Multiple introns in a deep-sea Annelid (Decemunciger: Ampharetidae) mitochondrial genome

Angelo F. Bernardino1, Yuanning Li2, Craig R. Smith3 & Kenneth M. Halanych2

Wood falls provide episodic fluxes of energy to the sea floor that are degraded by a species-rich benthic fauna. Part of this rich diversity includes annelid polychaetes but unfortunately, our understanding of such fauna is limited and their genetic variability and evolutionary origins remain poorly known. In this study, we sequenced complete mitochondrial genomes from three congeneric Decemunciger (Ampharetidae) individuals that had colonized multiple wood falls in the deep (~1600 m) NE Pacific Ocean. Mitochondrial gene order within Decemunciger was similar to the three other available Terebellomorpha genomes, consistent with the relatively conserved nature of mitochondrial genomes within annelids. Unexpectedly, we found introns within the cox1, nad1 and nad4 genes of all three genomes assembled. This is the greatest number of introns observed in annelid mtDNA genomes, and possibly in bilaterians. Interestingly, the introns were of variable sizes suggesting possible evolutionary differences in the age and origins of introns. The sequence of the introns within cox1 is similar to Group II introns previously identified, suggesting that introns in the mitochondrial genome of annelids may be more widespread then realized. Phylogenetically, Decemunciger appears to be a sister clade among current vent and seep deep-sea Ampharetinae.

Ampharetid polychaetes are tube-dwelling annelids that are abundant on shallow-marine and deep-sea continental margins, with some species showing adaptations to sulfide-rich sediments near cold seeps and organic falls, including wood-falls and whale carcasses4–7. In organic-fall and cold-seep habitats, these polychaetes can show remarkable abundances and diversity and may be important for organic-matter degradation8, 9. However, as with many other deep-sea taxa, there is limited understanding of their diversity and evolution, requiring additional study including use of informative molecular markers10, 11. Despite their high diversity and abundance in the deep-sea, a limited number of polychaete taxa have been molecularly characterized from deep-sea ecosystems and from chemosynthetic habitats12–14.

Advances in phylogenetic and evolutionary understanding of Annelida has been made using comparative mitogenomics15–17. Annelids, like other bilaterians, typically have 37 mitochondrial genes18–20. Recent descriptions of mitochondrial genomes from several annelid lineages revealed marked differences in gene order that are helping to resolve phylogenetic relationships, even though some inconsistencies between sequence data and phylogenies remain14, 20, 21. There are currently about 90 complete annelid mitochondrial DNA sequences (mtDNA) published14, 22, with many underrepresented lineages, making broad scale mitogenomic comparisons limited given the extremely high number of species in the deep-sea11. For instance, in the family Ampharetidae, there are only two incomplete mitochondrial genomes reported (Eclysippe vanelli and Auchenoplax crinita)18. Descriptions of new mtDNA genomes can help to clarify phylogenetic relationships among closely related lineages and also to discover less frequent genome features such as the presence of Group II introns23. The phylogeny of Terebelliformia includes two clades, one with Ampharetidae, Alvinellidae and Pectinariidae and the other with Terebellidae and Trichobranchidae24. Ampharetidae is a sister group to Alvinellidae based on current

1Universidade Federal do Espírito Santo, Grupo de Ecologia Bêntica, Departamento de Oceanografia, Av. Fernando Ferrari, 514, Vitória, ES, 29075-910, Brazil. 2Auburn University, Department of Biological Sciences, 101 Life Sciences Building, Auburn, AL, 36849, USA. 3Department of Oceanography, SOEST, University of Hawaii at Manoa, 1000 Pope Road, Honolulu, HI, 96822, USA. Correspondence and requests for materials should be addressed to A.F.B. (email: angelo.bernardino@ufes.br) or K.M.H. (email: ken@auburn.edu)
sizes between 14,414 and 22,058 bp\(^{14, 19, 20, 22, 30}\). Although the mitochondrial genome size varied slightly among species (Table 1), with 2,300 bp of introns (Fig. 1; Table 1). Other previously studied annelids have mtDNA genomes in lower eukaryotes, microbes, algae and higher plants, and are reported to contain genes with mobile capability\(^{20–28}\). Within Bilateria metazoans, group II introns were first described in the mitochondrial genome of the polychaete *Nephtys* sp.\(^{23}\), even though bilaterian mtDNA genomes were thought to be conserved in terms of gene content and lack introns\(^{27, 28} \). However, recent mitogenomic investigations have revealed a more common presence of Group II introns in the *cox1* mitochondrial gene in some Annelid worms, including two *Glycera* species and one myzostomid *Endomyzostoma* \(^{30, 31}\). Based on previous phylogenetic analysis, Richter *et al.* \(^{30}\) demonstrated a close phylogenetic relationship between *Nephtys* sp. and *Glycera* introns, but less similarity with one of the two *cox1* introns from *Glycera fallax*. The presence of introns in a few distantly-related annelid taxa makes mechanisms of intron acquisition and substitution rates of the relevant mtDNA regions unclear\(^{23}\). Although mitochondrial gene order is relatively conserved among annelids\(^{19}\), the presence or absence of such introns, their number and their association with unique or multiple genes with variable function suggests that annelid mitochondrial genomes may exhibit more variability than anticipated\(^{19, 20}\).

