Zoologica Scripta

Sipunculan phylogeny based on six genes, with a new classification and the descriptions of two new families

Gisele Y. Kawauchi, Prashant P. Sharma & Gonzalo Giribet

Submitted: 15 August 2011
Accepted: 27 November 2011
doi: 10.1111/j.1463-6409.2011.00507.x

Kawauchi, G. Y., Sharma, P. P. & Giribet, G. (2012). Sipunculan phylogeny based on six genes, with a new classification and the descriptions of two new families. —*Zoologica Scripta*, *41*, 186–210.

The internal phylogeny of Sipuncula has proven elusive, with the monophyly of multiple traditional groups in question. Repeated attempts to infer sipunculan relationships have attained discordant results, possibly owing to fragmentary molecular sequence data sets. We reassessed the phylogeny of Sipuncula using a six-gene data set and with larger target amplicons of certain loci. We additionally dated the molecular phylogeny employing recently discovered fossil taxa to constrain node ages. Our multilocus data set recovers six major clades of Sipuncula across multiple analytical treatments. Some groups considered suspect in previous studies are vindicated (e.g. Aspidosiphonidae), but most traditional sipunculan families were recovered as para- or polyphyletic groups, especially Sipunculidae, whose members appear in three distinct clades. To redress the dissonance between the current classification and the phylogeny of Sipuncula, we provide a new classification of the group, wherein (i) we erect two new families, Siphonosomatidae fam. nov. and Antillesomatidae fam. nov.; (ii) Phascolionidae and Themistidae are synonymized with Golfingiidae, **new synonymies** (iii) *Phascolopsis* is transferred to Golfingiidae, **new familial assignment**; and (iv) *Lithacrosiphon* is synonymized with *Aspidosiphon*, **new synonymy**. We observe that the origins of all families recognized are ancient, dating at least to the Mesozoic.

Corresponding author: *Gisele Y. Kawauchi, Museum of Comparative Zoology & Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA. E-mail: gykawa@gmail.com*
Prashant P. Sharma, Museum of Comparative Zoology & Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA. E-mail: psharma@fas.harvard.edu
Gonzalo Giribet, Museum of Comparative Zoology & Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA. E-mail: ggiribet@oeb.harvard.edu

Introduction

Sipunculans are exclusive marine, non-segmented, coelomate worms traditionally held to be a protostome phylum (Hyman 1959; Stephen & Edmonds 1972; Cutler 1994). The evolutionary origin and phylogenetic position of Sipuncula have long been contentious. Molecular studies during the last 10 years have added growing to a consensus that sipunculans are either closely related to or even fall within annelids (Boore & Staton 2002; Staton 2003; Jennings & Halanych 2005; Bleidorn *et al.* 2006; Tzetlin & Purschke 2006; Struck *et al.* 2007, 2011; Dunn *et al.* 2008; Hejnol *et al.* 2009; Mwinyi *et al.* 2009; Sperling *et al.* 2009; Zrzavý *et al.* 2009; Dordel *et al.* 2010). Developmental studies on Sipuncula have revealed aspects of segmentation during neurogenesis, corroborating the close relationship between Sipuncula and Annelida (Kristof *et al.* 2008, 2011; Wanninger *et al.* 2009). In spite of major and ongoing efforts to place Sipuncula in the metazoan tree of life, the exact position of this group of worms with respect to annelids remains uncertain (e.g. Edgecombe *et al.* 2011).

The systematics of Sipuncula is complex, as it has been variously ranked at different taxonomic levels – family, order, class and phylum (Saiz Salinas 1993; Cutler 1994). In the middle of the 20th century, phylum status was definitively accepted for the group, as 'Sipunculida' (Hyman 1959). The current name, Sipuncula, was proposed by Stephen (1964) and restated by Stephen &

Edmonds (1972) in the first monograph compiling all species described prior to about 1970. In addition, the authors classified the 320 species described in four families. During the 1980s, Cutler and collaborators contributed with extensive generic revisions (Cutler & Jurczak 1975; Cutler & Murina 1977; Cutler 1979, 1984; Cutler & Cutler 1982, 1983, 1985a,b, 1986, 1987, 1988, 1989, 1990; Cutler *et al.* 1982, 1983; Gibbs *et al.* 1983) and with the first seminal internal phylogenetic reconstructions (Cutler & Gibbs 1985; Gibbs & Cutler 1987). This first attempt to apply phylogenetic methods (using morphological characters and at a time when no fossil taxa were known) was somewhat hampered by inadequate outgroup comparison, ambiguous character polarity and a paucity of useful characters. The resulting classification consisted of two classes, four orders, six families and 17 genera (Cutler & Gibbs 1985; Gibbs & Cutler 1987) (Fig. 1).

Subsequently, Cutler synthesized the most updated monograph of sipunculans, including general observations on sipunculan biology (Cutler 1994). The classification therein was based in the same set of characters used in the two previous analyses, but the polarization process was based on a revised hypothetical ancestral sipunculan, resulting in several characters with opposite polarity compared to the previous works. In any case, the extensive revisions by Cutler and collaborators over 19 years reduced the original 320 species (Stephen & Edmonds 1972) to 149. The resulting system became widely accepted for the group and has been in place until today.

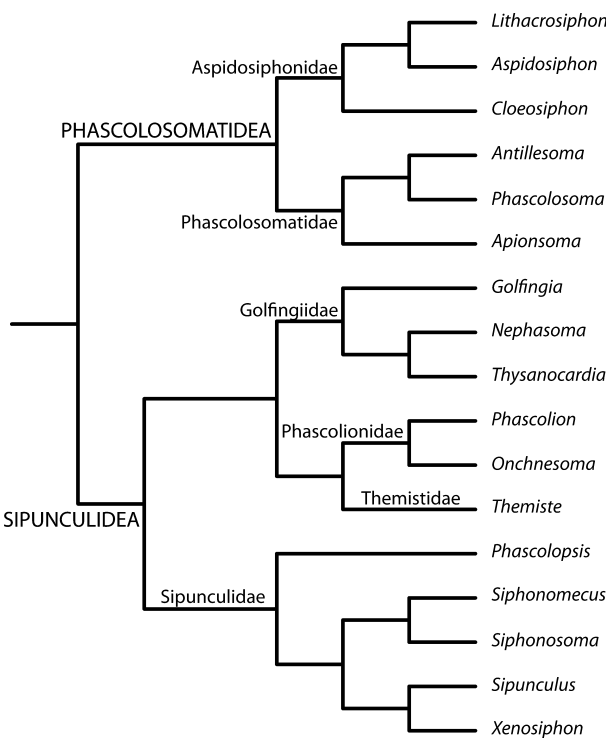

Fig. 1 Traditional classification of Sipuncula, redrawn from Cutler & Gibbs (1985).

The relationships of Sipuncula were recently examined in four successive phylogenetic studies, based on molecular sequence data, morphological data and/or the combination of the two (Maxmen *et al.* 2003; Staton 2003; Schulze *et al.* 2005, 2007). The monophyly of the group was strongly supported, but the four studies obtained discordant topologies that were sensitive to analytical treatments. In the most recent analysis, the topology was not consistent with a basal split between the two classes (Sipunculidea and Phascolosomatidea), and none of the families was monophyletic.

The phylogenies obtained in the four studies behooved changes in the higher-level systematics of Sipuncula. But Schulze *et al.* (2005) postponed changes to the classification on grounds of insufficient morphological and molecular data. The phylogeny proposed by Schulze *et al.* (2007) recovered a similar phylogeny from the previous study, and the authors recommended future changes, but did not take taxonomic action toward a new classification.

In addition, the dearth of fossil sipunculan specimens has obstructed a satisfactory understanding of the evolution of Sipuncula through time. Propitiously, three preserved fossils were recently discovered from the Lower Cambrian Maotianshan Shale in south-east China, placing the occurrence of crown-group Sipuncula at 520 Myr ago (Huang *et al.* 2004). The three fossils present evident similarities to modern sipunculans, particularly with species of Golfingiidae. However, the new fossils were classified as a sister group of the class Sipunculidea, based on the Cutler & Gibbs (1985) classification, which is inconsistent with topologies from phylogenetic results (Maxmen *et al.* 2003; Schulze *et al.* 2005, 2007).

To establish an updated classification of Sipuncula, we reinvestigated the molecular phylogeny of this group, implementing three strategies for phylogenetic refinement: (i) increasing the amount of 28S rRNA sequenced from ca. 300 to 2200 bp; (ii) sequencing the nuclear protein-encoding gene histone H4 and the mitochondrial ribosomal gene 16S rRNA; and (iii) increasing the requisite quantity of sequence data for a terminal to be included in the study, so as to reduce the influence of missing data. The phylogeny we obtain serves as a framework for a revised classification of Sipuncula comprised of monophyletic families exclusively. We additionally provide for the first time data on the timing of sipunculan diversification.

Materials and methods

Collection and taxon sampling

The majority of samples were previously collected for the preceding work (Maxmen *et al.* 2003; Schulze *et al.* 2005, 2007), but many of these were re-studied here. A collecting trip to New Caledonia (by G.Y.K. and Claudio G. Tiago, 2007) added several individuals to previous collections. *Nephasoma rimicola* and *Aspidosiphon* cf. *spiralis* (collected in New Caledonia); *Nephasoma* cf. *abyssorum* and *Themiste alutacea* (from MCZ DNA collection); and *Nephasoma columbaris* (collected in Florida, United States) are all new terminals added in this study. Outgroup taxa were chosen among the spiralian phyla Mollusca, Annelida, Entoprocta, Nemertea and Brachiopoda, as outlined in recent phylogenetic analyses of Metazoa (e.g. Dunn *et al.* 2008; Hejnol *et al.* 2009). Collecting data are provided in Appendix 1.

Molecular methods

Total DNA was extracted from a fragment (for large specimens) or the entire body (for small specimens). Purified genomic DNA was used as a template for polymerase chain reaction (PCR) amplification. Molecular markers consisted of two nuclear ribosomal genes (complete 18S rRNA and a 2.2-kb fragment of 28S rRNA); two nuclear protein-encoding genes (histones H3 and H4); one mitochondrial ribosomal gene (16S rRNA); and one mitochondrial protein-encoding gene (cytochrome *c* oxidase subunit I). The complete 18S rRNA was amplified according to Schulze *et al.* (2007). Partial 28S rRNA was amplified in three fragments, using the primers 28Srd1a-28Srd5b, 28Srd1a-28Srd4b or 28Ssip1-28Srd5b for the first fragment, 28Sa-28Sb for the second fragment and 28Srd4.8a-28Srd7b1 for the third fragment. The total length of the 28S rRNA amplicon is approximately 2267 bp. The other gene regions were amplified as a single fragment: cytochrome *c* oxidase subunit I (820 bp), 16S rRNA (ca. 598 bp), histone H3 (327 bp), and histone H4 (160 bp). PCRs were performed in 25-μL volume according to standard protocols with annealing temperatures between 34° C and 54° C for coding genes and 40° C and 59° C for ribosomal genes. Primer sequences are indicated in Table 1.

Visualization by agarose gel electrophoresis and direct sequencing were conducted as described by Schulze *et al.* (2007). Chromatograms obtained from the automatic sequencer were read and sequences assembled using the sequence editing software Sequencher™ (Gene Codes Corporation, Ann Arbor, MI, USA). Sequence data were edited in Se-Al v. 2.0a11 (Rambaut 1996). The 18S, 28S, 16S and H3 sequences have been deposited in GenBank under the accession nos. JN864950–JN865167 and JN869397–JN869399 and H4 sequences in EMBL under the accession nos. HE605122–HE605201 (Appendix 2).

Table 1 List of primer sequences used for amplification and sequencing, with original references of the primers sequences

Primer	Sequence	Author
16S rRNA		
16Sa	5′ – CGC CTG TTT ATC AAA AAC AT – 3′	Xiong & Kocher (1991)
16Sbr	5′ – CCG GTT TGA ACT CAG ATC ATG – 3′	Xiandong *et al.* (2008)
18S rRNA		
1F	5′ – TAC CTG GTT GAT CCT GCC AGT AG – 3′	Giribet *et al.* (1996)
3F	5′ – AGG CTC CCT CTC CGG AAT CGA AC – 3′	Giribet *et al.* (1996)
4R	5′ – GAA TTA CCG CGG CTG CTG G – 3′	Giribet *et al.* (1996)
9R	5′ – GAT CCT TCC GCA GGT TCA CCT AC – 3′	Giribet *et al.* (1996)
18Sa2.0	5′ – ATG GTT GCA AAG CTG AAA C – 3′	Giribet *et al.* (1996)
18Sbi	5′ – GAG TCT CGT TCG TTA TCG GA – 3′	Giribet *et al.* (1996)
28S rRNA		
28Ssip1	5′ – CCC YAG TAA CGG CGA GTA – 3′	This study
28Sa	5′ – GAC CCG TCT TGA AAC ACG GA – 3′	Whiting *et al.* (1997)
28Srd4b	5′ – CCT TGG TCC GTG TTT CAA GAC – 3′	Edgecombe & Giribet (2006)
28Srd5b	5′ – CCA CAG CGC CAG TTC TGC TTA C – 3′	Schwendinger & Giribet (2005)
28Srd4.8a	5′ – ACC TAT TCT CAA ACT TTA AAT GG – 3′	Schwendinger & Giribet (2005)
28Srd7b1	5′ – GAC TTC CCT TAC CTA CAT – 3′	Schwendinger & Giribet (2005)
COI		
LCO1490	5′ – GGT CAA CAA ATC ATA AAG ATA TTG G – 3′	Folmer *et al.* (1998)
HCOoutout	5′ – GTA AAT ATA TGR TGD GCT C – 3′	Prendini *et al.* (1998)
Histone H3		
H3aF	5′ – ATG GCT CGT ACC AAG CAG ACV GC – 3′	Colgan *et al.* (1998)
H3aR	5′ – ATA TCC TTR GGC ATR ATR GTG AC – 3′	Colgan *et al.* (1998)
Histone H4		
H4F2S	5′ – CKY TTI AGI GCR TAI ACC ACR TCC AT – 3′	Pineau *et al.* (2005)
H4F2er	5′ – TSC GIG AYA ACA TYC AGG GIA TCA C – 3′	Pineau *et al.* (2005)

Phylogenetic analysis

Both Bayesian inference (BI) and maximum likelihood (ML) analyses were conducted on static alignments, which were inferred as follows. Sequences of ribosomal genes were aligned using MUSCLE v. 3.6 (Edgar 2004) with default parameters and subsequently treated with GBLOCKS v. 0.91b (Castresana 2000) to cull positions of ambiguous homology. Sequences of the protein-encoding gene COI and histones H3 and H4 were aligned using MUSCLE v. 3.6 with default parameters as well, but alignments were confirmed using protein sequence translations prior to treatment with GBLOCKS v.0.91b. No gaps were permitted within blocks for any data partition. The size of data matrices for each gene prior and subsequent to treatment with GBlocks is provided in Appendix 3.

Bayesian inference (BI) analyses were conducted using MRBAYES v. 3.1.2 (Ronquist & Huelsenbeck 2003; Huelsenbeck & Ronquist 2005) with a unique GTR model of sequence evolution with corrections for a discrete gamma distribution and a proportion of invariant sites (GTR + Γ + I) specified for each partition, as selected in MODELTEST v. 3.7 (Posada & Crandall 1998; Posada 2005) under the Akaike Information Criterion (Posada & Buckley 2004). Default priors were used starting with random trees; two runs, each with three hot and one cold Markov chains, were run until the average deviation of split frequencies reached <0.01 (12 000 000 generations). After burn-in samples were discarded, sampled trees were combined in a single majority consensus topology, and the percentage of nodes was taken as clade posterior probabilities.

