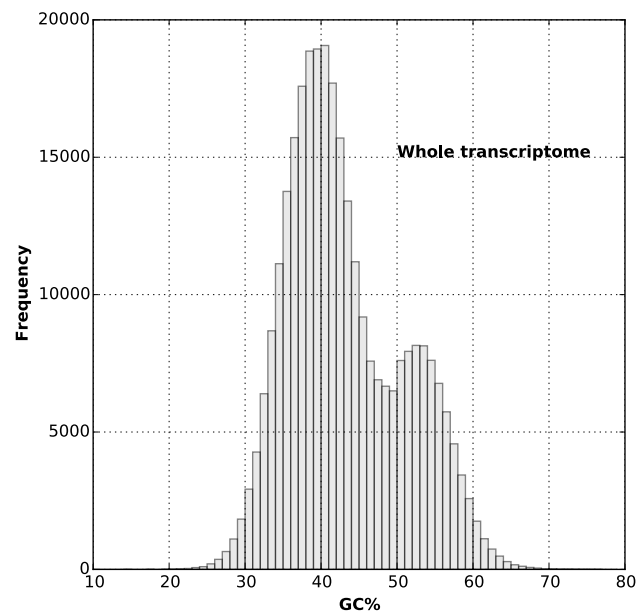
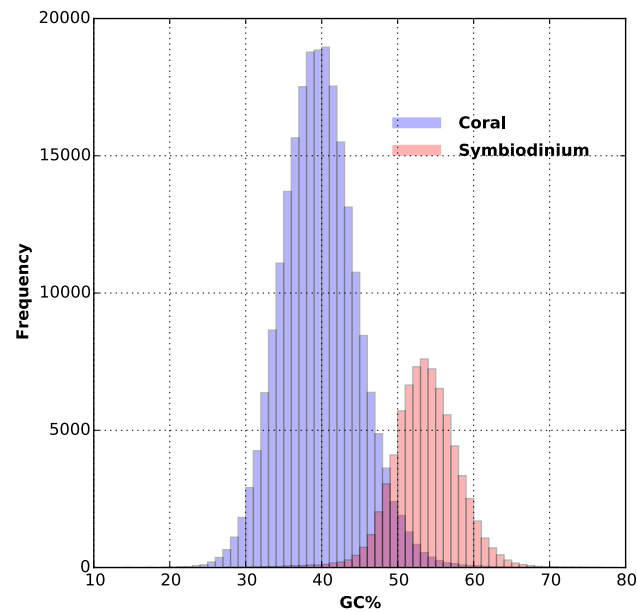


### Additional file 3

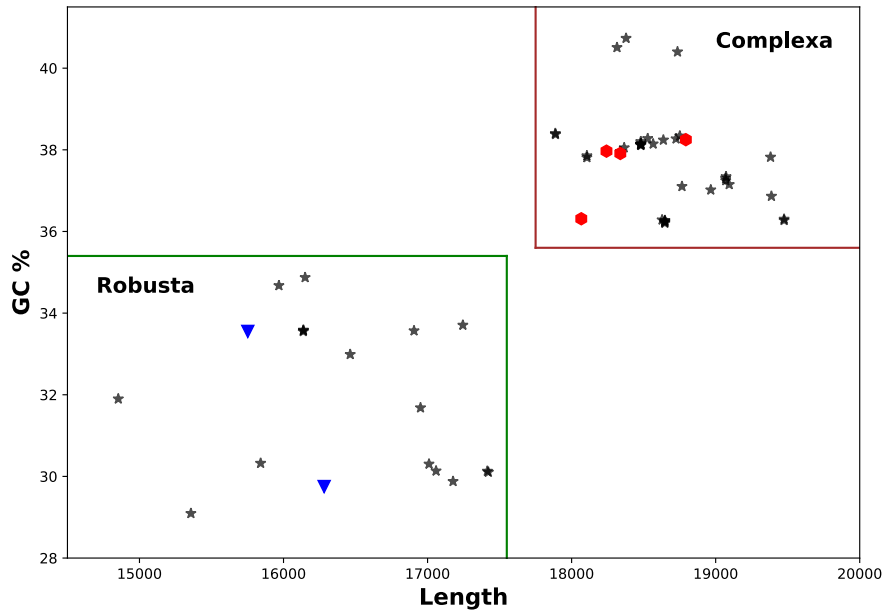
(a)



