Supplementary Material

Ctenophore relationships and their placement as the sister group to all other animals

Nathan V. Whelan, Kevin M. Kocot, Tatiana P. Moroz, Krishanu Mukherjee, Peter Williams, Gustav Paulay, Leonid L. Moroz, Kenneth M. Halanych.

Supplementary Discussion

Phylogenetics

Inferred relationships among ctenophores using datasets generated to test relationships among metazoan phyla (Fig. 2, Supplementary Figs. S1-S15) resulted in nearly identical relationships as those inferred with the ctenophore specific datasets (Figs. 2, Supplementary Figs. S16-S19). When relationships differed (e.g., placement of *Dryodora glandiformis*) they were less well supported on trees inferred with datasets Metazoa_ than conflicting nodes on trees generated with the ctenophore specific datasets (i.e., datasets Cteno_; Figs. 2, 3, Supplementary Figs. S1-19). The ctenophore centric datasets had more genes overall and less genes missing from ctenophore species than the metazoan datasets (Supplementary Table S3), which likely explains the more robust placement of ctenophore species in analyses using the ctenophore centric datasets.

Inferred ctenophore relationships were identical for each ctenophore-centric dataset and analytical method (Fig. 2, S15-S18; Extended Data Table 4; all tree files have been deposited on FigShare). Removing outgroups had no effect on inferred relationships, indicating no effect of outgroup choice on ingroup relationships (see tree files deposited on FigShare for trees without outgroups). Such similarities between datasets with different amounts and types of potential causes of systematic error pruned suggest robust phylogenetic hypotheses of ctenophore relationships (Fig. 2, S15-S18).

Model Performance and inferred non-bilaterian relationships

Past phylogenomic studies that have criticized the ctenophores-sister hypothesis have invariably argued that sponges must be the sister lineage to all other animals because trees inferred with site-heterogeneous CAT models have recovered sponges sister to all other animals[13,20,62,87,88]. However, multiple recent studies using CAT models, including the present study, have recovered ctenophores sister to all other animals[11,24,41]. Nevertheless, the argument that CAT models should be used for

phylogenomic inference has deeper flaws[24]. Generally, when a study has increased taxon sampling for any given group, compared to previous studies, trees inferred with site-heterogeneous CAT models and site-homogeneous models are often found to be congruent (see [24] for examples), even if past studies with less taxon sampling recovered incongruent trees with CAT models and other models. Our results are testament to this pattern: some past studies that used both site-homogenous models and CAT models resulted in disagreement on the placement of Ctenophora[20,62,86,87]. However, with greater ctenophore taxon sampling than previous studies, we recovered ctenophores as the sister group to all other animals when using both site-homogeneous models and CAT models (Figs. 2, ED1, S1-S14). This pattern has also been seen among studies analyzing the phylogenetic placement of acoels and *Xenotrubella*[42,88]. We are aware of no instance where phylogenetic inference with CAT-GTR and partitioning on datasets with increased sampling compared to earlier datasets produced congruent trees that also matched those inferred with CAT models on datasets with lower taxon sampling. Rather, trees match those inferred with site-homogeneous models, as seen here with the placement of Ctenophores. Thus, the logical conclusion is that CAT models can often be less accurate than other substitution models at inferring accurate trees, particularly when taxon sampling is limited for critical lineages.

The above should not be misconstrued as an argument against site-heterogeneous models, but merely an argument against models that often recover incorrect relationships and happen to be site-heterogeneous. Moreover, a well-conceived model could be poorly implemented in end-user programs. Current implementations of both CAT-F81 and CAT-GTR do not accurately model site-heterogeneity, as heterogeneity inferred by CAT models arbitrarily scales with dataset size[24]. This is no more realistic than using partitioned site-homogeneous models[24]. In fact, it may be less realistic. Moreover, no one should expect that combining an infinite mixture of sites with equal exchangeability rates among amino acids, as done with CAT-F81, would allow a substitution model to perform well in phylogenetic inference. Equal exchangeabilities among amino acids is simply an unrealistic assumption. Well performing site-

heterogeneous models that are computationally tractable would be beneficial to the field, but CAT-F81 is conclusively unrealistic and results in less accurate trees than CAT-GTR and partitioning. Therefore, the conclusion of Simion et al. [21] that sponges are the sister group to all other extant animals, which was entirely based off analyses with CAT-F81 is flawed.

Molecular Clock Analyses

A time-calibrated tree was inferred with BEAST 2[68] using a relaxed molecular clock (Supplementary Fig. 15). Although the inferred age of some nodes (e.g., the MRCA of sampled bilaterians; Supplementary Fig. 15) are younger than what has been inferred in past studies[73], we were most interested in the inferred age of Ctenophora relative to well-known diversification events. Thus, even with some uncertainty in the age of extant ctenophores, we can still test the 65 MYA bottleneck hypothesis[12,13] and approximately date the ctenophore MRCA with the molecular clock based tree inferred here (Supplementary Fig. 15).

The relative age of the MRCA of extant ctenophores was considerably younger than that of the respective MRCAs of Porifera, Cnidaria, and Bilateria (Supplementary Fig. S15). However, the MRCA of extant ctenophores was inferred as older than the age of the MRCA of *Hemithris digitata* + *Capitella teleta* (~476-551 MYA[73]), but younger than the origin of protostomes (~578-653 MYA[73]). Given the confidence interval associated with the inferred timing of extant ctenophore diversification (Supplementary Fig. 15) and previously hypothesized ages of bilaterian nodes[73], the MRCA of extant ctenophores is most likely no younger than 250 MYA. This age estimate is much older than the previously hypothesized 65 MYA age of crown group ctenophores[12,13]. Even though timing of ctenophore diversification inferred here is rather imprecise, we can reject a species-diversity bottleneck associated with the K-T extinction (~65 MYA). That said, based on the diversity of putative ctenophore fossils that are not morphologically similar to any known, extant species[14-16], plus the observation that

the extant ctenophore MRCA is considerably younger than both the MRCA of sponges and the MRCA of cnidarians (Supplementary Fig. S15), our analysis is consistent with a potentially large loss of diversity in the ctenophore lineage after its split from other Metazoa. We hypothesize that this loss of diversity, or bottleneck, occurred prior to or during the Permian-Triassic extinction[30]. However, we cannot rule out that it may have occurred farther in the past as cydippid fossils are known from the Devonian[90,91]. Future studies will be essential for more precisely testing this hypothesis with additional fossil calibrations and greater metazoan taxon sampling.

Ancestral State Reconstruction

As noted in the methods, characteristics of sampled ctenophores were assigned to each species using previous descriptive work and/or personal observations of individuals we collected (Supplementary Table S5). In some instances, previously reported character states were either unclear or contradictory, and we detail those issues below.

We could find no confirmed report of Platyctenida possessing the ability of bioluminescence, and we have never observed bioluminescence when collecting platyctenids at night. The site of bioluminescence in at least some ctenophores is below their comb rows[36], but all platyctenids lose their comb rows during development (except *Ctenoplana*, which we were unable to collect). Therefore, most platyctenids may simply lose the ability of bioluminescence during development. To account for this uncertainty, platyctenids collected here were coded as ambiguous concerning their character state for bioluminescence (Supplementary Table S5). The ability of bioluminescence has also not been explicitly addressed in the literature for *Pukia falcata*. We have observed this species at night, but we have not observed bioluminescence. Given this, and the placement of *P. falcata* as nested in a clade with other species that are not bioluminescent (Supplementary Fig. S24), we coded *P. falcata* as lacking bioluminescence (Supplementary Table S5).

Most character states for feeding mode were obtained from Haddock[38] with three exceptions. First, we coded *Cestum veneris* as capturing food primarily with tentacles rather than lobes. Although the ribbon morphology of cestids is derived from an ancestor with body lobes (Fig. 3), as hypothesized by Haddock[38], we argue that food capture by cestids is ultimately done with tentacles as noted by Matsumoto and Harbison[78] and Stretch[80]. Therefore, *Cestum veneris* was coded as using tentacles as its primary means of food capture. Second, according to the original species description[77], *Lobatolampea tetragona* feeds similarly to *Cestum veneris* and was coded as using primarily tentacles for feeding. Finally, even though *Dryodora glandiformis* possesses tentacles that they may use to sense stimuli, including food, we coded their primary food capture method as engulfing. There are no reports of *Dryodora glandiformis* physically capturing its prey with tentacles, and we doubt the simplified tentacles of *Dryodora glandiformis* could be used to capture the larvaceans it exclusively feeds upon. More broadly, one could argue that all species with lobes, except *Ocyropsis* because adults lack tentacles, use tentacles as adults in some fashion for food capture, rather than just their lobes. Thus, one could conceivably code feeding mode in a much finer manner. However, we were interested in broad evolutionary patterns so we coded character states as primary food capture mode rather than splitting feeding and food capture mode into many different character states that would have provided little insight into macroevolutionary patterns.

The ancestral state reconstruction analyses reported in the main text (Figs. 3-5, Supplementary Figs. S20-S22, S24, S25) ignored uncertainty in both relationships and branch length. In order to estimate how uncertainty in branch-length may, or may not, affect ancestral state reconstruction, we used MrBayes 3.2.6[90] to generate a posterior distribution of trees for dataset Cteno_RCFV_LB. A full analysis in MrBayes would not have converged in a reasonable time frame, so relationships were constrained based on the topology inferred using Cteno_RCFV_LB and RAxML (Fig. 2), but branch lengths were estimated. The dataset was partitioned following best-fit partitions as inferred with PartitionFinder. We

used two runs with four metropolis coupled MCMC chains to estimate branch lengths, and each run was sampled every 1000 generations for 2.68×10^6 generation; we also sampled across model space using rjMCMC (MrBayes command nst=mixed) because not all best-fit models were implemented in MrBayes. Convergence was tested using the MrBayes sump command and a burn-in of 25%; standard deviation of split frequencies was 0.00 and potential scale reduction factor of each parameter was 1.0, indicating convergence of independent Bayesian runs. Joint posterior probabilities of ancestral states at each node was inferred as described in the methods section, but 50 trees from the post-burn in posterior distribution of trees was used and only 1,000 MCMC generations of stochastic mapping were run for each tree in the posterior distribution; this was done to limit required computational time. Incorporating branch-length uncertainty into ancestral state reconstruction did not have a meaningful effect on inferred states (data available on FigShare). We chose to emphasize the analysis where uncertainty was ignored for two reasons: 1) forcing topological constraints on the MrBayes analyses was less than ideal and merely done for computational reasons, 2) many of the best-fit models (e.g., LG) are not implemented in MrBayes, possibly resulting in less accurate branch length estimates than those inferred with RAxML.

Ribosomal Gene Tree

Despite great efforts to sample as many ctenophore lineages as possible, obtaining tissue samples suitable for transcriptome sequencing was not possible for some lineages. We were also unable to photograph every sampled individual before preserving tissue. Therefore, we also assembled an 18S rRNA tree using sequences obtained from GenBank and transcriptomes sequenced here, when possible (Supplementary Table S6); we were unable to recover reasonably complete 18S rRNA genes from some transcriptomes. The 18S rRNA gene tree was inferred with RAxML using the GTR+Γ substitution model, and nodal support was assessed with 1,000 fast bootstrap replicates (Supplementary Fig. S23).

Specimens sequenced here with useable 18S sequences were recovered as close relatives of individuals from the same species that were sequenced in past studies (Supplementary Fig. S23)[12,13]. This is evidence that these species identifications were accurate, or at least consistent with those of previous workers. The inferred 18S rRNA tree also suggests possible identifications for some specimens we were not able to name. For example, we sequenced an unidentified *Pleurobrachia* sp. Florida, USA that has an 18S sequence that is nearly identical to that of a specimen of *Pleurobrachia brunnea* sequenced by Simion et al.[13].

Most deep nodes in the 18S tree had low BS support (<50), but no strongly-supported nodes were in conflict with our transcriptome based trees (Figs. 3-5, Supplementary Figs. S16-S19, S23). Consistent with our phylogenomic analyses, the monotypic family Pukiidae (*Pukia falcata*) is nested within Pleurobrachiidae on the 18S gene tree. Analysis of 18S supports the paraphyly of Mertensiidae, albeit with poor BS support. Although 18S appears useful for confirming species ID, the general lack of support for most nodes illustrates the usefulness of the transcriptome-based phylogenomic approach used here for inferring relationships among ctenophores.

References

86 Philippe, H. *et al.* Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* **9**, e1000602, (2011).

87 Philippe, H. *et al.* Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* **19**, 706-712, (2009).

88 Simakov, O. *et al.* Hemichordate genomes and deuterostome origins. *Nature* **527**, 459-465, (2015).

89 Ronquist, F. *et al.* MrBayes 3.2: efficient bayes inference and model choice across a large model space. *Syst. Biol.* **61**, 539-542, (2012).

90 Stanley, G.D. & Stürmer, W. The first fossil ctenophore from the lower Devonian of West Germany. *Nature*. **303**, 518-520, (1983).

