Molecular Phylogenetics and Evolution 83 (2015) 174–183

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Re-evaluating the phylogeny of Sipuncula through transcriptomics

Sarah Lemer [a,*,1], Gisele Y. Kawauchi [a,b,1], Sónia C.S. Andrade [a,c], Vanessa L. González [a,d], Michael J. Boyle [e], Gonzalo Giribet [a]

[a] Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA
[b] CEBIMar, Universidade de São Paulo, Praia do Cabelo Gordo, São Sebastião, São Paulo, Brazil
[c] Departamento de Zootecnia, ESALQ-USP, Piracicaba, São Paulo, Brazil
[d] Department of Invertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013, USA
[e] Smithsonian Tropical Research Institute (STRI), Naos Marine Laboratories, Panama 0843/03092, Panama

ARTICLE INFO

Article history:
Received 21 July 2014
Revised 17 October 2014
Accepted 23 October 2014
Available online 30 October 2014

Keywords:
Annelida
Peanut worms
Phylogenomics
Systematics

ABSTRACT

Sipunculans (also known as peanut worms) are an ancient group of exclusively marine worms with a global distribution and a fossil record that dates back to the Early Cambrian. The systematics of sipunculans, now considered a distinct subclade of Annelida, has been studied for decades using morphological and molecular characters, and has reached the limits of Sanger-based approaches. Here, we reevaluate their family-level phylogeny by comparative transcriptomic analysis of eight species representing all known families within Sipuncula. Two data matrices with alternative gene occupancy levels (large matrix with 675 genes and 62% missing data; reduced matrix with 141 genes and 23% missing data) were analysed using concatenation and gene-tree methods, yielding congruent results and resolving each internal node with maximum support. We thus corroborate prior phylogenetic work based on molecular data, resolve outstanding issues with respect to the familial relationships of Aspidosiphonidae, Antillesomatidae and Phascolosomatidae, and highlight the next area of focus for sipunculan systematics.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Sipuncula is a clade of unsegmented, coelomate marine worms (commonly known as peanut worms) that inhabit a diversity of benthic substrates in all major ocean basins across polar, temperate and tropical latitudes. The number of recognized species ranges from a systematic compilation of approximately 320 (Stephen and Edmonds, 1972), to a revised number of 149, including the introduction of new family-level clades (Cutler, 1994). Efforts to identify and name sipunculan families have progressed for more than a century, beginning with the use of several non-distinct group names (Baird, 1868; Pickford, 1947; Åkesson, 1958), followed by the establishment of four distinct families, Sipunculidae, Golfingiidae, Phascolosomatidae and Aspidosiphonidae (Stephen and Edmonds, 1972), and an increase to six families with the addition of Themistidae and Phascolionidae (Cutler and Gibbs, 1985). Over the past three decades, internal relationships within and among sipunculan clades have been inferred through numerical analyses of taxonomic characters (Cutler and Gibbs, 1985), DNA sequence data (Maxmen et al., 2003; Staton, 2003), and the combined use of morphological and molecular characters (Schulze et al., 2005, 2007). More recently, an extended dataset of six gene loci was analysed, building upon previous molecular approaches, and proposing a revised classification system with the following six sipunculan families: Sipunculidae, Golfingiidae, Siphonosomatidae, Antillesomatidae, Phascolosomatidae and Aspidosiphonidae (Kawauchi et al., 2012). While the basic structure of the sipunculan tree has been consistent across most of these studies, relationships among the families Antillesomatidae, Phascolosomatidae and Aspidosiphonidae remain inconclusive, as well as both the composition and branching patterns of genera within Golfingiidae.

For several reasons, resolving taxonomic relations within Sipuncula is more constructive and appropriate than ever. First, sipunculans traditionally have been considered a distinct animal phylum (Sedgwick, 1898; Hyman, 1959; Clark, 1969; Stephen and Edmonds, 1972; Rice, 1985; Saiz-Salinas, 1993; Cutler, 1994; Valentine, 1997; Strand et al., 2010). However, a series of molecular hypotheses show accumulative support for the inclusion of sipunculans within the annelid radiation (McHugh, 1997; Boore and Staton, 2002; Struck et al., 2007; Dunn et al., 2008; Struck et al., 2011), as one of the earliest diverging annelid lineages (Dordel et al., 2010; Hejnol et al., 2009; Struck et al., 2011; Weigert et al., 2014), suggesting that Annelida should be recognized as one of the most biologically diverse clades within Spiralia

* Corresponding author.
E-mail address: sarah.lemer@gmail.com (S. Lemer).
[1] Authors contributed equally to this study.

http://dx.doi.org/10.1016/j.ympev.2014.10.019
1055-7903/© 2014 Elsevier Inc. All rights reserved.

(also referred to as Lophotrochozoa), and Bilateria. Second, with evidence of adult sipunculans in the Lower Cambrian (Huang et al., 2004), and the oldest annelid fossil representatives (Weigert et al., 2014), it appears that at least one unsegmented annelid body plan has persisted for the past 520 Myr. In this context, the morphogenetic origin, or loss, of segmentation within Annelida is a mystery that will continue to stimulate research on body plan evolution. Thus far, it has been suggested that ancestral remnants of segmentation are reflected in neuronal architecture of the ventral nerve cord during larval development of sipunculans (Kristof et al., 2008, 2011; Wanninger et al., 2009), which certainly warrants additional, more comprehensive studies. Third, sipunculans are valuable research organisms for reproductive biology (Rice, 1973, 1989, 1993; Reunov and Rice, 1993; Adrianov and Maiorova, 2010), comparative development (Åkesson, 1958; Rice, 1967, 1975, 1988; Schulze and Rice, 2009a) and life history character reconstruction and evolution (Jägersten, 1972; Rice, 1976, 1985). They are also emerging as important non-model organisms for evolutionary and developmental biology, or evo-devo (Schulze and Rice, 2009b; Wanninger et al., 2005, 2009; Wanninger, 2008; Boyle and Seaver, 2010; Boyle and Rice, 2014). Fourth, due to an extended larval phase described for several species within multiple families (Scheltema and Hall, 1965, 1975; Rice, 1976, 1981; Scheltema and Rice, 1990), sipunculans constitute an interesting group for studying dispersal within and between widely separated oceanic regions, which is a topic addressed in several recent studies of cosmopolitanism in the marine realm (Staton and Rice, 1999; Kawauchi and Giribet, 2010; Kawauchi and Giribet, 2013; Schulze et al., 2012; Young et al., 2012; Lemer and Planes, 2014). These and similar studies rely upon robust phylogenetic hypotheses, from species to family-level relationships, to provide a stable evolutionary framework for critical re-interpretation of previous research, and to guide future investigations. This is particularly relevant to our project, considering that sipunculan familial diversity may extend as far back as the Mesozoic (Kawauchi et al., 2012).

In 1985 Cutler and Gibbs proposed a testable phylogenetic model of sipunculan relationships using taxonomic characters to erect new families, orders and classes. In essence, that work was...

> *"... an attempt to apply some of the extant phylogenetic methodology and logic to a phylum of poorly known, soft-bodied marine invertebrates for which there is no fossil record, an inadequate outgroup comparison on which to root character polarities, and only a modest number of useful characters."*

Today, sipunculans are generally well known, they have a distinct fossil record, adequate outgroup comparisons are plentiful, and both the number and nature of useful characters have radically changed. In addition, genomic and transcriptomic resources are now more frequently being applied to resolve phylogenetic relationships at both broad and narrow taxonomic levels. Next-generation sequencing (NGS) now enables us to generate large data sets for many species at a relatively low cost. The systematic community has necessarily progressed from candidate genes through EST-based methods to 454 and Illumina-based transcriptome and genome datasets to resolve major relationships among the animal phyla (e.g., Dunn et al., 2008; Hejnol et al., 2009; Nosenko et al., 2013; Ryan et al., 2013; Moroz et al., 2014). Long outstanding issues have been resolved for within-phylum relationships among arthropods, molluscs, and annelids, to mention just some of the largest animal phyla (e.g., Meusemann et al., 2010; Kocot et al., 2011; Smith et al., 2011; Struck et al., 2011; von Reumont et al., 2012; Andrade et al., 2014; Weigert et al., 2014). A third wave now focuses on resolving lower-level phylogenetic questions, relying almost entirely on Illumina-based technology (e.g., Johnson et al., 2013; Kocot et al., 2013; Wheat and Wahlberg, 2013; Dell'Ampio et al., 2014; Fernández et al., 2014a, 2014b). The time is prime to address the phylogeny of Sipuncula through phylogenomic techniques. Thus, our main objective is to revisit the branching pattern of sipunculan families using a novel phylogenomic approach to address outstanding issues among them. For this, we sequenced, assembled and analysed eight sipunculan transcriptomes, including all six sipunculan families, and found a close correlation with prior studies while further resolving previous outstanding questions with high support.

