Revised version: July 5th, 2016

**Phylogenetic relationships among superfamilies of Neritimorpha
(Mollusca: Gastropoda)**

Juan E. Uribe¹, Don Colgan², Lyda R. Castro³, Yasunori Kano⁴, and
Rafael Zardoya^{1*}

¹Museo Nacional de Ciencias Naturales (MNCN-CSIC), José Gutiérrez
Abascal 2, 28006, Madrid, Spain

²AMRI, The Australian Museum, 1 William Street, Sydney 2010,
Australia

³Grupo de y Ecología Molecular, Universidad
del Magdalena, Santa Marta, Colombia

⁴Atmosphere and Ocean Research Institute, The University of Tokyo, 5-
1-5 Kashiwanoha, Kashiwa, Chiba 277-8564, Japan.

*Corresponding author: rafaz@mncn.csic.es

ABSTRACT

Despite the extraordinary morphological and ecological diversity of Neritimorpha, few studies have focused on the phylogenetic relationships of this lineage of gastropods, which includes four extant superfamilies: Neritopsoidea, Hydrocenoidea, Helicinoidea, and Neritoidea. Here, the nucleotide sequences of the complete mitochondrial genomes of *Georissa bangueyensis* (Hydrocenoidea), *Neritina usnea* (Neritoidea), and *Pleuropoma jana* (Helicinoidea) and the nearly complete mt genomes of *Titiscania* sp. (Neritopsoidea), and *Theodoxus fluviatilis* (Neritoidea) were determined. Phylogenetic reconstruction using probabilistic methods were based on mitochondrial (13 protein coding genes and two ribosomal rRNA genes), nuclear (partial 28S rRNA, 18S rRNA, actin, and histone H3 genes) and combined sequence data sets. All phylogenetic analyses except one converged on a single, highly supported tree in which Neritopsoidea was recovered as the sister group of a clade including Helicinoidea as the sister group of Hydrocenoidea and Neritoidea. This topology agrees with the fossil record and supports at least three independent invasions of land by neritimorph snails. The mitochondrial genomes of *Titiscania* sp., *G. bangueyensis*, *N. usnea*, and *T. fluviatilis* share the same gene organization previously described for *Nerita* mt genomes whereas that of *P. jana* has undergone major rearrangements. We sequenced about half of the mitochondrial genome of another species of Helicinoidea, *Viana regina* and confirmed that this species shares the highly derived gene order of *P. jana*.

Keys words: Mitogenomic phylogeny, Helicinoidea, Hydrocenoidea, Neritoidea, Neritopsoidea.

1. Introduction

With approximately 2,000 living species (Fukumori and Kano, 2014; Richling, 2014), the Neritimorpha—also known as Neritopsina—forms a long-recognized and distinct lineage of Gastropoda that appears to be the remnant of an early radiation (Ponder and Lindberg, 1997; Lindberg, 2008). The earliest undisputed fossils of Neritimorpha date back to the Carboniferous (Kaim and Sztajner, 2005). However, its origin could be much older if platyceratid snails from the Ordovician to Devonian periods (Bandel and Frýda, 1999) are interpreted as stem lineages of the group (Frýda et al., 2008, 2009). Despite its current relatively low species richness, the group is characterized by an astonishing diversity of morphologies (besides spiral or conical shells, these snails have also evolved limpet- and slug-like forms) and the adaptation to very different habitats, thus resembling at a smaller scale the extraordinary morphological and ecological diversity achieved by all gastropods (Kano et al., 2002; Lindberg, 2008). Of marine evolutionary origin, they are currently found in deep-sea hydrothermal vents and seeps, as well as in submarine caves, but have particularly radiated in the intertidal and shallow subtidal zones of tropical regions (Warén and Bouchet, 2001; Frey and Vermeij, 2008; Kano and Kase, 2008). Additionally, several groups have independently invaded brackish, surface or underground freshwater and terrestrial (including arboreal) habitats (Kano et al., 2002; Lindberg, 2008; Schilthuizen et al., 2012; Fukumori and Kano, 2014).

Bouchet et al. (2005) classified the living species of Neritimorpha into four superfamilies: Neritopsoidea, Hydrocenoidea, Helicinoidea, and Neritoidea. The superfamily Neritopsoidea includes the marine families Neritopsidae and Titiscaniidae. The extant Neritopsidae are represented exclusively by a few species in the type genus *Neritopsis* from submarine caves and similar cryptic voids, but this family presents the

oldest fossil record among living neritimorph superfamilies (Kano et al., 2002; Kaim and Sztajner, 2005). The Titiscaniidae, also represented by a few species in the single genus *Titiscania*, have adapted to life in even smaller interstices by evolving a slug-like phenotype with complete loss of the shell. Because of their highly divergent morphology, titiscaniids were initially placed in their own superfamily Titiscanoidea (Bergh, 1890), but more recent studies suggest that they originated in the early Cenozoic as a specialized lineage within Neritopsidae or even within *Neritopsis* (Kano et al., 2002). The superfamily Hydrocenoidea has one family, Hydrocenidae, which includes two widely-accepted genera (*Hydrocena* and *Georissa*) with presumably over a hundred terrestrial species in forest litter, on limestone crops or deep inside karstic caves (Schilthuizen et al., 2012). The superfamily Helicinoidea includes three terrestrial families, namely Helicinidae, Proserpinidae and Proserpinellidae (Thompson, 1980; Richling, 2004, 2014), as well as the aquatic family Neritiliidae, whose members live in freshwater streams, underground waters (e.g., *Neritilia* and *Platynrerita*; Kano and Kase, 2004) and subtidal marine caves (e.g., *Pisulina* and *Siaesella*; Kano and Kase, 2002, 2008). Finally, the superfamily Neritoidea includes the Phenacolepadidae and Neritidae. The former family is mainly composed of limpet-like species that inhabit tidal flats and mangrove swamps (e.g., *Phenacolepas* and *Plesiothyreus*; Fretter, 1984), shipworm tunnels in sunken wood (*Phenacolepas*; Kano et al., 2013), or deep-sea hydrothermal vent fields (*Shinkailepas* and *Olgasoralis*; Warén and Bouchet, 2001). A monotypic genus *Bathynrerita* from methane seeps in the Gulf of Mexico also belongs to Phenacolepadidae, regardless of its retention or re-acquisition of a coiled shell (Holthuis, 1995; Fukumori and Kano, 2014). The family Neritidae (the nerites) has radiated into brackish and freshwater habitats, and includes also the genus *Neritodryas*, which is partially terrestrial or even arboreal although some species retain

99 planktotrophic early development (Little, 1990; Kano and Kase, 2002). Several genera
100 within Neritidae such as *Nerita*, *Clithon* and *Neritina* comprise numerous species
101 (Holthuis, 1995; Frey and Vermeij, 2008).

102 Only a few studies have specifically addressed relationships within Neritimorpha
103 (Holthuis, 1995; Kano et al., 2002). The most complete phylogenetic studies on the
104 internal relationships of the group were based on 57 morphological characters
105 (Holthuis, 1995) and the partial sequences of the 28S rRNA gene (Kano et al., 2002).
106 The monophyly of Neritimorpha is widely accepted and supported by numerous
107 morphological synapomorphies (Ponder and Lindberg, 1997). However, only characters
108 related to the reproductive system might be very useful for determining the
109 phylogenetic relationships among the superfamilies (Kano et al., 2002). The 28S rRNA
110 molecular phylogeny recovered Neritopsoidea as the sister group of the remaining three
111 superfamilies. Among these, Neritoidea and Helicinoidea were sister groups to the
112 exclusion of Hydrocenoidea (Kano et al., 2002).

113 Thus far, the mitochondrial (mt) genomes available for Neritimorpha, all
114 belonging to the genus *Nerita*, are only nearly complete, lacking the putative control
115 region (Castro and Colgan, 2010; Arquez et al., 2014). These mt genomes have been
116 used as representatives of Neritimorpha to determine its relative phylogenetic position
117 within Gastropoda, and have allowed the description of the mt genome order of *Nerita*,
118 which is similar to the inferred ancestral gastropod genome organization (Osca et al.,
119 2014a) and nearly identical (except on a tRNA gene transposition) to the order found in
120 some non-gastropod molluscs such as the cephalopod *Octopus vulgaris* Cuvier, 1797
121 (Yokobori et al., 2004). In this study, we aimed to (1) sequence the complete or almost
122 complete mt genomes of representatives of all superfamilies of Neritimorpha, (2)
123 determine whether the mt genome organization described in *Nerita* is common to the

other neritimorph lineages, (3) reconstruct the phylogenetic relationships of the living superfamilies, (4) compare the mt-based phylogeny with the nuclear-based phylogeny of Kano et al. (2002), and (5) to infer several evolutionary trends within the group, which may be at the origin of its extraordinary morphological and ecological diversity.

2. Material and Methods

2.1. Samples and DNA extraction

In order to have representation of the four neritimorph superfamilies, one specimen each of *Georissa banguyensis* Smith, 1895 (Hydrocenoidea), *Neritina usnea* (Röding, 1798) and *Theodoxus fluviatilis* (Linnaeus, 1758) (Neritoidea), *Pleuropoma jana* (Cox, 1872) and *Viana regina* (Morelet, 1849) (Helicinoidea), and *Titiscania* sp. (Neritopsoidea) was used for this study (See Table 1, for details on the locality, collector, and voucher ID of each sample). Except *P. jana*, which was held at -80° C after collection, all samples were stored in 100% ethanol and total genomic DNA was isolated from up to 30-60 mg of foot tissue following a standard phenol/ chloroform extraction.

2.2. PCR amplification and sequencing

The mt genomes (except *P. jana*, which was sequenced at the Australian Museum) were PCR amplified following a three-step approach. First, fragments of the *cox1*, *rrnL* and *cox3* genes were PCR amplified using the primers respectively detailed in Folmer et al. (1994), Palumbi et al. (1991) and Boore and Brown (2000). The standard PCR reaction contained 2.5 µl of 10x buffer, 1.5 µl of MgCl₂ (25 mM), 0.5 µl of dNTP (2.5 mM each), 0.5 µl of forward primer (10 mM), 0.5-1 µl (20-100 ng) of template DNA, 0.2 µl of Taq DNA polymerase, 5 PRIM (Hb, rg, G, r, y) and 1.5 µl of distilled water.

water up to 25 μ l. The following program was used: denaturing step at 94°C for 60 s; 45 cycles of denaturation at 94°C for 30 s, annealing using 44°C (*cox1*) and 52°C (*rrnL* and *cox3*) for 60 s and extension at 72°C for 90 s; a final extension step at 72°C for 5 min. Second, the amplified PCR fragments were sequenced using Sanger sequencing, and new primers were designed (see Appendix for primer sequences) for amplifying outwards from the short fragments in the next step. Third, the remaining mtDNA was amplified in two overlapping fragments by long PCR using the newly designed primers. The long PCR reaction contained 2.5 μ l of 10x LA Buffer II (Mg^{2+} plus), 3 μ l of dNTPs (2.5 mM each), 0.5 μ l of each primer (10 mM), 0.5-1 μ l (20-100 ng) of template DNA, 0.2 μ l TaKaRa LA Taq DNA polymerase (5 units/ μ l), and sterilized distilled water up to 25 μ l. The following PCR conditions were used: a denaturing step at 94°C for 60 s; 45 cycles of denaturation at 98°C for 10 s, annealing using a gradient of 50-65°C for 30 s and extension at 68°C for 60 s per kb; and a final extension step at 68°C for 12 min.

