The mitochondrial genome of *Ifremeria nautili* and the phylogenetic position of the enigmatic deep-sea Abysochrysoidea (Mollusca: Gastropoda)

David Osca, José Templado, Rafael Zardoya

Museo Nacional de Ciencias Naturales (MNCN-CSIC), José Gutiérrez Abascal
2, 28006, Madrid, Spain

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

Abstract

The complete nucleotide sequence of the mitochondrial (mt) genome of the deep-sea vent snail *Ifremeria nautiliei* (Gastropoda: Abysochrysoidea) was determined. The double stranded circular molecule is 15,664 pb in length and encodes for the typical 37 metazoan mitochondrial genes. The gene arrangement of the *Ifremeria* mt genome is most similar to genome organization of caenogastropods and differs only on the relative position of the *trnW* gene. The deduced amino acid sequences of the mt protein coding genes of *Ifremeria* mt genome were aligned with orthologous sequences from representatives of the main lineages of gastropods and phylogenetic relationships were inferred. The reconstructed phylogeny supports that *Ifremeria* belongs to Caenogastropoda and that is closely related to hypsogastropod superfamilies. Results were compared with a reconstructed nuclear-based phylogeny. Moreover, a relaxed molecular-clock timetree calibrated with fossils dated the divergence of Abysochrysoidea in the Late Jurassic-Early Cretaceous indicating a relatively modern colonization of deep-sea environments by these snails.

1. Introduction

Totally independent from sunlight, deep-sea hydrothermal vent and cold seep ecosystems are based on chemoautotrophic bacteria, which either have symbiotic relationships with or are the food source of complex metazoan communities mainly dominated by vestimentiferan tubeworms (siboglinid polychaetes), vesicomyid and bathymodiolid mussels, provannid snails, and bresiliid shrimps ([Van Dover, 2000](#); [Van Dover et al., 2002](#); [Little and Vrijenhoek, 2003](#)).

Several authors ([Newman, 1985](#); [Tunnicliffe, 1992](#); [McArthur and Tunnicliffe, 1998](#)) advocated for ancient (Paleozoic) origins of deep-sea hydrothermal vent and cold seep macrofauna based on the morphology of extant taxa inhabiting these ecosystems, the observed high levels of endemic and relict species, and the potential role of hydrothermal vents and cold seeps as refugia during mass extinctions thanks to their energy autonomy. Although it is true that living deep-sea fauna have significantly longer geologic ranges than other marine fauna ([Kiel and Little, 2006](#)), the rich fossil record of e.g. vent and seep mollusks goes back to the Late Jurassic-Early Cretaceous, suggesting relatively modern (Late Mesozoic or even Cenozoic) origins from non-vent ancestors ([Jacobs and Lindberg, 1998](#); [Warén and Bouchet, 2001](#); [Little and Vrijenhoek, 2003](#); [Kiel and Little, 2006](#)). Moreover, early molecular studies also supported relatively younger origin, although based on relatively short sequences and poorly resolved phylogenies that used calibrated strict molecular clocks ([Peek et al., 1997](#); [McArthur and Koop, 1999](#); [Shank et al., 1999](#); [Little and Vrijenhoek, 2003](#)). In order to discriminate among competing hypotheses, it would be important to reconstruct robust phylogenies based on

multi-gene data sets, to use relaxed-molecular clocks ([Yang et al., 2013](#)), and to select groups with a rich fossil record ([Little and Vrijenhoek, 2003](#)).

Snails of the superfamily Abysochrysoidea (Gastropoda: Caenogastropoda) commonly occur in deep-sea environments including hydrothermal vents, hydrocarbon seeps, and sites of organic deposition such as sunken wood and whale-falls ([Johnson et al., 2010](#)). These gastropods represent a relatively ancient lineage that has persisted since at least the Middle-Late Jurassic ([Kaim et al., 2008](#); [Kiel et al., 2008](#); [Kiel, 2010](#)). Therefore, this group of enigmatic gastropods constitutes a nice model system to test alternative hypotheses on the origin of the deep-sea hydrothermal vent and cold seep fauna. The superfamily Abysochrysoidea, as currently constructed ([Sasaki et al., 2010](#)) includes two extant families: Abysochrysoidea, with the only extant genus *Abysochryso* and the extinct genus *Humtulpisia* ([Kiel, 2008](#)), and Provannidae, which includes the extant genera *Ifremeria*, *Provanna*, *Alviniconcha*, *Cordesia* and *Desbruyeresia* ([Sasaki et al., 2010](#)) and the Late Pliocene fossil genus *Kaneconcha* ([Kaim et al., 2012](#)). Furthermore, the extinct family Hokkaidoconchidae ([Kaim et al., 2008](#)) has been also included in this superfamily, as well as the enigmatic extant genus *Rubyspira*, discovered from whale falls ([Johnson et al., 2010](#)). Phylogenetic relationships within the superfamily were reconstructed by ([Johnson et al., 2010](#)) based on a multi-gene data set, suggesting the paraphyly of Provannidae. In this study, a relaxed molecular clock dated the origin of Abysochrysoidea in the Early Cretaceous and lineage split into three major clades (*Provanna*; *Rubyspira*, *Abysochryso*, and *Desbruyeresia*; *Alviniconcha* and *Ifremeria*) during the Mid Cretaceous ([Johnson et al., 2010](#)). *Ifremeria* and *Alviniconcha* are known exclusively from

hydrothermal vents and they are the only known members of this superfamily to host chemosymbiotic bacteria in their gills ([Kaim et al., 2012](#))

As regards to its relative phylogenetic position within Gastropoda, the Abysochrysoidea were initially considered by ([Warén and Ponder, 1991](#)) and ([Warén and Bouchet, 2001](#)) as relict members of the otherwise extinct Loxonematoidea, as suggested previously by ([Houbrick, 1979](#)). Later, ([Bouchet and Rocroi, 2005](#)) placed them within the "Zygopleuroid group" of uncertain systematic position within the Caenogastropoda. On the other hand, phylogenetic analysis by ([Ponder and Lindberg, 1997](#)) and ([Ponder et al., 2008](#)); see Fig. 13.16) based on morphological characters placed Provannidae as a sister group of Littorinidae and closely related to Eatoniellidae and Cingulopsidae. Likewise, phylogenetic studies based on mitochondrial 16S rRNA gene also placed tentatively Provannidae as a sister group of Littorinidae ([Colgan et al., 2007](#)). Moreover, sperm ultrastructure suggested affinities between these two families ([Healy, 1990](#)). Nevertheless, the systematic position of Abysochrysoidea remains unclear and requires clarification with further genetic studies providing robust support.

Here, we report the complete nucleotide sequence of the mitochondrial (mt) genome of *Ifremeria nautilei*, the only known species of the genus. This large-sized snail is one of the dominant gastropods of the hydrothermal vent fields of the south-west Pacific, and is becoming a model system of Abysochrysoidea for population genetic and phylogeographic studies ([Kojima et al., 2000](#); [Suzuki et al., 2006](#); [Thaler et al., 2011](#); [Hidaka et al., 2012](#)). The new mt genome sequence data of *I. nautilei* were combined with corresponding available data from other Caenogastropoda in order to elucidate its relative

phylogenetic position. In addition, we investigated the evolution of mt genome rearrangements within Caenogastropoda, including the reported translocation of the tRNA^{TRP} gene in the mt genome of *I. nautili* (shared with *Alviniconcha*) with respect to *Provanna*, *Abyssochrysos* and *Debruyeresia* ([Hidaka et al., 2012](#)). Finally, we used a timetree of gastropods that was calibrated based on fossil data to test between competing hypotheses on the origin of Abyssochrysoidea.