2. Materials and methods

2.1. Taxon sampling, cDNA library construction and next-generation sequencing

Live specimens of eight species representing all sipunculan families were collected by MJB and GYK: *Antillesoma antillarum*, *Phascolosoma perlucens*, *Aspidosiphon parvulus*, *Nephasoma pellucidum*, *Phascolion cryptum*, *Siphonosoma cumanenses* and *Sipunculus nudus* (although for this study *S. nudus* transcriptome was retrieved from Riesgo et al., 2012); *Phascolopsis gouldii* was obtained from Marine Biological Specimens (Woods Hole, Massachusetts). Information about the sampling localities can be found in the MCZ online collections database (http://mczbase.mcz.harvard.edu) and in Table 1. In addition, 10 taxa were chosen as outgroups from which 3 were collected live for this study (*Baseodiscus unicolor*, *Argonomertes australiensis* and *Chaetopterus* sp.); 3 transcriptomes were retrieved from Riesgo et al. (2012; *Chiton olivaceus*, *Hormogaster samnitica* and *Octopus vulgaris*); 2 transcriptomes were retrieved from GenBank (*Owenia fusiformis* and *Magelona johnstoni*), and 2 transcriptomes were provided directly by Weigert et al. (2014; *Eurythoe complanata* and *Paramphinome jeffreysii*).

All samples were sent alive to the laboratory, and flash frozen in liquid nitrogen, or fixed in RNA*later*® (Life Technologies, Carlsbad, CA, USA) and stored at −80 °C. Total RNA was extracted using TRIzol (Life Sciences) following the manufacturer's protocol. In brief, tissue fragments were disrupted with a drill in 1000 ml total of TRIzol. After 5 min incubating at room temperature (RT), 100 ml of bromochloropropane was mixed by vortexing and incubated at RT for 10 min. The samples were then centrifuged at 16,000 rpm for 15 min at 4 °C. The upper aqueous layer was recovered, mixed with 500 ml of isopropanol, and incubated at RT for 10 min. Samples were centrifuged again for 15 min at 16,000 rpm at 4 °C, in order to precipitate total RNA. The pellet was washed twice in 1 ml of 75% isopropanol and centrifuged at 16,000 rpm at 4 °C for 15 min and 5 min respectively, air dried and eluted in 30μL of RNA Storage solution (Ambion). Purification of mRNA was performed using the Dynabeads (Invitrogen) following the manufacturer's instructions. Finally, mRNA was eluted in 15 ml of Tris–HCl buffer and quality was measured with a pico RNA assay in an Agilent 2100 Bioanalyzer (Agilent Technologies). Final mRNA quantity per extraction was measured with a RNA assay in Qubit fluorometer (Life Technologies).

The TruSeq RNA Sample Preparation kit (Illumina Inc., San Diego, California, USA) was used to construct the cDNA libraries of all the collected sipunculans, following the manufacturer's instructions. For *B. unicolor*, *A. australiensis* and *Chaetopterus* sp., libraries were constructed as described in Riesgo et al. (2012). Each library was marked with a distinct index to allow pooling for sequencing. Each library concentration was measured with a dsDNA High Sensitivity (HS) assay in a Qubit fluorometer (Invitrogen); quality and size selection was assessed with an HS DNA assay in an Agilent 2100 Bioanalyzer (Agilent Technologies). Finally, the samples were run on the Illumina HiSeq 2500 platform with paired-end reads of 150 bp at the FAS Center for Systems Biology at Harvard University.

Table 1
Species included in the analysis, including new and publicly available data. Illumina paired-end sequencing was used to produce all the data. The public archive used was NCBI. Transcriptomes marked with * were directly provided by Weigert et al. (2014). Voucher accession numbers beginning with IZ, MAL or DNA are at the Harvard Museum of Comparative Zoology.

Species	Specimen voucher/SRA number	Sampling location	N raw reads	N reads after filtering	Assembler	Ncontigs (>199 bp)	n50	Longuest contig	N contigs > 999 bp	Total length (bp)	N peptide sequences retained
Antillesoma antillarum	IZ-130189/SRR1646260	Fort Pierce, FL, USA	45,797,278	40,063,902	Trinity	62,043	367	11,036	14,739	4,905,310	17045
Aspidosiphon parvulus	IZ-46449/SRR1646391	Fort Pierce, FL, USA	56,513,388	50,464,330	Trinity	60,773	500	15,313	7671	50,633,146	13419
Nephasoma pellucidum	IZ-46448/SRR1646439	Fort Pierce, FL, USA	27,193,126	22,939,761	Trinity	73,784	994	30,397	19,728	76,540,449	22693
Phascolion cryptum	IZ-46450/SRR1646440	Fort Pierce, FL, USA	24,962,058	23,049,775	Trinity	40,797	546	12,774	5717	33,911,131	13346
Phascolopsis gouldii	IZ-130398/SRX755857	Woodshole, MA, USA	162,419,226	104,759,625	Trinity	92,555	502	7702	8312	69,316,935	13852
Phascolosoma perlucens	IZ-46453/SRR1646442	Fort Pierce, FL, USA	44,005,999	37,712,203	Trinity	47,068	560	15,688	7226	40,614,030	12488
Siphonosoma cumanense	IZ-46451/SRR1646441	Fort Pierce, FL, USA	19,663,280	18,362,312	Trinity	9839	418	7824	1084	8,702,726	3480
Sipunculus nudus	IZ-130438/SRR619011	Fort Pierce, FL, USA	195,601,190	34,173,928	Trinity	74,929	382	5278	2305	52,391,407	7656
Outgroups											
Baseodiscus unicolor	IZ-135322/SRR1505175	Bocas del Toro, Panama	175,593,324	78,906,444	Trinity	616,533	547	21,057	25,404	181,281,613	6435
Argonemertes australiensis	IZ-135314/SRR1506999, SRR1507000, SRR1507001	Tasmania, Australia	128,371,852	39,646,942	Trinity	142,931	590	11,311	8344	49,955,925	15973
Chiton olivaceus	MAL-378064/SRR618506	Tossa de Mar, Girona, Spain	82,814,428	55,901,966	Trinity	327,201	524	9463	12,958	93,638,412	22648
Hormogaster samnitica	GEL6[a]/SRR618446	Gello, Toscana, Italy	53,956,780	31,623,984	Trinity	296,395	874	11,234	30,546	110,940,165	28829
Octopus vulgaris	DNA106283/SRR331946	Blanes Bay, Spain	94,283,86	16,501,336	Trinity	146,680	647	14,344	9796	4,777,5665	14733
Chaetopterus sp.	G397/SRX755856	NA	42,587,754	12,694,056	Trinity	39,872	339	5232	669	24,810,493	5772
Owenia fusiformis	SRR1222288	NA		56,363,524	Trinity	33,302	772	15,436	7171	33,205,306	13212
Magelona johnstoni	SRR122229	NA		9,611,241	Trinity	5346	367	11,036	487	4,905,310	1251
*Eurythoe complanata**	NA	NA		379,418,476	Velvet	279,454	1435	16,344	106,367	379,418,476	35925
*Paramphinome jeffreysii**	SRR1257731	NA		35,568,312	CLC	48,014	539	8253	6337	35,568,312	21630

[a] Material deposited in the Department of Zoology and Physical Anthropology, Universidad Complutense de Madrid. See methods for details on sample preparation protocols.

