Molecular Phylogenetics and Evolution 85 (2015) 221–229

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Mitogenomics reveals phylogeny and repeated motifs in control regions of the deep-sea family Siboglinidae (Annelida)

Yuanning Li [a,*], Kevin M. Kocot [a,b], Christoffer Schander [c], Scott R. Santos [a], Daniel J. Thornhill [a], Kenneth M. Halanych [a,*]

[a] Department of Biological Sciences & Molette Biology Laboratory for Environmental and Climate Change Studies, Auburn University, Auburn, AL 36849, USA
[b] School of Biological Sciences, University of Queensland, St. Lucia, QLD 4067, Australia
[c] University of Bergen Museum, N-5020 Bergen, Norway

ARTICLE INFO

Article history:
Received 3 June 2014
Revised 6 December 2014
Accepted 13 February 2015
Available online 24 February 2015

Keywords:
Mitogenomic
Osedax
Control region
Size variations

ABSTRACT

Deep-sea tubeworms in the annelid family Siboglinidae have drawn considerable interest regarding their ecology and evolutionary biology. As adults, they lack a digestive tract and rely on endosymbionts for nutrition. Moreover, they are important members of chemosynthetic environments including hydrothermal vents, cold seeps, muddy sediments, and whale bones. Evolution and diversification of siboglinids has been associated with host-symbiont relationships and reducing habitats. Despite their importance, the taxonomy and phylogenetics of this clade are debated due to conflicting results. In this study, 10 complete and 2 partial mitochondrial genomes and one transcriptome were sequenced and analyzed to address siboglinid evolution. Notably, repeated nucleotide motifs were found in control regions of these mt genomes, which may explain previous challenges of sequencing siboglinid mt genomes. Phylogenetic analyses of amino acid and nucleotide datasets were conducted in order to infer evolutionary history. Both analyses generally had strong nodal support and suggest *Osedax* is most closely related to the Vestimentifera + *Sclerolinum* clade, rather than Frenulata, as recently reported. These results imply *Osedax*, the only siboglinid lineage with heterotrophic endosymbionts, evolved from a lineage utilizing chemoautotrophic symbionts.

1. Introduction

Deep-sea tubeworms, Siboglinidae, are annelids that typically lack a digestive tract and instead rely on endosymbionts for nutrition. Because of their highly unusual morphology and nutritional mode with respect to other annelids, siboglinids were previously classified in the phyla Pogonophora and Vestimentifera. To date, approximately 180 species of Siboglinidae have been described, with four major siboglinid lineages being recognized: Vestimentifera, Frenulata, *Sclerolinum* (Monilifera), and *Osedax* (Fig. 1). Each group is generally associated with a specific habitat type (Schulze and Halanych, 2003; Halanych, 2005; Thornhill et al., 2008; Hilário et al., 2011), with Vestimentiferans typically living in reducing habitats (e.g. hydrothermal vents or cold seeps), frenulates on muddy sediments, *Sclerolinum* species found on decaying organic matter (e.g. woods or ropes) or mud volcanoes (Hilário et al., 2011) and *Osedax* inhabiting large vertebrate (i.e., whale) bones (Rouse et al., 2004; Glover et al., 2005). While the first three of these groups host chemoautotrophic gammaproteobacteria as endosymbionts (Thornhill et al., 2008), with the exception of methanotrophs in *Siboglinum poseidoni* (Schmaljohann, 1991). Heterotrophic gammaproteobacteria belonging to the order Oceanospirillales are harbored by *Osedax* species (Goffredi et al., 2005).

To date, morphological (Rouse, 2001; Schulze, 2003) and molecular (Halanych et al., 2001; Rouse et al., 2004; Rousset et al., 2004; Glover et al., 2005, 2013) approaches applied towards understanding siboglinid phylogeny suggest: (1) siboglinids form a monophyletic clade, (2) Vestimentifera *Sclerolinum*, *Osedax* and Frenulata are all monophyletic lineages, and (3) *Sclerolinum* is sister to Vestimentifera. Nonetheless, important aspects of siboglinid evolutionary history are still debated. A recent study (Glover et al., 2013; Rouse et al., 2015), using combined nuclear small subunit (*18S*) ribosomal DNA (rDNA), 28S (*28S*) rDNA, histone H3 (*H3*) gene, mitochondrial large subunit (*16S*) rDNA and cytochrome oxidase subunit I (*COI*) data inferred *Osedax* as sister to Frenulata

* Corresponding authors at: 101 Rouse Life Sciences Bldg., Auburn University, AL 36849, USA. Fax: +1 334 844 2333.
E-mail addresses: yzl0084@auburn.edu (Y. Li), k.kocot@uq.edu.au (K.M. Kocot), santos@auburn.edu (S.R. Santos), thornhill.dan@gmail.com (D.J. Thornhill), ken@auburn.edu (K.M. Halanych).

http://dx.doi.org/10.1016/j.ympev.2015.02.008
1055-7903/

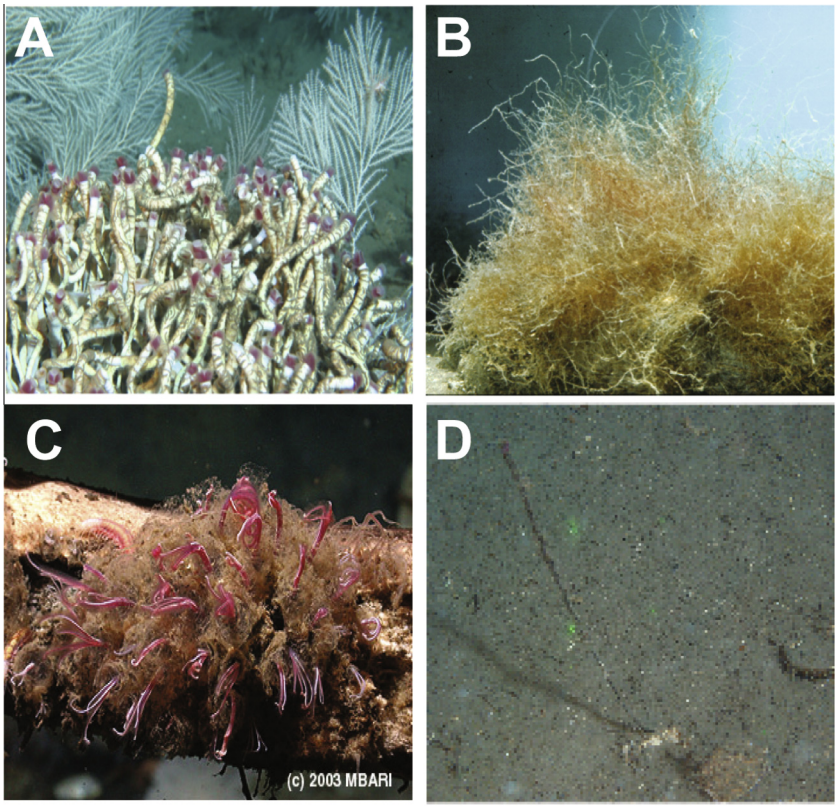

Fig. 1. Major siboglinid lineages and their habitat preferences. (A) *Lamellibrachia* growing near a hydrocarbon seep. (B) *Sclerolinum* inhabiting decaying ropes (Image courtesy of Eve Southward). (C) Bone-eating *Osedax* worms living on a piece of dead gray whale bone in Monterey Canyon (Image courtesy of Monterey Bay Aquarium Research Institute). (D) Frenulata species growing in deep-sea muddy habitats.

rather than the Vestimentifera + *Sclerolinum* clade, in contrast to previous reports (Rouse et al., 2004; Glover et al., 2005). Furthermore, habitat preference has been hypothesized to have proceeded from deep-sea muddy environments (similar to where frenulates inhabit) to more specialized reducing environments such as hydrocarbon cold seeps and hydrothermal vents (Schulze and Halanych, 2003). However, nodal support values within Vestimentifera have been generally low (Halanych, 2005), obscuring our understanding of habitat shifts as well as the evolutionary history of the group in general. Thus, important aspects of siboglinid evolution are still elusive and additional phylogenetic analyses are needed towards elucidating them.

