Supplemental Data

Phylogenomics Revives Traditional Views on Deep Animal Relationships

Herve Philippe, Romain Derelle, Philippe Lopez, Kerstin Pick, Carole Borchiellini, Nicole Boury-Esnault, Jean Vacelet, Emmanuelle Renard, Evelyn Houliston, Eric Queinnec, Corinne Da Silva, Patrick Wincker, Herve Le Guyader, Sally Leys, Daniel J. Jackson, Fabian Schreiber, Dirk Erpenbeck, Burkhard Morgenstern, Gert Worheide, and Michael Manuel

Supplemental Experimental Procedures

Data assembly

Twenty-three new orthologous genes were added to the 172 gene alignments used in previous studies [1-5]. These new markers were selected using the following procedure. Newly sequenced sponge ESTs were blasted against a reference bilaterian proteome (*Daphnia pulex* predicted proteins). Only genes showing significant match with ESTs from at least two different sponge groups (among Calcispongia, Demospongiae, Homoscleromorpha and Hexactinellida) were retained for further analyses. Then, sequences belonging to multigene families were pooled together, and genes from various representative metazoan species were added, for phylogenetic analyses. Gene families for which these analyses revealed ambiguous orthology relationships were discarded. After further elimination of the genes already present in our pre-existing 172 gene alignments, there were 23 potential new phylogenetic markers remaining.

These (172+23) alignments were updated with our new EST sequences as well as with newly available sequences downloaded from the Trace Archive (http://www.ncbi.nlm.nih.gov/Traces/) and the EST Database (http://www.ncbi.nlm.nih.gov/dbEST/) of GenBank at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) using the program forty (Denis Baurain, personal communication) that is based on blast similarity searches and new features of the program ED from the MUST package [6]. Unambiguous aligned regions were automatically detected and removed using the program GBlocks [7] and this selection was manually refined using the program ED. The list of genes with number of amino-acid positions and of species is reported in Table S1. The alignments are provided as Supplementary Information.
Gene selection and concatenation were performed thanks to the program SCAFOS [8]. Only the 128 genes that have been sequenced for more than 36 species were retained. For the concatenation of these genes, SCAFOS allows the selection of sequences according to their degree of divergence using the ML distance matrix computed under a WAG+F model by TREE-PUZZLE [9]. It also permits optimising the percentage of missing data per taxa by creating chimerical sequences for species belonging to the same taxonomic group (see list below). The resulting alignment of 128 genes and 55 species for 30,257 unambiguously aligned positions is provided as Supplementary Information (file CONCATENATION). In this supermatrix, ribosomal proteins represent 10,767 positions and non ribosomal proteins 19,490 positions.

Species names and list of chimerical Operational Taxonomic Units (OTUs)

Below are listed all species used as terminal taxa in this study (in the figures, only genus names are indicated). Species that have been newly sequenced for this study are indicated by an asterisk. In a few cases, to increase the amount of information, we created chimerical sequences between closely related taxa. These chimerical OTUs have been named by using the genus name of the most represented species (underlined genus names in the list below, with the most represented species in bold). Moreover, when several species were available for a given OTU, the slowest evolving sequence was selected, an approach shown to reduce the impact of long branch attraction artefact [8].:

Acropora: ***Acropora millepora**, Acropora palmata*
Allomyces macrogynus
Amoebidium parasiticum
Amphimedon queenslandica
Anoplodactylus eroticus
Aplysia californica
Batrachochytrium dendrobatidis
Branchiostoma: ***Branchiostoma floridae**, Branchiostoma belcheri, Branchiostoma lanceolatum*
Capitella sp.
Capsaspora owczarzaki
*Carteriospongia foliascens**
Ciona intestinalis
*Clytia hemisphaerica**
Crassostrea: ***Crassostrea gigas**, Crassostrea virginica*
Cyanea capillata
Danio rerio
Daphnia: ***Daphnia pulex**, Daphnia magna*
Ephydatia: ***Ephydatia muelleri**, Ephydatia fluviatilis*
Euperipatoides: ***Euperipatoides kanangrensis**, Epiperipatus sp., Peripatus sp.*
Euprymna: ***Euprymna scolopes**, Idiosepius paradoxus*
Helobdella: ***Helobdella robusta**, Haementeria depressa*
*Heterochone calyx**
Hydractinia echinata
Hydra magnipapillata
Ixodes scapularis
*Leucetta chagosensis**
Memniopsis leidyi
Metridium senile
Molgula tectiformis
Montastrea faveolata
Monosiga brevicollis
Monosiga ovata
Nasonia vitripennis
Nematostella vectensis
*Oopsacas minuta**
Oscarella: ***Oscarella carmela**, Oscarella lobularis**
Pedicellina cernua
Pediculus humanus
Petromyzon marinus
Phycomyces blakesleeanus
*Pleurobrachia pileus**
Podocoryne carnea
Proterospongia sp.
Rhizopus orizae
Saccoglossus kowalevskii
Scutigera coleoptrata
Sphaeroforma arctica

Spizellomyces punctatus
Strongylocentrotus purpuratus
Suberites: ***Suberites domuncula****, Suberites fuscus*
*Sycon raphanus**
Trichoplax adhaerens
Tubifex tubifex
Xenoturbella bocki

Missing data

The construction of a supermatrix containing a reasonable number of taxa unavoidably implies a certain amount of missing data (Table S2). In our concatenated dataset the number of amino acid residues available for the most incomplete species is nevertheless already large with 3,617 positions for *Oopsacas*. The complete dataset comprised 30,257 unambiguously aligned positions with a mean of 22,124 (73%) amino acid residues per taxa (Table S2). Among the 55 terminal taxa included in our study, 24 (44%) are complete or nearly complete (<= 5% of missing data), and 29 (53%) have > 80% of data.