2.2. Transcriptome assembly and identification of coding regions

Demultiplexed Illumina HiSeq 2500 sequencing results were retrieved in FASTQ format from the sequencing facility (Bauer Core – Harvard University) via FTP and in SRA format from GenBank for *Owenia fusiformis* and *Magelona johnstoni*. Raw reads for *Eurythoe complanata* and *Paramphinome jeffreysii* were not available, thus we used previously assembled data for these two samples. Each one of the other 16 samples was quality filtered and adapter trimmed using two different software packages. For *O. fusiformis* and *M. johnstoni*, we used Trimgalore version 0.3.3 (Wu et al., 2011), a tool incorporating both CutAdapt and FastQC. All reads with an average quality score lower than 30 based on a Phred scale, and shorter than 25 bp, were discarded. For all the other samples, we used SeqyClean (https://bitbucket.org/izhbannikov/seqyclean/downloads) to filter and trim all reads with a minimum Phred score set to 30 and a minimum size set to 65 bp. Vector contaminants were identified and filtered using the UniVec online database (http://www.ncbi.nlm.nih.gov/tools/vecscreen/univec/). In order to reduce the number of potential chimeric transcripts and the computational time for the assembly, ribosomal RNA (rRNA) was filtered out using Bowtie 1.0.0 (Langmead et al., 2009) by building a bowtie index using all known Annelida and Spiralia rRNA sequences that were downloaded from GenBank. All reads that did not align with the rRNA index were stored in fasta format as single files.

D*e novo* assemblies were conducted for each sample with Trinity (Grabherr et al., 2011; Haas et al., 2013) using paired read files and default parameters. Raw reads and assembled sequences have been deposited in the National Center for Biotechnology Information Sequence Read Archive and Transcriptome Shotgun Assembly databases (NCBI-SRA). *E. complanata* and *P. jeffreysii* assemblies were obtained from Weigert et al. (2014), who utilized Velvet and CLC, respectively. Reduction of redundant reads in each of the 18 raw assemblies was performed with CD-HIT version 4.6 (Fu et al., 2012) using a threshold of 98% global similarity. Reduced assemblies were then processed in TransDecoder (Haas et al., 2013) to identify candidate open reading frames within the transcripts. Predicted peptides were filtered for isoforms by selecting only one peptide per putative unigene with a custom Python script. This process chooses the longest open reading frame per trinity subcomponent, thus reducing variation within coding regions, caused by alternative splicing, closely related paralogs and allelic diversity. Filtered peptide sequences with all final candidate open reading frames were retained as multifasta files.

2.3. Orthology assignment and matrix construction

Stand-alone OMA v0.99u (Altenhoff et al., 2011; Altenhoff et al., 2013) was utilized to assign predicted open reading frames into orthologous groups across all samples. Contrary to best-hit approaches based on scores, the OMA algorithm uses evolutionary distances, considers distance inference uncertainty and differential gene losses, and includes many-to-many orthologous relations; making it more advantageous (Roth et al., 2008). The ortholog matrix is constructed from all-against-all Smith–Waterman protein alignments. The program identifies the "stable pairs," verifies them, and checks against potential paralogous genes before clustering cliques of stable pairs as groups of orthologs. All input files were single-lined multifasta files, and the parameters.drw file specified retained all default settings with the exception of "MaxTimePerLevel," which was set at 3600. The all-by-all local alignment process was parallelized across 100 cores of a single node once all the input pre-processing steps were achieved on a single core (to avoid risk of collision).

Two data matrices were generated for phylogenetic analyses: the first one was constructed by selecting the OMA ortholog groups containing 9 or more taxa. The second matrix was constructed by selecting the ortholog groups containing 13 or more taxa, thus increasing gene occupancy and reducing the amount of missing data. The ortholog group selection based on minimum taxon occupancy was performed using a custom Python script. The selected orthogroups for each matrix (675 and 141, respectively) were aligned individually using MUSCLE version 3.6 (Edgar, 2004). To account for alignment uncertainty and increase the signal-to-noise ratio of the data, we applied a probabilistic character masking to each alignment with ZORRO (Wu et al., 2012), using default parameters and FastTree 2.1.4 (Price et al., 2010) to construct guide trees. In all of the alignments, positions that were assigned a confidence score below the threshold of 5 by ZORRO were discarded, using a custom Python script. Ortholog groups for each matrix were concatenated using Phyutility 2.6 (Smith and Dunn, 2008). All the custom Python scripts used in this study were designed by C. Laumer and are deposited on the public online database GitHub (https://github.com/claumer).

2.4. Phylogenetic analyses

Maximum Likelihood inferences were conducted with PhyML-PCMA (Zoller and Schneider, 2013) as in Fernández et al. (2014b), except that we selected 10 principal components along with empirical amino acid frequencies in the analyses. PhyML-PCMA estimates a model through the use of a principal component analysis. The obtained principal components describe the substitution rates that covary the most among different protein families. In other words, the principal components define a semi-empirically determined parameterization for an amino acid substitution model specific to each data set (Zoller and Schneider, 2013). Bayesian inferences were conducted with ExaBayes version 1.21 with openmpi version 1.64 (The Exelixis Lab, http://sco.h-its.org/exelixis/web/software/exabayes/). ExaBayes implements a Markov chain Monte Carlo (MCMC) sampling approach similar to those in BEAST (Drummond and Rambaut, 2007) or MrBayes (Ronquist et al., 2012). However, it is better adapted for large datasets due to its ability to parallelize each independent run, each chain and the data (i.e., unique site patterns of the alignment). We used the amino acid model prior (aaPR), a discrete model prior, which mixes a combination of 18 models of evolution. Four independent Markov chain Monte Carlo chains (MCMC) were run for 1,000,000 generations, sampling every 100th generation. The first 2,500 trees (25%) were discarded as burn-in for each MCMC run prior to convergence (i.e., when maximum discrepancies across chains < 0.1). The small dataset (constructed with ortholog groups containing 13 or more taxa) was subjected to additional Bayesian analyses in PhyloBayes (Lartillot et al., 2009) using the CAT-GTR mixture model (Lartillot and Philippe, 2004) and two independent Markov chains. Convergence was tested using the "bpcomp" program in the PhyloBayes suite. Chains were considered to have converged when the "Maxdiff" between the two independent chains was <0.2 (see PhyloBayes manual).

To test for putative gene incongruence, we inferred individual gene trees for each ortholog group included in each of the two matrices using RAxML 7.7.5 (Berger et al., 2011) and the PROTGAMMALG4X model of selection. All individual best-scoring trees were concatenated per matrix (one file for the 50% taxon-occupancy matrix containing 675 genes, and one for the 75% taxon-occupancy matrix containing 141 genes) and fed into SuperQ v1.1 (Grünewald et al., 2013) in order to visualize intergene conflict. SuperQ decomposes all gene trees into quartets to infer a supernetwork where edge lengths are assigned based on quartet frequencies; it was run using the 'balanced' edge-weight

optimization function with no filter. The resulting supernetworks were visualized with SplitsTree v.4.13.1 (Huson and Bryant, 2006).

3. Results

3.1. Assembly statistics and orthology assignment

A total of 18 transcriptomes, of which 14 were newly sequenced, were used in this study to infer the phylogeny of sipunculans. A summary of the assembly statistics is shown in Table 1. In brief, after assembling each transcriptome with Trinity, the number of contigs longer than 199 bp ranged from 5346 (for *Magelona johnstoni*) to 616,533 (for *Baseodiscus unicolor*) with a n50 ranging from 339 (for *Chaetopterus* sp.) to 1435 (for *Eurythoe complanata*).

The number of peptide sequences retained per species after redundancy reduction, open reading frame prediction, selection of the longest open reading frame per putative unigene and isoform filtration ranged from 1251 (for *M. johnstoni*) to 35,925 (for *E. complanata)*. The orthology assignments of peptide sequences performed with OMA resulted in 49,648 orthogroups. From these orthogroups, we generated 2 data subsets to conduct all subsequent analyses: (1) Taxon50: a matrix of orthogroups containing each a minimum occupancy of at least 9 taxa, thus representing a 50% taxon-occupancy matrix and (2) Taxon75: a matrix of orthogroups containing each a minimum occupancy of at least 13 taxa, thus representing a 75% taxon-occupancy matrix. The total number of orthogroups was 675 in Taxon50 and 141 in Taxon75. The number of orthogroups represented per taxon varied from 99 to 623 for Taxon50, and from 30 to 139 for Taxon75 (Table 2). The length of each matrix after concatenation of the aligned orthogroups was 149,565 amino acids for Taxon50 and 27,798 amino acids for Taxon75, after the probabilistic character masking performed with ZORRO. In general, and for both datasets, the gene coverage per ingroup taxon had a maximum of 23% of missing data for *Phascolopsis gouldii* in Taxon75, and 62% missing data for *Siphonosoma cumanense* in Taxon50 (Table 2, visual representation in Fig. 1). In both datasets, the highest values of missing data were found in the outgroups, with *B. unicolor* being the taxon with the most missing data (75% in Taxon50 and 79% in Taxon75).

