Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 42 (2007) 171–192
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Phylogeny of sipunculan worms: A combined analysis of four gene regions and morphology

Anja Schulze [a,*,1], Edward B. Cutler [a], Gonzalo Giribet [a,b]

[a] *Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA*
[b] *Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA*

Received 23 November 2005; revised 11 June 2006; accepted 16 June 2006
Available online 5 July 2006

Abstract

The intra-phyletic relationships of sipunculan worms were analyzed based on DNA sequence data from four gene regions and 58 morphological characters. Initially we analyzed the data under direct optimization using parsimony as optimality criterion. An implied alignment resulting from the direct optimization analysis was subsequently utilized to perform a Bayesian analysis with mixed models for the different data partitions. For this we applied a doublet model for the stem regions of the 18S rRNA. Both analyses support monophyly of Sipuncula and most of the same clades within the phylum. The analyses differ with respect to the relationships among the major groups but whereas the deep nodes in the direct optimization analysis generally show low jackknife support, they are supported by 100% posterior probability in the Bayesian analysis. Direct optimization has been useful for handling sequences of unequal length and generating conservative phylogenetic hypotheses whereas the Bayesian analysis under mixed models provided high resolution in the basal nodes of the tree.

Keywords: Sipuncula; Direct optimization; Bayesian; Doublet model

1. Introduction

Sipuncula (peanut worms or star worms) contain roughly 150 exclusively marine species and are currently recognized as a phylum. Sipunculan systematics is rooted in a long and convoluted history, as outlined in Cutler (1994); Maxmen et al. (2003); Staton (2003) and Schulze et al. (2005). Today, annelid affinities are suggested by both mitochondrial gene arrangement data and DNA sequence analysis.

Using mitochondrial gene order data, both Boore and Staton (2002) and Bleidorn et al. (2005) place the sipunculan *Phascolopsis gouldii* in a clade with annelids. The latter find high support for a close relationship with the orbiniid polychaete *Orbinia latreillii*. However, taxon sampling for mitochondrial genomes of lophotrochozoans is still limited. Apart from *P. gouldii*, only annelid, mollusc and brachiopod mitochondrial genomes are considered in the analyses. In addition, only about 50% of the mitochondrial genome of *P. gouldii* genome has been sequenced. Jennings and Halanych (2005) conclude that the mitochondrial gene order is very conserved and of limited use for analyzing relationships within annelids. On the other hand, the correspondence of the partial mitochondrial gene sequence of *P. gouldii* with the highly conserved annelid sequence is noteworthy.

Several recent publications focusing on systematics of annelids based on DNA sequence data included sipunculans and unanimously placed them in the annelid ingroup, although the number of outgroups was sometimes limited (Brown et al., 1999) and the branch support was generally low for deep nodes in the tree (Bleidorn et al., 2003a,b; Hall et al., 2004). Better resolution and support for deep relationships was achieved by Telford et al. (2005) who presented a phylogeny of the Bilateria, based on 72 complete 18S ribosomal RNA sequences. The sequences were

* Corresponding author. Present address: Smithsonian Marine Station, 701 Seaway Drive, Fort Pierce, FL 34949, USA.
E-mail address: anjaschulze@bigfoot.com (A. Schulze).
[1] Current address: Department of Marine Biology, Texas A&M University at Galveston, 5007 Avenue U, Galveston, TX 77551.

1055-7903/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2006.06.012

analyzed under mixed models for the stem and loop regions of the ribosomal molecule, applying a doublet model that accounts for correlation of substitutions in the complementary stem regions. The included sipunculan sequence grouped with the seven annelids, as the sister taxon to a sabellid. Peterson and Eernisse (2001) and Staton (2003) also concluded that Sipuncula were more closely related to annelids than to molluscs contradicting the hypothesis of Scheltema (1996). On the other hand, total evidence analyses of ribosomal genes and morphology have suggested that sipunculans are the sister group to a clade including annelids and molluscs (Giribet et al., 2000). Our recent analyses of approximately 2400 base pairs from three gene regions and morphology strongly support the monophyly of Sipuncula but provide no resolution at a higher level (Maxmen et al., 2003; Schulze et al., 2005).

The fossil record for sipunculans is generally sparse but Huang et al. (2004) reported three exquisitely preserved fossil sipunculan species from the Lower Cambrian Maotianshan Shale in southwest China. This finding places the origin of the Sipuncula to more than 520 Myr ago despite the fact that the group presents little morphological variation and low species diversity. The affinity of the fossils to modern sipunculans was supported by the presence of a retractable introvert, a mouth surrounded by tentacles, a caudal appendage (sometimes present in modern Golfingiidae), a U-shaped gut with an anus on the anterior trunk and hooks, wrinkles and papillae on the body, but interestingly, the gut lacks the typical helical morphology of the modern members of the group.

As in our previous analyses (Maxmen et al., 2003; Schulze et al., 2005), the focus of the present paper is the internal phylogeny of the Sipuncula. Since submission of our latest analyses (Schulze et al., 2005), we have increased the number of sipunculan species from 29 to 52, representing more than one third of all known sipunculan species and representatives of all but one of the 17 currently recognized genera. The only missing one is *Siphonomecus*, a monotypic genus only known from the southeastern United States. *Siphonomecus multicinctus* is rare to emerge from its deep sediment burrows and we have not yet been able to obtain material appropriate for DNA analysis.

The phylogenetic hypotheses that we are generating will enable us to test the current taxonomy of the Sipuncula as proposed by Cutler and Gibbs (1985) and Gibbs and Cutler (1987) who classified the group into seventeen genera, six families, four orders and two classes. The two classes, Sipunculidea and Phascolosomatidea, are morphologically distinguished by their tentacle arrangement: in Sipunculidea the tentacles form a circle around a central mouth, whereas in Phascolosomatidea they are arranged in a horseshoe shape around the nuchal organ. For the diagnoses of genera, families and orders, see Gibbs and Cutler (1987) and references to taxonomic revisions therein. However, the monophyly of both classes has been questioned in recent phylogenetic analyses (Maxmen et al., 2003; Staton, 2003; Schulze et al., 2005), and the new hypothesis regards certain characters of Sipunculidea as plesiomorphies for the phylum.

The phylogeny proposed in this study will also serve as a framework to detect and interpret evolutionary trends within the phylum. In addition, by including multiple representatives of widespread species from different geographic locations whenever possible, we are performing initial tests for the phylogenetic cohesiveness of species and the putative existence of hidden species diversity. This will allow us in the future to choose appropriate species for population genetic and phylogeographic studies.

2. Materials and methods

2.1. Collections and taxon sampling

Upon collection in the field, all sipunculan specimens were frozen or fixed in 70–100% ethanol or in one case in isopropanol (*Xenosiphon branchiatus*) prior to DNA extraction. Outgroup taxa, chosen among molluscs, annelids, entoprocts and nemerteans, were treated in the same fashion. Collection data are listed in Appendix A. In total, we present data for 99 sipunculan individuals belonging to 52 recognized morphospecies.

2.2. Sequence generation

Total DNA was extracted from a piece of tissue, preferably from the retractor muscles, using the DNeasy tissue kit (Qiagen), following the instructions of the manufacturer. The desired gene regions were amplified from the DNA templates using polymerase chain reaction (PCR) (see Table 1 for primers). The complete 18S rRNA was amplified in three fragments, using the primers 1F-3R, 1F-4R or 1F-5R for the first, 3F-18Sbi for the second and 18Sa2.0-9R for the last fragment. Occasionally, the second fragment was split up into 3F-5R and 4F-18Sbi, 4F-7R or 4F-8R. The total length of the 18S rRNA gene is approximately 1800 bp, resulting in an implied alignment (see below) of 2053 bp. Three other gene regions (lengths excluding amplification primers) were amplified as single fragments: the D3 fragment of the nuclear 28S ribosomal RNA (ca. 310 bp; implied alignment 409 bp), the nuclear coding gene for histone H3 (327 bp) and the mitochondrial coding gene for cytochrome *c* oxidase subunit I (650 bp). PCRs were performed in 25 μl volume according to standard protocols with the annealing temperature varying between 40 °C for coding genes and up to 50 °C for ribosomal genes. PCR products were visualized in 1–1.5% agarose gels and cleaned using the GENECLEAN II kit (Bio 101) or the QIAquick PCR purification kit (Qiagen). Sequence reactions were performed in 10 μl volume, using 1 μl (if cleaned with GENECLEAN) or 5 μl (if cleaned with QIAquick) of template, 1 μM of primer, 2 μl of ABI BigDye™ Terminator v3.0 (Applied Biosystems) and 2 μl of halfTERM Dye Terminator reagent (Genpak). Sequence reactions were performed with the same thermal cycler as for PCR following

Table 1
Primers used for PCR amplification and cycle sequencing

Primer	Sequence	Reference
COI		
LCO1490	5′ GGT CAA CAA ATC ATA AAG ATA TTG G 3′	Folmer et al. (1994)
HCO2198	5′ TAA ACT TCA GGG TGA CCA AAA AAT CA 3′	Folmer et al. (1994)
COI-7	5′ ACN AAY CAY AAR GAY ATY GGN AC 3′	Saito et al. (2000)
COI-D	5′ TCN GGR TGN CCR AAN ARY CAR AA 3′	Saito et al. (2000)
H3		
H3aF	5′ ATG GCT CGT ACC AAG CAG AC(ACG) GC 3′	Colgan et al. (1998)
H3aR	5′ ATA TCC TT(AG) GGC AT(AG) AT(AG) GTG AC 3′	Colgan et al. (1998)
28S rRNA		
28Sa	5′ GAC CCG TCT TGA AAC ACG GA 3′	Whiting et al. (1997)
28Sb	5′ TCG GAA GGA ACC AGC TAC TA 3′	Whiting et al. (1997)
18S rRNA		
1F	5′ TAC CTG GTT GAT CCT GCC AGT AG 3′	Giribet et al. (1996)
3R	5′ AGG CTC CCT CTC CGG AAT CGA AC 3′	Giribet et al. (1996)
3F	5′ GTT CGA TTC CGG AGA GGG A 3′	Giribet et al. (1996)
4R	5′ GAA TTA CCG CGG CTG CTG G 3′	Giribet et al. (1996)
4F	5′ CCA GCA GCC GCG CTA ATT C 3′	Giribet et al. (1996)
5R	5′ CTT GGC AAA TGC TTT CGC 3′	Giribet et al. (1996)
7R	5′ GCA TCA CAG ACC TGT TAT TGC 3′	Giribet et al. (1996)
8R	5′ ACG GGC GGT GTG TAC 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. (1999)
18Sbi	5′ GAG TCT CGT TCG TTA TCG GA 3′	Giribet et al. (1999)

ABI standard protocols. After cleanup of the sequence reactions using Edge Biosystems gel filtration cartridges, the sequences were analyzed with an ABI PRISM® 3100 Genetic Analyzer.

2.3. Analyses

Electrochromatograms were visualized in Sequencher™ 4.0. Forward and reverse fragments were assembled and, in the case of 18S rRNA, several fragments joined into a single sequence. From each sequence, external primers were cropped and discarded. Sequences were subsequently edited in MacGDE (Smith et al., 1994; Linton, 2005) and/or BioEdit Sequence Alignment Editor (Hall, 1994). All new sequences were deposited in GenBank under accession numbers DQ299950 through DQ300172 (Table 2).

Phylogeny reconstruction followed two different approaches: (1) A direct optimization approach (Wheeler, 1996), as implemented in the computer program POY (Wheeler et al., 2004, 2006), using parsimony as the optimality criterion. (2) A Bayesian approach with mixed models estimated for each independent partition, as implemented in the program MrBayes 3.1.1 (Ronquist and Huelsenbeck, 2003). For both types of analyses, all partitions were analyzed separately and in combination. The parsimony jackknife tree of the morphological data alone was calculated in PAUP* (Swofford, 2003). We performed 1000 jackknife replicates, using the heuristic search option. For each heuristic search, 100 replicates of random taxon addition were performed with tree bisection and reconnection as the branch-swapping algorithm. Branches with less than 50% jackknife support were discarded. Both morphology-based trees were rooted with *Sipunculus nudus*, because we could not score other outgroups for the characters specific to sipunculans.

The direct optimization approach allows the analysis of sequences of unequal length without prior alignment. The alignment and tree generation are performed simultaneously in a dynamic programming environment by taking into account the same parameters (e.g., for transversion-to-transition ratios and gap penalties) throughout the entire analysis; this is known as "one-step phylogenetics". In addition we performed a sensitivity analysis in which we tested multiple parameter sets (Giribet, 2003; Wheeler, 1995). For each data partition, 12 parameter sets were analyzed with the transversion-to-transition ratios of 1, 2, 4 and 8 and indel-to-transversion ratios of 1, 2, and 4.

Tree searches were conducted in parallel (using PVM—Parallel Virtual Machine) on a cluster of 30 dual-processor nodes (between 1 and 2.4 GHz) assembled at Harvard University (darwin.oeb.harvard.edu). Commands for load balancing of spawned jobs were in effect to optimize parallelization procedures (-parallel, -dpm, -jobspernode 2, -multirandom). Initially trees were built through a random addition sequence procedure (20 replicates) followed by a combination of branch-swapping steps (SPR "subtree pruning and regrafting" and TBR "tree bisection and reconnection"), and continuing with tree fusing (Goloboff, 1999, 2002) in order to further improve tree length. While SPR and TBR allow branch rearrangement within a given tree, tree fusing allows exchanging of branches of identical composition among different trees, as in other simulated evolutionary algorithms (Moilanen, 1999, 2001). Discrepancies between heuristic and actual tree length calculations

Table 2
Taxon sampling and GenBank accession numbers for each sequenced locus

Species	*MCZ Catalogue #*	*18S rRNA*	*28S rRNA*	*Histone H3*	COI
Sipunculidae					
Phascolopsis gouldii	DNA100199	AF123306	AF519272	AF519297	DQ300134
Siphonosoma cumanense	DNA100235	AF519241	AF519271	AF519296	
Siphonosoma cumanense	DNA100464	DQ300001		DQ300088	DQ300155
Siphonosoma cumanense	DNA100622	AY326201	AY445139	AY326296	DQ300156
Siphonosoma cumanense	DNA100991	DQ300002	DQ300047	DQ300089	DQ300157
Siphonosoma vastum	DNA100625	DQ300003	AY445137	AY326297	DQ300158
Sipunculus (*S.*) *norvegicus*	DNA101069	DQ300004		DQ300090	DQ300159
Sipunculus (*S.*) *nudus*	DNA100234	DQ300005			DQ300160
Sipunculus (*S.*) *nudus*	DNA100245	AF519239	AF519269		
Sipunculus (*S.*) *nudus*	DNA100246	AF519240	AF519270	AF519295	DQ300161
Sipunculus (*S.*) *nudus*	DNA100468	DQ300006	DQ300048	DQ300091	DQ300162
Sipunculus (*S.*) *nudus*	DNA100629	DQ300007		DQ300092	DQ300163
Sipunculus (*S.*) *nudus*	DNA100993	DQ300008	DQ300049	DQ300093	DQ300164
Sipunculus (*S.*) *phalloides*	DNA101337	DQ300009		DQ300094	DQ300165
Sipunculus (*S.*) *polymyotus*	DNA101121	DQ300010		DQ300095	DQ300166
Xenosiphon branchiatus	DNA101086	DQ300016	DQ300050	DQ300101	DQ300172
Golfingiidae					
Golfingia elongata	DNA100465	DQ299969	DQ300031	DQ300065	DQ300121
Golfingia elongata	DNA100466	AF519242		DQ300066	DQ300122
Golfingia elongata	DNA101003	DQ299970			DQ300123
Golfingia elongata	DNA101066	DQ299971		DQ300067	DQ300124
Golfingia elongata	DNA101081	DQ299972		DQ300068	DQ300125
Golfingia margaritacea	DNA100738	DQ299973	DQ300032	DQ300069	DQ300126
Golfingia vulgaris	DNA100207	AF519244	AF519273		DQ300127
Nephasoma diaphanes	DNA101072	DQ299975		DQ300071	DQ300128
Nephasoma flagriferum	DNA100439	AF519243		AF519299	
Nephasoma flagriferum	DNA100440	DQ299976	DQ300033	DQ300072	DQ300129
Nephasoma flagriferum	DNA101071	DQ299977		DQ300073	DQ300130
Nephasoma pellucidum	DNA101009	DQ299978			DQ300131
Thysanocardia catherinae	DNA101068	DQ300015		DQ300099	
Thysanocardia nigra	DNA100606	AF519247	AF519274	DQ300100	
Themistidae					
Themiste (*T.*) *dyscrita*	DNA101095	DQ300011			DQ300167
Themiste (*T.*) *hennahi*	DNA100627	DQ300012		DQ300096	DQ300168
Themiste (*L.*) *lageniformis*	DNA100229	AF519249	AF519276	AF519302	DQ300169
Themiste (*L.*) *minor*	DNA100210	AF519250	F519277	AF519303	
Themiste (*L.*) *minor*	DNA101083	DQ300013		DQ300097	DQ300170
Themiste (*T.*) *pyroides*	DNA101084	DQ300014		DQ300098	DQ300171
Phascolionidae					
Onchnesoma steenstrupii	DNA101080	DQ299979	DQ300034	DQ300074	
Phascolion (*L.*) *cryptum*	DNA101007	DQ299980	DQ300035	DQ300075	DQ300132
Phascolion (*I.*) *gerardi*	DNA101002	DQ299981		DQ300076	
Phascolion (*P.*) *psammophilum*	DNA101006	DQ299982	DQ300036		DQ300133
Phascolion (*P.*) *strombus*	DNA100101	AF519248	AF519275	AF519301	
Phascolion (*P.*) *strombus*	DNA100739	DQ299983			
Phascolion (*P.*) *strombus*	DNA101077	DQ299984		DQ300077	
Phascolosomatidae					
Antillesoma antillarum	DNA100390	AF519259	AF519286	AF519311	
Antillesoma antillarum	DNA100759	DQ299950			
Antillesoma antillarum	DNA101008	DQ299951		DQ300051	DQ300102
Apionsoma (*A.*) *misakianum*	DNA100231	AF519260	AY445142		
Apionsoma (*A.*) *misakianum*	DNA100737	DQ299952	DQ300017	DQ300052	DQ300103
Apionsoma (*A.*) *murinae*	DNA100446	DQ299953	DQ300018		
Apionsoma (*E.*) *pectinatum*	DNA100624	AY326293	AY445142	AY326300	DQ300104
Phascolosoma (*P.*) *agassizii*	DNA101096	DQ299985	DQ300037	DQ300078	DQ300135
Phascolosoma (*P.*) *albolineatum*	DNA100396	AF519251	AF519278		DQ300136
Phascolosoma (*F.*) *capitatum*	DNA101070	DQ299986		DQ300079	DQ300137
Phascolosoma (*P.*) *granulatum*	DNA100201	AF519252	AF519279	AF519304	DQ300138
Phascolosoma (*P.*) *granulatum*		X79874			
Phascolosoma (*P.*) *nigrescens*	DNA100621	AY326292	AY445140	AY326299	DQ300139