We sequenced mitochondrial genomes of an abundant ampharetid (*Decemunciger* sp.) sampled from wood-fall blocks experimentally implanted for 12 months at ~1600 m depth on the East Pacific US margin. We detected differences in mitochondrial gene order relative to previously reported Terebellomorpha mt genomes\(^{18}\). Unexpectedly, we detected three intragenetic regions within *cox1* (Group II intron), *nad1* and *nad4* genes. Furthermore, we conducted a phylogenetic analysis of Ampharetidae based on available mt genomes and transcriptomic data to further explore ampharetid evolutionary history.

### Results and Discussion

#### Genome assemblies and description.

Using Illumina sequence data from three individuals of a deep-sea ampharetid annelid abundant on wood-falls in the deep NE Pacific, we assembled complete mitochondrial genomes. The three individuals were morphologically identified to potential new species of *Decemunciger*, and all three assembled genomes had a 100% identical *cox1* gene. There is no previous molecular data to confirm the identity of *Decemunciger* sampled in wood blocks separated by over 400 km on the Oregon-Washington margin, with the paratype described from the Atlantic\(^{32}\). Using a BLAST-based approach\(^{33}\), we identified mtDNA contigs that were roughly 15,000–16,000 bp in size from the genome assembly. The integrity of these contigs was confirmed by mapping sequence reads to the assembly \(^{34}\). *Decemunciger* sp. mt genome has 16,703–16,974 bp without introns, which is similar to the ampharetids *Eclysippe vanelli* (16,547; EU239687\(^{18} \)) and is slightly longer than the other ampharetid *Auchenoploax crinita* (13,759 bp; FJ976041 incomplete) and the Terebellomorpha *Pista cristata* (15,894 bp; EU239688). The complete mtDNA of *Decemunciger* sp. is approximately 19 kb long (19,003 to 19,274 bp; Table 1), with 2,300 bp of introns (Fig. 1; Table 1). Other previously studied annelids have mtDNA sizes between 14,414 and 22,058 bp\(^{14, 19, 20, 22, 30}\). Although the mitochondrial genome size varied slightly among our three specimens, the intergenic region between *nad2* and *cox1* showed the greatest variation.

|        | *Decemunciger* sp. A3359 AC KY742027 | *Decemunciger* sp. A3372-1 AC KY774370 | *Decemunciger* sp. A3372-2 AC KY774371 |
|--------|-------------------------------------|----------------------------------------|----------------------------------------|
| mtDNA size (bp) | 19,274                             | 19,096                                 | 19,003                                 |
| Coverage  depth | 258x                               | 1514x                                  | 1394x                                  |

Table 1. Genome size, coverage, coverage depth and base composition of assembled *Decemunciger* sp. Mitochondrial genomes. AC - GenBank accession numbers.

molecular analysis from mitochondrial and nuclear genes\(^{6, 7, 11, 25}\), but the taxonomy within the family is complex due to morphological variability. There is only limited phylogenetic work within Ampharetidae, but the subfamily Ampharetinae host several species adapted to chemosynthetic deep-sea ecosystems\(^{7, 25}\).

Group II introns are self-splicing mobile genetic elements typically found in mitochondrial and other organelle genomes in lower eukaryotes, microbes, algae and higher plants, and are reported to contain genes with mobile capability\(^{20–28}\). Within Bilateria metazoans, group II introns were first described in the mitochondrial genome of the polychaete *Nephtys* sp.\(^{23}\), even though bilaterian mtDNA genomes were thought to be conserved in terms of gene content and lack introns\(^{27, 28} \). However, recent mitogenomic investigations have revealed a more common presence of Group II introns in the *cox1* mitochondrial gene in some Annelid worms, including two *Glycera* species and one myzostomid *Endomyzostoma* \(^{30, 31}\). Based on previous phylogenetic analysis, Richter *et al.* \(^{30}\) demonstrated a close phylogenetic relationship between *Nephtys* sp. and *Glycera* introns, but less similarity with one of the two *cox1* introns from *Glycera fallax*. The presence of introns in a few distantly-related annelid taxa makes mechanisms of intron acquisition and substitution rates of the relevant mtDNA regions unclear\(^{23}\). Although mitochondrial gene order is relatively conserved among annelids\(^{19}\), the presence or absence of such introns, their number and their association with unique or multiple genes with variable function suggests that annelid mitochondrial genomes may exhibit more variability than anticipated\(^{19, 20}\).