Phylogenetic analyses of mitochondrial (mt) genomes have proven useful in resolving phylogenetic relationships across a wide range of metazoans (e.g. Osigus et al., 2013; Miya et al., 2001). In Bilateria, mt genomes are circular, usually range from 14 to 17 kb (but see Shao et al., 2009; Osigus et al., 2013; Boore, 1999) and typically possess 37 genes: 13 protein-coding genes (i.e., *ATP6*, *ATP8*, *COX1–3*, *COB*, *NAD1–6* and *NAD4l*), two ribosomal RNA genes, 22 tRNA genes, and the control region (also called the unknown [UNK] region or D-loop). Within Annelida, arrangement of these genes is relatively conserved (Jennings and Halanych, 2005; Vallès and Boore, 2006; Zhong et al., 2008), but, as of March 2014, complete mitochondrial genomes were only publicly available (i.e., in GenBank) for 17 annelid species. Furthermore, only two partial siboglinid mitochondrial genomes have so far been sequenced: *Galathealinum brachiosum* (Boore and Brown, 2000) and *Riftia pachyptila* (Jennings and Halanych, 2005). In both cases, difficulties with recovering the control region, despite considerably effort, were reported. This region of the mt genome putatively plays a role in controlling transcription and replication of mitochondrial genes (Shadel and Clayton, 1997; Boore and Brown, 2000), which may explain this situation.

Recent advances in high-throughput sequencing and bioinformatics allow novel approaches to sequencing whole mt genomes. To further explore siboglinid phylogeny, including placement of *Osedax* as either sister to Vestimentifera + *Sclerolinum* or Frenulata, and to understand the structure of siboglinid mitochondrial control region, we sequenced mt genomes from representatives of all major siboglinid lineages. These efforts included 10 complete and two partial mt genomes, and well as one transcriptome, to explore siboglinid evolutionary history.

2. Materials and methods

2.1. Specimen collection and mitochondrial genome sequencing

Specimen information is shown in Table 1. All were either stored frozen at −80 °C or preserved in 80–100% non-denatured ethanol following collection. Due to a limiting amount of tissue from *Osedax mucofloris*, only RNA was extracted since (1) mitochondrial protein-coding and ribosomal RNA genes, which were used in reconstructing the phylogeny of this family, can be recovered from whole transcriptome sequencing (Neto et al., 2000) and (2) gene order of mt genomes within siboglinids is highly conserved (see Sections 3 and 4). RNA was extracted from *Osedax mucofloris* using TRIzol (Invitrogen) and purified using the RNeasy kit (Qiagen) with on-column DNase digestion. Complimentary DNA (cDNA) libraries were constructed using the SMART cDNA library construction kit (Clontech). Total genomic DNA was extracted from all other samples using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's protocols.

For *Escarpia spicata*, *Seepiophila jonesi*, *Galathealinum* sp. and *O. mucofloris*, sequencing of genomic DNA or cDNA were performed

Table 1
Specimen data for sequenced taxa.

Species	Clade	Specimen collection		
		Location	Depth (m)	GPS coordinates
Riftia pachyptila	Vestimentifera	East Pacific Rise	~2522	N9°50.89′ W104°17.49′
Tevnia jerichonana	Vestimentifera	East Pacific Rise	~2537	N9°47.13′ W104°16.13′
Oasisia alvinae	Vestimentifera	East Pacific Rise	~2630	N9°48.12′ W103°56.12′
Ridgeia piscesae	Vestimentifera	Hulk, Canada	2190	N47°56.95′ W104°16.87′
Seepiophila jonesi	Vestimentifera	Mississippi Canyon, US	754	N28°11.58′ W89°47.94′
Escarpia spicata	Vestimentifera	Mississippi Canyon, US	754	N28°11.58′ W89°47.94′
Lamellibrachia luymesi	Vestimentifera	Mississippi Canyon, US	754	N28°11.58′ W89°47.94′
Sclerolinum brattstromi	Monilifera	Storfjorden Fjord, Norway	660	N62°27.26′ E6°47.57′
Siboglinum fiordicum	Frenulata	Skoge Inlet, Norway	36	N60°16.17′ E5°05.53′
Spirobrachia sp.	Frenulata	Aleutian Trench, US	4890	N57°27.39′ W148°00.01′
Galathealinum sp.	Frenulata	Mississippi Canyon, US	754	N28°11.58′ W89°47.94′
Siboglinum ekmani	Frenulata	Storfjorden Fjord, Norway	515	N62°23.30′ E6°54.58′
Osedax mucofloris	*Osedax*	Near Bergen, Norway	N/A	On artificial whale fall

by The Genomic Services Lab at the Hudson Alpha Institute in Huntsville, Alabama using Illumina (San Diego, California) 2 × 100 paired-end TruSeq protocols on a Illumina HiSeq 2000 platform. For *Tevnia jerichonana*, *Oasisia alvinae*, *Ridgeia piscesae*, *Sclerolinum brattstromi*, *Siboglinum fiordicum*, *Siboglinum ekmani*, sequencing libraries were prepared from total genomic DNA using Illumina's Nextera DNA sample preparation kit and run on an Illumina Miseq sequencer with 2 × 250 paired-end reads in the Molette Laboratory, Department of Biological Sciences Department, Auburn University.

2.2. Mitochondrial genome assemblies and annotation

All Illumina paired-end genomic sequence data were assembled *de novo* using Ray 2.2.0 (Boisvert et al., 2010) with k-mer = 31 (value chosen based on comparing a range of k-mer values relative to final assembly). To identify putative mt contigs, BLASTN (Altschul et al., 1997) was performed on contigs produced by Ray using the partial mt genome from *Riftia pachyptila* (GenBank Accession AY741662, Jennings and Halanych, 2005) as the query sequence. In 10 out 12 cases, the top-hitting contigs identified via BLASTN represented the entire mitochondrial genome (~15 kbp; Table 2). For *R. piscesae* and *S. ekmani*, however, 2 and 3 partial contigs were recovered, respectively. In an attempt to join these partial contigs, Price 1.0.1 (Ruby et al., 2013) was employed to extend existing contigs by iteratively adding raw sequence reads to the contig ends as appropriate using default settings. For *R. piscesae*, two mt contigs increased in size by 14 bp and 89 bp, respectively. In the case of *S. ekmani*, two contigs were bridged together into a single contig. The raw paired-end reads were then remapped to each of their respective draft mt genomes using Bowtie 2–2.2.1 (Langmead and Salzberg, 2012), and visualized in Tablet 1.13.12.17 (Milne et al., 2013).