Table 2 (*continued*)

Species	*MCZ Catalogue #*	*18S rRNA*	*28S rRNA*	*Histone H3*	COI
Phascolosoma (*P.*) *nigrescens*	DNA100736	DQ299987	DQ300038	DQ300080	DQ300140
Phascolosoma (*P.*) *nigrescens*	DNA100822	DQ299988	DQ300039	DQ300081	DQ300141
Phascolosoma (*P.*) *nigrescens*	DNA101010	DQ299989	DQ300040		DQ300142
Phascolosoma (*P.*) *nigrescens*	DNA101082	DQ299990	DQ300041		DQ300143
Phascolosoma (*P.*) *noduliferum*	DNA100208	AF519253	AF519280	AF519305	DQ300144
Phascolosoma (*P.*) *perlucens*	DNA100228	AF519254	AF519281	AF519306	DQ300145
Phascolosoma (*P.*) *perlucens*	DNA100233	DQ299991			
Phascolosoma (*P.*) *perlucens*	DNA100395	DQ299992		DQ300082	DQ300146
Phascolosoma (*P.*) *perlucens*	DNA100748	DQ299993	DQ300042		DQ300147
Phascolosoma (*P.*) *perlucens*	DNA100819	DQ299994			DQ300148
Phascolosoma (*P.*) *perlucens*	DNA100829	DQ299995	DQ300043	DQ300083	DQ300149
Phascolosoma (*P.*) *scolops*	DNA100373	AF519255	AF519282	AF519309	
Phascolosoma (*P.*) *scolops*	DNA100394	DQ299996			DQ300150
Phascolosoma (*P.*) *scolops*	DNA100735	DQ299997		DQ300084	DQ300151
Phascolosoma (*P.*) *scolops*	DNA100813	DQ299998	DQ300044	DQ300085	DQ300152
Phascolosoma (*P.*) *stephensoni*	DNA100203	DQ299999	DQ300045	DQ300086	
Phascolosoma (*P.*) *stephensoni*	DNA100209	AF519257	AF519284	AF519307	
Phascolosoma (*P.*) *stephensoni*	DNA100469	AF519256	AF519283	AF519310	DQ300153
Phascolosoma (*P.*) *stephensoni*	DNA100485	AF519258	AF519285	AF519308	
Phascolosoma (*P.*) *turnerae*	DNA100230	DQ300000	DQ300046	DQ300087	DQ300154
Aspidosiphonidae					
Aspidosiphon (*A.*) *albus*	DNA101017	DQ299954		DQ300053	DQ300105
Aspidosiphon (*A.*) *albus*	DNA101336	DQ299955		DQ300054	
Aspidosiphon (*A.*) *elegans*	DNA100977	DQ299956	DQ300019	DQ300055	
Aspidosiphon (*A.*) *elegans*	DNA101016	DQ299957	DQ300020	DQ300056	DQ300106
Aspidosiphon (*P.*) *fischeri*	DNA100620	AY326294		AY326301	DQ300107
Aspidosiphon (*P.*) *fischeri*	DNA100981	DQ299958	DQ300021		DQ300108
Aspidosiphon (*A.*) *gosnoldi*	DNA101014	DQ299959	DQ300022	DQ300057	DQ300109
Aspidosiphon (*A.*) *gracilis schnehageni*	DNA101087	DQ299960	DQ300023	DQ300058	DQ300110
Aspidosiphon (*P.*) *laevis*	DNA100467	AF519261	DQ300024	DQ300059	DQ300111
Aspidosiphon (*P.*) *laevis*	DNA100992	DQ299961			DQ300112
Aspidosiphon (*A.*) *misakiensis*	DNA100205	AF119090	AF519288	AF519312	
Aspidosiphon (*A.*) *muelleri*	DNA100206	DQ299962	DQ300025	DQ300060	DQ300113
Aspidosiphon (*P.*) *parvulus*	DNA100202	AF119075	DQ300026	DQ300061	
Aspidosiphon (*P.*) *parvulus*	DNA100375	DQ299963		DQ300062	DQ300114
Aspidosiphon (*P.*) *parvulus*	DNA100982	DQ299964	DQ300027	DQ300063	DQ300115
Aspidosiphon (*P.*) *steenstrupii*	DNA100232	AF519262	AF519291	AF519315	DQ300116
Aspidosiphon (*P.*) *steenstrupii*	DNA100372	DQ299965	DQ300028		DQ300117
Aspidosiphon (*P.*) *steenstrupii*	DNA100391	DQ299966	DQ300029		DQ300118
Aspidosiphon (*P.*) *steenstrupii*	DNA100630	DQ299967		DQ300064	DQ300119
Cloeosiphon aspergillus	DNA100393	AF519263	AF519292	AF519316	
Cloeosiphon aspergillus	DNA100825	DQ299968	DQ300030		DQ300120
Lithacrosiphon cristatus	DNA100623	AY326295	AY445142	AY326302	
Lithacrosiphon cristatus	DNA100986	DQ299974		DQ300070	
Nemertea					
Amphiporus sp.		AF119077	AF519265	AF519293	AJ436899
Argonemertes australiensis		AF519235	AF519264	AF519293	AY428840
Mollusca					
Lepidopleurus cajetanus		AF120502	AF120565	AY070142	AF120626
Rhabdus rectius		AF120523	AF120580	AY070144	AY260826
Yoldia limatula		AF120528	AF120585	AY070149	AF120642
Annelida					
Lumbrineris latreilli		AF519238	AF519267	AF185253	AY364855
Lamellibrachia spp.		AF168742		AF185235	U74055
Owenia fusiformis		AF448160	AY428824	AY428832	AY428839
Urechis caupo		AF119076	AF519268	X58895	U74077
Lumbricus terrestris		AJ272183	—	AF185262	NC001673
Entoprocta					
Loxosomella murmanica		AY218100	AY218129	AY218150	AY218083

were addressed by adjusting slop values (-checkslop 10). While doing tree refinements using TBR, -checkslop *n* accepts all trees that are within *n* tenths of a percent of the current minimum value. For example, -checkslop 10 accepts all trees up to 1% above the current minimum length while doing TBR.

POY facilitates efficient sensitivity analysis (Wheeler, 1995; Giribet, 2003). All data sets (individual genes and different combinations of genes) were analyzed under 12 parameter sets, for a range of indel-to-transversion ratios and transversion-to-transition ratios (see Table 3). Implied alignments—a topological-unique "alignment" or synapomorphy scheme (Wheeler, 2003; Giribet, 2005)—can be easily generated for each tree. Only when multiple partitions are combined in a single analysis will reciprocally inform each other about the homology statement—and this occurs only for those partitions that are not analyzed as prealigned (i.e., 18S rRNA and 28S rRNA).

To identify the optimal parameter set we employed a character-congruence technique which is a modification of the ILD (Incongruence Length Difference) metric developed by Mickevich and Farris (1981; see also Farris et al. 1995), as proposed by Wheeler (1995) (Table 3). The value is calculated for each parameter set by subtracting the sum of the scores of all partitions from the score of the combined analysis of all partitions, and normalizing it for the score of the combined length. Although the reliability of the ILD measure employed here has been questioned because it may show a trivial minimum in circumstances in which partitions are given disproportionate weights (Aagesen et al., 2005), this is not the case here. The ILD has been interpreted as a meta-optimality criterion for choosing the parameter set that best explains all partitions in combination, the one that maximizes overall congruence and minimizes character conflict among all the data (Giribet, 2003). But this congruence maximization comes from the individual partitions which are most congruent with the combined one. This parameter set was given special consideration in the analysis of data from each individual gene and is referred to throughout this paper as the "optimal parameter set". Additionally, we discuss results from the strict consensus of all parameter sets explored, which has been interpreted as a measure of stability to parameter choice, as applied in statistical sensitivity analyses (Wheeler, 1995; Giribet, 2003). Nodal support for all topologies was measured by parsimony jackknifing (Farris, 1997; Farris et al., 1996).

Table 3
Tree lengths for the different partitions analyzed (18S, 18S rRNA; 28S, 28S rRNA; H3, histone H3; COI, cytochrome *c* oxidase subunit I; MOR, morphological data; MOL, 4 loci combined; TOT, morphology + 4 loci combined) and congruence value (ILD) for the combined analysis of morphology + 4 molecular loci combined at different parameter sets (left column)

	18S	28S	H3	COI	MOR	MOL	TOT	ILD
111	**3256**	**721**	**1191**	**5539**	**154**	**11662**	**11909**	**0.08800**
121	4878	1101	1712	8134	308	17256	17731	0.09012
141	8033	1798	2712	13056	616	28052	29030	0.09697
181	14229	3185	4688	22848	1232	49487	51480	0.10291
211	3677	846	1191	5582	308	12333	12800	0.09344
221	5627	1323	1712	8192	616	18428	19363	0.09776
241	9482	2258	2712	13184	1232	30338	32227	0.10423
281	17122	4099	4688	23127	2464	54117	57865	0.11000
411	4311	1037	1191	5595	616	13297	14239	0.10457
421	6868	1674	1712	8217	1232	20380	22207	0.11276
441	11946	2917	2712	13227	2464	34202	38004	0.12467
481	22061	5368	4688	23190	4928	61769	69254	0.13023

The first numeral used in the parameter set column corresponds to the ratio between indel-to-transversion and the following two numbers correspond to the ratio between transversion-to-transition; e.g., 111 is equal weights; 121 corresponds to a indel-to-transversion ratio of 1 and a transversion-to-transition ratio of 2:1—so indels have a cost of 2, transversions have a cost of 2 and transitions have a cost of 1 (for a list of the specific step matrices that this involves see Giribet et al., 2002: App. 4). Optimal ILD value is indicated in bold.

In order to evaluate the potential effect of treating each gap as an independent character (e.g. Giribet and Wheeler, 1999; Simmons and Ochoterena, 2000), we ran the combined analysis of all data using a non-linear (affine) gap function, where the gap opening value was higher (2) than that of the gap extension (1). This was used for generating the implied alignment to be used in the subsequent Bayesian analysis.

We performed the Bayesian analysis under mixed models with the complete dataset of the morphological data and all four gene regions, based on the implied alignment from the POY analyses. For this purpose, the 18S rRNA sequence was divided into two partitions corresponding to stem and loop regions. To identify secondary structure features, the annotated sequence for *Phascolosoma granulatum* (GenBank Accession No. X79874) was downloaded from the European Ribosomal Database (Van de Peer et al., 2000) and used as an annotation reference for the remaining taxa. A list of 388 nucleotide pairings was assembled manually for the stem regions. The stem regions were analyzed under a doublet model with a single rate parameter and sixteen states (Schöninger and von Haeseler, 1994), representing all possible nucleotide pairings. Currently, the doublet model cannot be tested for goodness of fit against other models with common software programs, such as Modeltest (Posada and Crandall, 1998) or MrModeltest (Nylander, 2004), but recent studies have shown that the 16-state doublet model significantly outperformed simpler, 4-state, single-nucleotide models for stem regions of ribosomal sequences (Savill et al., 2001; Telford et al., 2005). No secondary structure information was used for the fragment of 28S rRNA because (1) no appropriate sipunculan reference sequence was available, and (2) only ca. 10% of the total 28S rRNA gene was sequenced for this study, resulting in few nucleotide pairings in this fragment.

All four gene regions were separately tested for the appropriateness of 24 different models of nucleotide substitution in MrModeltest 2.2 (Nylander, 2004). The same test was performed on the 18S rRNA loop regions only. The data sets were submitted to four hierarchical likelihood ratio tests (hLRT). In addition, the Akaike Information

Criterion (AIC) was calculated for each of the 24 models. For the 18S loop regions, one hLRT favored a general time reversible model but the remaining three hLRTs as well as the AIC favored a symmetrical model, which assumes equal base frequencies (Zarkihk, 1994). The symmetrical model was implemented with corrections for a discrete gamma distribution of substitution rates (four categories) (G) and a proportion of invariable sites (I). For the complete 18S rRNA and the other three gene regions, all hLRTs and the AIC favored the General Time Reversible (GTR) model (Tavaré, 1986) which was also implemented with corrections for a gamma distribution and a proportion of invariable sites. The 58 morphological characters (Appendix B, C) were analyzed under the discrete likelihood model proposed by Lewis (2001), as implemented in MrBayes 3.1. This model is similar to a Jukes–Cantor model (Jukes and Cantor, 1969), except that the number of states can vary from 2 to 10.