2. Materials and methods

2.1. DNA extraction, PCR amplification and sequencing

A specimen of *I. nautili* collected by Hans Tøre Rapp (University of Bergen, Norway) in Lau Basin (Pacific Ocean) in 2009 was used for this study. The voucher is stored at the tissue/DNA collection of the National Museum of Natural Sciences in Madrid (Spain) with No. MNCN/ADN: 59262

Total genomic DNA was isolated following a standard phenol-chloroform extraction. Two sets of specific PCR primer pairs were designed for long PCR: (1) IN16S.F (5'-CCGCGGTACTCTGACCGTGCAA AGG-3') and IN16S.R (5'-AGAAGCTCTGAATGTTCTCAGTCGCCCCA-3'), which were derived from the *rrnL* sequence available at GenBank accession No. GQ290492.1 ([Johnson et al., 2010](#)); and (2) primers INCOX3.F (5'-AGGTTCTTGCTGGCCCCCTTCTGGA-3') and INCOX3.R (5'-GCAAAAAGCCCTTGAATAGCCTCTCTCCGA-3'), which were designed upon the sequencing of a DNA fragment that was PCR amplified using universal *cox3* primers ([Boore and Brown, 2000](#)). These two sets of primers allowed the amplification of the remaining mtDNA in two overlapping fragments (7725 pb and 7938 pb, respectively) using the TaKaRa LA-PCR kit. Long PCR

amplifications 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 and 0,5 µl Taq DNA polymerase (5 units/µl) and adding sterilized distilled water up to 50 µl. The following profile was 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 and the two overlapping fragments were pooled together in equimolar concentrations for massive parallel sequencing. An indexed DNA library (with inserts of 200-500 bp) was constructed using the TrueSeq DNA Sample Kit (Illumina, San Diego, CA, USA) and following manufacturer's instructions at the Sequencing Service of the Biomedicine and Biotechnology Institute of Cantabria (IBBTEC; Santander, Spain). The *Ifremeria* indexed library was pooled together with several indexed libraries from other mt genome and transcriptome projects into a single Illumina lane (HiSeq2000) and sequenced using the 100bp pair-ended technique at Macrogen (Seoul, Korea). The complete mt genome sequence reported in this paper has been deposited in NCBI GenBank under accession number KC757644.

2.2. Genome assembly and annotation

Raw reads from each of the projects run in the single Illumina lane were pulled apart by their unique indexes used in library construction. Prior to assembly, the quality of the raw Illumina reads was checked with FastQC v.0.10.1 ([Andrews, 2010](#)); all adapter sequences were removed with SeqPrep v1.1 ([St John, 2011](#)) and reads were trimmed and filtered out according to their quality scores using

PRINSEQ v.0.20.3 ([Schmieder and Edwards, 2011](#)). The CLC Genomics Workbench® v.5.5.1 (<http://www.clcbio.com>; CLC bio, Aarhus, Denmark) was used for the *de novo* assembly of the mt genome. The probabilistic variant calling method implemented in CLC Genomics Workbench was used to check and correct possible disagreements among reads that mapped onto the mt genome sequences. Most (90%) of the mtDNA was assembled directly from the HiSeq (short-read) data. However, a fragment between *trnG* and *rrnL* genes had low coverage statistics in the assembly and was PCR amplified independently using primers IN_tGly.F, 5'-TCTGCTTGGAAGGCAACTGTAC-3' and IN16S.R (see above). The amplified PCR fragment was Sanger sequenced to close the circular mt genome using the PCR primers and a walking primer, IN12S.F, 5'-TGGACGTCAGATCAAGGTGCAACCA-3'.

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 mollusk mt genomes, and assumed to extend to the boundaries of adjacent genes ([Boore et al., 2005](#)).

The gene order of the mt genome of *I. nautiliei* was compared to the gene arrangement of other reported mollusk mitochondrial genomes (Table 1): one cephalopod, *Octopus vulgaris* ([Yokobori et al., 2004](#)), and 14 gastropods: one Vetigastropoda: *Haliotis rubra* ([Maynard et al., 2005](#)), one Neritimorpha: *Nerita*

melanotragus ([Castro and Colgan, 2010](#)), one Patellogastropoda: *Lottia digitalis* ([Simison et al., 2006](#)), six Heterobranchia: *Roboastra europea* ([Grande et al., 2002](#)), *Pyramidella dolabrata* and *Onchidella celtica* ([Grande et al., 2008](#)), *Albinaria caerulea* ([Hatzoglou et al., 1995](#)), *Aplysia californica* ([Knudsen et al., 2006](#)), *Pupa strigosa* ([Kurabayashi and Ueshima, 2000](#)) and five Caenogastropoda: *Oncomelania hupensis* ([Zhao et al., 2010](#)), *Dendropoma maximum* ([Rawlings et al., 2010](#)) and *Cymatium parthenopeum*, *Bolinus brandaris* and *Conus borgesii* ([Cunha et al., 2009](#)).

2.3. Sequence alignment

The mtDNA sequence of *I. nautiliei* was aligned to the orthologous sequences of 14 complete mt genomes of representatives of main gastropod orders and one cephalopod (Table 1). Derived amino acid sequences corresponding to all 13 mt protein-coding genes were downloaded from GenBank using MitoBank ([Abascal et al., 2007](#)). Due to the relative high substitution rates of mtDNA, 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 order to ameliorate saturation.

In addition, a nuclear data set was constructed by retrieving from GenBank the nucleotide sequences of the nuclear genes 18S, 28S (segments d1 and d6), and histone H3 ([Johnson et al., 2010](#)) of the representatives of the main orders of gastropods and one cephalopod (Table 1). 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 1). 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) or the nucleotide sequences of the three analyzed genes (nuclear data set) were aligned separately using MAFFT ([Kato et al., 2002](#)) with default settings. Ambiguously aligned positions were removed using Gblocks, v.0.19b ([Castresana, 2000](#)) under default settings. Finally, the different single alignments were concatenated.

2.4. Phylogenetic analyses

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 (as judged by plots of ML scores and low SD of split frequencies) to prevent sampling before reaching stationarity. 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. For the mt data set

the partitions tested were: all genes combined; all genes separated; all genes separated except *nad4/nad4L* and *atp8/atp6*; genes grouped by subunits i.e., *cox*, *atp*, *nad* and *cob*. For the nuclear data set the two partitions tested were: all genes combined versus all genes separated. 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 posterior probabilities (BI).

2.5 Estimation of divergence times

We used BEAST v.1.7 ([Drummond and Rambaut, 2007](#)) to perform a Bayesian estimation of divergence times among major gastropod lineages based on the mt amino acid data set. This software was used to infer tree topology, branch lengths, and nodal ages. For the clock model, we selected the lognormal relaxed-clock model, which allows rates to vary among branches without any a priori assumption of autocorrelation between adjacent branches. For the tree prior, we employed a Yule process of speciation. The ingroup was considered monophyletic with respect to the outgroup (*O. vulgaris*). We employed the partitions selected by Partition Finder Protein (see above), and the mtREV model for all partitions as selected by PROTTEST ([Abascal et al., 2005](#)) among the ones available in BEAST. The final Markov chain was run twice for 50 million generations, sampling every 1,000 generations and the first 5 million was discarded as part of the burn-in process, according to the convergence of chains checked with Tracer v.1.5. ([Rambaut and Drummond, 2007](#)). The effective sample size of all the parameters was above 200.

The posterior distribution of the estimated divergence times was obtained by specifying five calibration points as priors for divergence times of certain splits, using a lognormal distribution of prior probability. Fossils provided hard minimum bounds (offset) and mean and standard deviations (SD) were chosen so that the 95% probability limit corresponds to a soft maximum bound. Details on fossil dates for each calibration point are: a prior of 390 MY was used for the divergence of the Heterobranchia, based on the oldest known fossil of the Heterobranchia, documented from the Middle Devonian ([Bandel and Heidelberger, 2002](#)). Caenogastropods become recognized in the Early Carboniferous ([Bandel, 1997](#)) and thus a prior of 350 MY was used as its divergence time. Although possible stem group representatives of modern neogastropods are the Triassic-Cretaceous Maturifusidae ([Kaim, 2004](#)), the earliest record of muricoidean gastropods is known from the early Cretaceous. Hence, a normally distributed estimate prior of 140 MY was used for the divergence of the Neogastropoda from the other caenogastropods lineages. Within Heterobranchs, Pyramidelloidea are known since the Early Jurassic ([Kaim, 2004](#)), hence a prior of 190 MY was used for the divergence of the Pyramidelloidea from basal pulmonates. Likewise, Acteonidae dates back to the Mid Jurassic ([Kaim, 2004](#)), so we considered a prior of 170 MY for the divergence time of the Acteonoidea from the other major groups of opisthobranchs.

