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

Authors: Philipp H. Schiffer[1], Helen E. Robertson[1], Maximilian J. Telford[1*]

Affiliations: [1] 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

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

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

Results

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

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

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

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

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

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

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

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

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

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

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

Discussion

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

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

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

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

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

Annotating mitochondrial genomes.

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

Mitochondrial Phylogenetics

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

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

Ranking genes according to support for monophyletic Annelida.

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

Acknowledgements:

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

Author Contributions

Conceived the study: MJT. Planned the study: MJT and PHS. Assembled the data sets: PHS. Analysed the data: PHS and MJT. Drafted the manuscript: MJT and PHS. Analysed mitochondrial data: HER.

Declaration of Interests:

The authors declare no competing financial interests.

Figure Legends

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.

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

A. Alignment of the mitochondrial NAD5 gene from selected protostomes, deuterostomes, and outgroups, highlighting derived substitutions and amino acid deletions shared by the orthonectids, dicyemids, and other protostomes.
B. A mitochondrial bayesian phylogeny based on 2969 positions places orthonectids and dicyemids inside Lophotrochozoa, but the unlikely assemblage of *Intoshia linei* and flatworms with annelids suggest this is affected by systematic error.
C. Mitochondrial bayesian phylogeny omitting the long branching taxa including *Dicyema* gives some support for a position of *Intoshia* within Annelida.D. Order of the *Intoshia* nad1, nad6, and cob mitochondrial genes in comparison to the early branching annelid *Owenia fusiformis*, the pleistoannelid ground plan and the lophotrochozoan ground plan (see ref [17]).

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

A. A bayesian phylogeny 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 JP. (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).

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 [8] 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).

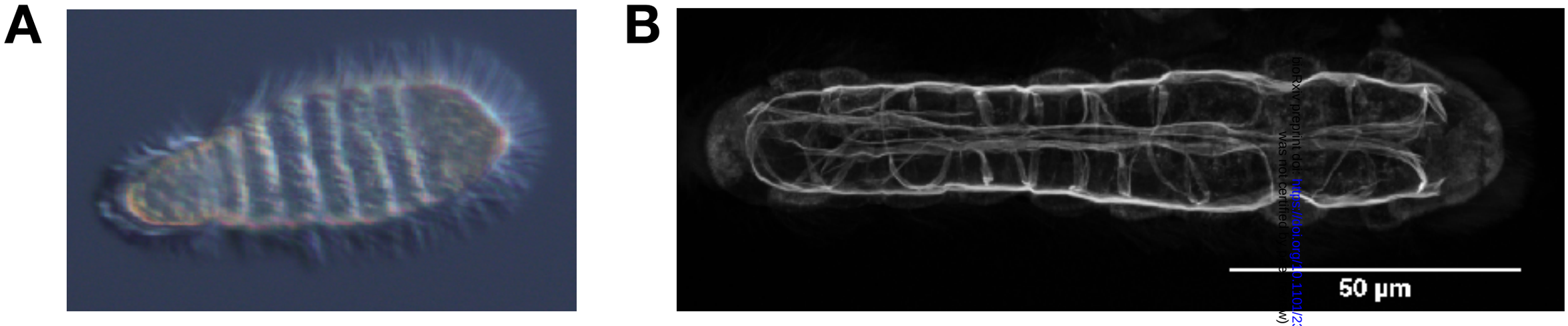


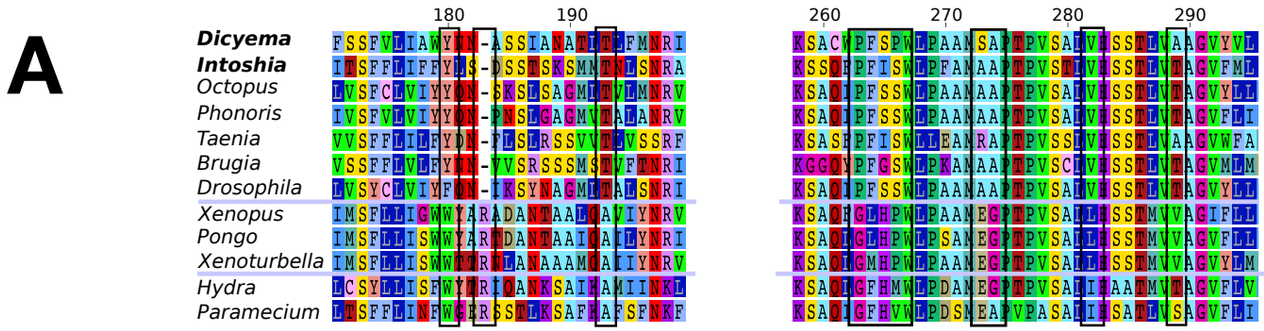
Fig. 1: The mesozoans *Intoshia variabili* and *Dicyema typus*

