Molecular Phylogenetics and Evolution 151 (2020) 106892

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

More is needed—Thousands of loci are required to elucidate the relationships of the 'flowers of the sea' (Sabellida, Annelida)

Ekin Tilic[a,b,*], Erfan Sayyari[c], Josefin Stiller[a,d], Siavash Mirarab[c], Greg W. Rouse[a,*]

[a] Scripps Institution of Oceanography, UC San Diego, La Jolla, CA, USA
[b] Institute of Evolutionary Biology and Animal Ecology, University of Bonn, Germany
[c] Department of Electrical and Computer Engineering, UC San Diego, La Jolla, CA, USA
[d] Centre for Biodiversity Genomics, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark

ARTICLE INFO

Keywords:
Phylogenomics
Transcriptomics
Sabellidae
Fabriciidae
Serpulidae

ABSTRACT

Sabellida is a well-known clade containing tube-dwelling annelid worms with a radiolar crown. Iterative phylogenetic analyses over three decades have resulted in three main clades being recognized; Fabriciidae, Serpulidae and Sabellidae, with Fabriciidae proposed as the sister group to Serpulidae. However, relationships within Sabellidae have remained poorly understood, with a proliferation of genera. In order to obtain a robust phylogeny with optimal support, we conducted a large-scale phylogenomic analysis with 19 new sabellid transcriptomes for a total of 21 species. In contrast to earlier findings based on limited DNA data, our results support the position of Fabriciidae as sister taxon to a Sabellidae + Serpulidae clade. Our large sampling within Sabellidae also allows us to establish a stable phylogeny within this clade. We restrict Sabellinae to a subclade of Sabellidae and broaden the previously monotypic Myxicolinae to include *Amphicorina* and *Chone*. We tested the robustness of species tree reconstruction by subsampling increasing numbers of genes to uncover hidden support of alternative topologies. Our results show that inclusion of more genes leads to a more stable topology with higher support, and also that including higher divergence genes leads to stronger resolution.

1. Introduction

Sabellida is a diverse group of tube-dwelling annelids with radiolar crowns. These generally sessile worms use their crown for respiration and to gather suspended food particles from the water column, as the remaining body stays within the tube. The often colorful appearance of these crowns and their rapid withdrawal into the safety of the tubes, have for long attracted scientists and SCUBA divers alike, resulting in their common names fan worms, feather-duster worms and the affectionate general epithet "flowers of the sea" (Jones, 1973) (Fig. 1).

Three family-ranked taxa are grouped within Sabellida: Serpulidae Rafinesque, 1815; Sabellidae Latreille, 1825 and Fabriciidae Rioja, 1923 (see Kupriyanova and Rouse, 2008). Traditionally the clade Sabellida included additional taxa. Sabellida *sensu* Fauchald (1977) and Knight-Jones (1981) also contained Sabellariidae, which was based on their shared chaetal inversion. However, Kieselbach and Hausen (2008) showed significant differences in the specific chaetal arrangement of Sabellidae and Sabellariidae, suggesting an independent evolution of this pattern. Phylogenomic data suggests that Sabellariidae forms a clade with Spionidae (Andrade et al., 2015; Helm et al., 2018; Weigert and Bleidorn, 2016). Rouse and Fauchald (1997) also placed Siboglinidae (formerly outside Annelida, as Pogonophora and Vestimentifera) and Oweniidae inside Sabellida. These taxa no longer form a clade with Sabellidae, Fabriciidae and Serpulidae according to molecular data (Andrade et al., 2015; Helm et al., 2018; Weigert and Bleidorn, 2016). Following this evidence, in this study we restrict the name Sabellida to the clade composed of Fabriciidae, Sabellidae and Serpulidae.

The close affinity of serpulids, the calcareous-tube worms, and sabellids (previously including fabriciids) has long been accepted and supported by a variety of morphological features such as the unique kind of chaetal inversion (Kieselbach and Hausen, 2008), an inverted faecal groove, and a radiolar crown (Dales, 1962; Fauchald, 1977; Fitzhugh, 1989). The phylogenetic relationships within Sabellida have been less clear from morphological data matrices and PCR-based sequence data (Brown et al., 1999; Capa et al., 2011a; Colgan et al., 2006; Dales, 1962; Fauchald, 1977; Fitzhugh, 1989; Kupriyanova and Rouse, 2008; Rouse and Fauchald, 1997; Rousset et al., 2007, 2004; Smith, 1991; Struck et al., 2007; Zrzavý et al., 2009). Though the monophyly of Serpulidae is well supported (Kupriyanova et al., 2006), the former delineation of Sabellidae (including fabriciids) was shown to be

* Corresponding authors at: Scripps Institution of Oceanography, UC San Diego, La Jolla, CA, USA (E. Tilic).
E-mail addresses: etilic@evolution.uni-bonn.de (E. Tilic), grouse@ucsd.edu (G.W. Rouse).

https://doi.org/10.1016/j.ympev.2020.106892
Received 4 May 2020; Received in revised form 9 June 2020; Accepted 10 June 2020
Available online 18 June 2020
1055-7903/

Fig. 1. Live photographs of some of the specimens included in this study. Sabellidae: **A** *Sabella pavonina* Savigny, 1822 **B** *Myxicola affinis* Bush, 1905 **C** *Amphiglena lindae* Rouse & Gambi, 1997 **D** *Sabellastarte magnifica* (Shaw, 1800) **E** *Bispira* cf. "*melanostigma*" **F** *Eudistylia vancouveri* (Kinberg, 1866) Fabriciidae: **G** *Manayunkia occidentalis* Atkinson et al., 2020 **H** *Novafabricia brunnea* (Hartman, 1969) **I** *Fabricinuda* sp. Serpulidae: *Salmacina australis* Haswell, 1885. (All photos G. Rouse, except for **A** © Fredrik Pleijel).

paraphyletic based on sequences from three nuclear genes (Kupriyanova and Rouse, 2008). Therefore, Fabriciinae were removed from Sabellidae, and elevated to Fabriciidae and Sabellidae membership was reduced to the previous subfamily Sabellinae (Kupriyanova and Rouse, 2008). Further, the relationships within Sabellidae have not been resolved with confidence (Capa et al., 2011a). Sabellida has been the focus of interest with regards to life history evolution since they show a wide range of reproductive modes and body sizes (Rouse and Fitzhugh, 1994; Kupriyanova et al., 2001; Rouse et al., 2006). However, without a reliable phylogeny for the group, robust inferences about transformations in life history characters cannot be made.

Transcriptomic datasets of hundreds or thousands of genetic loci have recently transformed the understanding of the annelid tree of life (Andrade et al., 2015; Helm et al., 2018; Struck et al., 2015; Weigert et al., 2014; Weigert and Bleidorn, 2016) and have suggested that some relationships cannot be confidently resolved with a small number of loci (Stiller et al., 2020). A phylogenomic analysis is still missing for the three lineages of Sabellida and within Sabellidae.