91 Stanley, G.D. & Stürmer, W. A new fossil ctenophore discovered by X-rays. *Nature*. **328**, 61-63,

Table S1: Taxon sampling for ctenophore-centric phylogenetic analyses

Phylum	Family	Species	Collection Locality^	Latitude (°)	Longitude (°)	Raw Reads	Assembled Transcripts	NCBI SRA, BioProject, or other Accession
Ctenophora								
	Euplokamididae	*Euplokamis dunlapae*	-	-	-	34,151,349	160,775	SRR777663
	Coeloplanidae	*Coeloplana astericola*	-	-	-	20,842,610	111,307	SRR786490
		Benthoplana meteoris	Luzon Island, Philippines	13.93	120.61	52,070,847	29,620	PRJNA396415
		Vallicula sp.	-	-	-	24,545,186	159,357	SRR786489
	Pleurobrachiidae	*Hormiphora californica*	-	-	-	32,337,982	79,758	SRR1992642
		Hormiphora palmata	Kona Coast, Hawaii, USA	19.617	-156.085	47,328,392	50,591	PRJNA396415
		Pleurobrachia bachei	-	-	-	N/A	19,524	http://neurobase.rc.ufl.edu/pleurobrachia/
		Pleurobrachia pileus	New Jersey, USA	40.383	-73.801	29,169,744	56,415	PRJNA396415
		Pleurobrachia pileus	-	-	-	25,313,211	197,803	SRR789901
		Pleurobrachia sp.	Florida, USA	30.337	-81.661	17,434,098	43,885	PRJNA396415
		Pleurobrachia sp.	South Carolina, USA	32.029	-79.725	16,271,430	49,172	PRJNA396415
		Pleurobrachia sp.	South Carolina, USA	32.029	-79.725	39,919,758	59,827	PRJNA396415
	Pukiidae	*Pukia falcata*	Queensland, Australia	-27.25	153.25	94,637,042	117,469	PRJNA396415
	Mertensiidae	*Callianira antarctica*	Antarctica	-65.095	-63.168	61,446,864	68,595	PRJNA396415
		Mertensiidae sp. (Washington, USA)	-	-	-	23,727,123	134,815	SRR786492
		Mertensiidae sp.	Antarctica	-64.65	-62.397	54,973,357	97,238	PRJNA396415
	Dryodoridae	*Dryodora glandiformis*	-	-	-	20,634,583		SRR777788
	Beroidae	*Beroe ovata*	South Carolina, USA	32.029	-79.725	15,579,602	51,295	PRJNA396415
		Beroe sp.	Queensland, Australia	-27.25	153.25	44,128,606	58,258	PRJNA396415
		Beroe forskalii	South Carolina, USA	32.553	-79.308	37,374,286	121,008	PRJNA396415
		Beroe sp.	Antarctica	-64.406	-61.916	52,294,485	37,475	PRJNA396415
		Beroe abyssicola	-	-	-	22,722,322		SRR777787
	Bolinopsidae	*Bolinopsis ashleyi*	Queensland, Australia	-27.25	153.25	47,700,132	72,847	PRJNA396415
		Bolinopsis infundibulum	Washington, USA	48.545	-123.012	32,028,806	143,811	PRJNA396415
		Mnemiopsis leidyi	-	-	-	N/A	16,548	https://kona.nhgri.nih.gov/mnemiopsis/
		Mnemiopsis leidyi	New Jersey, USA	39.717	-73.598	28,950,980	71,599	PRJNA396415
		Mnemiopsis mccradyi	Florida, USA	-	-	67,036,948	149,455	PRJNA396415
	Eurhamphaeidae	*Eurhamphaea vexilligera*	Bimini, Bahamas	25.733	-79.25	44,812,330	47,832	PRJNA396415
	Cestidae	*Cestum veneris*	Bimini, Bahamas	25.733	-79.25	146,947,617	99,132	PRJNA396415
	Ocyropsidae	*Ocyropsis crystallina*	North Carolina, USA	34.444	-75.972	20,398,482	143,811	PRJNA396415
		Ocyropsis sp.	Florida, USA	26.709	-80.064	65,159,704	116,128	PRJNA396415
		Ocyropsis crystallina guttata	Bimini, Bahamas	25.733	-79.25	43,217,528	95,632	PRJNA396415
	Lobatolampeidae	*Lobatolampea tetragona*	Luzon Island, Philippines	13.93	120.61	66,945,305	72,868	PRJNA396415
	Unidentified	Lobata sp.	Punta Arenas, Chile	-53.17	-70.907	51,629,293	57,045	PRJNA396415
		Cydippida sp. (Washington, USA)‡	-	-	-	21,688,585	174,129	SRR786491
		Cydippida sp.	Antarctica	-63.439	-55.453			PRJNA396415
		Cydippida sp.	Maryland, USA	37.863	-74.329	41,738,834	128,628	PRJNA396415
		Ctenophora sp. (larval specimen)‡	Florida, USA	30.337	-81.661	23,206,288	50,273	PRJNA396415
Porifera		*Sycon coactum*		-	-	N/A	70,220	http://dx.doi.org/10.7910/DVN/24737
		Latrunculia apicalis		-	-	12,691,254	76,210	SRR1915755
		Spongilla alba		-	-	N/A	56,696	http://dx.doi.org/10.7910/DVN/24737
		Amphimedon quenslandica		-	-	N/A	63,542	dbEST
Cnidaria		*Eunicella verrucosa*		-	-	70,071,835	32,637	SRR1324944; SRR1324945
		Nematostella vectensis		-	-	N/A		Joint Genome Institute
		Hydra vulgaris		-	-	N/A	45,250	dbEST
		Agalma elegans		-	-	53,998,182	217,596	SRR871526

‡Taxa excluded from ancestral state reconstruction

^Not provided for taxa sequenced elsewhere

Tables S2: Taxon sampling for Metazoa phylogenetic analyses

Taxon	NCBI or other accesion
Ichthyosporea	
Amoebidium parasiticum †	www.broadinstitute.org/annotation/genome/multicellularity_project/MultiHome.html
Sphaeroforma arctica †	www.broadinstitute.org/annotation/genome/multicellularity_project/MultiHome.html
Filasteria	
Capsaspora owczarzaki †	www.broadinstitute.org/annotation/genome/multicellularity_project/MultiHome.html
Ministeria vibrans †	SRR343051
Choanoflagellata	
Acanthoeca sp.	SRR1294413; SRR1296844
Monosiga brevicolis	Joint Genome Institute
Monosiga ovata	NCBI dbEST
Salpingoeca pyxidium	SRR1915694
Salpingoeca rosetta	www.broadinstitute.org/annotation/genome/multicellularity_project/MultiHome.html
Ctenophora	
Bereo forskalii	PRJNA396415
Beroe abyssicola	SRR777787
Beroe ovata	PRJNA396415
Beroe sp. Antarctica	PRJNA396415
Beroe sp. Queensland, Australia	PRJNA396415
Bolinopsis ashleyi	PRJNA396415
Bolinopsis infundibulum ‡	PRJNA396415
Callianira antarctica	PRJNA396415
Cestum veneris	PRJNA396415
Coeloplana astericola	SRR786490
Ctenophora sp. Florida, USA	PRJNA396415
Cydippida sp. Maryland, USA	PRJNA396415
Cydippida sp. Washington, USA	SRR786491
Dryodora glandiformis ‡	SRR777788
Euplokamis dunlapae	SRR777663
Eurhamphaea vexilligera	PRJNA396415
Hormiphora californica	PRJNA396415
Hormiphora palmata	PRJNA396415
Lobata sp. Punta Arenas‡	PRJNA396415
Lobatolampea tetragona‡	PRJNA396415
Mertensiidae sp. Antarctica	PRJNA396415
Mertensiidae sp. Washington, USA	PRJNA396415
Mnemiopsis leidyi	SRR789900
Ocyropsis crystallina	PRJNA396415
Ocyropsis sp. Florida, USA	PRJNA396415
Ocyropsis crystallina guttata	PRJNA396415
Pleurobrachia bachei ‡	neurobase.rc.ufl.edu.pleurobrachia
Pleurobrachia pileus	PRJNA396415
Pleurobrachia sp.	SRR789901
Vallicula sp.	SRR786489
Porifera	
Amphimedon queenslandica	NCBI dbEST
Aphrocallistes vastus	Evolution and Research Archive
Chondrilla nucula	http://dx.doi.org/10.7910/DVN/24737
Cliona varians	SRR1391011
Corticium candelabrum	http://dx.doi.org/10.7910/DVN/24737
Crella elegans	http://dx.doi.org/10.5061/dryad.50dc6/3
Hyalonema populiferum	SRR1916923
Ircinia fasciculata	http://dx.doi.org/10.7910/DVN/24737
Kirkpatrickia variolosa	SRR1916957
Latrunculia apicalis	SRR1915755
Mycale phytophylla	SRR1711043
Oscarella carmela	www.compagen.org
Petrosia ficiformis	http://dx.doi.org/10.7910/DVN/24737
Pseudospongosorites suberitoides	http://dx.doi.org/10.7910/DVN/24737
Rossella fibulata	SRR1915835
Spongilla alba	http://dx.doi.org/10.7910/DVN/24737
Sycon ciliatum	ERR592861
Sycon coactum	http://dx.doi.org/10.7910/DVN/24737
Sympagella nux	SRR1916581
Placozoa	
Trichoplax adhaerens	Joint Genome Institute
Cnidaria	
Abylopsis tetragona	SRR871525
Acropora digitifera	DRR055157
Agalma elegans	SRR871526
Aiptasia pallida	SRR696721; SRR696732; SRR696745
Bolocera tuediae	SRR504347
Craseo lathetica	SRR871529
Eunicella verrucosa	SRR1324944; SRR1324945
Hormathia digitata	SRR504348
Hydra oligactis	SRR040466; SRR040467; SRR040468; SRR040469
Hydra viridissima	SRR040470; SRR040471; SRR040472; SRR040473
Hydra vulgaris	NCBI dbEST
Nanomia bijuga	SRR871527
Nematostella vectensis	Joint Genome Institute
Periphylla periphylla	SRR191582
Physalia physalia	SRR871528
Bilateria	
Capitella teleta	Joint Genome Institute
Daphnia pulex	Joint Genome Institute
Drosophila melanogaster	HaMStR Core Orthologs
Hemithris digitata	SRR1611556
Homo sapiens	HaMStR Core Orthologs
Strongylocentrotus purpuratus	InParanoid Database

‡Species excluded from CAT-GTR analyses to facilitate Bayesian convergence

Supplementary Table S3: Phylogenetic datasets

Dataset	Number of Genes	Number of Amino Acids	Missing data (%)	Genes that may cause systematic error removed
Metazoa_full	224	68,082	46.25	Paralogs
Metazoa_RCFV_relaxed	205	65,347	46.98	Paralogs, heterogeneous genes
Metazoa_RCFV_strict	116	43,324	50.20	Paralogs, heterogeneous genes
Metazoa_LB_relaxed	164	51,211	45.70	Paralogs, long-branched genes
Metazoa_LB_strict	156	46,959	45.62	Paralogs, long-branched genes
Metazoa_RCFV_LB_relaxed	149	49,051	46.45	Paralogs, long-branched genes, heterogeneous genes
Metazoa_RCFV_LB_strict	74	28,759	49.50	Paralogs, long-branched genes, heterogeneous genes
Metazoa_Choano$^{\alpha}$	234	75,840	48.79	Paralogs
Metazoa_Choano_RCFV_relaxed$^{\alpha}$	161	59,699	50.27	Paralogs, heterogeneous genes
Metazoa_Choano_RCFV_strict$^{\alpha}$	127	49,405	52.81	Paralogs, heterogeneous genes
Metazoa_Choano_RCFV_strict_Bayes$^{\alpha}$	127	49,388	50.34	Paralogs, heterogeneous genes, unstable taxa
Metazoa_Choano_LB_relaxed$^{\alpha}$	164	51,211	45.70	Paralogs, long-branched genes
Metazoa_Choano_LB_strict$^{\alpha}$	156	46,959	45.62	Paralogs, long-branched genes
Metazoa_Choano_RCFV_LB_relaxed$^{\alpha}$	149	49,051	46.45	Paralogs, long-branched genes, heterogeneous genes
Metazoa_Choano_RCFV_LB_strict$^{\alpha}$	74	28,759	49.50	Paralogs, long-branched genes, heterogeneous genes
Ctenophore_full*	350	98,844	43.55	Paralogs
Ctenophore_RCFV*	280	84,187	43.57	Paralogs, heterogeneous genes
Ctenophore_ LB*	268	78,100	43.49	Paralogs, long-branched genes
Ctenophore_RCFV_LB*	217	68,194	44.02	Paralogs, long-branched genes, heterogeneous genes

*Datasets with outgroups removed were also analyzed

$^{\alpha}$Non-choanoflagellate outgroups were excluded during orthology determination and downstream dataset filtering

Supplementary Table S4: Fossil Calibrations for molecular clock analyses that failed to reach convergence

