

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

# **Title: Molecular data from Orthonectid worms show they are highly degenerate members of phylum Annelida not phylum Mesozoa.**

**Authors:** Philipp H. Schiffer<sup>1</sup>, Helen E. Robertson<sup>1</sup>, Maximilian J. Telford<sup>1\*</sup>

**Affiliations:** <sup>1</sup> Centre for Life's Origins and Evolution, Department of Genetics, Evolution and Environment, University College London, Gower Street, London, WC1E 6BT, UK

\*Correspondence to: [m.telford@ucl.ac.uk](mailto:m.telford@ucl.ac.uk)

## **Summary:**

The Mesozoa are a group of tiny, extremely simple, vermiform endoparasites of various marine animals (Fig. 1). There are two recognised groups within the Mesozoa: the Orthonectida (Fig. 1a,b; with a few hundred cells including a nervous system made up of just 10 cells [1]) and the Dicyemids (Fig. 1c; with at most 42 cells [2]). They are classic 'Problematica' [3] - the name Mesozoa suggests an evolutionary position intermediate between Protozoa and Metazoa (animals) [4] and implies their simplicity is a primitive state, but molecular data have shown they are members of Lophotrochozoa within Bilateria [5-8] which would mean they derive from a more complex ancestor. Their precise phylogenetic affinities remain uncertain, however, and ascertaining this is complicated by the very fast evolution observed in genes from both groups, leading to the common systematic error of Long Branch Attraction (LBA) [9]. Here we use mitochondrial and nuclear gene

24 sequence data, and show beyond doubt that both dicyemids and orthonectids are members of  
25 the Lophotrochozoa. Carefully addressing the effects of systematic errors due to unequal  
26 rates of evolution, we show that the phylum Mesozoa is polyphyletic. While the precise  
27 position of dicyemids remains unresolved within Lophotrochozoa, we unequivocally identify  
28 orthonectids as members of the phylum Annelida. This result reveals one of the most extreme  
29 cases of body plan simplification in the animal kingdom; our finding makes sense of an  
30 annelid-like cuticle in orthonectids [1] and suggests the circular muscle cells repeated along  
31 their body [10] may be segmental in origin.

32

### 33 **Results**

34 Using a new assembly of available genomic and transcriptomic sequence data we identified an  
35 almost complete mitochondrial genome from *Intoshia linei* (2 ribosomal RNAs, 20 transfer  
36 RNAs and all protein coding genes apart from *atp8*) and recovered 9 individual mitochondrial  
37 gene containing contigs from *Dicyema japonicum* and from a second unidentified species  
38 (*Dicyema sp.*; *cox1*, 2, 3; *cob*; and *nad1*, 2, 3, 4, 5). *Cob*, *nad3*, *nad4*, and *nad5* had not  
39 previously been identified in any *Dicyema* species. All protostomes studied possess a unique,  
40 derived combination of amino acid signatures and conserved deletions in their mitochondrial  
41 NAD5 genes. Comparing the NAD5 protein coding regions of *Intoshia* and *Dicyema* to those  
42 of other Metazoa shows that both share almost all of the conserved protostome signatures [11]  
43 (Fig. 2a). This signature is significantly more complex than the two amino acids of the  
44 Lox5/DoxC signature from *Dicyema* previously published [11-13] and shows beyond doubt  
45 that both groups are protostomes.

46 It has been suggested that mesozoans are derived from the parasitic neodermatan flatworms. If  
47 this were correct mesozoans would be expected to share two changes in mitochondrial genetic

48 code that unite all rhabditophoran flatworms, where the triplet AAA codes for Asparagine (N)  
49 rather than the normal Lysine (K) and ATA codes for Isoleucine (I) rather than the usual  
50 Methionine (M) [14]. We inferred the mitochondrial genetic codes for *Dicyema* and *Intoshia*.  
51 Both groups have the standard invertebrate mitochondrial code arguing against a relationship  
52 with the parasitic rhabditophoran platyhelminths (table S1).

53 We next aligned the mitochondrial genes of *Intoshia* and three species of *Dicyema* with  
54 orthologs from a diversity of other Metazoans and concatenated these to produce a matrix of  
55 2,969 reliably aligned amino acids from 69 species. Phylogenetic analyses of this  
56 comparatively small data set is not expected to be as reliable as a much larger set of nuclear  
57 genes and aspects of the topology and observed branch lengths suggest it was affected by LBA  
58 (Fig. 2b). To reduce the effects of LBA on the inference of the affinities of the mesozoans we  
59 removed the taxa with the longest branches and considered the position of the dicyemids and  
60 orthonectid separately (as both are very long branched). We were unable to resolve the position  
61 of the dicyemids (although they are clearly lophotrochozoans), but found some support for  
62 placing the orthonectid *Intoshia linei* with the annelids (Fig. 2c and figures S1, 2). *Intoshia*  
63 *linei* has a unique mitochondrial gene order although the order of the genes nad1, nad6, and  
64 cob match that seen in the Lophotrochozoan ground plan and the early branching annelid  
65 *Owenia* (Fig. 2d).

66 We next assembled a data set of 469 orthologous genes, 227,187 reliably aligned amino acids,  
67 from 45 species of animals including *Intoshia linei* and two species of *Dicyema*. After  
68 removing positions in the concatenated alignment with less than 50% occupancy we had an  
69 alignment length of 190,027 amino acids and average completeness of ~68%. *Intoshia linei*  
70 was 65% complete, while *Dicyema japonicum* and *Dicyema sp.* were 77% and 43% complete  
71 respectively (table S2). We conducted a bayesian phylogenetic analyses of these data with the  
72 site heterogeneous CAT+G4 model in Phylobayes [15]. To provide an additional, conservative

73 estimate of clade support and to enable further analyses in a practical time frame, we also used  
74 jackknife subsampling. For each jackknife analysis we took 50 random subsamples of 30,000  
75 amino acids each and ran 2,000 cycles (phylobayes CAT+G4) per sample. All 50 subsamples  
76 were summarised into a single tree with the first 1800 trees from each excluded as ‘burnin’  
77 [16].

78 We observed strong support for a clade of Lophotrochozoa (excluding Rotifers) including both  
79 dicyemids and the orthonectid (Bayesian Posterior Probability (PP) = 1.0; Jackknife Proportion  
80 (JP) = 0.97) (Fig. 3a). The dicyemids and orthonectids were not each other’s closest relatives;  
81 the position of the dicyemids within the Lophotrochozoa was not resolved; they were not the  
82 sister group of the platyhelminths nor of the gastrotrichs in our analysis. The position of the  
83 orthonectid *Intoshia*, in contrast, was resolved as being within the clade of annelids (Fig. 2a PP  
84 = 0.97; JP = 0.74).

85 We next asked whether there was any effect from long branched dicyemids on the strength of  
86 support for inclusion of *Intoshia* within the Annelida - *Intoshia* also being a long-branched  
87 taxon. Repeating our jackknife analyses with dicyemids excluded increased the support for  
88 *Intoshia* as an annelid from JP = 0.74 to JP = 0.86 (Fig. 3b) showing that when the expected  
89 LBA between *Dicyema* spp and *Intoshia* is prevented, there is stronger support for including  
90 the orthonectid in Annelida. An equivalent analysis omitting *Intoshia* did not help to resolve  
91 the position of dicyemids (figure S3).

92 To test further the support for *Intoshia* being a member of Annelida, we reasoned that an  
93 analysis restricted to genes showing the strongest signal supporting monophyletic Annelida  
94 should give stronger support to *Intoshia* within Annelida but only if it is indeed a member of  
95 the clade; if not, support should decrease when using this subset of genes. We first removed all  
96 mesozoan sequences from each individual gene alignment and reconstructed a tree for each  
97 gene. We ranked these gene trees according to the proportion of all annelids present in a given

98 gene data set that were observed united in a clade. We concatenated the genes (now including  
99 mesozoans) from strongest supporters of monophyletic Annelida to weakest. We repeated our  
100 jackknife analyses using the best quarter of genes. An analysis of the genes that most strongly  
101 support monophyletic Annelida results in an increase support for inclusion of *Intoshia* within  
102 Annelida from JP = 0.74 to JP = 0.94 (Fig. 3c).

103 Our results suggest that recent findings of a close relationship between *Intoshia* and *Dicyema*  
104 and the linking of both these taxa to rapidly evolving gastrotrichs and platyhelminths [7,8] is  
105 due to long branch attraction. To test this prediction we exaggerated the expected effects of  
106 LBA on our own data set by using less well fitting models. We first conducted cross validation  
107 comparing the site heterogeneous CAT+G4 model we have used to the site homogenous  
108 LG+G4 and show that LG+G4 is a significantly less good fit to our data (CAT+G4 is better  
109 than LG+G4:  $\Delta \ln L = 9787 \pm 249.265$ ). We used the less well fitting LG+G4 model to  
110 reanalyse the jackknife replicates of a data set including our four most complete annelids. We  
111 observed a topology clearly influenced by LBA in which long branched taxa including  
112 flatworms, annelids, rotifers and nematodes were grouped. We also observed within this 'LBA  
113 assemblage' the two longest branched clades, dicyemids and the orthonectid as each other's  
114 closest relatives. As a further test we reanalysed the published data set [8] which had linked  
115 orthonectid and dicyemid with platyhelminths and gastrotrichs. When we removed the most  
116 obvious source of LBA - the long branched dicyemid - we found that the orthonectid *Intoshia*  
117 was, as expected, found not with platyhelminths or gastrotrichs but with the two annelids  
118 present in this data set, again providing evidence of the effects of long branch attraction (Fig  
119 4).