The main aim of this study is to resolve the phylogeny of Sabellida, first to elucidate the relationships of the three main lineages and with specific focus on the diversity of Sabellidae. Nineteen new sabellid transcriptomes were sequenced for this study resulting in a sequence matrix containing up to 3015 orthologous genes and around one million amino acid positions. We find a topology among the Sabellida lineages that differs from previous analyses based on few genes. We therefore conducted a thorough methodological analysis using up to 13,393 gene trees to test what effect adding more and more gene trees had on the robustness of the new sabellid phylogeny.

Table 1
Voucher information and SRA accession numbers for the transcriptome data. Vouchers were deposited in the Scripps Institution of Oceanography Benthic Invertebrate Collection (SIO BIC). BioProject accession number PRJNA630201.

Clade	Species	Sampling locality/Source	Voucher SIO BIC	SRA number
Fabriciidae	*Fabricinuda* sp.	Florida, USA	A12371	**SRR11710118**
Fabriciidae	*Manayunkia occidentalis* Atkinson, Batholomew & Rouse 2020	Oregon, USA	A12115	**SRR11710117**
Fabriciidae	*Novafabricia brunnea* (Hartman, 1969)	California, USA	A1787	**SRR11710107**
Serpulidae	*Pomatostegus stellatus* (Abildgaard, 1789)	Belize	A12373	**SRR11710106**
Serpulidae	*Salmacina tribranchiata* (Moore, 1923)	California, USA	A12372	**SRR11710105**
Sabellidae	*Acromegalomma coloratum* (Chamberlin, 1919)	California, USA	A12374	**SRR11710104**
Sabellidae	*Amphicorina anneae* (Rouse, 1994)	Florida, USA	A12376	**SRR11710103**
Sabellidae	*Amphiglena* cf. *mediterranea*	Italy	A9476	**SRR11710102**
Sabellidae	*Anamobaea orstedii* Krøyer, 1856	Belize	A12384	**SRR11710101**
Sabellidae	*Bispira* cf. *"melanostigma"* 1	Belize	A12380	**SRR11710100**
Sabellidae	*Bispira* cf. *"melanostigma"* 2	Belize	A12381	**SRR11710116**
Sabellidae	*Bispira* cf. *turneri*	California, USA	A12382	**SRR11710115**
Sabellidae	*Branchiomma conspersum* (Ehlers, 1887)	Belize	A12378	**SRR11710114**
Sabellidae	*Chone* sp.	Belize	A12377	**SRR11710113**
Sabellidae	*Eudistylia vancouveri* (Kinberg, 1866)	Washington, USA	A10039	**SRR11710112**
Sabellidae	*Hypsicomus* sp.	Florida, USA	A12383	**SRR11710111**
Sabellidae	*Myxicola affinis* Bush, 1905	California, USA	A8662	**SRR11710110**
Sabellidae	*Pseudopotamilla oculifera* (Leidy, 1855)	Massachusetts, USA	A12375	**SRR11710109**
Sabellidae	*Sabella pavonina* Savigny, 1822	(Andrade et al., 2015)	A1015	SRR2005708
Sabellidae	*Sabellastarte magnifica* (Shaw, 1800)	Belize	A12379	**SRR11710108**
Spionidae	*Prionospio* sp.	(Andrade et al., 2015)	A5920	SRR2017831

2. Material and methods

2.1. Taxon sampling and transcriptome sequencing

Nineteen species of Sabellida were sampled for *de novo* transcriptome sequencing in addition to a published transcriptome. The available transcriptome of *Prionospio* sp. (Spionidae) was selected as an outgroup based on recent phylogenomic studies that supported Spionidae as sister to Sabellida (Andrade et al., 2015; Struck et al., 2015). Specimen details, voucher information and accession numbers for the sequence data are summarized in Table 1. *Amphiglena* has cryptic diversity in the Mediterranean (Tilic et al., 2019). The COI sequence extracted from the transcriptome falls within the *A. mediterranea* clade III in Tilic et al. (2019) and since the sequenced specimen was not collected from the type locality of *A. mediterranea* (Leydig, 1851) we have chosen to use the name *Amphiglena* cf. *mediterranea* for this terminal.

Tissues were preserved in RNAlater, quickly after the animals were collected and stored at −80 °C until RNA extraction. For larger specimens, only the anterior region of the animals were sampled. For minute specimens (Fabriciidae, *Amphicorina anneae* and *Amphiglena* cf. *mediterranea*) multiple whole individuals were pooled during extraction. RNA extractions were performed from Trizol, using Direct-zol RNA Miniprep Kit with an in-column DNase treatment (Zymo Research) mRNA was isolated with Dynabeads mRNA Direct Micro Kit (Invitrogen).

RNA concentration was estimated using Qubit RNA broad range assay kit, and quality was assessed using RNA ScreenTape with an Agilent 4200 TapeStation on an Agilent Bioanalyzer 2100. The measurements were used to customize library preparation protocols following manufacturer's instructions. Library preparation was performed with a KAPA-Stranded RNA-Seq kit, targeting an insert size in the range of 200–300 bp. Quality, concentration and molecular weight distribution of libraries were assessed using a DNA ScreenTape, a Bioanalyzer 2100. Libraries were sequenced in multiplexed 100 bp paired-end runs using Illumina HiSeq 4000, with 8 libraries per lane, resulting in an average sequencing depth of 35.5 million reads (range: 26.8–44.7 million). In order to minimize read crossover, we employed 10 bp sequence tags designed to be robust to indel and substitution errors (Faircloth and Glenn, 2012). All sequence data have been deposited in the NCBI sequence read archive (SRA) (Table 1) with BioProject accession number PRJNA630201.

2.2. Transcriptome assembly and phylogenetic analyses

Sequence adapters and low-quality regions were removed from the raw reads of each species using Trimmomatic v.0.36 (Bolger et al., 2014) with default parameters. All transcriptomes were *de novo* assembled and orthologous gene sequences were identified using the automated pipeline Agalma v.1.0.1 (Dunn et al., 2013; Guang et al., 2017). In brief, Agalma assembles transcripts with Trinity (Grabherr et al., 2011), maps reads with Bowtie2 (Langmead and Salzberg, 2012) and identifies orthologs using the topology of gene trees and DendroPy (Sukumaran and Holder, 2010). Assembly statistics and the number of orthologous genes identified by Agalma are summarized in Table S1.

Maximum Likelihood (ML) phylogenetic analyses were conducted using concatenated sequence matrices with 80% (1,070,842 amino acid (AA) sites; 3,015 genes) and 90% (435,040 AA sites; 1207 genes) occupancy matrices. Unpartitioned sequence matrices were analyzed with RAxML v.8.2.10 (Stamatakis, 2014) on the CIPRES Science Gateway (Miller et al., 2010), using the PROTGAMMAAUTO model setting and 1000 non-parametric bootstrap inferences with 10 distinct randomized maximum parsimony starting trees.

Furthermore, amino acid substitution model testing and partition merging (MF + MERGE) and subsequent ML analyses were run in IQ-TREE v.1.6 (Nguyen et al., 2015) for both supermatrices with ultrafast bootstrap approximation (UFboot) with 1000 replicates (Hoang et al., 2017).