Node (MRCA)	Calibration Shape	mean	sigma	alpha	beta	offset
Metazoa	normal	750	35	-	-	-
Cnidaria	gamma	-	-	2.0	2.0	529
Bilateria	gamma	-	-	2.0	2.0	554
Duetrostomia	gamma	-	-	2.0	2.0	515
Porifera	gamma	-	-	2.4	3.3	535
Protostomia	gamma	-	-	2.0	2.0	552

Supplementary Table S5: Traits of extant taxa used for ancestral state reconstruction

Species	Body Plan	Tentacles (Present/Absent)	Tentacles in Adults (Present/Absent)	Smooth muscles (Present/Absent)	Striated muscles (Present/Absent)	Separate Sexes (Present/Absent)	Benthic or Pelagic	Biolumenescence (Present/Absent)	Feeding Mode
Benthoplana meteoris	P	P	P	P	A	A	B	P,A	T
Beroe abyssicola	N	A	A	P	A	A	Pe	P	E
Beroe forskalii	N	A	A	P	A	A	Pe	P	E
Beroe ovata	N	A	A	P	A	A	Pe	P	E
Beroe sp. Antarctica	N	A	A	P	A	A	Pe	P	E
Beroe sp. Queensland, Australia	N	A	A	P	A	A	Pe	P	E
Bolinopsis ashleyi	L	P	P	P	A	A	Pe	P	L
Bolinopsis infundibulum	L	P	P	P	A	A	Pe	P	L
Callianira antarctica	C	P	P	P	A	A	Pe	P	T
Cestum veneus	R	P	P	P	A	A	Pe	P	T
Coeloplana astericola	P	P	P	P	A	A	B	P,A	T
Cydippida sp. Antarctica	C	P	P	P	A	A	Pe	P	T
Cydippida sp. Maryland	C	P	P	P	A	A	Pe	P	T
Dryodora glandiformis	C	P	P	P	A	A	Pe	P	E
Euplokamis dunlapae	C	P	P	P	P	A	Pe	P	T
Eurhamphaea vexilligera	L	P	P	P	A	A	Pe	P	L
Hormiphora californica	C	P	P	P	A	A	Pe	P	T
Hormiphora palmata	C	P	P	P	A	A	Pe	P	T
Lobata sp. Punta Arenas, Chile	L	P	P	P	A	A	Pe	P	L
Lobatolampea tetragona	L	P	P	P	A	A	B	P	T
Mertensiidae sp. Washington, USA	C	P	P	P	A	A	Pe	P	T
Mnemiopsis leidyi	L	P	P	P	A	A	Pe	P	L
Mnemiopsis mccradyi	L	P	P	P	A	A	Pe	P	L
Mnemiopsis sp. New Jersey, USA	L	P	P	P	A	A	Pe	P	L
Ocyropsis crystallina	L	P	A	P	A	P	Pe	P	L
Ocyropsis sp. Bimini, Bahamas	L	P	A	P	A	P	Pe	P	L
Ocyropsis sp. Florida, USA	L	P	A	P	A	P	Pe	P	L
Pleurobrachia bachei	C	P	P	P	A	A	Pe	A	T
Pleurobrachia pileus (1)	C	P	P	P	A	A	Pe	A	T
Pleurobrachia pileus (2)	C	P	P	P	A	A	Pe	A	T
Pleurobrachia sp. Forida, USA	C	P	P	P	A	A	Pe	A	T
Pleurobrachis sp. (1) South Carolina, USA	C	P	P	P	A	A	Pe	A	T
Pleurobrachis sp. (2) South Carolina, USA	C	P	P	P	A	A	Pe	A	T
Pukia falcata	C	P	P	P	A	A	Pe	A	T
Vallicula sp.	P	P	P	P	A	A	B	P,A	T
Agalma elegans	-	-	-	P	P	-	-	-	
Hydra vulgaris	-	-	-	P	A	-	-	-	
Nematostella vectensis	-	-	-	P	A	-	-	-	
Eunicella verrucosa	-	-	-	P	A	-	-	-	
Amphimedon queenslandica	-	-	-	A	A	-	-	-	
Spongilla lacustris	-	-	-	A	A	-	-	-	
Latrunculia apicalis	-	-	-	A	A	-	-	-	
Sycon coactum	-	-	-	A	A	-	-	-	

Body Plan: P = Platyctenid, N = Nuda, C = Cydippida, L = Lobata, R = Ribbon
Feeding Mode: T = Tentacles, L = Lobes, E = Engulfing

Taxon	NCBI Accession or other source
Pleurobrachia_pileus	AF293678
Pleurobrachia_bachei	AF293677
Hormiphora_plumosa	AF293676
Hormiphora_sp	AF100944
Euplokamis_sp	HE647719.2
Mertensia_ovum	FJ668937
Mertensia_ovum	AF293679
Haeckelia_beehleri	AF293673
Haeckelia_rubra	AF293674
Ctenophora_sp1_Podar_et_al_2001	AF293676
Ctenophora_sp2_Podar_et_al_2001	AF293680
Ctenophora_sp3_Podar_et_al_2001	AF293681
Beroe_forskalii	AF293697
Beroe_forskalii	AF293698
Beroe_cucumis	AF293695
Beroe_cucumis	AF293699
Beroe_cucumis	D15068
Beroe_ovata	AF293694
Beroe_gracilis	AF293696
Beroe_abyssicola	JN673817
Coeloplana_bocki	HQ435813
Coeloplana_anthostella	HQ435810
Vallicula_multiformis	AF293684
Lampocteis_cruentiventer	KF202290
Bolinopsis_infundibulum	AF293687
Leucothea_pulchra	AF293688
Ocyropsis_maculata	AF293689
Ocyropsis_crystallina_crystallina	AF293690
Ocyropsis_crystallina_guttata	AF293691
Ctenophora_sp4_Podar_et_al_2001	AF293686
Cestum_veneris	AF293692
Velamen_parallelum	AF293693
Thalassocalyce_inconstans	AF293685
Mnemiopsis_leidyi	AF293700
Coeloplana_aginae	AF358112
Charistephane_fugiens	AF293682
Charistephane_fugiens	AF358113
Coeloplana_bannwarthii	AF293683
Deiopea_kaloktenota_Simion_et_al_2014	KJ754160
Lampea_pancerina_Simion_et_al_2014	KJ754155
Pleurobrachia_brunnea_Simion_et_al_2014	KJ754154
Cestum_veneris_Simion_et_al_2014	KJ754161
Pleurobrachia_pileus_Simion_et_al_2014	KJ754153
Mnemiopsis_leidyi_Simion_et_al_2014	KJ754158
Coeloplana_cf_meteoris_Simion_et_al_2014	KJ754157
Beroe_forskalii_Simion_et_al_2014	KJ754156.1
Leucothea_multicornis_Simion_et_al_2014	KJ754159
Pleurobrachia_sp_Moroz_et_al_2015	MF599304
Vallicula_multiformis_Moroz_et_al_2015	MF599305
Bolinopsis_infundibulum_Moroz_et_al_2015	MF599306
Euplokamis_dunlapae_Moroz_et_al_2015	MF599307
Mertensiidae_sp_FHL_Moroz_et_al_2015	MF599308
Coeloplana_astericola_Moroz_et_al_2015	MF599309
Beroe_abyssicola_Moroz_et_al_2015	MF599310
Dryodora_glandiformis_Moroz_et_al_2015	MF599311
Hormiphora_palmata	MF599312
Pleurobrachia_pileus	MF599313
Lobatolampaea_tetragona	MF599314
Beroe_sp_Antarctica	MF599315
Beroe_spQueensland_Australia	MF599316
Beroe_gracilis	MF599317
Beroe_ovata	MF599318
Bolinopsis_ashleyi	MF599319
Ctenophora_sp_Florida_USA	MF599320
Mertensiidae_sp_Antarctica	MF599321
Pukia_falcata	MF599322
Ctenophora_sp_Bahamas	MF599323
Lobata_sp_Punta_Arenas_Chile	MF599324
Cestum_veneris	MF599325
Pleurobrachiidae_sp_Gulf_of_Mexico	MF599326
Ocyropsis_sp	MF599327
Hormiphora_californiensis	MF599328
Ocyropsis_sp_Florida_USA	MF599329
Mnemiopsis_mccradyi	MF599330
Mnemiopsis_sp_New_Jersey_USA	MF599331
Ocyropsis_crystallina	MF599332
Callianira_antarctica	MF599333
Pleurobrachia_sp_1_South_Carolina_USA	MF599334
Pleurobrachia_sp_2_South_Carolina_USA	MF599335
Cydippida_sp_Maryland	MF599336
AEGINA	Mallatt et al. 2010
AGLAUROPSIS	Mallatt et al. 2010
AMBYSTOMA	Mallatt et al. 2010
AMPHITRITE	Mallatt et al. 2010
ANEMONIA	Mallatt et al. 2010
ANTIPATHES	Mallatt et al. 2010
ARION	Mallatt et al. 2010
ATOLLA	Mallatt et al. 2010
AXINELLA	Mallatt et al. 2010
CALLIBAETIS	Mallatt et al. 2010
CALLIPALLENE	Mallatt et al. 2010
CARYBDEA	Mallatt et al. 2010
CATOSTYLUS	Mallatt et al. 2010
CEREBRATULUS	Mallatt et al. 2010
CHAETOPLEURA	Mallatt et al. 2010
CHRYSAORA	Mallatt et al. 2010
CLYTIA	Mallatt et al. 2010
CRATEROMORPHA	Mallatt et al. 2010
CYANEA	Mallatt et al. 2010
EPHYDATIA	Mallatt et al. 2010
FABIENNA	Mallatt et al. 2010
GLYCERA	Mallatt et al. 2010
HALICLYSTUS	Mallatt et al. 2010
HOMARUS	Mallatt et al. 2010
HYDRA	Mallatt et al. 2010
HYDRACTINIA	Mallatt et al. 2010
ICHTHYOPHONUS	Mallatt et al. 2010
LUBOMIRSKIA	Mallatt et al. 2010
MELICERTISSA	Mallatt et al. 2010
METRIDIUM	Mallatt et al. 2010
MISUMENOPS	Mallatt et al. 2010
MONOSIGA	Mallatt et al. 2010
MONOSIGAOVATA	Mallatt et al. 2010
MONTASTREA	Mallatt et al. 2010
MUCOR	Mallatt et al. 2010
MYCALE	Mallatt et al. 2010
NAUSITHOE	Mallatt et al. 2010
NECTOPYRAMIS	Mallatt et al. 2010
NEMATOSTELLA	Mallatt et al. 2010
OOPSACAS	Mallatt et al. 2010
OSCARELLA	Mallatt et al. 2010
PACHYDICTYUM	Mallatt et al. 2010
PANTACHOGON	Mallatt et al. 2010
PETROMYZON	Mallatt et al. 2010
PHORONIS	Mallatt et al. 2010
PODOCORYNE	Mallatt et al. 2010
PODURA	Mallatt et al. 2010
PORPITA	Mallatt et al. 2010
PROTEROSPONGIA	Mallatt et al. 2010
RAJA	Mallatt et al. 2010
RHABDOCALYPTUS	Mallatt et al. 2010
RHIZAXINELLA	Mallatt et al. 2010
SACCHAROMYCES	Mallatt et al. 2010
SACCOGLOSSUS	Mallatt et al. 2010
SALPINGOECA	Mallatt et al. 2010
SCRIPPSIA	Mallatt et al. 2010
SCUTIGERA	Mallatt et al. 2010
STRONGYLOCENTROTUS	Mallatt et al. 2010
TENEBRIO	Mallatt et al. 2010
TETHYA	Mallatt et al. 2010
TRACHYCLADUS	Mallatt et al. 2010
TRICHOPLAX	Mallatt et al. 2010
TRIOPS	Mallatt et al. 2010
AMPHIMEDON	Mallatt et al. 2010
SUBERITES	Mallatt et al. 2010
SYCON	Mallatt et al. 2010
HETEROCHONE	Mallatt et al. 2010
LEUCETTA	Mallatt et al. 2010
LEUCOSOLENIA	Mallatt et al. 2010

Figure S1: Phylogeny inferred with RAxML and dataset Metazoan_full. Nodes have 100% BS unless otherwise noted.

Figure S2: Phylogeny inferred with RAxML and dataset Metazoan_LB_strict. Nodes have 100% BS unless otherwise noted.

Figure S3: Phylogeny inferred with RAxML and dataset Metazoan_LB_relaxed. Nodes have 100% BS unless otherwise noted.

Figure S4: Phylogeny inferred with RAxML and dataset Metazoan_RCFV_strict. Nodes have 100% BS unless otherwise noted.

Figure S5: Phylogeny inferred with RAxML and dataset Metazoan_RCFV_relaxed. Nodes have 100% BS unless otherwise noted.

Figure S6: Phylogeny inferred with RAxML and dataset Metazoan_RCFV_LB_strict. Nodes have 100% BS unless otherwise noted.

Figure S7: Phylogeny inferred with RAxML and dataset Metazoan_RCFV_LB_relaxed. Nodes have 100% BS unless otherwise noted.