Under these conditions, the impact of missing data on phylogenetic inference can be considered as negligible (see [1, 10, 11, 12]). Wiens & Moen [12] used simulations to demonstrate that phylogenetic accuracy in Bayesian analyses is almost unaffected even when 50% of the taxa have 95% of missing data (the remaining 50% of the taxa being complete), as long as the overall number of characters is large, near-maximal accuracy being obtained with 2,000 characters. Our data set contains a considerably larger number of characters (30,257), the number of complete taxa stands within the order of magnitude of 50%, and the remaining (incomplete) taxa have considerably less than 95% of missing data on average, so that based on the conclusions of Wiens & Moen (2008), we can reasonably assume that missing data will not impact phylogenetic accuracy. Indeed, the five species with less than 10,000 amino acids - *Oopsacas* (3617), *Cyanea* (7432), *Montastrea* (6584), *Carteriospongia* (8026), *Heterochone* (8663) – are all robustly located (see Fig. 1).

Orthology check

The 128 genes are in single-copy in most of the opisthokonts (few recent duplications are observed mainly in vertebrates and *Drosophila*). They are therefore likely orthologous. Yet, to further validate their orthology, we inferred single gene phylogenies using TreeFinder [13] with a WAG+Γ_8 model. To reduce stochastic errors that are important for single genes, we excluded species with more than 50% missing data (this is not a stringent criterion, since two species can have no overlap). Then, for each gene, we retained only bipartitions supported by a Bootstrap Proportion (BP) higher than 70% (called testable bipartitions) and tested if these partitions are congruent with the tree based on the concatenation. There are 1,229 testable bipartitions and only 80 conflicts (6.5%). This is less than the expected error rate (if we assume that bootstrap is not conservative, 30% is expected; this is still acceptable if we follow refs. [14-16] that suggest that a BP of 70% corresponds to an error rate of 5%, since we would then expect 5%). More importantly, the vast majority of conflicts (see Table S3) correspond to minor local rearrangements (i.e. Nearest Neighbor Interchange, or NNI) errors (58). The majority of the NNI errors correspond to cases where there is a long basal branch (i.e. *Allomyces*, *Mnemiopsis*, and *Clytia*), cases that are expected to be prone to artefacts. There are also 5 quasi-NNI (i.e. a

move of two nodes instead of a single to resolve the conflict) and 6 cases of obvious LBA. Therefore, it remains only 11 conflicts for which there are no obvious explanations related to phylogenetic reconstruction errors. Yet, these eleven conflicts cannot be easily explained by paralogy. This analysis strongly suggests that single genes are in good agreement with the concatenation and therefore that paralogy does not play a significant negative role in our inference.

Phylogenetic analyses

PhyloBayes analyses were performed with the CAT+Γ_4 mixture model, which accounts for across-site heterogeneities in the amino-acid replacement process [17]. This model is implemented in a MCMC framework by the program PHYLOBAYES version 2.3 (http://www.lirmm.fr/mab/). Two independent runs were performed with a total length of 15,000 cycles (250 topological moves per cycle) with the same operators as in Lartillot et al. [18]. The first 5,000 points were discarded as burn-in, and the posterior consensus was computed on the 10,000 remaining trees.
We applied a standard bootstrap procedure [19]: 100 pseudo-replicates were generated with SEQBOOT [20]; each dataset was analysed with Phylobayes, trees were collected after the initial burn-in period and a consensus tree was computed by phylobayes; finally, a consensus tree was inferred from these 100 consensus trees using CONSENSE to compute the bootstrap support values for each node. For computing time reason, we performed only 10,000 cycles after verifying that this value is sufficient for the complete dataset. In addition, we used a conservative burn-in of 5,000 (manual verification of a few replicates indicates that the burn-in is generally less than 3,000).

References

1. Philippe, H., Snell, E.A., Bapteste, E., Lopez, P., Holland, P.W., and Casane, D. (2004). Phylogenomics of eukaryotes: impact of missing data on large alignments. Mol Biol Evol *21*, 1740-1752.
2. Philippe, H., Lartillot, N., and Brinkmann, H. (2005). Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. Mol Biol Evol *22*, 1246-1253.
3. Delsuc, F., Brinkmann, H., Chourrout, D., and Philippe, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature *439*, 965-968.
4. Baurain, D., Brinkmann, H., and Philippe, H. (2007). Lack of resolution in the animal phylogeny: closely spaced cladogeneses or undetected systematic errors? Mol Biol Evol *24*, 6-9.
5. Lartillot, N., and Philippe, H. (2008). Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. Philos Trans R Soc Lond B Biol Sci *363*, 1463–1472.
6. Philippe, H. (1993). MUST, a computer package of Management Utilities for Sequences and Trees. Nucleic Acids Res. *21*, 5264-5272.
7. Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol *17*, 540-552.
8. Roure, B., Rodriguez-Ezpeleta, N., and Philippe, H. (2007). SCaFoS: a tool for Selection, Concatenation and Fusion of Sequences for phylogenomics. BMC Evol Biol *7 Suppl 1*, S2.
9. Schmidt, H.A., Strimmer, K., Vingron, M., and von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics *18*, 502-504.