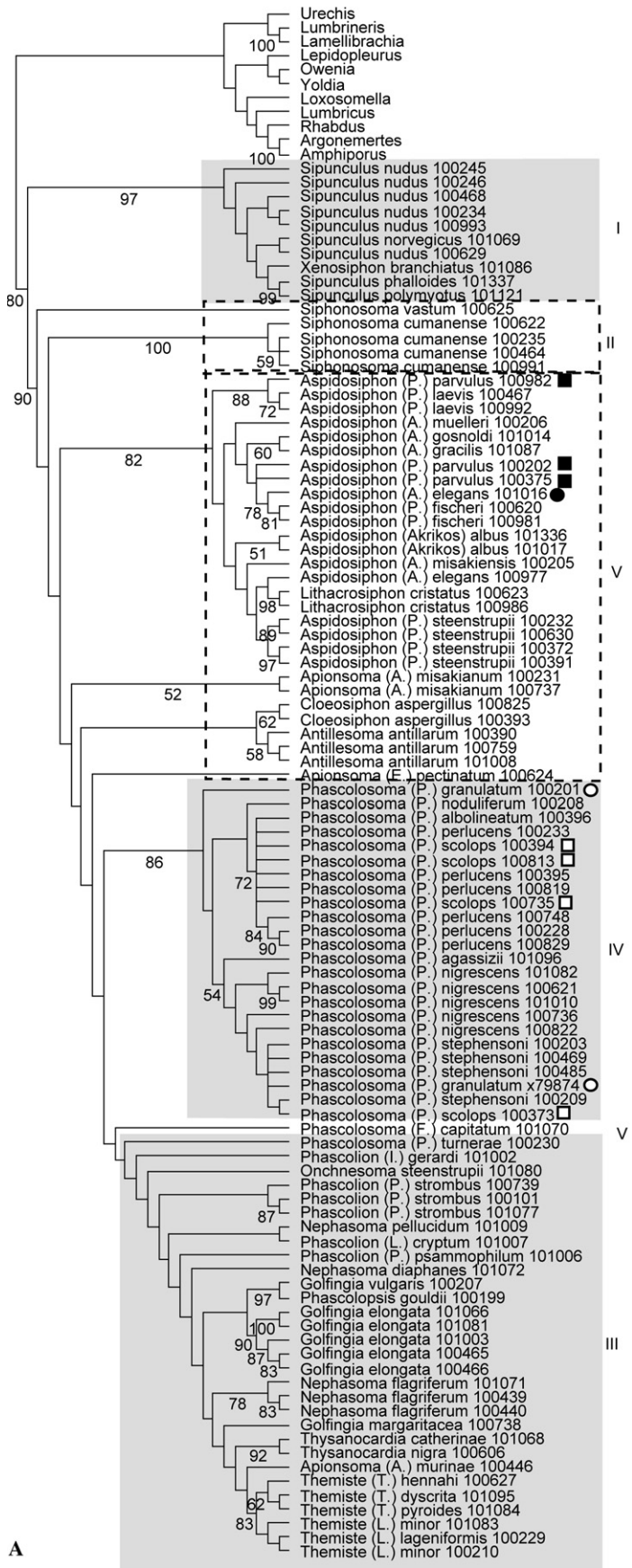
Phylogenetic analysis was performed using Bayesian statistics with a Metropolis coupled Markov Chain Monte Carlo algorithm as implemented in MrBayes 3.1.1 (Ronquist and Huelsenbeck, 2003). Two runs with four chains each were performed simultaneously for 1,500,000 generations, sampling trees every 500 generations. The temperature parameter was set to 0.08. We chose a lower temperature parameter than the default setting of 0.2, because the cold chain barely swapped among the four chains of a run under the default setting. The value of 0.08 was chosen after several initial trials under different temperature parameters. It increased the number of cold chain swaps and stationarity was achieved sooner than with higher parameter values. The initial 500,000 generations were discarded as burn-in. To test for the effect of the doublet model, the 18S rRNA partition was analyzed both under a single model (GTR+I+G) as well as under mixed models for stem and loop regions (doublet and symmetrical model, respectively).

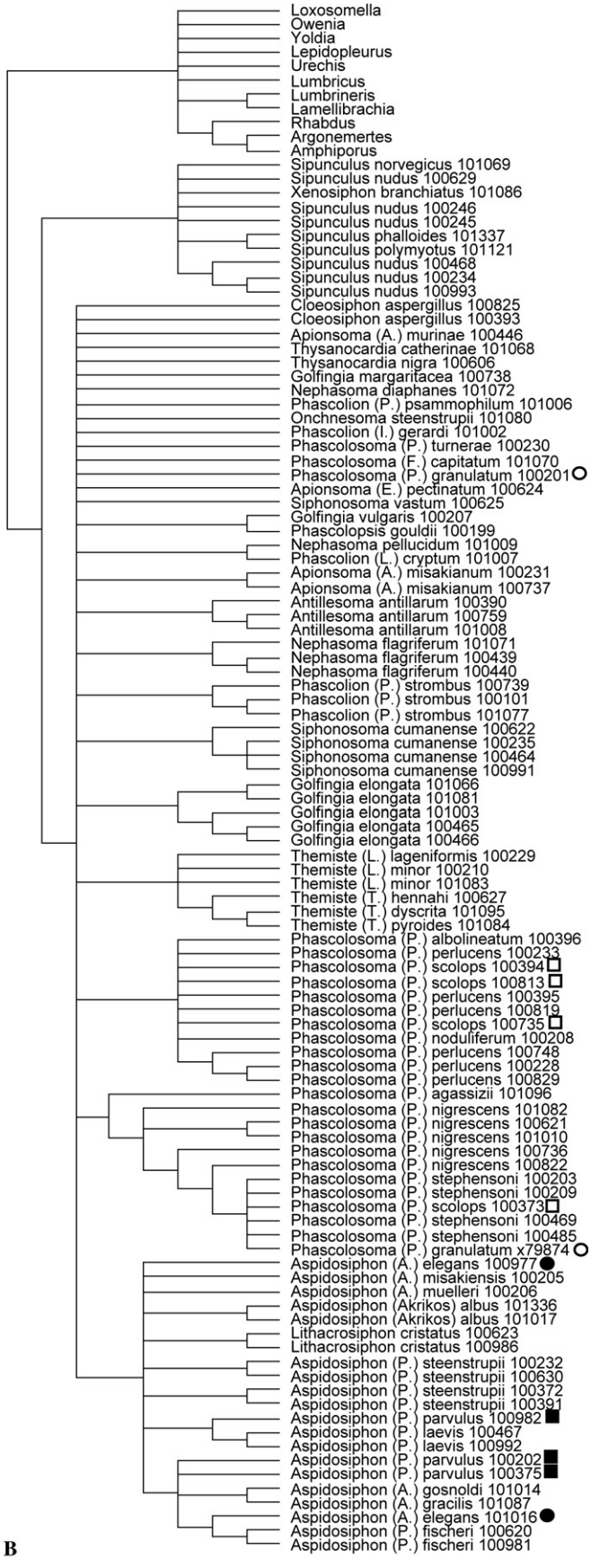
3. Results

For the POY analysis, parameter set "111" which specifies a ratio of 1:1 for both transversion-to-transition and indel-to-transversion ratios was determined to be the optimal parameter set, because it minimized overall incongruence as measured by the ILD test (Table 3). The combined analysis of all four gene regions and the morphological dataset resulted in 30 most parsimonious trees of length 11,909. The strict consensus of all most parsimonious trees for the optimal parameter set with jackknife support values is shown in Fig. 1A. Fig. 1B illustrates the strict consensus of all parameter sets for the combined analysis of all data. All parameter sets identify monophyly of Sipuncula, and a major split uniting the two sipunculidean genera *Sipunculus* and *Xenosiphon* (97% jackknife support; JF hereafter) as the sister group to all other sipunculans (90% JF). This clade shows paraphyly of *Siphonosoma* and identifies several clades, some well supported and stable to parameter variation, such as the clade including all *Aspidosiphon* + *Lithacrosiphon* or a clade including most species of *Phascolosoma*, except for *P. capitatum* and *P. turnerae*. Monophyly of the genus *Themiste* is also well supported. Finally, *Phascolopsis gouldii* and *Golfingia vulgaris* form a clade throughout the entire parameter space and have a JF of 97%, one of the highest values in the analysis.

The Bayesian analysis of 18S rRNA under mixed models resulted in significantly higher log likelihoods-lnL than those of the analysis under a single model. Among the 4000 sampled trees (2000 from each of the two runs), the best likelihood-lnL score was −16988.8 (mean: -17037.7 ± 16.5) under mixed models and −17761.00 (mean: -17820.00 ± 27.9) under a single model.

The 50% majority rule consensus tree resulting from the combined Bayesian analysis is shown in Fig. 2. The tree topology is identical to the tree excluding morphological data. The analysis resulted in five major clades of which four are supported by 100% posterior probability (pp hereafter). The five clades are referred to as clade I through V for comparative purpose and are also indicated in Fig. 1A. Clade I is the sister group to all other sipunculans and contains all representatives of the genera *Xenosiphon* and *Sipunculus*, with *Xenosiphon branchiatus* nested within *Sipunculus*. Clade II includes the two *Siphonosoma* species, with *S. vastum* as the sister group to the four representatives of *S. cumanense*. Clade III includes representatives of nine different genera. Of those, only *Themiste* and *Thysanocardia* are monophyletic. *Phascolion, Nephasoma* and *Golfingia* are clearly polyphyletic. The monophyly of *Onchnesoma* could not be tested because only one species (*O. steenstrupii*) was included in the analysis. *Phascolopsis gouldii* represents a monotypic genus. *Phascolosoma turnerae* and *Apionsoma murinae* are part of larger genera that fall into other clades in the tree. Clade IV includes all shallow-water *Phascolosoma* species. *P. turnerae* and *P. capitatum,* both deep-sea inhabitants, are included in clades III and V, respectively. Clade V is only supported by 56% pp, and appears polyphyletic in the direct optimization analysis. It contains representatives of six genera. *Phascolosoma capitatum* and the two *Apionsoma* species form the basal branches in the clade. The two representatives of *Apionsoma misakianum* form a clade with *A. pectinatum*, although with low pp (50%). The rest of the clade is well resolved: with 100% pp, *Cloeosiphon*, *Antillesoma*, *Lithacrosiphon* and *Aspidosiphon* form a clade. The two representatives of *Cloeosiphon aspergillus*, three representatives of *Antillesoma antillarum* and two representatives of *Lithacrosiphon cristatus* all form clades. *Lithacrosiphon* and *Aspidosiphon* form a highly supported clade (100% pp). *Lithacrosiphon* is nested within *Aspidosiphon*. Sipunculidea, as defined in Gibbs and Cutler (1987), does not appear as monophyletic. Clades IV + V basically correspond to the Phascolosomatidea, but *Apionsoma murinae* and *Phascolosoma turnerae*, two species that morphologically fall into Phascolosomatidea, appear in clade III.


Urechis
Lumbrineris
Lamellibrachia
Lepidopleurus
Owenia
Yoldia
Loxosomella
Lumbricus
Rhabdus
Argonemertes
Amphiporus
Sipunculus nudus 100245
Sipunculus nudus 100246
Sipunculus nudus 100468
Sipunculus nudus 100234
Sipunculus nudus 100993
Sipunculus norvegicus 101069
Sipunculus nudus 100629
Xenosiphon branchiatus 101086
Sipunculus phalloides 101337
Sipunculus polymyotus 101121
Siphonosoma vastum 100625
Siphonosoma cumanense 100622
Siphonosoma cumanense 100235
Siphonosoma cumanense 100464
Siphonosoma cumanense 100991
Aspidosiphon (P.) parvulus 100982
Aspidosiphon (P.) laevis 100467
Aspidosiphon (P.) laevis 100992
Aspidosiphon (A.) muelleri 100206
Aspidosiphon (A.) gosnoldi 101014
Aspidosiphon (A.) gracilis 101087
Aspidosiphon (P.) parvulus 100202
Aspidosiphon (P.) parvulus 100375
Aspidosiphon (A.) elegans 101016
Aspidosiphon (P.) fischeri 100620
Aspidosiphon (P.) fischeri 100981
Aspidosiphon (Akrikos) albus 101336
Aspidosiphon (Akrikos) albus 101017
Aspidosiphon (A.) misakiensis 100205
Aspidosiphon (A.) elegans 100977
Lithacrosiphon cristatus 100623
Lithacrosiphon cristatus 100986
Aspidosiphon (P.) steenstrupii 100232
Aspidosiphon (P.) steenstrupii 100630
Aspidosiphon (P.) steenstrupii 100372
Aspidosiphon (P.) steenstrupii 100391
Apionsoma (A.) misakianum 100231
Apionsoma (A.) misakianum 100737
Cloeosiphon aspergillus 100825
Cloeosiphon aspergillus 100393
Antillesoma antillarum 100390
Antillesoma antillarum 100759
Antillesoma antillarum 101008
Apionsoma (E.) pectinatum 100624
Phascolosoma (P.) granulatum 100201
Phascolosoma (P.) noduliferum 100208
Phascolosoma (P.) albolineatum 100396
Phascolosoma (P.) perlucens 100233
Phascolosoma (P.) scolops 100394
Phascolosoma (P.) scolops 100813
Phascolosoma (P.) perlucens 100395
Phascolosoma (P.) perlucens 100819
Phascolosoma (P.) scolops 100735
Phascolosoma (P.) perlucens 100748
Phascolosoma (P.) perlucens 100228
Phascolosoma (P.) perlucens 100829
Phascolosoma (P.) agassizii 101096
Phascolosoma (P.) nigrescens 101082
Phascolosoma (P.) nigrescens 100621
Phascolosoma (P.) nigrescens 101010
Phascolosoma (P.) nigrescens 100736
Phascolosoma (P.) nigrescens 100822
Phascolosoma (P.) stephensoni 100203
Phascolosoma (P.) stephensoni 100469
Phascolosoma (P.) stephensoni 100485
Phascolosoma (P.) granulatum x79874
Phascolosoma (P.) stephensoni 100209
Phascolosoma (P.) scolops 100373
Phascolosoma (F.) capitatum 101070
Phascolosoma (P.) turnerae 100230
Phascolion (I.) gerardi 101002
Onchnesoma steenstrupii 101080
Phascolion (P.) strombus 100739
Phascolion (P.) strombus 100101
Phascolion (P.) strombus 101077
Nephasoma pellucidum 101009
Phascolion (L.) cryptum 101007
Phascolion (P.) psammophilum 101006
Nephasoma diaphanes 101072
Golfingia vulgaris 100207
Phascolopsis gouldii 100199
Golfingia elongata 101066
Golfingia elongata 101081
Golfingia elongata 101003
Golfingia elongata 100465
Golfingia elongata 100466
Nephasoma flagriferum 101071
Nephasoma flagriferum 100439
Nephasoma flagriferum 100440
Golfingia margaritacea 100738
Thysanocardia catherinae 101068
Thysanocardia nigra 100606
Apionsoma (A.) murinae 100446
Themiste (T.) hennahi 100627
Themiste (T.) dyscrita 101095
Themiste (T.) pyroides 101084
Themiste (L.) minor 101083
Themiste (L.) lageniformis 100229
Themiste (L.) minor 100210
I
II
V
IV
V
III
A


Loxosomella
Owenia
Yoldia
Lepidopleurus
Urechis
Lumbricus
Lumbrineris
Lamellibrachia
Rhabdus
Argonemertes
Amphiporus
Sipunculus norvegicus 101069
Sipunculus nudus 100629
Xenosiphon branchiatus 101086
Sipunculus nudus 100246
Sipunculus nudus 100245
Sipunculus phalloides 101337
Sipunculus polymyotus 101121
Sipunculus nudus 100468
Sipunculus nudus 100234
Sipunculus nudus 100993
Cloeosiphon aspergillus 100825
Cloeosiphon aspergillus 100393
Apionsoma (A.) murinae 100446
Thysanocardia catherinae 101068
Thysanocardia nigra 100606
Golfingia margaritacea 100738
Nephasoma diaphanes 101072
Phascolion (P.) psammophilum 101006
Onchnesoma steenstrupii 101080
Phascolion (I.) gerardi 101002
Phascolosoma (P.) turnerae 100230
Phascolosoma (F.) capitatum 101070
Phascolosoma (P.) granulatum 100201
Apionsoma (E.) pectinatum 100624
Siphonosoma vastum 100625
Golfingia vulgaris 100207
Phascolopsis gouldii 100199
Nephasoma pellucidum 101009
Phascolion (L.) cryptum 101007
Apionsoma (A.) misakianum 100231
Apionsoma (A.) misakianum 100737
Antillesoma antillarum 100390
Antillesoma antillarum 100759
Antillesoma antillarum 101008
Nephasoma flagriferum 101071
Nephasoma flagriferum 100439
Nephasoma flagriferum 100440
Phascolion (P.) strombus 100739
Phascolion (P.) strombus 100101
Phascolion (P.) strombus 101077
Siphonosoma cumanense 100622
Siphonosoma cumanense 100235
Siphonosoma cumanense 100464
Siphonosoma cumanense 100991
Golfingia elongata 101066
Golfingia elongata 101081
Golfingia elongata 101003
Golfingia elongata 100465
Golfingia elongata 100466
Themiste (L.) lageniformis 100229
Themiste (L.) minor 100210
Themiste (L.) minor 101083
Themiste (T.) hennahi 100627
Themiste (T.) dyscrita 101095
Themiste (T.) pyroides 101084
Phascolosoma (P.) albolineatum 100396
Phascolosoma (P.) perlucens 100233
Phascolosoma (P.) scolops 100394
Phascolosoma (P.) scolops 100813
Phascolosoma (P.) perlucens 100395
Phascolosoma (P.) perlucens 100819
Phascolosoma (P.) scolops 100735
Phascolosoma (P.) noduliferum 100208
Phascolosoma (P.) perlucens 100748
Phascolosoma (P.) perlucens 100228
Phascolosoma (P.) perlucens 100829
Phascolosoma (P.) agassizii 101096
Phascolosoma (P.) nigrescens 101082
Phascolosoma (P.) nigrescens 100621
Phascolosoma (P.) nigrescens 101010
Phascolosoma (P.) nigrescens 100736
Phascolosoma (P.) nigrescens 100822
Phascolosoma (P.) stephensoni 100203
Phascolosoma (P.) stephensoni 100209
Phascolosoma (P.) scolops 100373
Phascolosoma (P.) stephensoni 100469
Phascolosoma (P.) stephensoni 100485
Phascolosoma (P.) granulatum x79874
Aspidosiphon (A.) elegans 100977
Aspidosiphon (A.) misakiensis 100205
Aspidosiphon (A.) muelleri 100206
Aspidosiphon (Akrikos) albus 101336
Aspidosiphon (Akrikos) albus 101017
Lithacrosiphon cristatus 100623
Lithacrosiphon cristatus 100986
Aspidosiphon (P.) steenstrupii 100232
Aspidosiphon (P.) steenstrupii 100630
Aspidosiphon (P.) steenstrupii 100372
Aspidosiphon (P.) steenstrupii 100391
Aspidosiphon (P.) parvulus 100982
Aspidosiphon (P.) laevis 100467
Aspidosiphon (P.) laevis 100992
Aspidosiphon (P.) parvulus 100202
Aspidosiphon (P.) parvulus 100375
Aspidosiphon (A.) gosnoldi 101014
Aspidosiphon (A.) gracilis 101087
Aspidosiphon (A.) elegans 101016
Aspidosiphon (P.) fischeri 100620
Aspidosiphon (P.) fischeri 100981
B