Figure S8: Phylogeny inferred with RAxML and dataset Metazoan_Choano. Nodes have 100% BS unless otherwise noted.

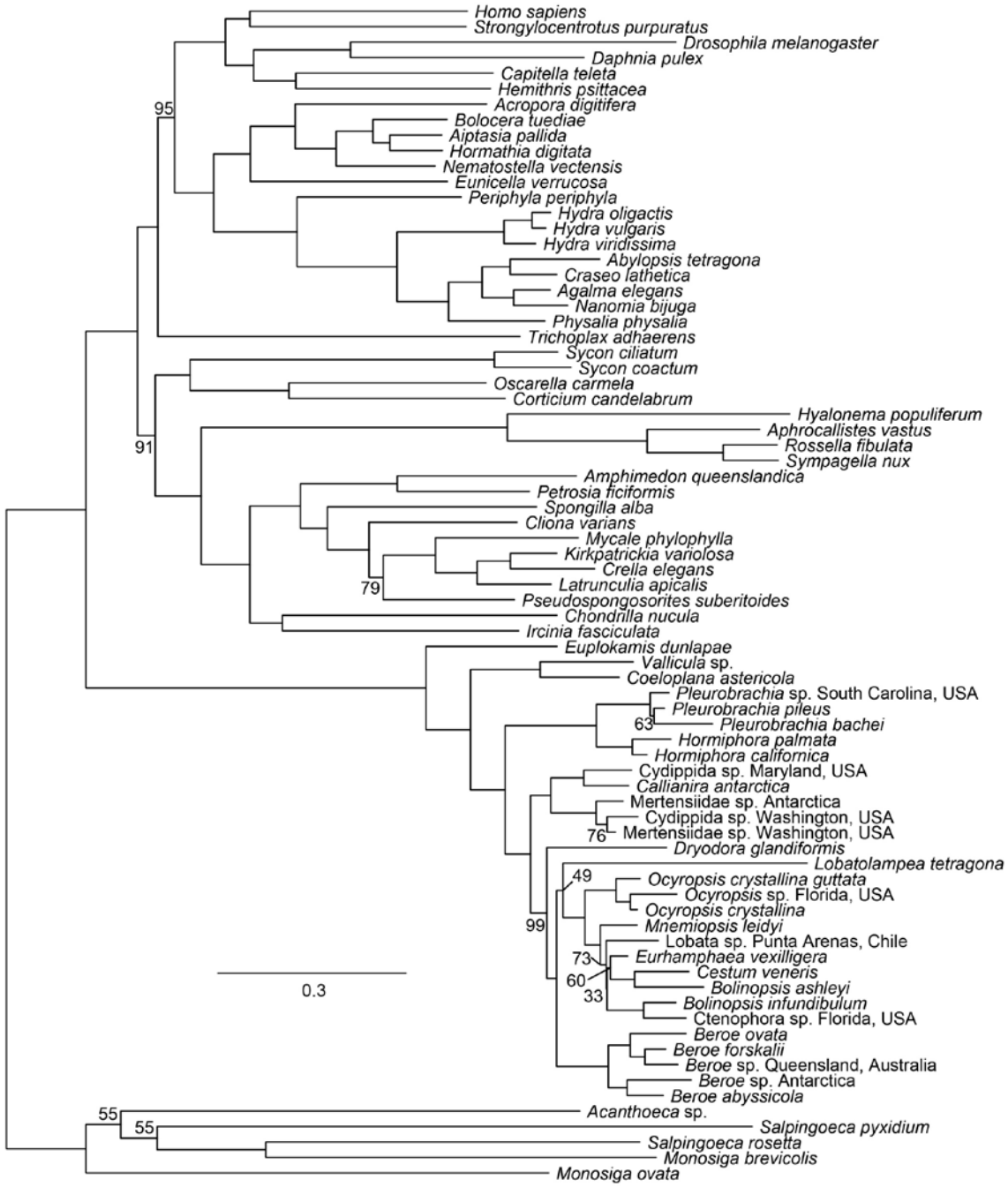


Figure S9: Phylogeny inferred with RAxML and dataset Metazoan_Choano_LB_strict. Nodes have 100% BS unless otherwise noted.

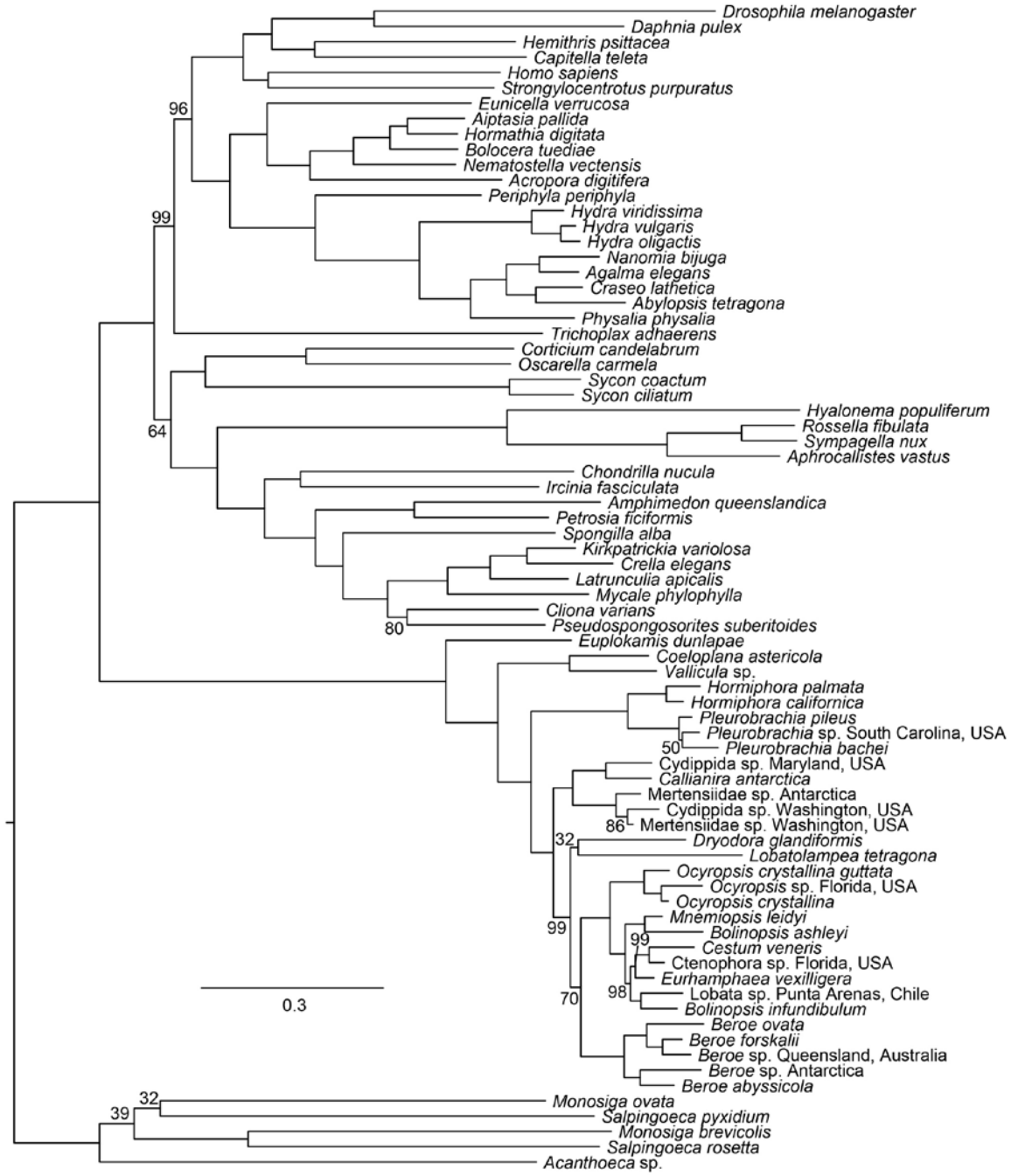


Figure S10: Phylogeny inferred with RAxML and dataset Metazoan_Choano_LB_relaxed. Nodes have 100% BS unless otherwise noted.

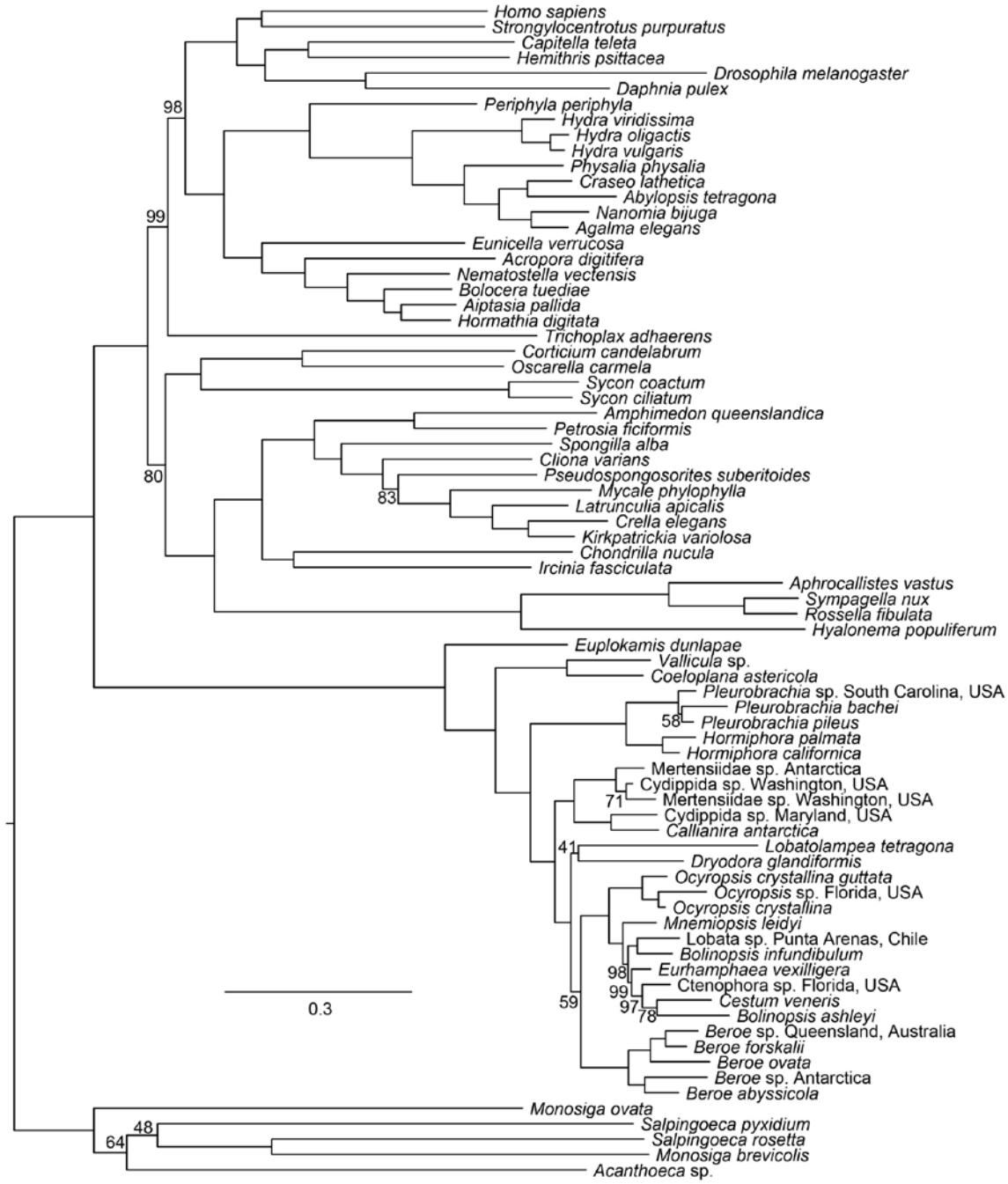


Figure S11: Phylogeny inferred with RAxML and dataset Metazoan_Choano_RCFV_strict. Nodes have 100% BS unless otherwise noted.