3. Results and Discussion

3.1. Mitochondrial genome organization and structural features

The complete mitochondrial genome of *Ifremeria nautiliei* was sequenced and assembled as a 15,664 pb circular molecule. The Illumina sequencing produced 289.66 Mb of sequence data (quality score Q30= 95%) and 81% of the raw reads (2,867,960) passed the adapter and quality trimming thresholds. The final assembly had a sequencing depth over 14000x. As is the case for most metazoans ([Boore, 1999](#)), the mtDNA encodes for 13 protein-coding genes, 22 tRNAs, and 2 rRNAs (Fig 1). The major strand encodes 29 out of the 37 genes (*cox1*, *cox2*, *trnD*, *atp6*, *atp8*, *rrnS*, *trnV*, *rrnL*, *trnL(uaa)*, *trnL(uag)*, *nad1*, *trnP*, *nad6*, *cob*, *trnS(uga)*, *nad4L*, *nad4*, *trnH*, *nad5*, *trnF*, *cox3*, *trnK*, *trnA*, *trnR*, *trnN*, *trnI*, *nad3*, *trnS(gcu)*, *nad2*). The genes *nad4*/*nad4L* and *atp8*/*atp6* overlapped, respectively. All protein-coding genes start with ATG with the exception of *nad4* that begins with ATA; *cox1*, *cox2*, *atp8*, *atp6*, *cob*, *cox3*, *nad3* and *nad2* end with TAA, and *nad4I*, *nad4* and *nad5* with TAG. The remaining genes (*nad1* and *nad6*) end with a single T, which presumably becomes functional by subsequent polyadenylation of the transcribed messenger RNAs ([Lopez Sanchez et al., 2011](#)). The range in size of the 22 tRNA genes is from 67 to 72 nucleotides. There are 27 noncoding regions, eight of them being longer than 30 base pairs, with a maximum of 56 bps (Fig. 1). The largest one was found adjacent to *cox3* gene, in a position that has been postulated as candidate to contain the control region in other gastropod mt genomes ([Kurabayashi and Ueshima, 2000](#)). Moreover, it has been suggested that some of these intergenic regions may be remnants of translocation events ([Hidaka et al., 2012](#)). All these intergenic regions show high A+T frequencies.

The gene order of the mt genome of *I. nautiliei* was compared against that of other gastropods and a cephalopod (Fig. 2). It was found to be very similar to

the gene arrangements reported for caenogastropod mt genomes (Fig. 2). Mitochondrial genome organization in gastropods is prone to rearrangements between main lineages but relatively stable within these, changes being restricted to tRNA genes ([Grande et al., 2008](#); [Cunha et al., 2009](#)). Therefore, a mt gene arrangement can be used as a fair proxy to assign any gastropod to one of the main lineages, and in the case of *I. nautiliei*, its assignment to the Caenogastropoda ([Cunha et al., 2009](#)). Within caenogastropods, the closest gene order is that shared by the mt genomes of *O. hupensis*, *C. parthenopeum* and *B. brandaris*, which only differs to that of *I. nautiliei* in the relative position of the *trnW* gene, which in the former is located between the *trnC* and *trnQ* genes and in the latter is found between the *trnM* and *trnY* genes (Fig. 2). The ancestral position of the *trnW* gene in gastropods was likely between the *trnY* and *trnQ* genes, as is found in some Cephalopoda ([Yokobori et al., 2004](#)), Polyplacophora ([Boore and Brown, 1994](#)) and some Bivalvia ([Plazzi et al., 2013](#)). Within gastropods, Vetigastropoda, Neritimorpha, and Caenogastropoda, have the *trnC* normally between *trnY* and *trnW* (the relative position of the *trnC* could be a molecular synapomorphy indicating a common ancestor of the three groups) ([Castro and Colgan, 2010](#)) (Fig. 2). In *I. nautiliei*, the *trnW* has been transposed to the region between the *trnM* and *trnY* genes. This location is shared with genus *Alviniconcha* but not with other Abyssochrisoidea that show the typical arrangement of Caenogastropoda ([Hidaka et al., 2012](#)). Therefore, the transposition of the *trnW* likely occurred in the common ancestor of *Ifremeria* and *Alviniconcha* ([Hidaka et al., 2012](#)).

3.2 Phylogenetic analyses

The molecular phylogeny of Caenogastropoda was reconstructed based on the mt, nuclear and combined data sets using ML and BI. After removing ambiguously aligned positions, the final mt amino acid matrix contained 3,091 sites. Both, ML (-lnL= 53530.051406) and BI (-lnL= 61906.69 for run1; -lnL= 61907.30 for run2) analyses based on the mt dataset arrived at the same tree topology, and only showed differences in branch lengths and levels of support (Fig. 3). The reconstructed tree recovered two highly supported clades, one consisting of Patellogastropoda and Heterobranchia, and the other including Vetigastropoda as sister group of Neritimorpha and Caenogastropoda (Fig. 3). Within Caenogastropoda, the superfamily Vermetoidea was recovered as sister group of the remaining caenogastropod superfamilies (Fig. 3). The reconstructed tree placed *I. nautili* deep within Caenogastropoda, as sister group of Tonnoidea and Neogastropoda (Muricoidea and Conoidea) (Fig. 3) with high statistical support (83% bootstrap; 0.98 posterior probability).

The nuclear matrix was analyzed at the nucleotide level, and had a final length of 1,297 sites. The ML (-lnL= 7530.363630) and BI (-lnL= 7631.83 for run1; -lnL= 7632.10 for run2) analyses based on the nuclear dataset arrived at the same tree topology, and only showed differences in branch lengths and levels of support (Fig. 4). Patellogastropoda was placed as sister group of the remaining gastropods with high support (Fig. 4). Neritimorpha was recovered as sister group of Vetigastropoda whereas Caenogastropoda was grouped with Heterobranchia (Fig. 4). The statistical supports of both clades were low. The genus *Ifremeria* was placed within Caenogastropoda, although internal relationships within this group were unresolved (Fig. 4).

The combined data set had a final alignment of 4,388 positions. In this case, the ML (-ln L= 61607.400080) and BI (-ln L= 69984.47 for run1; -ln L= 69987.43 for run2) inferred topologies differed only in the relative position of Vetigastropoda (affecting the only two nodes in the tree without strong statistical support) (Fig. 5). Both trees recovered Patellogastropoda as sister group of Heterobranchia. In the ML tree, Vetigastropoda and Neritimorpha form a clade (with low support), which is the sister group of Caenogastropoda (70% bootstrap). In the BI tree, Vetigastropoda is recovered unusually as sister group of the remaining gastropods (whose relative relationships are the same as in the ML tree) but without support (not shown). In both trees, *Ifremeria* is recovered deep within Caenogastropoda as sister group of the superfamily Tonnoidea and Neogastropoda (superfamilies Muricoidea and Conoidea) (Fig. 5).