For each mitochondrial genome sequenced herein, the genome was composed of 37 genes, with all 13 protein-coding, 2 ribosomal rRNAs and 22 tRNAs\(^{29}\) (Fig. 1). All genes encoded on the same strand, typical of other annelids\(^{20}\). As observed in other Terebellomorpha, *Decemunciger* sp. mtDNA is AT rich (65.1% AT) in the coding regions (CDS) (Table 1). Mitochondrial gene orders of *Decemunciger* sp. mtDNA differ from *E. vanelli* in relation to positions of *nad4*, *nad4L* and *nad5* genes, and differs from *Terebellides stroemi* (Trichobranchiidae) and *Pista*
which was similar to other Group II introns reported in annelids23, 30. Although ORFs were not found in introns these introns coded a protein, but presented palindromic sequences at both ends (based on a blast search results). Introns in distinct mitochondrial genes from Bilaterians. Introns within the cox1 gene and size of the introns were identical within the three assembled genomes and none of the three mitochondrial genomes from this study, from published genomes in GenBank and from transcriptomic sequencing will very likely reveal new patterns of introns as usual mitochondrial barcoding (e.g. Ampharetidae). A recent analysis of Syllidae also showed marked variability on the order of protein encoding genes, with four distinct gene orders14. With only 89 complete mtDNAs sequenced from annelids14, 15, 19, 20, 22, more variation in gene orders will certainly be uncovered. Slight differences in the number of introns were also revealed in Decemunciger sp., if compared to previous Terebelliformia mtDNA. Terbellides stroemi and P. cristata have two copies of the methionine tRNA gene in their mtDNA, whereas only one copy was present in Decemunciger sp. mtDNA, as previously observed on the ampharetid E. vanelli38. Changes in the position of tRNAs between Decemunciger sp. and the other Terebellomorpha were also observed (Fig. 1), and are common in bilaterian mtDNAs39.

Introns in Decemunciger mtDNA. Mitochondrial genomes of the three Decemunciger sp. individuals revealed the presence of introns within the cox1, nad1 and nad4 genes, which is the first report to date of multiple introns in distinct mitochondrial genes from Bilaterians. Introns within the cox1, nad1 and nad4 genes were 1648, 390 and 262 bp long, respectively. All introns were the same size across the three assembled genomes and none of these introns coded a protein, but presented palindromic sequences at both ends (based on a blast search results). The cox1 intron contained a 390 bp ORF for an intron maturase 2 type II transcriptase (blastp e-value 7.68e-08), which was similar to other Group II introns reported in annelids23, 30. Although ORFs were not found in introns from nad1 and nad4 genes, these regions could possibly be derived form ancient transposable elements which have since lost any function. However, the intron maturase enzyme in the cox1 intron may assist transposition of these elements30. Another possibility is that the nad1 and nad4 introns are discontinuous parts of one transposable element split among those genes and can be trans-spliced to form a functional intron30, 36. These mechanisms have been observed in higher plants; if true here, would be the first known case of trans-complementation of introns in annelid mitochondrial genes.

The insertion position into the cox1 gene and size of the introns were identical within the three Decemunciger mitochondrial genomes sequenced. Multiple introns were first identified on mitochondrial genes (cox1 and nad5) of sea anemones (Group I intron37, and recently Group II introns have been reported on a cox1 gene of a Nephtys (Nephys sp.) and glycerid polychaetes23, 30. Intron sizes, their position within the cox1 gene and their coding protein sequences, differ between Nephtys sp., Decemunciger sp. and Glyceria spp., consistent with distinct episodes of intron gain in these annelid lineages23, 30. Phylogenetic differences in the ORF region between introns are evident (Fig. 2). Different insertion positions of introns within cox1 genes of Decemunciger sp., Nephtys sp. and Glyceria spp. may be a result of variable intronic target sites (IEP) within the mitochondrial genome (Fig. 2)35, 30. The cox1 intron in Nephtys sp. has 1819 bp, whereas it is slightly shorter (1647 bp) in Decemunciger sp. The Nephtys sp. intron has an ORF region of 525 bp coding a reverse transcriptase enzyme, whereas the 390 bp region within the Decemunciger sp. cox1 gene translates into a type II intron maturase enzyme. Amino acid sequences of both Nephtys sp. and Decemunciger sp. introns are also only 16% similar, further supporting independent events of insertion in a scenario of “late intron-gain” for annelids23, 30.