2.3. Gene tree analyses and data exploration

Amino acid sequences of 13,393 loci were used to reconstruct the gene trees using RAxML v. 8.2.11 with 20 initial randomized parsimony trees. We used the best protein evolution model inferred by the Perl script 'ProteinModelSelection.pl' bundled in RAxML. In all of our experiments, the species trees were estimated using ASTRAL v.5.6.9 (Zhang et al., 2018), and to measure support for the branches of the inferred species trees, the local posterior probabilities (LocalPP) values (Sayyari and Mirarab, 2016) were computed. We investigated the effects of the set of genes used for species tree estimation in two ways: *i*) We selected random subsets of genes to study the impact of the number of loci examined. *ii*) We curated subsets of genes with low to high divergence (measured by their tree diameter; i.e., maximum tip-to-tip branch lengths) to study how phylogenetic information in a species tree analyses is impacted by divergence, and how a biased sampling does or

does not impact results.

For the random selection, we chose 1%, 2%, 3%, … 90% of genes and repeated the process 20 times. To perform divergence-based selection, we computed all tip-to-tip (patristic) distances of each gene tree and used the 80-percentile of these tip-to-tip distances as an approximation of the diameter of the gene tree. We did not use the maximum value to avoid inflating tree diameter by including outlier tips with long branches (Uyen and Mirarab, 2018). We then sorted genes based on the ascending order of their approximate diameter and created subsets of data based on this order. A subset is characterized as $y \pm x\%$ and included genes from $(y - x)\%$ to $(y + x)\%$ of the ordered list of genes. Thus, $y$represents the bias in the gene selection (section in the ordered list of tree diameter) and x represents the percentage of genes selected. For example, the subset $25 \pm 10\%$ included genes in the 15% to 35% range of the ordered list. We explored y = 10%, 25%, 50%, 60%, and 75% and vary x between 0.5% and 45%. This way, we could select sets of gene trees of a specified range of tree diameters (y) and vary the number of gene trees (x).

We used DiscoVista (Sayyari et al., 2018) to examine the species tree topology focusing on the three main clades (Fabriciidae, Sabellidae, Serpulidae) but also on the subclades of Sabellidae (Myxicolinae, Amphiglenini, Myxicolini, Sabellinae) and their combinations. For each clade, we visualized whether it was observed in the species tree as a monophyletic group. For clades that were present, we show the LocalPP support. When a clade was not recovered (i.e., was not monophyletic), we collapsed branches with LocalPP of 0.95 or less and tested if the contracted tree was compatible with the monophyly of the clade. If the contracted tree was compatible with monophyly, we say that the monophyly is weakly rejected, and otherwise called it strongly rejected. When a clade of interest was not included in the main ASTRAL result, we used the constrained version of ASTRAL (Maryam and Mirarab, 2020) to enforce the monophyly of that clade and to compute its "hidden support" (localPP).

3. Results

3.1. Data analyses and matrix assembly

Assembly statistics and values to assess the quality of each transcriptome, together with the number of orthologous genes identified are summarized in Table S1. The smallest number of reads in the total data set was 26.8 million reads for *Sabellastarte magnifica* (Shaw, 1800) (assembled into 19,577 contigs), whereas the largest one was the outgroup *Prionospio* sp., with 68.6 million reads (assembled into 37,158 contigs).

Total number of orthologous genes identified across the terminals was 13,393 (ranging from 2,254 in *Prionospio* sp. to 8,193 in *Branchiomma conspersum* (Ehlers, 1887)). For phylogenetic reconstruction two sequence matrices were curated, one with 80% occupancy (1,070,842 AA sites; 3,015 genes) and the other with 90% occupancy (435,040 AA sites; 1207 genes). All taxa except for the outgroup terminal *Prionospio* sp. (990 genes) and the serpulid *Salmacina tribranchiata* (Moore, 1923) (1316 genes) had over 2000 loci (average: 5896.4, range 2254–8193) in the 80% occupancy matrix (Fig. 2). These are among the largest and most complete data matrices ever analyzed in a phylogenomic context for an annelid clade. The recovered tree topology was identical, regardless of the method and sequence matrix analyzed. As support was optimal for all inferred nodes, no further analyses with different occupancy matrices were evaluated.

3.2. Phylogeny reconstruction based on supermatrices

The unpartitioned RAxML analyses of both data matrices, as well as the partitioned IQTREE analyses using integrated model selection and partition merging, reconstructed the same topology with 100% bootstrap values at each node (Fig. 2). All three clades within Sabellida; Fabriciidae, Serpulidae and Sabellidae, were recovered as monophyletic. Serpulidae + Sabellidae formed a clade that was always the sister group to Fabriciidae (Fig. 2).

Within Fabriciidae, the freshwater *Manayunkia occidentalis* Atkinson et al., 2020 was supported as the sister taxon to a marine clade of *Fabricinuda* sp. and *Novafabricia brunnea* (Hartman, 1969). The two serpulid terminals included in the sampling (*Salmacina tribranchiata* (Moore, 1923) and *Pomatostegus stellatus* (Abildgaard, 1789)) were a clade.

Within Sabellidae we recovered two major clades: Sabellinae (*sensu stricto*, with the type genus *Sabella* Linnaeus, 1767); including *Branchiomma conspersum* (Ehlers, 1887), *Sabellastarte magnifica* (Shaw, 1800), *Bispira* spp., *Sabella pavonina* Savigny, 1822, *Hypsicomus* sp. and *Anamobaea orstedii* Krøyer, 1856. We have broadened the previously monotypic Myxicolinae (type genus *Myxicola*) to include two tribes (Fig. 2). Within Myxicolinae *Amphicorina anneae* (Rouse, 1994), *Chone* sp. and *Myxicola affinis* Bush, 1905 share the presence of a glandular girdle around chaetiger 2 and we make this clade the tribe Myxicolini. The clade consisting here of *Amphiglena* cf. *mediterranea*, *Acromegalomma coloratum* (Chamberlin, 1919), *Eudistylia vancouveri* (Kinberg, 1866) and *Pseudopotamilla oculifera* (Leidy, 1855) is named Amphiglenini (new tribe, type genus *Amphiglena*; diagnosed below).