Long-PCR products were purified by ethanol precipitation. Overlapping fragments from the same mt genome were pooled together in equimolar concentrations, and subjected to massive parallel sequencing. For each mt genome, an indexed library was constructed using the NEXTERA XT DNA library prep Kit (Illumina, San Diego, CA, USA) at AllGenetics (A Coruña, Spain). The constructed libraries were run in an Illumina HiSeq2000 (100 bp Paired-ends) at Macrogen (Seoul, Korea).

For the *P. jana* mt genome, next generation sequencing was performed on a CTAB DNA extraction (Saghai-Marooft et al., 1984) of one individual. RNase A treatment was used after the pellet from the initial precipitation with ethanol was re-suspended. This solution was re-extracted once with phenol/chloroform/*iso*-amyl alcohol 25:24:1 and once with chloroform/*iso*-amyl alcohol 24:1) following which DNA

was precipitated from the supernatant with an equal volume of *iso*-propanol. The pellet was washed with 70% ethanol and air dried.

In addition, the following partial nuclear genes were amplified from all species (except *P. jana*) with standard PCR reactions as described above for partial mitochondrial genes but using an annealing temperature of 50°C in all cases: 28S rRNA gene using primers LSU5 F (Littlewood et al., 2000) and LSU1600 R (Williams et al., 2003); 18S rRNA gene using primers 1F (Winnepeninckx et al., 1995) and 1100R (Williams et al., 2003); *actin* gene using primers FA 2 (5'-ATCTGGCATCACACCTTCTACAAC-3') and RA 2 (5'-ACAGTGTTTRGGRTACAAGTCTTTACG-3'); and *histone H3* gene using primers H3MF (5'-ATGGCTCGTACCAAGCAGACTGG-3') and H3MR (5'-TGGATGTCCTTGGGCATGATTGTTAC-3') (Colgan et al. (1998). Each amplified fragment was sequenced using the Sanger technology (Table 2).

Contigs containing the individual nuclear genes were identified in *P. jana* by using known neritimorph sequences to search the CLC Genomic assembly (see below) using FASTA3, ver. 3.5 (Pearson, 1999).

2.3. Genome assembly and annotation

The reads corresponding to the different mt genomes were sorted using the indexes. Adapter sequences were removed using SeqPrep (StJohn, 2011). Assembly was performed using the TRUFA webserver (Kornobis et al., 2015). The quality (randomness) of the sequencing was checked using FastQC v.0.10.1 (Andrews, 2010). Reads were trimmed and filtered out according to their quality scores using PRINSEQ v.0.20.3 (Schmieder and Edwards, 2011). Filtered reads were used for *de novo* assembly of mt genomes, searching for contigs with a minimum length of 3 kb. The complete

circular sequence of each mt genome was finally assembled by overlapping the various contigs in Sequencher 5.0.1. The assembled sequence was used as reference to map the original (raw) reads with a minimum identity of 99% using Geneious® 8.0.3.

For *P. jana*, sequencing was performed at the Australian Genomics Research Facility. After quality-checking with FastQC version 0.10.1 (Andrews, 2010), the raw reads were assembled using the CLC Genomics Workbench. A variety of parameter values were examined and the assembly provided by the set (word = 20, bubble = 30, algorithm = fast) that gave the longest value for N50 was used. Contigs representing the mt genome were identified by FASTA3, ver. 3.5 searches on a local installation (Pearson, 1999) and confirmed by using them as test sequences in BLAST searches of GenBank.

The new neritimorph mt genomes were annotated using the MITOS (Bernt et al., 2013) and DOGMA (Wyman et al., 2004) webserver. The 13 mt protein-coding genes were annotated by identifying their open reading frames using the invertebrate mt code. 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 gastropod mt genomes, and assumed to extend to the boundaries of adjacent genes (Boore et al., 2005).

2.4. Sequence alignment

Three different sequence data sets were constructed. The first one, hereafter designated the mt data set, was built aligning the newly determined mt genomes of Neritimorpha with the orthologous sequences from four *Nerita* species available in NCBI (www.ncbi.nlm.nih.gov/). Mt genomes of four Caenogastropoda, five

Vetigastropoda, and one Neomphalina were used as gastropod outgroups (Table 1). No Heterobranchia were included in the present analyses because of their well-known high mitochondrial evolutionary rates that result in a long-branch attraction artefact (Grande et al., 2008; Stöger and Schrödl, 2013; Arquez et al., 2014; Uribe et al., 2016). This mt data set included the nucleotide sequence alignments of the two mt rRNA genes, which were combined either with the nucleotide (excluding third codon positions to avoid saturation) or the deduced amino acid sequences of the 13 mt protein coding genes. The second data set, hereafter the nuclear data set, included two ribosomal (28S *rRNA* and 18S *rRNA*) and two protein-coding (*actin* and *histone H3*) genes. The nucleotide sequences for these genes produced in this work were aligned with orthologous sequences available in NCBI for Neritimorpha, Caenogastropoda and Vetigastropoda (Table 2). The third data set, named the combined data set, concatenated the mitogenomic and nuclear DNA sequence data. Protein-coding genes were aligned separately using Translator X (Abascal et al., 2010) whereas rRNA genes were aligned separately using MAFFT v7 (Katoh and Standley, 2013) with default parameters. Ambiguously aligned positions were removed using Gblocks v.0.91b (Castresana, 2000) with the following settings: minimum sequence for flanking positions: 85%; maximum contiguous non-conserved positions: 8; minimum block length: 10; gaps in final blocks: no. The generated single alignments were concatenated into the three data matrices using the ALTER webserver (Glez-Peña et al., 2010).

2.5. Phylogenetic analyses

Phylogenetic relationships were reconstructed using maximum likelihood (ML; Felsenstein, 1981) and Bayesian inference (BI; Huelsenbeck and Ronquist, 2001). ML analyses were conducted with RAxML v7.3.1 (Stamatakis, 2006) and default

parameters using the rapid hill-climbing algorithm and 10,000 bootstrap pseudoreplicates. BI analyses were conducted using MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003) and default parameters, running four simultaneous Monte Carlo Markov chains (MCMC) for 10 million generations, sampling every 1,000 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 BI runs were performed to increase the chance of adequate mixing by the MCMC and to increase the chance of detecting failure to converge, as determined using Tracer v1.6 (Rambaut et al. 2014).

The best partition schemes and best-fit models of substitution for the three data sets were identified using Partition Finder and Protein Partition Finder (Lanfear et al., 2012) according to the Bayesian information criterion (BIC; Schwarz, 1978). For mt protein-coding genes, the partitions tested were: all genes combined, all genes separated except *atp6-atp8* and *nad4-nad4L*, and genes grouped by subunits (*atp*, *cox*, *cob* and *nad*; see Appendix). In addition, at the nucleotide level the three above partition schemes were tested considering first and second codon positions separated. For the mt rRNA genes, the two genes combined or separated were tested. For the nuclear data set, the partitions tested were: all genes combined, all genes separated, and protein-coding genes grouped versus rRNA genes grouped. Partitions considering the three codon positions were also tested for the protein-coding nuclear genes. The best partitions selected for the mt and nuclear data sets were also used in the combined data set (see Appendix).

To check whether the presence of long branches was biasing the output of the BI analyses performed on the mt data set (with protein-coding genes analyzed at the amino acid level; see results), BI using the site-heterogeneous mixture CAT model (Lartillot

and Philippe, 2004) was performed as implemented in PhyloBayes MPI v.1.5 (Lartillot et al., 2013). The site-heterogeneous mixture CAT model assumes that the different sites of a protein evolve under distinct substitution processes and has proven to be less sensitive to (and to alleviate) long-branch attraction biases in some instances (Lartillot et al., 2007). BI was performed with 10 independent MCMC chains until convergence, sampling every cycle. Due to restrictions in the program, the protein coding genes (at the amino acid level) and the rRNA genes were analyzed separately under the best-fit CAT-GTR model, using the discrete gamma approximation to model among-site rate heterogeneity. The performance of the CAT-GTR+G model was assessed using a 10-fold cross-validation performed on subsamples of 6,000 non-constant positions randomly drawn from the original matrices. Convergence of analyses was checked a posteriori using the convergence tools implemented in PhyloBayes (maxdiff < 0.125, maximum discrepancy < 0.1 and effective size > 100; see Appendix). Posterior probabilities provided branch support for BI analyses.

2.6. Estimation of divergence times

The divergence dates between neritimorph clades were estimated using the mt data set at the nucleotide level and an uncorrelated relaxed molecular clock model in BEAST 1.7 (Drummond and Rambaut, 2007). This software was used to infer branch lengths and nodal ages on a fixed topology with regards to neritimorph internal phylogenetic relationships (the reconstructed ML/BI tree topology in Fig. 2). For the clock model, we selected the lognormal relaxed-clock model, which allows rates to vary among branches without any *a priori* assumption of correlation between adjacent branches. For the tree prior, we employed a Yule process of speciation. We employed

the partitions selected by Partition Finder (see appendix 2; selected models were simplified to HKY85 to allow convergence). The final Markov chain was run twice, each for 120 million generations, sampling every 10,000 generations with the first 1,200 of the sampled generations discarded as the burn-in, after confirming convergence of chains with Tracer v.1.5. (Rambaut and Drummond, 2007). The effective sample size of the majority of the parameters was > 200 (except the gamma distribution and the proportion of invariable site of *cox1*, *cox2* and *cox3* genes, first and second positions).

The tree was time-calibrated by setting the ages of (1) the first split within the Neritimorpha, i.e., between *Titiscania* sp. and other study taxa of the clade, and (2) the first split within *Nerita*, i.e., between *N. melanotragus* and three other congeners. The first calibration point was set at a minimum of 235 million years ago (Mya) with a 95% upper limit of 251 Mya (Gamma distribution, Shape: 1, Offset: 235, Scale: 5.3) based on the Triassic records of neritimorphs lacking the internal whorls of the teleoconch (Bandel and Frýda, 1999; Kaim and Sztajner, 2005). The retention of internal coiling is considered to be plesiomorphic and seen in *Neritopsis* (Neritopsoidea) alone among the extant Neritimorpha (Bandel and Frýda, 1999; Kano et al., 2002). The minimum bound was set by referring to the age of the Cassian Formation of the Italian Alps (Lower Carnian). This formation has yielded a variety of non-neritopsoid taxa that might have given rise to the Recent Neritoidea (Bandel, 2007).

The second calibration point, the divergence date of the lineage leading to *N. melanotragus*, was constrained at a minimum age of 49 Mya with a 95% upper limit of 56 Mya (Gamma distribution, Shape: 1, Offset: 49, Scale: 2.3) by referring to the fossil record and a previous molecular phylogeny of the genus *Nerita* (Frey and Vermeij, 2008). The crown group representatives of the genus have appeared consistently throughout the record since the late Paleocene and many of the extant subgenera existed

already in the Lutetian age (40–49 Mya; see Symonds, 2009 for figures of fossil specimens). *Nerita melanotragus* represents one of the earliest offshoots among the extant species of the genus (Frey and Vermeij, 2008).