3.2. Phylogenetic relationships within Sipuncula

All the maximum likelihood and Bayesian phylogenetic analyses conducted on large and small matrices (Taxon50 and Taxon75), including the PhyloBayes analysis conducted on the small matrix only, yielded the same tree topology (Fig. 2). Sipuncula appeared monophyletic with Sipunculidae, represented by *Sipunculus nudus*, being the sister group to all other families, and species. All internal

Table 2
Characteristics of the datasets used for phylogenetic inferences.

Species	N orthologs selected Taxon50	Missing data Taxon50 (%)	N orthologs selected Taxon75	Missing data Taxon75 (%)
Antillesoma antillarum	595	12	136	4
Aspidosiphon parvulus	529	22	135	4
Nephasoma pellucidum	623	8	139	1
Phascolion cryptum	562	17	135	4
Phascolopsis gouldii	409	39	108	23
Phascolosoma perlucens	571	15	136	4
Siphonosoma cumanense	256	62	119	16
Sipunculus nudus	360	47	117	17
Outgroups				
Baseodiscus unicolor	99	75	30	79
Argonemertes australiensis	302	55	97	31
Chiton olivaceus	402	40	110	22
Hormogaster samnitica	539	20	129	9
Octopus vulgaris	454	33	117	17
Chaetopterus sp.	152	77	77	45
Owenia fusiformis	505	25	137	3
Magelona johnstoni	164	76	99	30
Eurythoe complanata	505	25	82	42
Paramphinome jeffreysii	362	46	80	43

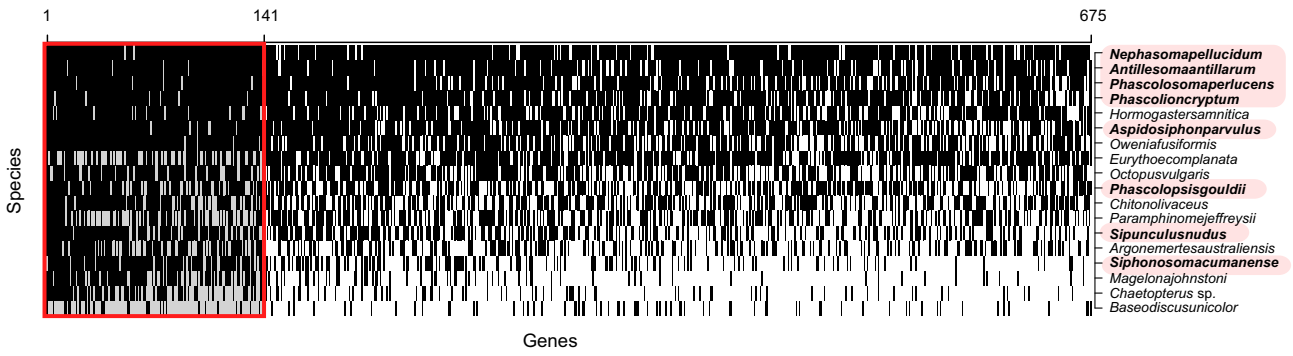

Fig. 1. Gene occupancy representation per species. A white cell indicates a non sampled genes. Taxa are sorted from the best (top) to worst (bottom) gene representation. *Nephasoma pellucidum* is the best-represented species, while *Baseodiscus unicolor* is the least represented one. Large matrix (Taxon50: 675 orthogroups) is represented as the larger box, and the reduced subset appears boxed in red (Taxon75: 141 orthogroups). Sipunculan species are highlighted. Although the small matrix contains fewer genes it has overall better gene occupancy (less white cells).

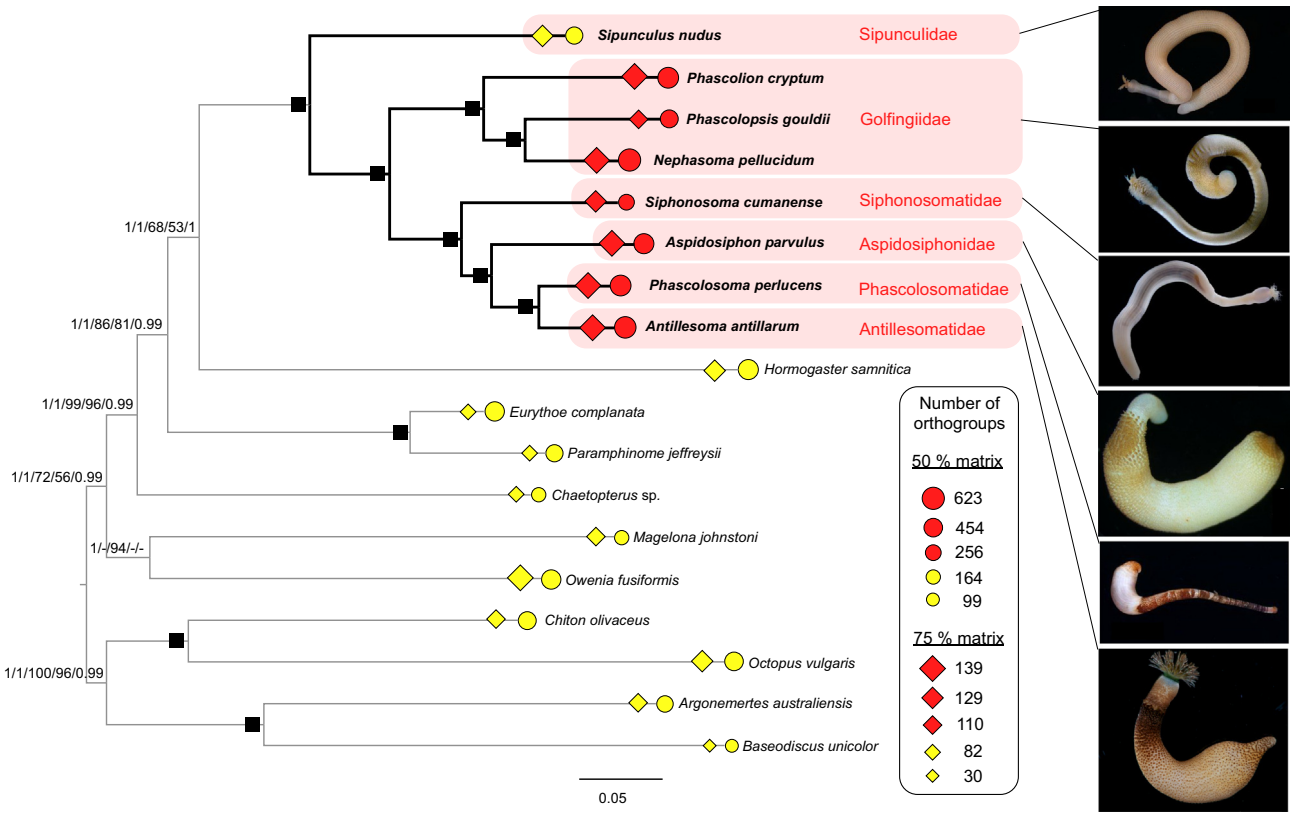

Fig. 2. Phylogenetic hypothesis based on the large data matrix analysed in PhyML-PCMA ($\ln L = -1{,}499{,}882.493745$). The area of the circles and diamonds at each tip is proportional to the number of genes present in the Taxon50 and Taxon75 matrices, respectively. Red circles and diamonds indicate new transcriptomes sequenced for this study. Support values (bootstrap values or posterior probabilities) are plotted as follows: large matrix ExaBayes/small matrix ExaBayes/large matrix PhyML/small matrix PhyML/small matrix PhyloBayes. Black squares at nodes indicate maximum support in all five analyses. All five analyses (ML and Bayesian) recovered the same tree topology with maximum nodal support within Sipuncula. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

nodes within Sipuncula had maximum support (100% bootstrap values and posterior probabilities of 1), visualized by black squares at each internal node on Fig. 2. All internal relationships were congruent with previous sipunculan phylogenetic studies, and the one remaining area of uncertainty, which was the relationship between Aspidosiphonidae, Phascolosomatidae and Antillesomatidae, is now fully resolved.