10. Wiens, J.J. (2003). Missing data, incomplete taxa, and phylogenetic accuracy. Syst Biol *52*, 528-538.
11. Wiens, J.J. (2005). Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? Syst Biol *54*, 731-742.
12 Wiens, J.J. and Moen, D.S. (2008). Missing data and the accuracy of Bayesian phylogenetics. J Syst Evol *46*, 307-314.
13. Jobb, G., von Haeseler, A., and Strimmer, K. (2004). TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol. Biol. *4*, 18.
14. Hillis, D.M., and Bull, J.J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst biol *42*, 182-192.
15. Zharkikh, A., and Li, W.H. (1992). Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. Mol Biol Evol *9*, 1119-1147.
16. Zharkikh, A., and Li, W.H. (1992). Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences: II. Four taxa without a molecular clock. J Mol Evol *35*, 356-366.
17. Lartillot, N., and Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Molecular Biology and Evolution *21*, 1095-1109.
18. Lartillot, N., Brinkmann, H., and Philippe, H. (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol *7 Suppl 1*, S4.
19. Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783-791.
20. Felsenstein, J. (2001). PHYLIP (Phylogene Inference Package). 3.6 Edition. (Distributed by the author, Department of Genetics, University of Washington, Seattle).

Table S1. List of the 128 genes used. New markers with respect to previous studies [1-5] are indicated in italics.

Gene	#positions	#OTUs
ar21	136	41
arc20	165	36
arp23	218	45
atpsynthalpha-a-mt	496	46
cct-A	504	38
cct-B	499	41
cct-D	479	43
cct-E	523	38
cct-G	513	39
cct-N	506	38
cct-T	451	38
cct-Z	501	41
cpn60-mt	508	40
crfg	412	36
ef2-EF2	777	48
eif5a	119	44
fibri	225	36
grc5	209	51
hsp70-E	583	44
hsp70-mt	569	36
hsp90-C	630	44
if1a	117	38
if2b	159	36
if2g	437	39
if4a-a	384	42
if4a-b	363	46
l12e-A	118	46
l12e-C	123	40
l12e-D	235	50
nsf1-G	383	36
nsf1-I	404	39
nsf1-J	385	38
nsf1-K	393	36
nsf1-M	411	38
nsf2-A	732	39
ornamtrans-a	355	37
psma-A	212	37
psma-B	235	36
psma-C	211	40
psma-D	218	43
psma-E	208	42
psma-F	221	37
psma-G	231	39
psmb-H	182	37
psmb-I	203	37
psmb-J	196	41
psmb-K	234	38
psmb-L	201	37
psmb-M	199	40
psmb-N	182	41
rpl1	213	51
rpl11b	169	52
rpl12b	163	51

rpl13	176	51
rpl14a	108	46
rpl15a	204	48
rpl16b	173	54
rpl17	155	51
rpl18	165	51
rpl19a	180	49
rpl2	248	52
rpl20	156	49
rpl21	154	50
rpl22	87	47
rpl23a	129	47
rpl24-A	117	51
rpl24-B	127	37
rpl25	119	50
rpl26	121	51
rpl27	135	51
rpl3	371	52
rpl30	105	47
rpl31	108	48
rpl32	129	50
rpl33a	104	50
rpl34	104	47
rpl35	119	49
rpl36	81	47
rpl37a	85	50
rpl38	63	38
rpl39	51	40
rpl42	103	49
rpl43b	89	48
rpl4B	297	52
rpl5	248	52
rpl6	140	52
rpl7-A	204	53
rpl9	167	51
rpp0	270	50
rps1	231	53
rps10	91	47
rps11	135	52
rps13a	151	49
rps14	134	52
rps15	138	49
rps16	137	51
rps17	107	48
rps18	152	51
rps19	130	51
rps2	214	52
rps20	105	50
rps22a	129	51
rps23	143	49
rps24	118	51
rps25	79	48
rps26	100	49
rps27	82	48
rps27a	58	54
rps28a	60	47
rps29	55	44
rps3	213	50

rps4	253	52
rps5	188	50
rps6	213	50
rps7	169	50
rps8	185	54
rps9	171	52
sadhchydrolase-E1	411	47
sap40	204	49
srs	367	36
stbproptase2a-b	299	38
suca	290	39
tif2a	242	40
vacaatpasepl21-a	144	37
vata	557	37
vatb	474	39
vate	195	36
vdac2	241	46

Table S2. Summary of the occurrence of missing data per taxa in the complete dataset (x = missing).