3. Results

3.1. Sequencing and assembly

The nucleotide sequences of the complete mt genomes of *Georissa banguyensis*, *Neritina usnea*, and *Pleuropoma jana* and the nearly complete mt genomes of *Titiscania* sp., and *Theodoxus fluviatilis* were determined (see Table 1 for lengths and Genbank accession numbers). In addition, about half of the mtDNA of *Viana regina* (7,634 bp) was sequenced to determine whether other Helicinoidea share the derived gene order of *Pleuropoma*. The HiSeq2000 sequencer produced a similar amount of data for *Titiscania* sp. (580,418 reads; 79.2 Mb), *T. fluviatilis* (786,306 reads; 602.2 Mb), and *V. regina* (509,154 reads; 137.2 Mb). However, fewer data (7,584 reads; 2 Mb) were produced for *G. banguyensis*. All these samples were run in lanes together with TruSeq RNA libraries (from other projects). Interestingly, *N. usnea* produced more data (10,113,794 reads; 2.74 Gb) and it was run in a lane together with Nextera DNA libraries (from other projects). After assembly, the average coverage for each mt genome was: 42x for *G. banguyensis*; 32,921x for *N. usnea*; 79x for *Titiscania* sp.; 602x for *T. fluviatilis*; 621x for *P. jana*; and 137x for *V. regina*. In the mt genomes of *Titiscania* sp. and *T. fluviatilis*, a region of less than 1kb including a cluster of tRNA genes and the putative control region could not be PCR amplified (see Appendix). The lengths and Genbank accession numbers for the nucleotide sequences of the partial nuclear genes of *Georissa banguyensis*, *Titiscania* sp, *Pleuropoma jana*, *Gibbula*

rarilineata (Michaud, 1829), and *Clanculus cruciatus* (Linnaeus, 1758) that were determined in this study are shown in Table 2.

3.2. Structural features and mitochondrial organization

The complete mt genomes of *G. banguueyensis*, *N. usnea*, and *P. jana* contain 13 protein coding, two ribosomal RNA, and 22 transfer RNA genes (Fig. 1). In addition, the mt genome of *P. jana* has an extra *trnM* gene. The gene arrangement of the mt genomes of *G. banguueyensis* and *N. usnea* is identical: the major strand encodes *cox1-3*, *atp6-8*, *nad3*, *nad2*, *trnD*, *trnT*, *trnS*(GCU), and the KARNI (*trnK*, *trnA*, *trnR*, *trnN*, and *trnI*) cluster whereas the minus strand encodes the remaining protein-coding genes (*nad5*, *nad4*, *nad4L*, *cob*, *nad6*, and *nad1*), the two ribosomal RNA genes (*rrnS* and *rrnL*), the *trnF*, *trnH*, *trnS*(UGA), *trnP*, *trnL*(UAA), *trnL*(UAG) genes and the MYCWQGE (*trnM*, *trnY*, *trnC*, *trnW*, *trnQ*, *trnG*, and *trnE*) cluster (Fig. 1). The nearly complete mt genomes of *Titiscania* sp. and *T. fluviatilis* generally have the same gene order as *N. usnea* and *G. banguueyensis* (but note that we failed to sequence the MYCWQGE tRNA gene cluster and the control region in the two former species). The gene order of *P. jana* is markedly different from that shared by the above mentioned four species. All genes are encoded by the major strand except the *trnT* gene. This may have been due to a single inversion event that affected the following genes: *trnF*, *nad5*, *trnH*, *nad4*, *nad4L*, *trnY*, *trnS*, *cob*, *nad6*, *trnP*, *nad1*, *trnL*(UAA), *trnL*(UAG), *rrnL*, *trnV*, *rrnS*, *trnM*, and *trnY* (Fig. 1). In addition, this mt genome shows independent rearrangements (inversions and translocations) of the *trnL*(UAG), *trnW*, *trnQ*, *trnG*, and *trnE* genes. The sequenced half of the mt genome of *V. regina* shares the same order of *P. jana*, except in the cluster KARNI, where *trnN* and *trnI* genes were translocated (KARIN; see Fig. 1).

The genes *nad4/nad4L* overlapped by seven bp in all mt genomes where this region was sequenced (all but *V. regina*). All determined protein-coding genes start their open reading frame with the codon ATG, except *nad3* and *nad4* of *P. jana* that start with ATA and TTG, respectively (see Appendix). The most commonly used stop codon was TAA. The *atp8* and *cox3* genes of *G. banguueyensis*, the *cox2* and *nad4L* genes of *N. usnea*, the *atp8*, *cox3*, *nad1*, *nad4*, and *nad6* genes of *Titiscania* sp., the *cox2* and *nad6* genes of *P. jana*, the *cox3* and *nad3* genes of *T. fluviatilis*, and the *atp6* and *cox1* genes of *V. regina* end with TAG (see Appendix). Other genes end with incomplete stop codons either TA- as the *nad3* and *nad5* genes of *G. banguueyensis* or T-- as the *cox2*, *nad4* and *nad2* genes of *G. banguueyensis*, the *cob*, *nad3*, *nad5*, and *nad6* genes of *P. jana*, the *nad4* and *nad2* genes of *N. usnea*, the *nad2* gene of *Titiscania* sp., the *nad2* and *nad5* genes of *T. fluviatilis*, and the *nad3* and *nad2* genes of *V. regina* (see Appendix). In the complete mt genomes, the longest intergenic regions were found between the *trnE* and *cox3* genes in *G. banguueyensis* (608 bp) and *N. usnea* (471 bp) and between the *trnF* and *trnL*(UAG) genes in *P. jana* (783 bp).

3.3. Phylogenetic relationships

The phylogenetic relationships of Neritimorpha were reconstructed based on the mt, nuclear, and combined data sets using probabilistic methods (Figs. 2 and 3). The alignment of the mt data set including protein-coding genes analyzed at the nucleotide level (only first and second codon positions) plus rRNA genes was 7,604 positions long. The ML ($-\ln L = 63,623.8$) and BI ($-\ln L = 63,529.6$ for run 1; $-\ln L = 63,528.9$ for run 2) phylogenetic analyses arrived at the same topology (Fig. 2) by using *Neomphalina* to root the tree following Uribe et al. (2016). ML and BI reconstructed trees both showed high statistical support for all nodes but one referring to an internal relationship within

the genus *Nerita* (Fig. 2). Neritimorpha was recovered monophyletic with maximal statistical support, and within Neritimorpha, the species *Titiscania* sp. representing the superfamily Neritopsoidea was recovered as the sister group of the remaining neritimorph superfamilies. The next clade that branched off within Neritimorpha included the species *P. jana* and *V. regina* as representatives of the superfamily Helicinoidea. The superfamilies Hydrocenoidea (*G. banguyensis*) and Neritoidea were recovered as sister groups. Within Neritoidea, *Neritina* and *Theodoxus* formed a clade that was the sister group of *Nerita*. Both Helicinoidea and Hydrocenoidea showed relatively long branches (Fig. 2). The protein-coding genes of the mt data set were also analyzed at the amino acid level combined with the rRNA genes. The final matrix was 4,484 positions long. The ML ($-\ln L = 57,096.7$) and BI ($-\ln L = 67,830.4$ for run 1; $-\ln L = 67,830.7$ for run 2) trees recovered similar topologies that only differed in the relative phylogenetic positions of superfamilies Hydrocenoidea and Helicinoidea. While ML recovered the topology shown in Figure 2, BI grouped Hydrocenoidea and Helicinoidea together with maximal support (see Appendix). However, BI analyses using the site-heterogeneous mixture CAT model as implemented in PhyloBayes recovered the topology shown in Figure 2.

The alignment of the nuclear data set was 2,962 positions long. The ML ($-\ln L = 10,841.8$) and BI ($-\ln L = 10,971.6$ for run1; $-\ln L = 10,973.3$ for run2) phylogenetic analyses arrived at the same topology (Fig. 3, inset). Vetigastropoda were used as outgroup since the analyzed nuclear genes were not available for Neomphalina. Within Neritimorpha, Neritopsoidea was recovered as the sister group of Helicinoidea plus Hydrocenoidea and Neritoidea (Fig. 3, inset). All nodes in this tree had maximal support in ML and BI, except that grouping Hydrocenoidea and Neritoidea (BP: 67%; PP: 0.94).

Interestingly, *G. bangueyensis* but not *P. jana* showed a long branch mainly due to accelerated substitution rates in the 28S rRNA gene.

The matrix of the combined data set was 10,566 positions long. The ML (-lnL = 49,365.4) and BI (-lnL = 49,355.5 for run1; -lnL = 49,355.8 for run2) phylogenetic analyses arrived at the same topology, with all nodes receiving maximal support in both analyses (Fig. 3).

According to the estimated divergence times (Fig. 4), the superfamily Helicinoidea branched off at 234 Mya (95% credible interval: 206–269 Mya) from the lineage leading to the Hydrocenoidea and Neritoidea. The split between the latter two superfamilies was estimated to have occurred at around 180 Mya (139–221 Mya) in the Jurassic period. Within-family divergences of the studied taxa were suggested to have started in the Late Cretaceous and the boundary with the Paleocene time (Helicinidae: 72 Mya; Neritidae: 65 Mya).

4. Discussion

4.1. Gene arrangement

Until the present work, no complete mt genome of any Neritimorpha was available, although up to four almost complete mt genomes lacking only the control region between the *trnE* and *cox3* genes had been determined for the genus *Nerita* (Castro and Colgan, 2010; Arquez et al., 2014). Here, we present three complete and two nearly complete mt genomes that together represent all known living superfamilies of Neritimorpha. The mt genomes of *Titiscania* sp., *G. bangueyensis*, *N. usnea*, and *T. fluviatilis* share the same gene organization previously described for *Nerita* mt genomes (Castro and Colgan, 2010; Arquez et al., 2014), indicating that three (Neritoidea, Hydrocenoidea, and Neritopsoidea) out of the four superfamilies have not undergone

major gene rearrangements during their evolutionary history. Moreover, this genome organization is very similar to the one found in Vetigastropoda (Williams et al., 2014; Uribe et al., 2016) with only minor differences in the relative positions of several tRNA genes (marked in red in Fig. 1), nearly identical (except on a tRNA gene transposition) to the gene order reported for the cephalopod *Octopus vulgaris* (Yokobori et al., 2004; Fig. 1), and only differing in a translocation of the *trnD* gene and an inversion of the *trnP* gene with respect to the mtDNA of *Katharina tunicata* (Wood, 1815) (Polyplacophora; Boore and Brown, 1994).

In contrast, the mt genome of *P. jana* (Helicinoidea) shows a markedly different gene order, which has been likely originated through a single inversion event that affected 17 genes. A second *trnM* gene in the upstream border of the inversion (Fig. 1) could represent a remnant of the rearrangement process. Moreover, in this mt genome the MYCWQGE cluster found in other gastropod mt genomes (Uribe et al., 2016) has been inverted, disrupted (e.g., the *trnG* gene has been inverted and translocated between the *cox2* and *trnD* genes) and rearranged. The sequencing of half the mt genome of *V. regina* confirmed that the peculiar gene order of *P. jana* including the large inversion, the inversion and translocation of the *trnG* gene, and the duplication of the *trnM* gene, is likely shared by all helicininids (Fig. 1). Interestingly, almost the same set of genes is involved in a gene rearrangement in the common ancestor of Caenogastropoda, where the integrity of the MYCWQGE cluster was nevertheless maintained (Cunha et al., 2009; Osca et al., 2014b, 2015). It is also noteworthy that the branch leading to the two sequenced species of Helicinoidea is the longest in the phylogenetic tree, indicating a significant acceleration of substitution rates. As in other reported cases (Shao et al., 2003; Stöger and Schrödl, 2013), this increase in substitution rate is correlated with an increase in rearrangement rates and a consequent major reorganization.