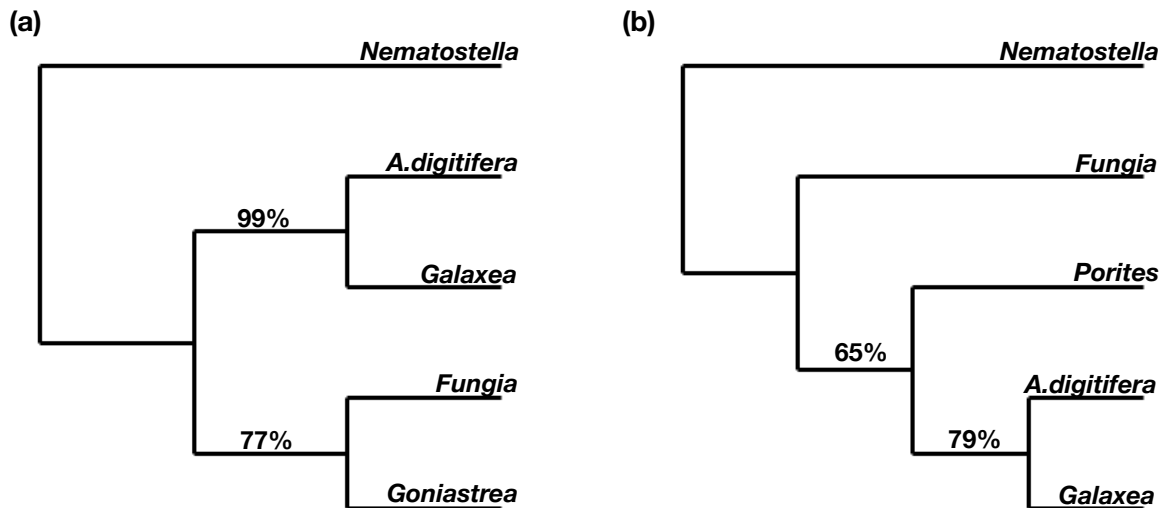
(b)



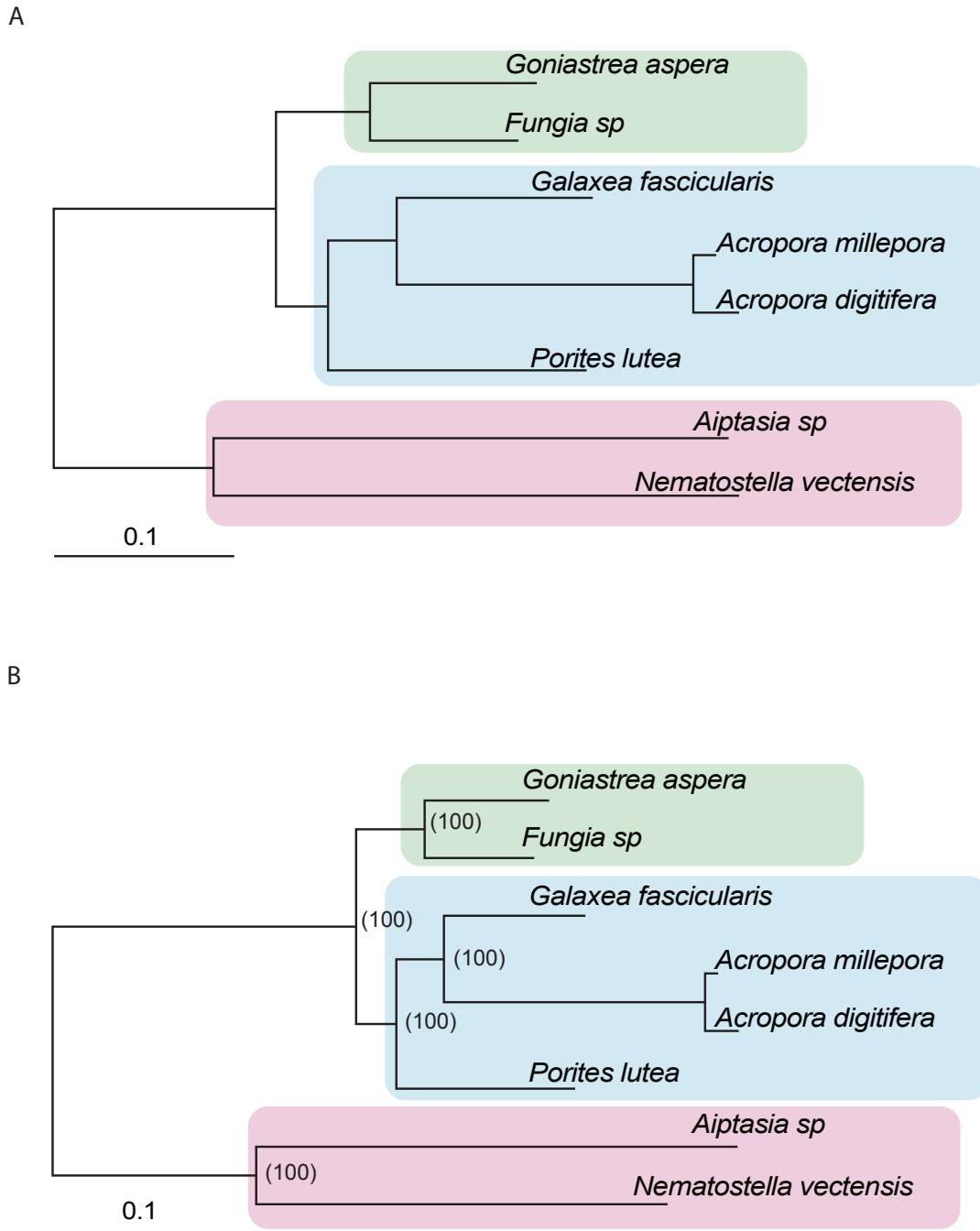
**Figure S1: GC content of *de novo* assembled transcripts before and after PSyTrans treatment.** (a) An apparent feature of Symbiodinium transcripts is that their GC content is higher than those from corals. Therefore, the GC content appears as a bimodal distribution from total RNA transcripts. (b) After PSyTrans separation, GC contents showed normal distributions from both coral and *Symbiodinium* transcripts.