Of the five clades as defined above, clade I is well supported in the direct optimization analysis under the optimal parameter set (97% JF) and stable across all parameters. Its topology differs from the topology of the Bayesian tree only with respect to *Sipunculus nudus* 100629 and *Sipunculus norvegicus* 101069. Clade IV has 86% jackknife support and it is also found under all parameter sets, with the exception of one parameter set that resolves the clade under some of the most parsimonious trees, but not all. Its topology is similar to the topology in the Bayesian analysis: *Phascolosoma noduliferum* is the sister taxon the other *Phascolosoma* species which group into two clades. Clade III appears in the consensus tree of the analysis under the optimal parameter set but has jackknife support below 50%. Furthermore, it is unstable to parameter set variation. However, there are several well-supported groups within the clade found in both analyses. Clades II and V are not supported in the direct optimization analysis, although there is high support for an *Aspidosiphon* + *Lithacrosiphon* clade which is also stable under different parameter sets.

Of the 21 species of which more than one individual was sequenced, eleven appeared as monophyletic in both analyses; one (*Aspidosiphon steenstrupii*) appeared as monophyletic only in the direct optimization analysis (Table 4).

Fig. 3 shows a tree based on a combination of the parsimony jackknife tree and the 50% majority rule consensus tree resulting from the Bayesian analysis of the morphological dataset. Both trees differ only in their degree of resolution. For the most part, the tree resulting from the Bayesian analysis shows higher resolution than the parsimony jackknife tree. However, the clade that includes all sipunculans except *Sipunculus* and *Xenosiphon* is weakly supported in the parsimony analysis only and remains unresolved in the Bayesian analysis. Of the five major clades identified in the Bayesian combined analysis (Fig. 2), only clade II also finds support in the morphological data with both types of analyses. The Bayesian analysis of the morphological data supports a monophyletic Phascolosomatidea.

4. Discussion

4.1. Phylogeny of the Sipuncula

Our analyses show that the large, sediment-burrowing species of the genera *Sipunculus, Xenosiphon* and *Siphonosoma* represent early branches in the sipunculan phylogenetic tree (clades I and II). *S. nudus* has long been used as a model sipunculan for physiological and biochemical studies, however, Cutler (1994) pointed out that the species is in many ways not a "typical" sipunculan because it differs from the majority of species in embryology, structure of the ventral nerve cord, coelomic urn cells, regeneration capabilities, osmoregulation and chromosome number. Because of these peculiarities, Cutler suspected that *S. nudus* would have a highly derived position in the sipunculan tree. Our Bayesian analysis shows that the *Sipunculus*/*Xenosiphon* clade is separated from the other sipunculans by long branch lengths, reflecting its morphological distinctness, but instead of being highly derived in the tree, the clade is the sister group to the remaining sipunculan species.

While the basal position of *Sipunculus* has been proposed previously (Maxmen et al., 2003; Schulze et al., 2005), our present study is the first to suggest that *Siphonosoma* is the next to branch off in the sipunculan tree. This is not only supported by the molecular data but also by the morphological data when analyzed under parsimony und rooted with *Sipunculus nudus.* In both previous studies, *Siphonosoma* appeared at the base of the Phascolosomatidea clade. *Siphonosoma* shows many morphological similarities with *Sipunculus* and *Xenosiphon*, such as the large elongated body with a short introvert, prominent papillae along the introvert and circular and longitudinal musculature split into bands. Among others, these similarities were the basis for the definition of the family Sipunculidae Rafinesque, 1814, but they appear to be plesiomorphic for the phylum. The family Sipunculidae is not supported in our analyses.

Clade III shows a high degree of morphological diversity. It contains representatives of nine different genera, only two of which—*Thysanocardia* and *Themiste*—are monophyletic according to our analyses. Both *Thysanocardia* and *Themiste* are morphologically clearly defined by their tentacular crown. In *Thysanocardia*, the tentacles are very numerous and extend some way along the introvert as festoons (Cutler, 1994). *Themiste* is the only sipunculan genus with branched tentacles, presumably an adaptation for filter feeding, whereas all other sipunculans have simple tentacles and seem to be primarily deposit feeders. We have not been able to detect any morphological trends among the other members of clade III. The number of introvert retractor muscles varies between one and four; introvert hooks may be absent or present; if present, they may be scattered or form rings.

Clade III basically corresponds to Golfingiiformes, except that *Phascolopsis gouldii*, previously thought to belong to Sipunculidae, and two species that morphologically fall into Phascolosomatiformes are included. Of the latter, *Apionsoma murinae*, is a small deep-water species of which we only had a single specimen. The specimen was identified by a sipunculan expert (J.I. Saiz Salinas) but it is a small, thread-like species

Fig. 1. (A) strict consensus tree of 30 most parsimonious trees (length 11,909) resulting from the direct optimization analysis of all data in POY under the optimal parameter set 111. Shaded boxes refer to clades defined in the Bayesian analysis (Fig. 2). Dashed boxes refer to non-monophyletic groupings that were monophyletic in the Bayesian analysis. Roman numerals correspond to clade designations in Fig. 2. Branch support is indicated by jackknife percentages (only >50%) at the nodes. (B) Strict consensus of most parsimonious trees found across all parameter sets analyzed. Squares and circles following species names indicate polyphyletic species.


Loxosomella
Owenia
Lepidopleurus
Yoldia
Rhabdus
Argonemertes
Amphiporus
Lumbricus
Urechis
Lumbrineris
Lamellibrachia
Sipunculus nudus 100245
Sipunculus nudus 100246
Sipunculus nudus 100468
Sipunculus nudus 100234
Sipunculus nudus 100993
Sipunculus nudus 100629
Sipunculus norvegicus 101069
Xenosiphon branchiatus 101086
Sipunculus phalloides 101337
Sipunculus polymyotus 101121
I
Siphonosoma vastum 100625
Siphonosoma cumanense 100622
Siphonosoma cumanense 100235
Siphonosoma cumanense 100464
Siphonosoma cumanense 100991
II
Phascolosoma turnerae 100230
Phascolion (I.) gerardi 101002
Nephasoma pellucidum 101009
Phascolion (P.) psammophilum 101006
Phascolion (L.) cryptum 101007
Onchnesoma steenstrupii 101080
Phascolion (P.) strombus 100739
Phascolion (P.) strombus 100101
Phascolion(P.) strombus 101077
Nephasoma flagriferum 101071
Nephasoma flagriferum 100439
Nephasoma flagriferum 100440
Golfingia vulgaris 100207
Phascolopsis gouldii 100199
Nephasoma diaphanes 101072
Golfingia elongata 101066
Golfingia elongata 101081
Golfingia elongata 101003
Golfingia elongata 100465
Golfingia elongata 100466
Golfingia margaritacea 100738
Thysanocardia catherinae 101068
Thysanocardia nigra 100606
Apionsoma (A.) murinae 100446
Themiste (T.) pyroides 101084
Themiste (T.) dyscrita 101095
Themiste (T.) hennahi 100627
Themiste (L.) minor 100210
Themiste (L.) lageniformis 100229
Themiste (L.) minor101083
III
Phascolosoma (P.) granulatum 100201
Phascolosoma (P.) noduliferum 100208
Phascolosoma (P.) scolops 100735
Phascolosoma (P.) perlucens 100233
Phascolosoma (P.) perlucens 100395
Phascolosoma (P.) scolops 100813
Phascolosoma (P.) albolineatum 100396
Phascolosoma (P.) scolops 100394
Phascolosoma (P.) perlucens 100819
Phascolosoma (P.) perlucens 100748
Phascolosoma (P.) perlucens 100228
Phascolosoma (p.) perlucens 100829
Phascolosoma (P.) nigrescens 101082
Phascolosoma (P.) nigrescens 100621
Phascolosoma (P.) nigrescens 101010
Phascolosoma (P.) nigrescens 100736
Phascolosoma (P.) nigrescens 100822
Phascolosoma (P.) stephensoni 100203
Phascolosoma (P.) granulatum x79874
Phascolosoma (P.) stephensoni 100485
Phascolosoma (P.) stephensoni 100469
Phascolosoma (P.) scolops 100373
Phascolosoma (P.) agassizii101096
Phascolosoma (P.) stephensoni100209
IV
Phascolosoma (F.) capitatum 101070
Apionsoma (E.) pectinatum 100624
Apionsoma (A.) misakianum 100231
Apionsoma (A.) misakianum 100737
Cloeosiphon aspergillus 100825
Cloeosiphon aspergillus 100393
Antillesoma antillarum 100390
Antillesoma antillarum 100759
Antillesoma antillarum 101008
Lithacrosiphon cristatus 100623
Lithacrosiphon cristatus 100986
Aspidosiphon (P.) steenstrupii 100232
Aspidosiphon (P.) steenstrupii 100630
Aspidosiphon (P.) steenstrupii 100372
Aspidosiphon (P.) steenstrupii 100391
Aspidosiphon (A.) elegans 100977
Aspidosiphon (A.) misakiensis 100205
Aspidosiphon (A.) albus 101336
Aspidosiphon (A.) albus 101017
Aspidosiphon (P.) parvulus 100982
Aspidosiphon (P.) laevis 100467
Aspidosiphon (P.) laevis 100992
Aspidosiphon (A.) muelleri 100206
Aspidosiphon (A.) gosnoldi 101014
Aspidosiphon (A.) gracilis 101087
Aspidosiphon (P.) parvulus 100202
Aspidosiphon (P.) parvulus 100375
Aspidosiphon (A.) elegans 101016
Aspidosiphon (P.) fischeri 100620
Aspidosiphon (P.) fischeri 100981
V
0.1

Table 4
Jackknife support or posterior probability (both in %) for species with multiple representatives in the two analyses of combined datasets presented in Figs. 1 and 2

Species	Number of individuals	Direct optimization analysis		Bayesian analysis	
		Branch support for monophyly (jackknife %)	Maximum branch support contradicting monophyly (jackknife %)	Branch support for monophyly (% posterior probability)	Maximum branch support contradicting monophyly (% posterior probability)
Antillesoma antillarum	3	58		100	
Apionsoma misakianum	2	52		100	
Aspidosiphon albus	2	51		99	
Aspidosiphon elegans	2	Nm	<50	Nm	100
Aspidosiphon fischeri	2	81		100	
Aspidosiphon laevis	2	72		100	
Aspidosiphon parvulus	3	Nm	88	Nm	100
Aspidosiphon steenstrupii	4	<50		Nm	54
Cloeosiphon aspergillus	2	62		100	
Golfingia elongata	5	90		100	
Lithacrosiphon cristatus	2	98		100	
Nephasoma flagriferum	3	78		100	
Phascolion strombus	3	<50		100	
Phascolosoma granulatum	2	Nm	54	Nm	99
Phascolosoma nigrescens	5	Nm	<50	Nm	99
Phascolosoma perlucens	6	Nm	<50	Nm	<50
Phascolosoma scolops	4	Nm	72	Nm	100
Phascolosoma stephensoni	4	Nm	<50	Nm	97
Siphonosoma cumanense	4	100		100	
Sipunculus nudus	6	Nm	<50	Nm	75
Themiste minor	2	Nm	<50	Nm	53

Numbers of individuals sequenced given in parentheses. NM, non-monophyletic.

for which identification is difficult. After fixation in ethanol and use of most of the tissue for DNA extraction, the identification can no longer be verified. This clearly demonstrates the need to keep adequately preserved voucher material in collections whenever possible. Furthermore, only approximately 1300 bp of 18S and 28S rRNA were sequenced for this species, potentially causing uncertainties in its placement. The second species that unexpectedly does not group with Phascolosomatidea is *Phascolosoma turnerae*. In both the direct optimization and the Bayesian analysis, it splits off early in clade III. *P. turnerae* is a deep-water wood-dwelling species (Rice, 1985). There is no doubt about its identification and the only gene region missing is COI. Although *P. turnerae* is unusual with respect to its depth range and habitat—most other *Phascolosoma* species either inhabit crevices of rocks or burrow into soft rocks in shallow water—it morphologically fits the description of the genus. Currently, our best explanation for its unexpected placement in our trees is that the morphological similarities between *P. turnerae* and other *Phascolosoma* species may be convergences, and this deserves further anatomical work.

Our placement of *P. gouldii* in clade III, and specifically as the sister species to *Golfingia vulgaris*, confirms the findings of Maxmen et al. (2003) and Schulze et al. (2005). *P. gouldii* is a species with a confusing taxonomic history. Prior to Cutler and Gibbs (1985) who placed it within Sipunculidae, it had been associated with species that are now considered Golfingiidae. Our findings therefore confirm original ideas about the evolutionary relationships of the species.

The division of the Sipuncula into the two major groups, Sipunculidea and Phascolosomatidea, is morphologically plausible. However, none of our analyses showed support for a monophyletic Sipunculidea, not even the analyses based on morphological data alone. The Bayesian analyses show strong support for a Phascolosomatidea clade, although the combined analysis excludes *Apionsoma murinae* and *Phascolosoma turnerae* from this clade, as discussed above. Other species with uncertain affinities in both analyses are *Phascolosoma* (*Fisherana*) *capitatum, Apionsoma* (*Edmondsius*) *pectinatum* and *Apionsoma* (*Apionsoma*) *misakianum*. In the Bayesian analysis, they form basal branches in clade V, but the branch support for such position is not significant. In the direct optimization analysis, *Phascolosoma capitatum* is the sister group to clade III. *P. capitatum* is morphologically distinguished from other *Phascolosoma* species by having the body wall musculature organized as a smooth sheet and not in bands. The affiliation of this species has changed repeatedly in the past: Stephen and Edmonds (1972) treated *Fisherana* as a distinct genus, Cutler (1979) associ-

Fig. 2. Fifty percent majority rule consensus tree resulting from the Bayesian analysis of four gene regions and morphology. Shaded boxes with roman numerals indicate major clades. Branch support is indicated as percent posterior probability. Asterisks indicate branch support of 100% posterior probability. Squares and circles following species names indicate polyphyletic species.