Phascolosomatidae, represented in our phylogeny by the species *Phascolosoma perlucens*, and Antillesomatidae, represented by *Antillesoma antillarum*, are sister groups. Aspidosiphonidae, represented by *Aspidosiphon parvulus*, is a sister group of the previous clade. All three species have low values of missing data (maximum being 22% in Taxon50 for *A. parvulus*; Table 2), suggesting that the observed pattern is not an artifact of missing data. In addition, the split networks constructed for each matrix, representing potential topological conflicts between individual gene trees, support this relationship (Fig. 3). Supernetworks for each matrix show that in most individual gene trees, *A. parvulus* is always separated from a group formed by *P. perlucens* and *A. antillarum* (Fig. 3a and b). Finally, all phylogenetic analyses recover Siphonosomatidae as the sister group of a clade formed by Aspidosiphonidae, Phascolosomatidae and Antillesomatidae, with Golfingiidae as the sister group of a clade formed by those four families.

Both split network analyses displayed a tree-like structure with topologies that were similar to that of the concatenated species tree (Figs. 2 and 3). The networks, however, indicated the presence of some gene conflict in the position of some outgroups, in particular for *M. johnstoni* and *O. fusiformis*, which is congruent with the concatenated species tree showing low nodal support for the position of these species (Fig. 2).

4. Discussion

For little more than a decade, attempts to resolve sipunculan relationships primarily utilized a small number of candidate genes from a common list of species (Maxmen et al., 2003; Staton, 2003; Schulze et al., 2005, 2007; Kawauchi et al., 2012). Our study represents the first use of entirely new sequence data, and a methodological departure from those earlier studies. Here, we employed state-of-the-art phylogenomic analyses of transcriptomes and recovered familial relationships that are, in many respects, similar to previous hypotheses (Fig. 2). Our results corroborate the proposed re-classification system of sipunculan families (Kawauchi et al., 2012), including the establishment of two new families (Siphonosomatidae and Antillesomatidae), and confirm the transference of *Phascolopsis*, a monotypic genus, to the single most inclusive sipunculan family, Golfingiidae. Collectively, multiple reassignments within the familial system highlight an important problem in sipunculan biology.

Within Sipuncula, there are four recognized life history patterns, including direct development, and indirect development with lecithotrophic and planktotrophic modes of larval formation (Rice, 1975, 1976, 1985; Boyle and Rice, 2014). As with previous molecular hypotheses (e.g., Maxmen et al., 2003; Schulze et al., 2007; Kawauchi et al., 2012) our transcriptome analyses confirm Sipunculidae as the sister clade to all other sipunculans. Development through a planktotrophic pelagosphera larva, unique among all metazoan larval types, is the only life history pattern observed thus far in Sipunculidae (Hatschek, 1883; Rice, 1988), suggesting planktotrophy as the plesiomorphic pattern of development within Sipuncula (Cutler, 1994). Interestingly, apart from a single

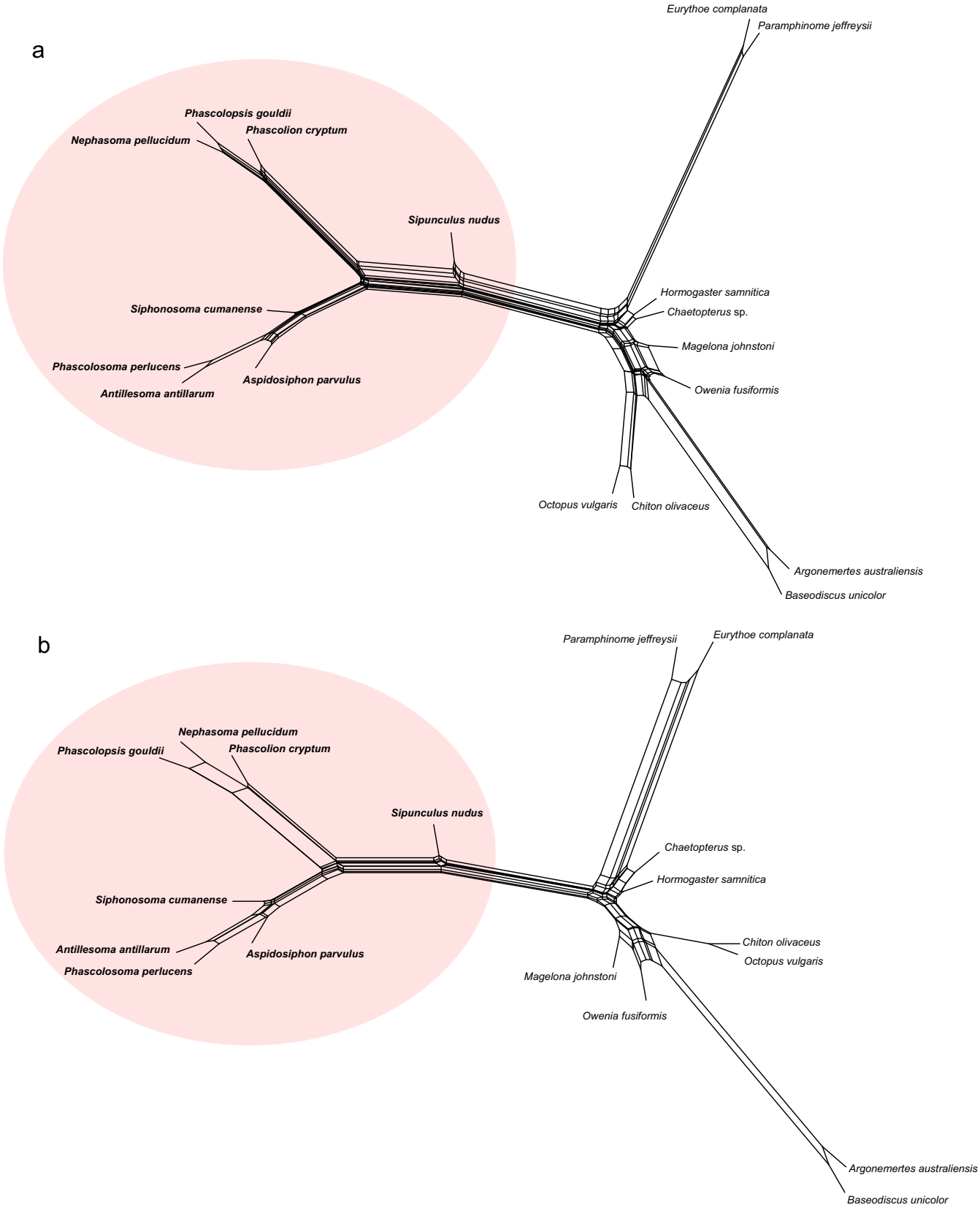

Fig. 3. Supernetwork representations of quartets derived from individual ML gene trees. (a) Large data matrix (Taxon50) and (b) small data matrix (Taxon75). Both supernetworks display a tree-like structure with a topology similar to that of the concatenated species trees. Phylogenetic conflict is detected for the relationship between *Magelona johnstoni* – *Owenia fusiformis* on supernetwork (a) (represented by an non-treelike split).

observation (Rice, 1970), planktotrophy is also the exclusive life history pattern in all other families (Siphonosomatidae, Aspidosiphonidae, Phascolosomatidae and Antillesomatidae) with one exception; Golfingiidae is the only clade in which all four life history patterns have been observed, and where the broadest diversity of larval and adult forms have been described (Rice, 1985; Pilger, 1987; Schulze et al., 2007; Schulze and Rice, 2009b; Kawauchi et al., 2012). Therefore, answers to fundamental questions about larval development, dispersal, speciation, and patterns of biodiversity may be found within this most genera-rich family. However, we have analysed transcriptome data for just three (*Phascolion*, *Phascolopsis* and *Nephasoma*) of seven genera within Golfingiidae, where internal relationships remain the least resolved, and where monophyly was previously recovered for only two genera, *Themiste* and *Thysanocardia* (Kawauchi et al., 2012). Supplemental sequencing and analyses are clearly required to resolve the internal relationships within this family.