3.3. Gene tree analyses and data exploration

Given the strong support of the sister group relationship of Sabellidae and Serpulidae found here, in contrast with a previous study based on a few genes (Kupriyanova and Rouse, 2008), we were interested in exploring whether this new relationship could be explained by the increased number of genetic loci and whether genes with lower or higher divergence provide a stronger resolution. The monophyly of Fabriciidae, Sabellidae, and Serpulidae was supported in all analyses with subsets of data except for one replicate with 1% of genes (=23 genes) chosen randomly (Fig. 3a). These three main clades also had high support across all subsets except occasionally when the number of genes was very small (i.e., 3% or lower). In most analyses, Sabellidae + Serpulidae were found as sister clades, but the recovery of this group and its support depended on the gene selection. With randomly selected genes, Sabellidae + Serpulidae continued to be supported in most analyses (Fig. 3a). However, in three (out of 600) runs, ASTRAL rejected this group with high support. In one of these three cases, surprisingly, Sabellidae was not monophyletic. To recover the Sabellidae + Serpulidae grouping consistently (i.e. across all replicates), more than 50% of the genes (> 1102 genes) were needed. Interestingly, to have full support in every random replicate, at least 90% of the genes were needed (> 1984 genes). Even with 80% of genes (1764 genes) randomly selected, Sabellidae + Serpulidae occasionally had less than perfect support. Among the few cases that did not recover Sabellidae + Serpulidae, most found Serpulidae + Fabriciidae instead, and the rest found Sabellidae + Fabriciidae. Part of the reason for increased consistency of results as the number of genes increase can be the pseudo-replication (81% of genes are expected to be identical in two selections of 90% of genes). Nevertheless, increased pseudo-replication does not explain increased localPP support when additional loci are sampled (Fig. 3a), which is a function of the number of genes and the amount of discordance.

With divergence-based subsampling, Sabellidae + Serpulidae was never strongly rejected and had strong support in all runs with more than 50% of genes (Fig. 3b). With low-divergence genes (i.e., y = 10%) Serpulidae was sister to Fabriciidae in about half of the replicates, and in the rest Sabellidae was sister to Serpulidae. However, the support was low both for recovering Sabellidae + Serpulidae (maximum 0.64 LocalPP) and for rejecting it. Thus, genes with the lowest divergence did not have enough signal to resolve this relationship. As we increased the divergence, we observed fewer cases that failed to put Sabellidae and Serpulidae together, and there were more cases with high support

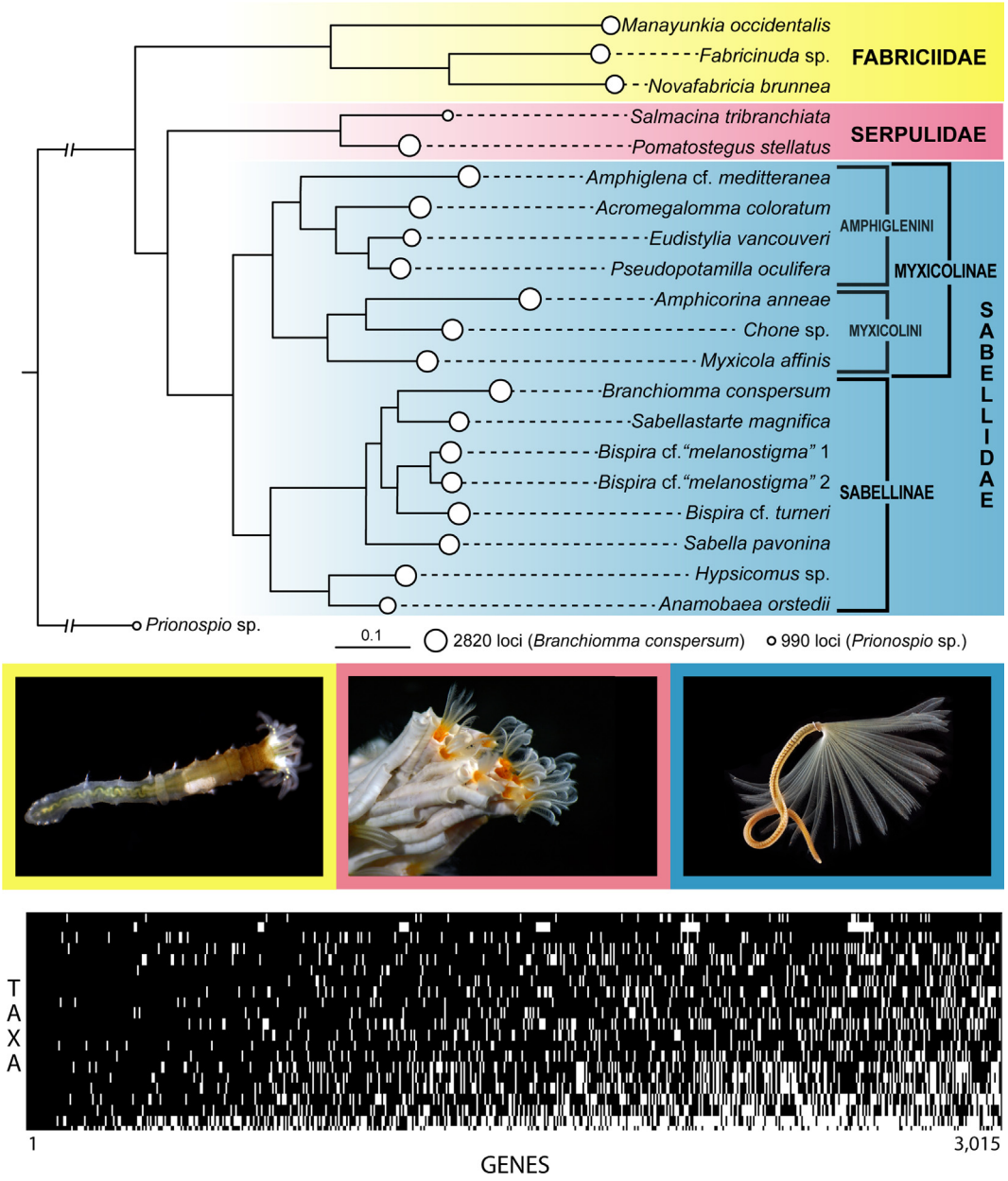

Fig. 2. Maximum likelihood tree showing the relationships of Sabellida. Topology and branch lengths are based on the concatenated 80% occupancy matrix (completeness shown at the bottom). Tree topology identical to the concatenated 90% occupancy matrix and to all 13,393 gene trees summarized with ASTRAL. All nodes had full bootstrap support and ASTRAL local posterior probabilities regardless of the dataset and inference method. Photos: *Manayunkia occidentalis* (Fabriciidae), *Salmacina australis* (Serpulidae), *Sabella pavonina* (Sabellidae).

for their monophyly. For example, with the highest levels of divergence (e.g., y = 75%), only one replicate that included 2% of genes failed to recover the Sabellidae + Serpulidae clade. The only condition when this group was recovered with high support with 20% of genes or less was when those genes had the highest levels of divergence (e.g., y = 75%). To summarize, inclusion of more genes leads to a more stable topology with higher support as expected, and also including higher divergence genes leads to stronger resolution.

We explored the stability of the sabellid species tree (Fig. 2) using constrained ASTRAL search to investigate "hidden support" for the main clades of the species tree in different gene subsets. Hidden support for alternative topologies (Fabriciidae + Sabellidae and Fabriciidae + Serpulidae) decreased with the number of loci sampled (Fig. 4a). We also detected little hidden support for alternative topologies (Fabriciidae + Sabellidae and Fabriciidae + Serpulidae) with highly divergent gene trees (Fig. 4b). However, some level of hidden support was observed for Fabriciidae + Serpulidae with low divergence genes. Monophyly of the subcaldes within Sabellidae (Sabellinae, Myxicolinae, Amphiglenini and Myxicolini) were always highly supported, regardless of divergence and number of genes used (Fig. 4a, b). Alternative combinations within Sabellidae (Amphiglenini + Sabellinae and Myxicolini + Sabellinae) had no support in any of the analyses.

