superfamily. A closer phylogenetic relationship of Volutidae (*Melo*) with Cancellarioidea and Conoidea than with Muricoidea was also obtained based on the entire 18S and partial nuclear H3, and mt *cox1*, *rrnL*, and *rrnS* (Zou et al., 2011). The taxonomic classification of Babyloniidae has been controversial. It was traditionally assigned to the superfamily Buccinoidea and, more recently included in Muricoidea (Bouchet and Rocroi, 2005). Phylogenetic analyses support close affinities of *Babylonia* to Volutidae and Olividae (Harasewych and Kantor, 2002) or to Turbinellidae (Zou et al., 2011) based on morphology and molecules, respectively.

The nuclear data set was analyzed at the nucleotide level for the partial 18S and 28S genes, and at the amino acid level for the H3 gene. After removing ambiguously aligned positions, the final matrix contained 1737 sites. The ML (-lnL= 7154.50) and BI (-lnL= 7319.54 for run1; -lnL= 7325.52 for run2) analyses based on the nuclear dataset arrived at similar tree topologies, and differences were concentrated on deeper nodes with low statistical support (Fig. 3). While the mt-based phylogeny recovered Neritimorpha + Vetigastropoda as sister group of Caenogastropoda (Fig. 2), the nuclear phylogeny placed together Heterobranchia and Caenogastropoda (Fig. 3) in support of

the traditional morphological hypothesis termed Apogastropoda (Ponder and Lindberg, 1997; Zapata et al., 2014). This discrepancy may be due to an LBA artifact associated to the long branches exhibited by heterobranch mt genomes (Grande et al. 2008; Stöger and Schrödl, 2013). The monophyly of Caenogastropoda was strongly supported both in ML and BI analyses whereas that of Architaenioglossa was not recovered (Fig. 3). The Ampullarioidea were placed as the sister group of the remaining Caenogastropoda with moderate (ML) to high (BI) statistical support (Fig. 3) in agreement with the BI tree inferred based on the mt data set but against the grouping of Ampullarioidea and Cyclophoroidea reported in previous phylogenetic analyses based on combined morphological and molecular data (Ponder et al., 2008). The monophyly of both Sorbeoconcha and Hypsogastropoda was supported but whether Cyclophoroidea or Viviparoidea are the sister group of Sorbeoconcha was unresolved (Fig. 3). Phylogenetic relationships of major lineages within Hypsogastropoda were largely unresolved (Fig. 3). This is not surprising since the available nuclear markers for the study are known to have relatively slow evolutionary rates, and the diversity of Hypsogastropoda has been proposed to arise from a rapid early radiation after the Permian–Triassic extinction (Colgan et al., 2007). Interestingly, vermetoids were clearly recovered within Hypsogastropoda (Fig. 3), indicating that the reported higher evolutionary rates associated to this group are restricted exclusively to the mt genome (Rawlings et al., 2010).

After removing ambiguously aligned positions, the final combined matrix contained 5165 (amino acid + nucleotide) sites. The ML (-lnL= 87130.89) and BI (-lnL= 108265.42 for run1; -lnL= 108265.64 for run2) analyses based on the combined dataset arrived at similar tree topologies, and differences were concentrated on deeper nodes with low statistical support (Fig. 4). Overall, the combined tree mostly reflected

the mt-based topology and showed similar levels of statistical support of the nodes, and thus of resolution (Fig. 4). The clade Neritimorpha + Vetigastropoda was the sister group of a highly supported Caenogastropoda clade (Fig. 4). Neither Architaenioglossa nor Sorbeoconcha were recovered whereas Hypsogastropoda (without vermetoids, which due to their long branch are recovered as sister group of the remaining caenogastropods) received high statistical support regardless of the method of phylogenetic inference (Fig. 4). Within Hypsogastropoda, the phylogenetic relationships of Abyssochrysoidea, Naticoidea, Rissooidea + Truncatelloidea, and the siphonate clade were unresolved in the ML tree but received maximal support in the BI tree, with Rissooidea + Truncatelloidea as sister group of a clade including Abyssochrysoidea + Naticoidea and the siphonates (Fig. 4). The relative position of Abyssochrysoidea in the combined tree is in agreement with a previous phylogeny based on fewer complete mt genomes (Osca et al., 2014b). Within siphonates, Cypraeoidea + Stromboidea are highly supported as the sister group of the remaining taxa regardless of the method of inference (Colgan et al., 2007; Zou et al., 2011). Cancellarioidea in the ML tree or Cancellarioidea + Volutidae in the BI tree are the next lineages branching off (Fig. 4). The relative phylogenetic position of Tonnoidea rejects the monophyly of Neogastropoda (Cunha et al., 2009). Phylogenetic relationships among Buccinoidea, Muricoidea (Babyloniidae), Muricoidea (Muricidae), Olivoidea, and Conoidea were unresolved (Fig. 4). As in the mt tree, the polyphyly of Muricoidea prompts for a major reorganization of this superfamily (Zou et al., 2011).

Overall, our results support the monophyly of Caenogastropoda, Hypsogastropoda (without Vermetoidea), and the siphonate clade. The taxonomic validity of Architaenioglossa (all analyses), Sorbeoconcha (BI with mt data, ML with combined data), Littorinimorpha (all analyses), Neogastropoda (all analyses), and Muricoidea (all

analyses) is seriously questioned. According to our results, Architaenioglossa should be considered a grade and paraphyletic with respect to Sorbeoconcha (or Hypsogastropoda depending on whether the monophyly of Sorbeoconcha is confirmed in further phylogenetic analyses). The relative phylogenetic positions of Cyclophoroidea, Ampullarioidea, Viviparoidea, and Cerithioidea remain unresolved (as well as that of Campaniloidea not included in the present phylogenetic analyses). Within Hypsogastropoda, Littorinimorpha should be considered a grade and paraphyletic with respect to Neogastropoda (including Tonnoidea). Other important superfamilies formerly included within Littorinimorpha such as e.g. Calyptraeidea, Littorinoidea, Xenophoroidea, Pterotracheoidea and Capuloidea need to be included in the phylogenetic analyses to further resolve this section of the caenogastropod tree, and to test whether any other of these superfamilies may show closer affinities to Neogastropoda. Moreover, the resolution of the relative phylogenetic position of the superfamilies Epitonioidea, Eulimoidea, and Triphoroidea (formerly included in Ptenoglossa) also await future inclusion of these taxa in molecular phylogenetic studies. For Neogastropoda, in order to reconcile taxonomy and the molecular phylogeny (if this proofs finally to be correct), the group could be redefined either including Tonnoidea (that would have secondarily lost the synapomorphies of Neogastropoda) or excluding Cancellarioidea and possibly Volutidae, which would indicate that the synapomorphies defining Neogastropoda would be instead homoplasies. In order to resolve the polyphyly of Muricoidea as traditionally defined, we propose to restrict Muricoidea to the family Muricidae and at least erect two new superfamilies Volutoidae and Babylonioidea. The phylogenetic position and taxonomic classification of other muricoid families (e.g. Harpidae, Mitridae, Turbinellidae) would await further phylogenetic studies.