Annotation of the 13 protein-coding genes (*ATP6*, *ATP8*, *COX1–3*, *COB*, *NAD1–6* and *NAD4l*) and two ribosomal RNAs was conducted with MITOS web server (Bernt et al., 2013) while tRNAs identified via the tRNAscan-SE web server (Lowe and Eddy, 1997; Schattner et al., 2005). Gene boundaries were examined and subsequently adjusted manually by comparison with sequenced siboglinid mt genomes (*Riftia pachyptila*, Jennings and Halanych, 2005 and *Galathealinum brachiosum*, Boore and Brown, 2000) in Artemis (Rutherford et al., 2000), with boundaries of control regions in each mt genome inferred by identifying flanking tRNA sequences. Sequences of control regions were tested for potential tandem repeats by RepeatMasker open-4.0.3 (Smit et al., unpublished data, www.repeatmasker.org, last accessed 01.08.14). Secondary structures of putative control regions and their thermodynamic properties were predicted using the mfold web server v2.3 (Zuker, 2003; mfold.rna.albany.edu).

To assemble the *Osedax* transcriptome, raw paired-end reads were first digitally normalized to a k-mer coverage of 30 using the normalize-by-median.py script (Brown et al., 2012; this step discards redundant data, thus decreasing memory usage). Remaining reads were then assembled using Trinity r2013-02-25 (Grabherr et al., 2011) with default settings. Mitochondrial protein-coding genes and ribosomal RNAs were identified by TBLASTX and BLASTN (Altschul et al., 1997), respectively (using the recovered siboglinid mt genomes above as queries). Since transcriptome data were used for *O. mucofloris*, tRNAs sequences as well as specific gene order information are lacking for this taxon. GenBank accession numbers for the above complete and partial mt genomes are provided in Table 2. Mitochondrial genome

Table 2
Taxa used in phylogenetic analysis.

Species	Clade	Mt genome size	GenBank number
Riftia pachyptila	Siboglinidae, Vestimentifera	14,987 complete	KJ789166
Riftia pachyptila (Genbank)	Siboglinidae, Vestimentifera	12,016 partial	AY741662
Tevnia jerichonana	Siboglinidae, Vestimentifera	14,891 complete	KJ789172
Oasisia alvinae	Siboglinidae, Vestimentifera	14,849 complete	KJ789164
Ridgeia piscesae	Siboglinidae, Vestimentifera	14,146 partial	KJ789165
Seepiophila jonesi	Siboglinidae, Vestimentifera	15,092 complete	KJ789168
Escarpia spicata	Siboglinidae, Vestimentifera	15,445 complete	KJ789161
Lamellibrachia luymesi	Siboglinidae, Vestimentifera	14,991 complete	KJ789163
Sclerolinum brattstromi	Siboglinidae, *Sclerolinum*	15,383 complete	KJ789167
Osedax mucofloris	Siboglinidae, *Osedax*	N/A	[a]
Siboglinum fiordicum	Siboglinidae, Frenulata	19,502 complete	KJ789170
Spirobrachia sp.	Siboglinidae, Frenulata	15,581 complete	KJ789171
Galathealinum sp.	Siboglinidae, Frenulata	14,779 complete	KJ789162
Galathealinum brachiosum (Genbank)	Siboglinidae, Frenulata	7568 partial	AF178679
Siboglinum ekmani	Siboglinidae, Frenulata	14,838 partial	KJ789169
Helobdella robusta	Hirudinea, Glossiphoniidae	7553 partial	AF178680
Lumbricus terrestris	Oligochaeta, Lumbricidae	14,998 complete	NC_001673
Orbinia latreillii	Scolecida, Orbiniidae	15,558 complete	NC_007933
Sipunculus nudus	Polychaeta, Sipunculidae	15,502 complete	NC_011826

[a] GenBank Numbers of 11 protein-coding and 1 ribosomal RNA genes of O. mucofloris are KJ806974, KJ806975, KJ806976, KJ806977, KJ806978, KJ806979, KJ806980, KJ806981, KJ806982, KJ806983, KJ806984 and KJ806985, respectively.

sequences (accession KJ789161–KJ789172) and individual *Osedax* genes (accession KJ806974–KJ806985) have been deposited to GenBank.

2.3. Southern blot

Based on results from the mt genome assemblies (see Sections 3 and 4), Southern blot analyses were conducted on a subset of the examined taxa to confirm whether mitochrondrial genomes were circular, rather than linear, in nature. Genomic DNA from *R. pachyptila*, *L. luymesi*, *E. spicata* and *Galathealinum* sp. was submitted to Transviragen in Chapel Hill, North Carolina for Southern blotting. Individuals of other species were too small to provide sufficient amounts of high quality DNA for Southern blotting. For each species, restriction enzymes that would only cut the mt genome once were identified. Based on mtDNA sequencing in hand, we used the enzymes *SbfI*, *Drd I*, *BamHI*, and *NDel* for *R. pachytila*, *L. luymesi*, *E. spicata*, and *Galathealinum* sp., respectively. With these enzyme-species combinations, a circular mitochondrial genome would produce a single band of roughly 15 Kb, whereas a linear genome would produce two bands in which the largest would be no more than 10.3 Kb in length. Whole genomic DNA was restricted to completion and run on a 0.7% TAE agarose gel adjacent to an unrestricted sample. The agarose gel was then blotted to a nitrocellulose filter and probed with an oligonucleotide designed to hybridize to CO1. The blot was visualized using the PCR DIG Probe Synthesis Kit (Roche).

2.4. Phylogenetic analyses

Nineteen Operational Taxonomic Units (OTUs) were included in the phylogenetic analyses. In addition to sequence data generated here, two partial siboglinid mitochondrial genome and four outgroup species were acquired from GenBank (Table 2). *Helobdella robusta*, *Lumbricus terrestris*, *Orbinia latreillii* and *Sipunculus nudus* were selected as outgroups based on data availability as well as current understanding of annelid evolutionary history (Struck et al., 2007, 2011; Weigert et al., 2014).

Two data sets were constructed – one being amino acid (AA) and the other being nucleotide (NUC) sequences. Nucleotide sequences were converted into amino acids using the standard invertebrate mitochondrial translation code implemented in MEGA 5.2 (Tamura et al., 2011). For amino acid and nucleotide data, each gene was individually aligned in MUSCLE 3.8.31 (Edgar, 2004), followed by manual correction. All genes were trimmed using the default settings in Gblocks 0.91b (Talavera and Castresana, 2007) to remove ambiguously aligned regions. Genes were then concatenated into final datasets using FASconCAT (Kück and Meusemann, 2010) for phylogenetic analysis. The NUC dataset consisted of nucleotide sequences of the 13 protein-coding and the 2 ribosomal RNA genes while the AA dataset included the amino acids sequences of the 13 protein-coding genes only.

Phylogenetic relationships of siboglinids were inferred using maximum likelihood (ML) in RAxML 7.3.8 (Stamatakis, 2006) and Bayesian inference (BI) in PhyloBayes MPI 1.4f (Lartillot et al., 2009). For phylogenetic analyses, ProtTest 2.4 (Abascal et al., 2005) was performed to evaluate all evolutionary models, however, since the MtZoa evolutionary model (Rota-Stabelli et al., 2009) for amino acid data is not available on ProtTest, we evaluated tree topologies based on MtZoa and MtArt + I + G (the best-fit model according to ProtTest) separately, and MtZoa was chosen as the best-fit model because it provided better likelihood scores and less computational time.

For ML analyses, both NUC and AA datasets were partitioned by gene. Analysis of the NUC dataset was done under the GTR (general time reversible) model of substitution rate with a gamma distribution (the GTRGAMMA option) while the AA dataset was analyzed using MtZoa model with a gamma distribution using empirical base frequencies (the PROTGAMMAMTZOAF option). Topological robustness for the ML analysis was evaluated with 100 replicates of nonparametric bootstrapping. Competing phylogenetic hypotheses for both datasets were evaluated using the Approximately Unbiased (AU) test (Shimodaira, 2002) in CONSEL 0.20 (Shimodaira and Hasegawa, 2001). Per site likelihoods values were determined using RAxML with same evolutionary models.