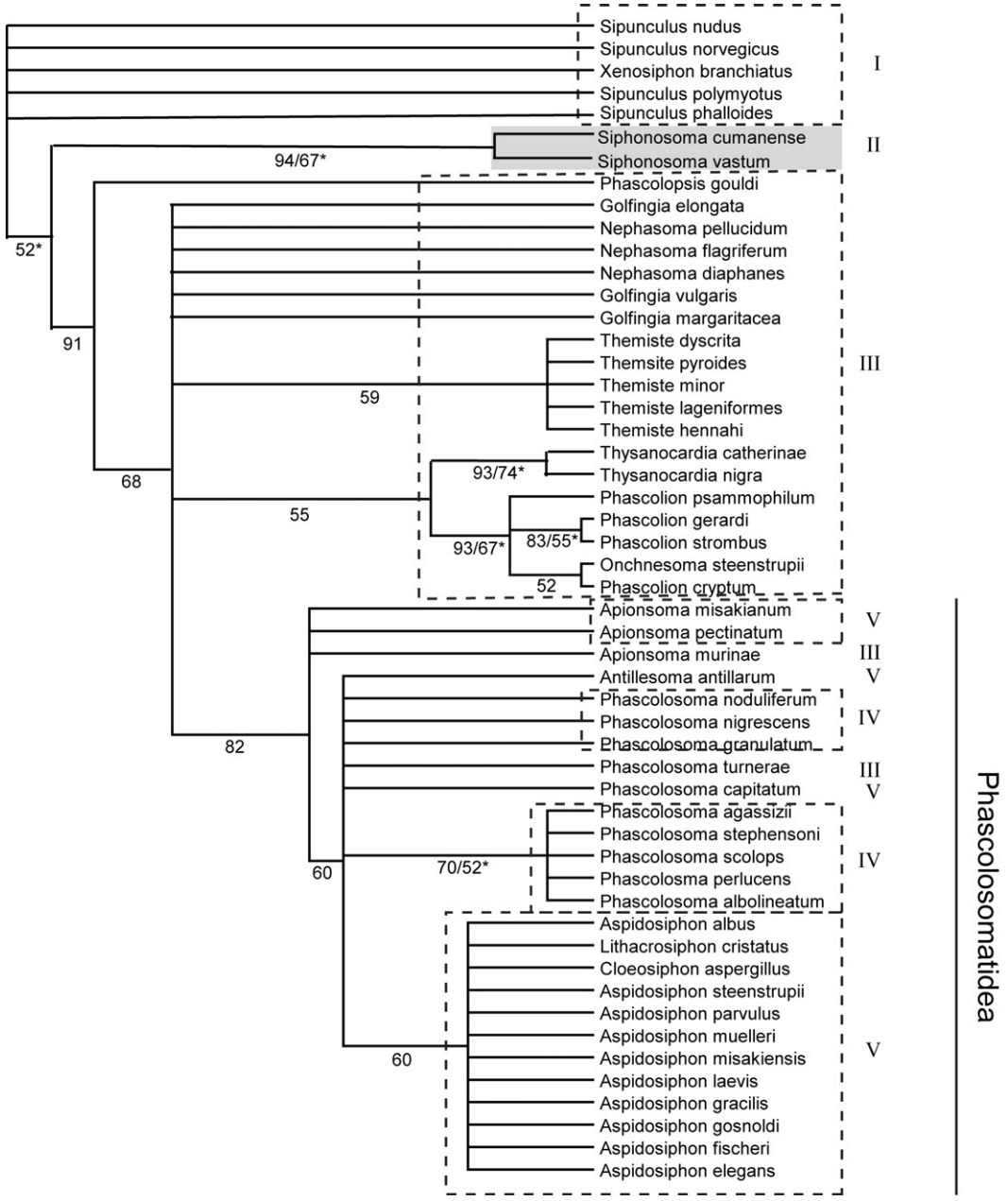


Fig. 3. Unrooted phylogenetic tree based on results of the analyses of morphological data alone. The parsimony and Bayesian analyses resulted in similar trees that only differed in degree of resolution. Branch support values are indicated as Bayesian posterior probabilities and parsimony jackknife percentages (the latter marked with asterisks). Roman numerals correspond to clade designations in Fig. 2. The shaded box corresponds to clade II in the Bayesian analysis. Dashed boxes refer to non-monophyletic groupings that were monophyletic in the Bayesian analysis.

ated *P. capitatum* with *Golfingia* (in the subgenus *Apionsoma*) and with *Apionsoma* (Cutler and Cutler, 1987). Later, assuming that the state of the body wall musculature was homoplastic, it was moved back into *Phascolosoma*, as originally proposed. *Apionsoma pectinatum* and *A. misakianum* belong to two different subgenera, *Edmondsius* and *Apionsoma*, respectively. While they form a weakly supported clade in the Bayesian analysis, there is no support for this clade in the direct optimization analysis. Similar to *Phascolosoma capitatum*, *A. pectinatum* has changed its generic affiliations in the past; it was originally described as a *Phascolosoma* by Keferstein (1867).

The two species *Antillesoma antillarum* (with three representatives) and *Cloeosiphon aspergillus* (with two representatives) clearly fall into clade V in the Bayesian analysis, and form a clade in the direct optimization analysis. The presence of an anal shield has in the past justified the inclusion of *Cloeosiphon* in the within Aspidosiphonidae (Cutler and Gibbs, 1985), although the anal shield is constructed very differently in this genus than in *Aspidosiphon* and *Lithacrosiphon*. It is not restricted to the dorsal side but extends all around the pre-anal trunk and is constructed of relatively large rhomboid plates. Our Bayesian analysis confirms that, despite these morphological differences, *Cloeosiphon* is indeed closely

related to the *Aspidosiphon*/*Lithacrosiphon* clade, however, *Antillesoma antillarum* also falls into this clade. *Antillesoma*, a monotypic genus, seems to be morphologically more similar to *Phascolosoma* than to the Aspidosiphonidae. It is lacking an anal shield, the main defining characteristic of the Aspidosiphonidae. The placement of *Antillesoma* in the Aspidosiphonidae has previously been suggested (Maxmen et al., 2003) and is gaining support with the addition of more data. The anal shield in *Lithacrosiphon* is distinctly cone-shaped but is restricted to the dorsal side as in *Aspidosiphon*. There are no other clear distinctions between *Lithacrosiphon* and *Aspidosiphon* and all our analyses agree that *Lithacrosiphon* is nested within *Aspidosiphon*. Within *Aspidosiphon*, there is no support for the monophyly of the three different subgenera *Akrikos, Aspidosiphon* and *Paraspidosiphon*.

4.2. Species

In our study, only slightly more than half of the species with multiple representatives appeared monophyletic (Table 4). It can be argued that biological species are not necessarily monophyletic. In some cases, species paraphyly may be due to a lack of resolution (e.g. *Themiste minor*, *Aspidosiphon steenstrupii*, and *Phascolosoma perlucens*). Species may also appear paraphyletic due to incomplete lineage sorting. This is to be expected when conserved genes are used to resolve relationships among very closely related species. However, of the non-monophyletic species, at least three (*Aspidosiphon elegans*, *A. parvulus*, and *P. scolops*) are clearly polyphyletic (see Figs. 1 and 2). None of the partitions or the combined analysis supports their monophyly. In a fourth polyphyletic species, *P. granulatum*, we cannot exclude a faulty identification in one of the two specimens, since the 18S rRNA sequence X79874 was downloaded from GenBank and the identification cannot be verified. All represented specimens from the other four species, however, are from our own collections.

Cryptic species are common in marine organisms (for reviews see Knowlton, 1993, 2000) and there is evidence for cryptic speciation in the sipunculan *Apionsoma misakianum* (Staton and Rice, 1999). However, it is unusual to find two morphologically cryptic species that are not sister species, as is the case in several of the sipunculan samples examined. All three polyphyletic species are geographically widespread and inhabit shallow-water hard substrates, mostly coral rubble. In *Aspidosiphon parvulus*, the two specimens from the Gulf of Mexico fall into a separate clade from the specimen from Belize. It is possible that the specimen from Belize represents *Paraspidosiphon spinososcutatus*, a species known from several Caribbean locations that Cutler (1973) synonymized with *Aspidosiphon parvulus*, due to the species' very similar morphology and ecological habits. *Paraspidosiphon* is not a valid genus any longer; it has been reduced to a subgenus of *Aspidosiphon* and includes all *Aspidosphon* species in which the longitudinal body wall musculature splits into bands. *A. parvulus* was originally described as having smooth body wall musculature but at closer inspection, Cutler (1973) detected longitudinal muscle bands in the type specimens and therefore placed the species in *Paraspidosiphon*. Other characters that originally distinguished the two species are introvert length, hook morphology and the pattern of furrows in the anal shield. According to Cutler (1973) all of these characters show too much variation within populations or even individuals—e.g. both single-pointed and double-pointed hooks occur in the same specimens—to warrant the status as separate species. However, in light of the molecular evidence, it would be worthwhile to re-evaluate the validity of *Aspidosiphon* (*Paraspidosiphon*) *spinososcutatus*.

Aspidosiphon elegans is another species with many junior synonyms. Cutler (1994) described a remarkable range in intraspecific morphological variation, but we cannot exclude the possibility that some of this variation actually represents interspecific differences. One of its junior synonyms is *Aspidosiphon brocki*, the only example of an asexually reproducing sipunculan.

The morphology of the anal shield and introvert hooks are the most important taxonomic characters in *Aspidosiphon*. The anal shield seems to function primarily as an operculum to seal the burrow after the introvert has been retracted into the trunk. The hooks are probably used to scrape epifauna off the rock surface when foraging. As the functions of hooks and anal shields seem to be the same across the genus, morphological convergence may be more common among *Aspidosiphon* species than originally thought. Another possible explanation for morphological convergence may be hybridization. There is currently no evidence for hybridization in sipunculans, but the possibility cannot be ruled out.

Similar to the polyphyletic *Aspidosiphon* species, *Phascolosoma scolops* has a number of junior synonyms. In our analyses, three of the four *P. scolops* specimens, from the Red Sea and the Indian Ocean, group relatively closely together in a poorly resolved clade (unresolved in the direct optimization analysis) with *P. perlucens* and *P. albolineatum*, but the fourth specimen from Hawaii clearly falls into a separate clade, together with *P. stephensoni* and *P. albolineatum*.

Given that most of the non-monophyletic species and all of the polyphyletic species fall into *Aspidosiphon* and *Phascolosoma*, we are currently focusing on the taxonomy of these two genera, using detailed morphological as well as molecular techniques.

4.3. Methodology

Generally, alignment precedes tree-building as a separate step in the analysis of sequence data. Direct optimization, as implemented in POY can avoid some of the problems associated with the analysis of sequences of unequal length by combining the alignment and tree building steps in a single, dynamic process. Alignment and tree building are performed under the same analytical parameters, and both are optimized according to the same optimality criterion. In our analyses, the optimality criterion was parsimony, although optimization in a maximum likelihood or Bayesian frame-

work is also possible (Fleissner et al., 2005; Redelings and Suchard, 2005; Wheeler, 2006). POY also allowed us to perform a sensitivity analysis to test the stability of clades under different parameters. On the other hand, when data are combined, POY does not allow for implementation of mixed models for the different partitions, except between molecular and morphological data. As its goal is to minimize overall tree length (under parsimony as the optimality criterion), all data are simultaneously submitted to an analysis under the overall optimal parameter set. However, it does not account for the possibility that different gene regions might evolve according to different models. This was therefore addressed in the Bayesian analysis under mixed models.

A doublet model (Schöninger and von Haeseler, 1994) accounts for correlated substitutions in the complementary strands of the stem regions in ribosomal molecules. Telford et al. (2005) used a novel permutation approach on bilaterian 18S rRNA and found that the likelihood of their trees improved significantly when the nucleotides in the stem regions were correctly paired and allowed to evolve under a doublet model as compared to the permutated data sets with unpaired nucleotides and a single model for the complete molecule. Their analysis under mixed models was clearly superior to simpler analyses that did not take into account the secondary structure. Similarly, the likelihoods for our 18S rRNA trees also significantly improved when mixed models for stem and loop regions were implemented. We were only able to apply the doublet model to the 18S rRNA stem regions. Even better resolution might be achieved if the full 28S rRNA sequences (and possibly mitochondrial ribosomal sequences) were sequenced and models of secondary structure could be applied for those. Bayesian analysis can overestimate branch support (e.g., Simmons et al., 2004); however, almost all of our basal nodes in the Bayesian analysis have maximum branch support and long branches separate most of our major clades.

4.4. Conclusions

Our analysis covers almost the entire diversity within the phylum and is by far the most comprehensive analysis of sipunculan phylogeny published to date. Yet some questions remain open, such as the phylogenetic affinities of some "stragglers", i.e. species that appear in different parts of the tree depending on the method or parameters of analysis.

Both our analyses strongly support the monophyly of Sipuncula and most of the same clades within Sipuncula. In particular, both agree that the *Sipunculus*/*Xenosiphon* clade is the sister group to all other sipunculans. This had been suggested in our previous analyses (Maxmen et al., 2003; Schulze et al., 2005) and has now gained further corroboration. Maxmen et al. (2003) also showed that the rooting between *Sipunculus* and the remaining sipunculans is not dependent on outgroup choice.

Direct optimization with POY has been a useful approach to handle the ribosomal sequences of unequal length and generating conservative estimates of phylogeny. The Bayesian analysis under mixed models provides high resolution with maximum branch support in the deep nodes of the tree.

Our study complements previous taxonomic work on sipunculans, such as Cutler's (1994) monograph and his previous generic revisions. Morphological data alone had reached their limitations due to the simple morphologies of these ancient worms. Molecular analyses have added a new perspective to sipunculan phylogeny. However, our study also pinpoints some taxonomic problems, especially in the genera *Phascolosoma* and *Aspidosiphon*. We are currently addressing these issues using detailed morphological and molecular methods.

Acknowledgments

We thank all individuals who contributed to the collection of specimens, as listed in Appendix A, without whom this study would have been impossible. We are especially grateful to Harlan Dean and Iñaki Saiz Salinas who have helped us in many ways through the years. This project was funded through a MarCraig Grant at Harvard University to G.G. and E.B.C., which included postdoctoral support to A.S. It was continued under a postdoctoral fellowship to A.S. at the Smithsonian Marine Station at Fort Pierce, FL (Contribution No. 658). Additional funding was provided by the Caribbean Coral Reef Foundation (Contribution No. 758) and by internal funds from Harvard University and the Museum of Comparative Zoology to G.G. We owe thanks to the captain and the crew of the R/V Oceanus and chief scientist K. Halanych (NSF EAR-0120646). We gratefully acknowledge Mary E. Rice, Cheryl Hayashi, Christoph Bleidorn and an anonymous reviewer for comments and suggestions that greatly improved this manuscript.

Appendix A. 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):** DNA100390—Phuket, Thailand, Jan. 31, 2001 (J. Hylleberg); DNA100759—Six Men's Bay, Barbados, June 27, 2002 (A. Schulze, J. I. Saiz-Salinas, E. B. Cutler); DNA101008—Bessie Cove, South Hutchinson Island, FL, USA, March 20, 2003 (A. Schulze).

***Apionsoma (Apionsoma) misakianum* (Ikeda, 1904):** DNA100231—Pickles Reef, Key Largo, USA, Nov. 27, 1993 (S. Taylor); DNA100737—Eilat, Israel, Sept, 30, 2000 (N. Ben-Eliahu).

***Apionsoma (Apionsoma) murinae*(E. Cutler, 1969):** DNA100446—Meteor station Me48/1#AT339, Antarctica, Nov. 30, 1999 (J.I. Saiz Salinas).

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

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

hardt); DNA101336—Capron Shoals, FL, USA, date unknown (A. Schulze, W. Lee, H., H. Reichardt).

***Aspidosiphon (Aspidosiphon) elegans* (Chamisso & Eysenhardt, 1821):** DNA100977—Carrie Bow Cay, Belize, April 17, 2003 (A. Schulze, M. E. Rice); DNA101016—R/V Sunburst cruise 520, Capron Shoals, FL, USA, March 11, 2003 (A. Schulze).

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

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

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

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

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

***Aspidosiphon (Paraspidosiphon) laevis* de Quatrefages, 1865:** DNA100467—Hungary Bay, Bermuda, Aug. 9, 2001 (T. Nishikawa); DNA100992—Twin Cays, Belize, April 20, 2003 (M. E. Rice, A. Schulze).

***Aspidosiphon (Paraspidosiphon) parvulus* Gerould, 1913:** DNA100202—unspecified locality, purchased from Gulf Specimens Co. [this specimen corresponds to the sequence erroneously published as *Themiste alutacea* by Giribet et al. (2000)]; DNA100375—unspecified locality, purchased from Gulf Specimens Co.; DNA100982—Twin Cays, Belize, April 20, 2003 (M. E. Rice, A. Schulze).

***Aspidosiphon (Paraspidosiphon) steenstrupii* Diesing, 1859:** DNA100232—Pickles Reef, Key Largo, USA, Nov. 27, 1993 (S. Taylor); DNA100372—Kewalo Reef, Honolulu, USA, Jan. 25, 2001 (J. Brock); DNA100391—Phuket, Thailand, Jan. 31, 2001 (J. Hylleberg); DNA100630—Bank Reef, Barbados, June 26, 2002 (J. I. Saiz Salinas, A. Schulze).