## 120 **Discussion**

121 We have analysed the first, almost complete mitochondrial genome sequence of an orthonectid  
122 mesozoan and added to the known mitochondrial genes of Dicyemida to provide two powerful

123 rare genomic changes. Our analyses of mitochondrial NAD5 gene sequences show  
124 unequivocally that both Dicyemida and Orthonectida are members of the protostomes and the  
125 absence of rhabditophoran flatworm mitochondrial genetic code changes rejects existing ideas  
126 that either group might be derived from parasitic flatworms. Both groups show unusually high  
127 rates of evolution and this required steps to test for and avoid the possible effects of long branch  
128 attraction, not least between the orthonectids and dicyemids.

129 Our mitochondrial data set and our large, taxonomically broad set of nuclear genes with a low  
130 percentage of missing data, analysed with well fitting, site heterogeneous models of sequence  
131 evolution, do not support the close relationship between orthonectids and dicyemids.  
132 Orthonectids are annelids and not members of the Mesozoa and the phylum Mesozoa *sensu*  
133 *lato* is an unnatural polyphyletic assemblage. We were unable to place the dicyemids more  
134 precisely and they may be considered a phylum in their own right. Experiments manipulating  
135 the expected effects of LBA strongly suggest previous phylogenies were affected by this  
136 important source of systematic error. Finding the orthonectids and dicyemids not closely  
137 associated demonstrates a remarkable instance of convergent evolution in two unrelated,  
138 miniaturised parasites.

139 The finding that the orthonectid *Intoshia* is a member of the Annelida shows that it has evolved  
140 its extraordinary simplicity by drastic simplification from a much more complex annelid  
141 common ancestor. Our phylogenetic analyses could not more precisely place *Intoshia* within  
142 the annelids, however, a short stretch of mitochondrial genes (*nad1*, *nad6*, *cob*) that are found  
143 in the same order as in the lophotrochozoan ancestor and in the early branching annelid *Owenia*  
144 *fusiformis* but not in the pleistoannelid ground plan argues for a position outside of the  
145 Pleistoannelida [17] (Fig 2d). Possible evidence of an ancestral segmented body plan is still  
146 apparent in the series of circular muscles regularly spaced along the antero-posterior axis of  
147 *Intoshia* (Fig 1b), along with similarly repeated bands of cilia (Fig 1 and ref [18]). Further

148 analysis of the genome, embryology and morphology of *Intoshia* or other orthonectids are  
149 predicted to show additional clues as to their cryptic annelidan ancestry.

## 150 **Methods**

151

### 152 Genome and transcriptome assemblies.

153 We downloaded genomic (*Intoshia linei*: SRR4418796, SRR4418797) and transcriptomic  
154 (*Dicyema sp.*: SRR827581; *Dicyema japonicum*: DRR057371) data from the NCBI Short  
155 Read Archives and DDBJ, and used Trimmomatic [19] to clean residual adapter sequences  
156 from the sequencing reads and to remove low quality bases. We used the clc assembly cell  
157 (clcBIO/Qiagen; v.5.0) to re-assemble the *I. linei* genome and the Trinity pipeline [20]  
158 (v.2.3.2) to assemble the *Dicyema sp.* and *D. japonicum* transcriptomes using default  
159 settings. We additionally assembled transcriptomes for *Phascolopsis gouldii*,  
160 *Spiochaetopterus sp.*, *Arenicola marina*, *Sabella pavonina*, *Magelona pitelkai*,  
161 *Pharyngocirrus tridentiger* and *Bonellia viridis* from SRA datasets (SRR1654498,  
162 SRR1224605, SRR2005653, SRR2005708, SRR2015609, SRR2016714, SRR2017645)  
163 using the same approach.

164

### 165 Identifying mitochondrial genome fragments.

166 Using mitochondrial gene protein coding sequences from flatworms as queries [21] we used  
167 tblastn [22] and blastp to search for *Dicyema sp.* and *D. japonicum* mitochondrial fragments  
168 in the Trinity RNA-Seq assemblies, and screened the *I. linei* genome re-assembly in a similar  
169 way. Positively identified ORFs were then blasted against NCBI nr to detect possible  
170 contamination from host species in the RNA-Seq data. For each *Dicyema sp.* gene-bearing  
171 contig, we also found additional contigs which had strongly matching blast hits to *Octopus* or

172 other cephalopods (or in some cases to the gastropod mollusc *Aplysia*) and we discarded  
173 these as likely contaminations.

174

#### 175 Annotating mitochondrial genomes.

176 Using blast we identified a 14.2kb mitochondrial contig in the assembled *I. linei* genome,  
177 which we annotated using MITOS [23]. The location of protein-coding genes were manually  
178 verified from MITOS prediction, and inferred to start from the first in-frame start codon  
179 (ATN, GTG, TTG, or GTT). The C-terminal of the protein-coding genes was inferred to be  
180 the first in-frame stop codon (TAA, TAG or TGA). We aligned the *Intoshia* and *Dicyema*  
181 NAD5 genes with those from 5 protostomes, 4 deuterostomes, and 2 non-bilaterian species in  
182 the Geneious software to visualise Protostome specific signatures in the sequence.

183

#### 184 Mitochondrial Phylogenetics

185 We grouped the mesozoan mitochondrial protein coding genes with their orthologs from 65  
186 other species selected to cover the diversity of the Metazoa including diploblasts,  
187 deuterostomes and ecdysozoans but with an emphasis on the diversity of Lophotrochozoa.  
188 We aligned each set of orthologs using Muscle [24] v3.8.31 using default parameters and  
189 trimmed these alignments to exclude unreliably aligned positions using TrimAl [25] (version  
190 1.2 rev 59 using default settings). Finally, we concatenated the trimmed alignments of all  
191 genes into a supermatrix of 2969 positions. We inferred a phylogeny with phylobayes (4.1b)  
192 under the CAT+G4 model. We ran 10 independent chains for 10,000 cycles each. We  
193 summarised all ten chains (bpcomp) discarding the first 8,000 trees from each as burnin. We  
194 reconstructed additional mitochondrial phylogenies omitting (i) the long branching flatworm  
195 species, (ii) all long branch taxa and also *Intoshia*, and (iii) long branch taxa and the *Dicyema*



196 species. Here and elsewhere we visualised and edited phylogenetic trees with FigTree  
197 (v1.4.3; <http://tree.bio.ed.ac.uk/software/figtree/>).

198

#### 199 Nuclear gene orthology determination

200 We chose to add the mesozoan data to sets of orthologous genes that were previously  
201 successfully used to infer lophotrochozoan phylogeny [26,27]. We first used Orthofinder [28]  
202 (v.1.0.8) to calculate orthologous relationships between the genes predicted for *I. linei* in the  
203 recent genome paper [8] and our *Dicyema* sp. gene predictions. To ensure robustness of the  
204 analysis we included several outgroup species (Supplementary table 2) In particular, as we  
205 were concerned about potential contamination by the hosts of the parasitic *Dicyema* we  
206 included the *Octopus bimaculoides* proteome. Since the published phylogenomic studies  
207 included few annelid species we added our own Trinity assemblies of several additional  
208 species (see above). We then extracted all orthologous groups containing the *Octopus* and the  
209 two mesozoan taxa from the Orthofinder output and inserted these sequences into the original  
210 alignments. This resulted in 590 orthologous groups. With the aid of OMA [29] and custom  
211 Perl scripts we filtered these groups to contain single copy orthologs of all species. We re-  
212 aligned each set of orthologs using clustal-omega [30]; we removed unreliably aligned  
213 positions from each alignment using TrimAl; finally we constructed individual gene trees  
214 from these trimmed alignments using phyml [31] (v20160207). Using Python code and the  
215 ETE3 toolkit we checked each tree for instances where sequences from *Octopus* and *Dicyema*  
216 sp. were each other's closest relatives (suggesting the sequence is an *Octopus* contaminant)  
217 and removed the 5 alignments where the trees had this topology from our set. We  
218 concatenated all single trimmed alignments of 45 taxa into a supermatrix of 227,646

219 positions. We used a custom script to eliminate all positions in the alignment with less than  
220 50% occupancy.

221

## 222 Nuclear Gene Phylogenomic analyses

223 Using the mpi version of phylobayes (in v.1.7) run over four independent chains for 5000  
224 cycles and discarding the first 4500 trees as burnin we reconstructed a phylogeny using this  
225 alignment under both the CAT+G4 model of molecular evolution. To provide a conservative  
226 measure of clade support and to test different data samples in a reasonable time we also  
227 reconstructed trees using 50 jackknife sub-samples of 30,000 positions each from the  
228 supermatrix. We used phylobayes 4.1c with the aid of the gnu-parallel command line tool  
229 [32] and the UCL HPC cluster. We used the CAT+G4 model, and also compared results  
230 from LG+G4. We ran phylobayes for 2000 cycles per jackknife sample which consistently  
231 resulted in a plateauing of the likelihood score. We summarised all 50 of these phylobayes  
232 analyses per model (using bpcomp) discarding the first 1800 sampled trees per jackknife as  
233 burnin. We also tested the effect of different species compositions in our dataset by  
234 performing phylobayes jackknife sampling with different subsets of taxa.