541

542 **4. Acknowledgements**

543 We are indebted to Gilberto Diosdado and Ivan Acevedo for the gift of a specimen of *N.*
544 *hebraeus* and of *C. adansoni*, respectively. We thank three anonymous reviewers for
545 their insightful comments on a previous version of the manuscript. We are grateful also
546 to Lourdes Alcaraz and Silvia Perea for their help in completing the *C. hidalgoi* mt
547 genome and to Juan Esteban Uribe for helping in the amplification and sequencing of
548 the control region of *C. adansoni*. We are grateful to Jesús Marco and Luis Cabellos
549 who provided access to the supercomputer Altamira at the Institute of Physics of
550 Cantabria (IFCA-CSIC), member of the Spanish Supercomputing Network, for
551 performing phylogenetic analyses. We thank Etienne Kornobis for helping with the
552 assembly programs and introducing us to the Altamira environment. This work was
553 supported by the Spanish Ministry of Science and Innovation (CGL2010-18216 and
554 CGL2013-45211-C2-2-P to RZ; BES-2008-009562 to DO).

5. References

- Abascal, F., Zardoya, R., Telford, M.J., 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* 38, W7-13.
- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle In: Petrov, B.N., Csaki, F. (Eds.), 2nd International Symposium on Information Theory Akademiai Kiado, Budapest, Hungary, pp. 267-281.
- Aktipis, S.W., Giribet, G., 2010. A phylogeny of Vetigastropoda and other “archaeogastropods”: re-organizing old gastropod clades. *Invertebr. Biol.* 129, 220-240.
- Arquez, M., Colgan, D., Castro, L.R., 2014. Sequence and comparison of mitochondrial genomes in the genus *Nerita* (Gastropoda: Neritimorpha: Neritidae) and phylogenetic considerations among gastropods. *Mar. Genom.* 15, 45-54.
- Bandel, K., 1993. Caenogastropoda during Mesozoic times. *Scripta Geol.* 2, 7-56.
- Bandel, K., 2006. Families of the Cerithioidea and related superfamilies (Palaeo-Caenogastropoda; Mollusca) from the Triassic to the Recent characterized by protoconch morphology - including the description of new taxa. *Freiberger Forschung. C* 511, 59-13.
- Bandel, K., Kiel, S., 2000. Earliest known (Campanian) members of the Vermetidae, Provannidae and Litiopidae (Cerithioidea, Gastropoda), and a discussion of their possible relationships. *Mitt. Geol.-Paläont. Inst. Univ. Hamburg* 84, 209-218.
- Bandyopadhyay, P.K., Stevenson, B.J., Cady, M.T., Olivera, B.M., Wolstenholme, D.R., 2006. Complete mitochondrial DNA sequence of a Conoidean gastropod, *Lophiotoma (Xenuroturris) cerithiformis*: Gene order and gastropod phylogeny. *Toxicon* 48, 29-43.

580 Bandyopadhyay, P.K., Stevenson, B.J., Ownby, J.-P., Cady, M.T., Watkins, M.,
 581 Olivera, B.M., 2008. The mitochondrial genome of *Conus textile*, *coxI*–*coxII*
 582 intergenic sequences and Conoidean evolution. *Mol. Phylogen. Evol.* 46, 215-223.
 583 Bernt, M., Bleidorn, C., Braband, A., Dambach, J., Donath, A., Fritzsche, G., Golombek,
 584 A., Hadrys, H., Jühling, F., Meusemann, K., Middendorf, M., Misof, B., Perseke,
 585 M., Podsiadlowski, L., von Reumont, B., Schierwater, B., Schlegel, M., Schrödl,
 586 M., Simon, S., Stadler, P.F., Stöger, I., Struck, T.H., 2013. A comprehensive
 587 analysis of bilaterian mitochondrial genomes and phylogeny. *Mol. Phylogen. Evol.*
 588 69, 352-364.
 589 Boore, J., Brown, W., 1994. Complete DNA sequence of the mitochondrial genome of
 590 the black chiton, *Katharina tunicata*. *Genetics* 138, 423 - 443.
 591 Boore, J.L., Brown, W.M., 2000. Mitochondrial Genomes of *Galathea*,
 592 *Helobdella*, and *Platynereis*: Sequence and Gene Arrangement Comparisons
 593 Indicate that Pogonophora Is Not a Phylum and Annelida and Arthropoda Are Not
 594 Sister Taxa. *Mol. Biol. Evol.* 17, 87-106.
 595 Bouchet, P., Rocroi, J.P., 2005. Classification and Nomenclator of Gastropod Families.
 596 *Malacologia* 47, 397.
 597 Brauer, A., Kurz, A., Stockwell, T., Baden-Tillson, H., Heidler, J., Wittig, I.,
 598 Kauferstein, S., Mebs, D., Stöcklin, R., Remm, M., 2012. The Mitochondrial
 599 Genome of the Venomous Cone Snail *Conus consors*. *PLoS ONE*
 600 7, e51528.
 601 Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their
 602 use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540-552.

603 Castro, L.R., Colgan, D.J., 2010. The phylogenetic position of Neritimorpha based on
604 the mitochondrial genome of *Nerita melanotragus* (Mollusca: Gastropoda). Mol.
605 Phylogen. Evol. 57, 918-923.

606 Colgan, D.J., Ponder, W.F., Beacham, E., Macaranas, J., 2007. Molecular phylogenetics
607 of Caenogastropoda (Gastropoda: Mollusca). Mol. Phylogen. Evol. 42, 717-737.

608 Colgan, D.J., Ponder, W.F., Beacham, E., Macaranas, J.M., 2003. Gastropod phylogeny
609 based on six segments from four genes representing coding or non-coding and
610 mitochondrial or nuclear DNA. Molluscan Res. 23, 123-148.

611 Colgan, D.J., Ponder, W.F., Eggler, P.E., 2000. Gastropod evolutionary rates and
612 phylogenetic relationships assessed using partial 28S rDNA and histone H3
613 sequences. Zool. Scr. 29, 29-63.

614 Cox, L.R., 1960. Thoughts on the classification of the Gastropoda. J. Molluscan Stud.
615 33, 239-261.

616 Cunha, R., Grande, C., Zardoya, R., 2009. Neogastropod phylogenetic relationships
617 based on entire mitochondrial genomes. BMC Evol. Biol. 9, 210.

618 Dotson, E.M., Beard, C.B., 2001. Sequence and organization of the mitochondrial
619 genome of the Chagas disease vector, *Triatoma dimidiata*. Insect Mol. Biol. 10,
620 205-215.

621 Felsenstein, J., 1981. Evolutionary trees from DNA sequences: A maximum likelihood
622 approach. J. Mol. Evol. 17, 368-376.

623 Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the
624 bootstrap. Evolution 39, 783-791.

625 Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for
626 amplification of mitochondrial cytochrome c oxidase subunit I from diverse
627 metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3, 294-299.

628 Glez-Peña, D., Gómez-Blanco, D., Reboiro-Jato, M., Fdez-Riverola, F., Posada, D.,
629 2010. ALTER: program-oriented conversion of DNA and protein alignments.
630 Nucleic Acids Res. 38, W14-W18.

631 Grande, C., Templado, J., Lucas Cervera, J., Zardoya, R., 2002. The Complete
632 Mitochondrial Genome of the Nudibranch *Roboastra europaea* (Mollusca:
633 Gastropoda) Supports the Monophyly of Opisthobranchs. Mol. Biol. Evol. 19,
634 1672-1685.

635 Grande, C., Templado, J., Zardoya, R., 2008. Evolution of gastropod mitochondrial
636 genome arrangements. BMC Evol. Biol. 8, 61.

637 Harasewych, M.G., Adamkewicz, S.L., Plassmeyer, M., Gillevet, P.M., 1998.
638 Phylogenetic relationships of the lower Caenogastropoda (Mollusca, Gastropoda,
639 Architaenioglossa, Campaniloidea, Cerithioidea) as determined by partial 18s
640 rDNA sequences. Zool. Scr. 27, 361-372.

641 Harasewych, M.G., Kantor, Y.I., 2002. On the morphology and taxonomic position of
642 *Babylonia* (Neogastropoda: Babyloniidae). . Bolletino Malacologico Supp. 4, 19-
643 36.