472

473 ***Phylogenetic relationships and radiation of Neritimorpha***

474 All phylogenetic analyses performed in this study except one converged on the
475 same topology, which received high statistical support. The only exception (BI based on
476 the mt data set with protein-coding genes analyzed at the amino acid level) may be
477 considered a spurious result related to a long-branch attraction artefact between
478 Hydrocenoidea and Helicinoidea, which, in fact, was ameliorated when the data was
479 analyzed using the site-heterogeneous mixture CAT model (Lartillot and Philippe, 2004;
480 Lartillot et al., 2007). The monophyly of Neritimorpha is undisputed (Kano et al., 2002;
481 Lindberg, 2008) and also supported by our phylogenetic analyses. Within Neritimorpha,
482 the Neritopsoidea is recovered in our phylogenetic analyses as the sister group of the
483 other three superfamilies. This result is in agreement with morphological evidence
484 (Holthuis, 1995) as well as with the only available molecular phylogeny of
485 Neritimorpha, which was based on the 28S rRNA gene (Kano et al., 2002). The
486 Neritopsoidea have the oldest fossil record of all living neritimorph superfamilies,
487 dating back at least to the Triassic and probably earlier (Bandel, 2000, 2007; Kaim and
488 Sztajner, 2005), and show important conchological and anatomical characteristics. The
489 shell of Neritopsoidea retains internal whorls as well as the central axis of coiling,
490 which have been reabsorbed in the other superfamilies (Solem, 1983; Bandel and Frýda,
491 1999). Among many other anatomical discrepancies, the male pallial gonoduct of
492 Neritoposidea has open lamellae, which have been fused ventrally closing the duct in
493 the other superfamilies (Holthuis, 1995; Kano et al., 2002; Kano and Kase, 2002).

494 With regards to the phylogenetic relationships among the remaining three
495 superfamilies, our results conflict with those in the previous 28S rRNA phylogeny
496 (Kano et al., 2002). While our present analyses place *Georissa* (as a representative of

Hydrocenoidea) in a sister position to Neritoidea, the 28S rRNA phylogeny recovered the genus as the second earliest offshoot of the Recent Neritimorpha (Kano et al., 2002). However, the latter relationship was only moderately supported in their ML analysis (Kano et al., 2002), due to a long branch leading to *Georissa* that biased the phylogenetic reconstruction. Actually, the addition of the other extant genus of the superfamily, *Hydrocena*, renders the 28S rRNA topology quite the same as the present topology based on mt genomes (Y. Kano and I. Richling, unpublished data).

Our divergence time estimates based the molecular data and two fossil-based calibration points suggested a Jurassic origin of the Hydrocenoidea, in accordance with the disjunct geographic distribution of its living taxa; the species of *Hydrocena* occur in southern Europe and on eastern Atlantic islands while *Georissa* has a wide Indo-West Pacific distribution, hence possibly suggesting a Pangaeian origin for the family (Bandel, 2000). The estimated date of the divergence does not contradict with the assumption that *Schwardtina cretacea* (Tausch, 1856) from fresh- or brackish-water deposits of the Late Cretaceous (Bandel and Riedel, 1994) constitutes the earliest representative of Hydrocenoidea (Kano et al., 2002). This species shows an intriguing resemblance to the extant Hydrocenidae in the protoconch and teleoconch shapes, irrespective of its aquatic habitat (Bandel, 2000; Fukumori and Kano, 2014). The earliest Neritoidea that can be unambiguously classified into extant genera also appeared in the Late Cretaceous or in the Paleocene (Bandel and Kiel, 2003; Frey and Vermeij, 2008). Although a number of older Mesozoic taxa have been assigned to Neritidae based solely on the plesiomorphic hemispherical shell (Kano et al., 2002), we suggest based on the time-calibrated phylogeny that neritids diverged in the Cretaceous period and subsequently colonized freshwater habitats.

521 The present topology further emphasizes the phenotypic plasticity in neritimorph
522 evolution. The astonishing diversity seen in the shell morphology and anatomy of the
523 group is evidently the outcome of their major ecological radiation (Kano et al., 2002;
524 Lindberg, 2008). The comparison of the morphological traits and even the identification
525 of character homology are particularly difficult between the aquatic and terrestrial taxa,
526 due to drastic changes in the body plan to adapt to a life in the air (Little, 1990; Kano
527 and Kase, 2002). Based on their 28S rRNA phylogeny, Kano et al. (2002) proposed that
528 the evolutionary history of the four superfamilies could only be traced morphologically
529 by different conditions of the female reproductive system: the Neritopsoidea and
530 Hydrocenoidea exhibit a monaulic condition with a single gonopore, which might
531 represent a plesiomorphic character state, while the others have independent vaginal and
532 ootype openings, which are separately located in the Helicinoidea and adjoin in the
533 Neritoidea (e.g. Bourne, 1911; Fretter, 1984; Holthuis, 1995; Kano and Kase, 2002;
534 Richling, 2004). The presently recovered sister relationship between the monaulic
535 Hydrocenoidea and diaulic Neritoidea clearly rejects such a simple, gradual increase of
536 complexity in the female reproductive system. To our knowledge, none of distribution
537 patterns of morphological traits adequately explains the relationships among the extant
538 superfamilies of Neritimorpha as reconstructed in our phylogeny.

539 In any case, our results agree with those of Kano et al. (2002) and with
540 morphological studies (Bourne, 1911; Haszprunar, 1988; Little, 1990) in supporting that
541 the transition from marine to terrestrial mode of life, despite being complex processes
542 involving numerous adaptations at morphological, physiological and ecological levels,
543 has occurred more than twice convergently within the evolutionary history of
544 Neritimorpha. Besides the independent terrestrialization of Hydrocenoidea,
545 Helicinoidea and *Neritodryas*, the Paleozoic *Dawsonella meeki* most probably

represents an early offshoot of Neritimorpha as one of the first terrestrial gastropods (Bourne, 1911; Solem, 1983). The Carboniferous occurrence of the species, however, long precedes the presumed Triassic split between the Neritopsoidea and the other extant superfamilies (Bandel and Frýda, 1999; Kaim and Sztajner, 2005; Bandel, 2007), hence exemplifying another line of land invasion (Bourne, 1911; Kano et al., 2002).

5. Concluding remarks

Phylogenetic information contained in complete mt genome sequences is useful in resolving phylogenetic relationships within major lineages of gastropods. Here we sequenced complete mtDNAs of representatives of the four superfamilies of Neritimorpha and recovered a highly supported phylogenetic tree at this taxonomic level that is congruent with nuclear-based inferences. Hence, the reconstructed phylogeny provides a statistically robust framework for a variety of comparative studies within the group as exemplified here by the inference of multiple invasions of land. Future mt genome phylogenetic studies incorporating a denser taxon sampling at the family level will additionally contribute to a more refined understanding of the evolutionary processes that generated the astonishing morphological diversity and habitat adaptation observed in Neritimorpha.

Acknowledgements

We want to express our gratitude to Iker Irisarri (University of Konstanz) for his seminal work in this project. We are indebted to Yolanda Camacho (Universidad of Costa Rica) for her gift of several specimens of *Titiscania* sp.; José Templado (MNCN-CSIC) for the samples of *Theodoxus fluviatilis*; Jane Herrera Uria (Museo Nacional de Historia Natural de Cuba) and Thierry Backeljau (RBINS) for the individuals of *Viana*

regina; Moises Arquez (Universidad Simón Bolívar) for the samples of *Neritina usnea*, and Menno Schilthuizen (Naturalis Biodiversity Center) for the specimens of *Georissa bangueyensis*. We are also grateful to Jesús Marco and Luis 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 (CGL2010-18216 and CGL2013-45211-C2-2-P to RZ; BES-2011-051469 to JEU) and the Japan Society for the Promotion of Science (26291077 and 15H04412 to YK). We are grateful for the critical reading of the original version of the manuscript by two anonymous reviewers.

Appendix A. Supplementary material

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.
- Andrews, S., 2010. FastQC. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- 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. Genomics* 15, 45-54.
- Bandel, K., 2000. The new family Cortinellidae (Gastropoda, Mollusca) connected to a review of the evolutionary history of the subclass Neritimorpha. *Neues Jahrb Geol* P-A 217, 111-129.
- Bandel, K., 2007. Description and classification of Late Triassic Neritimorpha (Gastropoda, Mollusca) from the St Cassian Formation, Italian Alps. *Bull. Geosci.* 82, 215-274.
- Bandel, K., Frýda, J., 1999. Notes on the evolution and higher classification of the subclass Neritimorpha (Gastropoda) with the description of some new taxa. *Geol. Palaeontol.* 33, 219-235.
- Bandel, K., Kiel, S., 2003. Relationships of Cretaceous Neritimorpha (Gastropoda, Mollusca), with the description of seven new species. *Bull. Czech Geol. Surv.* 78, 53-65.
- Bandel, K., Riedel, F., 1994. The late Cretaceous gastropod fauna from Ajka (Bakony Mountains, Hungary): a revision. *Ann. Nat. Hist. Mus. Wien A* 96, 1-65.
- Bergh, R., 1890. Die Titiscanien, eine Familie der rhipidoglossen Gastropoden. *Morphol. Jahrb.* 16, 1-26.

608 Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Frittsch, G., Pütz, J.,
 609 Middendorf, M., Stadler, P.F., 2013. MITOS: Improved *de novo* metazoan
 610 mitochondrial genome annotation. Mol. Phylogen. Evol. 69, 313-319.
 611 Boore, J.L., Brown, W.M., 1994. Complete DNA sequence of the mitochondrial genome
 612 of the black chiton, *Katharina tunicata*. Genetics 138, 423-443.
 613 Boore, J.L., Brown, W.M., 2000. Mitochondrial genomes of *Galathealinum*,
 614 *Helobdella*, and *Platynereis*: sequence and gene arrangement comparisons indicate
 615 that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa.
 616 Mol. Biol. Evol. 17, 87-106.
 617 Boore, J.L., Macey, J.R., Medina, M., 2005. Sequencing and comparing whole
 618 mitochondrial genomes of animals. Methods Enzymol. 395, 311 - 348.
 619 Bouchet, P., Rocroi, J.P., Frýda, J., Hausdorf, B., Ponder, W.F., Valdés, Á., A., W.,
 620 2005. Classification and nomenclator of gastropod families. Malacologia 47, 1-397.
 621 Bourne, G.C., 1911. Contributions to the morphology of the group Neritacea of the
 622 aspidobranch gastropods. Part II. The Helicinidae. Proc. Zool. Soc. Lond. 1911,
 623 759-809.
 624 Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their
 625 use in phylogenetic analysis. Mol. Biol. Evol. 17, 540-552.
 626 Castro, L.R., Colgan, D.J., 2010. The phylogenetic position of Neritimorpha based on
 627 the mitochondrial genome of *Nerita melanotragus* (Mollusca: Gastropoda). Mol.
 628 Phylogen. Evol. 57, 918-923.
 629 Colgan, D.J., McLauchlan, A., Wilson, G.D.F., Livingston, S.P., Edgecombe, G.D.,
 630 Macaranas, J., Cassis, G., Gray, M.R., 1998. Molecular phylogenetics of the
 631 Arthropoda: relationships based on histone H3 and U2 snRNA DNA sequences.
 632 Aust. J. Zool. 46, 419-437.