For BI analyses, the CAT model (Lartillot and Philippe, 2004) was employed for both NUC and AA datasets. The CAT model in Phylobayes is a site heterogeneous model that estimates site-specific substitution rates for the 4 nucleotides or 20 amino acids in an alignment. Thus, for BI analyses, the CAT + GTR and CAT + MtZoa models were employed in analyses of the NUC and AA datasets, respectively. Five parallel chains were each run for 25,000 generations, discarding the first 5000 generations as burn-in based on log likelihood scores for each chain once stationary was reached. A 50% majority rule consensus tree was computed from the remaining 20,000 trees from each chain, and nodal support was estimated in the post-burnin tree sample, with posterior probability values $\geqslant 0.95$ taken as significant (Huelsenbeck and Rannala, 2004). All phylogenetic analyses were conducted on the Auburn University Molette Laboratory's SkyNet server.

3. Results

3.1. Mt genome composition

Results from the high-throughput sequencing and contig assembly for the 12 mitochondrial genomes and transcriptome of *O. mucofloris* (missing the *NAD4l*, *ATP8* and small subunit *12S-rDNA* genes)

Table 3
Sequencing information.

Species	Number of mt contig(s) recovered	Total reads of mt contig(s)	Average sequencing depth of control region (X)	Average sequencing depth of mt genome (X)
Vestimentifera				
Riftia pachyptila	1	4632	45	71
Tevnia jerichonana	1	1484	5	22
Oasisia alvinae	1	2802	9	39
Ridgeia piscesae	2	1883	22	29
Seepiophila jonesi	1	32,811	98	220
Escarpia spicata	1	35,127	133	232
Lamellibrachia luymesi	1	4193	24	67
Sclerolinum				
Sclerolinum brattstromi	1	1504	21	23
Osedax				
Osedax mucofloris	12	120,480	N/A	802
Frenulata				
Siboglinum fiordicum	1	11,297	213	90
Spirobrachia sp.	1	8706	104	45
Galathealinum sp.	1	38,282	103	261
Siboglinum ekmani	3	1969	35	36

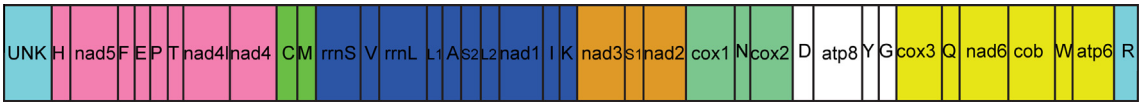

Fig. 2. Gene orders of mitochondrial genomes in all Siboglinidae sampled to date. Different colors show conserved gene clusters that were previously reported (Zhong et al., 2008). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Base composition and skewness measures.

Species	A + T%		3rd position%				GC-skew			AT-skew		
	Whole genome	Protein-coding genes	A	T	C	G	Whole genome	Protein-coding genes	3rd position	Whole genome	Protein-coding genes	3rd position
Vestimentifera												
Riftia pachyptila	66.76	66.31	42.13	36.81	18.68	2.39	−0.26	−0.31	−0.77	−0.06	−0.11	0.07
Tevnia jerichonana	64.51	63.53	39.47	32.57	24.51	3.18	−0.3	−0.35	−0.77	−0.06	−0.1	0.1
Oasisia alvinae	64.92	64.35	39.73	34.11	22.58	3.58	−0.28	−0.32	−0.73	−0.06	−0.11	0.08
Ridgeia piscesae	63.11	62.29	38.55	30.08	27.20	4.18	−0.05	−0.35	−0.73	−0.32	−0.09	0.12
Seepiophila jonesi	64.83	64.17	38.98	34.58	22.50	3.94	−0.29	−0.33	−0.7	−0.06	−0.11	0.06
Escarpia spicata	66.09	65.02	40.34	35.61	20.71	3.34	−0.26	−0.33	−0.72	−0.07	−0.11	0.06
Lamellibrachia luymesi	64.56	63.77	39.48	34.19	23.74	2.58	−0.3	−0.36	−0.8	−0.05	−0.11	0.07
Sclerolinum												
Sclerolinum brattstromi	64.9	64.46	39.79	32.63	23.45	4.13	−0.3	−0.33	−0.7	−0.05	−0.09	0.1
Osedax												
Osedax mucofloris	N/A	62.88	36.77	28.46	28.98	5.78	N/A	−0.35	−0.66	N/A	−0.12	0.13
Frenulata												
Siboglinum fiordicum	74.49	71.18	31.5	48.32	15.75	4.43	−0.27	−0.29	−0.56	−0.13	−0.25	−0.21
Spirobrachia sp.	74.57	73.57	39.21	45.68	12.62	2.5	−0.23	−0.02	−0.67	−0.11	−0.02	−0.08
Galathealinum sp.	77.93	77.07	39.03	52.29	7.05	1.63	−0.17	−0.83	−0.62	−0.14	−0.2	−0.15
Siboglinum ekmani	75.02	73.59	38.71	45.27	13.57	2.45	−0.25	−0.31	−0.69	−0.1	−0.14	−0.08

are presented in Table 3. Two mt genomes were partial; *R. piscesae* (14,146 bp) was missing the 12S-rDNA and trnV gene and *S. ekmani* (14,838 bp) was missing trnR as well as part of atp6 and the control region. Nine of the complete mt genomes were ~15 kbp in length, varying in size from 14,779 bp (*Galathealinum* sp.) to 15,581 bp (*Spirobrachia* sp.). However, *S. fiordicum* has a substantially larger mt genome (19,502 bp, see below). All genomes had the same gene order (Fig. 2; gene order unknown for *Osedax*).

All the mt genomes exhibited nucleotide and codon biases (Table 4). For example Vestimentifera, *Osedax*, and *Sclerolinum* were ~65% AT whereas Frenulata was ~75% AT. In all of these species, T was the most common base and G the least common base on the coding strand. Furthermore, the anti-G bias was especially pronounced in the third codon position, where G was only present at 1.63% (*Galathealinum* sp.) to 5.78% (*O. mucofloris*). GC-skew and AT-skew for a given strand were calculated as $(G - C)/(G + C)$ and $(A - T)/(A + T)$, respectively (Perna and Kocher, 1995), with negative values in skewness meaning the coding strand is enriched for T or C. In contrast, positive values infer more As and Gs. On the whole, AT-skew was slightly negative, or positive in the third codon position of vestimetiferans, and GC-skew was more negative than AT-skew (Table 4).

3.2. Protein-coding genes

All complete siboglinid mt genomes in this study possessed the 13 protein-coding genes, two ribosomal RNA genes, and 22 tRNA genes (Fig. 2) that are typical of bilaterian mt genomes (Boore, 1999). As in previously reported siboglinid (Boore and Brown, 2000; Jennings and Halanych, 2005) and other annelid mt genomes (e.g., Boore and Brown, 1995, 2000; Jennings and Halanych, 2005; Zhong et al., 2008; Shen et al., 2009), mitochondrial genes sequenced herein are transcribed from the same strand and gene order was conserved. The start and stop codon features of siboglinid mt genomes also showed patterns of bias (Table S1). For example, only ATG is used as an initiation codon, whereas most metazoan mt genomes use a combination of codons (i.e., ATA, ATC, GTG, GCC and GTT; Boore and Brown, 2000; Zhong et al., 2008). Most genes end with the stop codon TAA or TAG. Nevertheless, an incomplete termination codon, either a single T or TA, was observed for several protein gene sequences (Table S1).