Nephys sp. and Decemunciger sp. represent distinct lineages among Annelida, which likely inherited introns from separate viral vectors. The limited presence of introns may also suggest a high rate of intron loss among lineages. The loss of introns in genomes is generally related to fast replication rates observed, for example, in microbes in a process known as “genome streamlining”27. Since mitochondrial DNA is considered to possess a fast evolutionary rate39, introns may be rapidly removed from mitochondrial genes. Further complete mtDNA sequencing will very likely reveal new patterns of introns as usual mitochondrial barcoding (e.g. cox1) in marine invertebrates are based on short (about 600 bp) sequences that would not detect these introns.

Ampharetid phylogeny based on mtDNA. Amino acid (AA) sequences of protein coding genes from the three mitochondrial genomes from this study, from published genomes in GenBank and from transcriptomic...
data (see Table 2) were used to reconstruct a phylogenetic relationship of *Decemunciger* sp. within Ampharetidae. Phylogenetic relationships of ampharetids were inferred using maximum likelihood (ML) analysis from a dataset with the 10 protein-coding and 2 rRNA mitochondrial genes (see methods). The dataset contained 3,024 amino acid residues after trimming using Gblocks and the resulting ML analysis yielded a tree topology with relatively high bootstrap support values for the division of Ampharetidae subfamilies Melinninae and Ampharetinae11, 18 (Fig. 3). Ampharetidae was recovered as a monophyletic group, but our analysis did not include Alvinellidae 7, 25. Melinninae and Ampharetinae were recovered as sister taxa, which supports current phylogenetic analysis 25. Ampharetinae was also recovered as a monophyletic clade with strong support in the amino acid dataset, consistent with previous molecular and morphological analyses7, 11, 18, 25. Whithin Ampharetinae, the *Decemunciger* lineage was sister to a strongly supported clade (bs = 100) comprised of *Eclysippe*, *Auchenoplax*, *Samytha* and *Amphisamytha* species (Fig. 3, Supplemental Fig. S1). *Decemunciger* has also marked morphological similarities (e.g. branchiae position and number) with the vent ampharetid genus *Paramytha* gen nov., which is a sister group to other vent/seep Ampharetinae clades based on *cox1*, *16S* and *18S* genes25, 32. In summary, our phylogenetic analysis support *Decemunciger* as within the Ampharetinae, within a clade comprised of several described species from chemosynthetic ecosystems in the North Atlantic and Arctic basins.

Methods

**Genome assembly, annotation and mapping.** Three Ampharetid specimens (A3359, A3372–1 and A3372–2) were collected from 1.5 kg blocks of douglas fir (*Pseudotsuga menziesi*) experimentally deployed on the seafloor for 15 months and recovered via acoustic release using the R/V Oceana. Ampharetid A3359 was sampled from one wood block recovered from 1605 m depth on Jun 22nd 2014 (43°54.22 N; 125°10.238 W), whereas ampharetids A3372–1 and A3372–2 were sampled from wood blocks recovered about 400 km north from the previous site at 1596 m depth on Jun 27th 2014 (47°57.462 N; 126°02.118 W). Morphological observations indicate that all the three specimens belonged to the ampharetid genus *Decemunciger* sp. Specimens were immediately preserved onboard in 95–100% ethanol and later transferred to Auburn University.

DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen) following manufacturer’s protocols. Sequencing of genomic DNA was performed by The Genomic Services Lab at the Hudson Alpha Institute in Huntsville, Alabama on an Illumina HiSeq 2500 platform (San Diego, California) using 2 x 150 paired-end v4 chemistry. Paired-end reads were assembled de novo using Ray 2.2.0 with k-mer = 3124. Contigs of interest where
identified by using blast with previously published terebellomorph mtDNA genomes against the assembled genomic data. Annotation of the 13 protein-coding genes, 2 ribosomal RNAs and tRNAs was conducted initially with MITOS web server, followed by manual genome annotation in Artemis. Start and stop positions of genes were confirmed by BLASTn and BLASTp searches against the partial mitochondrial genome from *Eclysippe vanelli* (GenBank Accession EU239687) as well as manual inspection.

The presence of introns within coding mitochondrial genes was confirmed by mapping the paired Illumina reads against the assembled mitochondrial genome to check for coverage in each coding region and near the intronic reads (Supplementary Fig. S2). Reads were mapped with Bowtie2, indexed and sorted with Samtools and visually checked with Tablet software. Identity on introns was aided by Blast searches when possible.