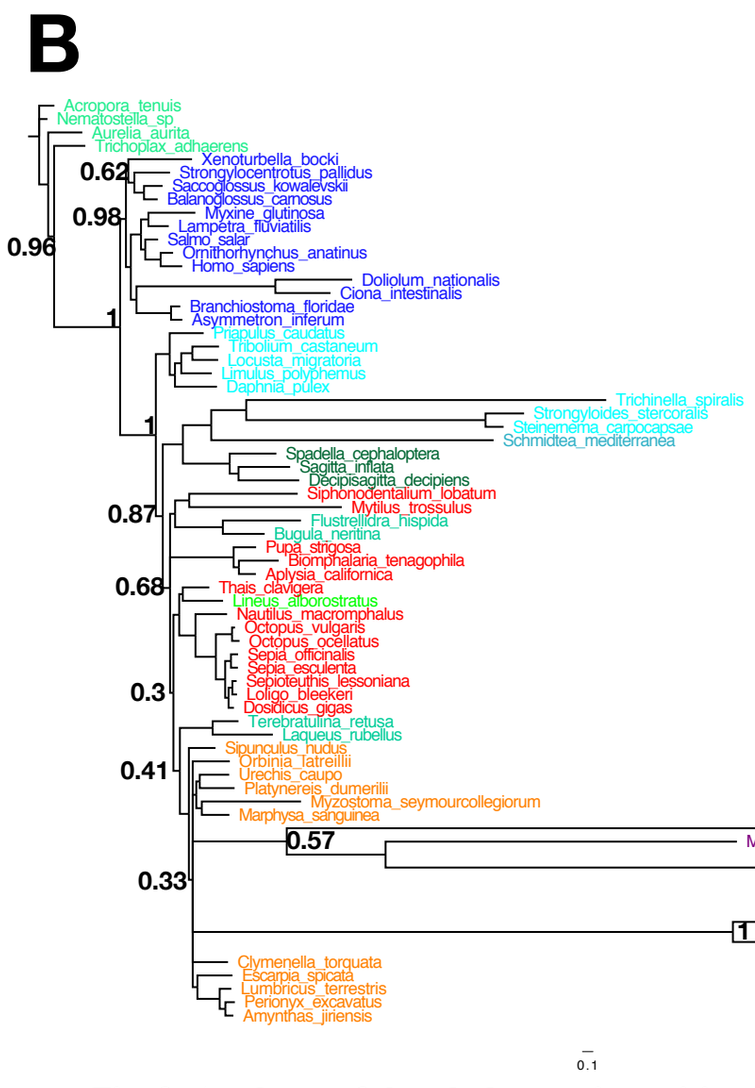
A. Differential Interference contrast micrograph of an *Intoshia variabili* female showing repeated bands of ciliated cells. Picture G. Slyusarev (St Petersburg State University, Russia).

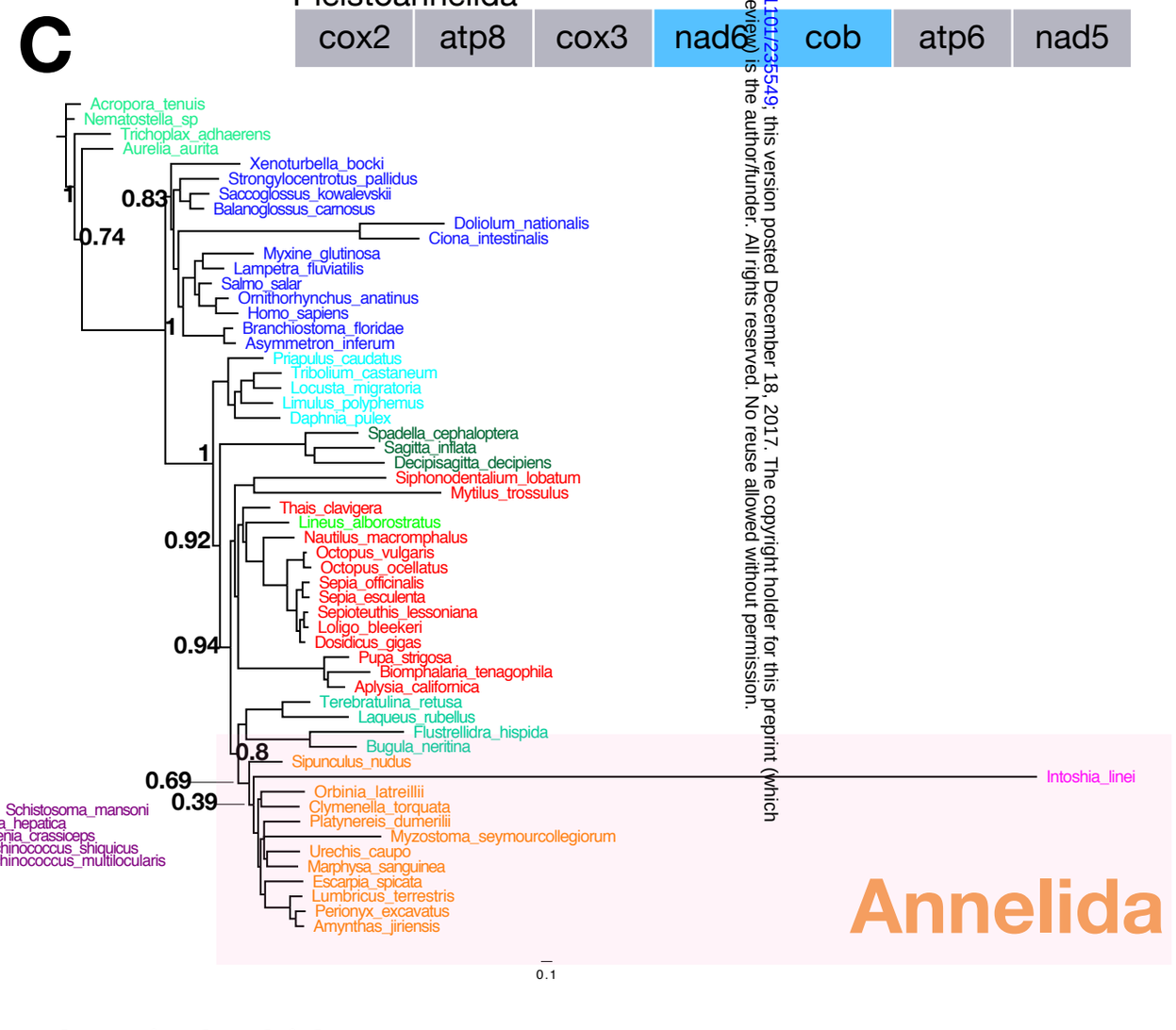
B. Confocal image of a phalloidin stained female specimen of Intoshia linei reveals repeated set of circular muscles. Picture G. Slyusarev (St Petersburg State Univ.).

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.