3.3. Control region

Putative control regions were identified in 12 mt genomes, occurring between *trnR* and *trnH* (Table 5). Although the exact boundaries for this region are difficult to precisely define, all sequences are AT-rich and contain simple repetitive or microsatellite-like motifs (also see Table 5). Notably, lengths of control regions are highly variable in size, ranging from 186 bp (*T. jerichonana*) to 4,737 bp (*S. fiordicum*). To further investigate size variability in the control region, raw paired-end reads were remapped to the putative control region (Table 3). In general, presence of repetitive motifs and hairpin-like secondary structures reduced read coverage in most cases. However, *S. fiordicum* putatively showed higher than average coverage in the control region. For example, one ~400 nt sequence was mapped at >2000× coverage compared to the 213× average across

Table 5
Structural features of control region.

Species	Size (bp) of control region	(A + T)%	Proportion of the mt genome (%)	Repeat motifs
Vestimentifera				
Riftia pachyptila	303	81.19	2.02	$(TA)_n$
Tevnia jerichonana	186	81.72	1.25	$(TA)_n$
Oasisia alvinae	147	66.67	0.99	$(TA)_n$
Ridgeia piscesae	309	67.31	2.18	$(TA)_n$
Seepiophila jonesi	381	76.12	2.52	$(TA)_n$
Escarpia spicata	741	82.05	4.8	$(TA)_n$; $(TATATG)_n$
Lamellibrachia luymesi	309	80.91	2.06	$(TA)_n$
Sclerolinum				
Sclerolinum brattstromi	654	63.61	4.25	$(TA)_n$
Osedax				
Osedax mucofloris	N/A	N/A	N/A	N/A
Frenulata				
Siboglinum fiordicum	4737	79.23	24.29	$(TA)_n$; $(CACA)_n$; $(CATA)_n$; $(TATATG)_n$; $(CA)_n$; $(CATATA)_n$; AT rich
Spirobrachia sp.	975	71.90	6.26	$(TA)_n$; $(GA)_n$; AT rich
Galathealinum sp.	240	87.92	1.62	$(TA)_n$; $(A)_n$; AT rich
Siboglinum ekmani	645	75.97	4.35	$(TA)_n$; AT rich

the rest of the control region. Potential secondary structure in this region has been reported in *Lumbricus terrestris* (Boore and Brown, 1995) and *Platynereis dumerilii* (Boore, 2001). However, mfold analyses failed to identify similar potential structures or even conservation of secondary structure among siboglinid mitochondrial control regions (Fig. S1).

Several of the mt genome contigs obtained from assemblies started and stopped in the control region. This could be due to the repetitive nature of the control region (repetitive elements can hinder assembly; Nagarajan and Pop, 2013) or if the molecule is linear. Either situation would offer an explanation on reported difficulties in sequencing the control region of annelids in general (see Section 1). To differentiate between these two possibilities, Southern blot experiments were performed. These experiments confirmed the circular nature of the siboglinid mt genomes (*R. pachyptila* and *E. spicata*) by exhibiting a restriction pattern consistent with a circular molecule being linearized rather than a linear molecule being cut into two fragments (Fig. S2). *Lamellibrachia luymesi* failed to cut and *Galathealinum* sp. DNA was too degraded for Southern blot analysis.

3.4. Phylogenetic analysis

The AA and NUC datasets contained 3813 and 13,923 parsimony-informative characters, respectively. Both ML and BI (Fig. 3) analyses of the two concatenated datasets yielded congruent tree topologies with high bootstrap support values (bs) or posterior probabilities (pp). In terms of higher-level relationships, both Vestimentifera and Frenulata were recovered as monophyletic clades with strong support (Fig. 3), consistent with previous molecular (Black et al., 1997; Halanych et al., 1998, 2001) and morphological analyses (Rouse, 2001; Schulze, 2003). Moreover, *Sclerolinum* was recovered sister to Vestimentifera (AA pp = 1.00, bs = 100; NUC pp = 0.98, bs = 100). Importantly, *Osedax* was placed sister to the *Sclerolinum* + Vestimentifera clade with moderate bootstrap support (bs = 90) in analyses of the AA dataset (but pp = 0.88) and stronger support (pp = 1.00, bs = 90) in analyses of the NUC dataset. Although less likely than our consensus topology (Fig. 3), the hypothesis of *Osedax* as sister to Frenulata was not explicitly rejected by AU test (Table 6). Notably, vestimentiferans generally had shorter branch lengths when compared to other clades. Within Frenulata, *Siboglinum* was not monophyletic; *Spirobrachia* and *Galatheallinum* were nested as a monophyletic clade within the paraphyletic *Siboglinum* (Fig. 3).

4. Discussion

In this study, analyses of siboglinid mitochondrial genomes support placement of *Osedax* as sister to a Vestimentiferan + *Sclerolinum* clade. Furthermore, although the overall gene order is similar to other annelids, the control regions of siboglinid mt genomes contain highly repetitive elements that generate substantial length variation among lineages.

4.1. Siboglinid phylogeny

Phylogenetic analyses supported traditional monophyletic clades of Vestimentifera and Frenulata. As seen in previous morphological (Rouse, 2001) and molecular analyses (Halanych et al., 2001; Rouse et al., 2004; Rousset et al., 2004; Glover et al., 2005, 2013), *Sclerolinum* is closely allied to Vestimentifera. Our analyses support the phylogenetic position of *Osedax* as sister to the Vestimentifera + *Sclerolinum* clade, rather than Frenulata, in agreement with previous studies based on combinations of nuclear *18S*, and mitochondrial *16S* or *COI* data (Glover et al., 2005; Rouse et al., 2004) but in contrary to the combined analysis of Glover et al. (2013) and Rouse et al. (2015). Although an AU test could not reject the hypothesis of *Osedax* as sister to Frenulata, likelihood scores and nodal support values support our consensus topology (Table 6). In addition, similar to Glover et al. (2013) and Rouse et al. (2015), placement of *Osedax* was variable in analyses based on a limited number of genes (i.e., *18S*, *16S* and *COI*; Hatleberg, Thornhill, Santos, and Halanych unpublished data). Although the mt genome is a single chromosome, it contains several genes with variable rates of evolution (Mueller, 2006). Our results imply that *Osedax*, which associates with heterotrophic endosymbionts and lives on whale bones, evolved from a lineage depending on chemoautotrophic symbionts that might have dwelled in deep-sea muddy sediments.

Within Vestimentifera, *Lamellibrachia* is sister to the remaining sampled vestimentiferans. Relatively low molecular diversity within vestimentiferans has been observed in all molecular studies to date (Halanych et al., 1998, 2001; Rouse et al., 2004; Rousset et al., 2004; Glover et al., 2005, 2013), including this study. This limited diversity may be due to selection pressures in extreme habitats or a recent evolutionary origin of this lineage (Halanych, 2005). Within Frenulata, *Siboglinum* is not monophyletic, consistent with previous molecular and morphological studies (reviewed by Halanych, 2005), suggesting that *Siboglinum* characters – for instance, possession of a single tentacle – are symplesomorphies.