***Cloeosiphon aspergillus* (de Quatrefages, 1865):** DNA100393—Phuket, Thailand, Jan. 21, 2001 (J. Hylleberg); DNA100825—Perrier's Rock, South Africa, Oct. 18, 2002 (R. Biseswar).

***Golfingia elongata* (Keferstein, 1862):** DNA100465—SW of Trunk Is., Harrington Sound, Bermuda, Aug. 8, 2001 (T. Nishikawa); DNA100466—South coast of Stock's Harbor, St. Davis Island, Bermuda, Aug. 7, 2001 (T. Nishikawa); DNA101003—Twin Cays, Belize, April 20, 2003 (M. E. Rice, A. Schulze); DNA101066—R/V Oceanus, Southern New England, USA, 40°27.299′N, 69°54.601′W, June 11, 2003 (A. Schulze); 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):** DNA100738—Kongsfjord Svalbard, Norway, June 23, 2002 (D. Hughes).

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

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

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

***Nephasoma flagriferum* (Selenka, 1885):** DNA100439—Meteor Station Me48/1#345/7, Antarctica, Nov. 11, 1999 (J.I. Saiz Salinas); DNA100440—Meteor Station Me48/1#349, Antarctica, Nov. 11, 1999 (J.I. Saiz Salinas); DNA101071—R/V Oceanus, Southern New England, USA, 39°47.230′N, 70°46.295′W, June 14, 2003 (A. Schulze).

***Nephasoma pellucidum* (Keferstein, 1865):** DNA101009—R/V Sunburst, cruise 526, 4 miles off Ft. Pierce, FL, USA, March 28, 2003 (A. Schulze).

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

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

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

***Phascolion (Phascolion) psammophilum* Rice, 1993:** DNA101006—R/V Sunburst cruise 523, Capron Shoals, FL, March 18, 2003 (A. Schulze).

***Phascolion (Phascolion) strombus* (Montagu, 1804):** DNA100101—Banyuls sur Mer, France, July 20, 2000 (G. Giribet); DNA100739—Kristineberg Marine Biological Station, Fiskebäckskil, Sweden, Dec. 31, 1997 (A. Okusu); 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):** DNA100199—Woods Hole, USA, Sept. 30, 1997 (Marine Biological Laboratory).

***Phascolosoma (Fisherana) capitatum* (Gerould, 1913):** DNA101070—R/V Oceanus, Southern New England, USA, 39°47.230′N, 70°46.295′W, June 14, 2003 (A. Schulze).

***Phascolosoma (Phascolosoma) agassizii* Keferstein, 1866:** DNA101096—Cape Arago (North Cove), OR, USA, Aug. 28, 2003 (M. E. Rice, S. Rumrill, C. Young).

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

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

***Phascolosoma (Phascolosoma) nigrescens* (Keferstein, 1865):** DNA100621—Six Mens Bay, Barbados, June 27, 2002 (A. Schulze, J.I. Saiz Salinas); DNA100736—Eilat, Israel, Sept. 30, 2002 (N. Ben-Eliahu); DNA100822—Per-

rier's Rock, South Africa, Oct. 18, 2002 (R. Biseswar); DNA101010—Bessie Cove, FL, USA, March 20, 2003 (A. Schulze); DNA101082—Broome, Australia, date unknown (G. Rouse).

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

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

***Phascolosoma (Phascolosoma) scolops* (Selenka & de Man, 1883):** DNA100373—Kewalo Reef, Honolulu, HI, USA, Jan. 25, 2001 (J. Brock); DNA100394—Phuket, Thailand, July 31, 2001 (J. Hylleberg); DNA100735—Eilat, Israel, Sept. 30, 2002 (N. Ben-Eliahu); DNA100813—Perrier's Rock, South Africa, Oct. 18, 2002 (R. Biseswar).

***Phascolosoma (Phascolosoma) stephensoni* (Stephen, 1942):** DNA100203—Cova Blava, Cabrera, Balearic Islands, Spain, May 31, 1997 (X. Turon); DNA100209—Nielsen Park Shore, Port Jackson, Sydney Harbor, NSW, Australia, April 12, 2000 (G. Giribet, P. Hutchings); DNA100469—Baileys Bay, Hamilton Island, Bermuda, Aug. 7, 2001 (E.B. Cutler); DNA100485—Terceira, Azores, Portugal, Oct. 31, 2001 (P. Wirtz).

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

***Siphonosoma cumanense* (Keferstein, 1867):** DNA100235—unspecified locality, Puerto Rico June 3, 1993 (J. Staton, H. Reichardt); DNA100464—Baileys Bay, Hamilton, and south coast of St. Davus, Bermuda, Aug. 7, 2001 (T. Nishikawa); DNA100622—Bath, Barbados, June 24, 2002 (A. Schulze, J.I. Saiz Salinas); DNA100991—Twin Cays, Belize, April 24, 2004 (M. E. Rice, A. Schulze).

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

***Sipunculus (Sipunculus) norvegicus* Danielssen, 1869:** 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:** DNA100234—Station 53F, No Name Cay, July 28, 1993 (J. Staton, H. Reichardt); DNA100245—near Arcachon (fishermen's locality unspecified), Oct. 30, 2000; DNA100246—unspecified locality, Vietnam, Oct. 30, 2000; DNA100468—South coast of Stock's Harbor, St David's Island, Aug. 7, 2001 (E. Cutler); DNA100629—Isla Taboguilla, off Panama City, Panama, June 20, 2002 (T. Nishikawa); DNA100993—Twin Cays, Belize, April 24, 2003 (M. E. Rice, A. Schulze).

***Sipunculus (Sipunculus) phalloides* (Pallas, 1774):** DNA101337—Ponta do Araça, Sao Sebastiao, Brazil, 23°49′02″S, 45°24′19″W (G.Y. Kawauchi).

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

***Themiste (Lagenopsis) lageniformis* (Baird, 1868):** DNA100229—Jack Island oyster beds, Fort Pierce, FL, USA, Oct. 18, 2000 (S. Reed).

***Themiste (Lagenopsis) minor* (Ikeda, 1904):** DNA100210—Nielsen Park Shore, Port Jackson, Sydney Harbor, April 12, 2000 (G. Giribet, P. Hutchings); DNA101083—unspecified locality, South Africa, Sept. 28, 2002 (G. Rouse).

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

***Themiste (Themiste) hennahi* Gray, 1828:** DNA100627—Bahı´a de Concepción, Lirquen Playa sector La Cata, Chile, April 26, 2001 (E. Tarifeño).

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

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

***Thysanocardia nigra* (Ikeda, 1904):** DNA100606—Lopez Island, WA, USA, May 17, 2002 (D. McHugh).

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

Appendix B. Description of morphological characters

The main sources for the following morphological characters were either direct observations or the following publications: Cutler (1994); Rice (1993) and Stephen and Edmonds (1972).

Characters 1-4 (tentacles): A crown of tentacles is usually present at the anterior end of the introvert. In a portion of the species, the tentacles encircle the mouth peripherally and the nuchal organ lies outside of that circle (1). In other species, the tentacles form an arc around the nuchal organ (2). Representatives of *Thysanocardia* are the only species in which both nuchal and peripheral tentacles are present. In *Themiste* species, 4-6 stem-like outgrowths give rise to numerous branches (3). In *Phascolion cryptum* and representatives of the genus *Thysanocardia* tentacles are not restricted to the very tip of the introvert but extend for a short distance along it (4). Character coding - 1. Peripheral tentacles: 0 = absent, 1 = present; 2. Nuchal tentacles: 0 = absent, 1 = present; 3. Branched tentacles: 0 = absent, 1 = present; 4. Tentacles along introvert: 0 = absent, 1 = present.

Charcters 5-8 (Nephridia): Nephridia are either paired or single (5). They usually form simple sacs but are distinctly bilobed in some species (6). They also vary with regard to their position relative to the anus (7) and may be

attached to the body wall (8). Character coding—5. Nephridia: 0 = paired, 1 = single; 6. Nephridial attachment: 0 = mostly unattached, 1 = at least 50% attached; 7. Nephridial shape: 0 = unilobed, 1 = bilobed. 8. Position of nephridiopores: 0 = anterior to anus, 1 = same level as anus, 2 = posterior to anus.

Characters 9-16 (body wall): In several sipunculan genera the coelom extends into the body wall (9). In *Siphonosoma* and *Siphonomecus* (the latter not included in this study) the extensions are sac-like whereas they form canals in *Sipunculus* and *Xenosiphon* (Ruppert and Rice, 1995) (10). The canals either run longitudinally between the longitudinal muscle bands (*Sipunculus*), or diagonally as short, subcutaneous canals (*Xenosiphon*) (11). The body wall consists of an outer layer of circular and an inner layer of longitudinal musculature. Both can either form continuous sheets or a variable number of bands (12, 13, 15). Muscle bands may be distinct for most of their length or anastomosing (14, 16). Character coding—9. Coelomic extensions in body wall: 0 = absent, 1 = present; 10. Type of coelomic extensions: 0 = sacs, 1 = canals; 11. Orientation of coelomic canals: 0 = longitudinal, 1 = in bands; 12. Longitudinal musculature: 0 = distinct, 1 = anastomosing; 13. Number of longitudinal muscle bands: 0 = commonly < 25; 1 = 25-40, 2 = > 40; 14. Longitudinal muscle bands: 0 = distinct, 1 = anastomosing. 15: Circular musculature: 0 = continuous sheet, 1 = in bands; 16. Circular muscle bands: 0 = distinct, 1 = anastomosing.

Characters 17-20 (anal shield): All currently recognized Aspidosiphonidae are characterized by a calcareous or horny protein shield shield at the anterior end of the trunk. However, its chemical composition, extend and morphology are variable among the species (17). The anal shield is cone-shaped in *Lithacrosiphon* and more or less flat in *Aspidosiphon* (18) and may be relatively smooth or bear distinct grooves (19). In both these genera it is restricted to the dorsal side. In *Cloeosiphon* it is composed of numerous calcareous plates that surround the anterior introvert ("pineapple shield") (20). Character coding—17. Dorsal shield: 0 = absent, 1 = present; 18. Shape of dorsal shield: 0 = flat, 1 = cone-shaped; 19. Grooves in dorsal shield: 0 = absent, 1 = present. 20. Pineapple shield: 0 = absent, 1 = present.

Characters 21-25 (spindle muscle): The spindle muscle is a slender, thread-like muscle that runs through the intestinal coil. Anteriorly, it is either attached to the body wall or on the rectum or wing muscle (22, 23). In most cases there is only one anterior point of insertion, except in *Siphonosoma* and two of the *Themiste* species included here where the muscle sends branches anteriorly with several points of insertion (24). Posteriorly, it attaches to the body wall or ends in the gut coil (25). Character coding—21. Spindle muscle: 0 = absent, 1 = present; 22. Anterior attachment of spindle muscle: 0 = body wall, 1 = on rectum. 23. Level of attachment on body wall: 0 = anterior to anus, 1 = same level as anus, 2 = posterior to anus. 24. Anterior roots of spindle muscle: 0 = one, 1 = two or more. 25. Posterior attachment of spindle muscle: 0 = posterior body wall, 1 = within gut coil.

Characters 26-40 (hooks): Other than in polychaetes, the hooks in sipunculans are non-chitinous epidermal structures (Andreae, 1882; Andrews, 1890; Voss-Foucart et al., 1977; Voss-Foucart et al., 1978). In representatives of the Sipunculidea, hooks, if present, are generally scattered, simple and only slightly curved. In the Phascolosomatidea hooks are usually arranged in rings, are sharply curved posteriorly and show a distinct internal anatomy when viewed with light microscopy. The number of rings of hooks varies greatly and can exceed 100. Bidentate hooks (30) are often found in the Aspidosiphonidae and are distinct from hooks in *Phascolosoma* (31). In both cases, there is a primary tooth and a secondary tooth, but whereas the secondary tooth in *Aspidosiphon* hooks is usually pointed, close to the primary tooth, similar in shape and only slightly smaller, in *Phascolosoma* it is closer to the base of the hook and always more blunt (somewhat pointed in *P. stephensoni* but still with broad base). The internal anatomy of *Phascolosoma* hooks may include an anterior clear triangle (33), a clear streak (34) and a posterior, crescent-shaped space (35). *Phascolosoma* hooks are also characterized by posterior basal structures (36) that can take the shape of toes, rootlets or warts (37). Warts are only found in *P. glabrum*, not included in this study. The angle of the hook relative to the body axis is usually greater than 90°, except in two *Phascolion* species, *Cloeosiphon aspergillus* and *Apionsoma misakianum* where it forms a smaller angle (38).

In Aspidosiphonidae, several zones of hooks can be distinguished along the introvert. In the distal zone, the hooks are laterally compressed and usually arranged in rings. In many cases, the anterior rings consist of bidentate hooks, followed by a zone of rings with unidentate hooks. Behind the hooks arranged in rings, there is often a zone of scattered hooks. These may include hooks that are similar in shape to the hooks in the rings, or pyramidal (39) and conical (40) hooks. Pyramidal hooks can be distinguished from laterally compressed hooks by their triangular base but intermediate forms between the two may be present. Conical hooks have a nearly circular cross section and are only found in the dorsal region of the introvert in *A. elegans* (Cutler and Cutler, 1989).

Character coding—26. Hooks on introvert in adults: 0 = absent, 1 = present; 27. Hooks in rings: 0 = absent, 1 = present; 28. Number of rings of hooks: 0 = < 50, 1 = > 50. 29. Scattered hooks: 0 = absent, 1 = present. 30. Bidentate hooks: 0 = absent, 1 = present. 31. Secondary tooth on hook: 0 = absent, 1 = present; 32. Shape of secondary tooth: 0 = blunt, 1 = pointed; 33. Anterior clear triangle: 0 = absent, 1 = present; 34. Clear streak: 0 = absent, 1 = present; 35. Crescent-shaped space: 0 = absent, 1 = present; 36. Posterior basal structures: 0 = absent, 1 = present; 37. Type of posterior basal structures: 0 = warts, 1 = rootlets; 38. Angle of hook: 0 = $< 90\%$, 1 = $> 90\%$. 39. Pyramidal hooks: 0 = absent, 1 = present. 40. Conical hooks: 0 = absent, 1 = present.

Character 41 (anus location): The anus is usually located dorsally at the anterior trunk. However, in two of the species included here, *Phascolion gerardi* and *Onchnesoma steenstrupii* it is shifted anteriorly onto the introvert (Rice, 1993; Shipley, 1892). Character coding—41. Location of anus: 0 = on anterior trunk, 1 = on introvert.

Character 42 (pigmented introvert bands): In *Phascolosoma* species, the dorsal side of the introvert is usually darker than the ventral side and the pigment is often distributed in distinct broad stripes. The presence or absence of these pigment bands is constant within species (Cutler and Cutler, 1990). Character coding—42. Pigmented introvert bands: 0 = absent, 1 = present.

Characters 43, 44 (contractile vessel villi): The contractile vessel is part of the tentacular coelomic system. It runs dorsally along the esophagus and has a coelomic lining. It contains hemocytes and is considered an analogue to a blood vascular system (Pilger, 1982). In some species, in particular of the genus *Themiste*, villi are present along the length of the vessel (Rice, 1993) (43). These vary in number and length: whereas they are relatively short in *Thysanocardia* species, *Antillesoma antillarum*, *Siphonosoma cumanense* and *Themiste* (*Lageniformis*) the villi are short and digitiform, they are long thread-like and fewer in number in *Themiste* (*Themiste*) (Edmonds, 1980). Character coding—43. Contractile vessel villi: 0 = absent, 1 = 1 present; 44. Type of contractile vessel villi: 0 = digitiform, 1 = elongate tubules.