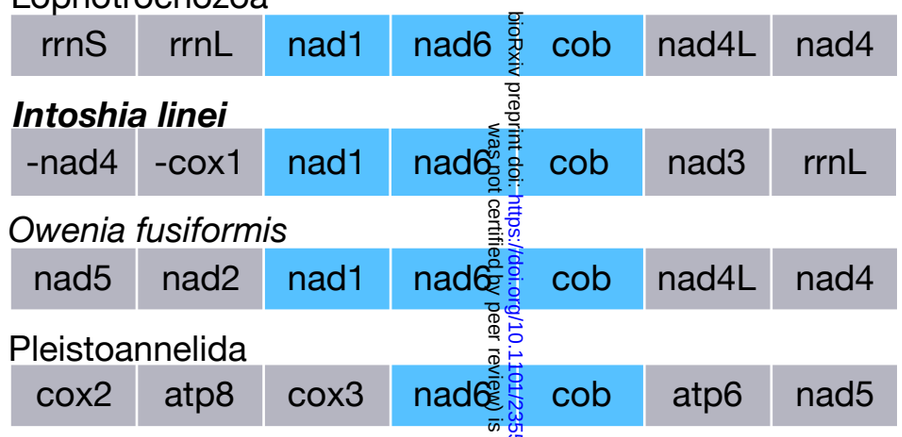


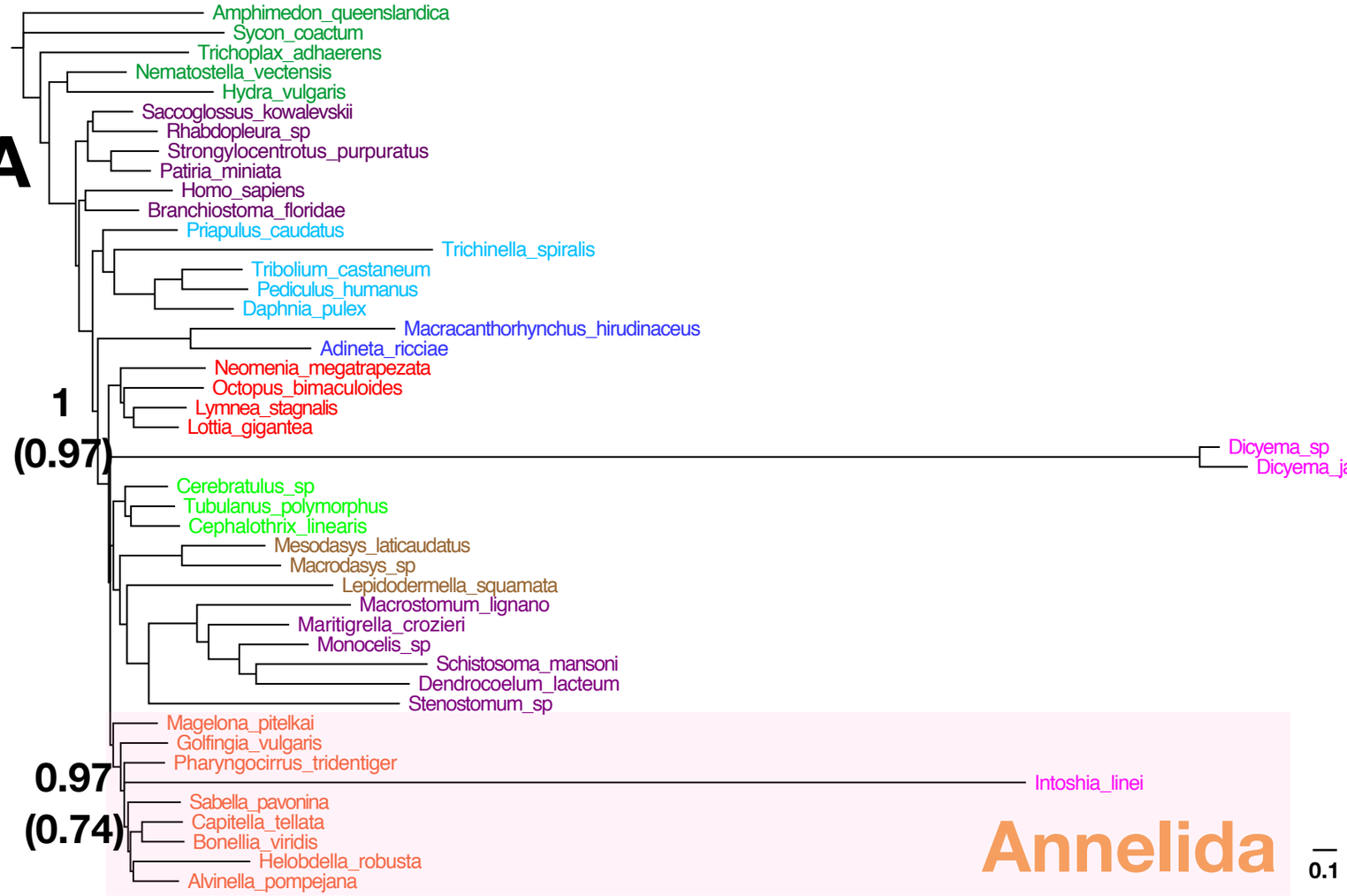
Fig. 2: Analyses of the phylogenetic positions of *Dicyema* and *Intoshia* based on mitochondrial gene sequences.

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

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

C. Mitochondrial bayesian phylogeny omitting the long branching taxa including *Dicyema* gives some support for a position of *Intoshia* within Annelida.

D. Order of the *Intoshia* nad1, nad6, and cob mitochondrial genes in comparison to the early branching annelid *Owenia fusiformis*, the pleistoannelid ground plan and the lophotrochozoan ground plan (see ref [17]).