Transcriptomic data generation and assembly for phylogenetic analysis. Upon collection, all specimens were either stored at −80 °C, in ethanol or preserved in RNAlater (Life Technologies Inc.). Due to a limiting amount of tissue, only RNA was extracted since mitochondrial protein-coding and ribosomal RNA genes, which were used in mitogenomic analysis, can be recovered from transcriptome sequencing. RNA extraction and cDNA preparation for high-throughput sequencing followed standard protocols. Total RNA was extracted using TRIzol (Invitrogen) and purified using the RNeasy kit (Qiagen) with on-column DNase digestion. Next, single strand cDNA libraries were reverse transcribed using the SMART cDNA Library Construction kit (Clontech) followed by double-stranded cDNA synthesis using the Advantage 2 PCR system (Clontech). Illumina sequencing library preparation and sequencing of *Lysippe labiata*, *Samytha sexcirrata*, *Samytha californiensis*, *Amphisamytha biculata*, *Amphicteis gunneri*, *Auchenoplax crinita* and *Melinna maculata* were performed by The Genomic Services Lab at the Hudson Alpha Institute in Huntsville, Alabama using 2 × 100 paired-end sequencing on an Illumina HiSeq 2000 platform (San Diego, California).

Prior to assembly, Illumina paired-end transcriptome sequence data were digital normalized to a k-mer coverage of 30 using normalize-by-median.py. Remaining reads were then assembled using Trinity with default settings. Mitochondrial protein-coding genes and ribosomal RNAs were identified by TBLSTX and BLASTN, respectively (using the recovered *E. vanelli* mt genome as query).

Phylogenetic analysis. Fourteen taxa were included in the phylogenetic analysis. *Pista cristata* (Terebellidae) and *Terebellides stroemi* (Trichobranchidae) were acquired from GenBank (Table 2) and selected as outgroups based on data availability as well as current understanding of annelid evolutionary history. To assist in phylogenetic analysis and check the previous incomplete assembly of the ampharetid mtDNA *Eclysippe vanelli*, we assembled a new complete mitochondrial genome from the ampharetid *E. vanelli*. The assembled *E. vanelli* genome has an identical gene order with the previous incomplete genome and a *cox1* amino acid identity of 99.8% with the *cox1* gene from the incomplete *E. vanelli* genome. We used the complete *E. vanelli* genes for phylogenetic analysis (indicated below), and included genes from transcriptomic assembly from seven other species of interest.

Our data set was based on amino acid sequences from 10 mitochondrial protein-coding genes (*cox1*, *cox2*, *cox3*, *cob*, *atp6*, *nad1*, *nad2*, *nad4*, *nad5*, *nad6*) and two ribosomal RNA genes (*rrnS* and *rrnL*). *nad4l*, *atp8* and *nad3* sequences were excluded due to limited number of recovered sequences from transcriptome data. Each of the 12 genes was individually aligned using MAFFT followed by manual correction. The selected genes were then trimmed using the default setting in Gblocks to remove ambiguously aligned regions. Genes were then concatenated into final supermatrix datasets using FASconCAT for downstream phylogenetic analysis. Phylogenetic relationships of ampharetids were inferred using maximum likelihood (ML) in RAxML. Prior to Figure 3. Maximum likelihood tree obtained when analyzing amino acid sequences from mtDNA protein coding genes. All nodes were supported with 100% bootstrap value (bs = 100) unless otherwise noted. Dashed lines indicate subfamilies represented within Ampharetidae.
ML analyses, PartitionFinder V1.1.152 was used to evaluate best-fit partition schemes and associated best-fit substitution models for both datasets. Topological robustness for the ML analysis was evaluated with 100 replicates of fast-bootstrapping.

**Intron phylogeny.** Phylogenetic position of group II introns was compared with the alignment of which built upon an analysis by Richter et al.30,33,34. The mitochondrial group II introns from *cox1* genes of the Annelids *Glyceria fallax*, *Glyceria unicorns* and Nephtys sp. were analyzed and compared to the *cox1* intron ORF from Decemunciger sp. and other chloplast and bacterial intronic ORFs. The Maximum likelihood analysis was conducted with RAxML v.8.0.5 under the substitution model LG + I + G + F. Bootstrap support values (>50%) from 1,000 pseudoreplicates are given at the nodes. Colorcodes were defined accordingly to Richter et al.30, where group II intron-encoded ORFs known from chloroplast genomes are highlighted in green, mitochondrial genomes in blue, and bacterial genomes in red. Genbank numbers from intron sequences used in this analysis are given in Richter et al.30.