Outside Golfingiidae, transcriptome analyses also recovered a distinct position for *Siphonosoma* (Siphonosomatidae), consistent with one of the new familial assignments proposed by Kawauchi et al. (2012). A distinct branch indicating the possibility of 'Siphonosomatidae' was also recovered in previous molecular hypotheses, which show alternative branching patterns, although this taxon was not formally named (Maxmen et al., 2003; Schulze et al., 2007). In all cases *Siphonosoma* was repositioned outside Sipunculidae, where it had been placed in an earlier classification scheme by parsimony analyses of morphological characters (Cutler and Gibbs, 1985). Comparative morphology previously suggested that *Siphonomecus*, another monotypic genus, was also a sister taxon of *Siphonosoma* within Sipunculidae (Cutler and Gibbs, 1985; Gibbs and Cutler, 1987). Because both genera share morphological characteristics distinct from *Sipunculus,* we predict that once sequence data are available for *Siphonomecus*, it will likely be reassigned to Siphonosomatidae, following Kawauchi et al. (2012). Additionally, transcriptome analyses recovered an internal bipartition showing Siphonosomatidae as the sister group to a larger clade consisting of three familial lineages: Aspidosiphonidae, Phascolosomatidae and Antillesomatidae (Fig. 2). The position of Siphonosomatidae, and respective branching patterns among the remaining families, received the highest support values in each of our phylogenetic analyses, and were clearly reflected in the ML gene trees (Fig. 3). A sister group relationship between Phascolosomatidae and Antillesomatidae was also supported in both gene-tree analyses (Fig. 3a and b), which separate *Aspidosiphon* from *Antillesoma* and *Phascolosoma* by a long edge. Although an analysis of the reduced matrix (Fig. 3b) showed some conflict regarding the position of *Aspidosiphon* (with respect to *Siphonosoma*), the large matrix (Fig. 3a) resolves this edge well, further supporting the phylogenetic analyses of the concatenated data (Fig. 2). These results finally resolve an outstanding topological controversy within the latest classification scheme, and support new familial assignments for both Siphonosomatidae and Antillesomatidae (Kawauchi et al., 2012).

One potential drawback when using large amounts of data like transcriptomes or genomes for phylogenetic reconstruction is the risk of increasing the amount of missing data. Missing data can have negative effects on phylogenetic reconstructions, such as inflating node support despite the absence of phylogenetic signal or producing misleading estimates of topology and branch lengths (e.g., Lemmon et al., 2009; Dell'Ampio et al., 2014) and it is thus recommended to graphically display the amount of missing data on phylogenetic trees or gene matrices (e.g., Roure et al. 2013), as we have done here (Figs. 1 and 2). Given our low levels of missing data and high matrix completeness (especially for the small matrix Taxon75: maximum of 23% of missing data for *Phascolopsis gouldii*), similar to other studies with comparable matrices and data analysis strategies (e.g., Andrade et al., 2014; Zapata et al., 2014), we think that our results should not be affected by gene occupancy or missing data artifacts. With the aim of having the most complete dataset possible, we optimized taxon sampling to represent all sipunculan families, ensuring that no major phylogenetic hypothesis was left untested for deep relationships within Sipuncula. However, we recommend adding more species per family in future studies in order to fully resolve internal relationships, especially within Golfingiidae. In the present case, increasing the amount of data from a few selected genes to hundreds of coding genes has enabled us to confirm previous molecular studies and resolve the last remaining controversies among sipunculan family relationships, supporting phylogenomics as an effective tool for resolving not only sipunculans, but also complex relationships within other spiralian clades, as shown by several recent studies (e.g., Smith et al., 2011; Kocot et al., 2011; Weigert et al., 2014).

In summary, regardless of which data sets were analyzed, or whether concatenation or gene-tree methods were utilized, our results agree in all aspects of the sipunculan phylogeny presented here (Figs. 2 and 3). Accordingly, the final hypothesis is strengthened by a combination of several factors, including prior assignment of orthologous gene groups, adequate representation of genes among the ingroup taxa, and co-assessment of both gene trees and species trees. We can thus conclude that after three decades of intense phylogenetic investigation, sipunculan familial relationships are markedly resolved. Furthermore, there is now a strong framework in place for reevaluating relationships among the multiple genera within Golfingiidae, and for pursuing outstanding questions on the evolutionary radiation and intriguing biology of these unique, unsegmented 'annelid' body plans that have persisted relatively intact since the Early Cambrian (Huang et al., 2004).

Acknowledgments

Many colleagues have assisted with fieldwork, specimens, and laboratory protocols, especially Marta Novo for the *Phascolopsis* library. FedEx is acknowledged for the reliable delivery of live sipunculans for RNA work. Rosa Fernández, Christopher Laumer and Horácio Montenegro assisted with analyses and scripts. The Bauer Core from the Faculty of Arts and Sciences at Harvard and the Research Computing Group, also from Harvard, are deeply acknowledged for their assistance at many stages of the data acquisition and analyses. We are grateful to Mary Rice for use of the Life Histories Laboratory at the Smithsonian Marine Station. The editor and two anonymous reviewers provided constructive criticism that helped to improve this article. This work was funded by the National Science Foundation, NSF #DEB-0844881 (Collaborative Research: Resolving old questions in Mollusc phylogenetics with new EST data and developing general phylogenomic tools) and NSF #DEB-0732903 (Collaborative Research: AToL: Phylogeny on the half-shell–Assembling the Bivalve Tree of Life) to G.G. This publication is Smithsonian Marine Station contribution no. 965.

References

Adrianov, A.V., Maiorova, A.S., 2010. Reproduction and development of common species of peanut worms (Sipuncula) from the Sea of Japan. Russ. J. Mar. Biol. 36, 1–15.

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. Undersökningar Över Öresnund XXXVIII, 249.

Altenhoff, A.M., Schneider, A., Gonnet, G.H., Dessimoz, C., 2011. OMA 2011: orthology inference among 1000 complete genomes. Nucleic Acids Res. 39, D289–D294.

Altenhoff, A.M., Gil, M., Gonnet, G.H., Dessimoz, C., 2013. Inferring hierarchical orthologous groups from orthologous gene pairs. PLoS ONE 8, e53786.

Andrade, S.C.S., Montenegro, H., Strand, M., Schwartz, M., Kajihara, H., Norenburg, J.L., Turbeville, J.M., Sundberg, P., Giribet, G., 2014. A transcriptomic approach to