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

635 Drummond, A., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by
636 sampling trees. BMC Evol. Biol. 7, 214.

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

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

642 Fretter, V., 1984. The functional anatomy of the neritacean limpet *Phenacolepas*
643 *omanensis* Biggs and some comparison with *Septaria*. J. Molluscan Stud. 50, 8-18.

644 Frey, M.A., Vermeij, G.J., 2008. Molecular phylogenies and historical biogeography of
645 a circumtropical group of gastropods (Genus: *Nerita*): Implications for regional
646 diversity patterns in the marine tropics. Mol. Phylogen. Evol. 48, 1067-1086.

647 Frýda, J., Nützel, A., Wagner, P., 2008. Paleozoic Gastropoda In: Ponder, W.F.,
648 Lindberg, D.R. (Eds.), Phylogeny and Evolution of the Mollusca. University of
649 California Press, Berkeley, pp. 239-270.

650 Frýda, J., Racheboeuf, P.R., Frýdová, B., Ferrová, L., Mergl, M., Berkyová, S., 2009.
651 Platyceratid gastropods—stem group of patellogastropods, neritimorphs or
652 something else? . Bull. Geosci. 84, 107-120.

653 Fukumori, H., Kano, Y., 2014. Evolutionary ecology of settlement size in planktotrophic
654 neritimorph gastropods. Mar. Biol. 161, 213–227.

655 Glez-Peña, D., Gómez-Blanco, D., Reboiro-Jato, M., Fdez-Riverola, F., Posada, D.,
656 2010. ALTER: program-oriented conversion of DNA and protein alignments.
657 Nucleic Acids Res. 38, W14-W18.

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

660 Haszprunar, G., 1988. On the origin and evolution of major gastropods group, with
661 special reference to the streptoneura. J. Molluscan Stud. 54, 367 - 441.

662 Holthuis, B.V., 1995. Evolution between marine and freshwater habitats: a case study of
663 the gastropod Neritopsina PhD thesis. University of Washington, Seattle.

664 Huelsenbeck, J., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogenetic trees.
665 Bioinformatics 17, 754-755.

666 Kaim, A., Sztajner, P., 2005. The opercula of neritopsid gastropods and their
667 phylogenetic importance. J. Molluscan Stud. 71, 211–219.

668 Kano, Y., Chiba, S., Kase, T., 2002. Major adaptive radiation in neritopsine gastropods
669 estimated from 28S rRNA sequences and fossil records. Proc. R. Soc. Biol. Sci. Ser.
670 B 269, 2457-2465.

671 Kano, Y., Fukumori, H., Brenzinger, B., Warén, A., 2013. Driftwood as a vector for the
672 oceanic dispersal of estuarine gastropods (Neritidae) and an evolutionary pathway
673 to the sunken-wood community. J. Molluscan Stud. 79, 378-382.

674 Kano, Y., Kase, T., 2002. Anatomy and systematics of the submarine-cave gastropod
675 *Pisulina* (Neritopsina, Neritiliidae). J. Molluscan Stud. 68, 365–384.

676 Kano, Y., Kase, T., 2004. Genetic exchange between anchialine cave populations by
677 means of larval dispersal: the case of a new gastropod species *Neritilia cavernicola*.
678 Zool. Scr. 33, 423-437.

679 Kano, Y., Kase, T., 2008. Diversity and distributions of the submarine-cave Neritiliidae
680 in the Indo-Pacific (Gastropoda: Neritimorpha). Org Divers Evol 8, 22-43.

681 Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software
 682 Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30, 772-
 683 780.

684 Kornobis, E., Cabellos, L., Aguilar, F., Frías-López, C., Rozas, J., Marco, J., Zardoya,
 685 R., 2015. TRUFA: A user-friendly web server for de novo RNA-seq analysis using
 686 cluster computing. *Evol. Bioinform.* 11, 97-104.

687 Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: Combined
 688 selection of partitioning schemes and substitution models for phylogenetic
 689 analyses. *Mol. Biol. Evol.* 29, 1695-1701.

690 Lartillot, N., Brinkmann, H., Philippe, H., 2007. Suppression of long-branch attraction
 691 artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol.*
 692 *Biol.* 7 Suppl. 1, S4.

693 Lartillot, N., Philippe, H., 2004. A Bayesian mixture model for across-site
 694 heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* 21, 1095-
 695 1109.

696 Lartillot, N., Rodrigue, N., Stubbs, D., Richer, J., 2013. PhyloBayes MPI: phylogenetic
 697 reconstruction with infinite mixtures of profiles in a parallel environment. *Syst.*
 698 *Biol.* 62, 611-615.

699 Laslett, D., Canbäck, B., 2008. ARWEN: a program to detect tRNA genes in metazoan
 700 mitochondrial nucleotide sequences. *Bioinformatics* 24, 172-175.

701 Lindberg, D.R., 2008. Patellogastropoda, Neritimorpha and Cocculinoidea: The low-
 702 diversity gastropod clades. In: Ponder, W.F., Lindberg, D.R. (Eds.), *Phylogeny and*
 703 *Evolution of the Mollusca*. University of California Press, Berkeley, pp. 271-296

704 Little, C., 1990. *The terrestrial invasion. An ecophysiological approach to the origin of*
 705 *land animals*. Cambridge University Press, Cambridge.

706 Littlewood, D.T.J., Curini-Galletti, M., Herniou, E.A., 2000. The Interrelationships of
 707 Proseriata (Platyhelminthes: Seriata) tested with molecules and morphology. *Mol.*
 708 *Phylogen. Evol.* 16, 449-466.

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

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

715 Osca, D., Templado, J., Zardoya, R., 2015. Caenogastropod mitogenomics. *Mol.*
 716 *Phylogen. Evol.* 93, 118-128.

717 Palumbi, S., Martin A, McMillan, W., Stice, L. Gr b w k G. 1991. Th p f '

718 guide to PCR. <http://palumbi.stanford.edu/SimpleFoolsMaster.pdf>, 1-45.

719 Pearson, W.R., 1999. Flexible sequence similarity searching with the FASTA3 program
 720 package. In: Misener, S., Krawetz, S.A. (Eds.), *Methods in Molecular Biology.*
 721 *Bioinformatics methods and protocols.* Humana Press, Totowa, New Jersey, pp.
 722 185–219.

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

725 Rambaut, A., Drummond, A.J., 2007. Tracer v1.4, Available from
 726 <http://beast.bio.ed.ac.uk/Tracer>.

727 Richling, R., 2004. Classification of the Helicinidae: Review of morphological
 728 characteristics based on a revision of the Costa Rican species and application to the
 729 arrangement of the Central American mainland taxa (Mollusca: Gastropoda:
 730 Neritopsina). *Malacologia* 45, 195-440.

731 Richling, R., 2014. Poorly explored jewels of the tropics: Estimating diversity in non-
 732 pulmonate land snails of the family Helicinidae (Gastropoda: Neritopsina). Am.
 733 Malacol. Bull. 32, 246-258.

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

736 Saghai-Marooof, M.A., K.M., S., Jorgensen, R.A., Allard, R.W., 1984. Ribosomal DNA
 737 spacer-length polymorphism in barley: mendelian inheritance, chromosomal
 738 location, and population dynamics. Proc. Natl. Acad. Sci. USA 81, 8014-8019.

739 Schattner, P., Brooks, A.N., Lowe, T.M., 2005. The tRNAscan-SE, snoscan and snoGPS
 740 web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33,
 741 W686-W689.

742 Schilthuizen, M., Rutten, E.M.J., Haase, M., 2012. Small-scale genetic structuring in a
 743 tropical cave snail and admixture with its above-ground sister species. Biol. J. Linn.
 744 Soc. 105, 727–740.

745 Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic
 746 datasets. Bioinformatics 27, 863-864.

747 Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461-464.

748 Sevigny, J.L., Kirouac, L.E., Thomas, W.K., Ramsdell, J.S., Lawlor, K.E., Sharifi, O.,
 749 Grewal, S., Baysdorfer, C., Curr, K., Naimie, A.A., Okamoto, K., Murray, J.A.,
 750 Newcomb, J.M., 2015. The mitochondrial genomes of the nudibranch mollusks,
 751 *Melibe leonina* and *Tritonia diomedea*, and their Impact on gastropod phylogeny.
 752 PLoS ONE 10, e0127519.

753 Shao, R., Dowton, M., Murrell, A., Barker, S.C., 2003. Rates of gene rearrangement and
 754 nucleotide substitution are correlated in the mitochondrial genomes of insects. Mol.
 755 Biol. Evol. 20, 1612 - 1619.

756 Solem, A., 1983. Lost or kept internal whorls: ordinal differences in land snails. J.
757 Molluscan Stud. Suppl. A 12, 172–178.

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

760 StJohn, J., 2011. SeqPrep. <https://github.com/jstjohn/SeqPrep>.

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

763 Symonds, M. F. 2009. The Neritopsidae and the Neritidae of the Bracklesham Group
764 (Early and Middle Eocene) of the Hampshire Basin. Cainozoic Res. 6, 37–51.

765 Thompson, F.G., 1980. Proserpinoid land snails and their relationships within the
766 Archaeogastropoda. Malacologia 20, 1–33.

767 Uribe, J.E., Kano, Y., Templado, J., Zardoya, R., 2016. Mitogenomics of
768 Vetigastropoda: insights into the evolution of pallial symmetry. Zool. Scr. 45, 145-
769 159.

770 Warén, A., Bouchet, P., 2001. Gastropoda and Monoplacophora from hydrothermal
771 vents and seeps; New taxa and records. The Veliger 44, 116-231.

772 Williams, S.T., Foster, P.G., Littlewood, D.T.J., 2014. The complete mitochondrial
773 genome of a turbinid vetigastropod from MiSeq Illumina sequencing of genomic
774 DNA and steps towards a resolved gastropod phylogeny. Gene 533, 38-47.

775 Williams, S.T., Reid, D.G., Littlewood, D.T.J., 2003. A molecular phylogeny of the
776 Littorininae (Gastropoda: Littorinidae): unequal evolutionary rates, morphological
777 parallelism, and biogeography of the Southern Ocean. Mol. Phylogenet. Evol. 28, 60-
778 86.

779 Winnepeninckx, B., Backeljau, T., De Wachter, R., 1995. Phylogeny of protostome
780 worms derived from 18S rRNA sequences. Mol. Biol. Evol. 12, 641-649.

781 Wyman, S.K., Jansen, R.K., Boore, J.L., 2004. Automatic annotation of organellar
782 genomes with DOGMA. *Bioinformatics* 20, 3252-3255.

783 Yokobori, S.-i., Fukuda, N., Nakamura, M., Aoyama, T., Oshima, T., 2004. Long-term
784 conservation of six duplicated structural genes in cephalopod mitochondrial
785 genomes. *Mol. Biol. Evol.* 21, 2034-2046.

786

Figure legends

Figure 1. Mitochondrial gene orders of the four superfamilies of Neritimorpha. The gene organization of the mt genomes of Cephalopoda, Vetigastropoda, and Caenogastropoda is shown for comparison. The genes encoded in the major and minor strands are shown in the top and bottom lines, respectively. Inversions are colored in yellow. Translocated genes are in green. Genes translocated and inverted are in blue. Genes duplicated are in orange. The red boxes in Vetigastropoda show the tRNA genes that rearrange in some lineages of this group with respect to Neritimorpha. Striped boxes indicate regions not sequenced in *Titiscania* and *Viana* (the region not sequenced in *Theodoxus* is of similar position and length to that of *Titiscania* but is not shown).