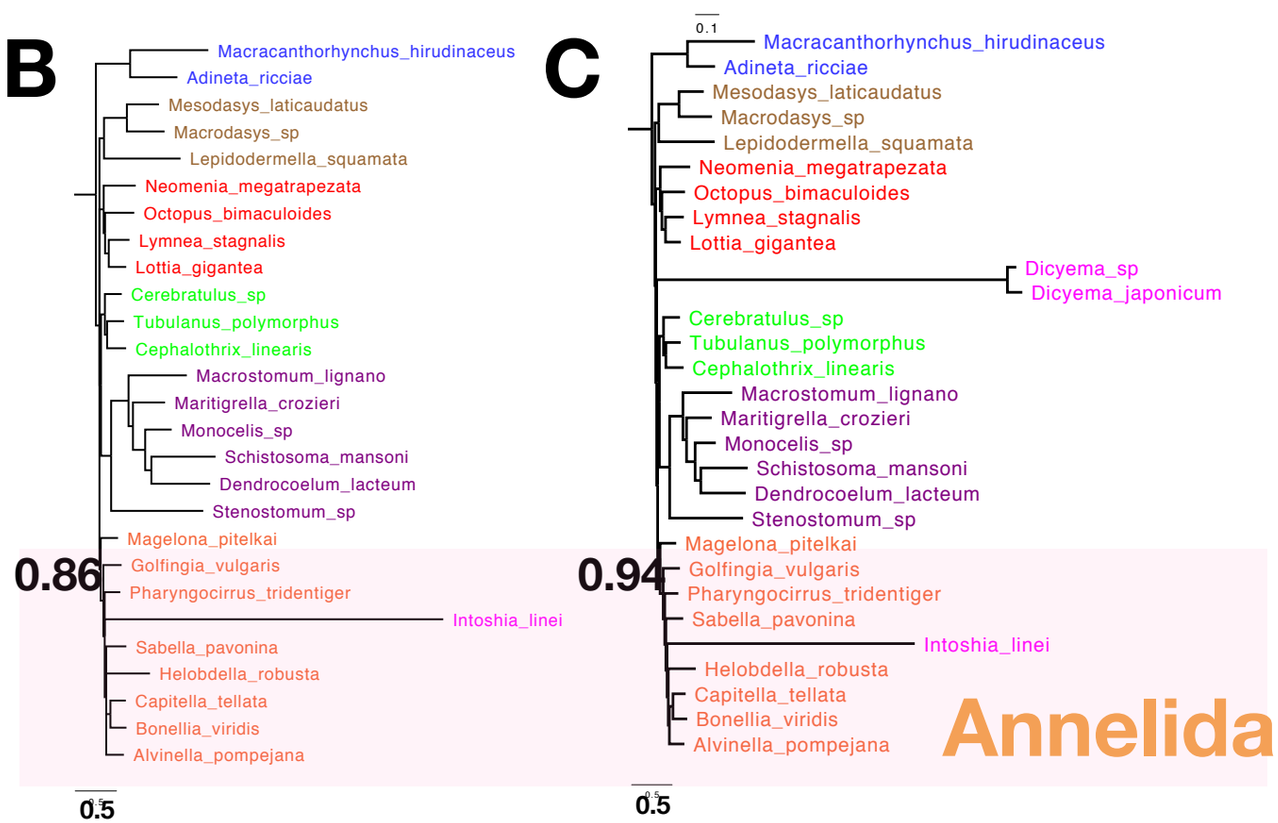


Fig. 3: Analyses of the phylogenetic positions of *Dicyema* and *Intoshia* based on nuclear gene sequences.

A. A bayesian phylogeny 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 JP. (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).