Characters 45-51 (introvert retractor muscles): This set of strong muscles insert anteriorly near the brain and are posteriorly attached to the body wall. The number of introvert retractors varies from a single column to four (45, 46) although all pelagosphera larvae or early juveniles seem to have four retractor muscles (A.S. pers. obs.). If four retractors are present in the adult, they are arranged in a dorsal and a ventral pair. The pairs are fused to various degrees (47) but separate origins on the body wall are usually discernable. If only one pair of retractors is present the muscles originate on the ventral side and are here regarded as ventral retractors. However, this needs further investigation as they might actually be the dorsal retractors that have shifted ventrally or a product of fusion between dorsal and ventral retractors. The point of origin of the muscles along the body wall also varies among species (48, 49). In some cases (*Antillesoma antillarum* and *Phascolosoma scolops*), there is evidence of fusion between the dorsal and the ventral retractor muscle on each side but this varies intraspecifically and is not included in this data set. The points of origin of the retractor muscles along the body wall vary greatly and can be in the anterior third of the trunk (e.g. *Sipunculus*), in the middle third (e.g. most *Phascolosoma* species) or near the posterior end (e.g. *Onchnesoma*). Character coding – 45. Introvert retractor muscles in adult: 0 = two pairs, 1 = less than two pairs; 46. Fusion of dorsal retractors: 0 = not fused (< 10%), 1 = partially fused, 2 = completely fused; 47. Fusion in ventral retractors: 0 = not fused (< 10%), 1 = partially fused, 2 = completely fused; 48. Fusion of dorsal and ventral retractors: 1 = absent, 2 = present; 49: Retractor column: 0 = absent, 1 = present; 50. Origin of dorsal retractor muscles: 0 = anterior 1/3 of body, 1 = middle 1/3 of body, 2 = posterior 1/3 of body; 51: Origin of ventral retractor muscles: 0 = anterior 1/3 of body, 2 = middle 1/3 of body; 3 = posterior 1/3 of body.

Character 52 (protractor muscle): In *Xenosiphon*, a pair of short muscles connects the introvert near the brain with the anterior body wall at the level of the anus. Character coding—52. Protractor muscle: 0 = absent, 1 = present.

Character 53 (ratio of introvert/trunk length): The ratio between the fully extended introvert and the trunk length is here roughly divided into three categories. Although some intraspecific variation is often observed, most species clearly fall into one of the categories. Character coding—53. Ratio of introvert/trunk length: 0 = < 0.75, 1 = 0.75-2, 2 = > 2.

Character 54 (holdfast papillae with hardened borders): These are specialized papillae found in *Phascolion* species that inhabit abandoned shells of gastropods, scaphopods or foraminiferans. The papillae are usually located in the posterior or mid-trunk region and have sclerotinized borders. The borders may surround the anterior margin of a round papilla (e.g. *P. hedraeum*), be U-shaped or V-shaped (e.g. *Phascolion cryptum*), or form pointed projections (*P. caupo*). Character coding—54. Holdfast papillae with hardened borders: 0 = absent, 1 = present.

Character 55 (caudal appendage): This tail-like projection at the posterior end of the trunk occurs in *Nephasoma flagriferum* and two *Golfingia* species. The latter are not included in this dataset, so that the character is here an autapomorphy for *Nephasoma flagriferum*. Character coding—55. Caudal appendage: 0 = absent, 1 = present.

Characters 56-58 (nuchal organ): Nuchal organs are probably chemosensory organs known also in annelids. In sipunculans, nuchal organs are located close to the tentacles at the anterior introvert in sipunculans (Åkesson, 1958; Rice, 1993). The nuchal organs of only one species have been examined with ultrastructural methods (Purschke, 1997). Åkesson (1958) originally concluded that nuchal organs in annelids and sipunculans are homologous. Later authors came to the opposite conclusion (Purschke, 1997, 2002) but suggested that more studies on sipunculans be undertaken to answer the question with more certainty. Three main designs are here distinguished in sipunculans (56): simple pits as in *Sipunculus* species, ciliated bands as in *Phascolion cryptum* and *P. gerardi* or patches (all other species where known). The number of patches varies between one and two (57). If only a single patch is present, it might be triangular to heart-shaped, distinctly bilobed or multilobed (58). Character coding—56. Nuchal organ: 0 = pit, 1 = ciliated band, 2 = patch or patches; 57. Number of nuchal patches: 0 = one, 1 = two; 58. Shape of nuchal patch: 0 = triangular, 1 = bilobed, 2 = multi-lobed.

Appendix C. Morphological data matrix for the 58 morphological characters described in Appendix B. "x" indicates inapplicable characters while "?" indicates missing observations

	1 2 3 4 5 6 7	8 9 10 11 12 13	14 15 16 17 18 19 20 21 22	23	24 25 26 27 28 29 30 31	32 33 34 35 36 37 38	39 40 41 42 43 44 45 46 47	48 49 50	51	52 53	54 55 56 57 58
Antillesoma antillarum	0 1 0 0 0 1 0	2 0 x x 1 0	1 0 x 0 x 0 x x 0 1 0	1	0 0 0 x x x x x	x x x x x x x	0 0 0 0 1 0 0 1 1	1 0 1	1	0 0	0 0 2 0 1
Apionsoma (A.) misakianum	0 1 0 0 0 0 1	0 0 x x 0 x	x 0 x 0 x 0 x x 0 1 0	0	0 0 1 1 0 0 0 0 0	x 0 0 0 0 0 x 1	0 0 0 0 0 0 x 0 0 0	0 0 1	0	0 2	0 0 ? ? ?
Apionsoma (A.) murinae	0 1 0 0 0 0 0 0,1	1 0 x x 0 x	x 0 x 0 x 0 x x 0 1 0	0	0 0 1 1 0 0 0 0 0	x 0 0 0 0 0 x 0	0 0 0 0 0 0 x 0 0 0	0 0 1,2	1,2	0 2	0 0 ? ? ?
Apionsoma (E.) pectinatum	0 1 0 0 0 1 1	1 0 x x 1 0,1	1 0 x 0 x 0 x x 0 1 0	0	0 1 1 1 0 0 0 0 0	x 0 0 0 0 0 x 0	0 0 0 0 0 0 x 0 0 0	0 0 0	0	0 2	0 0 2 0 1
Aspidosiphon (A.) albus	0 0 0 0 0 1 0	1 0 x x 0 x	x 0 x 1 0 0 0 1 0	?	? 0 0 x x x x x x	x x x x x x x	0 0 0 0 0 0 x 1 x 2	x 0 x	2	0 2	0 0 2 0 ?
Aspidosiphon (A.) elegans	0 1 0 0 0 1 0	2 0 x x 0 x	x 0 x 1 0 0 0 1 0	2	0 0 1 1 1 1 1 0	x 0 0 0 0 0 x 0	0 1 0 0 0 0 x 1 x 1	x 0 x	2	0 2	0 0 2 0 0
Aspidosiphon (A.) gosnoldi	0 1 0 0 0 1 0	? 0 x x 0 x	x 0 x 1 0 0 0 1 0	0	0 0 1 1 1 1 1 0	x 0 0 0 0 0 x 0	1 0 0 0 0 0 x 1 x 1	x 0 x	2	0 1,2	0 0 2 0 ?
Aspidosiphon (A.) gracilis	0 1 0 0 0 1 0	? 0 x x 0 x	x 0 x 1 0 0 0 1 0	?	? 0 1 1 1 1 0 0	x 0 0 0 0 0 x 0	1 0 0 0 0 0 x 1 x 1	x 0 x	2	0 1	0 0 2 0 ?
Aspidosiphon (A.) misakiensis	0 1 0 0 0 1 0	2 0 x x 0 x	x 0 x 1 0 0 0 1 0	0	0 0 1 1 1 1 1 0	x 0 0 0 0 0 x 0	0 0 0 0 0 0 x 1 x 1	x 0 x	2	0 1	0 0 2 0 ?
Aspidosiphon (A.) muelleri	0 1 0 0 0 1 0	1 0 x x 0 x	x 0 x 1 0 1 0 1 0	0	0 0 1 1 1 1 1 0	x 0 0 0 0 0 x 0	1 0 0 0 0 0 x 1 x 1	x 0 x	2	0 1	0 0 2 0 0
Aspidosiphon (P.) fischeri	0 1 0 0 0 1 0	0 0 x x 1 0	1 0 x 1 0 0 0 1 0	0	0 0 1 1 1 1 1 0	x 0 0 0 0 0 x 0	1 0 0 0 0 0 x 1 x 2	x 0 x	2	0 1	0 0 2 0 0
Aspidosiphon (P.) laevis	0 1 0 0 0 1 0	? 0 x x 1 1,2	1 0 x 1 0 1 0 1	0,1 0,2	1 0 1 1 1 1 0 0	x 0 0 0 0 0 x 0	0 0 0 0 0 0 x 1 x 2	x 0 x	2	0 1	0 0 2 0 0
Aspidosiphon (P.) parvulus	0 1 0 0 0 1 0	? 0 x x 1 0	1 0 x 1 0 0 0 1 0	0	0 0 1 1 1 1 1 0	x 0 0 0 0 0 x 0	1 0 0 0 0 0 x 1 x 2	x 0 x	2	0 1	0 0 2 0 ?
Aspidosiphon (P.) steenstrupii	0 1 0 0 0 1 0	1 0 x x 1 0	1 0 x 1 0 0 0 1 0	0	0 0 1 1 1 1 1 0	x 0 0 0 0 0 x 0	1 0 0 0 0 0 x 1 x 2	x 0 x	2	0 1	0 0 2 0 0
Cloeosiphon aspergillus	0 1 0 0 0 1 0	1 0 x x 0 x	x 0 x 0 x x 1 1 1	x	x 0 1 1 1 0 1 0	x 0 0 0 0 0 x 1	0 0 0 0 0 0 x 1 x 1	x 0 x	2	0 1	0 0 2 0 ?
Golfingia elongata	1 0 0 0 0 0 0	1 0 x x 0 x	x 0 x 0 x x 0 1 1	x	x 1 1 0 x 0 0 0	x 0 0 0 0 0 x 0	0 0 0 0 0 0 x 0 0 0,1	0 0 0	1	0 0	0 0 2 0 ?
Golfingia margaritacea	1 0 0 0 0 0 0	1 0 x x 0 x	x 0 x 0 x x 0 1 1	x	x 1 0 x x x x x	x x x x x x x	0 0 0 0 0 0 x 0 0 0	0 0 0	1	0 0	0 0 2 1 0
Golfingia vulgaris	1 0 0 0 0 0 0	0 0 x x 0 x	x 0 x 0 x x 0 1 1	x	x 1 1 0 x 0 0 0	x 0 0 0 0 0 x 0	0 0 0 0 0 0 x 0 0 0	0 0 0	1	0 0,1	0 0 2 1 0
Lithacrosiphon cristatus	0 1 0 0 0 0 0	2 0 x x 1 0	1 0 x 1 1 1 0 1 0	0	0 0 1 1 1 0 1 0	x 0 0 0 0 0 x 0	0 0 0 0 0 0 x 1 x 1	x 0 x	2	0 1	0 0 2 0 ?
Nephasoma diaphanes	1 0 0 0 0 0 0	1 0 x x 0 x	x 0 x 0 x x 0 1 ?	?	? 1 1 0 x 1 0 0	x 0 0 0 0 0 x 0	0 0 0 0 0 0 x 1 x 1	x 0 x	1,2	0 0	0 0 ? ? ?
Nephasoma flagriferum	1 0 0 0 0 0 0	1 0 x x 0 x	x 0 x 0 x x 0 1 0	0	0 1 0 x x x x x	x x x x x x x	0 0 0 0 0 0 x 1 x 1	x 0 x	1	0 1	0 1 2 ? ?
Nephasoma pellucidum	1 0 0 0 0 0 0	1 0 x x 0 x	x 0 x 0 x x 0 1 1	x	x 1 1 0 x 1 0 0	x 0 0 0 0 0 x 0	0 0 0 0 0 0 x 1 x 1	x 0 x	1	0 0	0 0 2 1 0
Onchnesoma steenstrupii	0 0 0 0 1 0 0	2 0 x x 0 x	x 0 x 0 x x 0 0 x	x	x x x 0 x x x x x	x x x x x x x	0 0 1 0 0 x 1 x 2	x 1 x	2	0 2	0 0 2 0 0
Phascolion (I.) gerardi	1 0 0 0 1 1 0	2 0 x x 0 x	x 0 x 0 x x 0 0 x	x	x x x 1 0 x 1 0 0	x 0 0 0 0 0 x 0	0 0 1 0 0 x 0 2 1	0 0 2	2	0 1	0 0 1 x x
Phascolion (L.) cryptum	1 0 0 1 1 0 0	2 0 x x 0 x	x 0 x 0 x x 0 0 x	x	x x x 0 x x x x x	x x x x x x x	0 0 0 0 0 0 x 1 x 2	x 1 x	2	0 1	1 0 1 x x
Phascolion (P.) psammophilum	1 0 0 0 1 0 0	2 0 x x 0 x	x 0 x 0 x x 0 0 x	x	x x x 0 x x x x x	x x x x 0 x 0	0 0 0 0 0 0 x 0 2 2	0 0 2	2	0 1	0 0 2 0 0
Phascolion (P.) strombus	1 0 0 0 1 1 0	2 0 x x 0 x	x 0 x 0 x x 0 0 x	x	x x x 1 0 x 1 0 0	x 0 0 0 0 0 x 0	0 0 0 0 0 0 x 0 2 2	0 0 2	2	0 1	1 0 2 1 0
Phascolopsis gouldi	1 0 0 0 0 0 0	2 0 x x 1 1	1 0 x 0 x x 0 1 0	0	0 1 0 x x x x x	x x x x x x x	0 0 0 0 0 0 x 0 0 0	0 0 0	1	0 0	0 0 2 ? ?

(*continued on next page*)