4. Discussion

Our phylogenomic study using a large dataset with 21 transcriptomes resolved the main relationships within Sabellida with strong support (Figs. 2, 3). Our results confidently placed Fabriciidae as sister to a clade of Serpulidae + Sabellidae. This result differs from previously published studies that found Fabriciidae to be the sister group to Serpulidae, though based on few genes (Kupriyanova and Rouse,

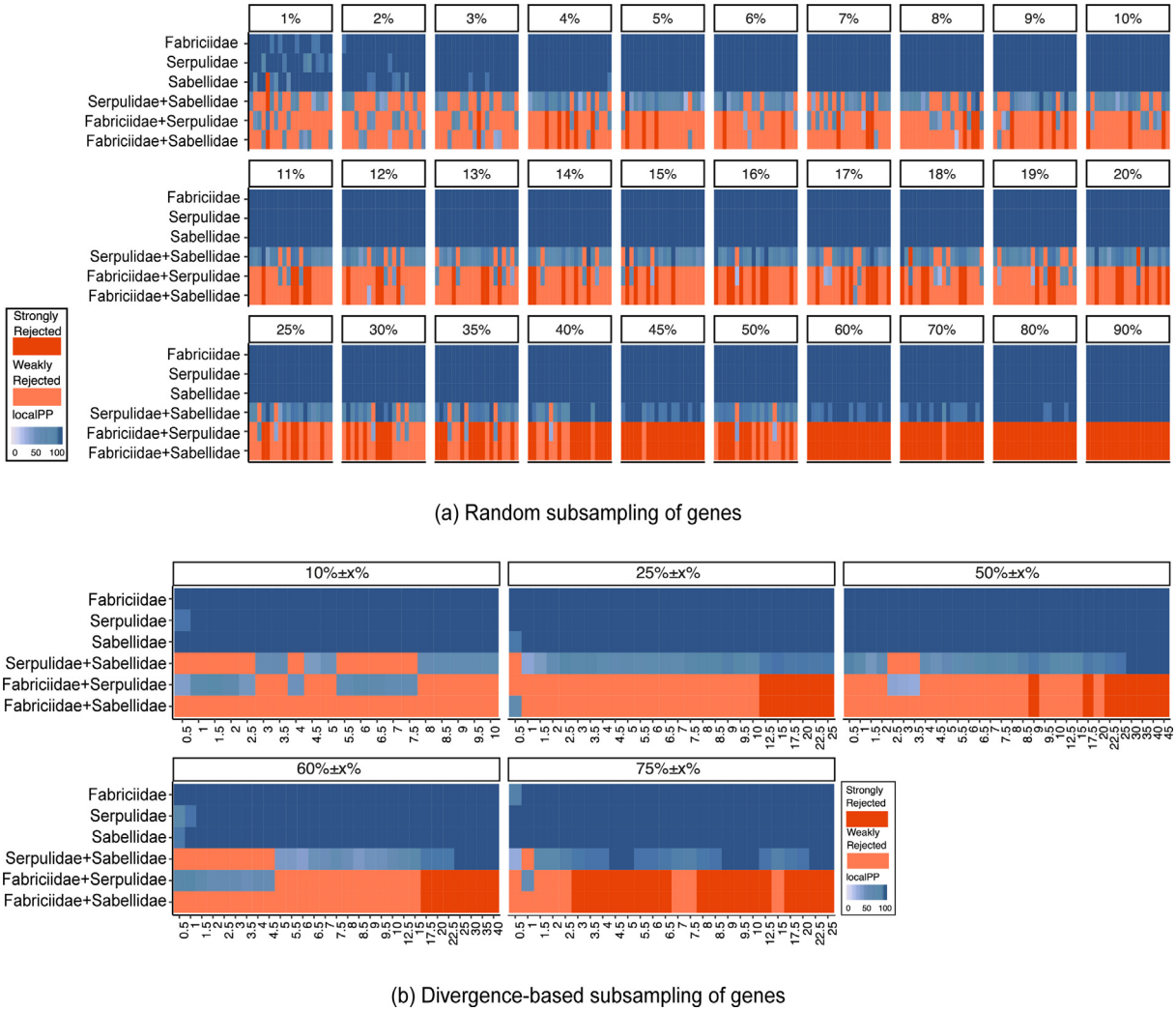

Fig. 3. Summary of ASTRAL species trees obtained from subsets of gene trees. We subsampled genes either at random (a) or based on similar divergence levels (b). For each subsample, we show the recovery of the main groups in the ASTRAL tree, using shades of blue to show LocalPP for recovered groups and light or dark orange for clades rejected weakly (< 0.95 LocalPP) or strongly, respectively. (a) Random sampling of a certain percentage of genes (boxes), repeated 20 times each (different columns). (b) Divergence-based sampling; we chose genes from (y-x)% to (y + x)% in the list of genes sorted ascendingly according to the 80%-percentile gene tree tip-to-tip distance. Boxes: levels of divergence (y); x-axis: half the number of genes selected (x). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2008, Capa et al., 2011a). Our analyses suggest that this discrepancy is due to there being insufficient data in these previous studies (Fig. 3). In fact, it is not surprising to see the change in the fabriciid phylogenetic position as the gene tree analyses showed that more than 50% of the genes were necessary to consistently recover Sabellida + Serpulidae (Fig. 3a). When low divergence genes were sampled, or only a small number of genes were analyzed (less than 3%), there was some hidden support for the Fabriciidae + Serpulidae relationship. This may explain the seemingly high support (91 bs ML, 1.0 pp) in Kupriyanova and Rouse's (2008) phylogeny based on three nuclear genes. Fabriciidae + Serpulidae in Capa et al. (2011a) was not highly supported (and had even less data than Kupriyanova and Rouse, 2008) and Fitzhugh's support for Sabellidae, including Sabellinae and Fabriciinae, was based on only a few chaetal characters (Fitzhugh, 1989, 1991). Our findings further support the nomenclatural changes by Kupriyanova and Rouse (2008), who raised Fabriciidae to family rank, and are also crucially important for our interpretation of morphology and life-history evolution in Sabellidae.