Phylogenetic analyses reconstructed remarkably different topologies depending on the dataset analyzed. In recent days, Patellogastropoda is considered the sister group of the remaining gastropods (e.g., ([Ponder and Lindberg, 1997](#)), and as such is placed in the nuclear-based phylogeny. Instead, the mt-based and combined phylogenies recover Patellogastropoda as sister group of Heterobranchia with maximal support ([Grande et al., 2008](#); [Arques et al., 2014](#)). It has been suggested that this result could be due to a long-branch attraction (LBA) phenomenon ([Grande et al., 2008](#); [Stöger and Schrödl, 2013](#)). However, *Lottia* has a long branch, regardless of the marker, and the effect of the LBA would be to push the branch of Patellogastropoda next to the root not to Heterobranchia. Therefore, the unusual position of Patellogastropoda has to be associated to a mt-specific feature, which could be

the extraordinary level of gene rearrangement of the mt genome of *Lottia*, which could be introducing e.g. strand biases ([Hassanin et al., 2005](#)). Resolution of this particular node would require the sequencing of the complete mt genomes of other representatives of the other clades within Patellogastropoda (e.g. *Patella*). Another traditionally recognized hypothesis of gastropods is that suggesting a sister group relationship of Caenogastropoda and Heterobranchia (Apogastropoda) as proposed by ([Ponder and Lindberg, 1997](#)) based on morphology. This clade is recovered in the nuclear-based phylogeny, although without statistical support. Instead, the mt-based and combined phylogenies recover a highly supported clade including Caenogastropoda, Vetigastropoda, and Neritimorpha, which would be in agreement with the sharing of the relative position of the *trnC* gene by the three groups. Whether Neritimorpha and Vetigastropoda are sister groups (nuclear and combined, this study; ([Arques et al., 2014](#); [Williams et al., 2014](#)) or not (mt, this study; ([Aktipis and Giribet, 2010](#); [Castro and Colgan, 2010](#)) would remain an open question.

While the monophyly of Caenogastropoda is strongly supported both by morphological and molecular studies (e.g. ([Ponder and Lindberg, 1997](#); [Strong, 2003](#); [Colgan et al., 2007](#); [Williams et al., 2014](#)) as well as by our reconstructed trees, phylogenetic interrelationships within the clade at the family and superfamily levels remain largely unresolved, with many higher taxa probably being paraphyletic or even polyphyletic ([Colgan et al., 2007](#); [Cunha et al., 2009](#); [Williams et al., 2014](#)). In this regard, the mt-based and combined phylogenies recover Littorinimorpha as paraphyletic with respect to Neogastropoda with high statistical support. Moreover, the nuclear-based phylogeny finds Littorinimorpha to be polyphyletic (due to a closer relationship of Tonnoidea to Neogastropoda),

although with low statistical support. The relatively basal position of Vermetoidea in all trees is against morphological phylogenies (e.g., ([Strong, 2003](#)) and previous molecular studies ([Colgan et al., 2007](#)), and needs to be taken with caution due to the long branch of the vermetid genus *Dendropoma* and major gene order rearrangement of its mtDNA ([Rawlings et al., 2010](#)). Altogether, these results prompt for a revision of the Hypsogastropoda (Littorinimorpha+Neogastropoda) using both complete mt genomes and several nuclear genes, and with an extended taxon sampling that includes representatives of all superfamilies within the group (thus far, most reported complete mt genomes of caenogastropods have concentrated mainly on Neogastropoda (Cunha et al., 2009) with few representatives of more basal lineages such as Vermetidae-Vermetoidea (Rawling et al., 2010), or Pomatiopsidae-Rissooidea (Zhao et al., 2010)).

Ever since its description, the family Provannidae was postulated to belong to the Caenogastropoda, although with uncertain phylogenetic position ([Warén and Ponder, 1991](#); [Warén and Bouchet, 2001](#); [Bouchet and Rocroi, 2005](#)). According to our results, *Ifremeria* is nested deep within the Hypsogastropoda among the superfamilies that constitute the Littorinimorpha evolutionary grade. These results would be in agreement with a previous 16S-based phylogeny ([Colgan et al., 2007](#)) and observed affinities of *Provanna* and Littorinoidea at the sperm ultrastructure ([Healy, 1990](#)).

3.3 Divergence times

The reconstructed relaxed timetree (Fig. 6) placed the origin of gastropods about 517 MYA. The earliest fossil assigned to mollusks is *Kimberella*

([Fedonkin and Waggoner, 1997](#)) from the Late Precambrian that dates back to 558-555 MYA. Hence, our results support that the early diversification of the main classes of mollusks (including Gastropoda) occurred right after the origin of the phylum. The diversification of the crown group of Caenogastropoda started about 238 MYA, and a radiation pattern (accelerated rates of diversification) is detected between 172-140 MYA when the main superfamilies of derived caenogastropods originated. The divergence of the Abysochrysoidea from other caenogastropods occurred about 154 (144-168) MYA. This is in agreement with the oldest reliable fossil record of Abysochrysoidea, *Acanthostrophia acanthica* from Italy that traces back to the Middle Jurassic (Bajocian; 168-170 MYA) ([Kaim et al., 2008](#)). The earliest fossils of the families Provannidae and Hokkaidoconchidae are also found in seep deposits that date back to the Late Jurassic ([Kaim et al., 2008](#); [Kiel et al., 2008](#); [Kiel, 2010](#)). Overall, the position of *Ifremeria* in the reconstructed molecular phylogeny and the divergence time corresponding to that node in the timetree support that the origin of the Abysochrysoidea is relatively modern (Late Mesozoic-Early Cenozoic), in agreement with the fossil record ([Little and Vrijenhoek, 2003](#)) and within the range (93-228 MYA) inferred based on universal mt and nuclear fragments ([Johnson et al., 2010](#)). In this regard, it has been proposed that most of the deep-sea fauna went extinct during the Cenomanian-Turonian anoxic event in the Early Cretaceous, and that modern vent and seep faunas occupied the empty niches, expanding from shallower water refugia ([Jacobs and Lindberg, 1998](#)).

4. Acknowledgements

We are indebted to Hans Tøre Rapp for the gift of a specimen of *I. nautiliei*. We thank Silvia Perea for support with the BEAST software and Etienne Kornobis for help with the assembly of the mtDNA. Federico Abascal provided insightful comments on a previous version of the manuscript. We thank David A. Ray and two anonymous reviewers for their constructive suggestions. We are grateful for the technical assistance provided by J.M. García-Lobo, M. Cruz-Rodriguez, and I. Varela of the Servicio de Secuenciación Masiva (Universidad de Cantabria-IBBTEC) for the next-generation library preparation and handling. We are indebted to J. Marco and L. Cabellos who provided access to the supercomputer Altamira at the Institute of Physics of Cantabria (IFCA-CSIC), member of the Spanish Supercomputing Network, for performing phylogenetic analyses. This work was supported by the Spanish Ministry of Science and Innovation (CGL2007-60954 and CGL2010-18216 to RZ; BES-2008-009562 to DO).

5. References

- Abascal, F., Posada, D., Zardoya, R., 2007. MtArt: a new model of amino acid replacement for Arthropoda. *Mol. Biol. Evol.* 24, 1-5.
- Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21, 2104-2105.
- Aktipis, S.W., Giribet, G., 2010. A phylogeny of Vetigastropoda and other "archaeogastropods": re-organizing old gastropod clades. *Invert. Biol.* 129, 220-240.
- Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Arques, 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.* In press.
- Bandel, K., 1997. Higher classification and pattern of evolution of the Gastropoda. A synthesis of biological and paleontological data. *Cour.Forsch. Sencken.* 201, 57-81.
- Bandel, K., Heidelberger, D., 2002. A Devonian member of the subclass Heterostropha (Gastropoda) with valvatoid shell shape. *Neues Jahr. Geol. P-M* 2002, 533-550.
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsche, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* 69, 313-319.
- Boore, J., 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* 27, 1767 - 1780.
- Boore, J., Brown, W., 1994. Complete DNA sequence of the mitochondrial genome of the black chiton, *Katharina tunicata*. *Genetics* 138, 423 - 443.
- Boore, J., Brown, W., 2000. Mitochondrial genomes of *Galathea*, *Helobdella*, and *Platynereis*: sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. *Mol. Biol. Evol.* 17, 87 - 106.
- Boore, J., Macey, J., Medina, M., 2005. Sequencing and comparing whole mitochondrial genomes of animals. *Meth. Enzymol.* 395, 311 - 348.
- Bouchet, P., Rocroi, J.P., 2005. Classification and Nomenclator of Gastropod Familie. *Malacologia* 47, 397.
- Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540-552.
- Castro, L.R., Colgan, D.J., 2010. The phylogenetic position of Neritimorpha based on the mitochondrial genome of *Nerita melanotragus* (Mollusca: Gastropoda). *Mol. Phylogenet. Evol.* 57, 918-923.
- Colgan, D.J., Ponder, W.F., Beacham, E., Macaranas, J., 2007. Molecular phylogenetics of Caenogastropoda (Gastropoda: Mollusca). *Mol. Phylogenet. Evol.* 42, 717-737.
- Cunha, R., Grande, C., Zardoya, R., 2009. Neogastropod phylogenetic relationships based on entire mitochondrial genomes. *BMC Evol. Biol.* 9, 210.