235

## 236 Cross validation

237 We compared the fit of CAT+G4 and LG+G4 models to our data using cross validation as  
238 described in the phylobayes user manual. We ran 10 replicates and for each replicate we used  
239 a randomly selected 30,000 positions of the data as a training set and 10,000 randomly

240 selected positions as the test set. Log likelihood scores were averaged over the ten replicates  
241 using the sumcv command.

242

### 243 Ranking genes according to support for monophyletic Annelida.

244 We first removed all *Intoshia* and *Dicyema* sequences from each individual gene alignment.  
245 For each individual gene, we reconstructed a tree from the aligned protein coding sequences  
246 using Ninja [33]. Each tree was parsed using a custom script to find the proportion of  
247 annelids in the data set present in the largest clade of annelids found. The tree was given a  
248 score which was calculated as the number of annelids in the largest clade/total number of  
249 annelids on the tree. Trees with larger monophyletic annelid clades scored highest. The  
250 genes were then concatenated in order of their score. We took the first 25% of positions from  
251 this concatenation (those genes with the strongest signal supporting monophyletic annelids)  
252 and analysed jackknife replicates as before.

253

### 254 **Acknowledgements:**

255

256 The authors are grateful to Prof George Slyusarev (Saint Petersburg State University, Russia)  
257 for providing images of *Intoshia linei* and for sharing an unpublished book chapter and to Dr  
258 Tsai-Ming Lu and colleagues (Okinawa Institute of Science and Technology, Japan) for  
259 sharing data. The authors also thank Fraser Simpson (UCL) for help in editing Figure 1b. The  
260 research was funded by ERC grant (ERC-2012-AdG 322790) to MJT. Alignments have been  
261 deposited with Zenodo (XXX) and phylogenetic trees are available through treebase.org  
262 (XXX).

263

264

265 **Author Contributions**

266 Conceived the study: MJT. Planned the study: MJT and PHS. Assembled the data sets: PHS.

267 Analysed the data: PHS and MJT. Drafted the manuscript: MJT and PHS. Analysed

268 mitochondrial data: HER.

269

270 **Declaration of Interests:**

271 The authors declare no competing financial interests.

272

273

274 **Figure Legends**

275

276 **Fig. 1: The mesozoans *Intoshia variabili* and *Dicyema typus***

277

278 A. Differential Interference contrast micrograph of an *Intoshia variabili* female showing  
279 repeated bands of ciliated cells. Picture G. Slyusarev (St Petersburg State University,  
280 Russia).

281 B. Confocal image of a phalloidin stained female specimen of *Intoshia linei* reveals repeated  
282 set of circular muscles. Picture G. Slyusarev (St Petersburg State Univ.).

283 C. Rhombogen stage of a dicyemid (*Dicyema typus* from the Octopus) adapted from Hyman  
284 L.H. The Invertebrates: Protozoa through Ctenophora McGraw-Hill, New York 1940(19).  
285 Anterior to right in all images.

286

287 **Fig. 2: Analyses of the phylogenetic positions of *Dicyema* and *Intoshia* based on  
288 mitochondrial gene sequences.**

289

290 A. Alignment of the mitochondrial NAD5 gene from selected protostomes, deuterostomes,  
291 and outgroups, highlighting derived substitutions and amino acid deletions shared by the  
292 orthonectids, dicyemids, and other protostomes.

293 B. A mitochondrial bayesian phylogeny based on 2969 positions places orthonectids and  
294 dicyemids inside Lophotrochozoa, but the unlikely assemblage of *Intoshia linei* and  
295 flatworms with annelids suggest this is affected by systematic error.

296 C. Mitochondrial bayesian phylogeny omitting the long branching taxa including *Dicyema*  
297 gives some support for a position of *Intoshia* within Annelida. D. Order of the *Intoshia* nad1,  
298 nad6, and cob mitochondrial genes in comparison to the early branching annelid *Owenia*  
299 *fusiformis*, the pleistoannelid ground plan and the lophotrochozoan ground plan (see ref [17]).

300

301 **Fig. 3: Analyses of the phylogenetic positions of *Dicyema* and *Intoshia* based on nuclear**  
302 **gene sequences.**

303 A. A bayesian phylogeny reconstructed from 190,027 aligned amino acid positions analysed  
304 under the CAT+G4 model. Support values are from bayesian posterior probabilities (PP) and  
305 from 50 jackknifed sub-samples of 30,000 residues (JP support values in brackets). Both  
306 analyses reveal Mesozoa to be polyphyletic and place *Intoshia linei* in Annelida (see Supp  
307 Fig 4a for support values).

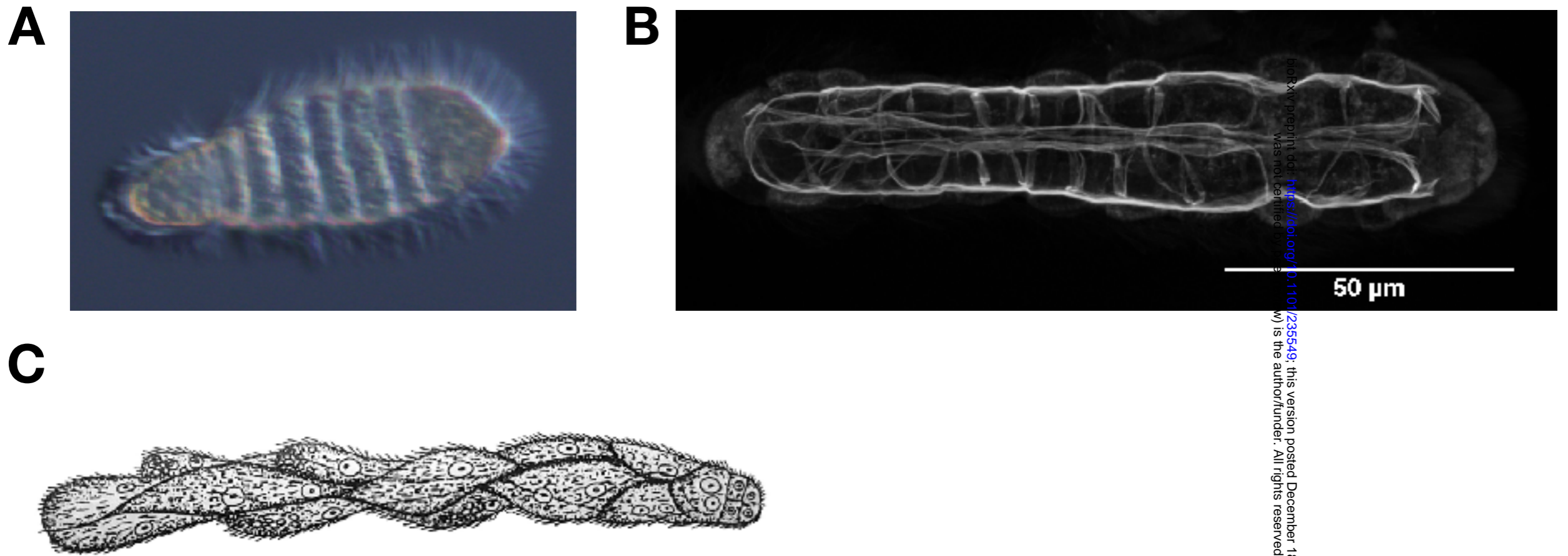
308 B. A repeat of the jackknife analysis omitting the long-branching *Dicyema* species eliminates  
309 the potential for LBA between *Intoshia* and *Dicyema*. This leads to an increase in the support  
310 for a position of *Intoshia* within Annelida from JP 0.74 to JP 0.86 JP. (Only lophotrochozoan  
311 part of the tree shown, see Supplementary Fig 4c for full tree).

312 C. Bayesian jackknife using CAT+G4 model using the best quarter of genes supporting  
313 monophyletic annelids leads to increased support for *Intoshia* within Annelida to JP 0.94  
314 even with the inclusion of the *Dicyema* species. (Only lophotrochozoan part of the tree  
315 shown, see Supplementary Fig 4d for full tree).

316

317 **Fig. 4: Reanalysis of a published data set addressing potential LBA between mesozoans**  
318 **supports annelid affinity for *Intoshia*.**

319 Repeating the analyses on a previously published data set [8] excluding the long branching  
320 *Dicyema* leads to *Intoshia* being placed with the annelids, showing the likely effect of LBA  
321 on the original analysis. Support values are bayesian posterior probabilities (PP).