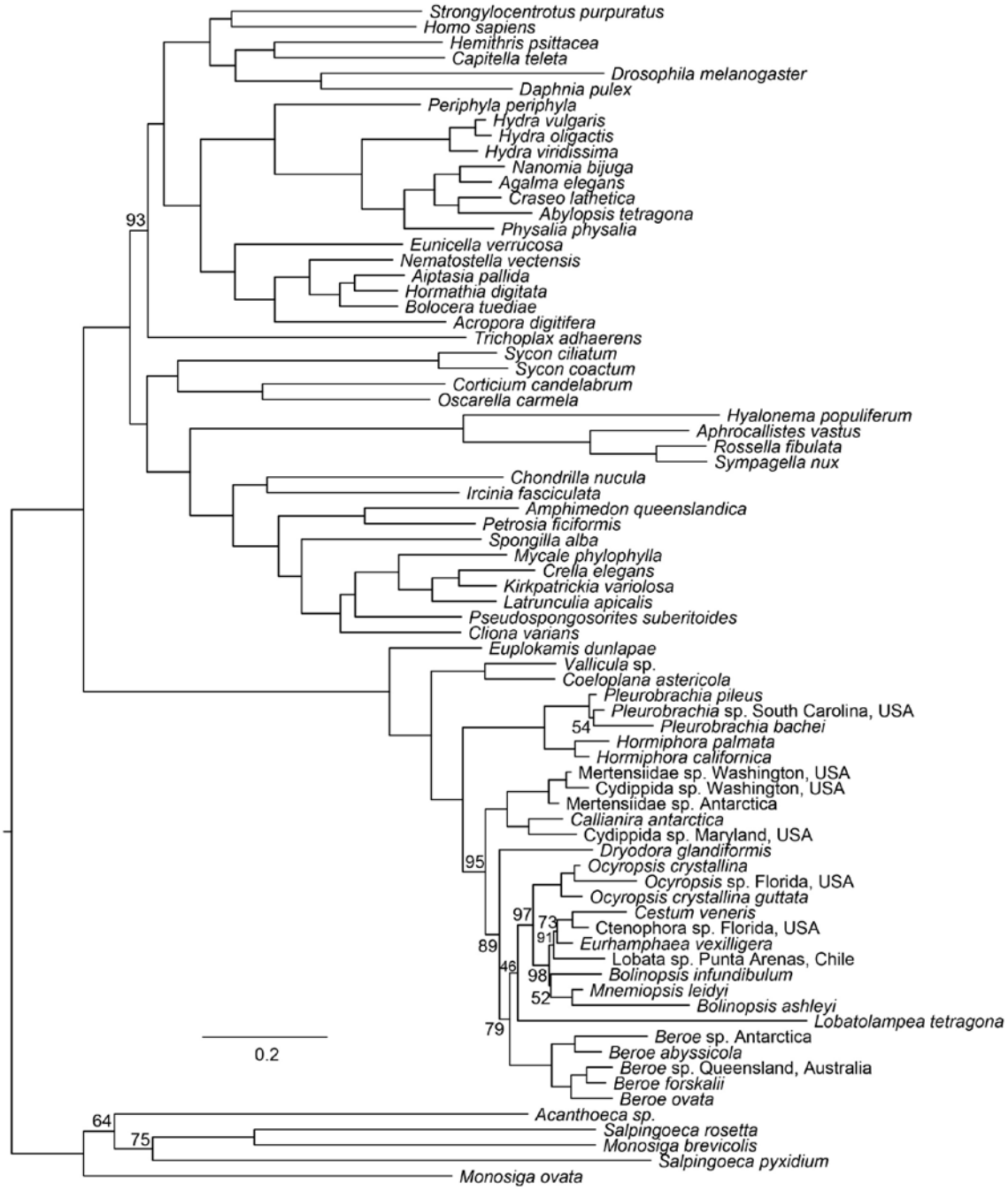


Figure S12: Phylogeny inferred with RAxML and dataset Metazoan_Choano_RCFV_relaxed. Nodes have 100% BS unless otherwise noted.

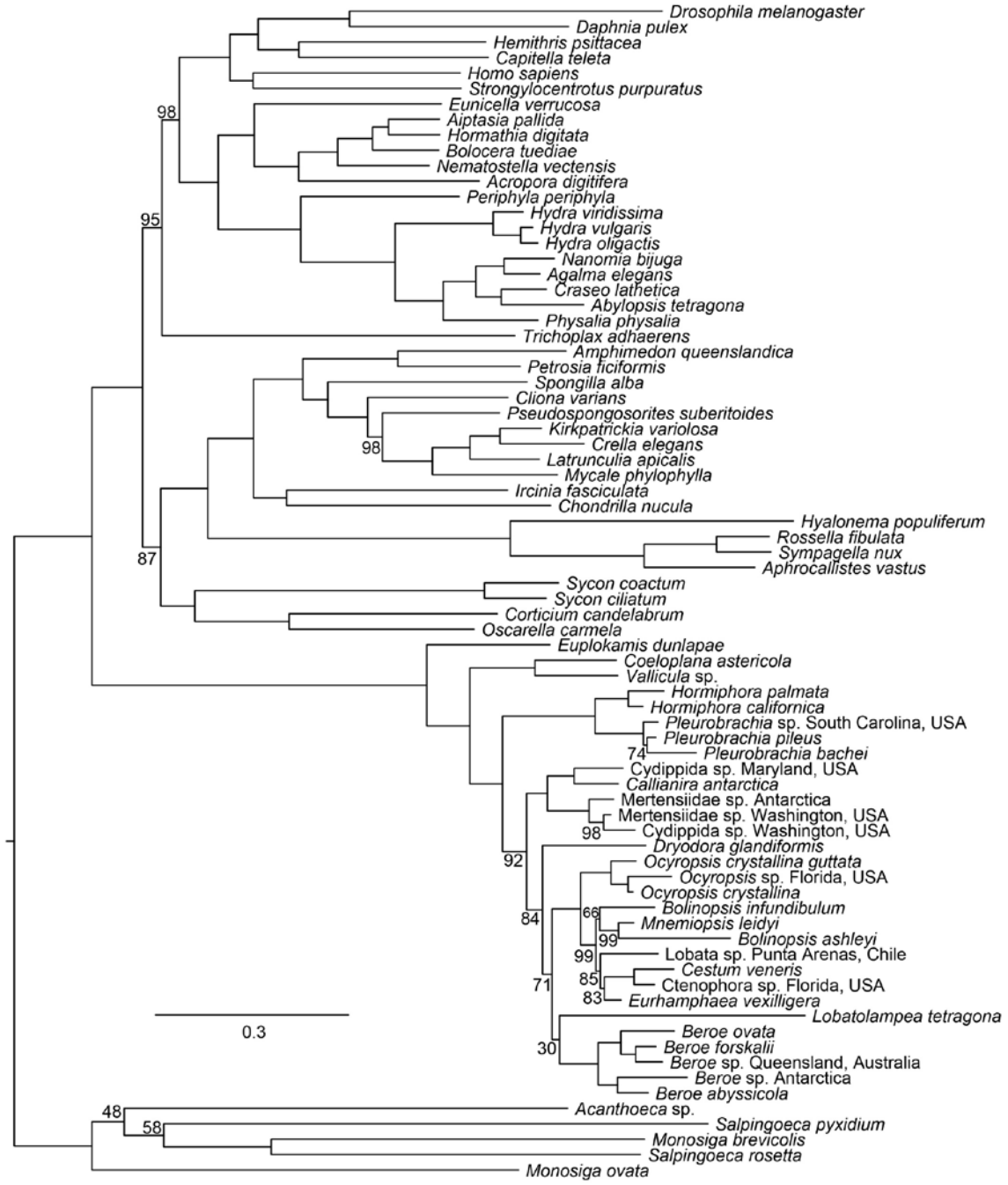
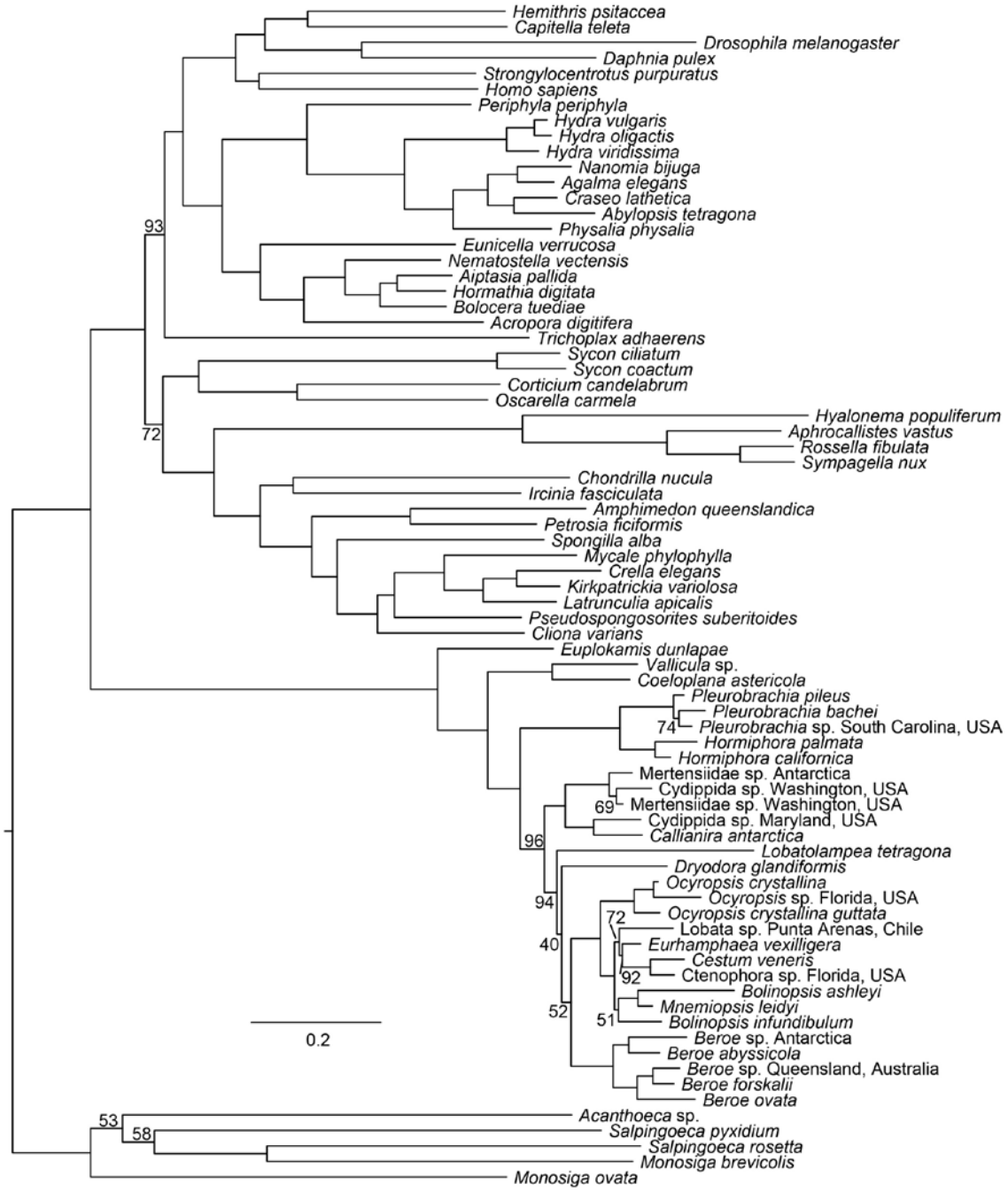
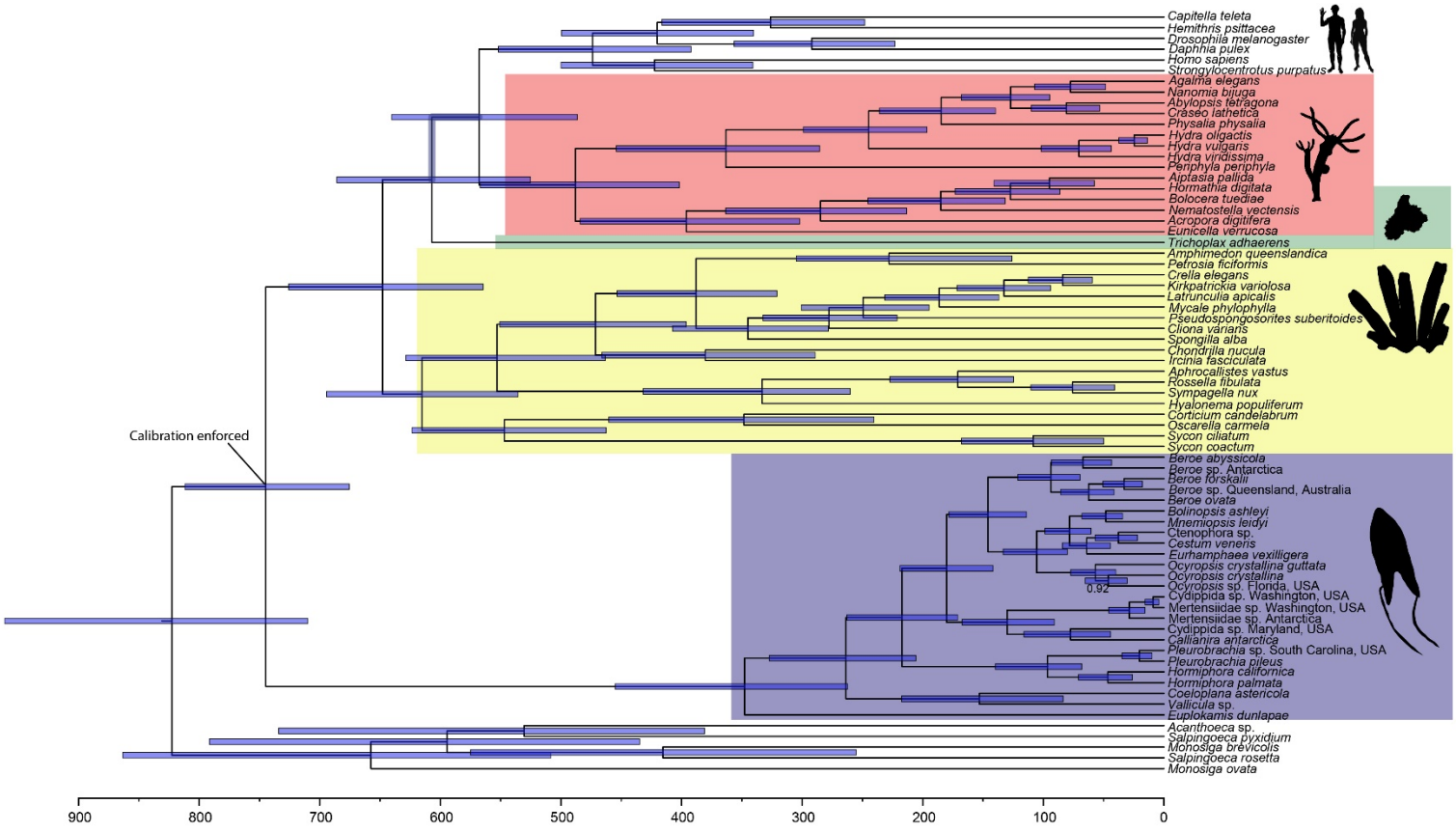


Figure S13: Phylogeny inferred with RAxML and dataset Metazoan_Choano_LB_RCFV_strict. Nodes have 100% BS unless otherwise noted.