644 Haszprunar, G., 1988. On the origin and evolution of major gastropod groups, with
645 special reference to the Streptoneura. J. Molluscan Stud. 54, 367-441.

646 Hayashi, S., 2005. The molecular phylogeny of the Buccinidae (Caenogastropoda:
647 Neogastropoda) as inferred from the complete mitochondrial 16S rRNA gene
648 sequences of selected representatives. Molluscan Res. 25, 85-98.

649 Healy, J.M., 1988. Sperm morphology in *Serpulorbis* and *Dendropoma* and its
650 relevance to the systematic position of the Vermetidae (Gastropoda). J. Molluscan
651 Stud. 54, 295-308.

652 Houbrick, R.S., 1988. Cerithioidean phylogeny. Malacol Rev supplement 4, 88–128.

653 Huelsenbeck, J., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic
 654 trees. *Bioinformatics* 17, 754-755.

655 Kantor, Y., 1996. Phylogeny and relationships of Neogastropoda. In: Taylor, J. (Ed.),
 656 Origin and evolutionary radiation of the Mollusca, pp. 221–230.

657 Katoh, K., Misawa, K., Kuma, K.i., Miyata, T., 2002. MAFFT: a novel method for
 658 rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids*
 659 *Res.* 30, 3059-3066.

660 Ki, J.-S., Lee, Y.-M., Jung, S.-O., Horiguchi, T., Cho, H.-S., Lee, J.-S., 2010.
 661 Mitochondrial genome of *Thais clavigera* (Mollusca: Gastropoda): Affirmation of
 662 the conserved, ancestral gene pattern within the mollusks. *Mol. Phylogen. Evol.* 54,
 663 1016-1020.

664 Knudsen, B., Kohn, A.B., Nahir, B., McFadden, C.S., Moroz, L.L., 2006. Complete
 665 DNA sequence of the mitochondrial genome of the sea-slug, *Aplysia californica*:
 666 Conservation of the gene order in Euthyneura. *Mol. Phylogen. Evol.* 38, 459-469.

667 Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X.,
 668 Wilson, A.C., 1989. Dynamics of mitochondrial DNA evolution in animals:
 669 amplification and sequencing with conserved primers. *Proc. Natl.l Acad. Sci. USA*
 670 86, 6196-6200.

671 Kurabayashi, A., Ueshima, R., 2000. Complete Sequence of the Mitochondrial DNA of
 672 the Primitive Opisthobranch Gastropod *Pupa strigosa*: Systematic Implication of
 673 the Genome Organization. *Mol. Biol. Evol.* 17, 266-277.

674 Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: Combined
 675 Selection of Partitioning Schemes and Substitution Models for Phylogenetic
 676 Analyses. *Mol. Biol. Evol.* 29, 1695-1701.

677 Lindberg, D.R., Ponder, W.F., 2001. The influence of classification on the evolutionary
 678 interpretation of structure a re-evaluation of the evolution of the pallial cavity of
 679 gastropod molluscs. *Org Divers Evol* 1, 273-299.

680 Lopez Sanchez, M.I.G., Mercer, T.R., Davies, S.M.K., Shearwood, A.-M.J., Nygård,
 681 K.K.A., Richman, T.R., Mattick, J.S., Rackham, O., Filipovska, A., 2011. RNA
 682 processing in human mitochondria. *Cell Cycle* 10, 2904-2916.

683 Márquez, E.J., Castro, E.R., Alzate, J.F., 2014. Mitochondrial genome of the
 684 endangered marine gastropod *Strombus gigas* Linnaeus, 1758 (Mollusca:
 685 Gastropoda). *Mitochondrial DNA*, 1-2.

686 Maynard, B., Kerr, L., McKiernan, J., Jansen, E., Hanna, P., 2005. Mitochondrial DNA
 687 Sequence and Gene Organization in the Australian Blacklip Abalone *Haliotis rubra*
 688 (Leach). *Mar. Biotechnol.* 7, 645-658.

689 McArthur, A., Harasewych, M., 2003. Molecular systematics of the major lineages of
 690 the Gastropoda. In: Lydeard, C., Lindberg, D.R. (Eds.), *Molecular Systematics and*
 691 *Phylogeography of Mollusks*. Smithsonian Books, Washington DC, pp. 140-160.

692 McComish, B.J., Hills, S.F.K., Biggs, P.J., Penny, D., 2010. Index-Free De Novo
 693 Assembly and Deconvolution of Mixed Mitochondrial Genomes. *Genome Biology*
 694 and *Evolution* 2, 410-424.

695 Neiman, M., Hehman, G., Miller, J.T., Logsdon, J.M., Taylor, D.R., 2010. Accelerated
 696 Mutation Accumulation in Asexual Lineages of a Freshwater Snail. *Mol. Biol.*
 697 *Evol.* 27, 954-963.

698 Núñez-Acuña, G., Aguilar-Espinoza, A., Gallardo-Escárate, C., 2013. Complete
 699 mitochondrial genome of *Concholepas concholepas* inferred by 454
 700 pyrosequencing and mtDNA expression in two mollusc populations. *Comparative*
 701 *Biochemistry and Physiology Part D: Genomics and Proteomics* 8, 17-23.

702 Oliverio, M., Modica, M.V., 2010. Relationships of the haematophagous marine snail
 703 *Colubraria* (Rachiglossa: Colubrariidae), within the neogastropod phylogenetic
 704 framework. Zool. J. Linn. Soc. 158, 779-800.

705 Osca, D., Irisarri, I., Todt, C., Grande, C., Zardoya, R., 2014a. The complete
 706 mitochondrial genome of *Scutopus ventrolineatus* (Mollusca: Chaetodermomorpha)
 707 supports the Aculifera hypothesis. BMC Evol. Biol. 14, 1-10.

708 Osca, D., Templado, J., Zardoya, R., 2014b. The mitochondrial genome of *Ifremeria*
 709 *nautiliei* and the phylogenetic position of the enigmatic deep-sea Abyssochrysoidea
 710 (Mollusca: Gastropoda). Gene 547, 257-266.

711 Palumbi, S., Martin A, McMillan, W., Stice, L., Grabowski, G., 1991. The simple fool's
 712 guide to PCR. <http://palumbi.stanford.edu/SimpleFoolsMaster.pdf>, 1-45.

713 Ponder, W.F., Colgan, D.J., Healy, J.M., Nützel, A., Simone, L.R.L., Strong, E.E.,
 714 2008. Caenogastropoda. In: Ponder, W.F., Lindberg, D.L. (Eds.), Molluscan
 715 Phylogeny and Evolution. University of California Press, Berkeley, pp. 331-383.

716 Ponder, W.F., Lindberg, D.R., 1997. Towards a phylogeny of gastropod molluscs: an
 717 analysis using morphological characters. Zool. J. Linn. Soc. 119, 83-265.

718 Rawlings, T., MacInnis, M., Bieler, R., Boore, J., Collins, T., 2010. Sessile snails,
 719 dynamic genomes: gene rearrangements within the mitochondrial genome of a
 720 family of caenogastropod molluscs. BMC Genomics 11, 440.

721 Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference
 722 under mixed models. Bioinformatics 19, 1572-1574.

723 San Mauro, D., Gower, D.J., Müller, H., Loader, S.P., Zardoya, R., Nussbaum, R.A.,
 724 Wilkinson, M., 2014. Life-history evolution and mitogenomic phylogeny of
 725 caecilian amphibians. Mol. Phylogen. Evol. 73, 177-189.