OTU	ar21.ali	arc20.ali	arp23.ali	atpsynthalpha-a-mt.a	cct-A.ali	cct-B.ali	cct-D.ali	cct-E.ali	cct-G.ali	cct-N.ali	cct-T.ali	cct-Z.ali	cpn60-mt.ali	crfg.ali	ef2-EF2.ali	eif5a.ali	fibri.ali	grc5.ali	hsp70-E.ali	hsp70-mt.ali	hsp90-C.ali	if1a.ali	if2b.ali	if2g.ali	if4a-a.ali	if4a-b.ali	l12e-A.ali	l12e-C.ali	l12e-D.ali	nsf1-G.ali	nsf1-I.ali	nsf1-J.ali	nsf1-K.ali	nsf1-M.ali	nsf2-A.ali	ornamtrans-a.ali	psma-A.ali	psma-B.ali	psma-C.ali
Acropora					x																		x										x					
Allomyces																																							
Amoebidium	x	x	x		x	x		x	x	x	x	x	x	x		x	x			x	x			x		x	x	x	x		x		x		x		x	
Amphimedon																																							
Anoplodactylus	x	x	x			x	x	x	x	x	x	x	x	x	x		x		x	x	x		x	x					x	x	x	x	x	x	x	x	x	x
Aplysia																																							
Batrachochytrium																																							
Branchiostoma																																							
Capitella																																							
Capsaspora																																							
Carteriospongia	x	x			x	x		x	x	x	x	x	x	x			x	x	x	x		x	x	x	x	x		x	x	x	x	x			x	x	x	x	
Ciona																																							
Clytia													x			x	x					x	x				x	x				x			x	x	x	x	x
Crassostrea																	x																						
Cyanea		x	x	x	x	x	x	x	x	x	x		x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Danio																																							
Daphnia																																							
Ephydatia	x	x		x	x		x			x	x	x	x			x	x		x	x		x	x	x			x		x		x	x	x			x	x	x
Euperipatoides	x	x			x	x	x	x	x	x	x		x	x			x			x	x	x	x	x	x	x	x	x		x	x		x	x	x	x	x	x
Euprymna																												x											
Helobdella																																							
Heterochone	x	x			x								x	x		x		x				x	x		x		x	x			x	x			x	x		
Hydra																																							
Hydractinia	x	x				x		x		x			x							x		x	x							x			x						x
Ixodes																																							
Leucetta					x	x	x	x	x	x	x	x	x	x	x		x		x	x				x	x						x	x	x	x	x	x	x	x	
Mertensiid	x	x	x		x	x	x	x	x	x		x		x	x		x		x	x	x	x	x	x					x	x	x	x	x	x	x	x	x	x
Metridium					x					x							x			x			x	x			x				x	x	x	x		x			x
Mnemiopsis										x									x	x	x				x						x			x	x				
Molgula																x																					x			
Monosiga																																							
Monosiga																																							
Montastraea	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x				x	x	x	x	x	x	x	x	x	x
Nasonia																																							
Nematostella																																							
Oopsacas	x	x		x	x	x	x	x		x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x
Oscarella		x		x					x		x			x									x	x			x	x	x	x							x	x
Pedicellina	x	x	x			x		x	x		x	x	x	x			x		x	x				x				x		x		x	x		x	x	x	x	x
Pediculus																																							
Petromyzon											x	x										x				x													
Phycomyces																																							
Pleurobrachia		x																			x	x					x												
Podocoryne	x	x						x	x					x			x			x		x	x	x						x	x			x	x	x	x		x	
Proterospongia		x	x	x			x	x	x	x		x		x		x			x	x	x	x	x	x	x		x			x	x					x	x	x	
Rhizopus																																							
Saccoglossus																																							
Scutigera	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x			x	x		x		x		x			x	x	x	x	x	x	x	x	x	x
Sphaeroforma	x				x			x		x			x									x		x					x			x	x					x
Spizellomyces																																							
Strongylocentrotus																																							
Suberites			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			x		x	x		x	x		x		x	x	x	x	x	x	x	x	x	x
Sycon		x	x	x	x	x	x	x		x	x	x		x		x	x			x		x		x	x		x		x	x	x	x	x	x	x	x	x	
Trichoplax																																							
Tubifex		x							x		x																			x						x			
Xenoturbella					x			x		x						x	x						x		x			x		x	x	x	x	x	x		x	x	

Table S2. Continued

OTU	psma-D.ali	psma-E.ali	psma-F.ali	psma-G.ali	psmb-H.ali	psmb-I.ali	psmb-J.ali	psmb-K.ali	psmb-L.ali	psmb-M.ali	psmb-N.ali	rpl1.ali	rpl11b.ali	rpl12b.ali	rpl13.ali	rpl14a.ali	rpl15a.ali	rpl16b.ali	rpl17.ali	rpl18.ali	rpl19a.ali	rpl2.ali	rpl20.ali	rpl21.ali	rpl22.ali	rpl23a.ali	rpl24-A.ali	rpl24-B.ali	rpl25.ali	rpl26.ali	rpl27.ali	rpl3.ali	rpl30.ali	rpl31.ali	rpl32.ali	rpl33a.ali	rpl34.ali	rpl35.ali	rpl36.ali
Acropora			X		X				X		X																						X					
Allomyces																																							
Amoebidium	X	X				X	X	X	X		X								X	X	X			X			X	X									X	X	
Amphimedon																																							
Anoplodactylus			X			X		X	X	X	X																	X											
Aplysia																																							
Batrachochytrium																																							
Branchiostoma																																							
Capitella																																							
Capsaspora																																							
Carteriospongia	X	X	X	X		X	X	X	X	X	X						X		X		X		X		X	X		X			X		X	X			X		X
Ciona																																							
Clytia			X		X		X					X	X	X		X								X	X	X		X		X			X	X	X	X	X	X	X
Crassostrea																																							
Cyanea	X	X		X	X	X		X	X	X	X										X	X			X													
Danio																																					X		
Daphnia																																							
Ephydatia		X	X	X	X	X	X		X	X	X					X	X			X								X	X										
Euperipatoides		X	X	X	X	X	X	X	X	X						X										X		X	X										
Euprymna					X											X																								
Helobdella																																					X		
Heterochone				X	X	X	X		X		X	X	X	X	X	X	X	X	X	X			X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X
Hydra																																							
Hydractinia	X		X	X			X	X				X																											
Ixodes																																							
Leucetta			X	X			X	X		X																		X											
Mertensiid	X	X	X	X	X	X	X		X	X	X																												
Metridium		X			X	X		X		X																								X		X			
Mnemiopsis				X																																			
Molgula																X																							
Monosiga																																							
Monosiga			X						X																			X											
Montastraea	X	X	X	X	X	X	X	X	X	X	X	X			X		X						X						X				X	X				X
Nasonia																																							
Nematostella																																							
Oopsacas			X	X	X	X			X	X			X	X	X		X		X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X		X	X
Oscarella		X	X	X	X	X	X			X					X	X	X				X					X	X	X		X	X		X	X	X				X
Pedicellina		X	X	X	X	X		X	X		X												X			X		X											X
Pediculus																																							
Petromyzon																																							
Phycomyces																																							
Pleurobrachia	X		X													X									X	X		X									X		
Podocoryne	X							X																				X											
Proterospongia	X	X		X	X			X	X	X						X	X					X	X	X	X			X	X			X	X		X	X		X	X
Rhizopus																																							
Saccoglossus																																							
Scutigera	X	X			X	X		X	X	X	X										X							X			X								
Sphaeroforma			X				X	X																															
Spizellomyces																																							
Strongylocentrotus																																							
Suberites	X	X	X	X	X	X	X	X	X	X	X																												
Sycon				X	X	X	X	X		X	X			X														X									X		
Trichoplax																																							
Tubifex					X				X																			X											
Xenoturbella	X		X		X	X		X	X		X														X												X		