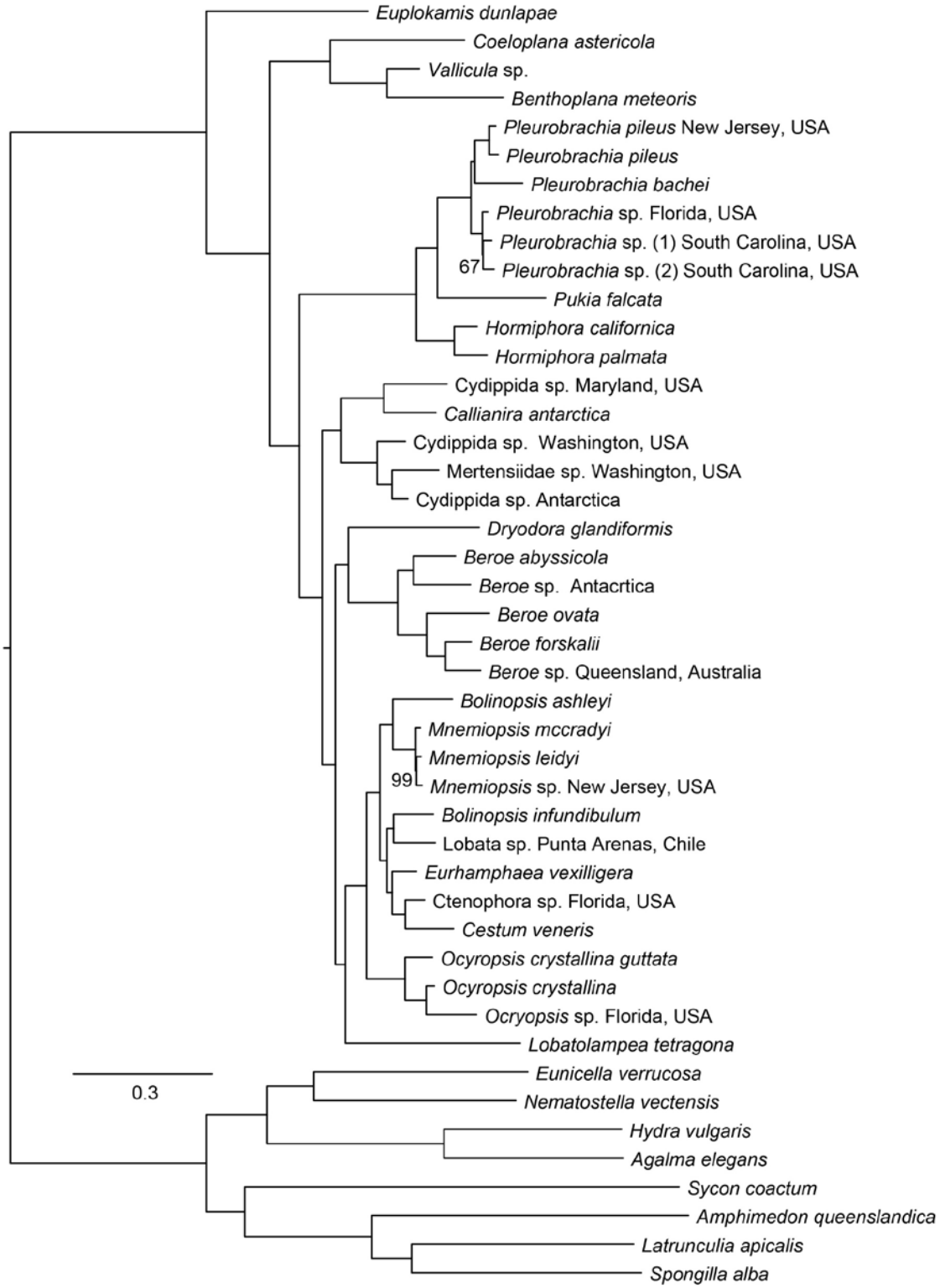
Figure S14: Phylogeny inferred with RAxML and dataset Metazoan_Choano_LB_RCFV_relaxed. Nodes have 100% BS unless otherwise noted.



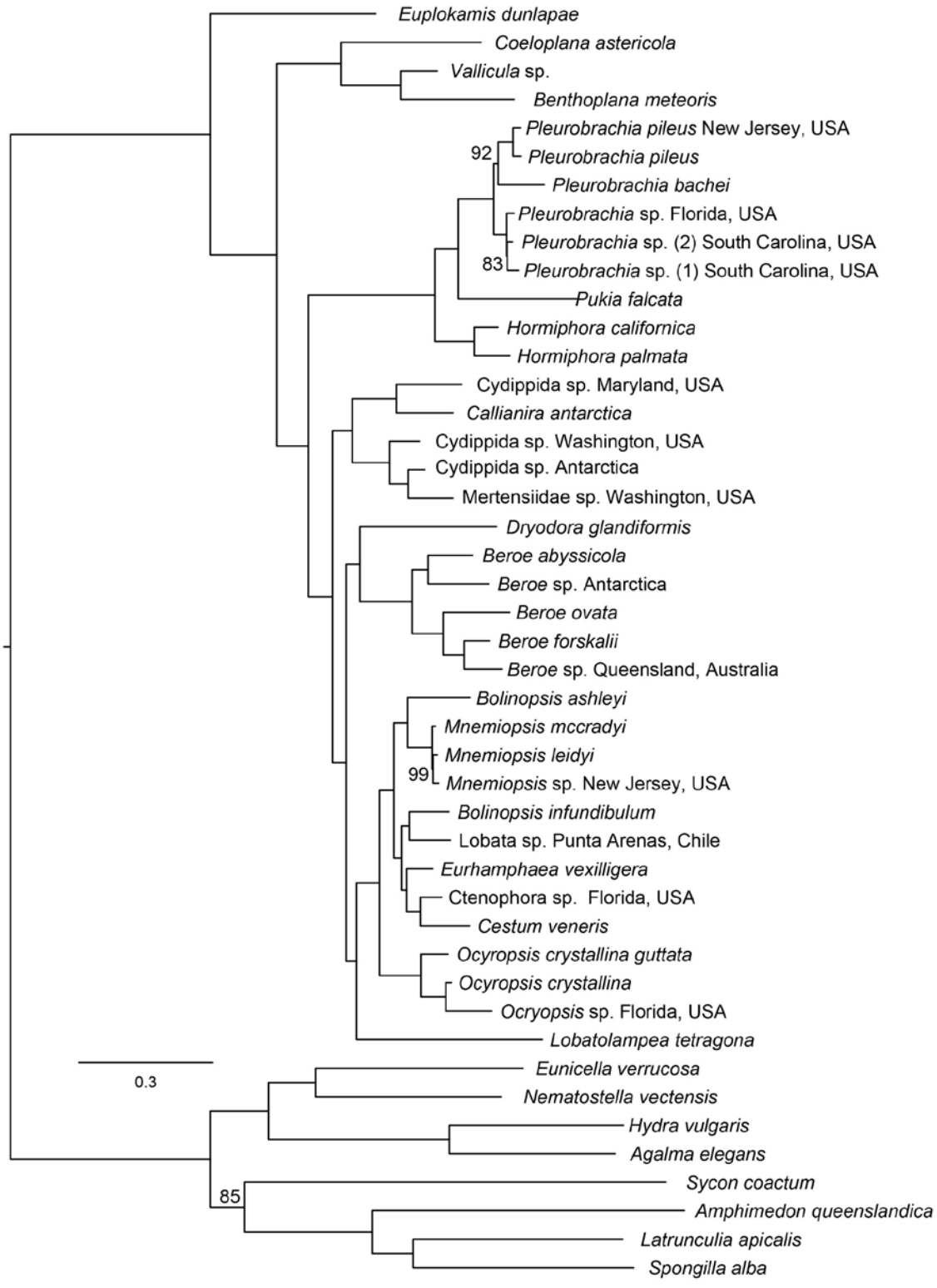
Supplementary Figure S15: Time-calibrated phylogeny inferred with BEAST2 and dataset metazoan_Choano_RCFV_strict in units of millions of years. Nodes have 1.00 PP unless otherwise noted. 95% confidence intervals of divergence time estimate are displayed on nodes.



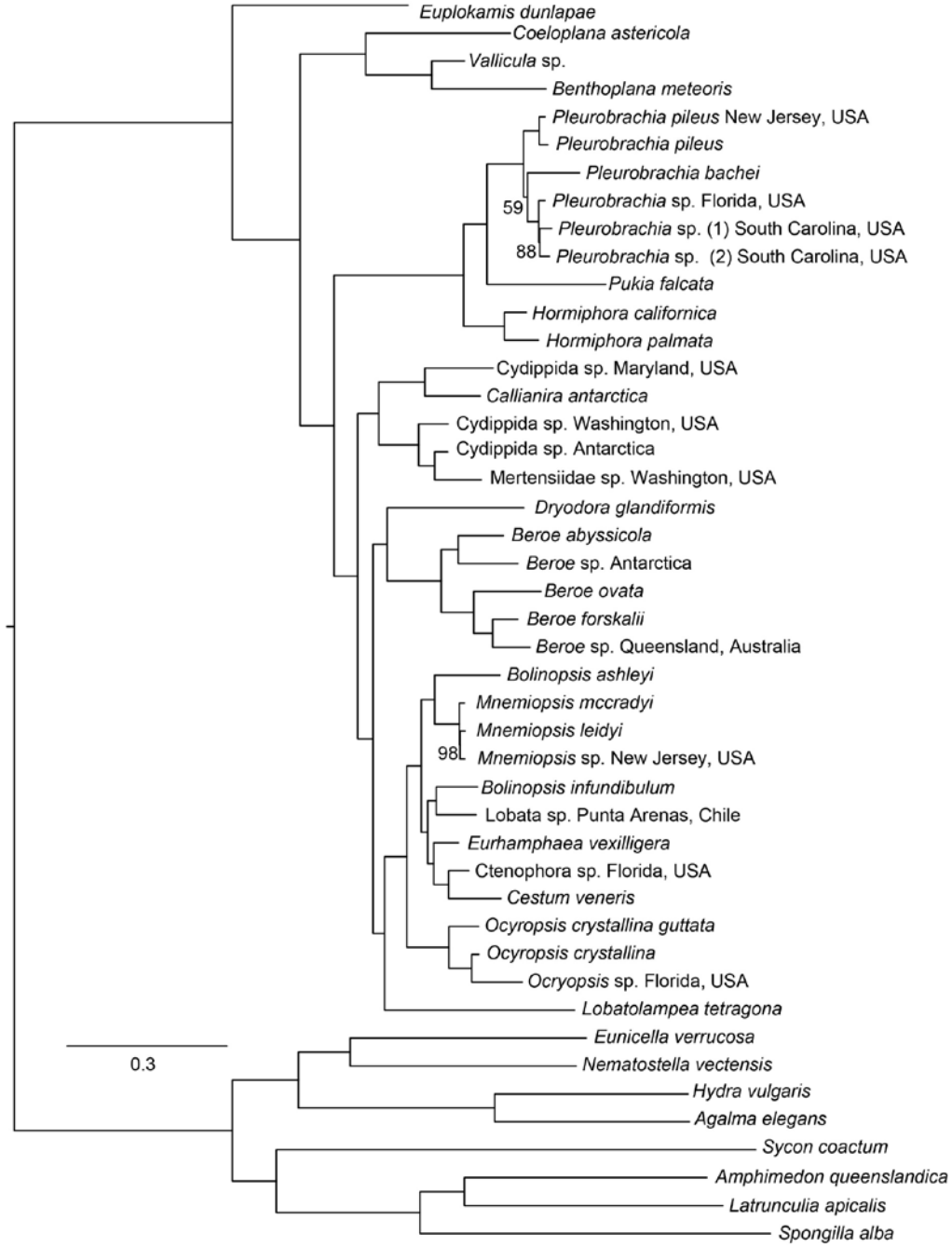
Supplementary Figure S16: Phylogeny inferred with RAxML and dataset Ctenophore_full. Nodes have 100% BS unless otherwise noted.