726 Shao, Z., Graf, S., Chaga, O.Y., Lavrov, D.V., 2006. Mitochondrial genome of the
727 moon jelly *Aurelia aurita* (Cnidaria, Scyphozoa): A linear DNA molecule encoding
728 a putative DNA-dependent DNA polymerase. *Gene* 381, 92-101.

729 Sigwart, J.D., Lindberg, D.R., 2014. Consensus and Confusion in Molluscan Trees:
730 Evaluating Morphological and Molecular Phylogenies. *Syst. Biol.* Epub, Dec2.

731 Simison, W., Lindberg, D., Boore, J., 2006. Rolling circle amplification of metazoan
732 mitochondrial genomes. *Mol Phylogenet Evol* 39, 562 - 567.

733 Simone, L.R.L., 2011. Phylogeny of the Caenogastropoda (Mollusca), based on
734 comparative morphology. *Arquivos de Zoologia* 42, 161-323.

735 Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic
736 analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688-2690.

737 Stöger, I., Schrödl, M., 2013. Mitogenomics does not resolve deep molluscan
738 relationships (yet?). *Mol Phylogenet Evol* 69, 376-392.

739 Strong, E.E., 2003. Refining molluscan characters: morphology, character coding and a
740 phylogeny of the Caenogastropoda. *Zool. J. Linn. Soc.* 137, 447-554.

741 Takano, T., Kano, Y., 2014. Molecular phylogenetic investigations of the relationships
742 of the echinoderm-parasite family Eulimidae within Hypsogastropoda (Mollusca).
743 *Mol. Phylogen. Evol.* 79, 258-269.

744 Thiele, J., 1929–35. *Handbuch der Systematischen Weichtierkunde*. Gustav Fischer,
745 Jena.

746 Vermeij G.J. 2007. The ecology of invasion: acquisition and loss of the siphonal canal
747 in gastropods. *Palaeobiol.* 33, 469-493.

748 Wang, M., Qiu, J.-W., 2014. Complete mitochondrial genome of the giant ramshorn
749 snail *Marisa cornuarietis* (Gastropoda: Ampullariidae). *Mitochondrial DNA*, 1-2.

750 Williams, S.T., Foster, P.G., Littlewood, D.T.J., 2014. The complete mitochondrial
 751 genome of a turbinid vetigastropod from MiSeq Illumina sequencing of genomic
 752 DNA and steps towards a resolved gastropod phylogeny. *Gene* 533, 38-47.
 753 Yamazaki, N., Ueshima, R., Terrett, J.A., Yokobori, S., Kaifu, M., Segawa, R.,
 754 Kobayashi, T., Numachi, K., Ueda, T., Nishikawa, K., Watanabe, K., Thomas, R.H.
 755 1997. Evolution of pulmonate gastropod mitochondrial genomes: comparisons of
 756 gene organizations of *Euhadra*, *Cepaea* and *Albinaria* and implications of unusual
 757 tRNA secondary structures. *Genetics* 145, 749-758.
 758 Yang, H., Zhang, J.-e., Luo, H., Luo, M., Guo, J., Deng, Z., Zhao, B., 2014. The
 759 complete mitochondrial genome of the mudsnail *Cipangopaludina cathayensis*
 760 (Gastropoda: Viviparidae). *Mitochondrial DNA*, 1-3.
 761 Zapata, F., Wilson, N.G., Howison, M., Andrade, S.C.S., Jörger, K.M., Schrödl, M.,
 762 Goetz, F.E., Giribet, G., Dunn, C.W., 2014. Phylogenomic analyses of deep
 763 gastropod relationships reject Orthogastropoda. *Proc. R. Soc. Lond. B* 281.
 764 Zeng, T., Yin, W., Xia, R., Fu, C., Jin, B., 2014. Complete mitochondrial genome of a
 765 freshwater snail, *Semisulcospira libertina* (Cerithioidea: Semisulcospiridae).
 766 *Mitochondrial DNA*, 1-2.
 767 Zhou, X., Chen, Y., Zhu, S., Xu, H., Liu, Y., Chen, L., 2014. The complete
 768 mitochondrial genome of *Pomacea canaliculata* (Gastropoda: Ampullariidae).
 769 *Mitochondrial DNA*, 1-2.
 770 Zou, S., Li, Q., Kong, L., 2011. Additional gene data and increased sampling give new
 771 insights into the phylogenetic relationships of Neogastropoda, within the
 772 caenogastropod phylogenetic framework. *Mol. Phylogen. Evol.* 61, 425-435.
 773

Legends to Figures

Fig. 1. Mitochondrial gene orders of main lineages of Caenogastropoda. One Neritimorpha (*Nerita tessellata*) and one Vetigastropoda (*Haliotis rubra*) are also shown for comparison. Genes translocated are colored. Inversions (indicated by the circular arrow) and transpositions of protein coding, tRNA and rRNA genes are depicted among the different taxa (except between Cerithioidea and Vermetoidea with respect to other caenogastropod lineages due to the high number of changes). Genes encoded by the minor strand are underlined.

Fig. 2. Phylogenetic relationships of Caenogastropoda based on the mt data set. The ML (A) and BI (B) phylograms are shown. Numbers at nodes are support values from ML (bootstrap proportions) and BI (posterior probabilities). Colors indicate classic higher taxonomic ranks of Caenogastropoda. Superfamilies within Caenogastropoda are indicated (* denotes that Muricoidea is recovered in three independent lineages corresponding to Muricoidea *sensu stricto*, Muricoidea-Volutidae, and Muricoidea-Babylonidae). Scale bar indicates substitutions/ site.

Fig. 3. Phylogenetic relationships of Caenogastropoda based on the nuclear data set. The ML (A: 50% strict consensus) and BI (B: 0.70 strict consensus) phylograms are shown. Numbers at nodes are support values from ML (bootstrap proportions) and BI (posterior probabilities). Colors indicate classic higher taxonomic ranks of Caenogastropoda. Superfamilies within Caenogastropoda are indicated (* denotes that Muricoidea is recovered in three independent lineages corresponding to Muricoidea *sensu stricto*, Muricoidea-

798 Volutidae, and Muricoidea-Babylonidae). Scale bar indicates substitutions/
799 site.

800 Fig. 4. Phylogenetic relationships of Caenogastropoda based on the combined data set.

801 The ML (A) and BI (B) phylograms are shown. Numbers at nodes are support
802 values from ML (bootstrap proportions) and BI (posterior probabilities). Colors
803 indicate classic higher taxonomic ranks of Caenogastropoda. Superfamilies
804 within Caenogastropoda are indicated (* denotes that Muricoidea is recovered
805 in three independent lineages corresponding to Muricoidea *sensu stricto*,
806 Muricoidea-Volutidae, and Muricoidea-Babylonidae). Scale bar indicates
807 substitutions/ site.