Table S2. Continued

OTU	rpl37a.ali	rpl38.ali	rpl39.ali	rpl42.ali	rpl43b.ali	rpl4B.ali	rpl5.ali	rpl6.ali	rpl7-A.ali	rpl9.ali	rpp0.ali	rps1.ali	rps10.ali	rps11.ali	rps13a.ali	rps14.ali	rps15.ali	rps16.ali	rps17.ali	rps18.ali	rps19.ali	rps2.ali	rps20.ali	rps22a.ali	rps23.ali	rps24.ali	rps25.ali	rps26.ali	rps27.ali	rps27a.ali	rps28a.ali	rps29.ali	rps3.ali	rps4.ali	rps5.ali	rps6.ali	rps7.ali	rps8.ali	rps9.ali
Acropora																																							
Allomyces																																						
Amoebidium		X						X			X		X						X										X			X						
Amphimedon																																						
Anoplodactylus																																						
Aplysia																																						
Batrachochytrium			X																											X	X						
Branchiostoma																																						
Capitella																																						
Capsaspora																																						
Carteriospongia		X	X										X				X							X			X		X		X	X	X					
Ciona																																						
Clytia		X	X	X	X					X			X		X		X		X	X			X	X	X	X	X	X	X		X	X				X		
Crassostrea		X																																				
Cyanea						X						X	X			X		X				X						X				X			X	X			X
Danio																												X										
Daphnia																																						
Ephydatia			X	X									X						X						X	X		X			X		X					
Euperipatoides					X																														X			
Euprymna		X	X	X	X																										X					X		
Helobdella																																						
Heterochone	X	X	X	X	X		X			X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X			X
Hydra																																						
Hydractinia																																						
Ixodes																																						
Leucetta	X		X																												X							
Mertensiid																																		X				
Metridium		X																																				
Mnemiopsis																																						
Molgula																																						
Monosiga																																						
Monosiga		X	X																											X								
Montastraea		X			X	X		X	X		X			X	X		X								X			X		X		X				X		X
Nasonia		X																																				
Nematostella																																						
Oopsacas	X	X	X		X	X	X	X	X	X	X			X	X	X	X	X	X	X	X	X		X		X				X	X	X	X	X	X	X	X	
Oscarella	X		X	X	X					X							X		X	X	X		X	X	X		X	X		X	X			X	X			
Pedicellina		X																																				
Pediculus		X	X																																			
Petromyzon																																						
Phycomyces			X																												X							
Pleurobrachia		X	X	X									X					X	X			X		X	X		X											
Podocoryne															X																								
Proterospongia	X		X			X							X	X						X						X						X			X			
Rhizopus																																						
Saccoglossus																																						
Scutigera											X																									X		
Sphaeroforma																																						
Spizellomyces																																						
Strongylocentrotus																																						
Suberites		X	X																																			
Sycon		X													X													X										
Trichoplax		X																																				
Tubifex																																						
Xenoturbella																															X							

Table S2. Continued

OTU	sadhchydrolase-E1.a	sap40.ali	srs.ali	stbproptase2a-b.ali	suca.ali	tif2a.ali	vacaatpasepl21-a.ali	vata.ali	vatb.ali	vate.ali	vdac2.ali	% missing positions
Acropora							X					24
Allomyces										X		2
Amoebidium				X	X	X	X	X	X	X		66
Amphimedon												1
Anoplodactylus	X			X	X	X	X	X	X	X		56
Aplysia												2
Batrachochytrium												6
Branchiostoma												0
Capitella												0
Capsaspora												3
Carteriospongia	X		X	X	X	X	X	X	X	X		73
Ciona												1
Clytia							X					32
Crassostrea			X									17
Cyanea	X	X	X	X	X	X	X	X	X			75
Danio												1
Daphnia												0
Ephydatia			X	X	X	X		X		X	X	57
Euperipatoides	X		X	X			X	X	X			60
Euprymna								X				25
Helobdella												5
Heterochone		X			X		X			X	X	71
Hydra												0
Hydractinia	X		X	X	X		X	X	X	X		41
Ixodes												3
Leucetta			X			X						51
Mertensiid			X	X	X	X	X	X	X	X	X	60
Metridium			X	X								28
Mnemiopsis					X							27
Molgula						X	X					5
Monosiga												1
Monosiga												5
Montastraea	X	X	X	X	X		X	X	X	X	X	78
Nasonia												0
Nematostella												0
Oopsacas	X	X	X	X		X		X	X	X	X	88
Oscarella				X		X	X	X		X		44
Pedicellina		X	X	X	X	X	X	X	X	X		51
Pediculus												3
Petromyzon												9
Phycomyces										X		2
Pleurobrachia												19
Podocoryne			X	X	X			X	X	X		44
Proterospongia			X		X	X	X		X	X	X	64
Rhizopus												1
Saccoglossus												2
Scutigera		X	X	X	X	X	X	X	X	X	X	62
Sphaeroforma			X									31
Spizellomyces												5
Strongylocentrotus												0
Suberites	X		X	X	X	X	X	X	X	X	X	60
Sycon			X		X		X	X	X	X	X	56
Trichoplax												2
Tubifex			X						X			19
Xenoturbella				X		X		X		X		39

Table S3. Summary of the conflicts between single and supermatrix phylogenies.