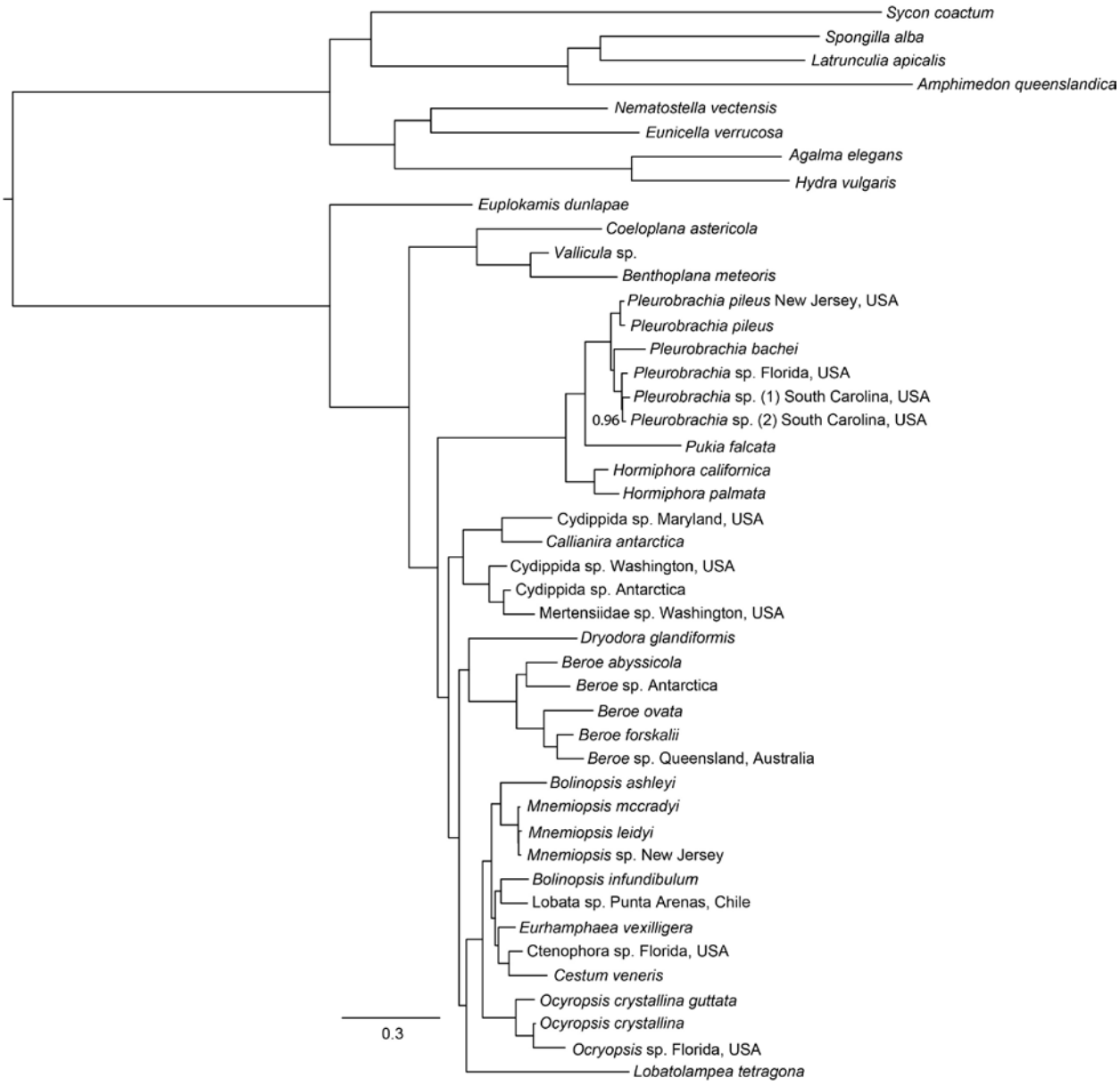
Supplementary Figure S17: Phylogeny inferred with RAxML and dataset Ctenophore_LB. Nodes have 100% BS unless otherwise noted.



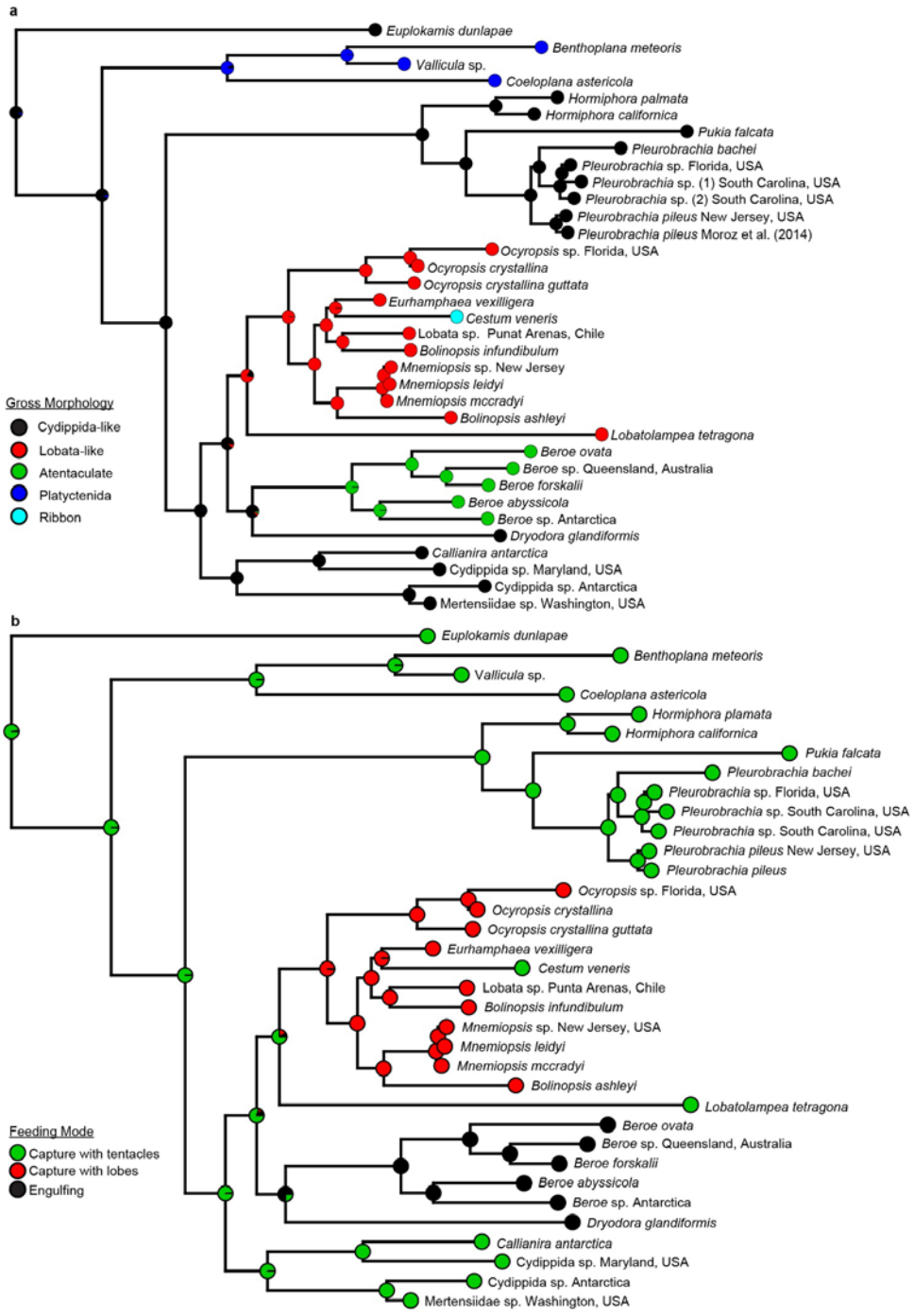
Supplementary Figure S18: Phylogeny inferred with RAxML and dataset Ctenophore_RCFV. Nodes have 100% BS unless otherwise noted.



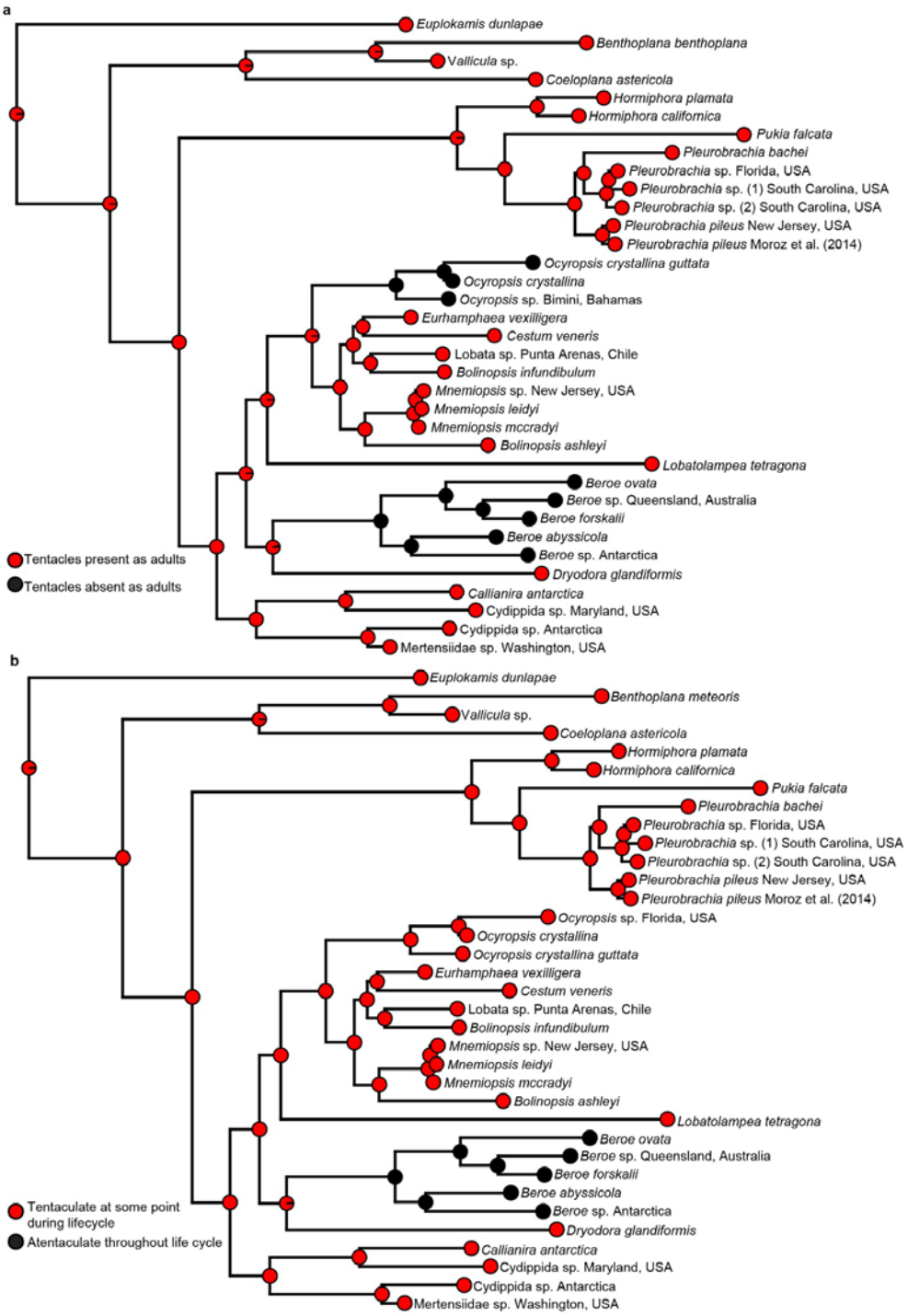
Supplementary Figure S19: Phylogeny inferred with PhyloBayes, the CAT-GTR substitution model and dataset Ctenophore_RCFV_LB. Nodes have 100% PP unless otherwise noted.