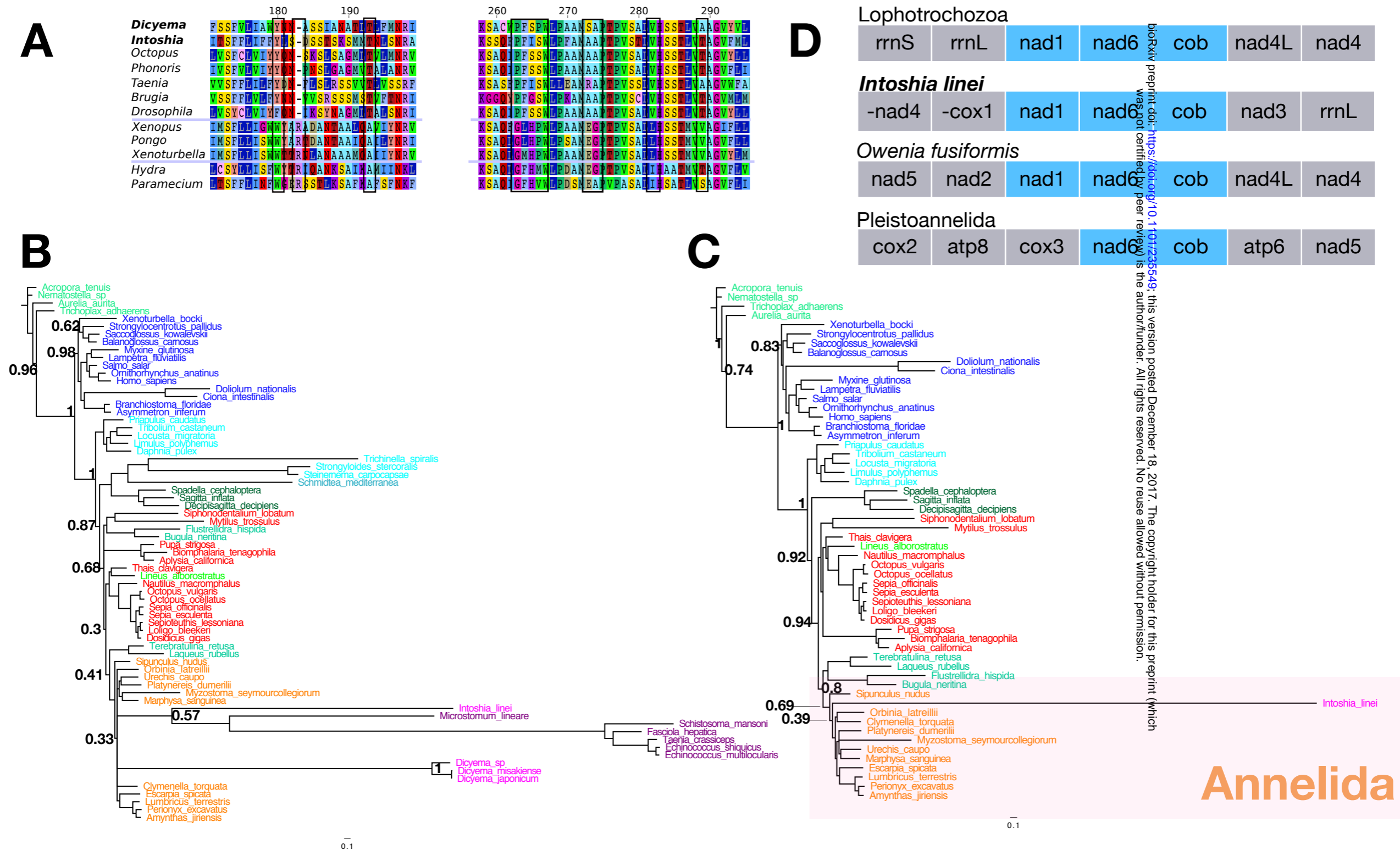
**Fig. 1: The mesozoans *Intoshia variabili* and *Dicyema typus***

A. Differential Interference contrast micrograph of an *Intoshia variabili* female showing repeated bands of ciliated cells. Picture G. Slyusarev (St Petersburg State University, Russia).

B. Confocal image of a phalloidin stained female specimen of *Intoshia linei* reveals repeated set of circular muscles. Picture G. Slyusarev (St Petersburg State Univ.).

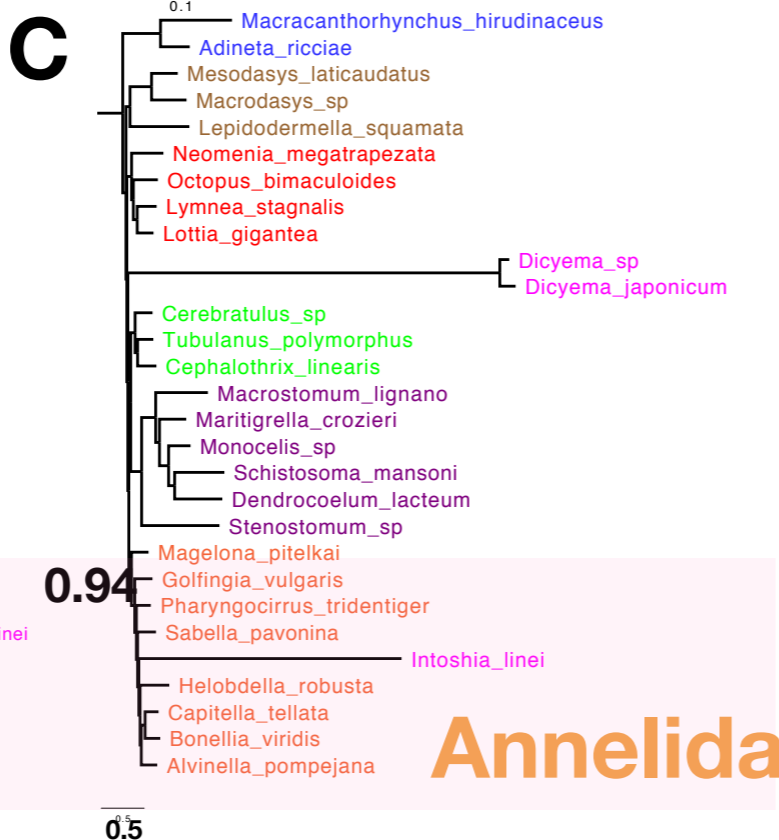
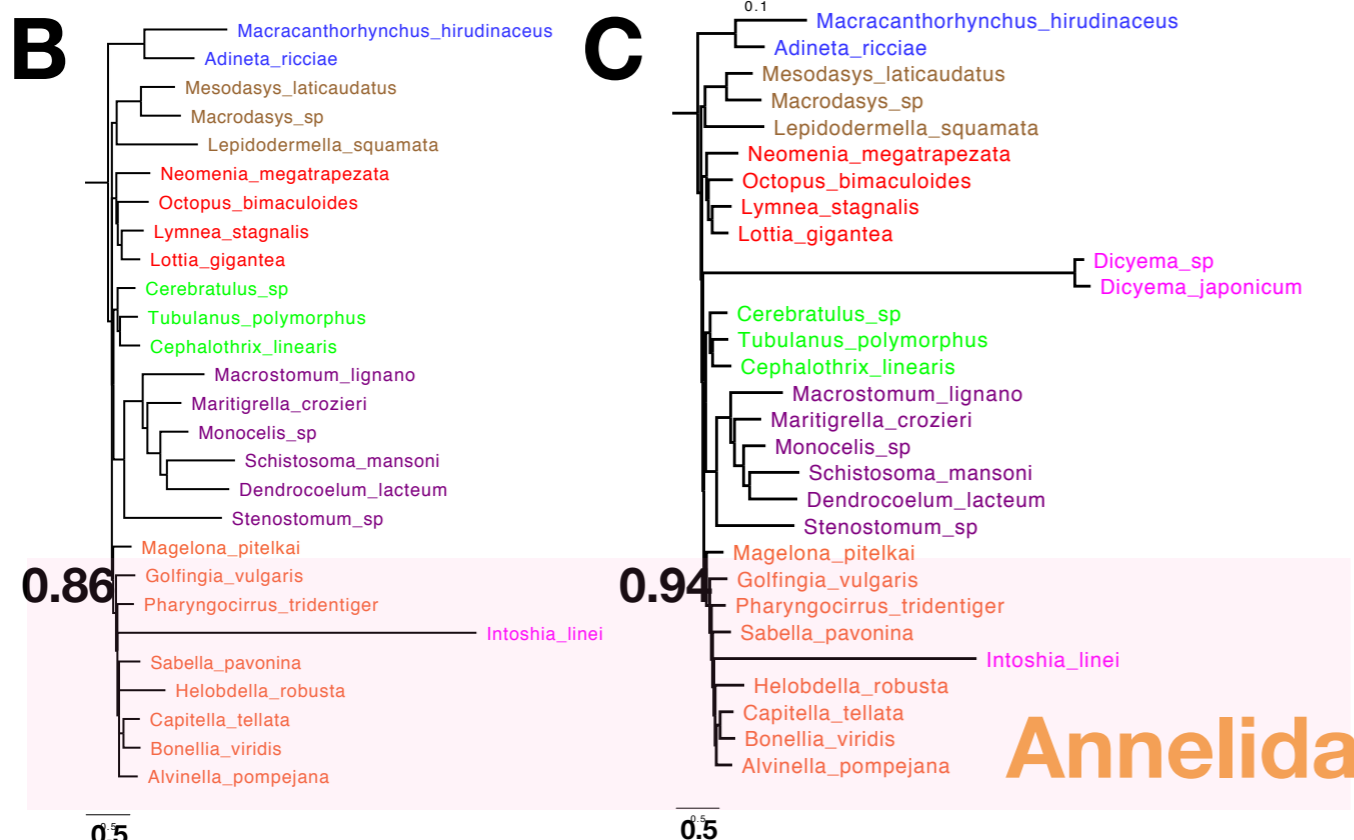
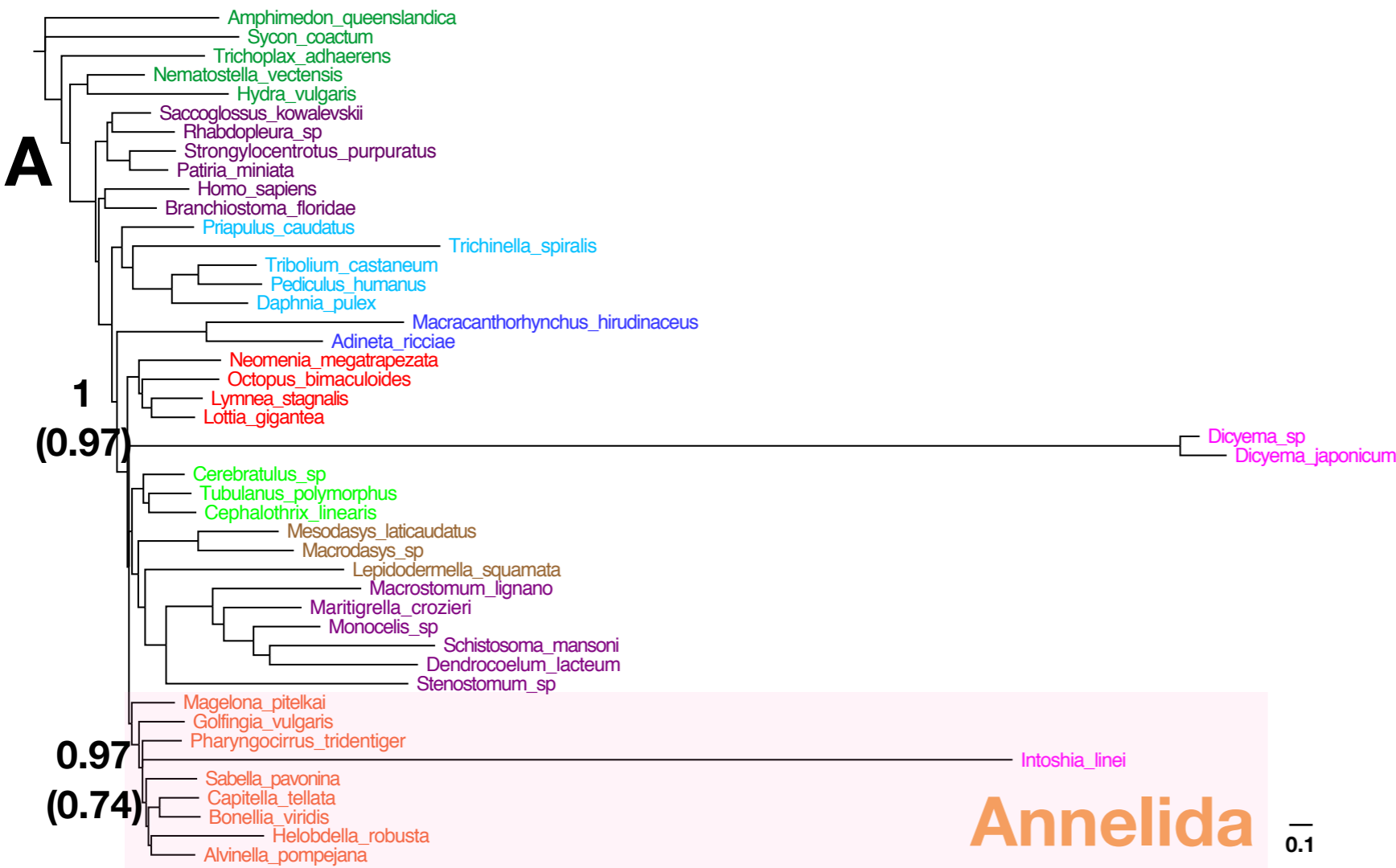
C. Rhombogen stage of a dicyemid (*Dicyema typus* from the Octopus) adapted from Hyman L.H. The Invertebrates: Protozoa through Ctenophora McGraw-Hill, New York 1940(19).