Maximum likelihood analysis was conducted using RAxML v. 7.2.7 (Stamatakis 2006) on 40 CPUs of a cluster at Harvard University, FAS Research Computing (odyssey.fas.harvard.edu). For the ML searches, a unique GTR model of sequence evolution with corrections for a discrete gamma distribution (GTR + Γ) was specified for each data partition, and 200 independent searches were conducted. Nodal support was estimated via the rapid bootstrap algorithm (1000 replicates; Stamatakis *et al.* 2008) using the GTR-CAT model, through the CIPRES v. 3 gateway, using the Abe Dell Intel 64 Linux teragrid cluster housed at the National Center for Supercomputing Applications (University of Illinois). Bootstrap resampling frequencies were thereafter mapped onto the optimal tree from the independent searches.

Parsimony analyses were based on a direct optimization approach (Wheeler 1996) using the program POY v. 4.1.2 (Varón *et al.* 2010). Each gene was analysed independently and in combination with all other molecular data. Tree searches were performed using the timed search function in POY, i.e., multiple cycles of (i) building multiple Wagner trees, (ii) subtree pruning and regrafting (SPR), (iii) tree bisection and reconnection (TBR), (iv) ratcheting (Nixon 1999) and (v) tree fusing (Goloboff 1999, 2002), on 24 CPUs of a cluster at Harvard University, FAS Research Computing (odyssey.fas.harvard.edu).

Timed searches of 6 h were run for the individual and combined analyses of all molecules under nine analytical parameter sets. A parameter space of two variables (indel/transversion ratio and transversion/transition ratio) was explored for six of these parameter sets; an additional two parameter sets incorporated variables for gap opening and affine gap costs, and the ninth parameter set deployed a mixed weighting scheme – ribosomal genes were weighted using the parameter set 3221 (indel opening cost = 3; indel extension cost = 1; transversions = transitions = 2; De Laet 2005), and protein-encoding genes were weighted using the parameter set 121 (indel cost = 2; transversion cost = 2; transition cost = 1), following previous exploration of a sipunculan data set in Sharma *et al.* (2011). Parameter set nomenclature follows Boyer & Giribet (2007). We undertook a sensitivity analysis (Wheeler 1995) of the nine parameter sets and used a modification of the incongruence length different (ILD) – the 'Wheeler ILD' or $_{W}$ILD (Mickevich & Farris 1981; Wheeler 1999; Sharma *et al.* 2011) – as a criterion for selecting a favoured parameter set. The results of the sensitivity analysis are shown in Table 2.

Two additional rounds of tree fusing taking all input trees from the previous round of analyses (Giribet 2007) were conducted for the combined analysis of molecular data under the favoured parameter set. Thereafter, the input trees from the previous round of analyses and the optimal trees from SATF were subjected to a 24-h timed search as before. Finally, all previous input trees, the optimal trees from SATF and the optimal trees from the extended timed search were subjected to 20 rounds of tree fusing under the favoured parameter set to check for heuristic stability. Nodal support for the optimal parameter set was estimated via jackknifing (250 replicates) with a probability of deletion of e^{-1} (Farris *et al.* 1996; Farris 1997).

Estimation of divergence times

Ages of clades were inferred using BEAST v. 1.6.1 (Drummond *et al.* 2002, 2006; Drummond & Rambaut 2007). We assigned the best-fitting models (a GTR model of sequence evolution with corrections for a discrete gamma distribution and a proportion of invariant sites, GTR + I + Γ) to each partition. Protein-encoding genes were partitioned into two sets by codon positions, separating third codon positions from the set of first and second positions; we erred on the side of over-parameterization to

Table 2 Tree lengths for different data partitions analysed and Wheeler ILD ($_W$ILD) values for the combined analysis of six molecular loci

	18S	28S	16S	COI	H3	H4	Combined	$_W$ILD
111	2782	6314	4290	6879	1249	368	23504	0.06900953
121	4186	9975	6633	9794	1770	526	35348	0.069706914
141	6924	16956	11076	15513	2781	828	58431	0.074498126
211	3156	7850	5053	6879	1249	368	26448	0.07157441
221	4878	12857	8045	9794	1770	526	40835	0.072609281
241	8272	22654	13780	15513	2781	828	69263	0.078469024
3211	4657	11621	7548	9794	1770	526	38635	0.070376602
3221	5717	12870	8760	13758	2498	736	47633	0.069153738
Mix	5717	12870	8760	9794	1770	526	42181	**0.065052986**

ILD, incongruence length different.

Boldface text marks the parameter set minimizing $_W$ILD. Individual data sets: 18S, 18S rRNA; 28S, 28S rRNA; 16S, 16S rRNA; COI, cytochrome *c* oxidase subunit I; H3, histone H3; H4, histone H4; Combined, combined data set (18S + 28S + 16S + COI + H3 + H4).

account for heterogeneity in substitution rates of codon positions in a taxon that originated in the Cambrian (Huang *et al.* 2004). An uncorrelated lognormal clock model was inferred for each partition, and a Yule speciation process was assumed for the tree prior. Other priors were sequentially optimized in a series of iterative test runs (data not shown). Markov chains were run for 50 000 000 generations, sampling every 1000 generations. Convergence diagnostics was assessed using TRACER v. 1.5 (Rambaut & Drummond 2009).

Sipunculan fossil taxa were used to calibrate divergence times. We constrained the origin of Golfingiiformes *sensu* Cutler to 520 Ma using the Cambrian sipunculans *Archaeogolfingia caudata* and *Cambrosipunculus tentaculatus* (Huang *et al.* 2004); a normal distribution with a standard deviation of 5 Myr was applied to this node to account for uncertainty in estimation of fossil ages. Given the topology we consistently recover in all analyses (Golfingiiformes *sensu* Cutler and part of Sipunculiformes forming a paraphyletic grade with respect to the remaining Sipuncula) and co-occurrence of *Archaeogolfingia* and *Cambrosipunculus* in the same Cambrian strata, the calibration is a conservative approach intended to avoid overestimating node ages insofar as we cannot place either fossil taxon in a derived crown group with confidence.

Taxonomy

Phylum **Sipuncula** Stephen 1964

Family **Sipunculidae** Rafinesque, 1814

Sipuncula Rafinesque, 1814: p. 32 (partim.).

Sipunculidae Baird, 1868: p. 77 (partim.); Stephen & Edmonds 1972: p. 19 (partim.); Cutler & Gibbs 1985: p. 166; Fig. 2 (partim.); Gibbs and Cutler 1987: p. 47 (partim.), Cutler 1994: p. 24 (partim.).

Type genus: *Sipunculus* Linnaeus, 1766.

Type species: *Sipunculus nudus* Linnaeus, 1766.

Valid genera[1] included: *Sipunculus* Linnaeus, 1766; *Xenosiphon* Fisher, 1947.

Diagnosis

Large sipunculans (trunk up to 45 cm in length). Introvert shorter than the trunk, covered with prominent papillae arranged irregularly. Hooks absent. Circular and longitudinal muscle layers divided into distinct bands. Body wall with coelomic extensions in the form of parallel longitudinal canals extending through most of the trunk length, or short diagonal canals limited in length to the width of one circular muscle band. Two protractor muscles may be developed (present in *S. mundanus*, *X. branchiatus*, *X. asconditus*).

Remarks

According to the traditional classification, *Xenosiphon* is sister to *Sipunculus*. This division has been called into question on the basis of phylogenetic study (Schulze *et al.* 2007), suggesting that *Xenosiphon* falls within *Sipunculus*. This same relationship is recovered in this study under the majority of analytical parameters or methodologies explored. However, the placement for *Xenosiphon* within *Sipunculus* is not supported in parsimony, and in two suboptimal parameter sets explored, we in fact recover a monophyletic *Sipunculus*. Morphologically, both genera are similar, but the presence of a pair of protractor muscles in *Xenosiphon* distinguishes them. The validity of the elusive ditypic *Xenosiphon* as a genus separate from *Sipunculus* remains to be investigated further.

Family **Golfingiidae** Stephen & Edmonds 1972

Sipunculidae Baird, 1868: p. 77 (partim.); Stephen & Edmonds 1972: p. 19 (partim.); Cutler & Gibbs 1985:

[1]For updated lists of synonymized genera consult WoRMS (World Register of Marine Species) online.

p. 166; Fig. 2 (partim.); Gibbs & Cutler 1987: p. 47 (partim.), Cutler 1994: p. 24 (partim.).

Golfingiidae Stephen & Edmonds 1972: p. 77 (partim.); Cutler & Gibbs 1985: p. 167 (partim.); Gibbs & Cutler 1987: p. 50 (partim.); Cutler 1994: p. 60 (partim.).

Themistidae Cutler & Gibbs 1985: p. 167; Fig. 2; Gibbs & Cutler 1987: p. 53; Cutler 1994: p. 140.

Phascolionidae Cutler & Gibbs 1985: p. 166-167; Fig. 2; Gibbs & Cutler 1987: p. 51 (partim.); Cutler 1994: p. 107.

Type genus: *Golfingia* Lankester, 1885.

Type species: *Golfingia macintoshii* Lankester, 1885 (= *Sipunculus vulgaris sensu* de Blainville, 1827).

Valid genera included: *Themiste* Gray, 1828, **new familial assignment**; *Phascolion* Théel, 1875, **new familial assignment**; *Golfingia* Lankester, 1885; *Onchnesoma* Koren & Danielssen, 1875, **new familial assignment**; *Nephasoma* Pergament, 1940; *Phascolopsis* (Fisher 1950), **new familial assignment**; *Thysanocardia* (Fisher 1950).

Diagnosis

Small to medium sipunculans (trunk no longer than 20 cm). Hooks may be deciduous, simple when present, not sharply curved and generally scattered, except in three species – *Golfingia elongata* (Keferstein, 1862), *Golfingia pectinatoides* E. Cutler and Cutler 1979, and *Nephasoma rimicola* (Gibbs, 1973) – where hooks are arranged in rings. Body wall with a continuous muscle layer, except in *Phascolopsis*, where the longitudinal muscles are divided in anastomosing bands.

Remarks

Golfingiidae *sensu* Gibbs and Cutler (1987) originally comprised the three genera *Golfingia*, *Nephasoma* and *Thysanocardia*. Our study confirms the previous finding by Schulze *et al.* (2007), where they recover a clade including nine different genera, with a high degree of morphological diversity. As in the latter study, *Thysanocardia* and *Themiste* were recovered as clades and both genera are morphologically clearly defined by the tentacular crown. This new family corresponds to Golfingiiformes (*sensu* Gibbs & Cutler 1987), barring *Phascolopsis*, a monotypic genus, which was placed within Sipunculidae by Cutler & Gibbs (1985). *Phascolopsis gouldii* has a complex taxonomic history and has been previously associated with species that are considered Golfingiidae (Fisher 1950). Our results confirm the finding of previous studies (Maxmen *et al. 2003*; Schulze *et al.* 2007) and corroborate the evolutionary relationship of *Phascolopsis* as well. The validity of the elusive monotypic *Phascolopsis* as a genus separate from *Golfingia* remains to be investigated. In our analysis, the relationships within this group remain poorly supported, and future revision of this family will require an increase in taxonomic sampling and the number of the target loci.

Family **Aspidosiphonidae** Baird, 1868

Aspidosiphonidae Baird, 1868: p. 100; Stephen & Edmonds 1972: p. 215; Cutler & Gibbs 1985: p. 166; Fig. 2, Gibbs & Cutler, 1987: p. 55; Cutler 1994: p. 199.

Loxosiphonidae Baird, 1868: p. 103.

Type genus: *Aspidosiphon* Diesing, 1851.

Type species: Aspidosiphon muelleri, Diesing, 1851.

Valid genera included: *Aspidosiphon* Diesing, 1851; *Cloeosiphon* Grübe, 1868.

Diagnosis

Generally small (up to 30 mm) sipunculans with smooth trunk and two retractor muscles. Introvert protruding at 45–90° angle ventral to main axis of trunk. Muscle layers smooth and continuous (most *Aspidosiphon* and *Cloeosiphon*), or with longitudinal muscle layer separated in anastomosing bundles (some *Aspidosiphon*). Anal shield consisting of hardened structure at anterior end or both ends of trunk.

Family **Phascolosomatidae** Stephen & Edmonds 1972

Sipunculidae Baird, 1868: p. 77 (partim.); Phascolosomatidae Stephen & Edmonds 1972: p. 269 (partim.); Gibbs & Cutler 1987: p. 54 (partim.); Cutler 1994: p. 156 (partim.).

Type genus: *Phascolosoma* Leuckart, 1828.

Type species: Phascolosoma granulatum Leuckart, 1828.

Valid genera included: *Phascolosoma* Leuckart, 1828; *Apionsoma* Sluiter, 1902.

Diagnosis

Small to medium sipunculans (trunk up to 12 cm in length). Hooks recurved, usually with internal structures, and closely packed in regularly spaced rings (absent in *Apionsoma trichocephalus* Sluiter, 1902). The external trunk wall is rough with obvious papillae. In *Apionsoma*, papillae are concentrated at the posterior end of the trunk. Internally, the longitudinal muscle is subdivided into anastomosing bands except in two subgenera – *Phascolosoma* (*Fisherana*) and *Apionsoma* (*Apionsoma*) – where this layer is thinner and continuous. Contractile vessel is smooth but may be large with bulbous pouches or swelling in a few *Phascolosoma*.

Remarks

Traditionally Phascolosomatidae consisted of the clade (*Phascolosoma* + *Apionsoma* + *Antillesoma*). In our study, we recover a sister relationship of *Phascolosoma* with *Apionsoma* under three analytical treatments (and with significant support in two of these), but a sister relationship of *Apionsoma* with Aspidosiphonidae in a fourth topology (albeit without support). Morphologically, *Apionsoma* lacks the anal shield, the defining synapomorphy of Aspidosiphonidae, which unambiguously disfavours the placement of *Apionsoma* as sister to, or nested within, aspidosiphonids.

Family **Siphonosomatidae fam. nov.**

Sipunculidae Baird, 1868: p. 77 (partim.); Stephen & Edmonds 1972: p. 19 (partim.); Cutler & Gibbs 1985: p. 166; Fig. 2 (partim.); Gibbs & Cutler, 1987: p. 47 (partim.), Cutler 1994: p. 24 (partim.).

Type genus: *Siphonosoma* Spengel, 1912 by present designation.

Type species: *Siphonosoma australe* (Keferstein, 1865), by subsequent designation; Gerould, 1913.

Genera included: *Siphonosoma* Spengel, 1912, **new familial assignment;** *Siphonomecus* Fisher, 1947, **new familial assignment.**

Diagnosis

Large to medium sipunculans (trunk up to 50 cm in length). Introvert much shorter than trunk with prominent conical papillae and/or hooks arranged in rings. Body wall with small, irregular saclike coelomic extensions. Circular and longitudinal muscle layers gathered into anastomosing, sometimes indistinct bands.

Remarks

Both *Siphonosoma* and *Siphonomecus* were originally placed in Sipunculidae (Cutler & Gibbs 1985; Gibbs and Cutler 1987), but this classification has been called into question on the basis of multiple phylogenetic studies (e.g. Maxmen *et al.* 2003; Schulze *et al.* 2005, 2007), which suggested that *Siphonosoma* was somehow related to Aspidosiphonidae or Phascolosomatidae (to date, *Siphonomecus* has not been available for molecular analysis). In this study, we obtain support for the sister relationship of *Siphonosoma* with (Aspidosiphonidae + Phascolosomatidae *sensu* Gibbs and Cutler (1987)); *Siphonosoma* is never recovered within or even sister to (*Sipunculus* + *Xenosiphon*) under any analytical parameters or methodologies explored.