4.2. Genome composition and structure

Prior to this study, only 17 complete annelid mt genomes, and two partial siboglinid mt genomes, had been reported. Siboglinid mitochondrial genomes exhibit the same gene order as most

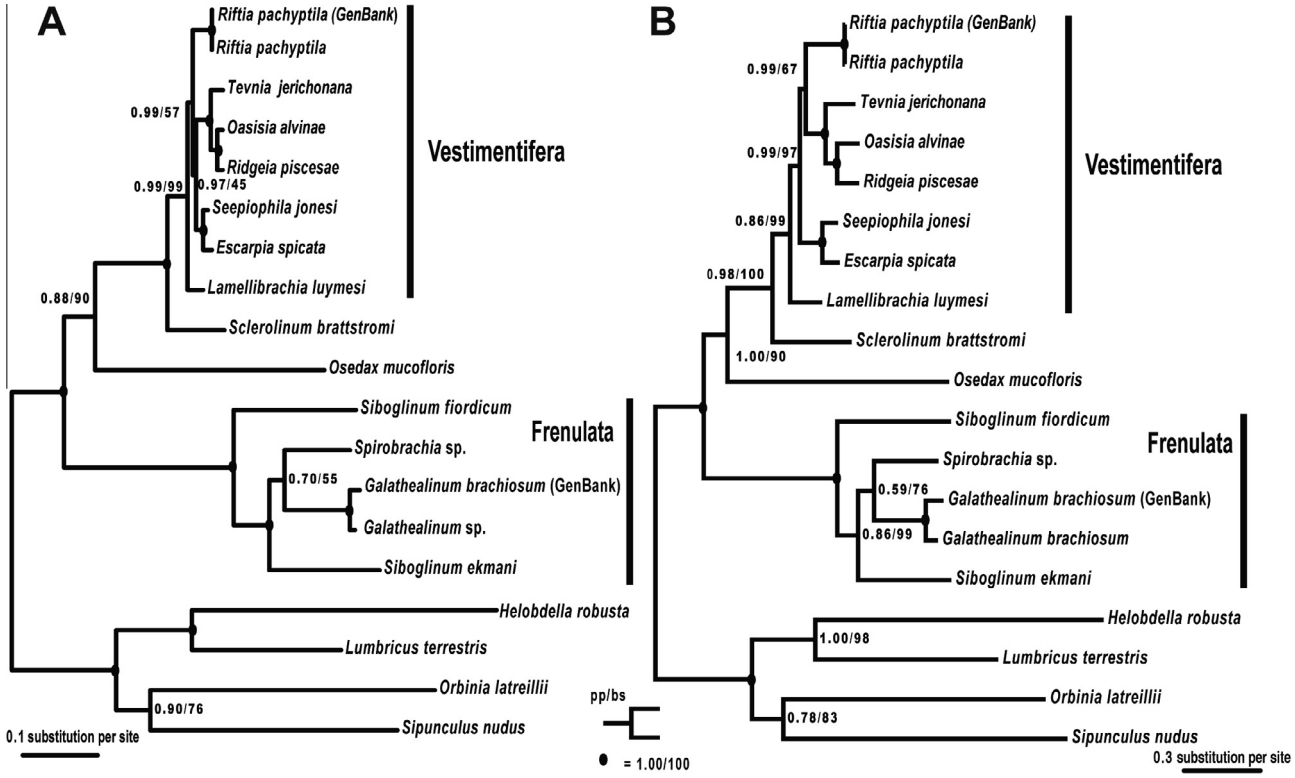

Fig. 3. Phylogenetic reconstructions of Siboglinidae based on (A) concatenated amino acids of the 13 mitochondrial protein-coding genes and (B) concatenated nucleotides of the 13 mitochondrial protein-coding and 2 ribosomal RNA genes. Majority rule (50%) consensus phylograms from each of the Bayesian analyses of the concatenated data matrices are shown. Values are shown next to nodes with posterior probabilities left and ML bootstrap support values right. Filled circles indicate fully supported nodes (bs = 100, pp = 1.00).

Table 6
AU tests of competing phylogenetic hypothesis.

Tree topology	AA dataset		NUC dataset	
	Log-likelihood	AU test (*P*-value)	Log-likelihood	AU test (*P*-value)
Osedax + Vestimentifera/*Sclerolinum*	−58555.90	0.882	−151727.36	0.914
Osedax + Frenulata	−58570.38	0.118	−151750.80	0.083

annelids (Fig. 2); this gene order could be assumed to be a synapomorphy for annelids in general (Jennings and Halanych, 2005; Zhong et al., 2008). Along with this and unlike most metazoans, all genes of annelid mt genomes are transcribed on the same strand, which was hypothesized to prevent inversions that occurred on the non-transcribed strand (Boore, 1999). Additionally, only ATG is used as a start codon, whereas most metazoan mt genomes use a variety of combinations (Boore, 1999). An incomplete stop codon, either a single T or TA, is also common for many protein-genes in the examined siboglinid mt genomes. Incomplete stop codons such as T or TA may be assigned to the adjacent down-stream gene, and then modified to a complete TAA stop codon via post-transcriptional polyadenylation (Zhong et al., 2008; Boore and Brown, 2000; Yuan et al., 2012). Similar patterns are also observed in molluscs (Hrbek and Farias, 2008; Yuan et al., 2012) and other groups (Ivey and Santos, 2007). Typically, all siboglinid mt genomes sampled to date are characterized by an anti-G bias that is relatively strong at the third codon position, where G is present at an average of only ~3.4%. The low G content may be a result of the asymmetrical replication of the mt genome (Clayton, 1982; Hrbek and Farias, 2008) or a tendency of mutational bias (Boore and Brown, 2000).

4.3. Control regions

The control region is typically the longest non-coding region in mt genomes and is believed to play a role in controlling mitochondrial replication and transcription (Clayton, 1991; Zhang and Hewitt, 1997; Boore, 1999). However, this region remains poorly characterized in annelids since previous efforts to sequence it in many annelid mt genomes have been unsuccessful (Boore and Brown, 2000; Jennings and Halanych, 2005; Zhong et al., 2008). We initially hypothesized that siboglinid mt genomes may be linear rather than circular for two reasons. First, the available published siboglinid mt genomes contained control regions that were difficult to amplify by long PCR primers, even though this technique has been successfully employed to amplify entire mt genomes in a variety of metazoan lineages (Yuan et al., 2012; Zhong et al., 2008; Shen et al., 2009). Secondly, a remapping-based approach identified highly repetitive motifs at the ends of the mt genome contigs similar to the telomeres of linear chromosomes (Zakian, 1995; Table 3). To test this hypothesis, Southern blot analyses were performed. Contrary to our predictions, Southern blots implied the mt genomes of siboglinid are circular. Thus, the most reasonable conclusion is that the control region has historically been difficult to amplify and sequence due to secondary structure formation, such as hairpins, among the tandem repeat regions.

The putative length of the control region among siboglinid mt genomes shows variability among taxa, ranging from 186 bp (*T. jerichonana*) to 4737 bp (*S. fiordicum*), which represents a nearly 25-fold difference in length. Of these cases, the unusually large control region of *S. fiordicum* may be due to false extensions occurring in *de novo* assembly due to this highly repetitive region or independent transposable elements insertions from the nuclear genome, or possibly both. Notably, repeated motifs are more

extensive, especially in *S. fiordicum*, compared to previously reported genomes (Boore and Brown, 2000; Jennings and Halanych, 2005; Zhong et al., 2008), and may again signal issues with artificial concatenation of repeats during assembly. Moreover, no obvious conserved regions or secondary structures (Fig. S2) have been observed in control regions across mt genomes from different siboglinid species and clades, except that all contained TA tandem repeats (Table 5). Since intronic microsatellites can affect gene transcription, mRNA splicing or export to cytoplasm (Li et al., 2004), these TA tandem repeats may have functional significance in the mt genomes of siboglinids. Interestingly, the presence of Type 2 transposons has been reported in other annelid mt genomes, which is virtually unknown among other bilaterians (Vallès et al., 2008). Although the present study represents a further step in the characterization of the mitochondrial control region in annelids, the function and reasons for variations or assembly artifacts in this region requires further study.