Figure 1

Architaenioglossa Littorinimorpha Neogastropoda Cerithioidea

1) Vetigastropoda 2) Neritimorpha 3) Tonnoidea, Cypraeoidea, Naticoidea, Truncatelloidea, Risssooidea, Abysochrysoidea, Stromboidea

4) Conoidea, Buccinoidea, Olivoidea, Cancellarioidea, Muricoidea

Figure 2

Figure 3

Figure 4

Table 1

Table 1. Complete caenogastropod mitochondrial (mt) genomes analyzed in this study

New mt genomes							
Species	Order	Length (bp)	GenBank Acc. No.	Voucher	Location	Collector	Year
<i>Naticarius hebraeus</i>	Naticoidea	15.384	KP716634	MNCN/ADN 65236	Castellón	Gilberto Diosdado	2009
<i>Galeodea echinophora</i>	Tonnoidea	15.388	KP716635	MNCN/ADN 65235	Valencia	David Osca	2008
<i>Erosaria spurca</i> [†]	Cypraeoidea	11.107	KP716636	MNCN/ADN 85223	Islas Columbretes	José Templado	2009
<i>Columbella adansoni</i>	Buccinoidea	16.272	KP716637	MNCN/ADN 85224	Playa de la Tejita (Tenerife)	Ivan Acevedo	2009
<i>Cochlostoma hidalgoi</i>	Cyclophoroidea	15.536	KP716638	MNCN/ADN 65237	Valle de Somiedo (La Malva)	José Templado	2009

GenBank mt genomes				
Species	Order	Length (bp)	GenBank Acc. No.	Publication
<i>Marisa cornuarietis</i>	Ampullarioidea	15.923	NC_025334	Wang and Qiu 2014
<i>Pomacea canaliculata</i>	Ampullarioidea	15.707	NC_024586	Zhou et al. 2014
<i>Cipangopaludina cathayensis</i>	Viviparoidea	17.157	NC_025577	Yang et al. 2014
<i>Semisulcospira libertina</i>	Cerithioidea	15.432	NC_023364	Zeng et al. 2014
<i>Ifremeria nautilei</i>	Abyssochrysoidea	15.664	NC_024642	Osca et al. 2014
<i>Tricula hortensis</i>	Rissooidea	15.179	NC_013833	Zhao et al. 2010
<i>Oncomelania hupensis</i>	Rissooidea	15.186	NC_012899	Zhao et al. 2010
<i>Potamopyrgus antipodarum</i>	Truncatelloidea	15,110	NC_020790	Neiman et al. 2010
<i>Potamopyrgus estuarinus</i>	Truncatelloidea	15,120	NC_021595	Neiman et al. 2010
<i>Thylacodes squamigerus</i>	Vermetoidea	15,544	NC_014588	Rawlings et al. 2010
<i>Ceraesignum maximum</i>	Vermetoidea	15,578	NC_014583	Rawlings et al. 2011
<i>Lobatus gigas</i>	Stromboidea	15.461	NC_024932	Marquez et al. 2014
<i>Bivetiella cancellata</i>	Cancellarioidea	16.648	NC_013241	Cunha et al. 2009
<i>Monoplex parthenopeus</i>	Tonnoidea	15,270	NC_013247	Cunha et al. 2009
<i>Ilyanassa obsoleta</i>	Buccinoidea	15.263	NC_007781	Simison et al. 2006
<i>Nassarius reticulatus</i>	Buccinoidea	15.271	NC_013248	Cunha et al. 2009
<i>Cymbium olla</i>	Muricoidea	15.375	NC_013245	Cunha et al. 2009
<i>Babylonia areolata</i>	Muricoidea	15.445	NC_023080	Chen and Ke 2013*
<i>Bolinus brandaris</i>	Muricoidea	15,380	NC_013250	Cunha et al. 2009
<i>Rapana venosa</i>	Muricoidea	15.272	NC_011193	Chandler et al. 2007*
<i>Concholepas concholepas</i>	Muricoidea	15.495	NC_017886	Núñez-Acuña et al. 2013
<i>Reishia clavigera</i>	Muricoidea	15.285	NC_010090	Ki et al. 2010
<i>Amalda northlandica</i>	Oливоidea	15.354	NC_014403	McComish et al. 2010
<i>Fusiturris similis</i>	Conoidea	15.595	NC_013242	Cunha et al. 2009
<i>Lophiotoma cerithiformis</i>	Conoidea	15,380	NC_008098	Bandyopadhyay et al. 2006
<i>Oxymeris dimidiata</i>	Conoidea	16.513	NC_013239	Cunha et al. 2009
<i>Cylinder textile</i>	Conoidea	15.562	NC_008797	Bandyopadhyay et al. 2008
<i>Pionoconus consors</i>	Conoidea	16.112	NC_023460	Brauer et al. 2012
<i>Africonus borgesii</i>	Conoidea	15.536	NC_013243	Cunha et al. 2009

[†]nearly complete; *Unpublished

Table 2

Table 2. Nuclear DNA sequences used in the phylogenetic analyses

Taxon		Nuclear gene fragments					
Lower	Higher	18S	28S_d1	28S_d6	28S_d4-d5	28S_B	H3
Caenogastropoda							
<i>Leptopoma perlucida</i>	Cyclophoroidea	DQ916519	AY296887	DQ916467	AF033758	AF033759	AF033696
—	Ampullarioidea	DQ916521	DQ916572	ASU78643	DQ916509	DQ916455	DQ916445
<i>Bellamya heudi</i>	Viviparoidea	DQ916520	U75863	U82423	AF033724	AH005970	AF033679
<i>Maoricolpus roseus</i>	Cerithioidea	DQ916523	DQ916573	DQ916469	DQ916510	DQ916456	DQ916446
<i>Ifremeria nautilei</i>	Abyssochrysoidea	GQ290545	GQ290527	GQ290564	—	—	GQ290545
<i>Oncomelania hupensis</i>	Rissooidea	DQ916528	AB611785	—	DQ916512	—	—
<i>Tricula</i>	Rissooidea	AF212908	EF394885	—	—	—	JQ082587
<i>Potamopyrgus</i>	Truncatelloidea	JN997449	AB930357	—	—	—	—
<i>Serpulorbis</i>	Vermetoidea	HQ833992	AY296890	DQ916475	AF033784	AF033785	AF033710
<i>Conuber melanostoma</i>	Naticoidea	DQ916533	DQ916580	DQ916580	DQ916515	DQ916463	DQ916452
—	Stromboidea	GU198749	—	AF033787	AF033786	AF033787	AF033711
<i>Sydaphera undulata</i>	Cancellarioidea	HQ833989	AY296901	DQ916489	AF033742	AF033743	AF033688
<i>Cypraea annulus</i>	Cypraeoidea	DQ916532	AY296892	DQ916479	AF033728	AF033729	AF033681
<i>Tonna cerevisina</i>	Tonnoidea	DQ916534	DQ916582	DQ916483	DQ916517	DQ916465	DQ916453
<i>Cabestana spengleri</i>	Tonnoidea	HM486915	AY296893	DQ916482	AF033738	AF033739	AF033686
<i>Ilyanassa obsoleta</i>	Buccinoidea	AY145379	AY145411	AY145411	--	AY145411	—
<i>Nassarius burchardi</i>	Buccinoidea	DQ916536	AY296897	DQ916486	AF033766	AF033767	AF033702
Volutidae	Muricoidea	HQ834042	DQ916584	—	DQ916518	DQ916466	HQ834182
<i>Babylonia areolata</i>	Muricoidea	HQ834021	—	EU399873	—	—	HQ834157
<i>Bolinus brandaris</i>	Muricoidea	DQ279944	DQ279986	DQ279986	—	—	DQ279944
<i>Concholepas concholepas</i>	Muricoidea	HM486913	EU391554	—	—	—	HM486913
<i>Dicathais orbita</i>	Muricoidea	DQ916537	AY296898	DQ916487	AF033746	AF033747	AF033690

<i>Reishia clavigera</i>	Muricoidea	X91979	HE584200	—	—	—	JN413189
<i>Oliva spicata</i>	Olivoidea	L78898	FM999144	—	—	—	—
<i>Rhizoconus miles</i>	Conoidea	DQ916538	AY296900	DQ916490	AF033734	AF033735	DQ916538
<i>Terebra textilis</i>	Conoidea	EU015525	KJ710222	—	—	—	EU015857
<i>Lophiotoma</i>	Conoidea	HQ834048	EU820176	—	—	—	HQ834048