A series of morphological characters divides true Sipunculidae from *Siphonosoma* and supports the placement of the latter with the monotypic genus *Siphonomecus*. For example, true Sipunculidae lack the anastomosing muscle bands and the posterior attachment of the spindle muscle. Moreover, whereas the extensions of the body wall in *Siphonosoma* and *Siphonomecus* are sac-like, those of *Sipunculus* and *Xenosiphon* form canals. The inclusion of *Siphosonoma* (or *Siphonomecus*) in Sipunculidae renders the latter para- or polyphyletic and obscures a clear diagnosis of Sipunculidae. Consequently, we favour a familial designation for *Siphonosoma* and *Siphonomecus*, which, taken together with the new familial designation of *Phascolopsis* (discussed above), renders Sipunculidae monophyletic. The validity of the elusive monotypic *Siphonomecus* as a genus separate from *Siphonosoma* remains to be investigated.

Family **Antillesomatidae fam. nov.**

Phascolosomatidae Stephen & Edmonds 1972: p. 269 (partim.); Cutler & Gibbs 1985: pp. 166–167; Fig. 2 (partim.); Gibbs & Cutler, 1987: p. 51 (partim.), Cutler 1994: p. 156 (partim.).

Type genus: *Antillesoma* Stephen & Edmonds 1972 by present designation.

Type species: *Antillesoma antillarum* (Grübe & Oersted, 1858), by monotypy.

Genera included: *Antillesoma* Stephen & Edmonds 1972 **new familial assignment.**

Diagnosis

Mid-sized sipunculan (trunk up to 8 cm). Distal part of the introvert smooth and white, proximal portion bears dark papillae and marked off by a distinctive collar. Hooks absent in adults but few hooks present in small individuals (<1 cm). Oral disc consisting of nuchal organ enclosed by numerous tentacles, which vary in number according to size (from 30 to 200 in adults). Body wall with longitudinal muscle layer gathered into anastomosing bands. Contractile vessel with many villi. Four introvert retractor muscles with the lateral pair often extensively fused.

Remarks

Currently, tentatively placed in Phascolosomatidae with *Phascolosoma* and *Apionsoma* (Gibbs and Cutler 1987), *Antillesoma* is a curious lineage whose phylogenetic placement is contentious. Maxmen *et al.* (2003) recovered *Antillesoma* nested within Aspidosiphonidae based on parsimony analysis of three genes. Upon adding a few terminals, Schulze *et al.* (2005) recovered *Antillesoma* sister to *Cloeosiphon* + (*Siphonosoma* + *Apionsoma*), to the exclusion of *Aspidosiphon*, based on parsimony analysis of the same three genes; adding morphological data to the analysis, they recovered the clade (*Cloeosiphon* + *Aspidosiphon* + *Apionsoma* + *Antillesoma*) – in neither case with nodal support >50%.

Schulze *et al.* (2007) thereafter redoubled sampling efforts, increasing the number of genes to four and the number of sipunculan exemplars to 99 (representing 52 species). Parsimony analysis of the molecular data set yielded the sister relationship between *Antillesoma* and *Cloeosiphon*, again without sufficient nodal support. Bayesian inference analysis of the same recovered *Antillesoma* nested within Aspidosiphonidae with a posterior probability of 0.95. Schulze *et al.* (2007) suggested that *Antillesoma* may require placement in Aspidosiphonidae, but granted that *Antillesoma* is morphologically more closely allied to Phascolosomatidae.

Antillesoma certainly differs from *Phascolosoma* and *Apionsoma* in bearing numerous villi on the contractile vessel. However, it also lacks the anal shield, the defining synapomorphy of Aspidosiphonidae. Furthermore, it is

unique among these genera in the absence of hooks in adults. In our phylogenetic analyses, we recover the traditional Phascolosomatidae (*Phascolosoma* + *Apionsoma* + *Antillesoma*) in two topologies and a sister relationship of *Antillesoma* with Aspidosiphonidae in another. In none of these is the placement supported and appears to be sensitive to analytical parameters. We thus consider *Antillesoma* to constitute an independent lineage and elevate it to family status so as to maintain the monophyly and diagnosability of previously described families.

Results

Phylogenetic analysis

Runs of MrBayes v. 3.1.2 reached stationarity in 2 000 000 generations; 3 000 000 generations (25%) were hence discarded as burn-in. The BI analysis with all taxa resulted in a tree topology with five clades, of which four were supported (Fig. 2). In order of proximity to the root, these clades are as follows: (*Sipunculus* + *Xenosiphon*) (PP = 1.00); Golfingiiformes *sensu* Cutler, but including *Themiste* and *Phascolopsis* (PP = 1.00); *Siphonosoma* (PP = 1.00); (*Antillesoma* + Aspidosiphonidae *sensu* Cutler) (PP = 0.79); and (*Apionsoma* + *Phascolosoma*) (PP = 0.99). In the first clade, *Xenosiphon* is nested within *Sipunculus*. Intergeneric relationships within the second clade ('Golfingiiformes') are largely obscure; barring the sister relationship between *Themiste* and *Thysanocardia*, which form supported, reciprocally monophyletic groups, the genera in this clade are generally para- or polyphyletic. Within the fourth clade, the sister relationship of *Antillesoma* with Aspidosiphonidae is unsupported, although the monophyly of the latter is not (PP = 0.98). We observe that *Lithacrosiphon cristatus* is deeply nested inside *Aspidosiphon*, rendering the latter paraphyletic (PP = 1.00, 1.00, 0.99) ('*Aspidosiphon cristatus*' in the figures, subsequent to taxonomic action, discussed below).

Maximum likelihood analysis resulted in a tree topology largely similar to the BI tree (Fig. 3). The notable topological difference is the placement of *Antillesoma*, which is recovered sister to (*Apionsoma* + *Phascolosoma*), again without support. Differences exist with respect to nodal support, particularly for nodes corresponding to interfamilial relationships, the monophyly of 'Golfingiiformes' and the monophyly of both *Apionsoma* and *Phascolosoma*.

Parsimony analysis under direct optimization yields a tree topology that agrees with most major aspects of the probabilistic analyses, but discords with BI and ML topologies in a number of respects (Fig. 4). The clade (*Sipunculus* + *Xenosiphon*) is recovered sister to the remaining Sipuncula (JF = 99%), but in this case, the placement for *Xenosiphon* within *Sipunculus* is not supported, justifying the maintenance of the two genera for the time being (in two suboptimal parameter sets explored, we in fact recover a monophyletic *Sipunculus*). The monophyly of 'Golfingiiformes' is recovered (JF = 54%), and relationships within this group remain poorly supported, as in the previous analyses. The mutual monophyly and sister relationship between *Themiste* and *Thysanocardia* are recovered under all parameter sets explored, but the sister relationship of the two, although stable to parameter variation, finds no resampling support. But other interfamilial relationships are not supported or demonstrably sensitive to parameter variation. Both Aspidosiphonidae *sensu* Cutler and *Siphonosoma* are recovered as polyphyletic, and *Apionsoma* is paraphyletic and not related to other phascolosomatids. However, as with *Sipunculus*, we do observe that under certain suboptimal parameter sets, *Siphonosoma* is recovered as monophyletic. Upon examining tree topologies found under the other parameter sets, we observe that the instability in the parsimony tree is largely caused by variability in the topological placement of *Antillesoma*, which is the clade that differs in placement between the ML and BI trees.

Estimation of divergence times

The run of BEAST reached stationarity after 7 000 000 generations; 10 000 000 generations (20%) were discarded as burn-in. The tree topology recovered by BEAST is similar to the ML topology, specifically with respect to the placement of *Antillesoma* as sister to Phascolosomatidae, but differs from the previous three topologies with respect to internal relationships of the five major clades (Fig. 5). Diversification of Sipuncula is dated at ca. 552 Ma. Diversification of major lineages is estimated as follows: (*Xenosiphon* + *Sipunculus*), 212 Ma; 'Golfingiiformes', 403 Ma; *Siphonosoma*, 289 Ma; Aspidosiphonidae, 283 Ma; (*Antillesoma* + *Apionsoma* + *Phascolosoma*), 279 Ma; and (*Apionsoma* + *Phascolosoma*), 244 Ma. The confidence intervals for estimated ages are imprecise, owing to the paucity of fossil taxa available to constrain the tree. A tree file of estimated ages and 95% highest posterior density (HPD) intervals for all nodes is provided in Data S1.

Although the placement of *Antillesoma* is unstable and somewhat unsupported, most interfamilial relationships are consistently recovered across analyses. A consensus of topologies across all analyses is shown in Fig. 6.

Discussion

The traditional classification of Sipuncula (Cutler & Gibbs 1985; Fig. 1) has been repeatedly challenged on the basis of molecular sequence data or the combination of molecular and morphological data (Maxmen *et al.* 2003; Schulze *et al.* 2005, 2007). Nevertheless, previous workers abstained from taxonomic action on account of inconsistencies between phylogenetic signal and morphological observations. For example, Maxmen *et al.* (2003) obtained alternative

Fig. 2 Bayesian inference analysis based on the combined data set and conducted in MrBayes v. 3.1.2 Numbers on nodes indicate posterior probabilities; asterisks correspond to PP = 1.00. Colours on branches correspond to new classification. For interpretation of references to color in figure legend, please refer to the Web version of this article.

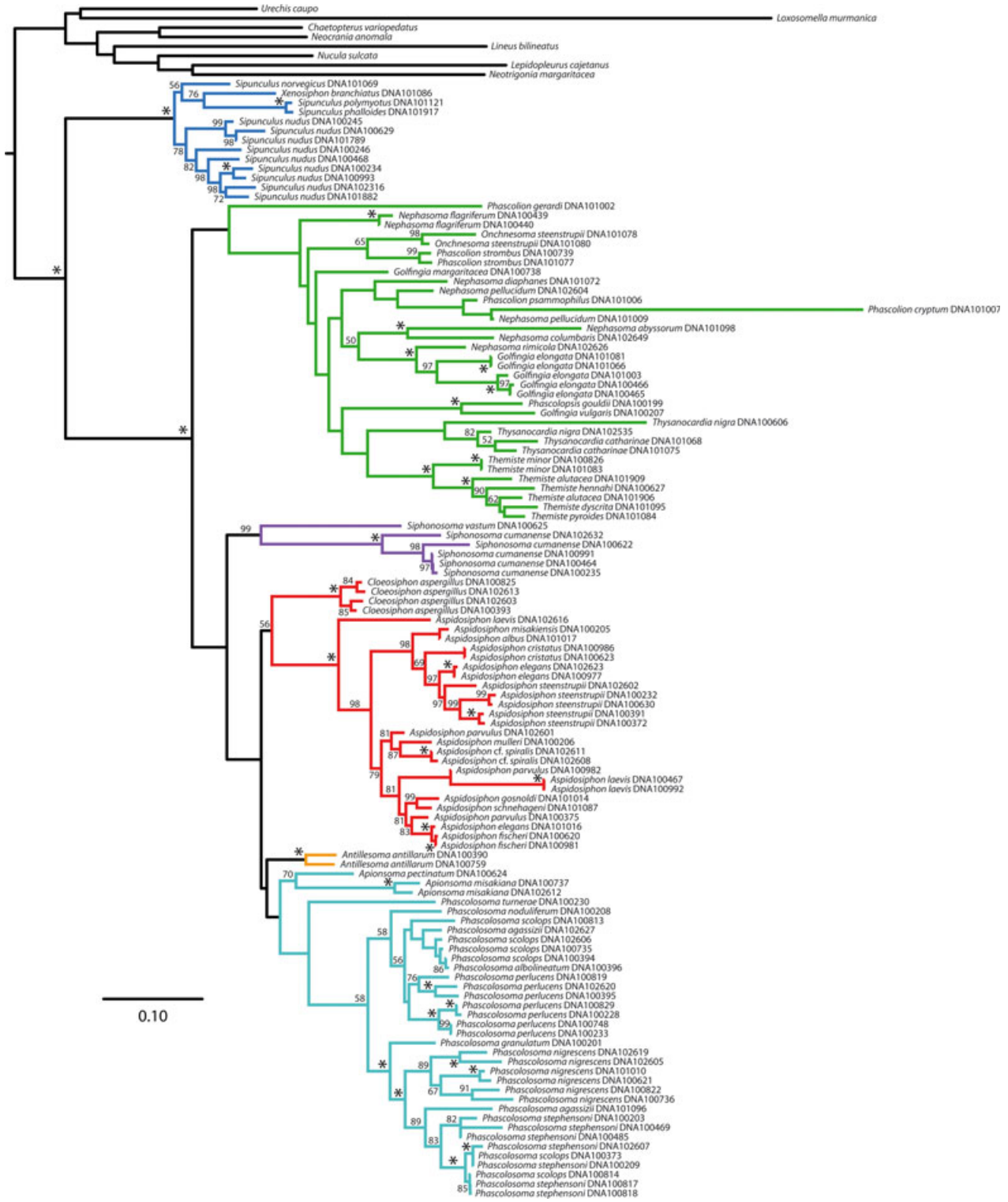

Fig. 3 Maximum likelihood analysis based on the combined data set and conducted in RAxML v. 7.2.7. Numbers on nodes indicate bootstrap resampling frequencies; asterisks correspond to BS = 100%. Colours on branches correspond to new classification. For interpretation of references to color in figure legend, please refer to the Web version of this article.