Anterior to right in all images.



bioRxiv preprint doi: <https://doi.org/10.1101/235549>; this version posted December 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Annelida**



**Fig. 3: Analyses of the phylogenetic positions of *Dicyema* and *Intoshia* based on nuclear gene sequences.**

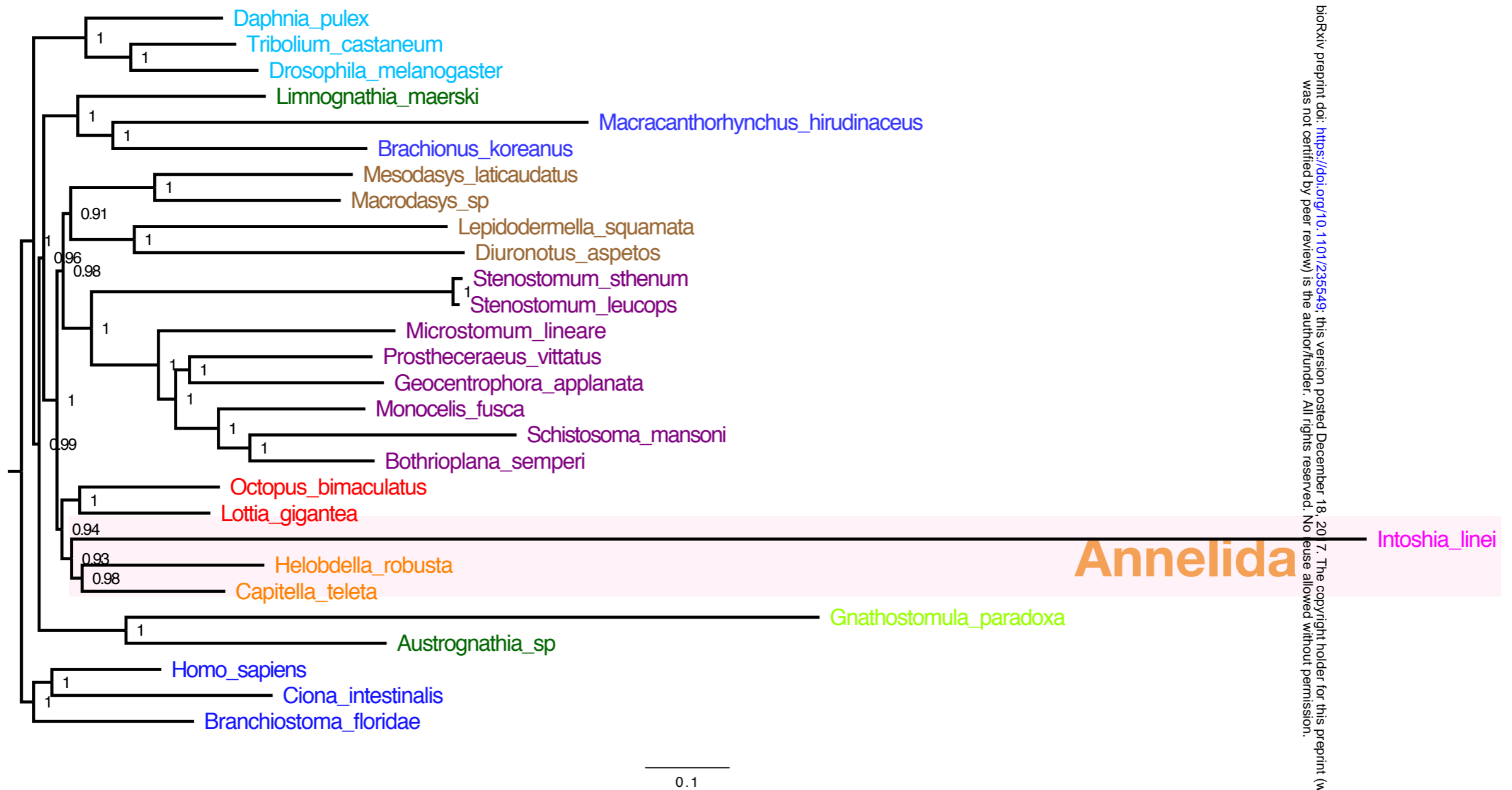
A. A bayesian phylogenetic tree reconstructed from 190,027 aligned amino acid positions analysed under the CAT+G4 model. Support values are from bayesian posterior probabilities (PP) and from 50 jackknifed sub-samples of 30,000 residues (JP support values in brackets). Both analyses reveal Mesozoa to be polyphyletic and place *Intoshia linei* in Annelida (see Supp Fig 4a for support values).

B. A repeat of the jackknife analysis omitting the long-branching *Dicyema* species eliminates the potential for LBA between *Intoshia* and *Dicyema*. This leads to an increase in the support for a position of *Intoshia* within Annelida from JP 0.74 to JP 0.86. (Only lophotrochozoan part of the tree shown, see Supplementary Fig 4c for full tree).