Outgroups

<i>Nerita atramentosa</i>	Neritoidea	DQ916542	AY296885	DQ916495	AF033764	AF033765	AF033701
<i>Tegula</i>	Trochoidea	GQ160801	HE800729	—	—	—	AY923979
Turbinidae	Trochoidea	AY698071	FR695685	—	—	—	AY923980
<i>Aplysia</i>	Anaspidea	DQ916539	AY296905	U78644	AF033716	AF033717	AF033675
—	Acteonoidea	AY427516	AY427481	FJ917243	—	DQ916566	EF133483
<i>Lottia</i>	Lottioidea	FJ977632	DQ248942	DQ248942	DQ248942	DQ248942	FJ977727
—	Chaetodermomorpha	X91977	AY145397	AY145397	—	—	AY377763

Appendix

1. Strategy for PCR amplification and sequencing of caenogastropod complete mitochondrial genomes

- *Columbella adansoni*

- Complete

1) Standard PCRs: *cob* and *cox3* (Boore and Brown, 2000)

2) Design of specific primers:

CaCob.F: 5'-agtgaatgggtatgaggaggattgcg-3'
CaCob.R: 5'-accgcaaatcctcctcataccattcca-3'

CaCOX3.F: 5'-tcatacaattcaggtgtcaaagggggttcg-3'
CaCOX3.R: 5'-cagttgctagaagccctgtaggcca-3'

3) Long PCRs:

CaCob.F>CACOX3.R	4000 bp	98°/30" (98°/10" - 57°/30" - 68°/12')*45 68°/10'
CaCOX3.F>CACob.R	+10000 bp	98°/30" (98°/10" - 57°/30" - 68°/12')*45 68°/10'

-Illumina sequencing of the CACob.F-CACOX3.R fragment failed and the following primers were used for nested PCR:

CaCobc.F: 5'-agtagggctgattgccatcctgt-3'
CaEnd5a.R: 5'-ttgggctgaaataggagct-3'

CaEnd5.F: 5'-agtcctatttcagcccaa-3'
CaCOX3.R (see above)

Coadnad5F 5'-ggcaacctcaattttaagattagctgc-3'
Coadcox3R 5'-ctttgacacctgaattgtatgaaatccc-3'

CaCobc.F>CAEnd5a.R	3000 bp	94°/60" (98°/10" - 57°/30" - 68°/4')*45 68°/10'
CaEnd5.F>CACOX3.R	1500 bp	94°/60" (98°/10" - 57°/30" - 68°/4')*45 68°/10'
Coadnad5F>Coadcox3R	1600 bp	94°/60" (98°/10" - 53°/30" - 68°/4')*45 68°/12'

-Primer walking:

ND4_seq.F: 5'-tccgagggaatgattacgtgtctagt-3'
Cob_seq.F: 5'-aggatatgggatcttgcgtcta-3'

ND5_seq.R: 5'-caatcccagccattagca-3'
ND5_seq2.R: 5'-agttatcctgcaagaactcaactc-3'

P1CAF-UJ: 5'-aaagataagagcatagcactgaagatgc-3'
P2CAR-UJ: 5'-atcactacatggttactacctatccagc-3'

-*Erosaria spurca*

- Nearly complete (missing 3000 bp between *cob* and *cox3*)

1) **Standard PCRs:** *cob* and *cox3* (Boore and Brown, 2000); *cox1* (Folmer et al, 1994)

2) **Design of specific primers:**

EsCOX1.F: 5'-tcctgccagtttagctggggct-3'
EsCob.R: 5'-gatggagagaaagtgcattccacgct-3'

EsCOX3.F: 5'-ccaagtctctaaaggactccggtgagga-3'
EsCOX1.R: 5'-tgggcaatgtttcccgttaaagg-3'

3) **Long PCRs:**

EsCOX1.F>EsCob.R	8000-10000 bp	98°/30" (98°/10" - 57°/30" - 68°/12")*45 68°/10'
EsCOX3.F>EsCOX1.R	3000-4000 bp	98°/30" (98°/10" - 57°/30" - 68°/12")*45 68°/10'

- Large fragment was sequenced with Illumina. Short fragment was sequenced through primer walking:

EsCOX3a.F: 5'-acagaagcctagcccctacgccagaatt-3'
EsCOX3a.R: 5'-atgatgcccgtcgagaagtgtgg-3'
Esnad2.R: 5'-gccaaaaattaggagacttgacccagc-3'
Es_tSer.R: 5'-agttagcagccctgcttactccaagtt-3'
Es_tArg.R: 5'-tgaagaaggactctaccgccaaggta-3'

- *Natica hebraeus*

- Complete

1) **Standard PCRs:** *nad4* (Shao et al., 2006); *rrnL* (Palumbi, 1996); *cox1* (Folmer et al, 1994)

2) **Design of specific primers:**

NhCOX1.F: 5'-tgattggtgggttgggaa-3'
Nh16S.R: 5'-tcagggtaccgcgcgcgtgaat-3'

Nh16S.F: 5'-gcggtaccctgaccgtgcaa-3'
Nhnd4.R: 5'-ccagcttgtaaactgtccggttga-3'

Nhnd4.F: 5'-gctatcaaccggaacgtttacaagctgga-3'
NhCOX1.R: 5'-gctcctaaaattgaagaagcaccagct-3'

3) **Long PCRs:**

NhCOX1.F>Nh16S.R	5000 bp	98°/30" (98°/10" - 55°/30" - 68°/7")*45 68°/10'
Nh16S.F>Nhnd4.R	4000-5000 bp	98°/30" (98°/10" - 63°/30" - 68°/7")*45 68°/10'
Nhnd4.F>NhCOX1.R	5000-6000 bp	98°/30" (98°/10" - 60°/30" - 68°/7")*45 68°/10'

- Full sequencing with Illumina

- *Cochlostoma hidalgoi*

1) **Standard PCRs:** *cob* and *cox3* (Boore and Brown, 2000); *cox1* (Folmer et al, 1994)

2) **Design of specific primers:**

ChCOX1.F: 5'-gccctgatatagctttccgcgcttaa-3'

ChCob.R: 5'-tcgacagcaaagccaccccaaact-3'

ChCob.F: 5'-agtttgggtggctttgctgtcga-3'

ChCOX3.R: 5'-gcaagaacctaattcaaccctaggcgcg-3'

ChCOX3.F: 5'-tgagcatatttcataggagcctcgcg-3'

ChCOX1.R: 5'-aaagcacaggatctccaccaccagct-3'

3) **Long PCRs:**

ChCOX1.F>ChCob.R	8000-10000 bp	98°/30" (98°/10" - 59°/30" - 68°/12')*45 68°/10'
ChCob.F>ChCOX3.R	5000-6000 bp	98°/30" (98°/10" - 57°/30" - 68°/12')*45 68°/10'
ChCOX3.F>ChCOX1.R	3000-4000 bp	98°/30" (98°/10" - 57°/30" - 68°/12')*45 68°/10'

- Shotgun sequencing plus nested PCRs:

Chnd1.F: 5'-tatgctgcgggtggctttgc-3'

ChCob.R: 5'-tcgacagcaaagccaccccaaact-3'

Chnd5.F: 5'-agagttttcccttacggtact-3'

ChCOX3a.R: 5'-acctcaacgtaatccgctggctacc-3'

Chnd5_def.F: 5'-ttagaccacggctgacttga-3'

ChCOX3b.R: 5'-cgtaagaccatgtagtcaacctgc-3'