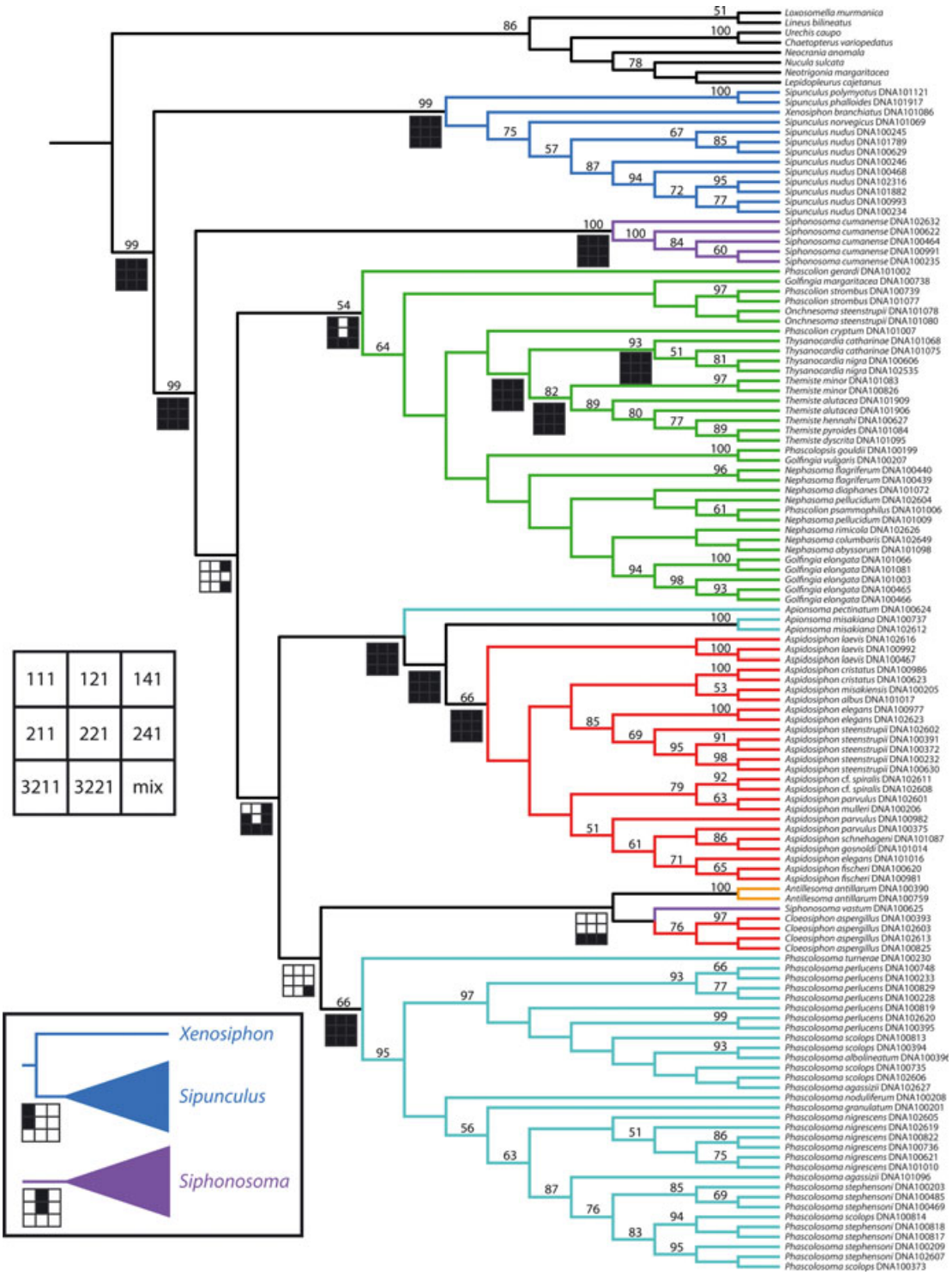

Fig. 4 Parsimony analysis under direct optimization based on the combined data set and conducted in POY v. 4.1.2. Numbers above nodes indicate jackknife resampling frequencies. Icons below nodes indicate results of sensitivity analysis to parameter variation, with black entries indicating recovery of corresponding node under a given parameter set. Parameter sets explored indicated to the left. Lower left inset: subtrees under certain suboptimal parameter sets recover monophyly of *Sipunculus* and *Siphonosoma*. Colours on branches correspond to new classification. For interpretation of references to color in figure legend, please refer to the Web version of this article.

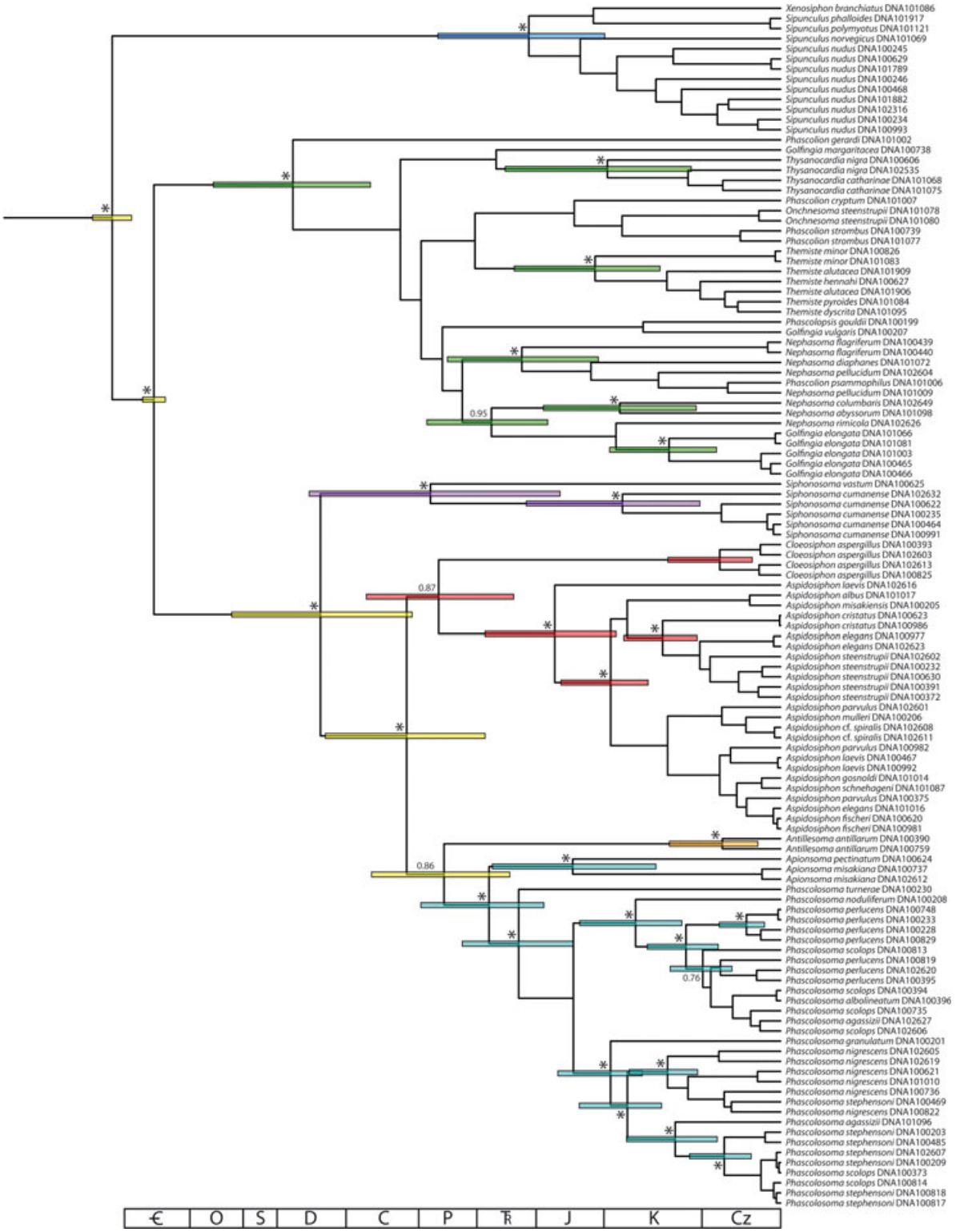

Fig. 5 Chronogram of Sipuncula inferred using BEAST V. 1.6.1 under an uncorrelated lognormal clock model and using fossil taxa for calibration (outgroups not shown). Numbers on nodes indicate posterior probabilities; asterisks correspond to PP = 1.00. Coloured bars indicate 95% highest posterior density intervals of node age estimates for selected groups. Colours of bars correspond to new classification. For interpretation of references to color in figure legend, please refer to the Web version of this article.

placements of *Apionsoma* that were contingent upon the molecular data set analysed and postulated that additional molecular sequence data would resolve the placement of this genus. Similarly, Schulze *et al.* (2007) recovered a paraphyletic Aspidosiphonidae *sensu* Gibbs and Cutler 1987 because of the inclusion of *Antillesoma*; a familial transfer in this case was unjustified because of the absence of anal shields – the putative synapomorphy of Aspidosiphonidae – in *Antillesoma*.

In the present study, we endeavoured to reassess the phylogeny of Sipuncula, employing three tactics to refine estimation of relationships. First, we increased the length of the 28S rRNA locus sequenced from 300 to 2200 bp, owing to the effectiveness of this marker for phylogenetic estimation of other marine invertebrate groups (e.g. Mallatt & Winchell 2002; Winchell *et al.* 2002; Fuchs *et al.* 2010; Jörger *et al.* 2010). Second, we sequenced two additional markers in pursuit of more robust phylogenetic signal. Finally, to address the potentially adverse effects of data incompleteness, we increased the threshold criterion for inclusion in the molecular data set. Specifically, we necessitated the availability of two amplicons of 18S rRNA and of 28S rRNA (operationally, at least 1150 and 1250 bp of sequence data, respectively) for a terminal to be included. (We did make exceptions for two well-represented genera, *Themiste* and *Thysanocardia*, for which shortfalls in sequence data in a particular gene were permitted if a conspecific terminal included sequence data from that gene; this circumvented the unique difficulty of amplifying 28S rRNA in these genera and was subsequently observed not to affect the monophyly of the genera.) We also re-sequenced suspect amplicons to address the possibility of contamination. If re-sequencing questionable amplicons and/or verification of species identification were not possible, certain terminals were excluded from the analysis *a priori* (e.g. potentially contaminated sequences of *Phascolosoma capitatum* could not be re-sequenced because of DNA degradation; the specimen of *Apionsoma murinae* is no longer available for identification because of the original extraction protocol).

The resulting topologies (Figs 2–5) are mostly congruent with respect to clade composition, but in the parsimony topology (Fig. 4), interfamilial relationships are sensitive to parameter variation or unsupported. Monophyly of Sipuncula is recovered across all analyses with high support, as is a sister relationship between the clade (*Sipunculus* + *Xenosiphon*) (henceforth Sipunculidae) and the remaining Sipuncula. The BI, ML and BEAST topologies (Figs 2, 3 and 5, respectively) uniformly recover the following four clades with variable support: (i) (Golfingiiformes *sensu* Gibbs and Cutler 1987 + *Phascolopsis*), henceforth Golfingiidae; (ii) *Siphonosoma*; (iii) Aspidosiphonidae *sensu* Gibbs and Cutler 1987; and (iv) the clade (*Apionsoma* + *Phascolosoma*), henceforth Phascolosomatidae. *Antillesoma* is related to either Phascolosomatidae (ML and BEAST) or Aspidosiphonidae (BI). Parsimony analysis recovers only Golfingiidae under the optimal parameter set; *Siphonosoma* is recovered as monophyletic but with low support in two suboptimal parameter sets explored. The majority of topologies recovers the relationships as follows: (Sipunculidae (Golfingiidae (*Siphonosoma* (Aspidosiphonidae + *Antillesoma* + Phascolosomatidae)))) (Figs 6–7).

Two traditional relationships previously considered suspect are vindicated by our results. For example, we recover the monophyly of *Phascolosoma* (PP_{BI} = 1.00, PP_{BEAST} = 1.00, 66% JF and 100% stability; Figs 2–4), previously recovered as paraphyletic because of the placement of *Phascolosoma turnerae* as sister to Golfingiidae (Schulze *et al.* 2007). Suspecting the maleficence of missing data, we redoubled sequencing efforts for this terminal for the 28S rRNA locus, amplified the entire target 2200-bp fragment for this species and presently recover it as sister to the remaining *Phascolosoma*. Similarly, the monophyly of Aspidosiphonidae, previously considered para- or polyphyletic (Schulze *et al.* 2005, 2007), was recovered in the majority of topologies, and in two with support (PP_{BI} = 0.98, BS = 56%; Figs 2, 3 and 5). We therefore maintain the systematic validity of these taxa.

However, consistent with phylogenetic studies heretofore, we observe discrepancies between the current classification of Sipuncula and the molecular phylogeny of the group, necessitating a series of taxonomic actions that we undertake here. The taxonomic changes are as follows:

1 As detailed in the Taxonomy section, we erect two new families, Siphonosomatidae fam. nov. and Antillesomatidae fam. nov., to recognize the distinctness of the constituent lineages and to maintain both the monophyly and the systematic utility of the remaining families.

2 As *Themiste*, *Phascolion* and *Onchnesoma* have repeatedly been placed in a clade with archetypal golfingiid genera (on the basis of molecular data or the combination of molecular and morphological data), we consider the families Themistidae, **new synonymy** and Phascolionidae, **new synonymy** to constitute junior subjective synonyms of Golfingiidae. Generic reassignment of many species remains to be carried out in future studies focusing on this clade.

3 Concordant with previous and present results, we transfer the monotypic genus *Phascolopsis* from Sipunculidae to Golfingiidae, **new familial assignment**. This action, in concert with the previous synonymy, renders Golfingiidae a monophyletic family. The non-monophyly of several constituent golfingiid genera suggests the need for extensive revision in this group (discussed later).

4 Multiple studies have repeatedly obtained a paraphyletic *Aspidosiphon* with respect to *Lithacrosiphon*. The two

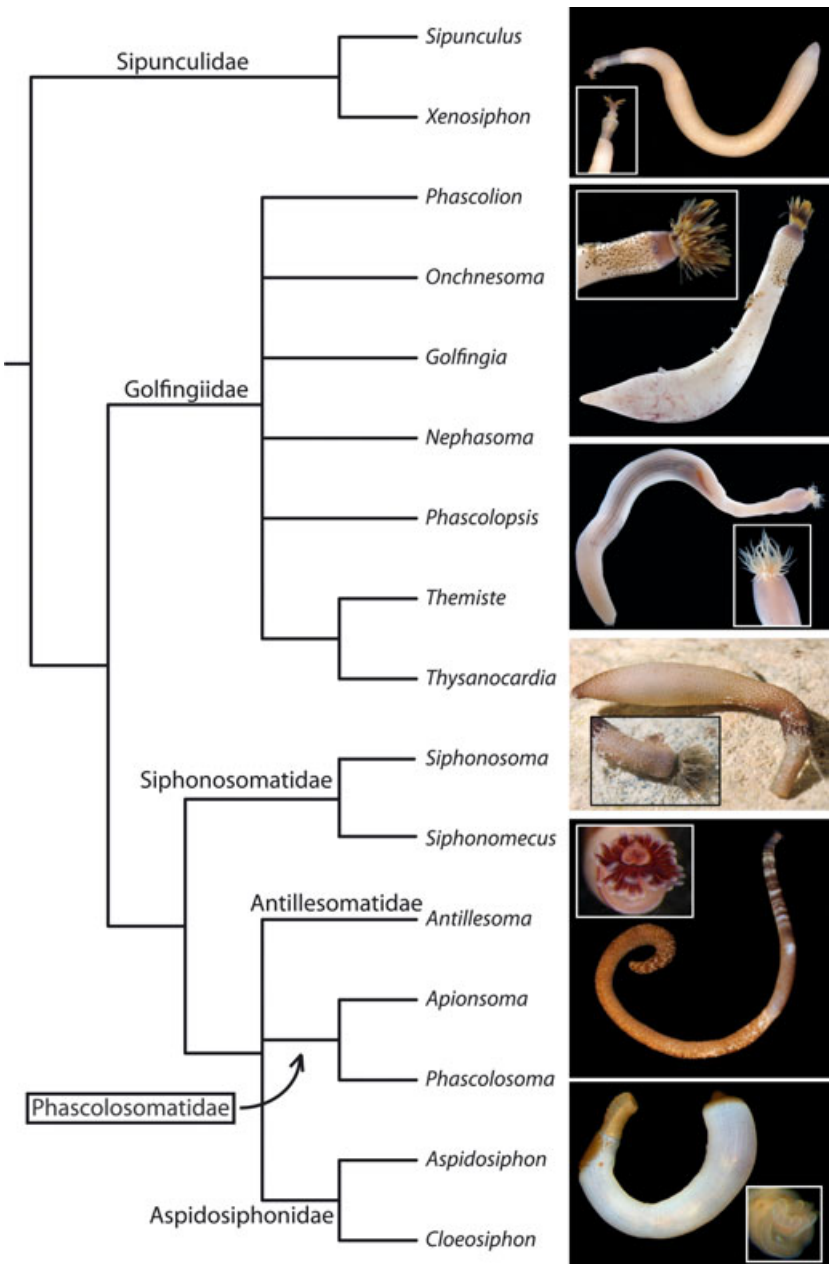

Fig. 6 New classification of Sipuncula. Cladogram reflects majority rule consensus of all tree topologies. Right, live exemplars of families. From top: *Sipunculus nudus*, *Themiste alutacea*, *Siphonosoma cumanense*, *Antillesoma antillarum*, *Phascolosoma nigrescens*, *Aspidosiphon fisheri*. Inset: Detail of tentacles.

genera were previously distinguished on the basis of anal shield shape, which we deem insufficient justification for the maintenance of *Lithacrosiphon* as a separate genus. We consider *Lithacrosiphon* to constitute a junior subjective synonym of *Aspidosiphon*, **new synonymy**. The former type species becomes *Aspidosiphon cristatus* Sluiter, 1902, **restored combination**, followed by the subspecies *Aspidosiphon cristatus lakshadweepensis* (Haldar, 1991), **new combination**. The second species of this genus becomes *Aspidosiphon maldivensis* Shipley, 1902, **restored combination**. This renders *Aspidosiphon* a monophyletic genus.