- Drummond, A., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214.
- Fedonkin, M.A., Waggoner, B.M., 1997. The Late Precambrian fossil *Kimberella* is a mollusc-like bilaterian organism. *Nature* 388, 868-871.
- Felsenstein, J., 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* 17, 368-376.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783-791.
- Grande, C., Templado, J., Cervera, J.L., Zardoya, R., 2002. The complete mitochondrial genome of the nudibranch *Roboastra europaea* (Mollusca: Gastropoda) supports the monophyly of opisthobranchs. *Mol. Biol. Evol.* 19, 1672-1685.
- Grande, C., Templado, J., Zardoya, R., 2008. Evolution of gastropod mitochondrial genome arrangements. *BMC Evol. Biol.* 8, 61.
- Hassanin, A., Léger, N., Deutsch, J., 2005. Evidence for Multiple Reversals of Asymmetric Mutational Constraints during the Evolution of the Mitochondrial Genome of Metazoa, and Consequences for Phylogenetic Inferences. *Syst. Biol.* 54, 277-298.
- Hatzoglou, E., Rodakis, G., Lecanidou, R., 1995. Complete sequence and gene organization of the mitochondrial genome of the land snail *Albinaria coerulea*. *Genetics* 140, 1353 - 1366.
- Healy, J.M., 1990. Taxonomic affinities of the deep-sea genus *Provanna* (Caenogastropoda): new evidence from sperm ultrastructure. *J. Mollusc. Stud.* 56, 119-122.
- Hidaka, H., Watanabe, H., Kano, Y., Kojima, S., 2012. Mitochondrial genome rearrangement in a hydrothermal vent-endemic lineage of provannid gastropods provides a new DNA marker for phylogeographical studies. *J. Mar. Biol. Assoc. UK* 93, 1053-1058.
- Houbrick, R.S., 1979. Classification and systematic relationships of the Aabysochrysidae, a relict family of bathyal snails (Prosobranchia: Gastropoda), Smithsonian Institution Press, Washington DC.
- Huelsenbeck, J., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754-755.
- Jacobs, D.K., Lindberg, D.R., 1998. Oxygen and evolutionary patterns in the sea: Onshore/offshore trends and recent recruitment of deep-sea faunas. *Proc. Natl. Acad. Sci. USA* 95, 9396-9401.
- Johnson, S.B., Warén, A., Lee, R.W., Kano, Y., Kaim, A., Davis, A., Strong, E.E., Vrijenhoek, R.C., 2010. *Rubyspira*, New Genus and Two New Species of Bone-Eating Deep-Sea Snails With Ancient Habits. *Biol. Bull.* 219, 166-177.
- Kaim, A., 2004. The evolution of conch ontogeny in Mesozoic open sea gastropods, Instytut Paleobiologii PAN im. Romana Kozłowskiego, Warszawa :.
- Kaim, A., Jenkins, R.G., Warén, A., 2008. Provannid and provannid-like gastropods from the Late Cretaceous cold seeps of Hokkaido (Japan) and the fossil record of the Provannidae (Gastropoda: Aabysochrysoidea). *Zool. J. Linn. Soc.* 154, 421-436.
- Kaim, A., Tucholke, B.E., Warén, A., 2012. A new Late Pliocene large provannid gastropod associated with hydrothermal venting at Kane Megamullion, Mid-Atlantic Ridge. *J. Syst. Palaeontol.* 10, 423-433.

- Katoh, K., Misawa, K., Kuma, K.i., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059-3066.
- Kiel, S., 2008. An Unusual New Gastropod from an Eocene Hydrocarbon Seep in Washington State. *J. Paleontol.* 82, 188-191.
- Kiel, S., 2010. The Fossil Record of Vent and Seep Mollusks, in: Kiel, S. (Ed.), *The Vent and Seep Biota*. Springer Netherlands, pp. 255-277.
- Kiel, S., Campbell, K.A., Elder, W.P., Little, C.S.T., 2008. Jurassic and Cretaceous gastropods from hydrocarbon seeps in forearc basin and accretionary prism settings, California. *Acta Palaeontol. Polon.* 53, 679-703.
- Kiel, S., Little, C.T.S., 2006. Cold-Seep Mollusks Are Older Than the General Marine Mollusk Fauna. *Science* 313, 1429-1431.
- Knudsen, B., Kohn, A.B., Nahir, B., McFadden, C.S., Moroz, L.L., 2006. Complete DNA sequence of the mitochondrial genome of the sea-slug, *Aplysia californica*: Conservation of the gene order in Euthyneura. *Mol. Phylogenet. Evol.* 38.
- Kojima, S., Segawa, R., Fujiwara, Y., Hashimoto, J., Ohta, S., 2000. Genetic Differentiation of Populations of a Hydrothermal Vent-Endemic Gastropod, *Ifremeria nautilei*, between the North Fiji Basin and the Manus Basin revealed by Nucleotide Sequences of Mitochondrial DNA. *Zool. Sci.* 17, 1167-74.
- Kurabayashi, A., Ueshima, R., 2000. Complete Sequence of the Mitochondrial DNA of the Primitive Opisthobranch Gastropod *Pupa strigosa*: Systematic Implication of the Genome Organization. *Mol. Biol. Evol.* 17, 266-277.
- Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Mol. Biol. Evol.* 29, 1695-1701.
- Laslett, D., Canbäck, B., 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* 24, 172-175.
- Little, C.T.S., Vrijenhoek, R.C., 2003. Are hydrothermal vent animals living fossils? *Trends Ecol. Evol.* 18, 582-588.
- Lopez Sanchez, M.I.G., Mercer, T.R., Davies, S.M.K., Shearwood, A.-M.J., Nygård, K.K.A., Richman, T.R., Mattick, J.S., Rackham, O., Filipovska, A., 2011. RNA processing in human mitochondria. *Cell Cycle* 10, 2904-2916.
- Maynard, B., Kerr, L., McKiernan, J., Jansen, E., Hanna, P., 2005. Mitochondrial DNA sequence and gene organization in Australian backup abalone *Haliotis rubra* (leach). *Mar. Biotechnol. (NY)* 7, 645 - 658.
- McArthur, A.G., Koop, B.F., 1999. Partial 28S rDNA Sequences and the Antiquity of Hydrothermal Vent Endemic Gastropods. *Mol. Phylogenet. Evol.* 13, 255-274.
- McArthur, A.G., Tunnicliffe, V., 1998. Relics and antiquity revisited in the modern vent fauna. *Geol.Soc., Lond., Special Publications* 148, 271-291.
- Newman, W.A., 1985. The abyssal hydrothermal vent invertebrate fauna: A glimpse of antiquity?, in: Jones, M.L. (Ed.), *The hydrothermal vents of the eastern Pacific: An overview.* . Bull. Biol. Soc. Washington, pp. 231-242.