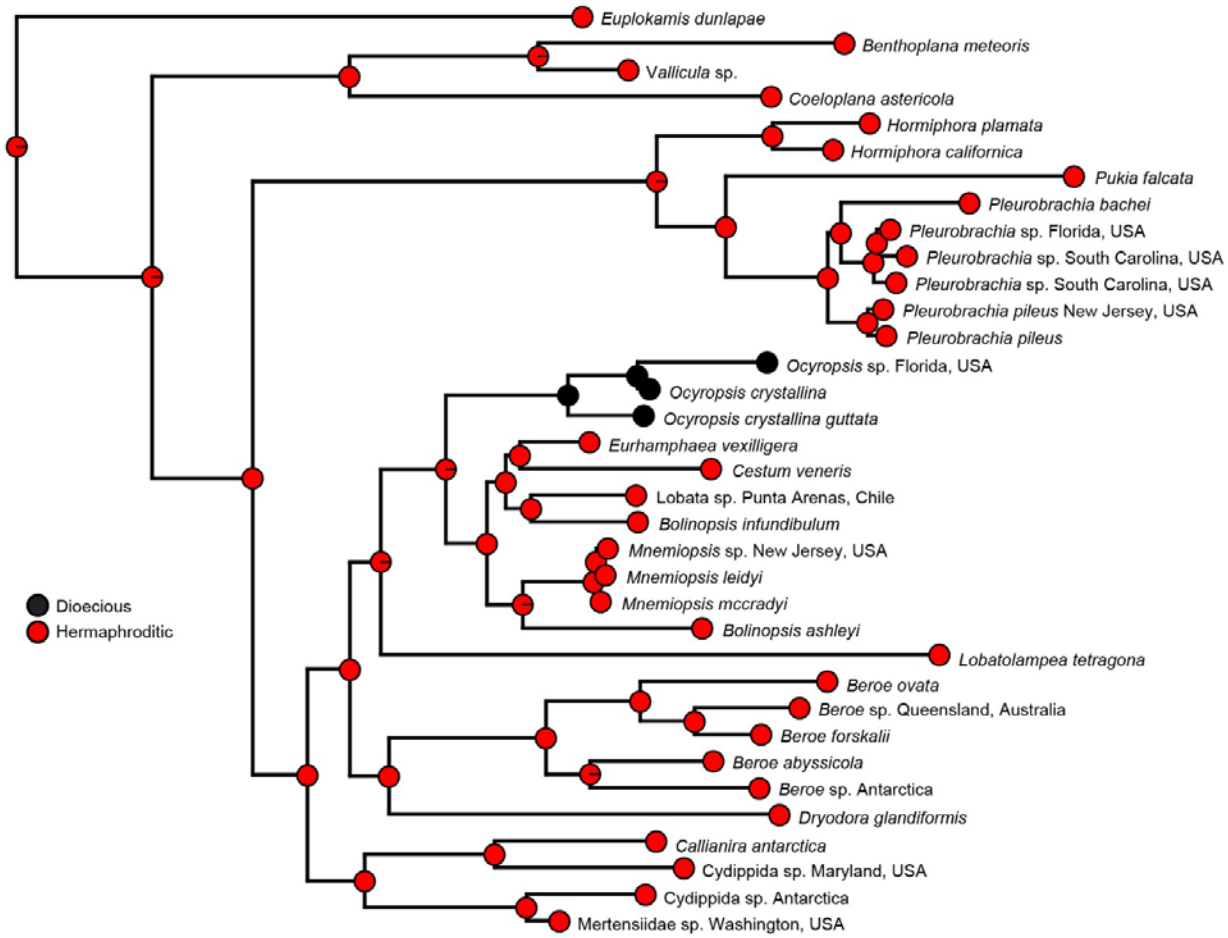
Supplementary Figure S20: Ancestral state reconstruction for a) general ctenophore body and b) primary feeding mode using phylogeny inferred with RAxML and dataset ctenophore RCFV_LB. Outgroups were not included in ancestral state reconstruction and are not figured. Nodes labeled with pie charts of posterior probability for ancestral state.



Supplementary Figure S21: Ancestral state reconstruction for a) presence of tentacles as adults and b) presence of tentacles at any life stage using phylogeny inferred with RAxML and dataset ctenophore RCFV_LB. Outgroups were not included in ancestral state reconstruction and are not figured. Nodes labeled with pie charts of posterior probability for ancestral state.

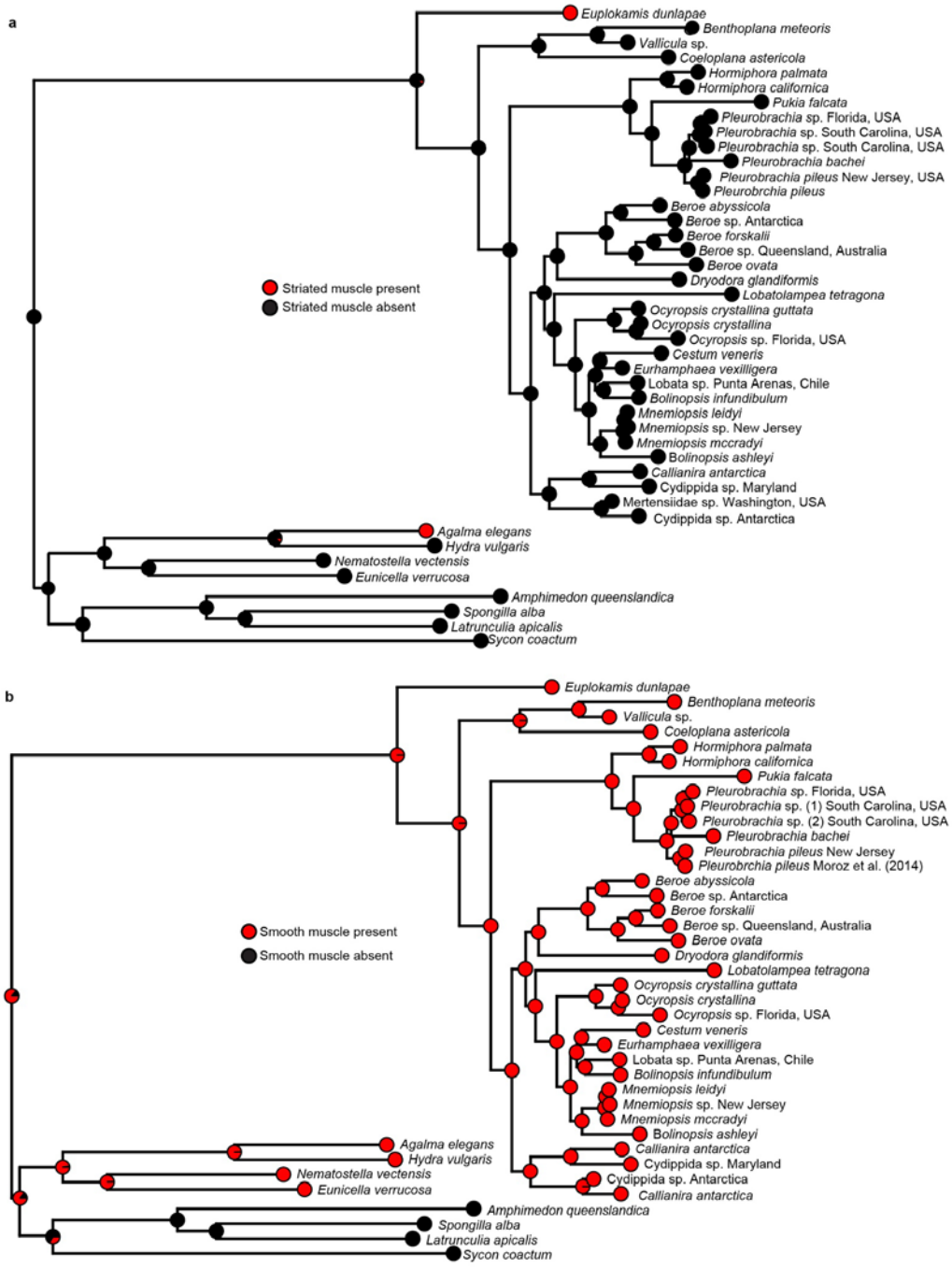


Supplementary Figure S22: Ancestral state reconstruction for whether species have separate sexes using phylogeny inferred with RAxML and dataset ctenophore RCFV_LB. Outgroups were not included in ancestral state reconstruction and are not figured. Nodes labeled with pie charts of posterior probability for ancestral state.

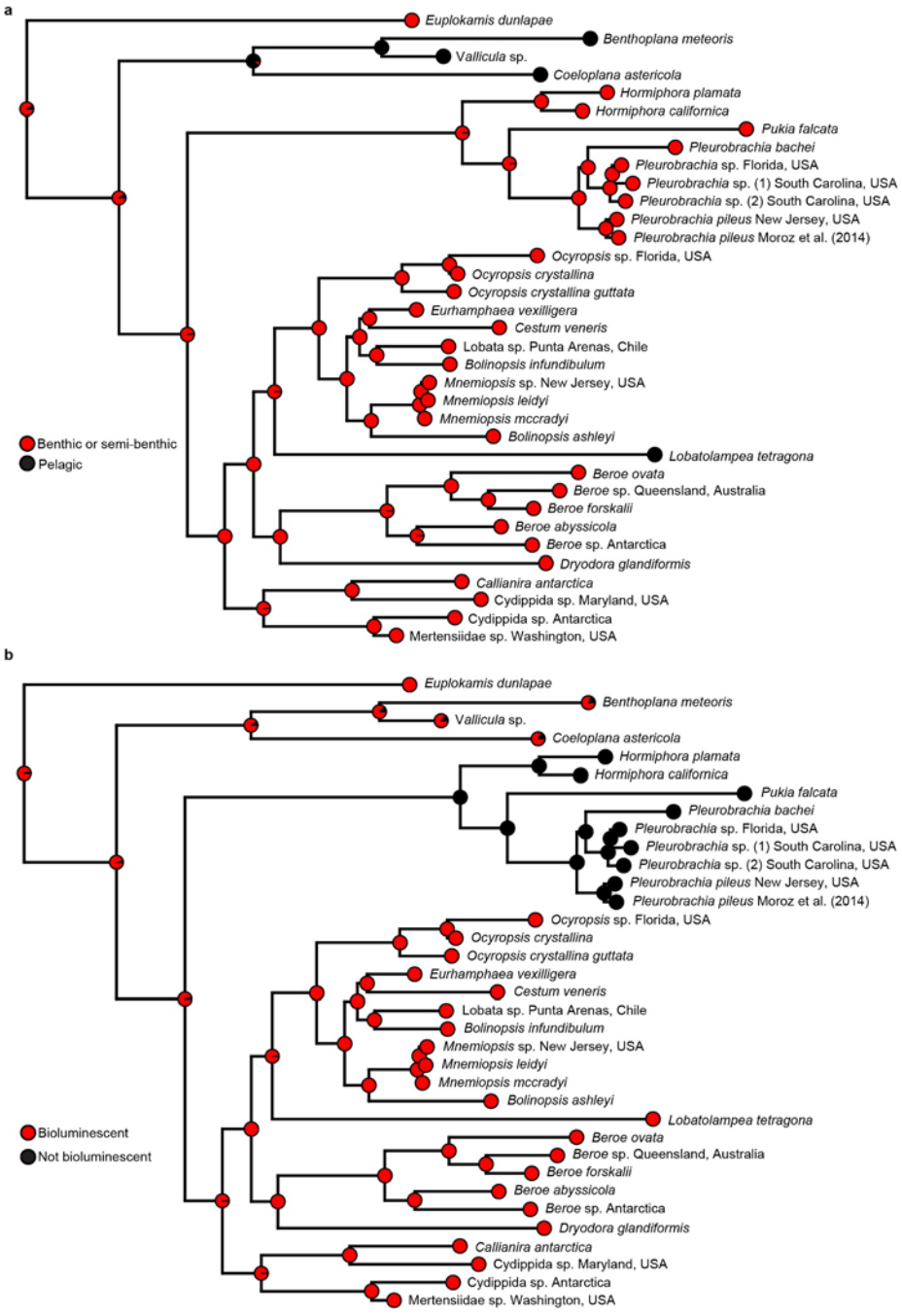


Supplementary Figure S23: Tree inferred with RAxML and 18S rRNA gene. Nodes with greater than BS values greater than 50 are labelled.

Supplementary Figure S24: Ancestral state reconstruction for a) presence of striated muscles and b) presence of smooth muscles using phylogeny inferred with RAxML and dataset ctenophore RCFV_LB. Outgroups were not included in ancestral state reconstruction and are not figured. Nodes labeled with pie charts of posterior probability for ancestral state.



Supplementary Figure S25: Ancestral state reconstruction for a) whether species were pelagic or benthic/semi-benthic and b) whether species have the ability to bioluminesce using phylogeny inferred with RAxML and dataset ctenophore RCFV_LB. Outgroups were not included in ancestral state reconstruction and are not figured. Nodes labeled with pie charts of posterior probability for ancestral state.



Supplementary Figure S26: Density plots of metrics indicating the degree to which OGs may cause systematic error. Shaded areas indicate genes that were removed to create certain datasets (see Extended Data Table 1). a) Dataset Metazoa_full. b) Dataset Metazoa_Choano. c) Dataset Ctenophore_full.