ribbon worm systematics (Nermertea): resolving the Pilidiophora problem. Mol. Biol. Evol. http://dx.doi.org/10.1093/molbev/msu253.
Baird, W.B., 1868. Monograph on the species of worms belonging to the subclass Gephyreae with a notice of such species as contained in the collection of the British Museum. Ann. Des Sci. Nat. 3sr. Zool. Proc. Zool. Soc. London, 76–114.
Berger, S.A., Krompass, D., Stamatakis, A., 2011. Performance, accuracy, and Web server for evolutionary placement of short sequence reads under maximum likelihood. Syst. Biol. 60, 291–302.
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.
Boyle, M.J., Rice, M.E., 2014. Sipuncula: an emerging model of spiralian development and evolution. Int. J. Dev. Biol. http://dx.doi.org/10.1387/ijdb.140095mb.
Boyle, M.J., Seaver, E.C., 2010. Expression of FoxA and GATA transcription factors correlates with regionalized gut development in two lophotrochozoan marine worms: *Chaetopterus* (Annelida) and *Themiste lageniformis* (Sipuncula). EvoDevo 1, 2.
Clark, R.B., 1969. Systematics and phylogeny: Annelida, Echiura, Sipuncula. Chem. Zool. 4, 1–68.
Cutler, E.B., 1994. The Sipuncula: Their Systematics, Biology, and Evolution. Cornell University Press, Ithaca.
Cutler, E.B., Gibbs, P.E., 1985. A phylogenetic analysis of higher taxa in the phylum Sipuncula. Syst. Zool. 34, 162–173.
Dell'Ampio, E., Meusemann, K., Szucsich, N.U., Peters, R.S., Meyer, B., Borner, J., Petersen, M., Aberer, A.J., Stamatakis, A., Walzl, M.G., Minh, B.Q., von Haeseler, A., Ebersberger, I., Pass, G., Misof, B., 2014. Decisive data sets in phylogenomics: lessons from studies on the phylogenetic relationships of primarily wingless insects. Mol. Biol. Evol. 31, 239–249.
Dordel, J., Fisse, F., Purschke, G., Struck, T.H., 2010. Phylogenetic position of Sipuncula derived from multi-gene and phylogenomic data and its implication for the evolution of segmentation. J. Zool. Syst. Evol. Res. 48, 197–207.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.
Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E.C., Rouse, G.W., Obst, M., Edgecombe, G.D., Sørensen, M.V., Haddock, S.H.D., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R.M., Wheeler, W.C., Martindale, M.Q., Giribet, G., 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452, 745–749.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Fernández, R., Hormiga, G., Giribet, G., 2014a. Phylogenomic analysis of spiders reveals nonmonophyly of orb-weavers. Curr. Biol. 24, 1772–1777.
Fernández, R., Laumer, C.E., Vahtera, V., Libro, S., Kaluziak, S., Sharma, P.P., Pérez-Porro, A.R., Edgecombe, G.D., Giribet, G., 2014b. Evaluating topological conflict in centipede phylogeny using transcriptomic data sets. Mol. Biol. Evol. 31, 1500–1513.
Fu, L.M., Niu, B.F., Zhu, Z.W., Wu, S.T., Li, W.Z., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152.
Gibbs, P.E., Cutler, E.B., 1987. A classification of the phylum Sipuncula. Bull. Brit. Mus. Nat. Hist. (Zool.) 52, 43–58.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q.D., Chen, Z.H., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.
Grünewald, S., Spillner, A., Bastkowski, S., Bogershausen, A., Moulton, V., 2013. SuperQ: computing supernetworks from quartets. IEEE/ACM Trans. Comput. Biol. Bioinform. 10, 151–160.
Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512.
Hatschek, B., 1883. Ueber Entwicklung von *Sipunculus nudus*. Arbeiten aus dem Zoologischen Institute der Universität-Wien und der Zoologischen Station in Triest 5, 61–140.
Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G.W., Edgecombe, G.D., Martinez, P., Baguñà, J., Bailly, X., Jondelius, U., Wiens, M., Müller, W.E.G., Seaver, E., Wheeler, W.C., Martindale, M.Q., Giribet, G., Dunn, C.W., 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc. Roy. Soc. B Biol. Sci. 276, 4261–4270.
Huang, D.-Y., Chen, J.-Y., Vannier, J., Saiz Salinas, J.I., 2004. Early Cambrian sipunculan worms from southwest China. Proc. Biol. Sci. 271, 1671–1676.
Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.
Hyman, L.H., 1959. Phylum sipunculida. In: Hyman, L.H. (Ed.), The Invertebrates, Smaller Coelomate Groups, vol. 5. McGraw Hill, New York, pp. 610–690.
Jägersten, G., 1972. Evolution of the Metazoan Life Cycle. Academic Press, London.
Johnson, B.R., Borowiec, M.L., Chiu, J.C., Lee, E.K., Atallah, J., Ward, P.S., 2013. Phylogenomics resolves evolutionary relationships among ants, bees, and wasps. Curr. Biol. 23, 2058–2062.
Kawauchi, G.Y., Giribet, G., 2010. Are there true cosmopolitan sipunculan worms? A genetic variation study within *Phascolosoma perlucens* (Sipuncula, Phascolosomatidae). Mar. Biol. 157, 1417–1431.
Kawauchi, G.Y., Giribet, G., 2013. *Sipunculus nudus* Linnaeus, 1766 (Sipuncula): cosmopolitan or a group of pseudo-cryptic species? An integrated molecular and morphological approach. Mar. Ecol. http://dx.doi.org/10.1111/maec.12104.
Kawauchi, G.Y., Sharma, P.P., Giribet, G., 2012. Sipunculan phylogeny based on six genes, with a new classification and the descriptions of two new families. Zool. Scr. 41, 186–210.
Kocot, K.M., Cannon, J.T., Todt, C., Citarella, M.R., Kohn, A.B., Meyer, A., Santos, S.R., Schander, C., Moroz, L.L., Lieb, B., Halanych, K.M., 2011. Phylogenomics reveals deep molluscan relationships. Nature 477, 452–456.
Kocot, K.M., Halanych, K.M., Krug, P.J., 2013. Phylogenomics supports Pan pulmonata: Opisthobranch paraphyly and key evolutionary steps in a major radiation of gastropod molluscs. Mol. Phylogenet. Evol. 69, 764–771.
Kristof, A., Wollesen, T., Wanninger, A., 2008. Segmental mode of neural patterning in Sipuncula. Curr. Biol. 18, 1129–1132.
Kristof, A., Wollesen, T., Maiorova, A.S., Wanninger, A., 2011. Cellular and muscular growth patterns during sipunculan development. J. Exp. Zool. Part B 316B, 227–240.
Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25.
Lartillot, N., Philippe, H., 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21, 1095–1109.
Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286–2288.
Lemer, S., Planes, S., 2014. Effects of habitat fragmentation on the genetic structure and connectivity of the black-lipped pearl oyster *Pinctada margaritifera* popualtions in french Polynesia. Mar. Biol. 161, 2035–2049.
Lemmon, A.R., Brown, J.M., Strangler-Hall, K., Lemmon, E.M., 2009. The effect of ambiguous data on phylogenomic estimates obtained by maximum likelihood and Bayesian inferences. Syst. Biol. 58, 130–145.
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. Phylogenet. Evol. 27, 489–503.
McHugh, D., 1997. Molecular evidence that echiurans and pogonophorans are derived annelids. Proc. Natl. Acad. Sci. USA 94, 8006–8009.
Meusemann, K., von Reumont, B.M., Simon, S., Roeding, F., Strauss, S., Kück, P., Ebersberger, I., Walzl, M., Pass, G., Breuers, S., Achter, V., von Haeseler, A., Burmester, T., Hadrys, H., Wägele, J.W., Misof, B., 2010. A phylogenomic approach to resolve the arthropod tree of life. Mol. Biol. Evol. 27, 2451–2464.
Moroz, L.L., Kocot, K.M., Citarella, M.R., Dosung, S., Norekian, T.P., Povolotskaya, I.S., Grigorenko, A.P., Dailey, C., Berezikov, E., Buckley, K.M., Ptitsyn, A., Reshetov, D., Mukherjee, K., Moroz, T.P., Bobkova, Y., Yu, F., Kapitonov, V.V., Jurka, J., Bobkov, Y.V., Swore, J.J., Girardo, D.O., Fodor, A., Gusev, F., Sanford, R., Bruders, R., Kittler, E., Mills, C.E., Rast, J.P., Derelle, R., Solovyev, V.V., Kondrashov, F.A., Swalla, B.J., Sweedler, J.V., Rogaev, E.I., Halanych, K.M., Kohn, A.B., 2014. The ctenophore genome and the evolutionary origins of neural systems. Nature 510, 109–114.
Nosenko, T., Schreiber, F., Adamska, M., Adamski, M., Eitel, M., Hammel, J., Maldonado, M., Müller, W.E., Nickel, M., Schierwater, B., Vacelet, J., Wiens, M., Wörheide, G., 2013. Deep metazoan phylogeny: when different genes tell different stories. Mol. Phylogenet. Evol. 67, 223–233.
Pickford, G.E., 1947. Sipunculida. Encyclopedia Britannica, vol. 20. University of Chicago, Chicago, pp. 717–718b.
Pilger, J.F., 1987. Reproductive biology and development of *Themiste lageniformis*, a parthenogenic sipunculan. Bull. Mar. Sci. 41, 59–67.
Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2-approximately maximum-likelihood trees for large alignments. PLoS ONE 5, e9490.
Reunov, A., Rice, M.E., 1993. Ultrastructural observations on spermatogenesis in *Phascolion cryptum* (Sipuncula). Trans. Am. Microsc. Soc. 112, 195–207.
Rice, M.E., 1967. A Comparative study of the development of *Phascolosoma agassizii*, *Golfingia pugettensis*, and *Themiste pyroides* with a discussion of developmental patterns in the Sipuncula. Ophelia 4, 143–171.
Rice, M.E., 1970. Asexual reproduction in a sipunculan worm. Science 167, 1618–1620.
Rice, M.E., 1973. Morphology, behavior, and histogenesis of the pelagosphera larva of *Phascolosoma agassizii* (Sipuncula). Smith. Contrib. Zool. 132, 1–51.
Rice, M.E., 1975. Observations on the development of six species of Caribbean Sipuncula with a review of development in the phylum. In: Proceedings of the International Symposium on the Biology of the Sipuncula and Echiura Belgrade. Naučno Delo Press, Kotor, Montenegro, pp. 35–49.
Rice, M.E., 1976. Larval development and metamorphosis in Sipuncula. Am. Zool. 16, 563–571.
Rice, M.E., 1981. Larvae adrift: patterns and problems in life histories of sipunculans. Am. Zool. 21, 605–619.
Rice, M.E., 1985. Sipuncula: developmental evidence for phylogenetic inference. In: Morris, S.C., George, J.D., Gibson, R., Platt, H.M. (Eds.), The Origins and Relationships of Lower Invertebrates. Oxford University Press, USA, p. 400.
Rice, M.E., 1988. Observations on development and metamorphosis of *Siphonosoma cumanense* with comparative remarks on *Sipunculus nudus* (Sipuncula, Sipunculidae). Bull. Mar. Sci. 42, 1–15.
Rice, M.E., 1989. Comparative observations of gametes, fertilization, and maturation in sipunculans. In: 23rd European Biology Symposium. Olsen and Olsen, Fredensborg, Denmark, pp. 167–182.
Rice, M.E., 1993. Sipuncula. In: Harrison, F.W., Rice, M.E. (Eds.), Microscopic Anatomy of Invertebrates, Onychophora, Chilopoda, and Lesser Protostomata, vol. 12. Wiley-Liss, New York, pp. 237–325.