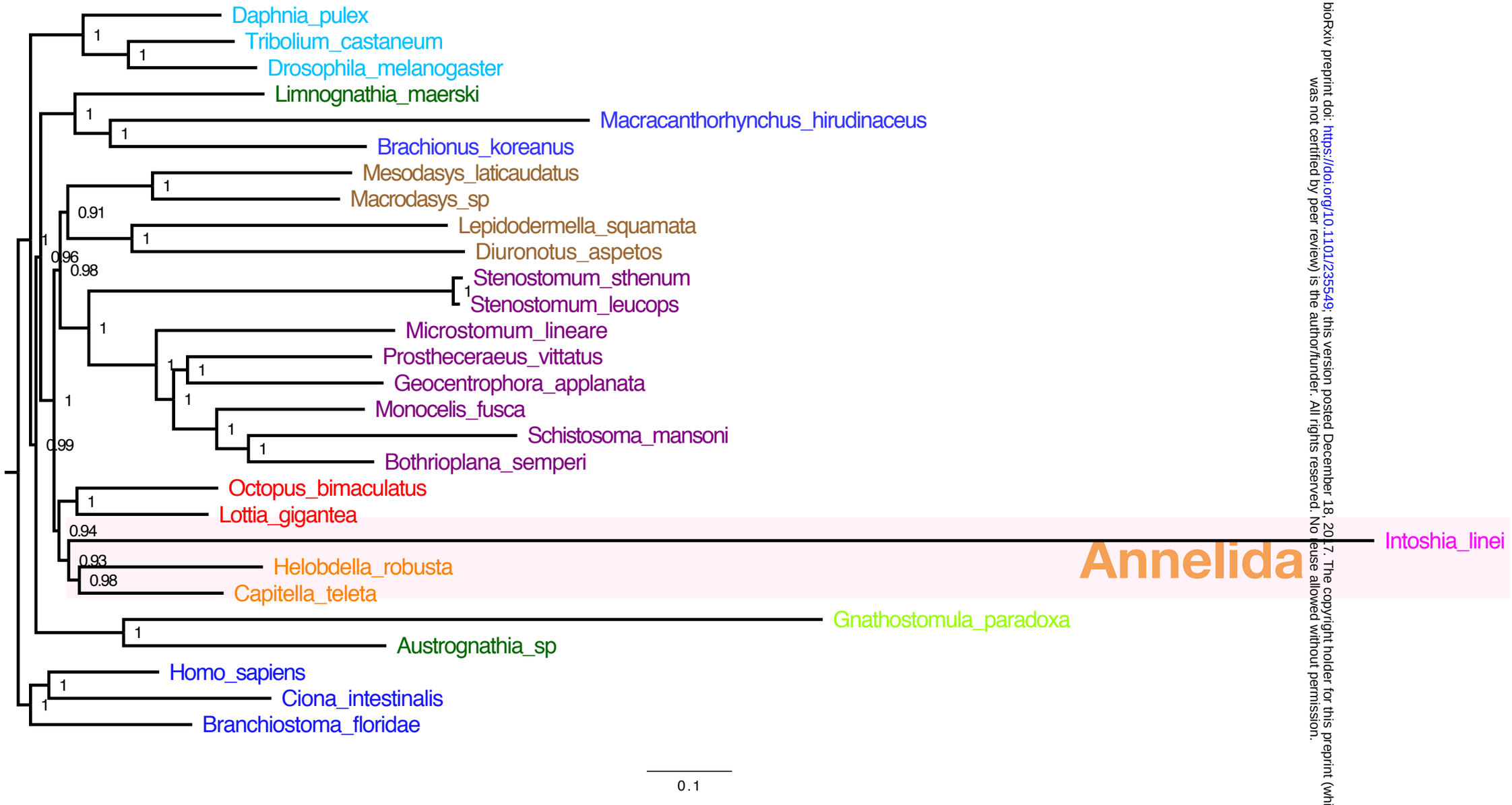


Fig. 4: Reanalysis of a published data set addressing potential LBA between mesozoans supports annelid affinity for *Intoshia*.

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).

Supplementary Tables

Supplementary Table 1

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

Supplementary Table 2

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

Supplementary Figures:

Figure S1. Related to Figure 2.
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.

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.

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.

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.

References:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[16] Simion P, Philippe H, Baurain D, Jager M, Richter DJ, Di Franco A, et al. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. Curr Biol 2017;27:958–67.

[17] Weigert A, Golombek A, Gerth M, Schwarz F, Struck TH, Bleidorn C. Evolution of mitochondrial gene order in Annelida. Mol Phyl Evol 2016;94:196–206.

[18] Slyusarev GS, Kristensen RM. Fine structure of the ciliated cells and ciliary rootlets of *Intoshia variabili* (Orthonectida). Zoomorphology 2002;122:33–9.

[19] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20.

[20] Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 2013;8:1494–512.

[21] Robertson HE, Lapraz F, Egger B, Telford MJ, Schiffer PH. The mitochondrial genomes of the acoelomorph worms *Paratomella rubra*, *Isodiametra pulchra* and *Archaphanostoma ylvae*. Sci Rep 2017;7:1847.

[22] Altschul SF, Gish W, Miller W, Myers EW. Basic local alignment search tool. Journal of Mol Biol 1990;215:403–10.

[23] Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, et al. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phyl Evol 2013;69:313–9.

[24] Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 2004;5:113.

[25] Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 2009;25:1972–3.

[26] Egger B, Lapraz F, Tomiczek B, Müller S, Dessimoz C, Girstmair J, et al. A transcriptomic-phylogenomic analysis of the evolutionary relationships of flatworms. Curr Biol 2015;25:1347–53.

[27] Struck TH, Wey-Fabrizius AR, Golombek A, Hering L, Weigert A, Bleidorn C, et al. Platyzoan paraphyly based on phylogenomic data supports a noncoelomate ancestry of spiralia. Mol Biol Evol 2014;31:1833–49.

[28] Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 2015;16:E9–13.

[29] Altenhoff AM, Škunca N, Glover N, Train C-M, Sueki A, Piližota I, et al. The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. Nucleic Acids Res 2015;43:D240–9.

[30] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 2011;7:1–6.

[31] Morrison DA. Increasing the efficiency of searches for the maximum likelihood tree in a phylogenetic analysis of up to 150 nucleotide sequences. Systematic Biol 2007;56:988–1010.

[32] Tange O. Gnu parallel-the command-line power tool. The USENIX Magazine; 2011.

[33] Wheeler TJ. Large-Scale Neighbor-Joining with NINJA. Algorithms in Bioinformatics, vol. 5724, Berlin, Heidelberg: Springer Berlin Heidelberg; 2009, pp. 375–89.

Figure S1. Related to Figure 2.