C. Bayesian jackknife using CAT+G4 model using the best quarter of genes supporting monophyletic annelids leads to increased support for *Intoshia* within Annelida to JP 0.94 even with the inclusion of the *Dicyema* species. (Only lophotrochozoan part of the tree shown, see Supplementary Fig 4d for full tree).

bioRxiv preprint doi: <https://doi.org/10.1101/2019.12.20.372677>; this version posted December 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.





bioRxiv preprint doi: <https://doi.org/10.1101/235549>; this version posted December 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Fig. 4: Reanalysis of a published data set addressing potential LBA between mesozoans supports annelid affinity for *Intoshia*.**

Repeating the analyses on a previously published data set [7] excluding the long branching *Dicyema* leads to *Intoshia* being placed with the annelids, showing the likely effect of LBA on the original analysis. Support values are bayesian posterior probabilities (PP).

322 **Supplementary Tables**

323

324 **Supplementary Table 1**

325 Predicted correspondence of nucleotide triplets to amino acids in *Intoshia* and three *Dicyema*  
326 species. For each triplet, the amino acid corresponding to the triplet in the standard  
327 invertebrate mitochondrial code is shown, the number of observations of the triplet to  
328 prediction is based on, the predicted amino acid and its score and finally the second highest  
329 scoring amino acid prediction. The triplets AAA and ATA are highlighted in green and likely  
330 errors highlighted in blue. Likely errors are mostly associated with very low numbers of  
331 observed GC rich triplets in these very AT rich mitochondrial genomes.

332

333 **Supplementary Table 2**

334 List of species used in the final phylogenetic analysis, data sources, and representation in the  
335 final alignment.

336

337

338 **Supplementary Figures:**

339

340 **Figure S1. Related to Figure 2.**

341 Phylogram and corresponding cladogram of a Bayesian analysis of our mitochondrial data set  
342 omitting the long-branching flatworm species. Phylobayes CAT+G4 model was run in 10  
343 independent runs for 10,000 cycles each on an alignment with 2969 positions and 8000 trees  
344 were discarded as burnin.

345

346 **Figure S2. Related to Figure 2.**

347 Phylogram and corresponding cladogram of a Bayesian analysis of our mitochondrial data set  
348 omitting *Intoshia linei*. Phylobayes CAT+G4 model was run in 10 independent runs for  
349 10,000 cycles each on an alignment with 2969 positions and 8000 trees were discarded as  
350 burnin.

351

352 **Figure S3. Related to Figure 3.**

353 A phylogram based on our analysis of the jackknifed dataset omitting *Intoshia linei*. Contrary  
354 to the improvement in placing *I. linei* observed when excluding the *Dicyema* species, the  
355 exclusion of *I. linei* does not lead to a better resolution of the *Dicyema* species' position. This  
356 can be seen as further evidence for the non-affiliation of orthonectids and dicyemids and the  
357 correct inference that orthonectids are part of Annelida.

358 **Figure S4. Related to Figure 3.**

359 A. Cladogram corresponding to Fig 3a showing all PP support values for the CAT+G4  
360 phylogeny based on the full alignment of 190,027 amino acid positions.

361

362 B. A cladogram including JP support values based on 50 jackknife subsamples of 30,000  
363 amino acid positions each independently analysed for 2000 cycles under the CAT+G4 model  
364 in phylobayes and summarised with the bpcomp command setting 1800 as burnin. As in the  
365 analysis of the full dataset *I. linei* is found within the annelids and phylum Mesozoa is found  
366 as an unnatural assemblage.

367

368 C. Cladogram corresponding to Fig 3b showing all support values.

369

370 D. Cladogram corresponding to Fig 3c showing all support values.

371

372 **References:**

373

374 [1] Slyusarev GS, Starunov VV. The structure of the muscular and nervous systems of  
375 the female *Intoshia linei* (Orthonectida). *Org Divers Evol* 2015;16:65–71.

376 [2] Furuya H, Hochberg FG, Tsuneki K. Cell number and cellular composition in  
377 infusoriform larvae of dicyemid mesozoans (Phylum Dicyemida). *Zool Sci*  
378 2004;21:877–89.

379 [3] Nielsen C. *Animal Evolution*. Oxford University Press; 2011.

380 [4] Dodson EO. A note on the systematic position of the Mesozoa. *Syst Zoo* 1956;5:37.

381 [5] Suzuki TG, Ogino K, Tsuneki K, Furuya H. Phylogenetic analysis of dicyemid  
382 mesozoans (phylum Dicyemida) from innexin amino acid sequences: Dicyemids are  
383 not related to Platyhelminthes. *J Parasitol* 2010;96:614–25.

384 [6] Hanelt B, Van Schyndel D, Adema CM, Lewis LA, Loker ES. The phylogenetic  
385 position of *Rhopalura ophiocomae* (Orthonectida) based on 18S ribosomal DNA  
386 sequence analysis. *Mol Biol Evol* 1996;13:1187–91.

387 [7] Lu T-M, Kanda M, Satoh N, Furuya H. The phylogenetic position of dicyemid  
388 mesozoans offers insights into spiralian evolution. *Zool Letts* 2017;3:419.

389 [8] Mikhailov KV, Slyusarev GS, Nikitin MA, Logacheva MD, Penin AA, Aleoshin VV, et  
390 al. The genome of *Intoshia linei* affirms orthonectids as highly simplified spirilians.  
391 *Curr Biol* 2016;26:1768–74.

392 [9] Philippe H, Brinkmann H, Copley RR, Moroz LL, Nakano H, Poustka AJ, et al.  
393 Acoelomorph flatworms are deuterostomes related to *Xenoturbella*. *Nature*  
394 2011;470:255–8.

395 [10] Slyusarev GS. Fine structure and development of the cuticle of *Intoshia variabli*  
396 (Orthonectida). *Acta Zool* 2000;81:1–8.

397 [11] Telford MJ, Copley RR. Improving animal phylogenies with genomic data. *Trends*  
398 *Genet* 2011;27:186–95.

399 [12] Telford MJ. Turning Hox “signatures” into synapomorphies. *Evol Dev* 2000;2:360–4.

400 [13] Kobayashi M, Furuya H, Holland PW. Dicyemids are higher animals. *Nature*  
401 1999;401:762–2.

402 [14] Telford MJ, Herniou EA, Russell RB, Littlewood DT. Changes in mitochondrial  
403 genetic codes as phylogenetic characters: two examples from the flatworms. *P Natl*  
404 *Acad Sci Usa* 2000;97:11359–64.

405 [15] Lartillot N, Blanquart S, Lepage T. PhyloBayes 3.3 a Bayesian software for  
406 phylogenetic reconstruction and molecular dating using mixture models. 2012.

- 407 [16] Simion P, Philippe H, Baurain D, Jager M, Richter DJ, Di Franco A, et al. A large  
408 and consistent phylogenomic dataset supports sponges as the sister group to all  
409 other animals. *Curr Biol* 2017;27:958–67.
- 410 [17] Weigert A, Golombek A, Gerth M, Schwarz F, Struck TH, Bleidorn C. Evolution of  
411 mitochondrial gene order in Annelida. *Mol Phyl Evol* 2016;94:196–206.
- 412 [18] Slyusarev GS, Kristensen RM. Fine structure of the ciliated cells and ciliary rootlets  
413 of *Intoshia variabilis* (Orthonectida). *Zoomorphology* 2002;122:33–9.
- 414 [19] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina  
415 sequence data. *Bioinformatics* 2014;30:2114–20.
- 416 [20] Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De  
417 novo transcript sequence reconstruction from RNA-seq using the Trinity platform for  
418 reference generation and analysis. *Nat Protoc* 2013;8:1494–512.
- 419 [21] Robertson HE, Lapraz F, Egger B, Telford MJ, Schiffer PH. The mitochondrial  
420 genomes of the acoelomorph worms *Paratomella rubra*, *Isodiametra pulchra* and  
421 *Archaphanostoma ylvae*. *Sci Rep* 2017;7:1847.
- 422 [22] Altschul SF, Gish W, Miller W, Myers EW. Basic local alignment search tool. *Journal*  
423 *of Mol Biol* 1990;215:403–10.
- 424 [23] Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritsch G, et al. MITOS:  
425 improved de novo metazoan mitochondrial genome annotation. *Mol Phyl Evol*  
426 2013;69:313–9.
- 427 [24] Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and  
428 space complexity. *BMC Bioinformatics* 2004;5:113.
- 429 [25] Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. TrimAl: a tool for automated  
430 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*  
431 2009;25:1972–3.
- 432 [26] Egger B, Lapraz F, Tomiczek B, Müller S, Dessimoz C, Girstmair J, et al. A  
433 transcriptomic-phylogenomic analysis of the evolutionary relationships of flatworms.  
434 *Curr Biol* 2015;25:1347–53.
- 435 [27] Struck TH, Wey-Fabrizius AR, Golombek A, Hering L, Weigert A, Bleidorn C, et al.  
436 Platyzoan paraphyly based on phylogenomic data supports a noncoelomate  
437 ancestry of spiralia. *Mol Biol Evol* 2014;31:1833–49.
- 438 [28] Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome  
439 comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*  
440 2015;16:E9–13.
- 441 [29] Altenhoff AM, Škunca N, Glover N, Train C-M, Sueki A, Piližota I, et al. The OMA  
442 orthology database in 2015: function predictions, better plant support, synteny view  
443 and other improvements. *Nucleic Acids Res* 2015;43:D240–9.
- 444 [30] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable  
445 generation of high-quality protein multiple sequence alignments using Clustal  
446 Omega. *Mol Syst Biol* 2011;7:1–6.
- 447 [31] Morrison DA. Increasing the efficiency of searches for the maximum likelihood tree  
448 in a phylogenetic analysis of up to 150 nucleotide sequences. *Systematic Biol*  
449 2007;56:988–1010.
- 450 [32] Tange O. Gnu parallel-the command-line power tool. *The USENIX Magazine*; 2011.
- 451 [33] Wheeler TJ. Large-Scale Neighbor-Joining with NINJA. *Algorithms in Bioinformatics*,  
452 vol. 5724, Berlin, Heidelberg: Springer Berlin Heidelberg; 2009, pp. 375–89.