The monophyly of each of the three major clades of Sabellida was highly supported (Figs. 3, 4), and can be further substantiated by a number of morphological apomorphies. All Fabriciidae have 'branchial hearts', an abdomen with 3 or 4 segments and gametogenesis only occurs in the thorax (Capa et al., 2011a; Fitzhugh, 1998, 1991, 1989). The calcareous tube found in all serpulids is a clear apomorphy for the group as well as the presence of a thoracic membrane and thoracic uncini with no manubrium (Kupriyanova et al., 2006). Sabellidae *sensu stricto* is unified by the presence of vacuolated, stabilizing cells inside the radioles (a radiolar skeleton), the dorsal fusion of the radiolar lobes and the presence of dorsal and ventral lips (see Fitzhugh, 1989; Capa et al., 2011b). There are no clear morphological synapomorphies for the clade Sabellidae + Serpulidae. Smith (1991) proposed a phylogeny of Sabellida based on morphology, and similar to our findings, placed Sabellidae (as Sabellinae) in a clade with Serpulidae, and Fabriciidae (as Fabriciinae) was the sister to that clade. However, in his study Fabriciinae included many genera (such as *Amphicorina*, *Jasminiera* etc.) that are now part of Sabellidae (Fitzhugh, 1989). It is also important to note that the synapomorphies Smith (1991) proposed for "Sabellinae + Serpulidae" were somewhat dubious (i.e. suspension-feeding only, presence of branchial eye-spots) and mainly based on absences (i.e. lack of pygidial eye-spots, absence of thoracic long shafted uncini). The position of *Manayunkia occidentalis* as the sister taxon to the remaining two fabriciids is consistent with the most recent

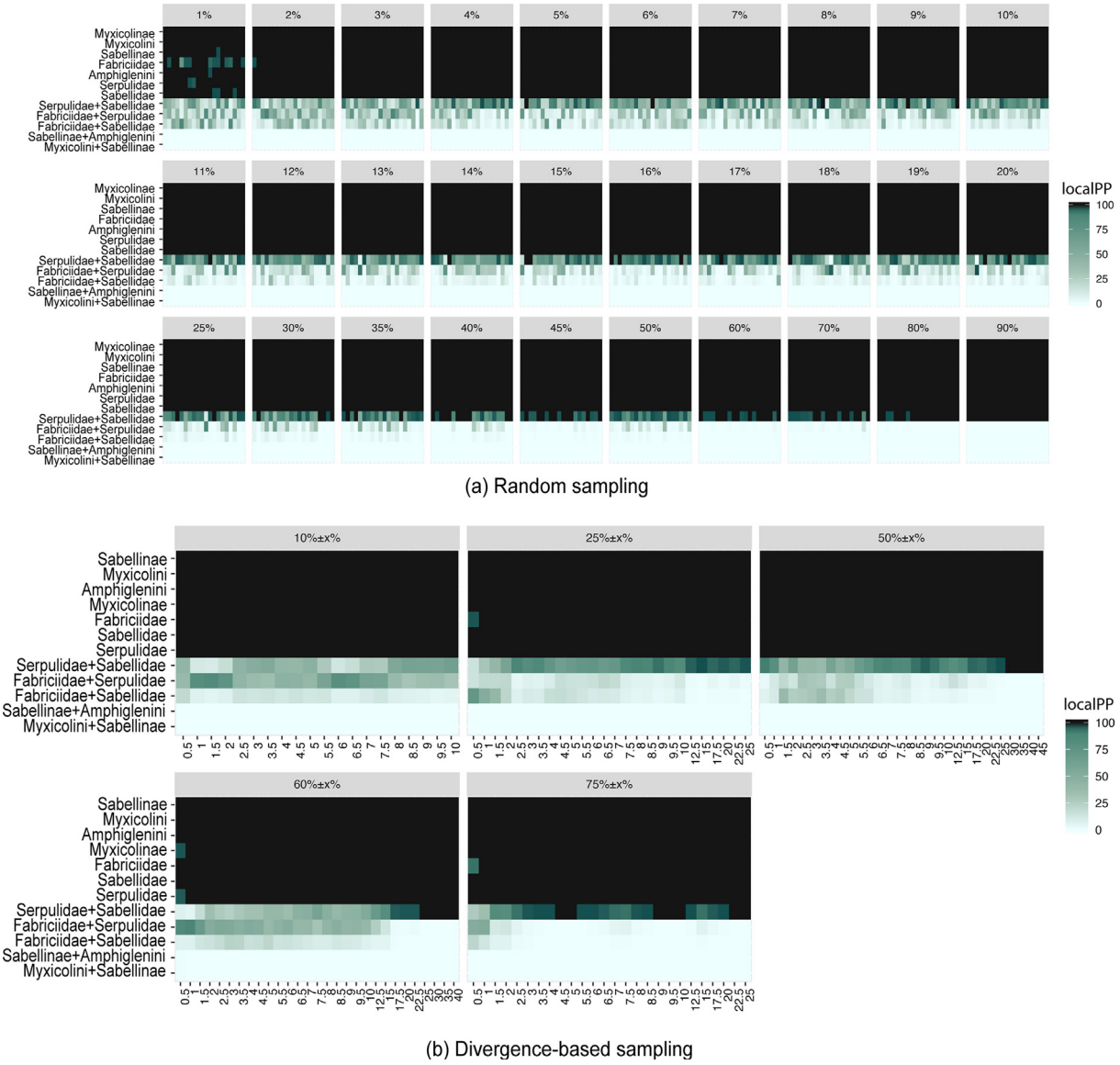

Fig. 4. Constrained ASTRAL trees on subsets of genes to interrogate hidden support. We inferred constrained ASTRAL species trees from subsets of genes while forcing the monophyly of three main hypothesised groups: Fabriciidae + Serpulidae, or Fabriciidae + Sabellidae, or Serpulidae + Sabellidae and of the three sabellid subclades: Myxicolinae, or Sabellinae + Amphiglenini, or Myxicolinii + Sabellinae. For every subsampled set of genes, we show LocalPP of the main groups in the constrained ASTRAL tree, using shades of green distributed in a log scale to emphasize differences among high support values; to distinguish, we use black for 100% LocalPP. Panels are labeled as the previous figure.

fabriciid phylogeny and the taxon sampling spans the major fabriciid clades (Huang et al., 2011). Similarly, the two serpulid species included in our analyses were chosen to span the serpulid phylogeny as currently understood (Kupriyanova et al., 2006).

Sabellidae was split into two major clades (Fig. 2) and we apply two previously erected subfamilial names for them; Sabellinae and Myxicolinae. In Capa et al. (2011a), two main clades of Sabellidae were also recovered though with low support; their Clade IB, which is equivalent to our Myxicolinae and Clade IIB, which is equivalent to our Sabellinae. Myxicolinae corresponds to the grade identified as IIa in Fitzhugh (1989) and Sabellinae to a grade comprising his clades IIIa and IV. Our analysis shows that the topology recovered within Sabellidae is highly stable and there is no hidden support for alternative topologies, regardless of the number of randomly subsampled genes and also regardless of divergence (Fig. 4). Within Myxicolinae, which had previously been a monotypic taxon for *Myxicola* and usually regarded as a junior synonym of Sabellinae (see Fitzhugh, 1989) we also name two clades. Myxicolini Rioja, 1923 contains *Myxicola*, *Amphicorina* and *Chone* in our tree (Fig. 2) and corresponds to Clade I in Capa et al. (2011a). The monophyly of Myxicolini is further supported as all three genera, and those in Clade I of Capa et al. (2011a) have a glandular girdle around their second chaetiger (Fitzhugh, 1989), a likely apomorphy for this clade. The other clade we recovered within Myxicolinae is referred to here as Amphiglenini new tribe:

Diagnosis: Sabellidae with radiolar crown with two or more rows of cells in radiolar skeleton; ventral lips absent or present; dorsal pinnular appendages absent or present; parallel lamellae absent or present; unpaired compound radiolar eyes absent or present. Posterior peristomial ring collar absent or present; paired, ventral basal flanges from posterior peristomial ring to ventral-most radioles present or absent. Companion chaetae usually present. Superior thoracic notochaetae broadly or narrowly hooded; broadly hooded chaetae in inferior thoracic neuropodia absent or present; elongate, broadly hooded chaetae in the anterior and posterior rows of abdominal neuropodia absent or present.