**References**
1. Bernardino, A. F. & Smith, C. R. Community structure of infaunal macrobenthos around vestimentiferan thackets at the San Clemente cold seep, NE Pacific. *Marine Ecology-an Evolutionary Perspective* 31, 608–621, doi:10.1111/j.1439-0485.2010.00389.x (2010).
2. Bernardino, A. F., Levin, L. A., Thurber, A. R. & Smith, C. R. Comparative Composition, Diversity and Trophic Ecology of Sediment Macrofauna at Vents, Seeps and Organic Falls. *Plos One* 7, doi:10.1371/journal.pone.0033515 (2012).
3. Bernardino, A. F., Smith, C. R., Baco, A., & Altamira, I. & Sumida, P. Y. G. Macrofaunal succession in sediments around kelp and wood falls in the deep NE Pacific and community overlap with other reducing habitats. *Deep-Sea Research Part I-Oceanographic Research Papers* 57, 708–723, doi:10.1016/j.dsr.2010.03.004 (2010).
4. Smith, C. R., Bernardino, A. F., Baco, A. & Harndies, A. & Altamira, I. Seven-year enrichment: macrofaunal succession in deep-sea sediments around a 30 tonne whale fall in the Northeast Pacific. *Marine Ecology Progress Series* 515, 133–149, doi:10.3354/meps10995 (2014).
5. Rouse, G. W. & Pleijel, F. *Polychaetes*. (Oxford University Press, 2001).
6. Reuscher, M., Fiege, D. & Wehe, T. Terebellomorph polychaetes from hydrothermal vents and cold seeps with the description of two new species of Terebellidae (Annelida: Polychaeta) representing the first records of the family from deep-sea vents. *Journal of the Marine Biological Association of the United Kingdom* 92, 997–1012, doi:10.1017/S0025315411000658 (2012).
7. Stiller, J. et al. Phylogeny, biogeography and systematics of hydrothermal vent and methane seep Amphipamithya (Ampharetidae, Annelida), with descriptions of three new species. *Systematics and Biodiversity* 11, 35–65, doi:10.1080/14772000.2013.772925 (2013).
8. Smith, C. R., Glover, A. G., Treude, T., Higgs, N. D. & Amon, D. J. Whale-fall ecosystems: recent insights into ecology, paleoecology and evolution. *Annual Review of Marine Science* 7, 571–596, doi:10.1146/annurev-marine-010213-135144 (2015).
9. Treude, T. et al. Biogeochemistry of a deep-sea whale fall: sulfate reduction, sulfide efflux and methanogenesis. *Marine Ecology Progress Series* 382, 1–21, doi:10.3354/meps09772 (2009).
10. Halanych, K. M. & Janosik, A. M. A review of molecular markers used for Annelid phylogenetics. *Integrative and Comparative Biology* 46, 533–543 (2006).
11. Zhong, M., Hansen, M., Nesnidal, A., Golombek, A. & Halanych, K. M. Escaping the symplesiomorphy trap: A multigene phylogenetic analysis for terebelliform annelids. *BMC Evolutionary Biology* 11, 369 (2011).
12. Snegirev, P. V. & Smith, C. R. A. Riot of species in an environmental calm: the paradox of the species-rich deep-sea floor. *Oceanography and Marine Biology: an Annual Review* 40, 311–342 (2002).
13. Appeltans, W. et al. The magnitude of global marine species diversity. *Current Biology* 22, 2189–2202, doi:10.1016/j.cub.2012.09.036 (2012).
14. Agudo, M. T. et al. Syllidae mitochondrial gene order is usually variable for Annelida. *Gene* 594, 89–96 (2016).
15. Weigert, A. & Bleidorn, C. Current status of annelid phylogeny. *Organisms Diversity and Evolution* 16, 345–362. doi:10.1007/s13127-016-0265-7 (2016).
16. Halanych, K. M. How our view of animal phylogeny was reshaped by molecular approaches: lessons learned. *Organisms Diversity and Evolution* 16, 319–328 (2016).

| Species               | Family         | Subfamily       | mtDNA genome                  | Transcriptome data     | Ref               |
|-----------------------|----------------|-----------------|-------------------------------|------------------------|------------------|
| *Pista cristata*      | Terebellidae   |                 | NC_011011.1                  |                        | Zhong et al.18    |
| *Terebellides stroemi*| Trichobranchidae |                | NC_011014                    |                        | Zhong et al.18    |
| *Decemunciger* sp A3559 | Ampharetidae | Ampharetinae   | this study KY742027          |                        |                  |
| *Decemunciger* sp A3372-1 | Ampharetidae | Ampharetinae   | this study KY774370          |                        |                  |
| *Decemunciger* sp A3372-2 | Ampharetidae | Ampharetinae   | this study KY774371          |                        |                  |
| *Amphipamithya bioculata* | Ampharetidae | Ampharetinae   | this study KY972369-KY972532 |                        |                  |
| *Samytha californiensis* | Ampharetidae | Ampharetinae   | this study KY972369-KY972532 |                        |                  |
| *Samytha sexcirrata*  | Ampharetidae   | Ampharetinae   | this study KY972369-KY972532 |                        |                  |
| *Melina maculata*     | Ampharetidae   | Melininae       | this study KY972369-KY972532 |                        |                  |
| *Aucheneplax crinita* | Ampharetidae   | Ampharetinae   | this study KY972369-KY972532 |                        | Zhong et al.18    |
| *Echylippe vanelli*   | Ampharetidae   | Ampharetinae   | this study                  |                        |                  |
| *Echylippe vanelli*   | Ampharetidae   | Ampharetinae   | EU239687                    |                        | Zhong et al.18    |
| *Amphicetes gunneri*  | Ampharetidae   | Ampharetinae   | this study KY972369-KY972532 |                        |                  |
| *Izylippe labiata*    | Ampharetidae   | Ampharetinae   | this study KY972369-KY972532 |                        |                  |