Figure 2. Phylogenetic relationships of Neritimorpha based on complete mt genomes (protein coding genes excluding third codon positions plus rRNA genes; 7,604 positions in total). The reconstructed ML phylogram using *Neomphalina* as outgroup is shown. Branch colors indicate main gastropod lineages. The names of the four neritimorph superfamilies are indicated. Numbers at nodes are statistical support values for ML (bootstrap proportions)/ BI (posterior probabilities). Red asterisks indicate nodes with maximal ML and BI support. Scale bar indicates substitutions/site. Drawings of Neritopsoidea, Helicinoidea, and Hydrocenoidea taken from Kano et al. (2002).

Figure 3. Phylogenetic relationships of Neritimorpha based on combined mt and nuclear data (10,566 positions long). The reconstructed ML phylogram using Vetigastropoda as outgroup is shown. Additionally, a ML phylogeny based solely on nuclear data (2,962 positions from *18S rRNA*, *28S rRNA*,

actin and *histone H3* genes) is shown in the inset. Branch colors indicate main gastropod lineages. The names of the four neritimorph superfamilies are indicated. For some tree tips, sequences were concatenated from species for which there is strong evidence for the monophyly of the higher taxon rank, which is indicated. Numbers at nodes are statistical support values for ML (bootstrap proportions)/ BI (posterior probabilities). Red asterisks indicate nodes with maximal ML and BI support. Scale bar indicates substitutions/site.

Figure 4. Chronogram of Neritimorpha based on mt data set (7,604 positions) and fossil-based calibration priors (A and B). A Bayesian uncorrelated relaxed lognormal clock was used in BEAST. Horizontal bars represent 95% credible intervals for time estimates; dates are in millions of years.

Figure 1

NERITIMORPHA

Neritoidea

Nerita, Neritina, and Theodoxus

Hydrocenoidea

Georissa

Helicinoidea

Pleuropoma

Viana

Neritopsoidea

Titiscania

CAENOCASTROPODA

VETIGASTROPODA

CEPHALOPODA

Octopus

Figure 2

Figure 3

Figure 4

Table 1. Complete mitochondrial (mt) genomes analyzed in this study. Length in bp, Genbank accession number, museum voucher, sampling location, habitat, and name of collector are provided

New mt genomes							
Species	Superfamily	bp	GenBank	Voucher	Location	Habitat	Collector
<i>Georissa banguyensis</i>	Hydrocenoidea	15.267	KU342664	MNCN/ ADN:86335	Sabah, Sandakan, Malaysia	Terrestrial	Menno Schilthuizen
<i>Neritina usnea</i>	Neritoidea	15.574	KU342665	MNCN/ ADN:86336	Cienaga de Mallorquin, Colombia	Estuarine	Moises Arquez
<i>Pleuropoma jana</i>	Helicinoidea	15.851	KU342666	AM/ EBU:65394	Yessabah Caves, New South Wales, Australia	Terrestrial	Michael Shea, Don Colgan, Sven Schreiter
<i>Theodoxus fluviatilis</i> *	Neritoidea	14.122	KU342667	MNCN/ ADN:86338	La Fuente del Mal Nombre, Malaga, España	Freshwater	José Templado
<i>Viana regina</i> *	Helicinoidea	7.557	KU342668	MNCN/ ADN:86339	Valle de Viñales, Pinar del Río, Cuba	Terrestrial	Jane Herrera
<i>Titiscania</i> sp.*	Neritopsoidea	15.046	KU342669	MNCN/ ADN:86337	Península de Santa Elena, Guanacaste, Costa Rica	Marine	Yolanda Camacho

GenBank mt genomes				
Species	Superfamily	bp	GenBank	Reference
<i>Haliotis rubra</i>	Haliotoidea	16.907	NC_005940	Maynard et al., 2005
<i>Lunella aff. cinerea</i>	Trochoidea	17.670	KF700096	Williams et al., 2014
<i>Angaria neglecta</i>	Angaroidea	19.470	KR297248	Uribe et al., 2016
<i>Lepetodrilus schrolli</i> *	Lepetodriiloidea	15.579	KR297250	Uribe et al., 2016
<i>Phasianella solida</i>	Phasianelloidea	16.698	KR297251	Uribe et al., 2016
<i>Ilyanassa obsoleta</i>	Buccinoidea	15.263	NC_007781	Simison et al., 2006
<i>Rapana venosa</i>	Muricoidea	15.272	NC_011193	Chandler et al., 2008
<i>Oncomelania hupensis</i>	Truncatelloidea	15.182	NC_013073	Li and Zhou 2009
<i>Africonus borgesii</i>	Conoidea	15.536	NC_013243	Cunha et al., 2009
<i>Nerita melanotragus</i> *	Neritoidea	15.261	GU810158	Castro and Colgan 2010
<i>Nerita fulgurans</i> *	Neritoidea	15.261	KF728888	Arquez et al., 2014
<i>Nerita tessellata</i> *	Neritoidea	15.741	KF728889	Arquez et al., 2014
<i>Nerita versicolor</i> *	Neritoidea	15.866	KF728890	Arquez et al., 2014
<i>Chrysomallon squamiferum</i>	Neomphaloidea	15.388	AP013032	Nakagawa et al., 2014

*incomplete mt genomes

Table 2. Nuclear sequence data analyzed in this study. Length in bp, Genbank accession number, museum voucher, and sampling location are provided

New sequences		<i>18S</i>		<i>28S</i>		<i>actin</i>		<i>H3</i>		Voucher	Location
Species	Superfamily	bp	Genbank	bp	Genbank	bp	Genbank	bp	Genbank		
<i>Pleuropoma jana</i>	Helicinoidea	1768	KU342679	3714	KU342680	300	KU342681	314	KU342682	AM/ EBU: 65394	Yessabah Caves, New South Wales, Australia
<i>Titiscania</i> sp.	Neritopsoidea	988	KU342683	1437	KU342684	533	KU342685	—	—	MNCN/DNA: 86337	Península de Santa Elena, Guanacaste, Costa Rica
<i>Georissa banguyensis</i>	Hydrocenoidea	971	KU342673	990	KU342674	488	KU342675	314	KU342676	MNCN/DNA: 86335	Sabah, Sandakan, Malaysia
<i>Gibbula rarilineata</i> [†]	Trochoidea	—	—	—	—	711	KU342677	314	KU342678	NHUK/DNA: 20080375	Lumbarda, Korcula I., Croatia
<i>Clanculus cruciatus</i>	Trochoidea	—	—	1415	KU342672	720	KU342671	314	KU342670	MNCN/DNA: 86340	Cabo de Palos, Murcia, Spain

GenBank sequences		<i>18S</i>		<i>28S</i>		<i>actin</i>		<i>H3</i>	
Species	Superfamily	bp	Genbank	bp	Genbank	bp	Genbank	bp	Genbank
<i>Titiscania limacina</i>	Neritopsoidea	—	—	—	—	—	—	327	KF527280
<i>Nerita</i> *	Neritoidea	957	AM048633	1479	AM048693	325	KM025036	—	—
<i>Gibbula rarilineata</i> [†]	Trochoidea	1182	AF534996	—	—	—	—	—	—
<i>Clanculus cruciatus</i>	Trochoidea	1733	AF120514	—	—	—	—	—	—
<i>Lepetodrilus elevatus</i> [#]	Lepetodrilioidea	1856	AY145381	3408	AY145413	—	—	376	AY923959
<i>Rapana venosa</i>	Muricoidea	1814	HQ834011	1477	HE584197	1565	KF410817	354	HQ834144
<i>Ilyanassa obsoleta</i>	Buccinoidea	1826	AY145379	3415	AY145411	297	AY766115	—	—

* mix of sequences of different species from the same genus in the nuclear and combined data sets

[†] mixed with the *Lunella* aff. *cinerea* mt genome in the combined analyses (as Trochoidea; see Fig. 3).

[#] mixed with the *Lepetodrilus schrolli* mt genome in the combined analyses (as *Lepetodrilus*; see Fig. 3).

Supplementary Material 1. Long PCR and primer walking primers

Georissa bangueyensis

Long PCR & Illumina

Primer	Sequence 5'-3'	Fragment (bp)
GBcox1F	GCTCCTGATATGGCTTTTCCTCGGTAA	cox1-rrnL (9527)
NS16SF	GAGAGATTGTGACCTCGATGTTGGACTAGG	
Geocox1R	AGAACCCCCCATGAGCCAAATTACT	rrnL-cox1 (5917)
NERmt16SR	AGTCCAACATCGAGGTCACAAWCTCTT	

Neritina usnea

Long PCR & Illumina

Primer	Sequence 5'-3'	Fragment (bp)
NUcos1FH	GAGCTTGGACAGCCTGGTGCTTTATTAGGGG	cox1-rrnL (9773)
SPN16sF	CGGTACCTAGTCCAACATCGAGGTCACA	
Nas16sR	AGGCTGGTATGAACGGTTTGACGAGA	rrnL-cox1 (5835)
SV-NU-cox1R	CCAAGCTCAGCTCGAATTAACAACTAAGAG	

Standard PCR

Primer	Sequence 5'-3'
NULeuR	TAAACCTAATGCACCTCTTCTGCCACC
NU16SF	CAAAAAGACGAGAAGACCCTGTTGAGC

Titiscania sp.

Long PCR & Illumina

Primer	Sequence 5'-3'	Fragment (bp)
TLmtcox1F	GGTTCGGTGGACTTAGCTATTTTTCACT	cox1-rrnL (10507)
NERmt16SF	CTGACYGTGCAAAGGTAGCATAATCAT	

Long PCR & Primer walking

Primer	Sequence 5'-3'	Fragment (bp)
COIRbruTit	CGTCAACAACATAGTAATGGCCCCAGC	cox3-cox1 (2930)
TLcox3F	AGTTACAGTTACTTGGGCGCATCATGC	
TLLys3F	TATAGGCAGAAGCATGAAGAGGAGGC	rrnL-rrnS (2322)
TLnad2R	ATCTCCAAAAAGAGCACCAGGTTGCC	
EctremFTit	AGAAGATAGAAACCAACCTGGCTCGC	
12Sb (Kocher et al., 1989)	GAGGGTGACGGGCGGTGTGT	
TL16s1F_w	ACCTCGTAATATGAAGTGCAGCCC	rrnL-rrnS (2322)
TL12sR_w	TTAGATGTTTCAGTTCAGTGCCTCGC	

Theodoxus fluviatilis

Long PCR & Illumina

Primer	Sequence 5'-3'	Fragment (bp)
GastCOX1F	CTGGWACAGGATGRACGTGTWATYC	cox1-rrnL (9844)
GAS116SF	TGCCTTTTAATTGAAGGCTGGWATGAATGGTTTRACG	