We therefore present a new classification of Sipuncula (Fig. 6) incorporating the aforementioned taxonomic actions and reflecting the most recent phylogeny of the group.

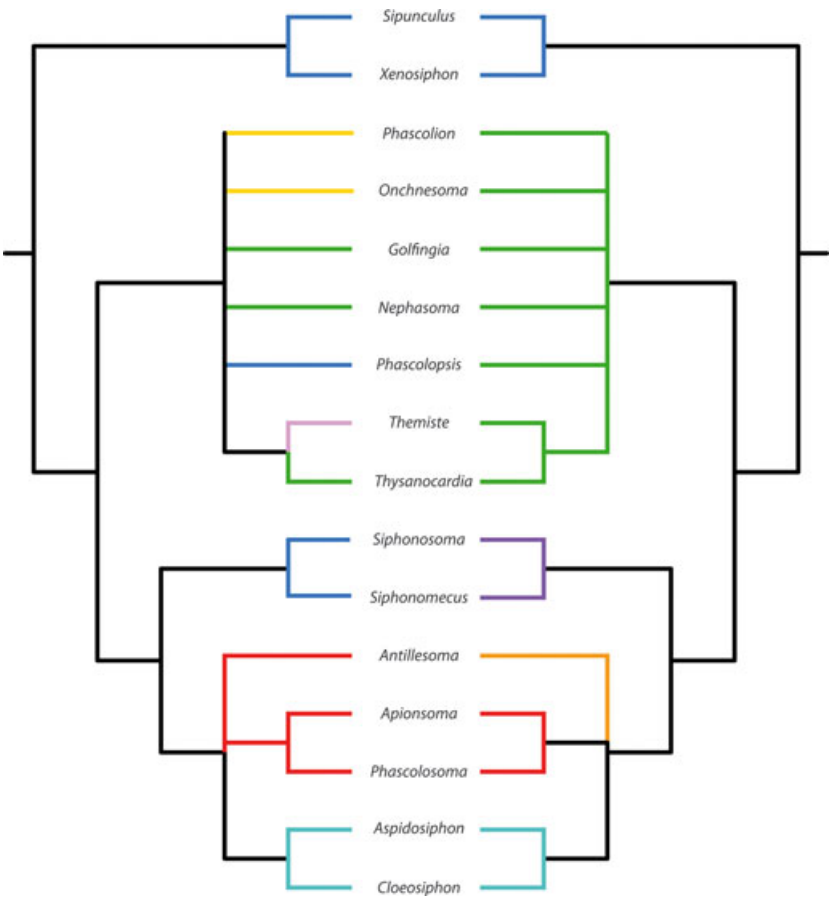

Fig. 7 Comparison of classification systems. Left, majority rule consensus cladogram, with colours corresponding to traditional classification (*sensu* Cutler & Gibbs 1985). Right, majority rule consensus cladogram, with colours corresponding to revised classification. Blue: Sipunculidae; gold: Phascolionidae; green: Golfingiidae; pink: Themistidae; red: Phascolosomatidae; cyan: Aspidosiphonidae; violet: Siphonosomatidae fam. nov.; orange: Antillesomatidae fam. nov. For interpretation of references to color in figure legend, please refer to the Web version of this article.

Family Sipunculidae Rafinesque, 1814
  Sipunculus Linnaeus, 1766
  Xenosiphon Fisher, 1947
Family Siphonosomatidae new family
  Siphonosoma Spengel, 1912 **new familial assignment**
  Siphonomecus Fisher, 1947 **new familial assignment**
Family Golfingiidae Stephen & Edmonds 1972
  Themiste Gray, 1828 **new familial assignment**
  Phascolion Théel, 1875 **new familial assignment**
  Golfingia Lankester, 1885
  Onchnesoma Koren & Danielssen, 1875 **new familial assignment**
  Nephasoma Pergament, 1940
  Phascolopsis (Fisher 1950) **new familial assignment**
  Thysanocardia (Fisher 1950)
Family Antillesomatidae new family
  Antillesoma (Stephen & Edmonds 1972) **new familial assignment**
Family Aspidosiphonidae Baird, 1868
  Aspidosiphon Diesing 1851
  Cloeosiphon Grübe, 1868
Family Phascolosomatidae Stephen & Edmonds 1972
  Phascolosoma Leuckart, 1828
  Apionsoma Sluiter, 1902

Consistent with the incidence of the earliest Sipuncula in Cambrian fossil beds (Huang *et al.* 2004), the molecular dating of the group indicates that the families, as defined here, are ancient groups, with origins dating to the Mesozoic or earlier (Fig. 5). Barring the case of *Antillesoma*, the diversification of the families also occurs in the Mesozoic or earlier. We observe that several clades identified as a single species are very old (e.g. *Siphonosoma cumanense*, 136 Ma; *Aspidosiphon steenstrupii*, 59.2 Ma), perhaps indicating the incidence of cryptic species (discussed below). However, as very few fossils are presently available, the 95% HPD intervals of these estimates are large, spanning entire geological periods. The discovery of additional sipunculan fossils, particularly from Mesozoic strata, is anticipated to limit the variance on divergence time estimates.

One of the outstanding issues in sipunculan systematics is the internal relationships of Golfingiidae as revised here. Beyond the mutual monophyly of *Themiste* and *Thysanocardia* (recovered across all analyses), constituent genera are largely non-monophyletic and internal relationships are unsupported. This likely results from missing data in the 28S rRNA data partition. Specifically, the first and most variable amplicon of 28S rRNA could not be sequenced for most *Nephasoma*, *Onchnesoma* and *Phascolion*, although these taxa were successfully sequenced for the remaining

(and largely conserved) ca. 1300 bp. While this merited inclusion of these taxa in the phylogeny, a lack of informative characters at this level may be driving the instability of golfingiid internal relationships. While generic revision in Golfingiidae is beyond the scope of the present study, future efforts are directed toward increasing golfingiid taxonomic sampling and the sequencing of informative gene fragments.

Another concern for sipunculan systematics is the frequent non-monophyly of species in the topologies recovered (e.g. *Phascolosoma scolops*, *P. stephensoni* and *Aspidosiphon parvulus*). The repeated incidence of this phenomenon and verification of sequence data suggest that incorrect identification, cryptic speciation or a combination of the two may be prevalent in this group of marine worms, where an unusually large proportion of species are cosmopolitan or spread across several ocean basins. Incorrect identification may stem from Cutler's final revision (Cutler 1994), wherein 320 species were reduced to 149 subsequent to drastic synonymies. The extensive synonymy of former species that are in fact valid may engender the non-monophyly of these taxa as presently defined. In one case, the test of a putatively cosmopolitan species' validity (*Phascolosoma perlucens*) demonstrated significant population structure and suggested that some synonyms of this species may have been valid entities (Kawauchi & Giribet 2010). Meticulous case-by-case examination of species' integrity, in concert with molecular tools, is therefore required for future systematic efforts.

Conclusions

Our multilocus molecular phylogeny corroborates the need for a revised classification of Sipuncula. *Apropos*, we provide a new hierarchical classification of the phylum, erecting two new families, and thereby rendering the six resulting families monophyletic entities. The consistency of our results across multiple analytical treatments attests to the phylogenetic informativeness of the 28S rRNA locus (as well as the additionally sequenced loci histone H4 and 16S rRNA) and to the potentially adverse effects of missing data in previous efforts. We anticipate that continued efforts toward completion of the 28S rRNA data set (with concomitant addition of amplicons not previously sequenced) may provide a definitive placement of Antillesomatidae fam. nov. and aid the revision of the genera within Golfingiidae.

Acknowledgements

We dedicate this study to the late Ed Cutler, great mentor and friend, and to Ann Covert who supported this sipunculan research through a MarCraig Memorial Fund Grant. Without them, none of this work would have been conducted or even have been possible. We are also indebted to Anja Schulze, Harlan Dean, Iñaki Saiz Salinas and Mary Rice, who through the years have helped us in many aspects of sipunculan taxonomy and systematics, and who have been great colleagues, provided samples and aided with identifications. Fieldwork to New Caledonia was supported by the Putnam Expedition Grants program of the Museum of Comparative Zoology. Claudio G. Tiago assisted with fieldwork in New Caledonia. David Paulaud from Direction de l' Environnement (Province Sud) facilitated collecting permits for New Caledonia. Bertrand Richer De Forges, Jean-Lou Justine and Claire Goiran (IRD-Noumea) assisted with logistics in Noumea. Mike Ethier and the Research Computing team (FAS, Harvard University) facilitated cluster operations. Per Sundberg and an anonymous reviewer provided comments that improved an earlier version of this manuscript. Part of this work was also funded by the protostome AToL grant #EF-0531757 to G.G.

References

Bleidorn, C., Podsiadlowski, L. & Bartolomaeus, T. (2006). The complete mitochondrial genome of the orbiniid polychaete *Orbinia latreillii* (annelida, Orbiniidae) – a novel gene order for annelida and implications for annelid phylogeny. *Gene*, 370, 96–103.

Boore, J. L. & Staton, J. L. (2002). The mitochondrial genome of the sipunculid *Phascolopsis gouldii* supports its association with Annelida rather than Mollusca. *Molecular Biology and Evolution*, 19, 127–137.

Boyer, S. L. & Giribet, G. (2007). A new model gondwanan taxon: Systematics and biogeography of the harvestman family Pettalidae (Arachnida, Opiliones, Cyphophthalmi), with a taxonomic revision of genera from Australia and New Zealand. *Cladistics*, 23, 337–361.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17, 540–552.

Colgan, D.J., McLauchlan, A., Wilson, G.D.F., Livingston, S.P., Edgecombe, G.D., Macaranas, J., Cassis, G. and Gray, M.R., (1998). Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. *Australian Journal of Zoology*, 46, 419–437.

Cutler, E. B. (1979). A reconsideration of the sipunculan taxa *Fisherana* Stephen, *Mitosiphon* Fisher and *Apionsoma* Sluiter. *Zoological Journal of the Linnean Society*, 65, 367–384.

Cutler, E. B. (1984). Proposed conservation of *Siphonosoma cumanense* (Keferstein, 1867) over *S. Edule* (Pallas, 1774) (Sipuncula). *Bulletin of Zoological Nomenclature*, 41, 62–64.

Cutler, E. B. (1994). *The Sipuncula. Their Systematics, Biology and Evolution*. Ithaca, NY: Cornell University Press.

Cutler, E. B. & Cutler, N. J. (1982). A revision of the genus *Siphonosoma* (Sipuncula). *Proceedings of the Biological Society of Washington*, 95, 748–762.

Cutler, E. B. & Cutler, N. J. (1983). An examination of the *Phascolosoma* subgenera *Antillesoma*, *Rueppellisoma* and *Satonus* (Sipuncula). *Zoological Journal of the Linnean Society*, 77, 175–187.

Cutler, E. B. & Cutler, N. J. (1985a). A revision of the genera *Phascolion* Theel and *Onchnesoma* Koren and Danielssen (Sipuncula). *Proceedings of the Biological Society of Washington*, 98, 809–850.

Cutler, E. B. & Cutler, N. J. (1985b). A revision of the genera *Sipunculus* and *Xenosiphon* (Sipuncula). *Zoological Journal of the Linnean Society*, 85, 219–246.

Cutler, E. B. & Cutler, N. J. (1986). A revision of the genus *Nephasoma* (Sipuncula: Golfingiidae). *Proceedings of the Biological Society of Washington*, 99, 547–573.

Cutler, E. B. & Cutler, N. J. (1987). A revision of the genus *Golfingia* (Sipuncula: Golfingiidae). *Proceedings of the Biological Society of Washington*, 100, 735–761.

Cutler, E. B. & Cutler, N. J. (1988). A revision of the genus *Themiste* (Sipuncula). *Proceedings of the Biological Society of Washington*, 101, 741–766.

Cutler, E. B. & Cutler, N. J. (1989). A revision of the genus *Aspidosiphon* (Sipuncula). *Proceedings of the Biological Society of Washington*, 102, 826–865.

Cutler, N. J. & Cutler, E. B. (1990). A revision of the subgenus *Phascolosoma* (Sipuncula, *Phascolosoma*). *Proceedings of the Biological Society of Washington*, 3, 691–730.

Cutler, E. B. & Gibbs, P. E. (1985). A phylogenetic analysis of higher taxa in the phylum Sipuncula. *Systematic Zoology*, 34, 162–173.

Cutler, E. B. & Jurczak, E. (1975). The sipunculan genus *Lithacrosiphon* Shipley, a taxonomic review. *Zoological Journal of the Linnean Society*, 56, 235–248.

Cutler, E. B. & Murina, V. V. (1977). On the sipunculan genus *Golfingia* lankester, 1885. *Zoological Journal of the Linnean Society*, 60, 173–189.

Cutler, E. B., Cutler, N. J. & Nishikawa, T. (1982). *Siphonosoma cumanense* and *S. edule* (sipuncula): a reevaluation of their separate status. *Publications of the Seto Marine Biological Laboratory*, 27, 265–267.

Cutler, E. B., Cutler, N. J. & Gibbs, P. E. (1983). A reconsideration of the *Golfingia* subgenera *Golfingiella* and *Siphonoides* (Sipuncula). *Proceedings of the Biological Society of Washington*, 96, 669–694.

De Laet, J. E. (2005). Parsimony and the problem of inapplicables in sequence data. In V. A. Albert (Ed.) *Parsimony, Phylogeny, and Genomics* (pp. 81–116). Oxford: Oxford University Press.

Dordel, J., Fisse, F., Purschke, G. & Struck, T. H. (2010). Phylogenetic position of Sipuncula derived from multi-gene and phylogenomic data and its implication for the evolution of segmentation. *Journal of Zoological Systematics and Evolutionary Research*, 48, 197–207.

Drummond, A. J. & Rambaut, A. (2007). Beast: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7, 214–221.

Drummond, A. J., Nicholls, G. K., Rodrigo, A. G. & Solomon, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*, 161, 1307–1320.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4, 699–710.

Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E.C., Rouse, G.W., Obst, M., Edgecombe, G.D., Sørensen, M.V., Haddock, S.H.D., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R.M., Wheeler, W.C., Martindale, M.Q. Giribet, G. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, 452, 745–749.

Edgar, R. C. (2004). Muscle: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5, 1–19.

Edgecombe, G. D. & Giribet, G. (2006). A century later – a total evidence re-evaluation of the phylogeny of scutigeromorph centipedes (Myriapoda: Chilopoda). *Invertebrate Systematics*, 20, 503–525.

Edgecombe, G. D., Giribet, G., Dunn, C. W., Hejnol, A., Kristensen, R. M. & Neves, R. C. (2011). Higher-level metazoan relationships: recent progress and remaining questions. *Organisms, Diversity & Evolution*, 11, 151–172.

Farris, J. S. (1997). The future of phylogeny reconstruction. *Zoologica Scripta*, 26, 303–311.

Farris, J. S., Albert, V. A., Källersjö, M., Lipscomb, D. & Kluge, A. G. (1996). Parsimony jackknifing outperforms neighbor-joining. *Cladistics*, 12, 99–124.