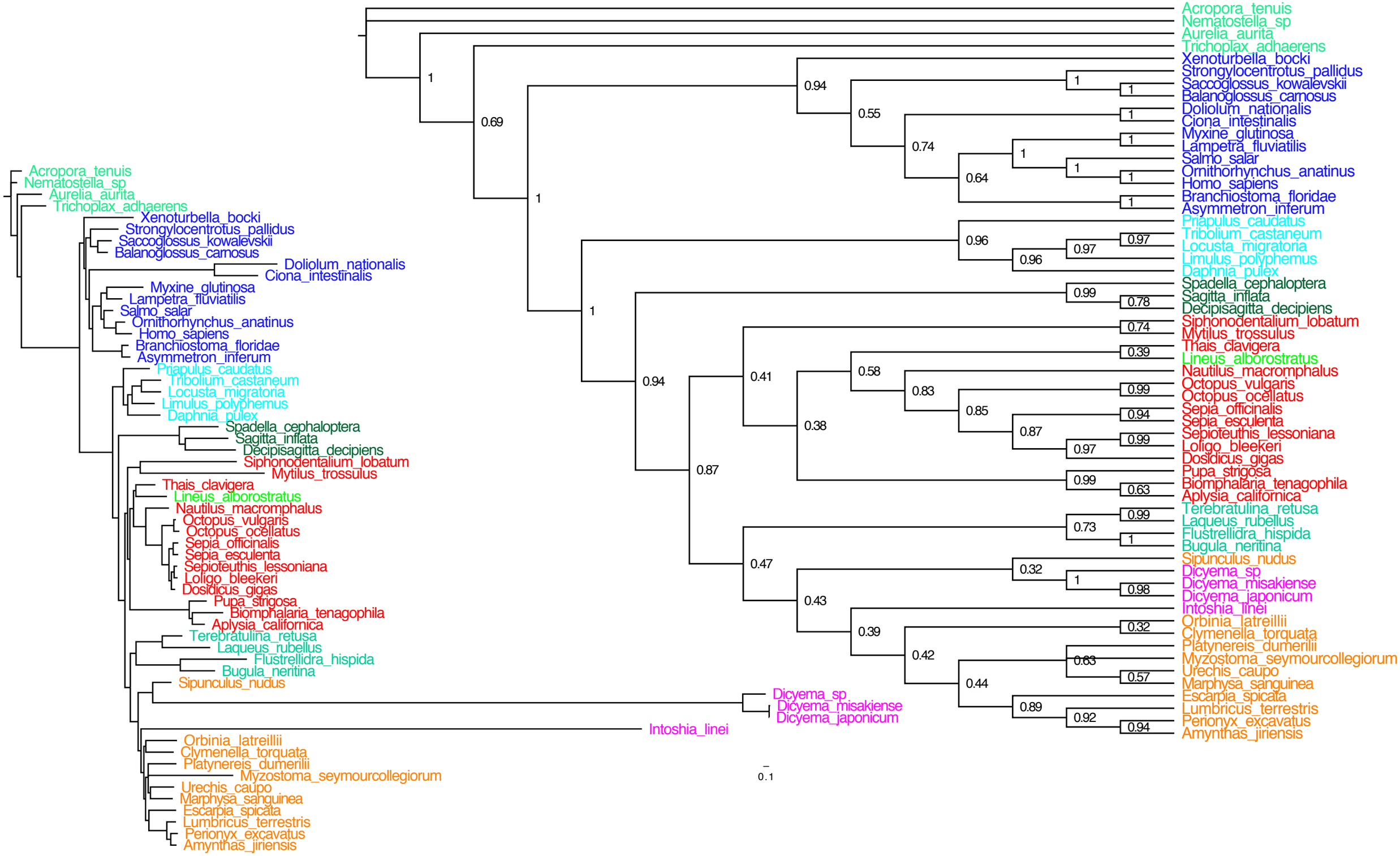


Figure S1:
Phylogram and corresponding cladogram of a Bayesian analysis of our mitochondrial data set omitting the long-branching flatworm species. Phylobayes CAT+G4 model was run in 10 independent runs for 10,000 cycles each on an alignment with 2969 positions and 8000 trees were discarded as burnin.

Figure S2. Related to Figure 2.

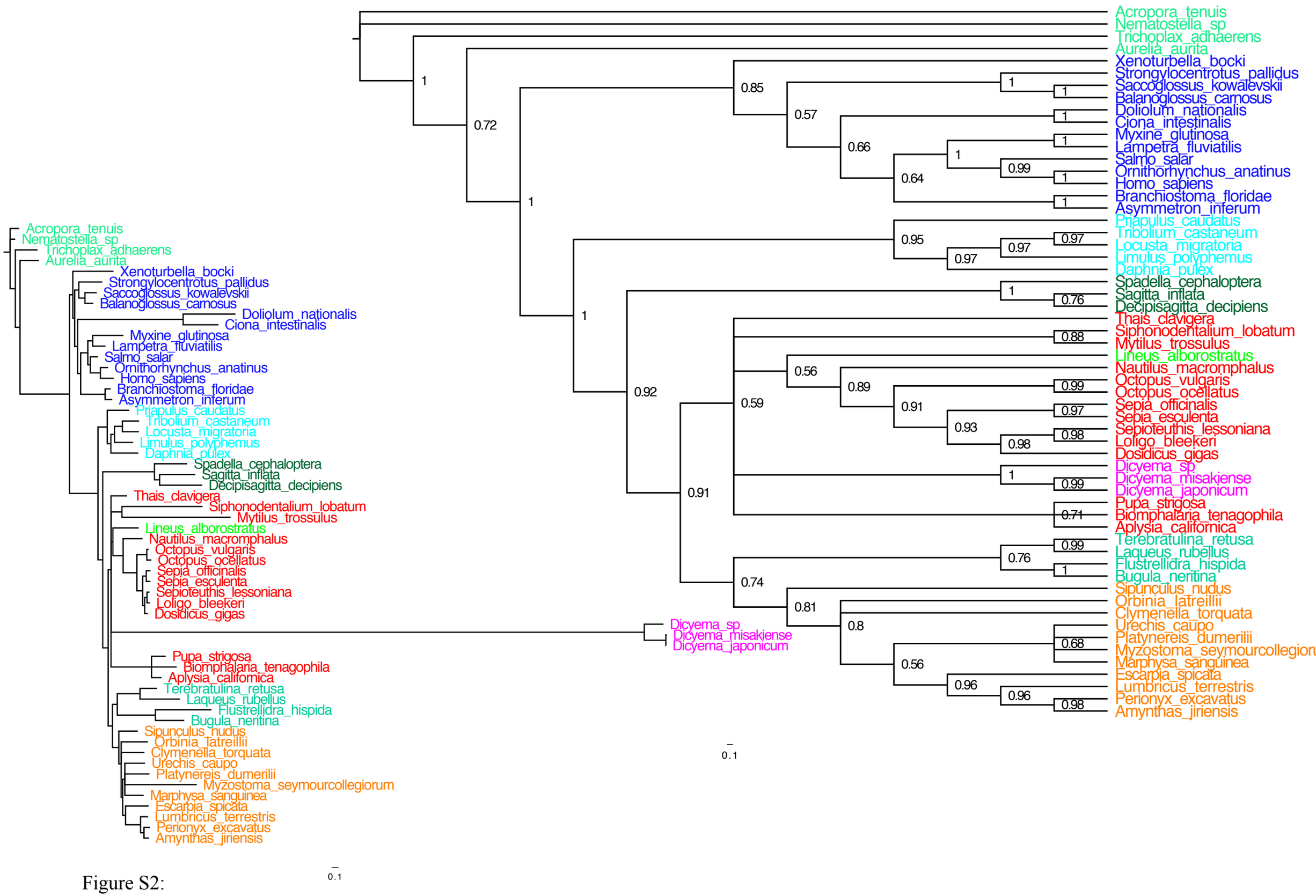


Figure S2:

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.