Table 2. List of taxa included in the Ampharetidae phylogenetic analysis, with genbank assession numbers and references to published sequences.
17. Valles, Y. & Boore, J. L. Lophotrochozoan mitochondrial genomes. Integrative and Comparative Biology 46, 544–557, doi:10.1093/icb/icp056 (2006).
18. Zhong, M., Struck, T. H. & Halanych, K. M. Three mitochondrial genomes of Terebelliformia (Annelida) worms and duplication of methionine tRNA. Gene 416, 11–21 (2008).
19. Jennings, R. M. & Halanych, K. M. Mitochondrial genomes of Clymenella torquata (Maldanidae) and Riftia pachyptila (Siboglinidae): Evidence for conserved gene order in Annelida. Molecular Biology and Evolution 22, 210–222 (2005).
20. Weigert, A. et al. Evolution of mitochondrial gene order in Annelida. Molecular Phylogenetics and Evolution 94, 196–206 (2016).
21. Aguado, M. T., Glasby, C. J., Schroeder, P. C., Weigert, A. & Bleidorn, C. The making of a branching annelid: an analysis of complete mitochondrial genome and ribosomal data of Ramiysillis multicaudata. Scientific Reports 5, 12072, doi:10.1038/srep12072 (2015).
22. Seixas, V. C., Russo, C. Ad. M. & Paiva, P. C. Mitochondrial genome of the Christmas tree worm Spirobranchus giganteus (Annelida: Serpulidae) reveals a high substitution rate among annelids. Gene 605, 43–53, doi:10.1016/j.gene.2016.12.024 (2017).
23. Valles, Y., Halanych, K. M. & Boore, J. L. Group II Introns Break New Boundaries: Presence in a Bilaterian’s Genome. PLOS One 3, e1488, doi:10.1371/journal.pone.0001488 (2008).
24. Rouse, G. W. & Fauchald, K. Cladistics and polychaetes. Zoologica Scripta 26, 139–204 (1997).
25. Kongsrud, J. A., Eilertsen, M. H., Alvestad, T., Kongshavn, K. & Rapp, H. T. New species of Ampharetidae (Annelida: Polycheata) from Arctic Loki Castle vent field. Deep-Sea Research II in press, doi:10.1016/j.dsr2.2016.08.015 (2017).
26. Robert, A. R. & Zimmerly, S. Group II intron retroelements function and diversity. Cитогенет Genom Res 110, 589–597 (2005).
27. Lambowitz, A. M. & Zimmerly, S. In Cold Spring Harb Perspectives in Biology Vol. 3 (eds J.F. Atkins, R.F. Gesteland & T.R. Ceci) a003166 (Cold Spring Harb Laboratory Press, 2011).
28. Huchon, D., Szitenberg, A., Shefer, S., Ilan, M. & Feldstein, T. Mitochondrial group I and group II introns in the sponge orders Agelasida and Axinellida. BMC Evolutionary Biology 15, 278, doi:10.1186/s12862-015-0556-1 (2015).
29. Boore, J. L. Animal mitochondrial genomes. Nucleic Acids Research 27, 1767–1780 (1999).
30. Richter, S., Schwarz, F., Hering, L., Boggermann, M. & Bleidorn, C. The utility of genome skimming for phylogenomic analyses as demonstrated for Glycerid relationships (Annelida: Glyceridae). Genome Biology and Evolution 7, 3443–3462, doi:10.1093/gbe/evw224 (2015).
31. Zhong, M. Applicability of mitochondrial genome data to Annelid phylogeny and the evolution of group II introns PhD thesis thesis, Auburn University (2009).
32. Zottoli, R. Two new genera of deep-sea polychaete worms of the family Ampharetidae and the role of one species in deep-sea ecosystems. Proceedings of the Biological Society of Washington 95, 48–57 (1982).
33. Altschul, S. F., Gish, W., Miller, W., Myers, E. M. & Lipman, D. J. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410 (1990).
34. Li, Y. et al. Mitogenomics reveals phylogeny and repeated motifs in control regions of the deep-sea family Siboglinidae (Annelida). Molecular Phylogenetics and Evolution 85, 221–229 (2015).
35. Irimia, M. & Roy, S. W. In Cold Spring Harb Perspectives in Biology (2014).
36. Boonen, L. Trans-splicing of pre-mRNA in plants, animals and protists. FEBS Journal 7, 40–46 (1993).
37. Beagley, C. T., Okada, N. A. & Wolstenholme, D. R. Two mitochondrial group I introns in a metazoan, the sea anemone Actinia equina: one intron contains genes for subunits 1 and 3 of NADH dehydrogenase. PNAS 93, 5619–5623 (1996).
38. Kooinin, E. V. The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? Biology Direct 1, 22, doi:10.1186/1745-6150-1-22 (2006).
39. Brown, W. M., George, M. & Wilson, A. C. Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences 76, 1967–1971 (1979).
40. Bernt, M. et al. MitoS: Improved de novo metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution 69, 313–319 (2013).
41. Rutherford, K. et al. Artemis: sequence visualization and annotation. Bioinformatics 16, 944–945, doi:10.1093/bioinformatics/16.10.944 (2000).
42. Langmead, B. & Salzberg, S. Fast gapped-read alignment with Bowtie2. Nature methods 9, 357–359 (2012).
43. Milne, I. et al. Using Tablet for visual exploration of second-generation sequencing data. Briefings in Bioinformatics 14, 193–202, doi:10.1093/bib/bbo123 (2013).
44. Dias Neto, E. et al. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. Proceedings of the National Academy of Sciences 97, 3491–3496, doi:10.1073/pnas.97.7.3491 (2000).
45. Li, Y. et al. Phylogenomics of tube worms (Siboglinidae, Annelida) and comparative performance of different reconstruction methods. Zoologica Scripta in press, doi:10.1111/zsc.12201 (2017).
46. Brown, C. T., Howe, A., Zhang, Q., Pyrkoza, A. B. & Brom, T. H. A reference-free algorithm for computational normalization of shotgun sequencing data. arXiv 1203, 4802 (2012).
47. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nature Biotechnology 29, 644–652, doi:10.1038/nbt.1883 (2011).
48. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30, 3059–3066 (2002).
49. Talavera, G. & Castresana, J. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. Systematic Biology 56, 565–577, doi:10.1080/10635150701421614 (2007).
50. Kuck, P. & Meusemann, K. FASconCAT: Convenient handling of data matrices. Molecular Phylogenetics and Evolution 56, 115–118, doi:10.1016/j.ympev.2010.04.024 (2010).
51. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690, doi:10.1093/bioinformatics/btl346 (2006).
52. Lanfear, R., Calcott, B., Ho, S. Y. W. & Guindon, S. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29, 1695–1701, doi:10.1093/molbev/msq207 (2012).
53. Zimmerly, S., Haussner, G. & Wu, X. C. Phylogenetic relationships among group II intron ORFs. Nucleic Acids Res 29, doi:10.1093/nar/gkx94 (2001).
54. Valles, Y., Halanych, K. M. & Boore, J. L. Group II introns break new boundaries: presence in a bilaterian’s genome. PLoS One 3, doi:10.1371/journal.pone.0001488 (2008).