- Peek, A.S., Gustafson, R.G., Lutz, R.A., Vrijenhoek, R.C., 1997. Evolutionary relationships of deep-sea hydrothermal vent and cold-water seep clams (Bivalvia: Vesicomidae): results from the mitochondrial cytochrome oxidase subunit I. *Mar. Biol.* 130, 151-161.
- Plazzi, F., Ribani, A., Passamonti, M., 2013. The complete mitochondrial genome of *Solemya velum* (Mollusca: Bivalvia) and its relationships with Conchifera. *BMC Genomics* 14, 409.
- Ponder, W.F., Colgan, D.J., Healy, J.M., Nützel, A., Simone, L.R.L., Strong, E.E., 2008. Caenogastropoda, in: Ponder, W.F. and Lindberg, D.R. (Eds.), *Phylogeny and Evolution of the Mollusca*. Univ. California Press, California, pp. 331-383.
- Ponder, W.F., Lindberg, D.R., 1997. Towards a phylogeny of gastropod molluscs: an analysis using morphological characters. *Zool. J. Linn. Soc.* 119, 83-265.
- Rambaut, A., Drummond, A.J., 2007. Tracer v1.4, Available from <http://beast.bio.ed.ac.uk/Tracer>.
- Rawlings, T., MacInnis, M., Bieler, R., Boore, J., Collins, T., 2010. Sessile snails, dynamic genomes: gene rearrangements within the mitochondrial genome of a family of caenogastropod molluscs. *BMC Genomics* 11, 440.
- Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572-1574.
- Sasaki, T., Waren, A., Kano, Y., Okutani, T., Fujikura, K., 2010. Gastropods from recent hot vent and cold seeps: systematics, Diversity and life strategies, in: S., K. (Ed.), *The vent and seep biota, topics in geobiology*. Springer, Dordrecht, The Netherlands, pp. 169-254.
- Schattner, P., Brooks, A., Lowe, T., 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33, W686 - W689.
- Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863-864.
- Shank, T.M., Black, M.B., Halanych, K.M., Lutz, R.A., Vrijenhoek, R.C., 1999. Miocene Radiation of Deep-Sea Hydrothermal Vent Shrimp (Caridea: Bresiliidae): Evidence from Mitochondrial Cytochrome Oxidase Subunit I. *Mol. Phylogenet. Evol.* 13, 244-254.
- Simison, W., Lindberg, D., Boore, J., 2006. Rolling circle amplification of metazoan mitochondrial genomes. *Mol. Phylogenet. Evol.* 39, 562 - 567.
- St John, J., 2011. SeqPrep 1.1. Tripping Adaptors and/or Merging paired reads with overlap into Single Reads. Available online at: <https://github.com/jstjohn/SeqPrep>.
- Stamatakis, A., 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688-2690.
- Stöger, I., Schrödl, M., 2013. Mitogenomics does not resolve deep molluscan relationships (yet?). *Mol. Phylogenet. Evol.* 69, 376-392.
- Strong, E.E., 2003. Refining molluscan characters: morphology, character coding and a phylogeny of the Caenogastropoda. *Zool. J. Linn. Soc.* 137, 447-554.
- Suzuki, Y., Kojima, S., Sasaki, T., Suzuki, M., Utsumi, T., Watanabe, H., Urakawa, H., Tsuchida, S., Nunoura, T., Hirayama, H., Takai, K.,

- Nealson, K.H., Horikoshi, K., 2006. Host-symbiont relationships in hydrothermal vent gastropods of the genus *Alviniconcha* from the Southwest Pacific. *Appl. Environ. Microbiol.* 72, 1388-93.
- Thaler, A.D., Zelnio, K., Saleu, W., Schultz, T.F., Carlsson, J., Cunningham, C., Vrijenhoek, R.C., Van Dover, C.L., 2011. The spatial scale of genetic subdivision in populations of *Ifremeria nautilei*, a hydrothermal-vent gastropod from the southwest Pacific. *BMC Evol. Biol.* 11, 372.
- Tunnicliffe, V., 1992. The nature and origin of the modern hydrothermal vent fauna. *Palaios* 7, 338-350.
- Van Dover, C.L., 2000. The ecology of deep-sea hydrothermal vents, Princeton University Press, Princeton.
- Van Dover, C.L., German, C.R., Speer, K.G., Parson, L.M., Vrijenhoek, R.C., 2002. Evolution and Biogeography of Deep-Sea Vent and Seep Invertebrates. *Science* 295, 1253-1257.
- Warén, A., Bouchet, P., 2001. Gastropoda and Monoplacophora from hydrothermal vents and seeps; New taxa and records. *Veliger* 44, 116-231.
- Warén, A., Ponder, W.F., 1991. New species, anatomy, and systematic position of the hydrothermal vent and hydrocarbon seep gastropod family Provannidae fam.n. (Caenogastropoda). *Zool. Scr.* 20, 27-56.
- Williams, S.T., Foster, P.G., Littlewood, D.T.J., 2014. The complete mitochondrial genome of a turbinid vetigastropod from MiSeq Illumina sequencing of genomic DNA and steps towards a resolved gastropod phylogeny. *Gene* 533, 38-47.
- Wyman, S., Jansen, R., Boore, J., 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20, 3252 - 3255.
- Yang, J.-S., Lu, B., Chen, D.-F., Yu, Y.-Q., Yang, F., Nagasawa, H., Tsuchida, S., Fujiwara, Y., Yang, W.-J., 2013. When Did Decapods Invade Hydrothermal Vents? Clues from the Western Pacific and Indian Oceans. *Mol. Biol. Evol.* 30, 305-309.
- Yokobori, S.-i., Fukuda, N., Nakamura, M., Aoyama, T., Oshima, T., 2004. Long-Term Conservation of Six Duplicated Structural Genes in Cephalopod Mitochondrial Genomes. *Mol. Biol. Evol.* 21, 2034-2046.
- Zhao, Q.-P., Zhang, S.H., Deng, Z.-R., Jiang, M.-S., Nie, P., 2010. Conservation and variation in mitochondrial genomes of gastropods *Oncomelania hupensis* and *Tricula hortensis*, intermediate host snails of *Schistosoma* in China. *Mol. Phylogenet. Evol.* 57, 215-226.

Legends to figures

Fig. 1. Gene organization of the mitochondrial genome of *Ifremeria nautiliei* (Abyssochrysoidea). Genes encoded by the opposite strand are underlined. The intergenic regions longer than 10 positions are shown.

Fig. 2. Mitochondrial gene arrangements of main lineages of gastropods and a cephalopod. Note major gene rearrangements among main gastropod lineages and relatively stable gene orders (with changes restricted to tRNA genes) within them.

Fig. 3. Phylogenetic relationships of Gastropoda based on the mt data set. The ML phylogram is shown. Numbers at nodes are support values from ML (bootstrap proportions; above) and BI (posterior probabilities; below). Dashes (-) indicate values below 70% bootstrap support or 0.90 Bayesian posterior probabilities. Names of major clades of gastropods are shown in capitals, and superfamilies within Caenogastropoda are indicated. Scale bar is substitutions/ site.

Fig. 4. Phylogenetic relationships of Gastropoda based on the nuclear data set. The ML phylogram is shown. Numbers at nodes are support values from ML (bootstrap proportions; above) and BI (posterior probabilities; below). Dashes (-) indicate values below 70% bootstrap support or 0.90 Bayesian posterior probabilities. Names of major clades of gastropods are shown in capitals, and superfamilies within Caenogastropoda are indicated. Scale bar is substitutions/ site.

Fig. 5. Phylogenetic relationships of Gastropoda based on the combined data set. The ML phylogram is shown. Numbers at nodes are support values

from ML (bootstrap proportions; above) and BI (posterior probabilities; below). Dashes (-) indicate values below 70% bootstrap support or 0.90 Bayesian posterior probabilities. The BI tree differs only in the relative phylogenetic position of *Haliotis*. Names of major clades of gastropods are shown in capitals, and superfamilies within Caenogastropoda are indicated. Scale bar is substitutions/ site.