Long PCR & Primer walking

Primer	Sequence 5'-3'	Fragment (bp)
Nercox3F	GAGGGAACWTTTCAAGGMTTTCATACTGG	cox3-cox1 (2761)
Gas2cox1R	ATTAAGCCGAGGAAAAGCCATATCAGGAGC	
TXPrimer2R	CGTATTCATCTCAATACATAACAAAAGCCC	rrnL-rrnS (1911)
Txnad2R	GCAGAAACTGAAATATAGTTCCAAACCC	
ExtremFNer	AAAGATAGAACTGACCTGGCTTACGCC	rrnL-rrnS (1911)
12Sb (Kocher et al., 1989)	GAGGGTGACGGGCGGTGTGT	
Tx12SR_w	CGTTGCTAATAGCAGGGTGTCTAATCC	

Viana regina

Long PCR & Illumina

Primer	Sequence 5'-3'	Fragment (bp)
Vivcox1F	CTTTGATGTTGAAGGTGCCAGATATGGC	cox1-rrnL (5071)
Viv16SRUJ	ATAAAGCTCAACAGGGTCTCATCGCC	
Vivcox1R	GGTGGATAAACAGTCCAACCAGTACC	cox3-cox1 (2676)
Vivcox3F	CTCATCATGCTTTAGTTGTTGGTTGTCG	

Supplementary Material 2. Evolutionary substitution models selected by Partition Finder

mt data set													
nucleotides	Partition (BIC = 99518.19)	Best Model	Alpha	pinvar	A <-> C	A <-> G	A <-> T	C <-> G	C <-> T	pi(A)	pi(C)	pi(G)	pi(T)
	1 All 1st codon positions of mt protein-coding genes	GTR+I+G	0.94	0.33	2.16	6.25	3.15	0.75	22.00	0.25	0.17	0.26	0.31
	2 All 2nd codon positions of mt protein-coding genes	GTR+G	0.73	—	0.68	2.60	0.67	5.61	2.01	0.17	0.21	0.16	0.43
	3 <i>rrnS-rrnL</i> genes	GTR+I+G	0.60	0.16	1.18	6.34	2.96	0.33	11.57	0.33	0.15	0.21	0.29
amino acids	Partition (BIC=86576.14)	Best Model	Alpha	pinvar	A <-> C	A <-> G	A <-> T	C <-> G	C <-> T	pi(A)	pi(C)	pi(G)	pi(T)
	1 All mt protein-coding genes in aminoacids level combined	MtArt+I+G+F*	0.56	0.15									
	2 <i>rrnS-rrnL</i> genes	GTR+I+G	0.47	0.17	0.45	2.87	0.93	0.07	5.51	0.33	0.15	0.21	0.29
nuclear data set	Partition (BIC=22139.97)	Best Model	Alpha	pinvar	A <-> C	A <-> G	A <-> T	C <-> G	C <-> T	pi(A)	pi(C)	pi(G)	pi(T)
	1 <i>18S</i> gene	TrNef+G	1000	—	0.91	1.88	0.48	0.60	3.94	0.25	0.21	0.26	0.25
	2 <i>28S</i> gene	GTR+I+G	0.31	0.00	0.57	1.80	1.41	1.62	6.62	0.22	0.25	0.32	0.19
	4 1st codon positions of <i>H3</i> and <i>actin</i> genes	GTR+G	0.58	—	4.41	0.54	6.11	0.18	17.63	0.27	0.23	0.31	0.16
	5 2nd codon positions of <i>H3</i> and <i>actin</i> genes	GTR+G	1.60	—	2.03	1.69	0.40	4.34	1.77	0.30	0.25	0.17	0.26
	6 3rd codon positions of <i>H3</i> and <i>actin</i> genes	GTR+G	0.96	—	8.75	49.06	39.18	0.12	28.36	0.10	0.41	0.28	0.19
combined data set	Partition (mt + nuclear)	Best Model	Alpha	pinvar	A <-> C	A <-> G	A <-> T	C <-> G	C <-> T	pi(A)	pi(C)	pi(G)	pi(T)
	1 All 1st codon positions of mt protein-coding genes	GTR+I+G	1.34	0.35	3.04	7.40	4.96	0.70	18.49	0.25	0.16	0.25	0.31
	2 All 2nd codon positions of mt protein-coding genes	GTR+G	0.68	—	1.08	2.43	0.75	5.62	2.27	0.17	0.21	0.16	0.44
	3 <i>rrnS-rrnL</i> genes	GTR+I+G	0.73	0.18	1.05	7.39	4.33	0.25	11.63	0.32	0.15	0.21	0.30
	4 <i>18S</i> gene	TrNef+G	1000	—	0.66	1.42	0.36	0.46	2.89	0.25	0.21	0.26	0.25
	5 <i>28S</i> gene	GTR+I+G	0.36	0.04	0.34	1.19	1.02	1.04	4.36	0.22	0.25	0.32	0.19
	6 1st codon positions of <i>H3</i> and <i>actin</i> genes	GTR+G	0.63	—	3.53	0.33	2.32	0.00	9.39	0.27	0.24	0.32	0.15
	7 2nd codon positions of <i>H3</i> and <i>actin</i> genes	GTR+G	1.69	—	3.08	0.19	0.45	8.16	1.38	0.30	0.25	0.17	0.26
	8 3rd codon positions of <i>H3</i> and <i>actin</i> genes	GTR+G	1.33	—	4.06	19.21	24.70	0.23	15.19	0.09	0.42	0.29	0.18
BEAST data set	Partition (BIC= 70900.79)	Best Model	Alpha	pinvar									
	1 1st codon positions of <i>atp6-8</i> genes	HKY+G	0.51	—									
	2 2nd codon positions of <i>atp6-8</i> genes	GTR+G	0.41	—									
	3 1st codon positions of <i>cob</i> gene	GTR+I+G	0.46	0.27									
	4 2nd codon positions of <i>cob</i> gene	GTR+G	0.22	—									
	5 1st codon positions of <i>cox1-3</i> genes	GTR+I+G	0.4	0.23									
	6 2nd codon positions of <i>cox1-3</i> genes	GTR+I+G	0.6	0.41									
	7 1st codon positions of <i>nad1-6</i> genes	GTR+I+G	1.11	0.25									
	8 2nd codon positions of <i>nad1-6</i> genes	GTR+G	0.4	—									
	9 <i>rrnS-rrnL</i> genes	GTR+I+G	0.46	3.45									

* implemented manually in MrBayes using the command prset aarevmatpr=fixed (replacement matrix)

Supplementary Material 3. Mitochondrial genome features

<i>Georissa bangueyensis</i>								
Gene	Type	Gene			Codon		Intergenic	Strand
		Start	Stop	Length	Start	Stop		
<i>cox1</i>	CDS	1	1539	1539	TAG	TAA	4	forward
<i>cox2</i>	CDS	1544	2228	687	TAG	T--	0	forward
<i>trnD</i>	tRNA	2229	2291	63			0	forward
<i>atp8</i>	CDS	2292	2447	156	TAG	TAG	9	forward
<i>atp6</i>	CDS	2457	3155	699	TAG	TAA	29	forward
<i>trnF</i>	tRNA	3185	3251	67			0	reverse
<i>nad5</i>	CDS	3252	4954	1704	TAG	T--	0	reverse
<i>trnH</i>	tRNA	4955	5018	64			0	reverse
<i>nad4</i>	CDS	5019	6381	1363	TAG	T--	-7	reverse
<i>nad4L</i>	CDS	6375	6665	291	TAG	TAA	5	reverse
<i>trnT</i>	tRNA	6671	6733	63			-1	forward
<i>trnS(tga)</i>	tRNA	6733	6798	66			25	reverse
<i>cob</i>	CDS	6824	7969	1146	TAG	TAA	1	reverse
<i>nad6</i>	CDS	7971	8468	498	TAG	TAA	3	reverse
<i>trnP</i>	tRNA	8472	8534	63			0	reverse
<i>nad1</i>	CDS	8535	9454	924	TAG	T--	0	reverse
<i>trnL(taa)</i>	tRNA	9455	9518	64			0	reverse
<i>trnL(tag)</i>	tRNA	9519	9581	63			0	reverse
<i>rrnL</i>	rRNA	9582	10810	1229			0	reverse
<i>trnV</i>	tRNA	10811	10873	63			0	reverse
<i>rrnS</i>	rRNA	10874	11620	747			0	reverse
<i>trnM</i>	tRNA	11621	11684	64			0	reverse
<i>trnY</i>	tRNA	11685	11747	63			1	reverse
<i>trnC</i>	tRNA	11749	11810	62			2	reverse
<i>trnW</i>	tRNA	11813	11880	68			-3	reverse
<i>trnQ</i>	tRNA	11878	11943	66			13	reverse
<i>trnG</i>	tRNA	11957	12019	63			0	reverse
<i>trnE</i>	tRNA	12020	12081	62			608	reverse
<i>cox3</i>	CDS	12690	13469	780	TAG	TAG	17	forward
<i>trnK</i>	tRNA	13487	13551	65			1	forward
<i>trnA</i>	tRNA	13553	13617	65			0	forward
<i>trnR</i>	tRNA	13618	13683	66			-1	forward
<i>trnN</i>	tRNA	13683	13748	66			0	forward
<i>trnI</i>	tRNA	13749	13816	68			0	forward
<i>nad3</i>	CDS	13817	14169	353	TAG	TA-	0	forward
<i>trnS(gct)</i>	tRNA	14170	14237	68			0	forward
<i>nad2[†]</i>	CDS	14238	15267	1030	TAG	T--	0	forward

Indetermined region

Intergenic regions with % A-T > 75 (only regions ≥ 25pb were calculated)

[†] Incomplete stop codon that requires polyadenylation; alternatively this gene may overlap with *cox1* until a complete stop codon is found

Supplementary Material 3 (cont.)

<i>Neritina usnea</i>								
Gene	Type	Gene			Codon		Intergenic	Strand
		Start	Stop	Length	Start	Stop		
<i>cox1</i>	CDS	1	1548	1548	TAG	TAA	11	forward
<i>cox2</i>	CDS	1560	2249	690	TAG	TAG	1	forward
<i>trnD</i>	tRNA	2251	2316	66			0	forward
<i>atp8</i>	CDS	2317	2481	165	TAG	TAA	6	forward
<i>atp6</i>	CDS	2488	3189	702	TAG	TAA	22	forward
<i>trnF</i>	tRNA	3212	3279	68			6	reverse
<i>nad5</i>	CDS	3286	4993	1708	TAG	TAA	0	reverse
<i>trnH</i>	tRNA	4994	5059	66			0	reverse
<i>nad4</i>	CDS	5060	6425	1366	TAG	T--	-7	reverse
<i>nad4L</i>	CDS	6419	6712	294	TAG	TAA	5	reverse
<i>trnT</i>	tRNA	6718	6785	68			6	forward
<i>trnS(tga)</i>	tRNA	6792	6856	65			5	reverse
<i>cob</i>	CDS	6862	7998	1137	TAG	TAA	12	reverse
<i>nad6</i>	CDS	8011	8517	507	TAG	TAA	1	reverse
<i>trnP</i>	tRNA	8519	8584	66			1	reverse
<i>nad1</i>	CDS	8586	9518	933	TAG	TAA	0	reverse
<i>trnL(taa)</i>	tRNA	9519	9586	68			12	reverse
<i>trnL(tag)</i>	tRNA	9599	9668	70			0	reverse
<i>rrnL</i>	rRNA	9669	10972	1304			0	reverse
<i>trnV</i>	tRNA	10973	11039	67			0	reverse
<i>rrnS</i>	rRNA	11040	11901	862			0	reverse
<i>trnM</i>	tRNA	11902	11968	67			5	reverse
<i>trnY</i>	tRNA	11974	12041	68			5	reverse
<i>trnC</i>	tRNA	12047	12110	64			0	reverse
<i>trnW</i>	tRNA	12111	12176	66			0	reverse
<i>trnQ</i>	tRNA	12177	12245	69			0	reverse
<i>trnG</i>	tRNA	12246	12311	66			2	reverse
<i>trnE</i>	tRNA	12314	12379	66			471	reverse
<i>cox3</i>	CDS	12851	13630	780	TAG	TAA	28	forward
<i>trnK</i>	tRNA	13659	13727	69			5	forward
<i>trnA</i>	tRNA	13733	13800	68			16	forward
<i>trnR</i>	tRNA	13817	13885	69			6	forward
<i>trnN</i>	tRNA	13892	13963	72			13	forward
<i>trnI</i>	tRNA	13977	14045	69			0	forward
<i>nad3</i>	CDS	14046	14399	354	TAG	TAA	5	forward
<i>trnS(gct)</i>	tRNA	14405	14472	68			0	forward
<i>nad2⁺</i>	CDS	14473	15574	1102	TAG	T--	-29	forward

Supplementary Material 3 (cont.)