Fisher, W. K. (1950). The sipunculid genus *Phascolosoma*. *Annals and Magazine of Natural History (ser. 12)*, 3, 547–552.

Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. C. (1994). DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3, 294–299.

Fuchs, J., Iseto, T., Hirose, M., Sundberg, P. & Obst, M. (2010). The first internal molecular phylogeny of the animal phylum Entoprocta (Kamptozoa). *Molecular Phylogenetics and Evolution*, 56, 370–379.

Gibbs, P. E. & Cutler, E. B. (1987). A classification of the phylum Sipuncula. *Bulletin of the British Museum Natural History (Zoology)*, 52, 43–58.

Gibbs, P. E., Cutler, E. B. & Cutler, N. J. (1983). A review of the genus *Thysanocardia* Fisher (Sipuncula). *Zoologica Scripta*, 12, 295–304.

Giribet, G. (2007). Efficient tree searches with available algorithms. *Evolutionary Bioinformatics*, 3, 341–356.

Giribet, G., Carranza, S., Baguñà, J., Riutort, M. and Ribera, C. (1996). First molecular evidence for the existence of a Tardigrada + Arthropoda clade. *Molecular Biology and Evolution*, 13, 76–84.

Goloboff, P. A. (1999). Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics*, 15, 415–428.

Goloboff, P. A. (2002). Techniques for analyzing large data sets. In R. Desalle, G. Giribet & W. Wheeler (Eds) *Techniques in molecular systematics and evolution* (pp. 70–79). Basel: Brikhäuser Verlag.

Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G. W., Edgecombe, G. D. *et al.* (2009). Assessing the root of bilaterian animals with scalable phylogenomic methods.

Proceedings of the Royal Society B-Biological Sciences, 276, 4261–4270.

Huang, D. Y., Chen, J. Y., Vannier, J. & Saiz Salinas, J. I. (2004). Early Cambrian sipunculan worms from southwest China. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 271, 1671–1676.

Huelsenbeck, J. P. & Ronquist, F. (2005). Bayesian analyses of molecular evolution using MrBayes. In R. Nielsen (Ed.) *Statistical Methods in Molecular Evolution* (pp. 183–226). New York, NY: Springer.

Hyman, L. H. (1959). The protostomatous coelomates – Phylum Sipunculida. In L. H. Hyman (Ed.) *The Invertebrates: Smaller Coelomate Groups* (pp. 610–696). Nova Iorque: MacGraw-Hill Book Company.

Jennings, R. M. & Halanych, K. M. (2005). Mitochondrial genomes of *Clymenella torquata* (Maldanidae) and *Riftia pachyptila* (Siboglinidae): evidence for conserved gene order in Annelida. *Molecular Biology and Evolution*, 22, 210–222.

Jörger, K. M., Stöeger, I., Kano, Y., Fukuda, H., Knebelsberger, T. & Schrödl, M. (2010). On the origin of Acochlidia and other enigmatic euthyneuran gastropods, with implications for the systematics of Heterobranchia. *BMC Evolutionary Biology*, 10, 323.

Kawauchi, G. Y. & Giribet, G. (2010). Are there true cosmopolitan sipunculan worms? A genetic variation study within *Phascolosoma perlucens* (Sipuncula, Phascolosomatidae). *Marine Biology*, 157, 1417–1431.

Kristof, A., Wollesen, T. & Wanninger, A. (2008). Segmental mode of neural patterning in Sipuncula. *Current Biology*, 18, 1129–1132.

Kristof, A., Wollesen, T., Maiorova, A. S. & Wanninger, A. (2011). Cellular and muscular growth patterns during sipunculan development. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution*, 316B, 227–240.

Mallatt, J. & Winchell, C. J. (2002). Testing the new animal phylogeny: first use of combined large-subunit and small-subunit rRNA gene sequences to classify the protostomes. *Molecular Biology and Evolution*, 19, 289–301.

Maxmen, A. B., King, B. F., Cutler, E. B. & Giribet, G. (2003). Evolutionary relationships within the protostome phylum Sipuncula: a molecular analysis of ribosomal genes and histone H3 sequence data. *Molecular Phylogenetics and Evolution*, 27, 489–503.

Mickevich, M. F. & Farris, J. S. (1981). The implications of congruence in *Menidia*. *Systematic Zoology*, 30, 351–370.

Mwinyi, A., Meyer, A., Bleidorn, C., Lieb, B., Bartolomaeus, T. & Podsiadlowski, L. (2009). Mitochondrial genome sequence and gene order of *Sipunculus nudus* give additional support for an inclusion of sipuncula into Annelida. *BMC Genomics*, 10, 27.

Nixon, K. C. (1999). The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics*, 15, 407–414.

Pineau, P., Henry, M., Suspene, R., Marchio, A., Dettai, A., Debruyne, W. *et al.* (2005). A universal primer set for PCR amplification of nuclear histone H4 genes from all animal species. *Molecular Biology and Evolution*, 22, 582–588.

Posada, D. (2005). Modeltest 3.7. Available via http://darwin.uvigo.es/.

Posada, D. & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, 53, 793–808.

Posada, D. & Crandall, K. A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics*, 14, 817–818.

Prendini, L., Weygoldt, P. & Wheeler, W. C. (2005). Systematics of the *Damon variegatus* group of african whip spiders (Chelicerata : Amblypygi): evidence from behaviour, morphology and DNA. *Organisms, Diversity & Evolution*, 5, 203–236.

Rambaut, A. E. (1996). Se-al sequence alignment editor. Available via http://evolve.zoo.ox.ac.uk/software.

Rambaut, A. & Drummond, A. J. (2009). Tracer v. 1.5. Available via http://beast.bio.ed.ac.uk/Tracer/.

Ronquist, F. & Huelsenbeck, J. P. (2003). Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, 1572–1574.

Saiz Salinas, J. I. (1993). *Sipuncula*. Madrid: Museuo Nacional de Ciencias Naurales CSIC.

Schulze, A., Cutler, E. B. & Giribet, G. (2005). Reconstructing the phylogeny of the Sipuncula. *Hydrobiologia*, 535/536, 277–296.

Schulze, A., Cutler, E. B. & Giribet, G. (2007). Phylogeny of sipunculan worms: a combined analysis of four gene regions and morphology. *Molecular Phylogenetics and Evolution*, 42, 171–192.

Schwendinger, P. J. & Giribet, G. (2005). The systematics of the south-east Asian genus *Fangensis rambla* (Opiliones : Cyphophthalmi: Stylocellidae). *Invertebrate Systematics*, 19, 297–323.

Sharma, P. P., Vahtera, V., Kawauchi, G. Y. & Giribet, G. (2011). Running $_{w}$ILD: the case for exploring mixed parameter sets in sensitivity analysis. *Cladistics*, 27, 538–549.

Sperling, E. A., Vinther, J., Moy, V. N., Wheeler, B. M., Semon, M., Briggs, D. E. G. *et al.* (2009). MicroRNAS resolve an apparent conflict between annelid systematics and their fossil record. *Proceedings of the Royal Society B: Biological Sciences*, 276, 4315–4322.

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

Stamatakis, A., Hoover, P. & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology*, 57, 758–771.

Staton, J. L. (2003). Phylogenetic analysis of the mitochondrial cytochrome c oxidase subunit I gene from 13 sipunculan genera: intra-and interphylum relationships. *Invertebrate Biology*, 122, 252–264.

Stephen, A. C. (1964). A revision of the classification of the phylum Sipuncula. *Annals and Magazine of Natural History*, 7, 457–462.

Stephen, A. C. & Edmonds, S. J. (1972). *The Phyla Sipuncula and Echiura*. London: Trustees British Mus. (Nat. Hist.).

Struck, T. H., Schult, N., Kusen, T., Hickman, E., Bleidorn, C., McHugh, D. *et al.* (2007). Annelid phylogeny and the status of Sipuncula and Echiura. *BMC Evolutionary Biology*, 7, 57.

Struck, T. H., Paul, C., Hill, N., Hartman, S., Hösel, C., Kube, M. *et al.* (2011). Phylogenomic analyses unravel annelid evolution. *Nature*, 471, 95–100.

Tzetlin, A. B. & Purschke, G. (2006). Fine structure of the pharyngeal apparatus of the pelagosphera larva in *Phascolosoma agassizii* (Sipuncula) and its phylogenetic significance. *Zoomorphology*, 125, 109–117.

Varón, A., Vinh, L. S. & Wheeler, W. C. (2010). POY version 4: phylogenetic analysis using dynamic homologies. *Cladistics*, 26, 72–85.

Wanninger, A., A., K. & Brinkmann, N. (2009). Sipunculans and segmentation. *Communicative & Integrative Biology*, 2, 56–59.

Wheeler, W. C. (1995). Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular-data. *Systematic Biology*, 44, 321–331.

Wheeler, W. (1996). Optimization alignment: the end of multiple sequence alignment in phylogenetics? *Cladistics*, 12, 1–9.

Wheeler, W. (1999). Fixed character states and the optimization of molecular sequence data. *Cladistics*, 15, 379–385.

Whiting, M. F., Carpenter, J. M., Wheeler, O. D. & Wheeler, W. C. (1997). The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Systematic Biology*, 46, 1–68.

Winchell, C. J., Sullivan, J., Cameron, C. B., Swalla, B. J. & Mallatt, J. (2002). Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. *Molecular Biology and Evolution*, 19, 762–776.

Xiandong, D., Chen, Z., Deng, Y. W., Wang, G. H. & Huang, R. L. (2008). Genetic diversity and population structure of the peanut worm (*Sipunculus nudus*) in southern China as inferred from mitochondrial 16S rRNA sequences. *Israeli Journal of Aquaculture: Bamidgeh*, 60, 237–242.

Xiong, B. & Kocher, T. D. (1991). Comparison of mitochondrial-DNA sequences of 7 morphospecies of black flies (Diptera, Simuliidae). *Genome*, 34, 306–311.

Zrzavý, J., Říha, P., Piálek, L. & Janouškovec, J. (2009). Phylogeny of Annelida (Lophotrochozoa): total-evidence analysis of morphology and six genes. *BMC Evolutionary Biology*, 9, 189.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Tree file of the evolutionary time tree of Sipuncula inferred from BEAST analysis.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Appendix 1 Collection data for specimens used in this study, in the following format: MCZ DNA voucher number, collecting location, collection date (collector)

Antillesoma antillarum (Grübe & Oersted, 1858): MCZ DNA100390—Phuket, Thailand, Jan. 31, 2001 (J. Hylleberg); MCZ DNA100759—Six Men's Bay, Barbados, June 27, 2002 (A. Schulze, J. I. Saiz Salinas, E. B. Cutler).

Apionsoma (*Apionsoma*) *misakianum* (Ikeda, 1904): MCZ DNA100737—Eilat, Israel, Sept. 30, 2000 (N. Ben-Eliahu); MCZ DNA102612—Îlot Maitre, New Caledonia, Nov. 16, 2007 (G. Y. Kawauchi and C. Tiago).

Apionsoma (*Edmondsius*) *pectinatum* (Keferstein, 1867): MCZ DNA100624—Six Mens Bay, Barbados, June 27, 2002 (A. Schulze, J. I. Saiz Salinas, E. B. Cutler).

Aspidosiphon (*Akrikos*) *albus* Murina, 1967: MCZ DNA101017—R/V Sunburst, cruise 521, Capron Shoals, FL, USA, Mar. 18, 2003 (A. Schulze, W. Lee, H. Reichardt).

Aspidosiphon (*Aspidosiphon*) *cristatus* (Sluiter, 1902): MCZ DNA100623—Bank Reef, Barbados, June 26, 2002 (A. Schulze, J. I. Saiz Salinas); MCZ DNA100986—Carrie Bow Cay, Belize, Apr. 17, 2003 (M. E. Rice, A. Schulze).

Aspidosiphon (*Aspidosiphon*) *elegans* (Chamisso & Eysenhardt, 1821): MCZ DNA100977—Carrie Bow Cay, Belize, Apr. 17, 2003 (A. Schulze, M. E. Rice); MCZ DNA101016—R/V Sunburst cruise 520, Capron Shoals, FL, USA, Mar. 11, 2003 (A. Schulze); MCZ DNA102633—Baie du Magenta, New Caledonia, Nov. 24, 2007 (G. Y. Kawauchi and C. Tiago).

Aspidosiphon (*Paraspidosiphon*) *fischeri* ten Broeke, 1925: MCZ DNA 100620—Martin's Bay, Barbados, June 21, 2002 (A. Schulze, J. I. Saiz Salinas, E. B. Cutler, G. Y. Kawauchi); MCZ DNA100981—Twin Cays, Belize, Apr. 24, 2003 (M. E. Rice, A. Schulze).

Aspidosiphon (*Aspidosiphon*) *gosnoldi* (E. B. Cutler, 1981): MCZ DNA101014—R/V Sunburst cruise 521, Capron Shoals, FL, USA, Mar. 18, 2003 (A. Schulze).

Aspidosiphon (*Aspidosiphon*) *gracilis schnehageni* (W. Fischer, 1913): MCZ DNA101087—Punta Moralia, Costa Rica, Aug. 27, 2003 (H.K. Dean, J. A. Vargas).

Aspidosiphon (*Paraspidosiphon*) *laevis* de Quatrefages, 1865: MCZ DNA100467—Hungary Bay, Bermuda, Aug. 9, 2001 (T. Nishikawa); MCZ DNA100992—Twin Cays, Belize, Apr. 20, 2003 (M. E. Rice, A. Schulze); MCZ DNA102616—Îlot Maitre, New Caledonia, Nov. 16, 2007 (G. Y. Kawauchi and C. Tiago).

Aspidosiphon (*Aspidosiphon*) *misakiensis* Ikeda, 1904: MCZ DNA100205—Cova Blava, Cabrera, Balearic Islands, Spain, May 31, 1997 (X. Turon).

Aspidosiphon (*Aspidosiphon*) *muelleri* Diesing, 1851: MCZ DNA100206—Banyuls-sur-Mer, France, July 19, 2000 (G. Giribet).

Aspidosiphon (*Paraspidosiphon*) *parvulus* Gerould, 1913: MCZ DNA100375—unspecified locality, purchased from Gulf Specimens Co.; MCZ DNA100982—Twin Cays, Belize, Apr. 20, 2003 (M. E. Rice, A. Schulze); MCZ DNA102601—Récif du Prony, New Caledonia, Nov. 15, 2007 (G. Y. Kawauchi and C. Tiago).

Aspidosiphon (*Aspidosiphon*) cf. *spiralis* Sluiter, 1902: MCZ DNA102608— Îlot Maitre, New Caledonia, Nov. 16, 2007 (G. Y. Kawauchi and C. Tiago); MCZ DNA102611— Îlot Maitre, New Caledonia, Nov. 16, 2007 (G. Y. Kawauchi and C. Tiago).

Aspidosiphon (*Paraspidosiphon*) *steenstrupii* Diesing, 1859: MCZ DNA100232—Pickles Reef, Key Largo, USA, Nov. 27, 1993 (S. Taylor); MCZ DNA100372—Kewalo Reef, Honolulu, USA, Jan. 25, 2001 (J. Brock); MCZ DNA100391—Phuket, Thailand, Jan. 31, 2001 (J. Hylleberg); MCZ DNA100630—Bank Reef, Barbados, June 26, 2002 (J. I. Saiz Salinas, A. Schulze); MCZ DNA102602—Récif du Prony, New Caledonia, Nov. 15, 2007 (G. Y. Kawauchi and C. Tiago).