Chnd1.F>ChCob.R	94°/60" (98°/10" - 57°/30" - 68°/4')*45 68°/5'
Chnd5.F>ChCOX3a.R	98°/30" (98°/10" - 57°/30" - 68°/2')*45 68°/5'
Chnd5_def.F>ChCOX3b.R	98°/30" (98°/10" - 57°/30" - 68°/2')*45 68°/3'

- And primer walking:

ChCOX2.F: 5'-tcccacaagcgacctcagaagcg-3'

Ch12Sb.R: 5'-tcatctttcgacgagagcga-3'

ChATP6.F: 5'-gtgaggatatcacaagtactcaga-3'

ChND5_def2.F: 5'-aatggccttgtaagtcgt-3'

- *Galeodea echinophora*

- Complete

1) Standard PCRs: *rrnS* (Kocher et al., 1989); *cox1* (Folmer et al., 1994); *cob* (Boore and Brown, 2000)

2) Design of specific primers:

GeCOX1.F: 5'-ttccttaccggtattagctggagctattactat-3'
Ge12S.R: 5'-acctttatataggctgcaccttgatctgacg-3'

Ge12S.F: 5'-acctgtctcataatcgacaatccgcgttacacc-3'
GeCob.R: 5'-ggaactttctaccgctcgctatttaatcccag-3'

GeCob.F: 5'-ttgaggagggttcgctattgataacgcaactct-3'
GeCOX1.R: 5'-tcaacagatcctccggcatgagctagattacc-3'

3) Long PCRs:

GeCOX1.F>Ge12S.R	3000-4000 bp	94°/1' (98°/10' - 68°/12')*40 72°/10'
Ge12S.F>GeCob.R	3500-4000 bp	94°/1' (98°/10' - 68°/12')*40 72°/10'
GeCob.F>GeCOX1.R	8000-10000 bp	94°/1' (98°/10' - 68°/12')*40 72°/10'

- Shotgun sequencing plus nested PCRs:

GeATP6.F: 5'-aatctcagccaccaaca-3'
Genad1.R: 5'-caattgctcgtagctctcc-3'

Ge16S.F: 5'-aggctctgaaacatgcac-3'
Genad1.R: 5'-caattgctcgtagctctcc-3'

GeCOX3.F: 5'-ggcatcatttggcttgaagcag-3'
GeCOX1a.R: 5'-tcgtctcctagtaaagctccaggtga-3'

Gend4.F: 5'-gcatgtgaagcaagtctcg-3'
Gend5a.R: 5'-taaagagccagcgcatcaa-3'

GeATP6.F>Genad1.R	94°/60" (98°/10" - 62°/30" - 68°/4')*45 68°/5'
Ge16.F>Genad1.R	94°/60" (98°/10" - 62°/30" - 68°/3')*45 68°/5'
GeCOX3.F>GeCOX1a.R	94°/60" (98°/10" - 62°/30" - 68°/4')*45 68°/5'
GeND4.F>GeND5a.R	94°/60" (98°/30" - 57°/30" - 68°/4')*45 68°/5'

- And primer walking:

Gecoba_int.F: 5'-cggtgagaaagttccatttcacg-3'
Gend5def.F: 5'-actgtagcagcatcattgc-3'
Gend5def.R: 5'-tgacttaatgtggatagcgc-3'

2. Annotation of the new caenogastropod mitochondrial genomes

<i>Cochlostoma hidalgoi</i>					
Gene		Position	Strand	Codon	Length (pb)
		From-To		Start/Stop	
<i>cox1</i>		1-1539	+	ATG/TAA	1539
<i>cox2</i>		1543-2229	+	ATG/TAA	687
<i>atp8</i>		2296-2463	+	ATG/TAG	168
tRNA-Asp	D - GUC	2480-2547	+		
<i>atp6</i>		2551-3263	+	ATG/T--	713
tRNA-Met	M - CAU	3336-3270	-		
tRNA-Gln	Q - UUG	3402-3339	-		
tRNA-Gly	G - UCC	3466-3404	-		
tRNA-Cys	C - GCA	3531-3472	-		
tRNA-Trp	W - UCA	3597-3533	-		
tRNA-Tyr	Y - GUA	3663-3600	-		
tRNA-Glu	E - UUC	3737-3674	-		
<i>12S</i>		3738-4699	+		962
tRNA-Val	V - UAC	4700-4765	+		
<i>16S</i>		4766-6081	+		1316
tRNA-Leu	L - UAG	6082-6147	+		
tRNA-Leu	L - UAA	6165-6231	+		
<i>nad1</i>		6232-7167	+	ATG/TAG	939
tRNA-Pro	P - UGG	7169-7237	+		
<i>nad6</i>		7238-7750	+	ATG/TAA	512
<i>cob</i>		7743-8880	+	ATG/T--	1140
tRNA-Ser	S - UGA	8881-8946	+		
<i>nad4l</i>		9003-9293	+	ATG/TAA	291
<i>nad4</i>		9287-10654	+	ATG/TAG	1367
tRNA-His	H - GUG	10658-10726	+		
<i>nad5</i>		10727-12440	+	ATG/T--	1713
tRNA-Phe	F - GAA	12459-12525	+		
tRNA Thr	T - UGU	12544-12607	+		
<i>cox3</i>		12856-13635	+	ATG/TAG	780
tRNA-Lys	K - UUU	13662-13734	+		
tRNA-Ala	A - UGC	13753-13817	+		
tRNA-Arg	R - UCG	13827-13895	+		
tRNA-Asn	N - GUU	13894-13959	+		
tRNA-Ile	I - GAU	13960-14028	+		
<i>nad3</i>		14029-14382	+	ATG/TAA	354
tRNA-Ser	S - GCU	14388-14454	+		
<i>nad2</i>		14455-15528	+	ATG/TAA	1074

<i>Naticarius hebraeus</i>					
Gene		Position	Strand	Codon	Length (pb)
		From-To		Start/Stop	
<i>cox1</i>		1-1536	+	ATG/TAA	1536
<i>cox2</i>		1553-2239	+	ATG/TAA	687
tRNA-Asp	D - GUC	2245-2312	+		
<i>atp8</i>		2314-2472	+	ATG/TAA	159
<i>atp6</i>		2475-3170	+	ATG/TAA	696
tRNA-Met	M - CAU	3269-3203	-		
tRNA-Tyr	Y - GUA	3342-3273	-		
tRNA-Cys	C - GCA	3408-3343	-		
tRNA-Trp	W - UCA	3475-3410	-		
tRNA-Gln	Q - UUG	3537-3477	-		
tRNA-Gly	G - UCC	3613-3545	-		
tRNA-Glu	E - UUC	3687-3621	-		
<i>12S</i>		3688-4658	+		971
tRNA-Val	V - CAC	4659-4726	+		
<i>16S</i>		4727-6158	+		1432
tRNA-Leu	L - UAA	6159-6226	+		
tRNA-Leu	L - UAG	6232-6300	+		
<i>nad1</i>		6301-7222	+	ATG/TAA	922
tRNA-Pro	P - UGG	7245-7312	+		
<i>nad6</i>		7314-7807	+	ATG/TAG	493
<i>cob</i>		7824-8963	+	ATG/TAA	1140
tRNA-Ser	S - UGA	8971-9037	+		
tRNA Thr	T - UGU	9123-9056	-		
<i>nad4l</i>		9131-9427	+	ATG/TAG	297
<i>nad4</i>		9421-10794	+	GTG/TAA	1374
tRNA-His	H - GUG	10801-10867	+		
<i>nad5</i>		10868-12569	+	ATG/TAA	1702
tRNA-Phe	F - GAA	12592-12658	+		
<i>cox3</i>		12692-13495	+	ATA/TAA	804
tRNA-Lys	K - UUU	13520-13591	+		
tRNA-Ala	A - UGC	13596-13662	+		
tRNA-Arg	R - UCG	13678-13746	+		
tRNA-Asn	N - GUU	13748-13817	+		
tRNA-Ile	I - GAU	13827-13897	+		
<i>nad3</i>		13898-14250	+	ATG/TA-	353
tRNA-Ser	S - GCU	14251-14320	+		
<i>nad2</i>		14320-15381	+	ATG/TAA	1062