Note added in proof

Further work in our lab confirms that *Osedax rubiplumus* possesses the same mitochondrial gene order as other siboglinids.

Author contributions

All authors helped conceive and design this study. K.M.H, C.S., K.M.K and D.J.T. collected specimens. K.M.K. and Y.L. prepared and sequenced Nextera libraries using the MiSeq. Y.L. conducted the bioinformatic and phylogenetic analyses with assistance from K.M.K. and K.M.H. Y.L. prepared the figures and submitted sequences to GenBank. All authors, except C.S. (deceased), contributed in the writing of the manuscript.

Acknowledgments

We thank Miquel Arnedo and two anonymous reviewers for helpful comments on the manuscript; Damien Waits and Amanda Shaver conducted RNA extraction and cDNA library preparation of *O. mucofloris*. This study was supported by awards from the U.S. National Science Foundation (NSF) to K.M.H., S.R.S., and D.J.T. (DEB-1036537 and IOS-0843473). Yuanning Li is supported by a scholarship from the China Scholarship Council (CSC) for studying and living abroad. This is Molette Biology Laboratory contribution #35 and Auburn University Marine Biology Program contribution #126.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2015.02.008.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105. http://dx.doi.org/10.1093/bioinformatics/bti263.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids Res. 25, 3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. Mitogenom. Metazoan Evol. 69, 313–319. http://dx.doi.org/10.1016/j.ympev.2012.08.023.

Black, M.B., Halanych, K.M., Maas, P.A.Y., Hoeh, W.R., Hashimoto, J., Desbruyères, D., Lutz, R.A., Vrijenhoek, R.C., 1997. Molecular systematics of vestimentiferan tubeworms from hydrothermal vents and cold-water seeps. Mar. Biol. 130, 141–149. http://dx.doi.org/10.1007/s002270050233.

Boisvert, S., Laviolette, F., Corbeil, J., 2010. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J. Comput. Biol. 17, 1519–1533. http://dx.doi.org/10.1089/cmb.2009.0238.

Boore, J.L., 1999. Animal mitochondrial genomes. Nucl. Acids Res. 27, 1767–1780. http://dx.doi.org/10.1093/nar/27.8.1767.

Boore, J.L., 2001. Complete Mitochondrial Genome Sequence of the Polychaete Annelid *Platynereis dumerilii*. Mol Biol Evol 18, 1413–1416.

Boore, J.L., Brown, W.M., 1995. Complete sequence of the mitochondrial DNA of the annelid worm *Lumbricus terrestris*. Genetics 141, 305–319.

Boore, J.L., Brown, W.M., 2000. Mitochondrial genomes of *Galathealinum*, *Helobdella*, and *Platynereis*: sequence and gene arrangement comparisons indicate that pogonophora is not a phylum and annelida and arthropoda are not sister taxa. Mol. Biol. Evol. 17, 87–106.

Brown, C.T., Howe, A., Zhang, Q., Pyrkosz, A.B., Brom, T.H., 2012. A reference-free algorithm for computational normalization of shotgun sequencing data. arXiv 1203.4802 [q-bio].

Clayton, D.A., 1982. Replication of animal mitochondrial DNA. Cell 28, 693–705. http://dx.doi.org/10.1016/0092-8674(82)90049-6.

Clayton, D.A., 1991. Replication and transcription of vertebrate mitochondrial DNA. Annu. Rev. Cell Biol. 7, 453–478.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32, 1792–1797. http://dx.doi.org/10.1093/nar/gkh340.

Glover, A.G., Källström, B., Smith, C.R., Dahlgren, T.G., 2005. World-wide whale worms? a new species of *Osedax* from the shallow north Atlantic. Proc. R. Soc. B 272, 2587–2592. http://dx.doi.org/10.1098/rspb.2005.3275.

Glover, A.G., Wiklund, H., Taboada, S., Avila, C., Cristobo, J., Smith, C.R., Kemp, K.M., Jamieson, A.J., Dahlgren, T.G., 2013. Bone-eating worms from the Antarctic: the contrasting fate of whale and wood remains on the Southern Ocean seafloor. Proc. R. Soc. B 280, 20131390. http://dx.doi.org/10.1098/rspb.2013.1390.

Goffredi, S.K., Orphan, V.J., Rouse, G.W., Jahnke, L., Embaye, T., Turk, K., Lee, R., Vrijenhoek, R.C., 2005. Evolutionary innovation: a bone-eating marine symbiosis. Environ. Microbiol. 7, 1369–1378. http://dx.doi.org/10.1111/j.1462-2920.2005.00824.x.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., 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. http://dx.doi.org/10.1038/nbt.1883.

Halanych, K.M., 2005. Molecular phylogeny of siboglinid annelids (a.k.a. pogonophorans): a review. Hydrobiologia 535–536, 297–307. http://dx.doi.org/10.1007/s10750-004-1437-6.

Halanych, K.M., Lutz, R.A., Vrijenhoek, R.C., 1998. Evolutionary origins and age of vestimentiferan tube worms. Cahiers de Biologie Marine 39, 355–358.

Halanych, K.M., Feldman, R.A., Vrijenhoek, R.C., 2001. Molecular evidence that *Sclerolinum brattstromi*: is closely related to vestimentiferans, not to frenulate pogonophorans (Siboglinidae, Annelida). Biol. Bull. 201, 65–75.

Hilário, A., Capa, M., Dahlgren, T.G., Halanych, K.M., Little, C.T.S., Thornhill, D.J., Verna, C., Glover, A.G., 2011. New perspectives on the ecology and evolution of siboglinid tubeworms. PLoS ONE 6, e16309. http://dx.doi.org/10.1371/journal.pone.0016309.

Hrbek, T., Farias, I.P., 2008. The complete mitochondrial genome of the pirarucu (*Arapaima gigas*, Arapaimidae, Osteoglossiformes). Genetics Mol. Biol. 31, 293–302. http://dx.doi.org/10.1590/S1415-47572008000200024.

Huelsenbeck, J., Rannala, B., 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Syst. Biol. 53, 904–913. http://dx.doi.org/10.1080/10635150490522629.

Ivey, J.L., Santos, S.R., 2007. The complete mitochondrial genome of the Hawaiian anchialine shrimp *Halocaridina rubra* Holthuis, 1963 (Crustacea: Decapoda: Atyidae). Gene 394, 35–44. http://dx.doi.org/10.1016/j.gene.2007.01.009.

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. http://dx.doi.org/10.1093/molbev/msi008.

Kück, P., Meusemann, K., 2010. FASconCAT: convenient handling of data matrices. Mol. Phylogenet. Evol. 56, 1115–1118. http://dx.doi.org/10.1016/j.ympev.2010.04.024.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Meth. 9, 357–359. http://dx.doi.org/10.1038/nmeth.1923.

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. http://dx.doi.org/10.1093/molbev/msh112.

Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286–2288. http://dx.doi.org/10.1093/bioinformatics/btp368.

Li, Y.-C., Korol, A.B., Fahima, T., Nevo, E., 2004. Microsatellites within genes: structure, function, and evolution. Mol. Biol. Evol. 21, 991–1007. http://dx.doi.org/10.1093/molbev/msh073.