Explanations	#conflicts
Nearest Neighbor Interchange	58 (*Acropora*=5, *Allomyces*=26, *Clytia*=4, *Mnemiopsis*=10)
Local rearrangements (two nodes)	5
Long Branch Attraction	6
No obvious explanations	11

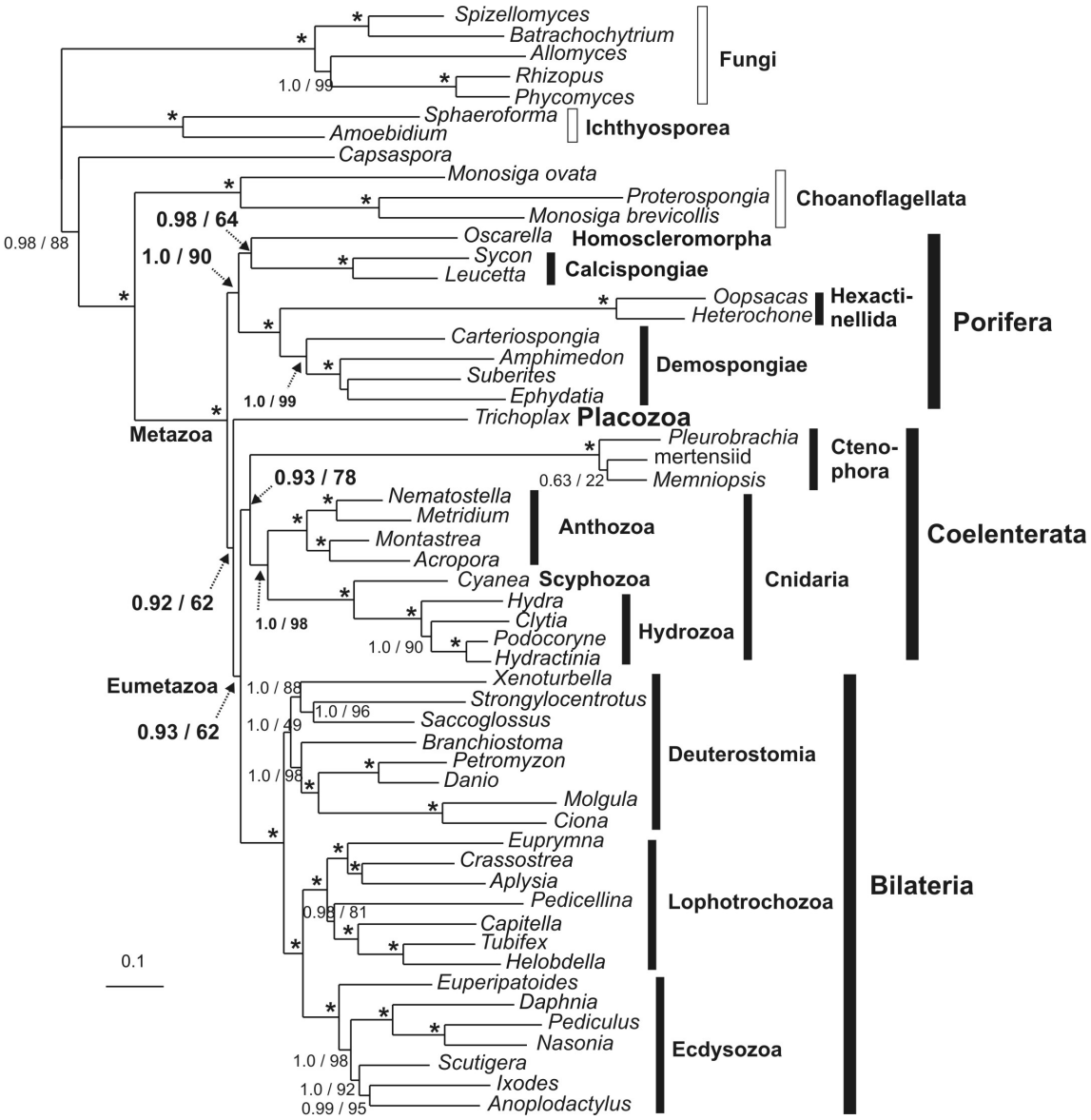

Figure S1. Tree resulting from the analysis of the full dataset (55 terminal taxa; "outgroup 1" in Fig. 1), identical to the tree shown in Fig. 1 but with Bayesian posterior probabilities (PP) and bootstrap proportions after 100 replicates (BP) (PP / BP). Nodes with maximal support values are indicated by an asterisk. Scale bar indicates number of changes per site.