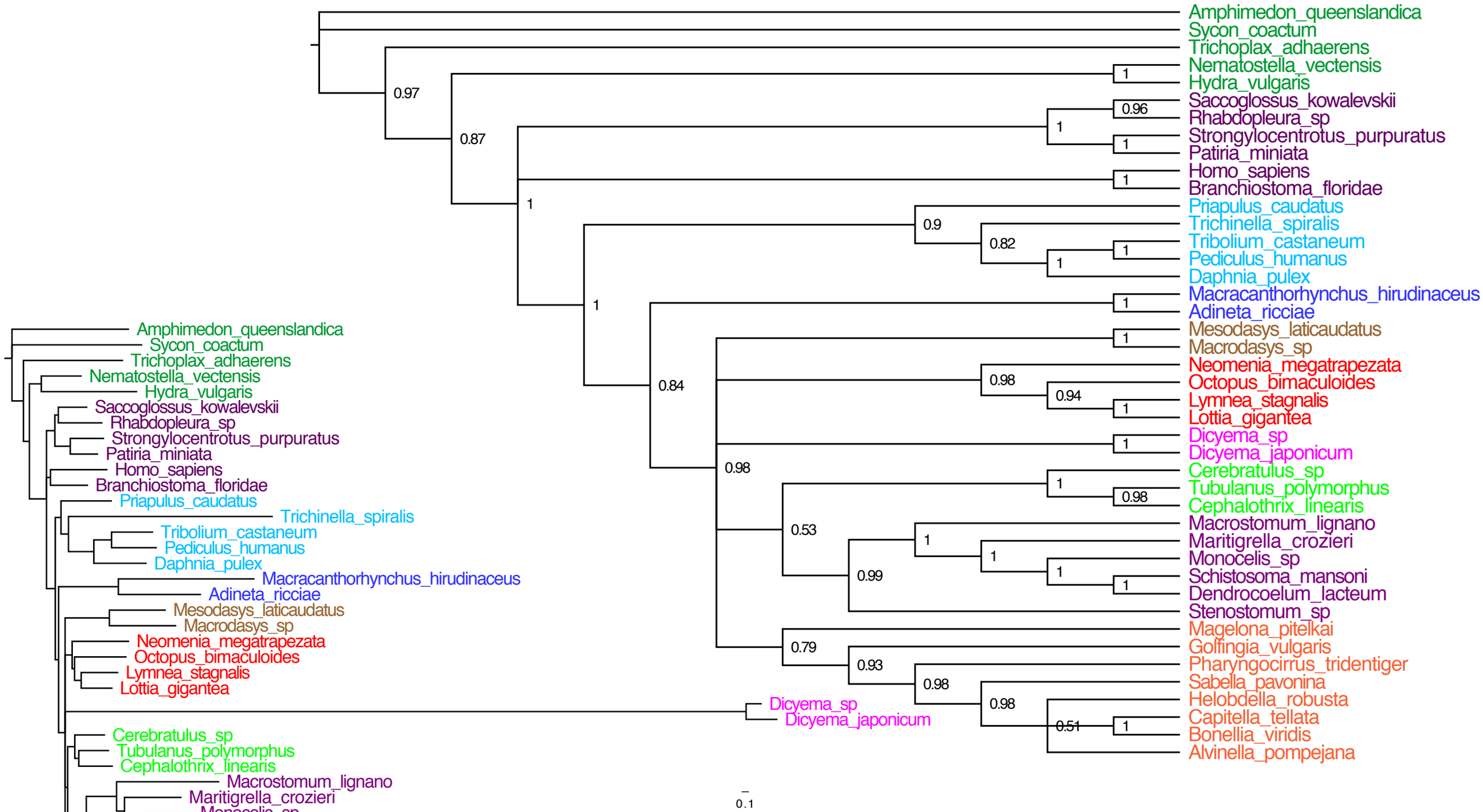
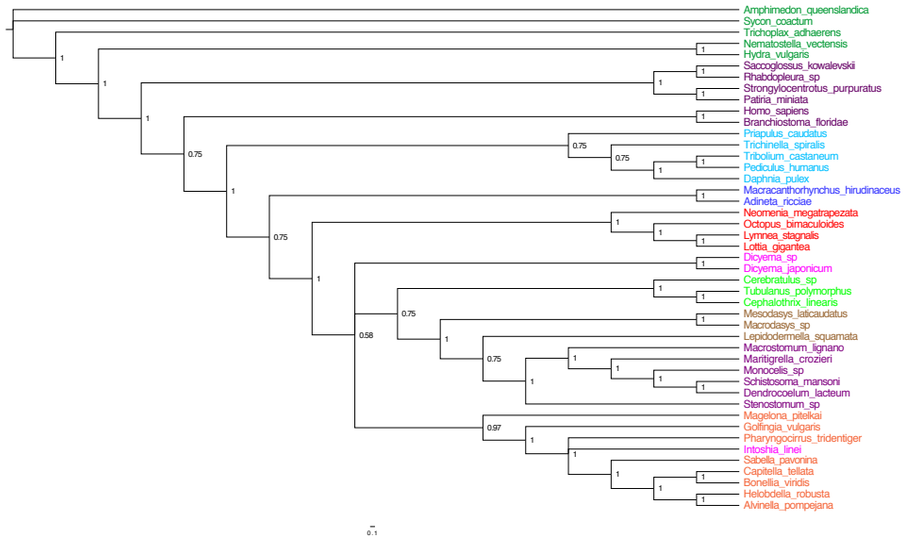
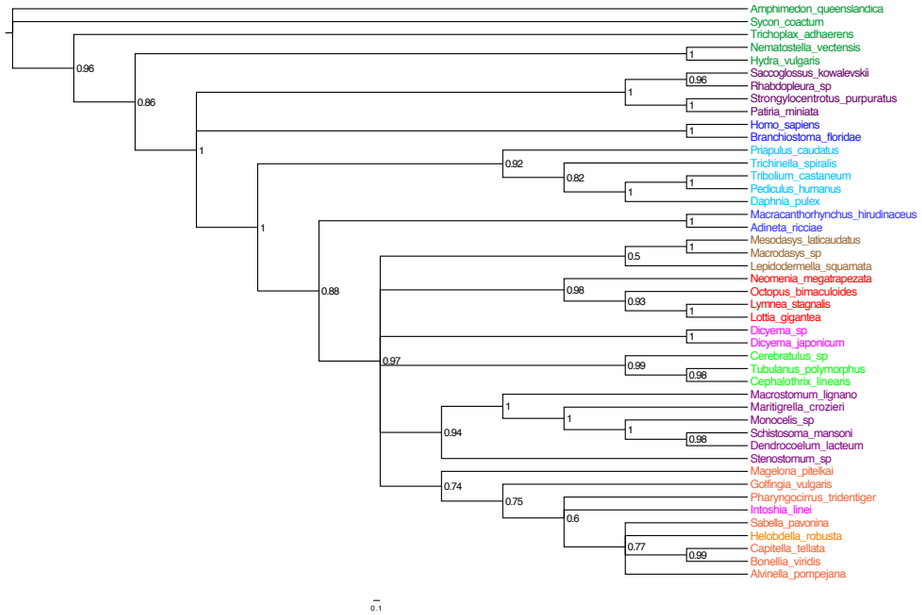