Type genus *Amphiglena* Claparède, 1864

Amphiglenini appears as a grade in Capa et al. (2011a) with respect to their Clade I, i.e. Myxicolinae, but none of the relevant nodes in their results have high support. In Fitzhugh (1989) Amphiglenini corresponds to Clades V, VI and VII, for which he identified a single apomorphy, the loss of a palmate membrane. The delineation and membership of Sabellinae, Myxicolinae, Myxicolini and Amphiglenini will be further explored in a future study employing high throughput sequencing and dense sampling of sabellid genera, in which we aim to elucidate the fine-scale phylogenetic relationships within sabellids (Tilic et al. in prep.). We will also assess the application of additional names within Sabellinae. Further work is also required to establish morphological support and synapomorphies for the major clades recovered within Sabellidae. Previous studies have shown that a detailed taxonomic revision of sabellid genera is required, as many of them appear to be paraphyletic (Capa et al., 2011a). By including a number of the main lineages of Sabellidae, this paper provides the necessary groundwork and backbone to fully resolve the phylogeny and systematics of this diverse taxon.

Evolution of reproductive modes in Sabellidae have been the focus of previous studies (e.g., Rouse and Fitzhugh, 1994; Rouse et al., 2006). The new sabellid phylogeny has major implications for how life-history data are interpreted and allows these questions to be addressed with strong confidence. In Rouse and Fitzhugh's (1994) study Sabellariidae was used as the outgroup, which have broadcast spawning and planktotrophic larvae. Both Sabellidae and Fabriciidae only show lecithotrophic larval development, whereas in Serpulidae both planktotrophic and lecithotrophic larvae are known. Various forms of brooding and broadcast spawning can be found in Serpulidae (Kupriyanova et al., 2001) as well as Sabellidae (Rouse and Fitzhugh, 1994). Fabriciidae, however are all small-sized brooders and this taxon would now appear to be the appropriate root for an analysis of the evolution of reproductive modes in Sabellidae and Serpulidae.

CRediT authorship contribution statement

Ekin Tilic: Conceptualization, Investigation, Formal analysis, Visualization, Writing - original draft. **Erfan Sayyari:** Methodology, Formal analysis, Visualization, Writing - review & editing. **Josefin Stiller:** Investigation, Formal analysis, Writing - review & editing. **Siavash Mirarab:** Conceptualization, Methodology, Writing - review & editing, Supervision. **Greg W. Rouse:** Conceptualization, Investigation, Writing - review & editing, Supervision.

Data availability statement

Data for this article is submitted to Mendeley Data: https://doi.org/10.17632/khjmr4mh34.2.

Acknowledgements

ET was supported by a personal postdoc fellowship of the German Research Foundation, DFG (TI 973/1-1). We gratefully acknowledge Dewy White for her support for this project. We also thank Avery Hatch and Jose Ignacio Carvajal for their help with the molecular lab work and to Charlotte Seid for handling and cataloging of specimens at the Benthic Invertebrate Collection of Scripps Institution of Oceanography. Thanks to Stephen Atkinson and Maria Cristina Gambi for providing samples of *Manayunkia occidentalis* and *Amphiglena* cf. *mediterranea*, respectively. Also, thanks to Phil Zerofski for collecting *Salmacina tribranchiata*, and to Nerida Wilson and Fredik Pleijel for help collecting *Amphicorina anneae* and *Fabricinuda* sp. Thanks also to Fredrik Pleijel for the photo of *Sabella pavonina*.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2020.106892.

References

Andrade, S.C.S., Novo, M., Kawauchi, G.Y., Worsaae, K., Pleijel, F., Giribet, G., Rouse, G.W., 2015. Articulating "archiannelids": Phylogenomics and annelid relationships, with emphasis on meiofaunal taxa. Mol. Biol. Evol. 32, 2860–2875. https://doi.org/10.1093/molbev/msv157.

Atkinson, S.D., Bartholomew, J.I., Rouse, G.W., 2020. The invertebrate host of salmonid fish parasites *Parvicapsula minibicornis* (Cnidaria: Myxozoa), is a novel fabriciid annelid, *Manayunkia occidentalis* sp. nov. (Sabbelida: Fabriciidae). Zootaxa 4751 (2), 310–320.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Brown, S., Rouse, G., Hutchings, P., Colgan, D., 1999. Assessing the usefulness of histone H3, U2 snRNA and 28S rDNA in analyses of polychaete relationships. Aust. J. Zool. 47, 499–516. https://doi.org/10.1071/ZO99026.

Capa, M., Hutchings, P., Aguado, T.M., Bott, N.J., 2011a. Phylogeny of Sabellidae (Annelida) and relationships with other taxa inferred from morphology and multiple genes. Cladistics 27, 449–469. https://doi.org/10.1111/j.1096-0031.2010.00341.x.

Capa, M., Nogueira, J.M. de M., Rossi, M.C.S., 2011b. Comparative internal structure of dorsal lips and radiolar appendages in Sabellidae (Polychaeta) and phylogenetic implications. J. Morphol. 272, 302–319. https://doi.org/10.1002/jmor.10914.

Colgan, D., Hutchings, P., Braune, M., 2006. A multigene framework for polychaete phylogenetic studies. Org. Divers. Evol. 6, 220–235. https://doi.org/10.1016/j.ode.2005.11.002.

Dales, R.P., 1962. The polychaete stomodeum and the interrelationships of the families of the Polychaeta. Proc. Zool. Soc. Lond. 139, 289–328. https://doi.org/10.1111/j.1469-7998.1962.tb01837.x.

Dunn, C.W., Howison, M., Zapata, F., 2013. Agalma: an automated phylogenomics workflow. BMC Bioinformatics 14, 330. https://doi.org/10.1186/1471-2105-14-330.

Faircloth, B.C., Glenn, T.C., 2012. Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. PLoS ONE. https://doi.org/10.1371/journal.pone.0042543.

Fauchald, K., 1977. The polychaete worms: definitions and keys to the orders, families and genera. Natural History Museum of Los Angeles County Los Angeles, CA, USA.