Cloeosiphon aspergillus (de Quatrefages, 1865): MCZ DNA100393—Phuket, Thailand, Jan. 21, 2001 (J. Hylleberg); MCZ DNA100825—Perrier's Rock, South Africa, Oct. 18, 2002 (R. Biseswar); MCZ DNA102603—Récif du Prony, New Caledonia, Nov. 15, 2007 (G. Y. Kawauchi and C. Tiago); MCZ DNA102613—Îlot Maitre, New Caledonia, Nov. 16, 2007 (G. Y. Kawauchi and C. Tiago).

Golfingia elongata (Keferstein, 1862): MCZ DNA100465—SW of Trunk Island, Harrington Sound, Bermuda, Aug. 8, 2001 (T. Nishikawa); MCZ DNA100466—South coast of Stock's Harbor, St. Davis Island, Bermuda, Aug. 7, 2001 (T. Nishikawa); MCZ DNA101003—Twin Cays, Belize, Apr. 20, 2003 (M. E. Rice, A. Schulze); MCZ DNA101066—R/V Oceanus, Southern New England, USA, 40°27.299′ N, 69°54.601′ W, June 11, 2003 (A. Schulze); MCZ DNA101081—R/V Oceanus, Southern New England, USA, 40°27.299′ N, 69°54.601′ W, June 11, 2003 (A. Schulze).

Golfingia margaritacea (Sars, 1851): MCZ DNA100738—Kongsfjord Svalbard, Norway, June 23, 2002 (D. Hughes).

Golfingia vulgaris (de Blainville, 1827): MCZ DNA100207—Banyuls sur Mer, France, July 19, 2000 (G. Giribet).

Nephasoma cf. *abyssorum* (Koren and Danielssen, 1875): MCZ DNA101098— unspecified locality, Antarctic, Nov. 1999 (J. I. Saiz Salinas).

Nephasoma diaphanes (Gerould, 1913): MCZ DNA101072—R/V Oceanus, Southern New England 40°20.410′ N, 70°46.765′ W, June 11, 2003 (A. Schulze).

Nephasoma columbaris (Kawauchi and Rice 2009): MCZ DNA102649—Sebastian Pinnacles, Fort Pierce, USA, May 1, 2005 (G. Y. Kawauchi).

Nephasoma flagriferum (Selenka, 1885): MCZ DNA100439— Meteor Station Me48/1#345/7, Antarctica, Nov. 11, 1999 (J. I. Saiz Salinas); MCZ DNA100440—Meteor Station Me48/1#349, Antarctica, Nov. 11, 1999 (J. I. Saiz Salinas).

Nephasoma rimicola (Gibbs, 1973): MCZ DNA102626—Baie du Magenta, New Caledonia, Nov. 24, 2007 (G. Y. Kawauchi and C. Tiago).

Nephasoma pellucidum (Keferstein, 1865): MCZ DNA101009—R/V Sunburst, cruise 526, 4 miles off Fort Pierce, USA, Mar. 28, 2003 (A. Schulze); MCZ DNA102604—Récif du Prony, New Caledonia, Nov. 15, 2007 (G. Y. Kawauchi and C. Tiago).

Onchnesoma steenstrupii Koren & Danielssen, 1875: MCZ DNA101078— R/V Oceanus, Southern New England, USA, June 15, 2003 (A. Schulze); MCZ DNA101080—R/V Oceanus - Station 23, Southern New England, USA, 39°56.172′ N, 69°34.563′ W, June 15, 2003 (A. Schulze).

Phascolion (*Isomya*) *gerardi* Rice, 1993: MCZ DNA101002—Pinnacles between Sand bores, south of Carrie Bow Cay and Curlew Bank, Belize, Apr. 21, 2003 (M. E. Rice, A. Schulze).

Phascolion (*Lesenka*) *cryptum* Hendrix, 1975: MCZ DNA101007—Harbor Branch Oceanographic Institution, Indian River Lagoon, Ft. Pierce, USA, Mar. 9, 2003 (A. Schulze).

Phascolion (*Phascolion*) *psammophilus* Rice, 1993: MCZ DNA101006—R/V Sunburst cruise 523, Capron Shoals, USA, Mar. 18, 2003 (A. Schulze).

Phascolion (*Phascolion*) *strombus* (Montagu, 1804): MCZ DNA100739—Kristineberg Marine Biological Station, Fiskebäckskil, Sweden, Dec. 31, 1997 (A. Okusu); MCZ DNA101077—R/V Oceanus, Southern New England, USA, 39°47.230′ N, 70°46.295′ W, June 14, 2003 (A. Schulze).

Phascolopsis gouldii (Portalés, 1851): MCZ DNA100199—Woods Hole, USA, Sept. 30, 1997 (Marine Biological Laboratory).

Phascolosoma (*Phascolosoma*) *agassizii* Keferstein, 1866: MCZ DNA101096—Cape Arago (North Cove), OR, USA, Aug. 28, 2003 (M. E. Rice, S. Rumrill, C. Young); MCZ DNA102627—Baie du Magenta, New Caledonia, Nov. 24, 2007 (G. Y. Kawauchi and C. Tiago).

Phascolosoma (*Phascolosoma*) *albolineatum* (Baird, 1868): MCZ DNA100396—Phuket, Thailand, Jan. 31, 2001 (J. Hylleberg).

Phascolosoma (*Phascolosoma*) *granulatum* Leuckart, 1828: MCZ DNA100201—Blanes, Girona, Catalonia, Spain, Aug. 12, 1997 (G. Giribet, C. Palacín).

Phascolosoma (*Phascolosoma*) *nigrescens* (Keferstein, 1865): MCZ DNA100621—Six Mens Bay, Barbados, June 27, 2002 (A. Schulze, J. I. Saiz Salinas); MCZ DNA100736—Eilat, Israel, Sept. 30, 2002 (N. Ben-Eliahu); MCZ DNA100822—Perrier's Rock, South Africa, Oct. 18, 2002 (R. Biseswar); MCZ DNA101010—Bessie Cove, FL, USA, Mar. 20, 2003 (A. Schulze); MCZ DNA102605—Récif du Prony, New Caledonia, Nov. 15, 2007 (G. Y. Kawauchi and C. Tiago); MCZ DNA102619—Îlot Maitre, New Caledonia, Nov. 16, 2007 (G. Y. Kawauchi and C. Tiago).

Phascolosoma (*Phascolosoma*) *noduliferum* Stimpson, 1855: MCZ DNA100208—Nielsen Park Shore, Port Jackson, Sydney Harbor, Australia, Apr. 12, 2000 (G. Giribet, P. Hutchings).

Phascolosoma (*Phascolosoma*) *perlucens* Baird, 1868: MCZ DNA100228—García House, Puerto Peñasco, Sonora, Mexico, Oct. 12, 2000 (M. K. Nishiguchi); MCZ DNA100233—Missouri Key, FL, USA Oct. 8, 1993 (J. Wise); MCZ DNA100395—Phuket, Thailand, Jan. 31, 2001 (J. Hylleberg); MCZ DNA100748—Bank Reef, Barbados, June 26, 2002 (A. Schulze, J. I. Saiz Salinas, E. B. Cutler, G. Y. Kawauchi); MCZ DNA100819—Perrier's Rock, South Africa, Oct. 18, 2002 (R. Biseswar); MCZ DNA100829—Farfan, Panama, June 19, 2002 (T. Nishikawa); MCZ DNA102620—Îlot Maitre, New Caledonia, Nov. 16, 2007 (G. Y. Kawauchi and C. Tiago).

Phascolosoma (*Phascolosoma*) *scolops* (Selenka & de Man, 1883): MCZ DNA100373—Kewalo Reef, Honolulu, HI, USA, Jan. 25, 2001 (J. Brock); MCZ DNA100394—Phuket, Thailand, July 31, 2001 (J. Hylleberg); MCZ DNA100735—Eilat, Israel, Sept. 30, 2002 (N. Ben-Eliahu); MCZ DNA100813—Perrier's Rock, South Africa, Oct. 18, 2002 (R. Biseswar); MCZ DNA100814—Park Rynie Beach, South Africa, Aug., 22, 2002 (R. Biseswar); MCZ DNA102606—Récif du Prony, New Caledonia, Nov. 15, 2007 (G. Y. Kawauchi and C. Tiago).

Phascolosoma (*Phascolosoma*) *stephensoni* (Stephen, 1942): MCZ DNA100203—Cova Blava, Cabrera, Balearic Islands, Spain, May 31, 1997 (X. Turon); MCZ DNA100209—Nielsen Park Shore, Port Jackson, Sydney Harbor, NSW, Australia, Apr. 12, 2000 (G. Giribet, P. Hutchings); MCZ DNA100469—Baileys Bay, Hamilton Island, Bermuda, Aug. 7, 2001 (E. B. Cutler); MCZ DNA100485—Terceira, Azores, Portugal, Oct. 31, 2001 (P. Wirtz); MCZ DNA100817— Park Rynie Beach, South Africa, Aug., 22, 2002 (R. Biseswar); MCZ DNA100818— Park Rynie Beach, South Africa, Sept., 09, 2002 (R. Biseswar); MCZ DNA102607—Récif du Prony, New Caledonia, Nov. 15, 2007 (G. Y. Kawauchi and C. Tiago).

Phascolosoma (*Phascolosoma*) *turnerae* Rice, 1985: MCZ DNA100230—Southwest Reef, Bahamas, Jan. 31, 2000 (S. Brooke, T. Griffin).

Siphonosoma cumanense (Keferstein, 1867): MCZ DNA100235—unspecified locality, Puerto Rico June 3, 1993 (J. Staton, H. Reichardt); MCZ DNA100464—Baileys Bay, Hamilton, and south coast of St. Davus, Bermuda, Aug. 7, 2001 (T. Nishikawa); MCZ DNA100622—Bath, Barbados, June 24, 2002 (A. Schulze, J. I. Saiz Salinas); MCZ DNA100991—Twin Cays, Belize, Apr. 24, 2004 (M. E. Rice, A. Schulze); MCZ DNA102632—Teremba Bay New Caledonia, Nov., 26 2007 (G. Y. Kawauchi and C. Tiago).

Siphonosoma vastum (Selenka & von Bülow, 1883): MCZ DNA100625—Bath, Barbados, June 24, 2002 (A. Schulze, J. I. Saiz Salinas, E. B. Cutler).

Sipunculus (*Sipunculus*) *norvegicus* Danielssen, 1869: MCZ DNA101069—R/V Oceanus, Southern New England, USA, 39°47.230′ N, 70°46.295′ W, June 14, 2003 (A. Schulze).

Sipunculus (*Sipunculus*) *nudus* Linnaeus, 1766: MCZ DNA100234—Station 53F, No Name Cay, July 28, 1993 (J. Staton, H. Reichardt); MCZ DNA100245—near Arcachon (Fishermen's locality unspecified), Oct. 30, 2000; MCZ DNA100246—unspecified locality, Vietnam, Oct. 30, 2000; MCZ DNA100468—South coast of Stock's Harbor, St David's Island, Aug. 7, 2001 (E. B. Cutler); MCZ DNA100629— Isla Taboguilla, off Panama City, Panama, June 20, 2002 (T. Nishikawa); MCZ DNA100993—Twin Cays, Belize, Apr. 24, 2003 (M. E. Rice, A. Schulze); MCZ DNA101789—Praia do Barreiro, São Sebastião, Brazil, June 17, 2003 (G. Y. Kawauchi); MCZ DNA101882—Ponta do Araçá, São Sebastião, Brazil, June 17, 2003 (G. Y. Kawauchi); MCZ DNA102316—Fort Pierce, USA, Mar. 1, 2006 (G. Y. Kawauchi and A. Schulze).

Sipunculus (*Sipunculus*) *phalloides* (Pallas, 1774): MCZ DNA101917—Ponta do Araçá, São Sebastião, Brazil, 23°49′02″ S, 45 °24′19 W (G. Y. Kawauchi) [this specimen corresponds to the sequence erroneously published as MCZ DNA101337 by Schulze *et al.* (2007)].

Sipunculus (*Sipunculus*) *polymyotus* Fisher, 1947: MCZ DNA101121—Pelican Beach, Belize, Oct. 24, 2002 (D. Felder, R. Robles).

Themiste (*Themiste*) *alutacea* (Grübe & Oersted, 1858): MCZ DNA101906—Cubagua Island, Venezuela, unspecified date, (I. Hernandez-Avila); MCZ DNA101909—Cubagua Island, Venezuela, unspecified date, (I. Hernandez-Avila).

Themiste (*Lagenopsis*) *minor* (Ikeda, 1904): MCZ DNA100826— Perrier's Rock, South Africa, Oct. 18, 2002 (R. Biseswar); MCZ DNA101083—unspecified locality, South Africa, Sept. 28, 2002 (G. Rouse).

Themiste (*Themiste*) *dyscrita* (Fisher, 1952): MCZ DNA101095—Cape Arago (North Cove), OR, USA, Aug. 29, 2002 (M. E. Rice, S. Rumrill, C. Young).

Themiste (*Themiste*) *hennahi* Gray, 1828: MCZ DNA100627—Bahia de Concepción, Lirquen Playa sector La Cata, Chile, Apr. 26, 2001 (E. Tarifeño).

Themiste (*Themiste*) *pyroides* (Chamberlin, 1920): MCZ DNA101084—Whiffen Spit, Vancouver Island, B.C., Canada, Sept. 9, 2003 (A. Schulze, M. E. Rice).

Thysanocardia catherinae (Grübe, 1868): MCZ DNA101068—R/V Oceanus, Southern New England, USA, 39°47.230′ N, 70°48.295′ W, June 14, 2003 (A. Schulze); MCZ DNA101075—R/V Oceanus, Southern New England, USA, S20, June 2003 (A. Schulze).

Thysanocardia nigra (Ikeda, 1904): MCZ DNA100606—Lopez Island, WA, USA, May 17, 2002 (D. McHugh); MCZ DNA102535—Turn Island, WA, USA, Aug. 28, 2007 (F. Brown and M. Rice).

Xenosiphon branchiatus (Fischer, 1895): MCZ DNA101086—Tamarindo Beach, Costa Rica, Feb. 8, 2003 (R. Quiros).