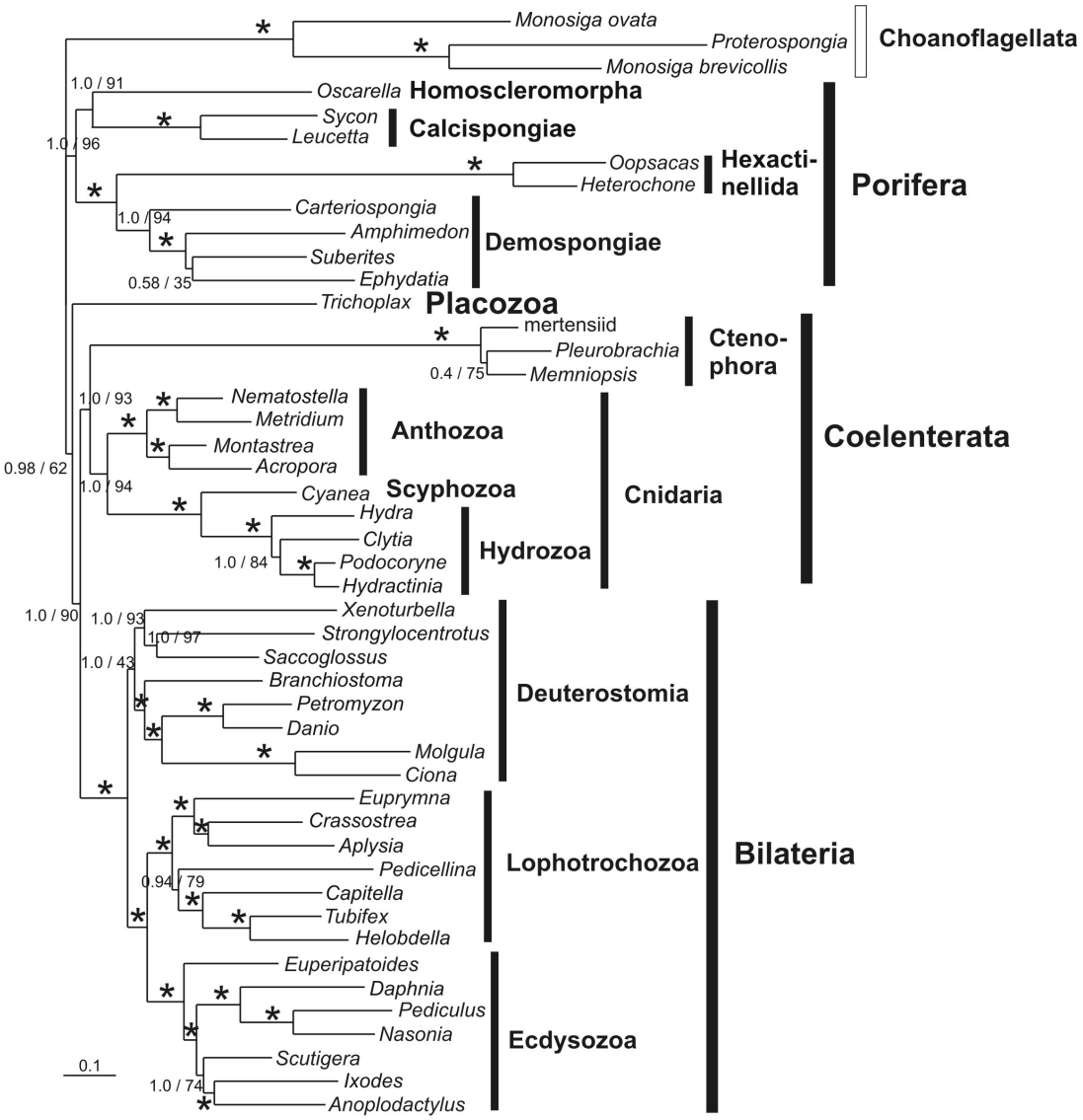

Figure S2. Tree resulting from the analysis rooted on choanoflagellates ("outgroup 2" in Fig. 1). Bayesian posterior probabilities (PP) and bootstrap proportions after 100 replicates (BP) are shown (PP / BP). Nodes with maximal support values are indicated by an asterisk. Scale bar indicates number of changes per site.