Riesgo, A., Andrade, S.C., Sharma, P.P., Novo, M., Pérez-Porro, A.R., Vahtera, V., González, V.L., Kawauchi, G.Y., Giribet, G., 2012. Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. Front. Zool. 9, 33.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.
Roth, A.C., Gonnet, G.H., Dessimoz, C., 2008. Algorithm of OMA for large-scale orthology inference. BMC Bioinform. 9, 518.
Roure, B., Baurain, D., Philippe, H., 2013. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. Mol. Biol. Evol. 30, 197–214.
Ryan, J.F., Pang, K., Schnitzler, C.E., Nguyen, A.D., Moreland, R.T., Simmons, D.K., Koch, B.J., Francis, W.R., Havlak, P., Smith, S.A., Putnam, N.H., Haddock, S.H.D., Dunn, C.W., Wolfsberg, T.G., Mullikin, J.C., Martindale, M.Q., Baxevanis, A.D., 2013. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. Science 342, 1242592.
Saiz-Salinas, J.I., 1993. Sipuncula. Museo Nacional de Ciencias Naturales, CSIC, Madrid.
Scheltema, R.S., Hall, J.R., 1965. Trans-oceanic transport of sipunculid larvae belonging to genus *Phascolosoma*. Am. Zool. 5, 100.
Scheltema, R.S., Hall, J.R., 1975. The dispersal of pelagosphaera larvae by ocean currents and the geographical distrubution of sipunculans. In: Rice, M.E., Todorovíc, M. (Eds.), Proceedings of The International Symposium on the Biology of the Sipunculan and Echiura Kotor, Yugoslavia, pp. 103–115.
Scheltema, R.S., Rice, M.E., 1990. Occurrence of teleplanic pelagosphera larvae of sipunculans in tropical regions of the Pacific and Indian Oceans. Bull. Mar. Sci. f47, 159–181.
Schulze, A., Rice, M.E., 2009a. Musculature in sipunculan worms: ontogeny and ancestral states. Evol. Dev. 11, 97–108.
Schulze, A., Rice, M.E., 2009b. *Nephasoma pellucidum*: a model species for sipunculan development? Smith. Contrib. Mar. Sci. 38, 209–217.
Schulze, A., Cutler, E.B., Giribet, G., 2005. Reconstructing the phylogeny of the Sipuncula. Hydrobiologia 535 (536), 277–296.
Schulze, A., Cutler, E.B., Giribet, G., 2007. Phylogeny of sipunculan worms: a combined analysis of four gene regions and morphology. Mol. Phylogenet. Evol. 42, 171–192.
Schulze, A., Maiorova, A., Timm, L.E., Rice, M.E., 2012. Sipunculan larvae and "cosmopolitan" species. Int. Comp. Biol. 52, 497–510.
Sedgwick, A., 1898. Sipunculoidea (Gephyrea, Achaeta). A Student's Textbook of Zoology. Swan Sonnenschein, London, pp. 534–539.
Smith, S.A., Dunn, C.W., 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. Bioinformatics 24, 715–716.
Smith, S., Wilson, N.G., Goetz, F., Feehery, C., Andrade, S.C.S., Rouse, G.W., Giribet, G., Dunn, C.W., 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. Nature 480, 364–367.
Staton, J.L., 2003. Phylogenetic analysis of the mitochondrial cytochrome c oxidase subunit 1 gene from 13 sipunculan genera: intra- and interphylum relationships. Invertebr. Biol. 122, 252–264.
Staton, J.L., 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.
Stephen, A.C., Edmonds, S.J., 1972. The Phyla Sipuncula and Echiura. Trustees of the British Museum (Natural History), London.
Strand, M., Samuelsson, H., Sundberg, P., 2010. Nationalnyckeln till Sveriges flora och fauna. Stjärnmaskar-Slemmaskar, Sipuncula-Nemertea. ArtDatabanken, SLU, Uppsala.
Struck, T.H., Schult, N., Kusen, T., Hickman, E., Bleidorn, C., McHugh, D., Halanych, K.M., 2007. Annelid phylogeny and the status of Sipuncula and Echiura. BMC Evol. Biol. 7, 11.
Struck, T.H., Paul, C., Hill, N., Hartmann, S., Hösel, C., Kube, M., Lieb, B., Meyer, A., Tiedemann, R., Purschke, G., Bleidorn, C., 2011. Phylogenomic analyses unravel annelid evolution. Nature 471, 95–98.
Valentine, J.W., 1997. Cleavage patterns and the topology of the metazoan tree of life. Proc. Nat. Acad. Sci. USA 94, 8001–8005.
von Reumont, B.M., Jenner, R.A., Wills, M.A., Dell'Ampio, E., Pass, G., Ebersberger, I., Meyer, B., Koenemann, S., Iliffe, T.M., Stamatakis, A., Niehuis, O., Meusemann, K., Misof, B., 2012. Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the possible sister group of Hexapoda. Mol. Biol. Evol. 29, 1031–1045.
Wanninger, A., 2008. Comparative lophotrochozoan neurogenesis and larval neuroanatomy: recent advances from previously neglected taxa. Acta Biol. Hung. 59 (1), 127–136.
Wanninger, A., Koop, D., Bromham, L., Noonan, E., Degnan, B.M., 2005. Nervous and muscle system development in *Phascolion strombus* (Sipuncula). Dev. Genes Evol. 215, 509–518.
Wanninger, A., Kristof, A., Brinkmann, N., 2009. Sipunculans and segmentation. Commun. Integr. Biol. 2, 56–59.
Weigert, A., Helm, C., Meyer, M., Nickel, B., Arendt, D., Hausdorf, B., Santos, S.R., Halanych, K.M., Purschke, G., Bleidorn, C., Struck, T.H., 2014. Illuminating the base of the annelid tree using transcriptomics. Mol. Biol. Evol. 31, 1391–1401.
Wheat, C.W., Wahlberg, N., 2013. Phylogenomic insights into the Cambrian Explosion, the colonization of land and the evolution of flight in Arthropoda. Syst. Biol. 62, 93–109.
Wu, Z.P., Wang, X., Zhang, X.G., 2011. Using non-uniform read distribution models to improve isoform expression inference in RNA-Seq. Bioinformatics 27, 502–508.
Wu, M., Chatterji, S., Eisen, J.A., 2012. Accounting for alignment uncertainty in phylogenomics. PLoS ONE 7, e30288.
Young, C.M., He, R., Emlet, R.B., Li, Y., Qian, H., Arellano, S.M., Van Gaest, A., Bennett, K.C., Wolf, M., Smart, T.I., Rice, M.E., 2012. Dispersal of deep-sea larvae from the intra-American seas: simulations of trajectories using ocean models. Integr. Comp. Biol. 52, 483–496.
Zapata, F., Wilson, N.G., Howison, M., Andrade, S.C.S., Jörger, K.M., Schrödl, M., Goetz, F.E., Giribet, G., Dunn, C.W., 2014. Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. Proc. Roy. Soc. B. 281, 20141739.
Zoller, S., Schneider, A., 2013. Improving phylogenetic inference with a semi empirical amino acid substitution model. Mol. Biol. Evol. 30, 469–479.