Fitzhugh, K., 1998. New fan worm genera and species (Polychaeta, Sabellidae, Fabriciinae) from the western Pacific, and cladistic relationships among genera. Zool. Scr. 27, 209–245. https://doi.org/10.1111/j.1463-6409.1998.tb00438.x.

Fitzhugh, K., 1991. Further revisions of the Sabellidae subfamilies and cladistic relationships among the Fabriciinae (Annelida: Polychaeta). Zool. J. Linn. Soc. 102, 305–332. https://doi.org/10.1111/j.1096-3642.1991.tb00004.x.

Fitzhugh, K., 1989. A systematic revision of the Sabellidae-Caobangiidae-Sabellongidae complex (Annelida: Polychaeta). Bull. Am. Mus. Nat. Hist. 192, 1–104.

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. 39, 644–652. https://doi.org/10.1038/nbt.1883.

Guang, A., Howison, M., Zapata, F., Lawrence, C., Dunn, C., 2017. Revising transcriptome assemblies with phylogenetic information in Agalma1.0. bioRxiv https://doi.org/10.1101/202416.

Helm, C., Beckers, P., Bartolomaeus, T., Drukewitz, S.H., Kourtesis, I., Weigert, A., Purschke, G., Worsaae, K., Struck, T.H., Bleidorn, C., 2018. Convergent evolution of the ladder-like ventral nerve cord in Annelida. Front. Zool. 15, 36. https://doi.org/10.1186/s12983-018-0280-y.

Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., Vinh, L.S., 2017. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522. https://doi.org/10.1093/molbev/msx281.

Huang, D., Fitzhugh, K., Rouse, G.W., 2011. Inference of phylogenetic relationships within Fabriciidae (Sabellida, Annelida) using molecular and morphological data. Cladistics 27, 356–379. https://doi.org/10.1111/j.1096-0031.2010.00343.x.

Jones, M.L., 1973. Flowers of Sea. Nat. Hist. 82 (3), 40–45.

Kieselbach, D., Hausen, H., 2008. Chaetal arrangement provides no support for a close relationship of Sabellidae and Sabellariidae (Annelida). J. Morphol. 117, 104–117. https://doi.org/10.1002/jmor.10537.

Knight-Jones, P., 1981. Behaviour, setal inversion and phylogeny of Sabellida (Polychaeta). Zool. J. Linn. Soc. 64, 201–240.

Kupriyanova, E.K., Macdonald, T.A., Rouse, G.W., 2006. Phylogenetic relationships within Serpulidae (Sabellida, Annelida) inferred from molecular and morphological data. Zool. Scr. 35, 421–439. https://doi.org/10.1111/j.1463-6409.2006.00244.x.

Kupriyanova, E.K., Nishi, E., Ten Hove, H.A., Rzhavsky, A.V., 2001. Live-history patterns in serpulimorph polychaetes: ecological and evolutionary perspectives. In: Gibson, R.N., Barnes, M., Atkinson, R.J.A. (Eds.), Oceanography and Marine Biology: An Annual Review. Taylor and Francis, Boca Raton, pp. 1–101.

Kupriyanova, E.K., Rouse, G.W., 2008. Yet another example of paraphyly in Annelida: molecular evidence that Sabellidae contains Serpulidae. Mol. Phylogenet. Evol. 46, 1174–1181. https://doi.org/10.1016/j.ympev.2007.10.025.

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

Maryam, R., Mirarab, S., 2020. Forcing external constraints on tree inference using ASTRAL. BMC Genom. 21, 218. https://doi.org/10.1186/s12864-020-6607-z.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing

Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA, pp. 1–8.
Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.
Rouse, G., Fitzhugh, K., 1994. Broadcasting fables: Is external fertilization really primitive? Sex, size, and larvae in sabellid polychaetes. Zool. Scr. 23, 271–312. https://doi.org/10.1111/j.1463-6409.1994.tb00390.x.
Rouse, G.W., Fauchald, K., 1997. Cladistics and polychaetes. Zool. Scr. 26, 139–204. https://doi.org/10.1111/j.1463-6409.1997.tb00412.x.
Rouse, G.W., Kupriyanova, E., and Nishi, E., 2006. Sabellida. In: Rouse, G.W., Pleijel, F. (Eds.), Reproductive Biology and Phylogeny of Annelida. pp. 521–564. (Science Publishers, Inc.: Enfield, New Hampshire).
Rousset, V., Pleijel, F., Rouse, G.W., Erséus, C., Siddall, M.E., 2007. A molecular phylogeny of annelids. Cladistics 22, 1–23. https://doi.org/10.1111/j.1096-0031.2006.00128.x.
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. https://doi.org/10.1111/j.1096-0031.2004.00039.x.
Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33, 1654–1668. https://doi.org/10.1093/molbev/msw079.
Sayyari, E., Whitfield, J.B., Mirarab, S., 2018. DiscoVista: Interpretable visualizations of gene tree discordance. Mol. Phylogenet. Evol. 122, 110–115. https://doi.org/10.1016/j.ympev.2018.01.019.
Smith, R.S., 1991. Relationships within the order Sabellida (Polychaeta). Ophelia Suppl 5, 249–260.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
Stiller, J., Tilic, E., Rousset, V., Pleijel, F., Rouse, G.W., 2020. Spaghetti to a Tree: A Robust Phylogeny for Terebelliformia (Annelida) Based on Transcriptomes, Molecular and Morphological Data. Biology 9 (4), 73. https://doi.org/10.3390/biology9040073.
Struck, T.H., Golombek, A., Weigert, A., Franke, F.A., Westheide, W., Purschke, G., Bleidorn, C., Halanych, K.M., 2015. The evolution of annelids reveals two adaptive routes to the interstitial realm. Curr. Biol. 25, 2000–2006. https://doi.org/10.1016/j.cub.2015.06.007.
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. https://doi.org/10.1186/1471-2148-7-57.
Sukumaran, J., Holder, M.T., 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26, 1569–1571. https://doi.org/10.1093/bioinformatics/btq228.
Tilic, E., Feerst, K., Rouse, G.W., 2019. Two new species of *Amphiglena* (Sabellidae, Annelida), with an assessment of hidden diversity in the Mediterranean. Zootaxa 4648 (2), 337–353. https://doi.org/10.11646/zootaxa.4648.2.8.
Uyen, M., Mirarab, S., 2018. TreeShrink: Fast and accurate detection of outlier long branches in collections of phylogenetic trees. BMC Genomics 19, 272. https://doi.org/10.1186/s12864-018-4620-2.
Weigert, A., Bleidorn, C., 2016. Current status of annelid phylogeny. Org. Divers. Evol. 16, 345–362. https://doi.org/10.1007/s13127-016-0265-7.
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. https://doi.org/10.1093/molbev/msu080.
Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. https://doi.org/10.1186/s12859-018-2129-y.
Zrzavý, J., Říha, P., Piálek, L., Janouskovec, J., 2009. Phylogeny of Annelida (Lophotrochozoa): total-evidence analysis of morphology and six genes. BMC Evol. Biol. 9, 189. https://doi.org/10.1186/1471-2148-9-189.