	1 2 3 4 5 6 7	8 9 10 11 12 13	14 15 16 17 18 19 20 21 22	23 24 25 26 27 28 29 30 31	32 33 34 35 36 37 38	39 40 41 42 43 44 45 46 47	48 49 50	51	52 53	54 55 56 57 58
Phascolosoma (F.) capitatum	0 1 0 0 0 1 0	1 0 x x 0 x	x 0 x 0 x x 0 1 0	0 0 0 1 1 0 0 0 0	x 0 1 0 0 x 0	0 0 0 0 0 0 x 0 0 1	1 0 1,2	2	0 1	0 0 2 0 1
Phascolosoma (P.) agassizii	0 1 0 0 0 1 0	1 0 x x 1 0,1	1 0 x 0 x x 0 1 0	0 0 0 1 1 0 0 0 0,1	0 1 1 0 1 0 0	0 0 0 0 1 0 x 0 0 0	0 0 1,2	1	0 1	0 0 2 0 1
Phascolosoma (P.) albolineatum	0 1 0 0 0 1 0	1 0 x x 1 1	1 0 x 0 x x 0 1 0	0 0 0 1 1 0 0 0 1	0 1 1 0 1 0 1	0 0 0 0 1 0 x 0 0 0	0 0 1	1	0 1	0 0 2 0 1
Phascolosoma (P.) granulatum	0 1 0 0 0 1 0	1 0 x x 1 0	1 0 x 0 x x 0 1 0	0 0 0 1 1 1 0 0 0,1	0 0 1 0 1 0 0	0 0 0 0 0 0 x 0 0 0	1 0 1	1	0 1	0 0 2 0 1
Phascolosoma (P.) nigrescens	0 1 0 0 0 1 0	1 0 x x 1 1	1 0 x 0 x x 0 1 0	0 0 0 1 1 1 0 0 0,1	0 0 1 0 1 0 0	0 0 0 0 1 0 x 0 0 0	0 0 1	1	0 1	0 0 2 ? ?
Phascolosoma (P.) noduliferum	0 1 0 0 0 1 0	1 0 x x 1 0	1 0 x 0 x x 0 1 0	0 0 0 1 1 1 0 0 0	x 0 1 0 1 0 0	0 0 0 0 1 0 x 0 0 0	0 0 1	1	0 1	0 0 2 ? ?
Phascolosoma (P.) perlucens	0 1 0 0 0 1 0	1 0 x x 1 0	1 0 x 0 x x 0 1 0	0 0 0 1 1 0 0 0 1	0 1 1 0 1 0 0	0 0 0 0 1 0 x 0 0 0	0 0 1	1	0 1	0 0 2 0 1
Phascolosoma (P.) scolops	0 1 0 0 0 1 0	1 0 x x 1 0	1 0 x 0 x x 0 1 0	0 0 0 1 1 0 0 0 0,1	0 1 1 0 1 0 0	0 0 0 0 1 0 x 0 0 0	1 0 1	1	0 1	0 0 2 ? ?
Phascolosoma (P.) stephensoni	0 1 0 0 0 1 0	1 0 x x 1 1	1 0 x 0 x x 0 1 0	0 0 0 1 1 1 0 0 1	1 1 1 1 1 0 0	0 0 0 0 1 0 x 0 0 0	0 0 1	1	0 1	0 0 2 0 ?
Phascolosoma (P.) tumerae	0 1 0 0 0 1 0	1 0 x x 1 1	1 0 x 0 x x 0 1 0	0 0 0 1 1 1 0 0 0	x 0 0 0 1 1 0,1	0 0 0 0 0 0 x 0 0 0	0 0 1	2	0 1	0 0 2 0 1
Siphonosoma cumanense	1 0 0 0 0 0 0	2 1 0 x 1 0	1 1 1 0 x x 0 1 0	0 1 0 0 x x x x x	x x x x x x x	0 0 0 0 1 0 0 0 0	0 0 0	0	0 0	0 0 ? ? ?
Siphonosoma vastum	1 0 0 0 0 1 0	0 1 0 x 1 0	1 1 1 0 x x 0 1 0	0 1 0 1 1 0 0 0 0	x 0 0 0 0 x 0	0 0 0 0 0 0 x 0 0 0	0 0 0	0	0 0	0 0 ? ? ?
Sipunculus norvegicus	1 0 0 0 0 0 0	0 1 1 0 1 0	0 1 0 0 x x 0 1 0	0 0 1 0 x x x x x	x x x x x x x	0 0 0 0 0 0 x 0 0 0	0 0 0	0	0 0	0 0 0 x x
Sipunculus nudus	1 0 0 0 0 0 0	0 1 1 0 1 1	0 1 0 0 x x 0 1 0	0 0 1 0 x x x x x	x x x x x x x	0 0 0 0 0 0 x 0 0 0	0 0 0	0	0 0	0 0 0 x x
Sipunculus phalloides	1 0 0 0 0 0 0	0 1 1 0 1 1	0 1 0 0 x x 0 1 0	0 0 1 0 x x x x x	x x x x x x x	0 0 0 0 0 0 x 0 0 0	0 0 0	0	0 0	0 0 0 x x
Sipunculus polymyotus	1 0 0 0 0 0 0	1 1 1 0 1 2	0 1 0 0 x x 0 1 0	0 0 1 0 x x x x x	x x x x x x x	0 0 0 0 0 0 x 0 0 0	0 0 0	0	0 0	0 0 0 x x
Themiste (L.) lagenlformes	1 0 1 0 0 0 0	1 0 x x 0 x	x 0 x 0 x x 0 1 0	2 1 1 0 x x x x x	x x x x x x x	0 0 0 0 1 0 1 x 1	x 0 x	2	0 0	0 0 2 0 2
Themiste (L.) minor	1 0 1 0 0 0 0	1 0 x x 0 x	x 0 x 0 x x 0 1 0	1 0 1 1 0 x 1 0 0	x 0 0 0 0 x 0	0 0 0 0 1 0 1 x 0	x 0 x	1	0 0	0 0 2 0 2
Themiste (T.) dyscrita	1 0 1 0 0 0 0	2 0 x x 0 x	x 0 x 0 x x 0 1 0	1 0 1 0 x x x x x	x x x x x x x	0 0 0 0 1 1 1 x 0	x 0 x	2	0 0	0 0 2 0 ?
Themiste (T.) hennahi	1 0 1 0 0 0 0	2 0 x x 0 x	x 0 x 0 x x 0 1 0	2 1 1 0 x x x x x	x x x x x x x	0 0 0 0 1 1 1 x 0	x 0 x	1	0 0	0 0 2 0 ?
Themiste (T.) pyroides	1 0 1 0 0 0 0	2 0 x x 0 x	x 0 x 0 x x 0 1 0	2 0 1 1 0 x 1 0 0	x 0 0 0 0 x 0	0 0 0 0 1 1 1 x 0	x 0 x	2	0 0	0 0 2 0 ?
Thysanocardia catherinae	1 1 0 1 0 0 0	0 0 x x 0 x	x 0 x 0 x x 0 1 0	1 0 1 0 x x x x x	x x x x x x x	0 0 0 0 1 0 1 x 1	x 0 x	2	0 1	0 0 2 0 0
Thysanocardia nigra	1 1 0 1 0 0 0	2 0 x x 0 x	x 0 x 0 x x 0 1 0	1 0 1 0 x x x x x	x x x x x x x	0 0 0 0 1 0 1 x 1	x 0 x	1,2	0 1	0 0 2 0 0
Xenosiphon branchiatus	1 0 0 0 0 0 0	2 1 1 1 1 1	0 1 0 0 x x 0 1 0	0 0 1 0 x x x x x	x x x x x x x	0 0 0 0 0 x 0 0 0	0 0 0	0	1 0	0 0 ? ? ?

References

Aagesen, L., Petersen, G., Seberg, O., 2005. Sequence length variation, indel costs and congruence in sensitivity analysis. Cladistics 21, 15–30.

Åkesson, B., 1958. A study of the nervous system of the sipunculoideae, with some remarks on the development of the two species *Phascolion strombi* Montagu and *Golfingia minuta* Keferstein. Undersokningar over Oresund 38, 1–249.

Andreae, J., 1882. Beitrage zur Anatomie und Histologie des *Sipunculus nudus* L. Zeitschr. Wiss. Zool. 36, 201–255.

Andrews, E.A., 1890. Notes on the anatomy of *Sipunculus gouldii* Pourtales. Stud. Biol. Lab. Johns Hopkins Univ. 4, 384–430.

Bleidorn, C., Vogt, L., Bartolomaeus, T., 2003a. A contribution to sedentary polychaete phylogeny using 18S rRNA sequence data. J. Zool. Syst. Evol. Res. 41, 186–195.

Bleidorn, C., Vogt, L., Bartolomaeus, T., 2003b. New insights into polychaete phylogeny (Annelida) inferred from 18S rDNA sequences. Mol. Phylogenet. Evol. 29, 279–288.

Bleidorn, C., Podsiadlowski, L., Bartolomaeus, T., 2005. 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. Mol. Biol. Evol. 19, 127–137.

Brown, S., Rouse, G., Hutchings, P., Colgan, D., 1999. Assessing the usefulness of histone H3, U2 snRNA and 28S rDNA in analyses of polychaete relationships. Austr. J. Zool. 47, 499–516.

Colgan, D.J., McLauchlan, A., Wilson, G.D.F., Livingston, S.P., Edgecombe, G.D., Macaranas, J., Cassis, G., Gray, M.R., 1998. Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. Austr. J. Zool. 46, 419–437.

Cutler, E.B., 1973. Sipuncula of the western North Atlantic. Bull. Amer. Mus. Nat. Hist. 152, 105–204.

Cutler, E.B., 1979. A reconsideration of the *Golfingia* subgenera *Fisherana* Stephen, *Mitosiphon* Fisher, and *Apionsoma* Sluiter (Sipuncula). Zool. J. Linn. Soc. 65, 367–384.

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

Cutler, E.B., Cutler, N.J., 1987. Deep-water Sipuncula from the Eastern Atlantic Ocean. Sarsia 72, 71–89.

Cutler, E.B., Cutler, N.J., 1989. A revision of the genus *Aspidosiphon* (Sipuncula, Aspidosiphonidae). Proc. Biol. Soc. Wash. 102, 826–865.

Cutler, N.J., Cutler, E.B., 1990. A revision of the subgenus *Phascolosoma* (Sipuncula, *Phascolosoma*). Proc. Biol. Soc. Wash. 3, 691–730.

Cutler, E.B., Gibbs, P.E., 1985. A phylogenetic analysis of higher taxa in the phylum Sipuncula. Syst. Zool. 34, 162–173.

Edmonds, S.J., 1980. A revision of the systematics of Australian sipunculans (Sipuncula). Rec. S. Austr. Mus. 18, 1–74.

Farris, J.S., 1997. The future of phylogeny reconstruction. Zool. Scr. 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.

Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1995. Testing significance of incongruence. Cladistics 10, 315–319.

Fleissner, R., Metzler, D., von Haeseler, A., 2005. Simultaneous statistical multiple alignment and phylogeny reconstruction. Syst. Biol. 54, 548–561.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Bio. Biotechnol. 3, 294–295.

Gibbs, P.E., Cutler, E.B., 1987. A classification of the phylum Sipuncula. Bull. Brit. Mus. Nat. Hist. (Zool.) 52, 43–58.

Giribet, G., 2003. Stability in phylogenetic formulations and its relationship to nodal support. Syst. Biol. 52, 554–564.

Giribet, G., 2005. Generating implied alignments under direct optimization using POY. Cladistics 21, 396–402.

Giribet, G., Carranza, S., Baguñà, J., Riutort, M., Ribera, C., 1996. First molecular evidence for the existence of a Tardigrada + Arthropoda clade. Mol. Biol. Evol. 13, 76–84.

Giribet, G., Distel, D.L., Polz, M., Sterrer, W., Wheeler, W.C., 2000. Triploblastic relationships with emphasis on the acoelomates and the position of Gnathostomulida, Cycliophora, Plathelminthes, and Chaetognatha: a combined approach of 18S rDNA sequences and morphology. Syst. Biol. 49, 539–562.

Giribet, G., Edgecombe, G.D., Wheeler, W.C., Babbitt, C., 2002. Phylogeny and systematic position of Opiliones: a combined analysis of chelicerate relationships using morphological and molecular data. Cladistics 18, 5–70.

Giribet, G., Rambla, M., Carranza, S., Baguñà, J., Riutort, M., Ribera, C., 1999. Phylogeny of the arachnid order Opiliones (Arthropoda) inferred from a combined approach of complete 18S and partial 28S ribosomal sequences and morphology. Mol. Phylogenet. Evol. 11, 296–307.

Giribet, G., Wheeler, W.C., 1999. On gaps. Mol. Phylogenet. Evol. 13, 132–143.

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: DeSalle, R., Giribet, G., Wheeler, W.C. (Eds.), Techniques in Molecular Systematics and Evolution. Birkhäuser Verlag, Basel, pp. 70–79.

Hall, K.A., Hutchings, P., Colgan, D.J., 2004. Further phylogenetic studies of the Polychaeta using 18S rDNA sequence data. J. Mar. Biol. Ass. UK 84, 949–960.

Hall, T. 1994. BioEdit Sequence Alignment Editor 6.0.7 (available at http://www.mbio.ncsu.edu/BioEdit/page2.html).

Huang, D., Chen, J., Vannier, J., Saiz Salinas, J.I., 2004. Early Cambrian sipunculan worms from southwest China. Proc. Roy. Soc. Lond. B 271, 1671–1676.

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

Jukes, T., Cantor, C., 1969. Evolution of protein molecules. In: Munro, H. (Ed.), Mammalian Protein Metabolism. Academic Press, New York.

Keferstein, W., 1867. Untersuchungen uber einige amerikanische Sipunculiden. Zeitsch. Wiss. Zool. 17, 44–55.

Knowlton, N., 1993. Sibling species in the sea. Ann. Rev. Ecol. Syst. 24, 189–216.

Knowlton, N., 2000. Molecular genetic analyses of species boundaries in the sea. Hydrobiologia 420, 73–90.

Lewis, P.O., 2001. A likelihood approach to estimating phylogeny from discrete morphological data. Syst. Biol. 50, 913–925.

Linton, E.W., 2005. MacGDE: Genetic Data Environment for MacOS X. Ver. 2.0. Software available at www.msu.edu/~lintone/macgde/.

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. Mol. Biol. Evol. 27, 489–503.

Mickevich, M.F., Farris, J.S., 1981. The implications of congruence in *Menidia*. Syst. Zool. 27, 143–158.

Moilanen, A., 1999. Searching for most parsimonious trees with simulated evolutionary optimization. Cladistics 15, 39–50.

Moilanen, A., 2001. Simulated evolutionary optimization and local search: introduction and application to tree search. Cladistics 17, S12–S25.

Nylander, J.A.A., 2004. MrModeltest v2. Program distributed by the author.

Pilger, J.F., 1982. Ultrastructure of the tentacles of *Themiste lageniformis* (Sipuncula). Zoomorph. 100, 143–156.

Peterson, K.J., Eernisse, D.J., 2001. Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. Evol. Dev. 3, 170–205.

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

Purschke, G., 1997. Ultrastructure of the nuchal organs in polychaetes (Annelida)—new results and review. Acta Zool. 78, 123–143.

Purschke, G., 2002. On the ground pattern of Annelida. Org. Divers. Evol. 2, 181–196.

Redelings, B.D., Suchard, M.A., 2005. Joint Bayesian estimation of alignment and phylogeny. Syst. Biol. 54, 410–418.

Rice, M.E., 1985. Description of a wood dwelling sipunculan, *Phascolosoma turnerae*, new species. Proc. Biol. Soc. Wash. 98, 54–60.

Rice, M.E., 1993. Sipuncula. In: Harrison, F.W., Rice, M.E. (Eds.), Microscopic Anatomy of Invertebrates: Onychophora, Chilopoda, and Lesser Protostomata. Wiley-Liss, New York, pp. 237–325.

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

Ruppert, E.E., Rice, M.E., 1995. A functional organization of dermal coelomic canals in *Sipunculus nudus* (Sipuncula) with a discussion of respiratory designs in sipunculans. Invert. Zool. 114, 51–63.

Saito, M., Kojima, S., Endo, K., 2000. Mitochondrial CO1 sequences of brachiopods: genetic code shared with protostomes and limits of utility for phylogenetic reconstruction. Mol. Phylogenet. Evol. 15, 331–344.

Savill, N.J., Hoyle, D.C., Higgs, P.G., 2001. RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum likelihood methods. Genetics 157, 399–411.

Scheltema, A.H., 1996. Phylogenetic position of Sipuncula, Mollusca and the progenetic Aplacophora. In: Taylor, J. (Ed.), Origin and Evolutionary Radiation of the Mollusca. Oxford University Press, pp. 53–58.

Schöninger, M., von Haeseler, A., 1994. A stochastic model and the evolution of autocorrelated DNA sequences. Mol. Phylogenet. Evol. 3, 2–240.

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

Shipley, A.E., 1892. On a new species of *Phymosoma*. Proc. Camb. Philos. Soc. 7, 77–78.

Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analysis. Syst. Biol. 49, 369–381.

Simmons, M.P., Pickett, K.M., Miya, M., 2004. How meaningful are Bayesian support values? Mol. Biol. Evol. 21, 188–199.

Smith, S.W., Overbeek, R., Woese, C.R., Gilbert, W., Gillevet, P.M., 1994. The Genetic Data Environment: an expandable GUI for multiple sequence analysis. Comput. Appl. Biosci. 10, 671–675.

Staton, J., Rice, M.E., 1999. Genetic Differentiation despite teleplanic larval dispersal: allozyme variation in sipunculans of the *Apionsoma misakianum* species complex. Bull. Mar. Sci. 65, 467–480.

Staton, J.L., 2003. Phylogenetic analysis of the mitochondrial cytochrome c oxidase subunit 1 gene from 13 sipunculan genera: intra-and interphylum relationships. Invert. Biol. 122, 252–264.

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

Swofford, D.L., 2003. PAUP*: Phylogenetic Analysis Using Parsimony (and other methods), version 4.0b10 for Macintosh. Sinauer, Sunderland, MA.

Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect. Math. Life Sci. 17, 57–86. 368–376.

Telford, M.J., Wise, M.J., Gowri-Shankar, V., 2005. Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the Bilateria. Mol. Biol. Evol. 22, 1129–1136.

Van de Peer, Y., De Rijk, P., Wuyts, J., Winkelmans, T., De Wachter, R., 2000. The European small subunit ribosomal RNA database. Nucleic Acids Res. 28, 175–176.

Voss-Foucart, M.F., Barzin, S., Jeuniaux, C., Bussers, J.C., 1977. Étude comparée de la composition chimique des regions souples et durcies de la cuticule de quatre espèces de siponculiens. Cah. Biol. Mar. 18, 135–145.

Voss-Foucart, M.F., Barzin, S., Toussaint, C., 1978. Étude comparée de la composition en acides amines d'extraits peptidiques de la cuticule de siponculiens et d'annelides polychetes. Arch. Zool. Exp. Gen. 118, 457–470.

Wheeler, W.C., 1995. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. Syst. Biol. 44, 321–331.

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

Wheeler, W.C., 2003. Implied alignment: a synapomorphy-based multiple-sequence alignment method and its use in cladogram search. Cladistics 19, 261–268.

Wheeler, W.C., 2006. Dynamic homology and the likelihood criterion. Cladistics 22, 157–170.

Wheeler, W.C., Gladstein, D., DeLaet, J., 2004–2006. POY: the optimization of alignment characters, version 3.0. American Museum of Natural History. Program and documentation available at ftp.amnh.org/pub/molecular.

Whiting, M.F., Carpenter, J.M., Wheeler, Q.D., 1997. The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. Syst. Biol. 46, 1–68.

Zarkihk, A., 1994. Estimation of evolutionary distances between nucleotide sequences. J. Mol. Evol. 39, 315–329.