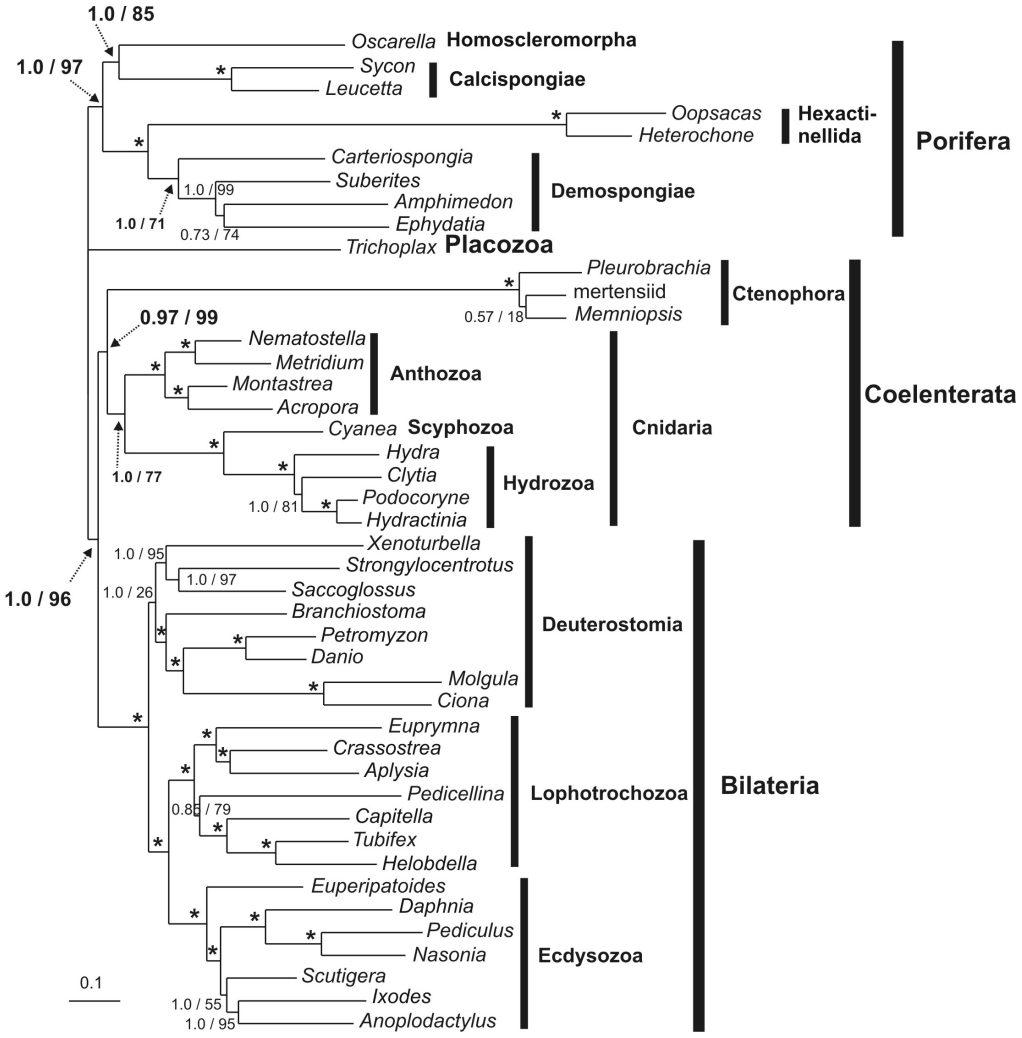

Figure S3. Tree resulting from the unrooted analysis. Bayesian posterior probabilities (PP) and bootstrap proportions after 100 replicates (BP) are shown (PP / BP). Nodes with maximal support values are indicated by an asterisk. Scale bar indicates number of changes per site.

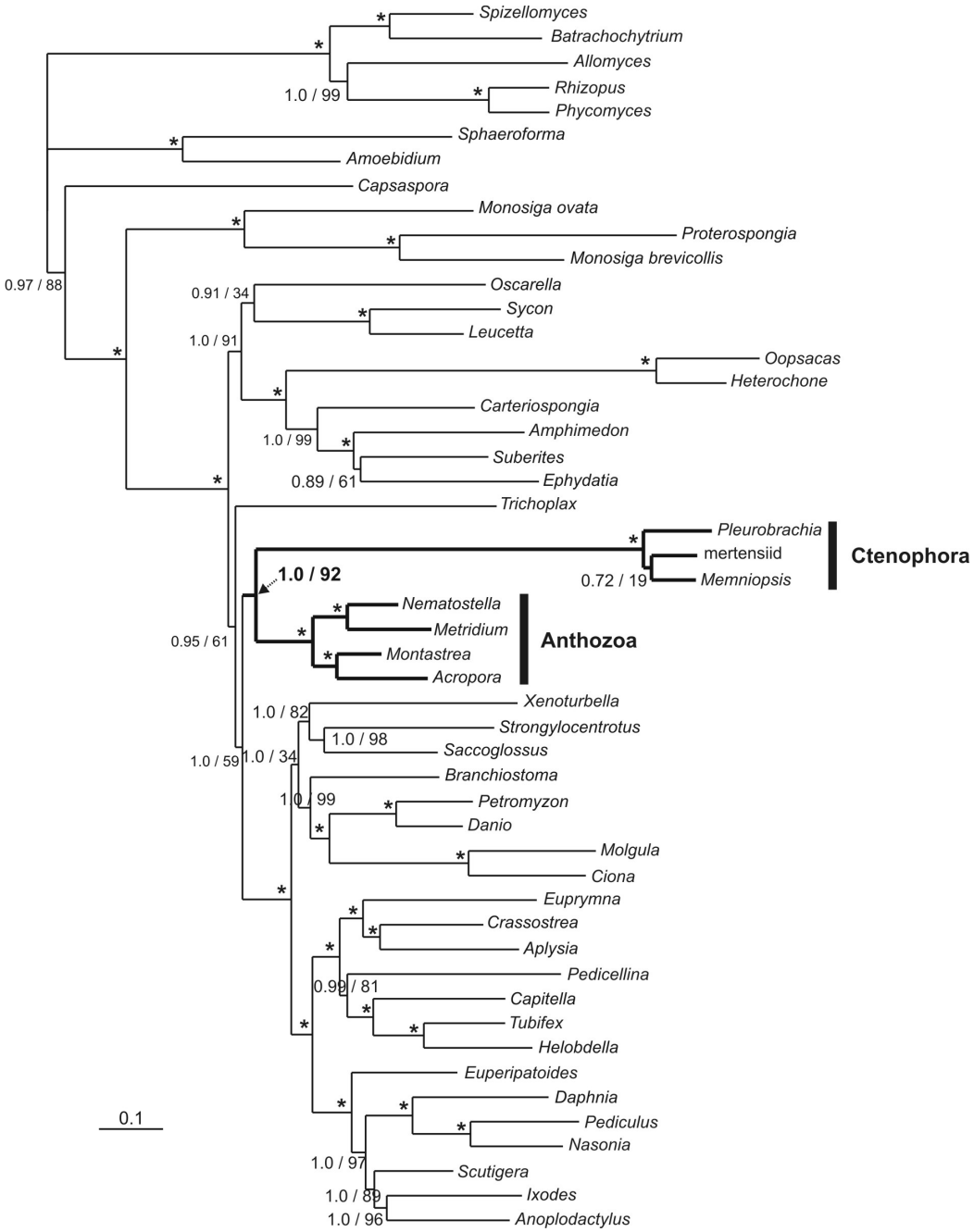

Figure S4. Phylogenetic analyses of same molecular data set as in Fig. 1 but following removal of the Medusozoa (hydrozoan and scyphozoan cnidarians). Bayesian posterior probabilities (PP) and bootstrap proportions after 100 replicates (BP) are shown (PP / BP). Nodes with maximal support values are indicated by an asterisk. Scale bar indicates number of changes per site.