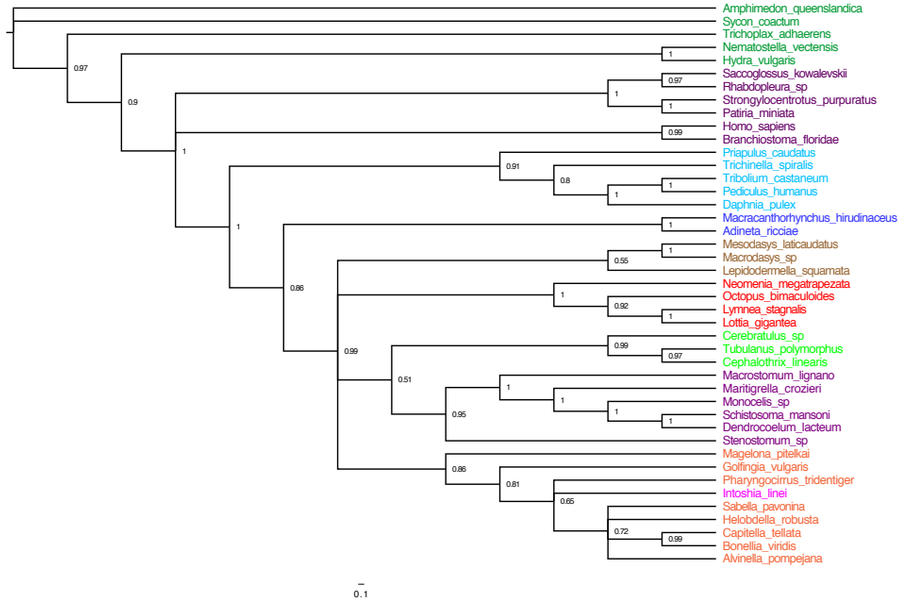
Figure S3:

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.







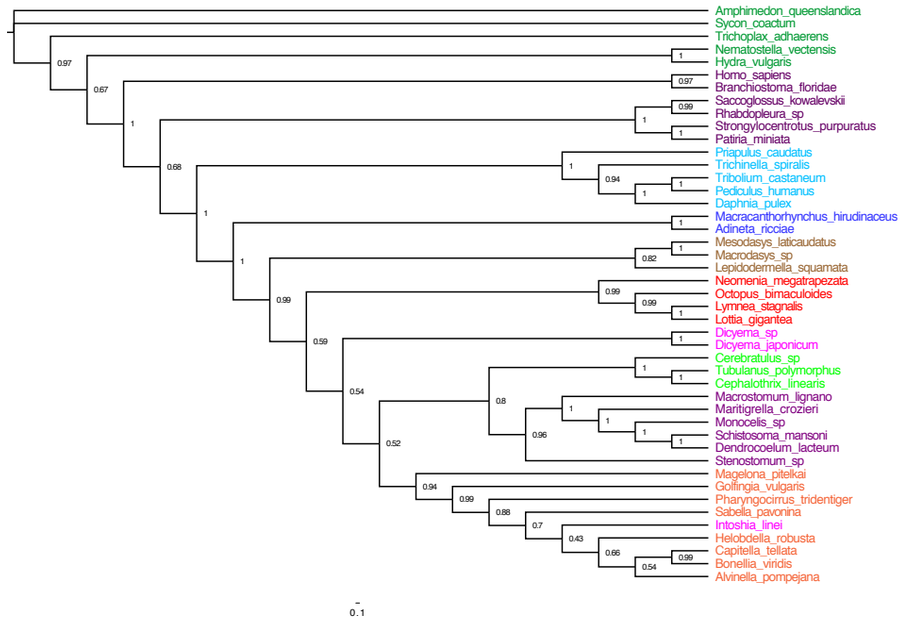


Figure S4:

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.