Lowe, T.M., Eddy, S.R., 1997. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucl. Acids Res. 25, 0955–964. http://dx.doi.org/10.1093/nar/25.5.0955.

Milne, I., Stephen, G., Bayer, M., Cock, P.J.A., Pritchard, L., Cardle, L., Shaw, P.D., Marshall, D., 2013. Using tablet for visual exploration of second-generation

sequencing data. Brief. Bioinform. 14, 193–202. http://dx.doi.org/10.1093/bib/bbs012.
Miya, M., Kawaguchi, A., Nishida, M., 2001. Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. Mol. Biol. Evol. 18, 1993–2009.
Mueller, R.L., 2006. Evolutionary rates, divergence dates, and the performance of mitochondrial genes in Bayesian phylogenetic analysis. Syst. Biol. 55 (2), 289–300. http://dx.doi.org/10.1080/10635150500541672.
Nagarajan, N., Pop, M., 2013. Sequence assembly demystified. Nat. Rev. Genet. 14, 157–167. http://dx.doi.org/10.1038/nrg3367.
Neto, E.D., Correa, R.G., Verjovski-Almeida, S., Briones, M.R.S., Nagai, M.A., da Silva, W., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S.L., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F.L., de Souza, S.J., Simpson, A.J.G., 2000. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. PNAS 97, 3491–3496. http://dx.doi.org/10.1073/pnas.97.7.3491.
Osigus, H.-J., Eitel, M., Bernt, M., Donath, A., Schierwater, B., 2013. Mitogenomics at the base of Metazoa. Mol. Phylogenet. Evol. 69, 339–351. http://dx.doi.org/10.1016/j.ympev.2013.07.016.
Perna, N.T., Kocher, T.D., 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41, 353–358. http://dx.doi.org/10.1007/BF00186547.
Rota-Stabelli, O., Yang, Z., Telford, M.J., 2009. MtZoa: a general mitochondrial amino acid substitutions model for animal evolutionary studies. Mol. Phylogenet. Evol. 52, 268–272. http://dx.doi.org/10.1016/j.ympev.2009.01.011.
Rouse, G.W., 2001. A cladistic analysis of Siboglinidae Caullery, 1914 (Polychaeta, Annelida): formerly the phyla Pogonophora and Vestimentifera. Zool. J. Linnean Soc. 132, 55–80. http://dx.doi.org/10.1111/j.1096-3642.2001.tb02271.x.
Rouse, G.W., Wilson, N.G., Worsaae, K., Vrijenhoek, R.C., 2015. A dwarf male reversal in bone-eating worms. Curr. Biol. 25, 236–241. http://dx.doi.org/10.1016/j.cub.2014.11.032.
Rouse, G.W., Goffredi, S.K., Vrijenhoek, R.C., 2004. *Osedax*: bone-eating marine worms with dwarf males. Science 305, 668–671. http://dx.doi.org/10.1126/science.1098650.
Rousset, V., Rouse, G.W., Siddall, M.E., Tillier, A., Pleijel, F., 2004. The phylogenetic position of Siboglinidae (Annelida) inferred from 18S rRNA, 28S rRNA and morphological data. Cladistics 20, 518–533. http://dx.doi.org/10.1111/j.1096-0031.2004.00039.x.
Ruby, J.G., Bellare, P., Derisi, J.L., 2013. PRICE: software for the targeted assembly of components of (meta) genomic sequence data. G3 (Bethesda) 3, 865–880. http://dx.doi.org/10.1534/g3.113.005967.
Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., Barrell, B., 2000. Artemis: sequence visualization and annotation. Bioinformatics 16, 944–945. http://dx.doi.org/10.1093/bioinformatics/16.10.944.
Schattner, P., Brooks, A.N., Lowe, T.M., 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucl. Acids Res. 33, W686–W689. http://dx.doi.org/10.1093/nar/gki366.
Schmaljohann, R., 1991. Oxidation of various potential energy sources by the methanotrophic endosymbionts of Siboglinum poseidoni Pogonophora. Mar. Ecol. Progr. Series 762, 143–148.
Schulze, A., 2003. Phylogeny of Vestimentifera (Siboglinidae, Annelida) inferred from morphology. Zool. Scripta 32, 321–342. http://dx.doi.org/10.1046/j.1463-6409.2003.00119.x.
Schulze, A., Halanych, K.M., 2003. Siboglinid evolution shaped by habitat preference and sulfide tolerance. Hydrobiologia 496, 199–205. http://dx.doi.org/10.1023/A:1026192715095.
Shadel, G.S., Clayton, D.A., 1997. Mitochondrial DNA Maintenance in Vertebrates. Annu. Rev. Biochem. 66, 409–435. http://dx.doi.org/10.1146/annurev.biochem.66.1.409.
Shao, R., Kirkness, E.F., Barker, S.C., 2009. The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse *Pediculus humanus*. Genome Res. 19, 904–912. http://dx.doi.org/10.1101/gr.083188.108.
Shen, X., Ma, X., Ren, J., Zhao, F., 2009. A close phylogenetic relationship between Sipuncula and Annelida evidenced from the complete mitochondrial genome sequence of *Phascolosoma esculenta*. BMC Genomics 10, 136. http://dx.doi.org/10.1186/1471-2164-10-136.
Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508. http://dx.doi.org/10.1080/10635150290069913.
Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247. http://dx.doi.org/10.1093/bioinformatics/17.12.1246.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690. http://dx.doi.org/10.1093/bioinformatics/btl446.
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, 57. http://dx.doi.org/10.1186/1471-2148-7-57.
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. http://dx.doi.org/10.1038/nature09864.
Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564–577. http://dx.doi.org/10.1080/10635150701472164.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739. http://dx.doi.org/10.1093/molbev/msr121.
Thornhill, D.J., Wiley, A.A., Campbell, A.L., Bartol, F.F., Teske, A., Halanych, K.M., 2008. Endosymbionts of *Siboglinum fiordicum* and the phylogeny of bacterial endosymbionts in Siboglinidae (Annelida). Biol. Bull. 214, 135–144.
Vallès, Y., Boore, J.L., 2006. Lophotrochozoan mitochondrial genomes. Integr. Comp. Biol. 46, 544–557. http://dx.doi.org/10.1093/icb/icj056.
Vallès, Y., Halanych, K.M., Boore, J.L., 2008. Group II introns break new boundaries: presence in a Bilaterian's genome. PLoS ONE 3, e1488. http://dx.doi.org/10.1371/journal.pone.0001488.
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. http://dx.doi.org/10.1093/molbev/msu080.
Yuan, Y., Li, Q., Yu, H., Kong, L., 2012. The complete mitochondrial genomes of six heterodont bivalves (Tellinoidea and Solenoidea): variable gene arrangements and phylogenetic implications. PLoS ONE 7, e32353. http://dx.doi.org/10.1371/journal.pone.0032353.
Zakian, V.A., 1995. Telomeres: beginning to understand the end. Science 270, 1601–1607.
Zhang, D.X., Hewitt, G.M., 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. Biochem. Syst. Ecol. 25, 99–120. http://dx.doi.org/10.1016/S0305-1978(96)00042-7.
Zhong, M., Struck, T.H., Halanych, K.M., 2008. Phylogenetic information from three mitochondrial genomes of Terebelliformia (Annelida) worms and duplication of the methionine tRNA. Gene 416, 11–21. http://dx.doi.org/10.1016/j.gene.2008.02.020.
Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucl. Acids Res. 31, 3406–3415. http://dx.doi.org/10.1093/nar/gkg595.