Fig. 6. Timetree with age estimates of major divergence events among gastropods, based on the mt data set, and using Bayesian relaxed dating methods (BEAST). The ingroup was constrained to be monophyletic with respect to the outgroup (*Octopus*). Horizontal bars represent 95% credibility intervals of relevant nodes, and calibration constraints are indicated on the corresponding nodes (A to E). Dates (and credibility intervals) are in millions of years.

Table 1. Mitochondrial genome and nuclear gene sequences analyzed in this study

GenBank ID		Taxa	Nuclear Genes Fragments				GenBank ID
Species	MitoGenome		18S	28S D1	28S D6	H3	
<i>Octopus vulgaris</i>	NC_006353	Cephalopoda	<i>Graneledone verrucosa</i>	<i>Octopus cyanaea</i>	<i>Granelodone pacifica</i>	<i>Octopus joubini</i>	AY557468; AJ310244; AY145407; AY377781
<i>Nerita melanotragus</i>	GU810158*	Neritimorpha	<i>Nerita atramentosa</i>	<i>Nerita atramentosa</i>	<i>Nerita atramentosa</i>	<i>Nerita atramentosa</i>	DQ916542; AY296885; DQ916495; AF033701
<i>Haliotis rubra</i>	NC_005940	Vetigastropoda	<i>Haliotis corrugata</i>	<i>Haliotis corrugata</i>	<i>Haliotis corrugata</i>	<i>Haliotis corrugata</i>	FJ977644; FJ977675; FJ977675; FJ977736
<i>Dendropoma maximum</i>	NC_014583	Vermetoidea	<i>Serpulorbis imbricatus</i>	<i>Serpulorbis sp.</i>	<i>Serpulorbis sp.</i>	<i>Serpulorbis sp.</i>	HQ833992; AY29689; DQ916475; AF033710
<i>Ifremeria nautiliei</i>	KC757644	Abyssochrysoidea	<i>Ifremeria nautiliei</i>	<i>Ifremeria nautiliei</i>	<i>Ifremeria nautiliei</i>	<i>Ifremeria nautiliei</i>	GQ290545; GQ290527; GQ290564; GQ290631
<i>Oncomelania hupensis</i>	NC_013073	Rissooidea	<i>Rissoina fasciata</i>	<i>Rissoina fasciata</i>	<i>Rissoina fasciata</i>	<i>Pisinna albizona</i>	DQ916528; DQ916576; DQ916474; DQ916448
<i>Cymatium parthenopeum</i>	NC_013247	Tonnoidea	<i>Tonna cerevesima</i>	<i>Tonna cerevesima</i>	<i>Tonna cerevesima</i>	<i>Tonna cerevesima</i>	DQ916534; DQ916582; DQ916483; DQ916453
<i>Bolinus brandaris</i>	NC_013250	Muricoidea	<i>Dicathais orbita</i>	<i>Dicathais orbita</i>	<i>Dicathais orbita</i>	<i>Dicathais orbita</i>	DQ916537; AY296898; DQ916487; AF033690
<i>Conus borgesii</i>	NC_013243	Conoidea	<i>Conus miles</i>	<i>Conus miles</i>	<i>Conus miles</i>	<i>Conus miles</i>	DQ916538; AY296900; DQ916490; AF033684
<i>Onchidella celtica</i>	NC_012376	Systellommatophora	<i>Onchidella floriana</i>	<i>Onchidium verruculatum</i>	<i>Onchidium verruculatum</i>	<i>Onchidium sp.</i>	AY427522; AY427487; DQ256742; AF033706
<i>Albinaria caerulea</i>	NC_001761	Stylommatophora	<i>Arion silvaticus</i>	<i>Arion silvaticus</i>	<i>Arion silvaticus</i>	<i>Albinaria puella</i>	AY145365; AY145392; AY145392; JX098162
<i>Pyramidella dolabrata</i>	NC_012435	Pyramidelloidea	<i>Boonea seminuda</i>	<i>Boonea seminuda</i>	<i>Boonea seminuda</i>	<i>Turbonilla lactea</i>	AY145367; AY145395; AY145395; EF133488
<i>Aplysia californica</i>	NC_005827	Aplysioidea	<i>Aplysia juliana</i>	<i>Aplysia juliana</i>	<i>Aplysia juliana</i>	<i>Aplysia juliana</i>	DQ916539; AY296905; U78644; AF033675
<i>Roboastra europaea</i>	NC_004321	Nudibranchia	<i>Dialula sp.</i>	<i>Hypselodoris infucata</i>	<i>Dialula sandigensis</i>	<i>Roboastra ricei</i>	AY145374; FJ917467; AY144352; HM162504
<i>Pupa strigosa</i>	NC_002176	Acteonoidea	<i>Pupa solidula</i>	<i>Pupa solidula</i>	<i>Rictaxis punctocaelatus</i>	<i>Pupa solidula</i>	AY427516; AY427481; FJ917243; EF133483
<i>Lottia digitalis</i>	NC_007782	Patellogastropoda	<i>Lottia gigantea</i>	<i>Lottia gigantea</i>	<i>Lottia gigantea</i>	<i>Lottia gigantea</i>	FJ977632; FJ977662; FJ977662; FJ977725