<i>Pleuropoma jana</i>								
Gene	Type	Gene			Codon		Intergenic	Strand
		Start	Stop	Length	Start	Stop		
<i>cox1</i>	CDS	1	1539	1539	TAG	TAA	13	forward
<i>cox2</i>	CDS	1553	2239	687	TAG	TAG	8	forward
<i>trnG</i>	tRNA	2248	2310	63			5	forward
<i>trnD</i>	tRNA	2316	2378	63			0	forward
<i>atp8</i>	CDS	2379	2537	159	TAG	TAA	11	forward
<i>atp6</i>	CDS	2549	3241	693	TAG	TAA	28	forward
<i>trnM</i>	tRNA	3270	3332	63			6	forward
<i>trnY</i>	tRNA	3339	3404	66			0	forward
<i>trnM</i>	tRNA	3405	3470	66			0	forward
<i>rrnS</i>	rRNA	3471	4299	829			0	forward
<i>trnV</i>	tRNA	4300	4365	66			0	forward
<i>rrnL</i>	rRNA	4366	5626	1261			0	forward
<i>trnL(taa)</i>	tRNA	5627	5690	64			0	forward
<i>nad1</i>	CDS	5691	6614	924	TAG	TAA	76	forward
<i>trnP</i>	tRNA	6691	6756	66			1	forward
<i>nad6</i>	CDS	6758	7247	490	TAG	TAG	-8	forward
<i>cob</i>	CDS	7255	8388	1141	TAG	T--	0	forward
<i>trnS(tga)</i>	tRNA	8389	8454	66			14	forward
<i>trnT</i>	tRNA	8469	8538	70			43	reverse
<i>nad4L</i>	CDS	8582	8881	300	TAG	TAA	-7	forward
<i>nad4</i>	CDS	8875	10242	1368	TTG	TAA	0	forward
<i>trnH</i>	tRNA	10243	10303	61			0	forward
<i>nad5</i>	CDS	10304	11996	1693	TAG	T--	0	forward
<i>trnF</i>	tRNA	11997	12063	67			783	forward
<i>trnL(tag)</i>	tRNA	12847	12910	64			3	forward
<i>trnW</i>	tRNA	12914	12980	67			26	forward
<i>trnE</i>	tRNA	13007	13068	62			2	forward
<i>trnC</i>	tRNA	13071	13134	64			7	forward
<i>trnQ</i>	tRNA	13142	13209	68			29	forward
<i>cox3</i>	CDS	13239	14018	780	ATG	TAA	11	forward
<i>trnK</i>	tRNA	14030	14087	58			0	forward
<i>trnA</i>	tRNA	14088	14154	67			0	forward
<i>trnR</i>	tRNA	14155	14216	62			2	forward
<i>trnN</i>	tRNA	14219	14285	67			8	forward
<i>trnI</i>	tRNA	14294	14359	66			3	forward
<i>nad3</i>	CDS	14363	14708	346	ATA	T--	0	forward
<i>trnS(gct)</i>	tRNA	14709	14776	68			1	forward
<i>nad2</i>	CDS	14778	15833	1056	TAG	TAA	18	forward

***Titiscania* sp.**

Gene	Type	Gene			Codon		Intergenic	Strand
		Start	Stop	Length	Start	Stop		
<i>cox1</i>	CDS	2307	3842	1536	ATG	TAA	9	forward
<i>cox2</i>	CDS	3852	4547	696	ATG	TAA	21	forward
<i>trnD</i>	tRNA	4569	4636	68			1	forward
<i>atp8</i>	CDS	4638	4817	180	ATG	TAG	53	forward
<i>atp6</i>	CDS	4871	5569	699	ATG	TAA	39	forward
<i>trnF</i>	tRNA	5609	5676	68			393	reverse
<i>nad5</i>	CDS	6070	7812	1743	ATG	TAA	2	reverse
<i>trnH</i>	tRNA	7815	7883	69			38	reverse
<i>nad4</i>	CDS	7922	9304	1383	ATG	TAG	-7	reverse
<i>nad4L</i>	CDS	9298	9594	297	ATG	TAA	10	reverse
<i>trnT</i>	tRNA	9605	9673	69			6	forward
<i>trnS(tga)</i>	tRNA	9680	9746	67			5	reverse
<i>cob</i>	CDS	9752	10888	1137	ATG	TAA	29	reverse
<i>nad6</i>	CDS	10918	11424	507	ATG	TAG	2	reverse
<i>trnP</i>	tRNA	11427	11496	70			45	reverse
<i>nad1</i>	CDS	11542	12477	936	ATG	TAG	1	reverse
<i>trnL(taa)</i>	tRNA	12479	12547	69			12	reverse
<i>trnL(tag)</i>	tRNA	12560	12627	68			0	reverse
<i>rrnL</i>	rRNA	12628	14042	1415			0	reverse
<i>trnV</i>	tRNA	14043	14110	68			0	reverse
<i>rrnS</i>	rRNA	14111	15046	936				reverse
<hr/>								
<i>cox3</i>	CDS	1	288	288	partial	TAG	66	forward
<i>trnK</i>	tRNA	355	427	73			9	forward
<i>trnA</i>	tRNA	437	505	69			18	forward
<i>trnR</i>	tRNA	524	592	69			6	forward
<i>trnN</i>	tRNA	599	670	72			1	forward
<i>trnI</i>	tRNA	672	739	68			2	forward
<i>nad3</i>	CDS	742	1101	360	ATG	TAA	47	forward
<i>trnS(gct)</i>	tRNA	1149	1216	68			0	forward
<i>nad2[†]</i>	CDS	1217	2306	109	ATG	T--	0	forward

Supplementary Material 3 (cont.)

Theodoxus fluviatilis

		Gene			Codon			
Gene	Type	Start	Stop	Length	Start	Stop	Intergenic	Strand
cox1	CDS	2479	4026	1548	ATG	TAA	10	forward
cox2	CDS	4037	4726	690	ATG	TAA	6	forward
trnD	tRNA	4733	4798	66			0	forward
atp8	CDS	4799	4963	165	ATG	TAA	8	forward
atp6	CDS	4972	5673	702	ATG	TAA	22	forward
trnF	tRNA	5696	5762	67			0	reverse
nad5	CDS	5763	7476	1714	ATG	T--	0	reverse
trnH	tRNA	7477	7541	65			0	reverse
nad4	CDS	7542	8907	1366	ATG	T--	-7	reverse
nad4L	CDS	8901	9194	294	ATG	TAA	4	reverse
trnT	tRNA	9199	9266	68			7	forward
trnS(tga)	tRNA	9274	9338	65			2	reverse
cob	CDS	9341	1048	114	ATG	TAA	9436	reverse
nad6	CDS	10485	10991	507	ATG	TAA	1	reverse
trnP	tRNA	10993	11056	64			1	reverse
nad1	CDS	11058	1199	933	ATG	TAA	10791	reverse
trnL(taa)	tRNA	11991	12058	68			0	reverse
trnL(tag)	tRNA	12059	12128	70			0	reverse
rrnL	rRNA	12129	13424	1296			0	reverse
trnV	tRNA	13425	13492	68			0	reverse
rrnS	rRNA	13493	14122	630			0	reverse
cox3	CDS	1	537	537	partial	TAG	26	forward
trnK	tRNA	564	631	68			5	forward
trnA	tRNA	637	704	68			19	forward
trnR	tRNA	724	792	69			12	forward
trnN	tRNA	805	879	75			4	forward
trnI	tRNA	884	951	68			0	forward
nad3	CDS	952	1305	354	ATG	TAG	3	forward
trnS(gct)	tRNA	1309	1376	68			0	forward
nad2 ⁺	CDS	1377	2478	1102	ATG	T--	-29	forward

[illegible]

Additional taxa included in the PhyloBayes analysis

Species	Order	Length (bp)	GenBank Acc. No.	Publication
<i>Tegula lividomaculata</i>	Trochoidea	17,375	KT207826	Uribe et al., 2015
<i>Bolma rugosa</i>	Trochoidea	17,432	KT207824	Uribe et al., 2015
<i>Lottia digitalis</i>	Lottioidea	26,835	NC_007782	Simison et al., 2006
<i>Peronia peronii</i>	Onchidioidea	13,968	NC_016181	White et al., 2011
<i>Galba pervia</i>	Lymnaeoidea	13,768	NC_018536	Liu et al., 2012
<i>Roboastra europaea</i>	Anadoridoidea	14,472	NC_004321	Grande et al., 2002
<i>Octopus vulgaris</i>	Neocoleoidea	15,744	NC_006353	Yokobori et al., 2004
<i>Scutopus ventrolineatus</i>	Scutopodidae*	14,662	NC_025284	Osca et al., 2014

*unassigned to a superfamily

References

- 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.
- Liu, G.-H., Wang, S.-Y., Huang, W.-Y., Zhao, G.-H., Wei, S.-J., Song, H.-Q., Xu, M.-J., Lin, R.-Q., Zhou, D.-H., Zhu, X.-Q., 2012. The complete mitochondrial genome of *Galba pervia* (Gastropoda: Mollusca), an intermediate host snail of *Fasciola* spp. *PLoS ONE* 7, e42172.
- Osca, D., Irisarri, I., Todt, C., Grande, C., Zardoya, R., 2014. The complete mitochondrial genome of *Scutopus ventrolineatus* (Mollusca: Chaetodermomorpha) supports the Aculifera hypothesis. *BMC Evol. Biol.* 14, 1-10.
- Simison, W.B., Lindberg, D.R., Boore, J.L., 2006. Rolling circle amplification of metazoan mitochondrial genomes. *Mol. Phylogen. Evol.* 39, 562 - 567.
- Uribe, J.E., Kano, Y., Templado, J., Zardoya, R., 2015. Mitogenomics of Vetigastropoda: insights into the evolution of pallial symmetry. *Zool. Scr.* doi: 10.1111/zsc.12146.
- White, T., Conrad, M., Tseng, R., Balayan, S., Golding, R., de Frias Martins, A., Dayrat, B., 2011. Ten new complete mitochondrial genomes of pulmonates (Mollusca: Gastropoda) and their impact on phylogenetic relationships. *BMC Evol. Biol.* 11, 295.
- 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.