# Figure S1. Related to Figure 2.

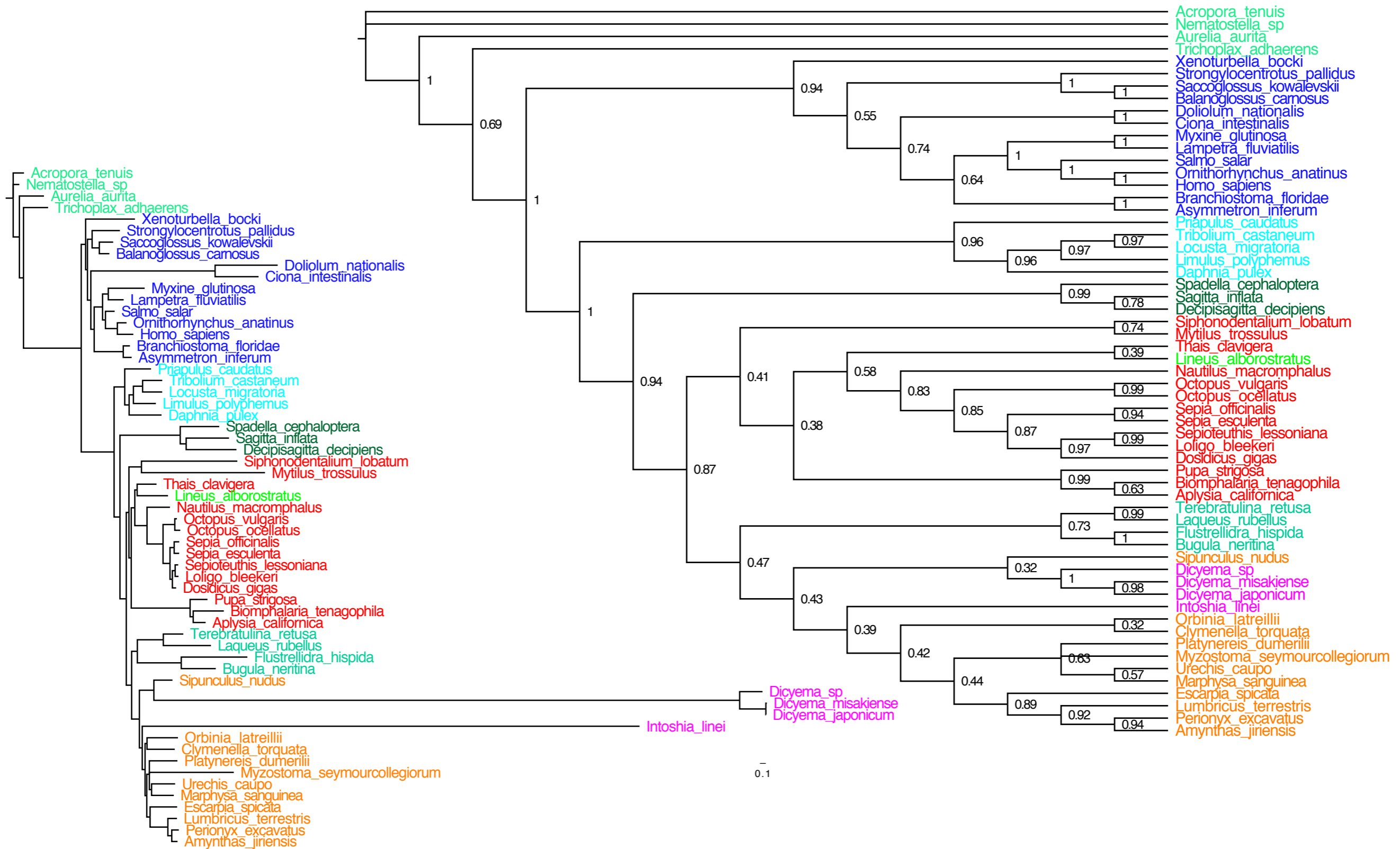


Figure S1:

Phylogram and corresponding cladogram of a Bayesian analysis of our mitochondrial data set omitting the long-branching flatworm species. Phylobayes CAT+G4 model was run in 10 independent runs for 10,000 cycles each on an alignment with 2969 positions and 8000 trees were discarded as burnin.

# Figure S2. Related to Figure 2.

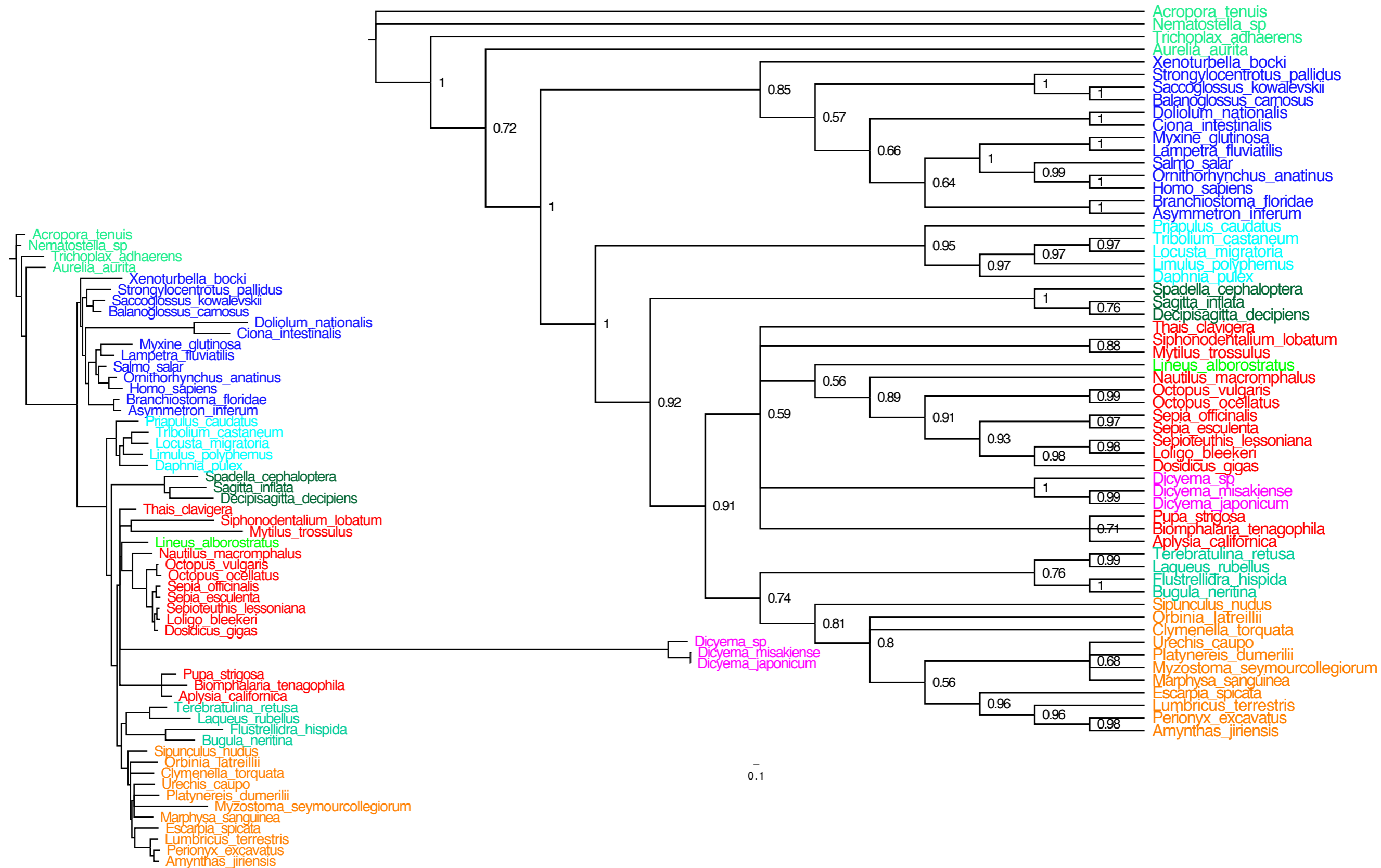


Figure S2:

0.1

Phylogram and corresponding cladogram of a Bayesian analysis of our mitochondrial data set omitting *Intoshia linei*. Phylobayes CAT+G4 model was run in 10 independent runs for 10,000 cycles each on an alignment with 2969 positions and 8000 trees were discarded as burnin.