Acknowledgements
We thank all members of the science teams and RV Oceanus crew of BOWLS cruises OC1304A and OC1406B for help at sea. Thanks also to Dr. Michael Reuscher for the identification of ampharetids. AFB was supported by a CNPq PDE grant 200504/2015-0. This work was supported by US National Science Foundation grant no. OCE-1155703 to CRS and no. OCE-1155188 to KMH for the BOWLS project and DEB-1036537 for the WormNet II project. This is Moleville Biology Laboratory contribution 62 and Auburn University Marine Biology Program contribution 155.
Author Contributions
Conceived and designed the experiments: C.R.S., K.M.H. Performed the experiments: A.F.B., Y.L., C.R.S., K.M.H. Analyzed the data: A.F.B., Y.L., K.M.H. Contributed reagents/materials/analysis tools: K.M.H., C.R.S. Wrote the paper: A.F.B., Y.L., C.R.S., K.M.H.

Additional Information
Supplementary information accompanies this paper at doi:10.1038/s41598-017-04094-w

Competing Interests: The authors declare that they have no competing interests.

Accession codes: Mitochondrial genomes are deposited in GenBank under accession codes KY742027, KY774370 and KY774371. Genes used in Ampharetid phylogenetic analysis are deposited under accession codes KY972369-KY972378.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017