* almost complete

Gastropoda	Cephalopoda			<i>Octopus vulgaris</i>	cox1	cox2	D	atp8	atp6	<u>F</u>	<u>nad5</u>	<u>H</u>	<u>nad4</u>	<u>nad4L</u>	T	<u>S</u>	<u>cob</u>	<u>nad6</u>	<u>P</u>	<u>nad1</u>	<u>L</u>	<u>L</u>	<u>rrnL</u>	<u>V</u>	<u>rrnS</u>	<u>M</u>	<u>C</u>	<u>Y</u>	<u>W</u>	<u>Q</u>	<u>G</u>	<u>E</u>	cox3	K	A	R	N	I	nad3	S	nad2
	Neritimorpha			<i>Nerita melanotragus</i>	cox1	cox2	D	atp8	atp6	<u>F</u>	<u>nad5</u>	<u>H</u>	<u>nad4</u>	<u>nad4L</u>	T	<u>S</u>	<u>cob</u>	<u>nad6</u>	<u>P</u>	<u>nad1</u>	<u>L</u>	<u>L</u>	rrnL	<u>V</u>	<u>rrnS</u>	<u>M</u>	<u>Y</u>	<u>C</u>	<u>W</u>	<u>Q</u>	<u>G</u>		cox3	<u>K</u>	<u>A</u>	<u>R</u>	<u>N</u>	<u>I</u>	nad3	S	nad2
	Vetigastropoda			<i>Haliotis rubra</i>	cox1	cox2	atp8	atp6	<u>F</u>	<u>nad5</u>	<u>H</u>	<u>nad4</u>	<u>nad4L</u>	T	<u>S</u>	<u>cob</u>	<u>nad6</u>	<u>P</u>	<u>nad1</u>	<u>L</u>	<u>L</u>	<u>rrnL</u>	<u>V</u>	<u>rrnS</u>	<u>M</u>	<u>Y</u>	<u>C</u>	<u>W</u>	<u>Q</u>	<u>G</u>	<u>E</u>	cox3	D	K	A	R	I	nad3	N	S	nad2
	Caenogastropoda	Littorinimorpha	Vermetoidea	<i>Dendropoma maximum</i>	cox1	D	atp8	atp6	N	nad5	K	A	cox3	S	nad2	R	<u>Y</u>	<u>M</u>	<u>C</u>	<u>W</u>	<u>Q</u>	<u>G</u>	<u>E</u>	rrnS	V	rrnL	L	L	nad1	P	nad6	cob	S	<u>T</u>	nad4L	nad4	H	F	I	nad3	cox2
		Unassigned	Abyssochrysoidea	<i>Ifremeria nautilei</i>	cox1	cox2	D	atp8	atp6	<u>M</u>	<u>W</u>	<u>Y</u>	<u>C</u>	<u>Q</u>	<u>G</u>	<u>E</u>	rrnS	V	rrnL	L	L	nad1	P	nad6	cob	S	<u>T</u>	nad4L	nad4	H	nad5	F	cox3	K	A	R	N	I	nad3	S	nad2
		Littorinimorpha	Rissooidea	<i>Oncomelania hupensis</i>	cox1	cox2	D	atp8	atp6	<u>M</u>	<u>Y</u>	<u>C</u>	<u>W</u>	<u>Q</u>	<u>G</u>	<u>E</u>	rrnS	V	rrnL	L	L	nad1	P	nad6	cob	S	<u>T</u>	nad4L	nad4	H	nad5	F	cox3	K	A	R	N	I	nad3	S	nad2
		Littorinimorpha	Tonnoidea	<i>Cymatium parthenopeum</i>	cox1	cox2	D	atp8	atp6	<u>M</u>	<u>Y</u>	<u>C</u>	<u>W</u>	<u>Q</u>	<u>G</u>	<u>E</u>	rrnS	V	rrnL	L	L	nad1	P	nad6	cob	S	<u>T</u>	nad4L	nad4	H	nad5	F	cox3	K	A	R	N	I	nad3	S	nad2
		Neogastropoda	Muricoidea	<i>Bolinus brandaris</i>	cox1	cox2	D	atp8	atp6	<u>M</u>	<u>Y</u>	<u>C</u>	<u>W</u>	<u>Q</u>	<u>G</u>	<u>E</u>	rrnS	V	rrnL	L	L	nad1	P	nad6	cob	S	<u>T</u>	nad4L	nad4	H	nad5	F	cox3	K	A	R	N	I	nad3	S	nad2
		Neogastropoda	Conoidea	<i>Comus borgesi</i>	cox1	cox2	D	atp8	atp6	<u>M</u>	<u>C</u>	<u>Y</u>	<u>W</u>	<u>Q</u>	<u>G</u>	<u>E</u>	rrnS	V	rrnL	L	L	nad1	P	nad6	cob	S	<u>T</u>	nad4L	nad4	H	nad5	F	cox3	K	A	R	N	I	nad3	S	nad2
	Heterobranchia	(Panpulmonata;Eupulmonata)	Systellommatophora	<i>Onchidella celitica</i>	cox1	V	rrnL	L	A	P	nad6	nad5	nad1	nad4L	cob	D	C	F	cox2	Y	W	G	H	<u>Q</u>	<u>L</u>	<u>atp8</u>	<u>N</u>	<u>atp6</u>	<u>R</u>	<u>E</u>	<u>rrnS</u>	<u>M</u>	<u>nad3</u>	<u>S</u>	S	nad4	<u>T</u>	<u>cox3</u>	I	nad2	K
		(Panpulmonata; Eupulmonata)	Stylommatophora	<i>Albinaria caerulea</i>	cox1	V	rrnL	L	A	nad6	P	nad5	nad1	nad4L	cob	D	C	F	cox2	Y	W	G	H	<u>Q</u>	<u>L</u>	<u>atp8</u>	<u>N</u>	<u>atp6</u>	<u>R</u>	<u>E</u>	<u>rrnS</u>	<u>M</u>	<u>nad3</u>	<u>S</u>	<u>T</u>	<u>cox3</u>	S	nad4	I	nad2	K
		(Panpulmonata)	Pyramidelloidea	<i>Pyramidella dolabrata</i>	cox1	V	rrnL	L	A	P	nad6	nad5	nad1	nad4L	cob	K	F	D	cox2	W	C	<u>Q</u>	<u>atp6</u>	<u>R</u>	<u>L</u>	<u>G</u>	<u>H</u>	<u>atp8</u>	<u>N</u>	<u>E</u>	<u>rrnS</u>	<u>M</u>	<u>nad3</u>	<u>S</u>	S	nad4	<u>T</u>	<u>cox3</u>	I	nad2	Y
		(Euopisthobranchia)	Aplysiomorpha	<i>Aplysia californica</i>	cox1	V	rrnL	L	A	P	nad6	nad5	nad1	Y	W	nad4L	cob	D	F	cox2	G	H	<u>Q</u>	<u>L</u>	<u>atp8</u>	<u>N</u>	C	<u>atp6</u>	<u>R</u>	<u>E</u>	<u>rrnS</u>	<u>M</u>	<u>nad3</u>	<u>S</u>	S	nad4	<u>T</u>	<u>cox3</u>	I	nad2	K
		(Nudipleura)	Nudibranchia	<i>Roboastra europaea</i>	cox1	V	rrnL	L	A	P	nad6	nad5	nad1	Y	W	nad4L	cob	D	F	cox2	G	H	C	<u>Q</u>	<u>L</u>	<u>atp8</u>	<u>N</u>	<u>atp6</u>	<u>R</u>	<u>E</u>	<u>rrnS</u>	<u>M</u>	<u>nad3</u>	<u>S</u>	S	nad4	<u>T</u>	<u>cox3</u>	I	nad2	K
(Lower Heterobranchia)		Acteonoidea	<i>Pupa strigosa</i>	cox1	V	rrnL	L	A	P	nad6	nad5	nad1	Y	W	nad4L	cob	D	F	cox2	G	H	<u>Q</u>	<u>L</u>	<u>atp8</u>	<u>N</u>	C	<u>atp6</u>	<u>R</u>	<u>E</u>	<u>rrnS</u>	<u>M</u>	<u>nad3</u>	<u>S</u>	S	nad4	<u>T</u>	<u>cox3</u>	I	nad2	K	
Patellogastropoda			Lottiioidea	<i>Lottia digitalis</i>	cox1	N	cox3	F	V	cox2	E	T	P	rrnL	M	rrnS	Y	<u>cob</u>	K	atp8	Q	nad4L	nad4	L	I	<u>R</u>	nad2	W	<u>nad3</u>	<u>atp6</u>	<u>A</u>	<u>G</u>	D	nad5	S	L	<u>H</u>	<u>nad1</u>	C	nad6

-LnL														
MITO (aa)					RAXML		MrBayes							
Scheme Name: groups per function					53530,0514		61920,16							
Subset	Best Model	Subset Partitions	alpha	pinvar										
1	MtArt+G	ATP6, ATP8	1,04E+00											
2	LG+I+G+F	COX1, COX2, COX3	7,34E-01	2,03E-01										
3	MtArt+I+G+F	CYTB	1,46E+00	2,24E-01										
4	MtArt+I+G+F	ND1, ND2, ND3, ND4, ND4L, ND5, ND6	1,51E+00	1,88E-01										
Nuclear (nt)														
Scheme Name: Separated					7530,36363		7652,523							
Subset	Best Model	Subset Partitions	alpha	pinvar	r(A<->C)	r(A<->G)	r(A<->T)	r(C<->G)	r(C<->T)	r(G<->T)	pi(A)	pi(C)	pi(G)	pi(T)
1	TIMef+G	18S	4,07E-01		1,28E-01	1,93E-01	5,78E-02	1,08E-01	4,13E-01	1,00E-01	2,68E-01	2,46E-01	2,68E-01	2,19E-01
2	TrN+G	28S_d1	3,37E-01		4,83E-02	1,62E-01	1,16E-01	8,10E-02	4,93E-01	9,96E-02	2,19E-01	2,64E-01	3,30E-01	1,87E-01
3	TrN+G	28S_d6	5,31E-01		8,43E-02	1,52E-01	8,99E-02	9,76E-02	5,09E-01	6,76E-02	2,28E-01	2,56E-01	3,45E-01	1,72E-01
4	GTR+I+G	H3	7,11E-01	6,28E-01	1,44E-01	2,90E-01	2,45E-01	1,77E-02	2,84E-01	1,92E-02	2,45E-01	3,13E-01	1,81E-01	2,60E-01
MITO (aa) + Nuclear (nt)														
Scheme Name: groups per function + separated					61607,4001		70007,973							