# Figure S3. Related to Figure 3.

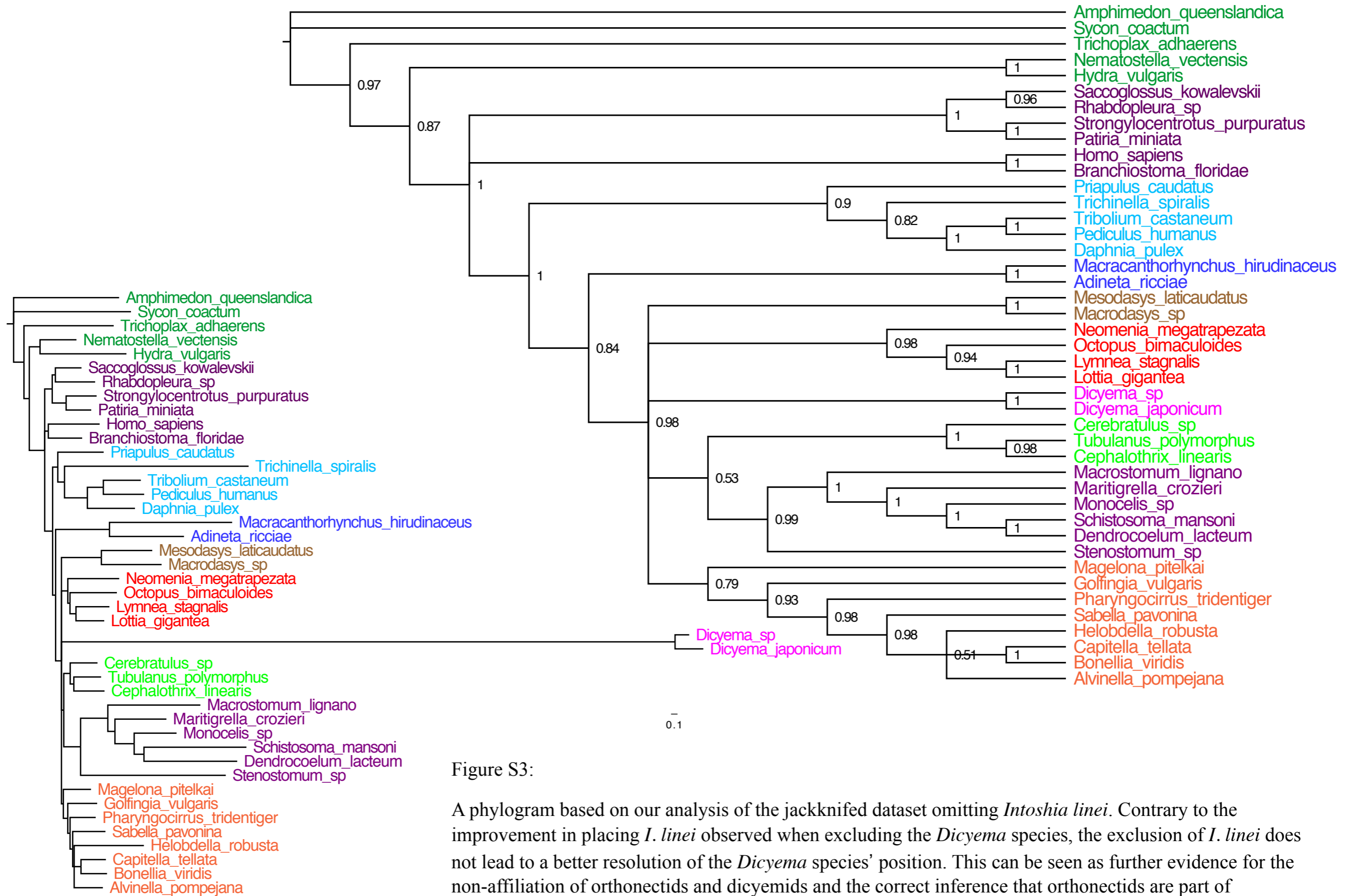


Figure S3:

A phylogram based on our analysis of the jackknifed dataset omitting *Intoshia linei*. Contrary to the improvement in placing *I. linei* observed when excluding the *Dicyema* species, the exclusion of *I. linei* does not lead to a better resolution of the *Dicyema* species' position. This can be seen as further evidence for the non-affiliation of orthonectids and dicyemids and the correct inference that orthonectids are part of Annelida.

# Figure S4. Related to Figure 3.

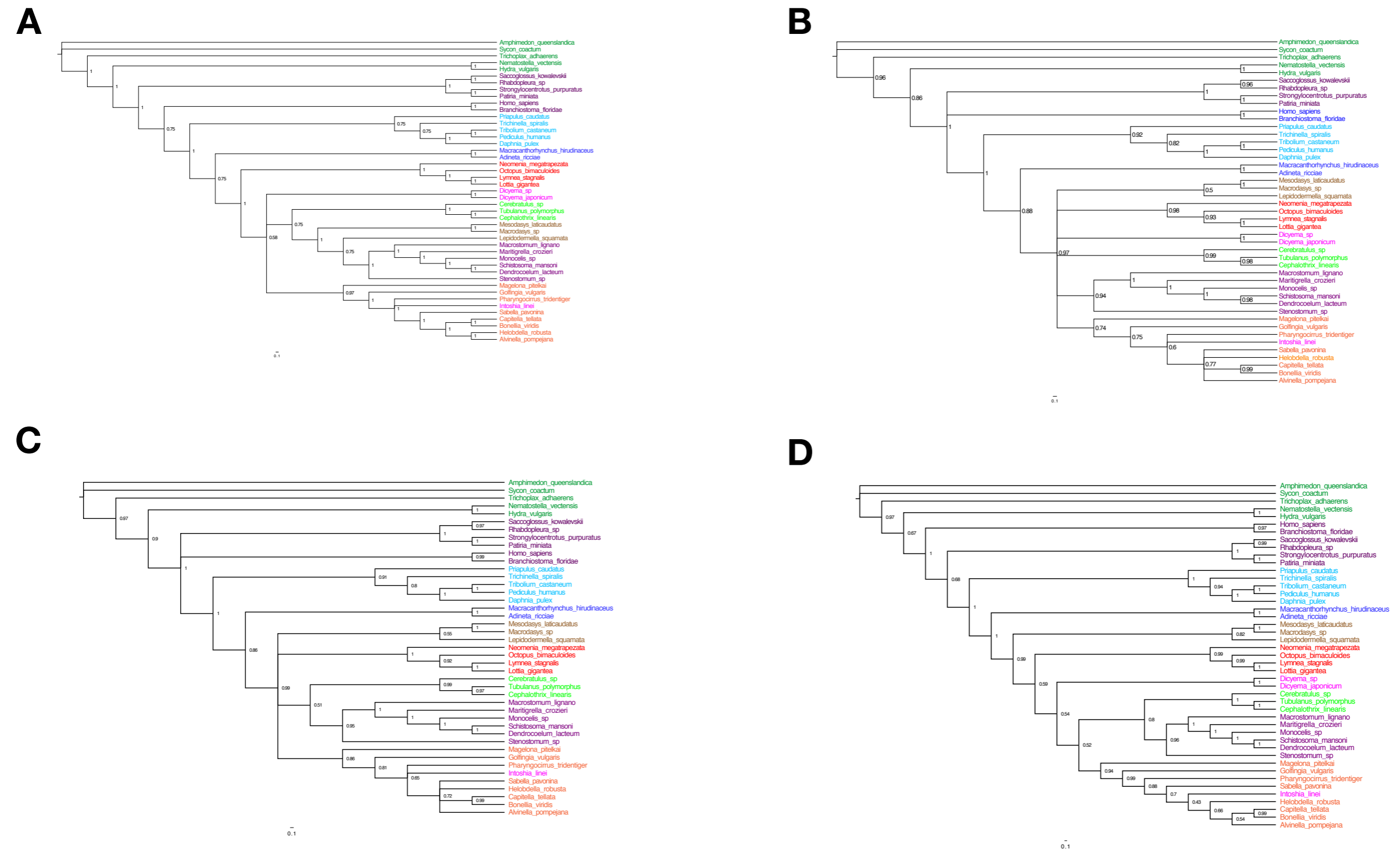


Figure S4:

- A. Cladogram corresponding to Fig 3a showing all PP support values for the CAT+G4 phylogeny based on the full alignment of 190,027 amino acid positions.
- B. A cladogram including JP support values based on 50 jackknife subsamples of 30,000 amino acid positions each independently analysed for 2000 cycles under the CAT+G4 model in phylobayes and summarised with the bpcomp command setting 1800 as burnin. As in the analysis of the full dataset I. linei is found within the annelids and phylum Mesozoa is found as an unnatural assemblage.
- C. Cladogram corresponding to Fig 3b showing all support values.
- D. Cladogram corresponding to Fig 3c showing all support values.