Appendix 2 GenBank and EMBL accession numbers for sequence fragments

Species	MCZ DNA Acc. no.	18S	28S	H3	H4	COI	16S
Sipunculidae							
Sipunculus (*S.*) *norvegicus*	DNA101069	DQ300004	JN865046	DQ300090	HE605172	DQ300159	–
Sipunculus (*S.*) *nudus*	DNA100234	DQ300005	JN865047	JN865127	–	DQ300160*	JN864959
Sipunculus (*S.*) *nudus*	DNA100245	AF519239	AF519269*	JN865148	HE605129	JN865105	JN864960
Sipunculus (*S.*) *nudus*	DNA100246	AF519240	AF519270*	AF519295	HE605130	DQ300161	JN864961
Sipunculus (*S.*) *nudus*	DNA100468	DQ300006	DQ300048*	DQ300091	HE605143	DQ300162*	JN864962
Sipunculu s (*S.*) *nudus*	DNA100629	DQ300007	JN865048	DQ300092	–	DQ300163*	JN864963
Sipunculus (*S.*) *nudus*	DNA100993	DQ300008	DQ300049*	DQ300093	HE605166	DQ300164*	JN864964
Sipunculus (*S.*) *nudus*	DNA101789	JN865019	JN865049	JN865132	HE605181	JN865106	–
Sipunculus (*S.*) *nudus*	DNA102316	JN865020	JN865050	JN865167	HE605186	JN865107	JN865004
Sipunculus (*S.*) *nudus*	DNA101882	JN869399	JN865051	JN865128	HE605182	JN865108	JN865000
Sipunculus (*S.*) *nudus*	DNA101917	DQ300009	JN865052	DQ300094	HE605185	DQ300165*	JN865003
Sipunculus (*S.*) *nudus*	DNA101121	DQ300010	JN865053	DQ300095	–	DQ300166	–
Xenosiphon branchiatus	DNA101086	DQ300016	DQ300050*	DQ300101	HE605178	DQ300172	JN864996
Siphonosomatidae							
Siphonosoma cumanense	DNA100235	AF519241	AF519271	AF519296	–	–	–
Siphonosoma cumanense	DNA100464	DQ300001	JN865054	DQ300088	HE605139	DQ300155	JN864971
Siphonosoma cumanense	DNA100622	AY326201	AY445139*	AY326296	HE605147	DQ300156*	JN864977
Siphonosoma cumanense	DNA100991	DQ300002	DQ300047*	DQ300089	–	DQ300157	–
Siphonosoma cumanense	DNA102632	JN865021	JN865055	JN865159	HE605200	–	JN865018
Siphonosoma vastum	DNA100625	DQ300003	AY445137	AY326297	–	DQ300158*	–
Golfingiidae							
Golfingia elongata	DNA100465	DQ299969	DQ300031*	DQ300065	HE605140	DQ300121	JN864972
Golfingia elongata	DNA100466	AF519242	JN865056	DQ300066	HE605141	DQ300122	JN864973
Golfingia elongata	DNA101003	DQ299970	JN865057	JN865156	–	DQ300123	JN864985
Golfingia elongata	DNA101066	DQ299971	JN865058	DQ300067	HE605170	DQ300124	–
Golfingia elongata	DNA101081	DQ299972	JN865059	DQ300068	–	DQ300125	–
Golfingia margaritacea	DNA100738	DQ299973	DQ300032*	DQ300069	HE605151	DQ300126	JN864981
Golfingia vulgaris	DNA100207	AF519244	AF519273*	–	HE605125	DQ300127	JN864954
Nephasoma cf. *abyssorum*	DNA101098	JN865022	–	JN865153	HE605180	JN865109	JN864999
Nephasoma diaphanes	DNA101072	DQ299975	JN865060	DQ300071	HE605173	DQ300128	JN864992
Nephasoma columbaris	DNA102649	JN869397	JN865061	JN865151	HE605201	JN865110	–
Nephasoma flagriferum	DNA100439	AF519243	–	AF519299	–	–	–

Appendix 2 Continued

Species	MCZ DNA Acc. no.	18S	28S	H3	H4	COI	16S
Nephasoma flagriferum	DNA100440	DQ299976	DQ300033*	DQ300072	HE605138	DQ300129*	–
Nephasoma pellucidum	DNA101009	DQ299978	–	DQ300131	–	DQ300130*	JN864987
Nephasoma pellucidum	DNA102604	JN865023	JN865062	JN865157	HE605189	JN865111	JN865008
Nephasoma rimicola	DNA102626	JN865024	JN865063	JN865134	HE605199	JN865114	–
Phascolopsis gouldii	DNA100199	AF123306	AF519272*	AF519297	HE605122	DQ300134*	JN864952
Thysanocardia catherinae	DNA101068	DQ300015	JN865064	DQ300099	HE605171	–	JN864991
Thysanocardia catherinae	DNA101075	JN865025	JN865065	–	–	–	JN864993
Thysanocardia nigra	DNA100606	AF519247	AF519274	–	–	–	–
Thysanocardia nigra	DNA102535	JN865026	JN865066	–	HE605187	–	–
Themiste (*T.*) *dyscrita*	DNA101095	DQ300011	JN865067	–	HE605179	DQ300167	JN864998
Themiste (*T.*) *hennahi*	DNA100627	DQ300012	JN865068	DQ300096	–	DQ300168	JN864979
Themiste (L.) *minor*	DNA101083	DQ300013	–	DQ300097	HE605176	DQ300170	JN864994
Themiste (L.) *minor*	DNA100826	JN865027	–	–	HE605160	JN865112	JN864982
Themiste (*T.*) *pyroides*	DNA101084	DQ300014	JN865069	DQ300098	HE605177	DQ300171	JN864995
Themiste alutacea	DNA101906	JN865028	–	JN865150	HE605183	–	JN865001
Themiste alutacea	DNA101909	–	JN865070	JN865149	HE605184	–	JN865002
Onchnesoma steenstrupii	DNA101080	DQ299979	DQ300034*	DQ300074	HE605175	JN865113	–
Onchnesoma steenstrupii	DNA101078	JN865029	JN865071	JN865146	HE605174	–	–
Phascolion (*L.*) *cryptum*	DNA101007	DQ299980	DQ300035*	DQ300075	–	DQ300132	–
Phascolion (I.) *gerardi*	DNA101002	DQ299981	–	DQ300076	HE605167	JN865123	JN864984
Phascolion (*P.*) *psammophilus*	DNA101006	DQ299982	DQ300036*	JN865163	HE605168	DQ300133*	JN864986
Phascolion (*P.*) *strombus*	DNA100739	DQ299983	JN865072	–	HE605152	–	–
Phascolion (*P.*) *strombus*	DNA101077	DQ299984	JN865073	DQ300077	–	–	–
Antillesomatidae							
Antillesoma antillarum	DNA100390	AF519259	AF519286*	AF519311	HE605133	JN865120	JN864968
Antillesoma antillarum	DNA100759	DQ299950	JN865074	JN865136	HE605154	GU230172	GU230180
Phascolosomatidae							
Apionsoma (*A.*) *misakianum*	DNA100737	DQ299952	DQ300017*	DQ300052	HE605150	DQ300103*	GU230178
Apionsoma (*A.*) *misakianum*	DNA102612	JN865030	JN865075	JN865155	HE605193	JN865118	JN865014
Apionsoma (*E.*) *pectinatum*	DNA100624	AY326293	AY445142*	AY326300	HE605149	DQ300104	JN864979
Phascolosoma (*P.*) *agassizii*	DNA101096	DQ299985	DQ300037*	DQ300078	–	DQ300135	–
Phascolosoma (*P.*) *agassizii*	DNA102627	JN865031	JN865076	JN865130	–	–	JN865017
Phascolosoma (*P.*) *albolineatum*	DNA100396	AF519251	AF519278	JN865166	HE605137	GU230175	GU230186
Phascolosoma (*P.*) *granulatum*	DNA100201	AF519252*	AF519279*	AF519304	HE605123	DQ300138	GU230181
Phascolosoma (*P.*) *nigrescens*	DNA100621	AY326292	AY445140	AY326299	–	DQ300139	–
Phascolosoma (*P.*) *nigrescens*	DNA100736	DQ299987	DQ300038*	DQ300080	–	DQ300140	–
Phascolosoma (*P.*) *nigrescens*	DNA100822	DQ299988	DQ300039*	D300081	HE605158	DQ300141*	GU230182
Phascolosoma (*P.*) *nigrescens*	DNA101010	DQ299989	DQ300040*	JN865137	HE605169	DQ300142*	–
Phascolosoma (*P.*) *nigrescens*	DNA102619	JN865032	JN865077	JN865131	HE605196	JN865122	JN865016
Phascolosoma (*P.*) *nigrescens*	DNA102605	JN865033	JN865078	JN865165	–	JN865121	JN865009
Phascolosoma (*P.*) *noduliferum*	DNA100208	AF519253	AF519280*	AF519305	–	DQ300144	JN864955
Phascolosoma (*P.*) *perlucens*	DNA100228	AF519254	AF519281*	AF519306	HE605127	DQ300145*	JN864957
Phascolosoma (*P.*) *perlucens*	DNA100233	DQ299991	JN865079	JN865160	HE605128	JN865124	–
Phascolosoma (*P.*) *perlucens*	DNA100395	DQ299992	JN865080	DQ300082	HE605136	GU190249	GU190305
Phascolosoma (*P.*) *perlucens*	DNA100748	DQ299993	DQ300042*	JN865141	HE605153	GU190253	GU190309
Phascolosoma (*P.*) *perlucens*	DNA100819	DQ299994	JN865081	JN865129	HE605157	GU190268	GU190322
Phascolosoma (*P.*) *perlucens*	DNA100829	DQ299995	DQ300043*	DQ300083	HE605161	DQ300149*	–
Phascolosoma (*P.*) *perlucens*	DNA102620	JN865034	JN865082	JN865135	HE605197	GU190296	GU190350
Phascolosoma (*P.*) *scolops*	DNA100373	AF519255	JN865083	AF519282	HE605132	AF519309	JN864966
Phascolosoma (*P.*) *scolops*	DNA100394	DQ299996	JN865084	–	HE605135	DQ300150	–
Phascolosoma (*P.*) *scolops*	DNA100735	DQ299997	JN865085	DQ300084	–	DQ300151**	–
Phascolosoma (*P.*) *scolops*	DNA100813	DQ299998	DQ300044*	DQ300085	–	DQ300152*	–
Phascolosoma (*P.*) *scolops*	DNA100814	JN865035	JN865086	JN865162	–	GU230173	GU230184
Phascolosoma (*P.*) *scolops*	DNA102606	JN865036	JN865087	JN865144	HE605190	–	JN865010
Phascolosoma (*P.*) *stephensoni*	DNA100203	DQ299999	DQ300045*	DQ300086	HE605124	–	–
Phascolosoma (*P.*) *stephensoni*	DNA100209	AF519257	AF519284*	AF519307	HE605126	–	JN864956
Phascolosoma (*P.*) *stephensoni*	DNA100469	AF519256	AF519283*	AF519310	HE605144	DQ300153	–

Appendix 2 Continued

Species	MCZ DNA Acc. no.	18S	28S	H3	H4	COI	16S
Phascolosoma (*P.*) *stephensoni*	DNA100485	AF519258	AF519285*	AF519308	HE605145	–	JN864975
Phascolosoma (*P.*) *stephensoni*	DNA100818	JN865037	JN865088	JN865133	HE605156	GU230174	GU230185
Phascolosoma (*P.*) *stephensoni*	DNA100817	–	JN865089	JN865138	HE605155	JN865119	–
Phascolosoma (*P.*) *stephensoni*	DNA102607	JN865038	JN865090	JN865125	HE605191	–	JN865011
Phascolosoma (*P.*) *turnerae*	DNA100230	DQ300000	DQ300046*	DQ300087	–	DQ300154	GU230183
Aspidosiphonidae							
Aspidosiphon (*A.*) *albus*	DNA101017	DQ299954	JN865091	DQ300053	–	DQ300105	JN864990
Aspidosiphon (*A.*) *cristatus*	DNA100623	AY326295	AY445141*	–	HE605148	–	JN864950
Aspidosiphon (*A.*) *cristatus*	DNA100986	DQ299974	JN865092	DQ300070	HE605164	–	JN864951
Aspidosiphon (*A.*) *elegans*	DNA100977	DQ299956	DQ300019*	DQ300055	HE605162	–	–
Aspidosiphon (*A.*) *elegans*	DNA102623	JN865039	JN865093	JN865147	HE605198	–	–
Aspidosiphon (*A.*) *elegans*	DNA101016	DQ299957	DQ300020*	DQ300056	–	DQ300106*	JN864989
Aspidosiphon (*P.*) *fischeri*	DNA100620	AY326294	JN865094	AY326301	HE605146	DQ300107	JN864976
Aspidosiphon (*P.*) *fischeri*	DNA100981	DQ299958	DQ300021*	JN865139	HE605163	DQ300108	–
Aspidosiphon (*A.*) *gosnoldi*	DNA101014	DQ299959	DQ300022*	DQ300057	–	DQ300109*	JN864988
Aspidosiphon (*A.*) *gracilis schnehageni*	DNA101087	DQ299960	DQ300023*	DQ300058	–	DQ300110*	JN864997
Aspidosiphon (*P.*) *laevis*	DNA100467	AF519261	DQ300024*	DQ300059	HE605142	DQ300111*	JN864974
Aspidosiphon (*P.*) *laevis*	DNA100992	DQ299961	JN865095	JN865140	HE605165	JN865115	–
Aspidosiphon (*P.*) *laevis*	DNA102616	JN865040	JN865096	JN865164	HE605195	GU230171	GU230179
Aspidosiphon (*A.*) *misakiensis*	DNA100205	AF119090	AF519288	–	–	–	–
Aspidosiphon (*A.*) *muelleri*	DNA100206	DQ299962	DQ300025*	DQ300060	–	DQ300113*	JN864953
Aspidosiphon (*P.*) *parvulus*	DNA100375	DQ299963	JN865097	DQ300062	–	DQ300114	JN864967
Aspidosiphon (*P.*) *parvulus*	DNA100982	DQ299964	DQ300027*	DQ300063	–	DQ300115	JN864983
Aspidosiphon (*P.*) *parvulus*	DNA102601	JN865041	JN865098	JN865145	HE605188	–	JN865005
Aspidosiphon (*A.*) cf. *spiralis*	DNA102608	JN865042	JN865099	JN865152	–	JN865116	JN865012
Aspidosiphon (*A.*) cf. *spiralis*	DNA102611	JN865043	JN865100	JN865158	HE605192	JN865117	JN865013
Aspidosiphon (*P.*) *steenstrupii*	DNA100232	AF519262	AF519291	AF519315	–	DQ300116	JN864958
Aspidosiphon (*P.*) *steenstrupii*	DNA100372	DQ299965	DQ300028*	JN865161	HE605131	DQ300117	JN864965
Aspidosiphon (*P.*) *steenstrupii*	DNA100391	DQ299966	DQ300029*	JN865126	HE605134	DQ300118	JN864969
Aspidosiphon (*P.*) *steenstrupii*	DNA100630	DQ299967	JN865101	DQ300064	–	DQ300119	JN864980
Aspidosiphon (*P.*) *steenstrupii*	DNA102602	JN869398	JN865102	JN865143	–	–	JN865006
Cloeosiphon aspergillus	DNA100393	AF519263	AF519292*	AF519316	–	–	JN864970
Cloeosiphon aspergillus	DNA100825	DQ299968	DQ300030*	–	HE605159	DQ300120	–
Cloeosiphon aspergillus	DNA102603	JN865044	JN865103	JN865154	–	–	JN865007
Cloeosiphon aspergillus	DNA102613	JN865045	JN865104	JN865142	HE605194	–	JN865015
Nemertea							
Lineus bilineatus		DQ279932.1	DQ279947.1	DQ279996.1	–	DQ280014.2	DQ280022.1
Mollusca							
Nucula sulcata		DQ279937.1	DQ279960.1	DQ280001.1	–	DQ280017.1	DQ280029.1
Lepidopleurus cajetanus		AF120502.1	FJ445776			AF120626.1	AY377585.1
Neotrigonia margaritacea		AF411690.1	DQ279963.1	AY070155.1	–	U56850.1	DQ093489.1
Annelida							
Chaetopterus variopedatus		U67324.1	AY145399.1	–	AF007904.1	AM503096.1	–
Urechis caupo		AF119076.1	AF519268.1	X58895.1	–	AY619711.1	–
Entoprocta							
Loxosomella murmanica		AY218100.1	DQ279950	AY218150.1	–	AY218083.1	–
Brachiopoda							
Neocrania anomala		DQ279934	DQ279949	DQ279997.1	–	–	DQ280024.1

Asterisks indicate sequences updated in GenBank.

Appendix 3 Length of gene partition alignments prior and subsequent to treatment with GBLOCKS v. 0.91b

Data partition	No. of positions after treatment with MUSCLE	No. of positions after treatment with GBlocks
16S rRNA	598	403
18S rRNA	1769	1769
28S rRNA	2267	1629
COI	820	657
Histone H3	327	327
Histone H4	159	159