<i>Galeodea echinophora</i>					
Gene		Position	Strand	Codon	Length (pb)
		From-To		Start/Stop	
<i>cox1</i>		1-1536	+	ATG/TAA	1536
<i>cox2</i>		1557-2241	+	ATG/T--	687
tRNA-Asp	D - GUC	2242-2309	+		
<i>atp8</i>		2310-2468	+	ATG/TAA	159
<i>atp6</i>		2474-3169	+	ATG/TAA	696
tRNA-Met	M - CAU	3277-3210	-		
tRNA-Tyr	Y - GUA	3363-3297	-		
tRNA-Cys	C - GCA	3448-3383	-		
tRNA-Trp	W - UCA	3514-3449	-		
tRNA-Gln	Q - UUG	3578-3514	-		
tRNA-Gly	G - UCC	3647-3580	-		
tRNA-Glu	E - UUC	3718-3648	-		
<i>12S</i>		3719-4687	+		969
tRNA-Val	V - CAC	4688-4755	+		
<i>16S</i>		4756-6147	+		1392
tRNA-Leu	L - UAG	6148-6216	+		
tRNA-Leu	L - UAA	6218-6286	+		
<i>nad1</i>		6287-7228	+	ATG/TAG	942
tRNA-Pro	P - UGG	7229-7297	+		
<i>nad6</i>		7299-7799	+	ATG/TAA	501
<i>cob</i>		7807-8946	+	ATG/TAG	1140
tRNA-Ser	S - UGA	8952-9016	+		
tRNA Thr	T - UGU	9088-9017	-		
<i>nad4l</i>		9097-9393	+	ATG/TAG	297
<i>nad4</i>		9387-10760	+	ATG/TAA	1373
tRNA-His	H - GUG	10774-10838	+		
<i>nad5</i>		10839-12560	+	ATG/TAA	1721
tRNA-Phe	F - GAA	12561-12630	+		
<i>cox3</i>		12686-13465	+	ATG/TAA	780
tRNA-Lys	K - UUU	13481-13554	+		
tRNA-Ala	A - UGC	13579-13646	+		
tRNA-Arg	R - UCG	13658-13726	+		
tRNA-Asn	N - GUU	13728-13795	+		
tRNA-Ile	I - GAU	13824-13893	+		
<i>nad3</i>		13898-14251	+	ATG/TAA	354
tRNA-Ser	S - GCU	14256-14323	+		
<i>nad2</i>		14324-15382	+	ATG/TAA	1059

<i>Erosaria spurca</i>					
Gene		Position	Strand	Codon	Length (pb)
		From-To		Start/Stop	
<i>cox1</i>		2484-4022	+	ATG/TAA	1539
<i>cox2</i>		4048-4734	+	ATG/TAA	687
tRNA-Asp	D - GUC	4743-4811	+		
<i>atp8</i>		4813-4971	+	ATG/TAA	159
<i>atp6</i>		4976-5671	+	ATG/TAA	696
tRNA-Met	M - CAU	5775-5708	-		
tRNA-Tyr	Y - GUA	5844-5777	-		
tRNA-Cys	C - GCA	5920-5851	-		
tRNA-Trp	W - UCA	5987-5921	-		
tRNA-Gln	Q - UUG	6050-5988	-		
tRNA-Gly	G - UCC	6121-6055	-		
tRNA-Glu	E - UUC	6193-6123	-		
<i>12S</i>		6194-7161	+		968
tRNA-Val	V - CAC	7162-7231	+		
<i>16S</i>		7232-8633	+		1402
tRNA-Leu	L - UAG	8634-8702	+		
tRNA-Leu	L - UAA	8706-8775	+		
<i>nad1</i>		8777-9718	+	ATA/TAG	942
tRNA-Pro	P - UGG	9719-9785	+		
<i>nad6</i>		9787-10290	+	ATG/TAA	504
<i>cob</i>		10296-11105	+	ATG/---	
tRNA-Ser	S - UGA	---			
tRNA Thr	T - UGU	---			
<i>nad4l</i>		---		---	---
<i>nad4</i>		---		---	---
tRNA-His	H - GUG	---			
<i>nad5</i>		---		---	---
tRNA-Phe	F - GAA	---			
<i>cox3</i>		1-600*(250pb absent)	+	---/TAA	---
tRNA-Lys	K - UUU	622-689	+		
tRNA-Ala	A - UGC	691-759	+		
tRNA-Arg	R - UCG	764-832	+		
tRNA-Asn	N - GUU	843-912	+		
tRNA-Ile	I - GAU	919-987	+		
<i>nad3</i>		993-1346	+	ATG/TAA	354
tRNA-Ser	S - GCU	1347-1414	+		
<i>nad2</i>		1415-2476	+	ATG/TAG	1059

<i>Columbella adansoni</i>					
Gene		Position	Strand	Codon	Length (pb)
		From-To		Start/Stop	
<i>cox1</i>		1-1533	+	ATG/TAG	1533
<i>cox2</i>		1567-2253	+	ATG/TAA	687
tRNA-Asp	D - GUC	2258-2325	+		
<i>atp8</i>		2326-2484	+	ATG/TAA	159
<i>atp6</i>		2495-3190	+	ATG/TAA	696
tRNA-Met	M - CAU	3227-3293	-		
tRNA-Tyr	Y - GUA	3299-3366	-		
tRNA-Cys	C - GCA	3367-3430	-		
tRNA-Trp	W - UCA	3431-3497	-		
tRNA-Gln	Q - UUG	3498-3559	-		
tRNA-Gly	G - UCC	3563-3630	-		
tRNA-Glu	E - UUC	3631-3695	-		
<i>12S</i>		3696-4650	+		955
tRNA-Val	V - CAC	4651-4719	+		
<i>16S</i>		4720-6067	+		1348
tRNA-Leu	L - UAG	6068-6136	+		
tRNA-Leu	L - UAA	6137-6205	+		
<i>nad1</i>		6206-7145	+	ATG/T--	940
tRNA-Pro	P - UGG	7146-7214	+		
<i>nad6</i>		7216-7716	+	ATG/TAA	501
<i>cob</i>		7719-8858	+	ATG/TAA	1140
tRNA-Ser	S - UGA	8865-8929	+		
tRNA Thr	T - UGU	8933-9000	-		
<i>nad4l</i>		9013-9309	+	ATG/TAG	297
<i>nad4</i>		9303-10674	+	ATG/T--	1377
tRNA-His	H - GUG	10675-10740	+		
<i>nad5</i>		10741-12462	+	ATG/TAA	1721
tRNA-Phe	F - GAA	12470-12538	+		
<i>cox3</i>		13632-14411	+	ATG/T--	774
tRNA-Lys	K - UUU	14427-14492	+		
tRNA-Ala	A - UGC	14512-14578	+		
tRNA-Arg	R - UCG	14580-14648	+		
tRNA-Asn	N - GUU	14651-14719	+		
tRNA-Ile	I - GAU	14721-14788	+		
<i>nad3</i>		14789-15142	+	ATG/TAA	354
tRNA-Ser	S - GCU	15143-15210	+		
<i>nad2</i>		15211-16269	+	ATG/TAA	1059

3. Partitions and model selection

[illegible]