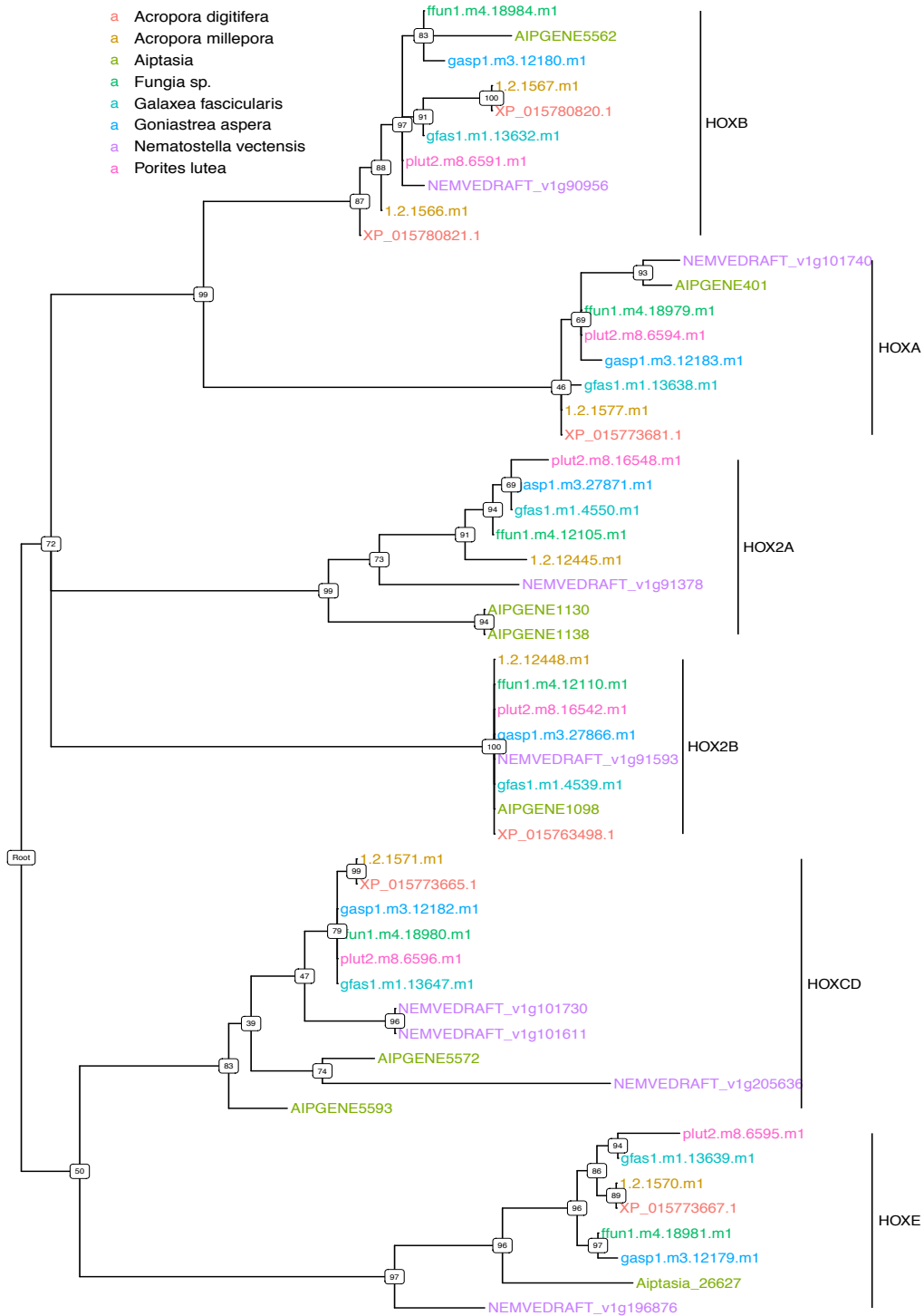
**Figure S2: GC content of mitochondrial genomes.** Black stars represent sequences obtained from the NCBI nucleotide database. Stars in the green box represent robust corals, and stars in the brown box represent complex corals. Blue triangles represent the two robust corals (*Fungia* and *Goniastrea*), and red pentagons represent the four complex corals (*A. digitifera*, *A. millepora*, *Galaxea* and *Porites*) used in the present study.



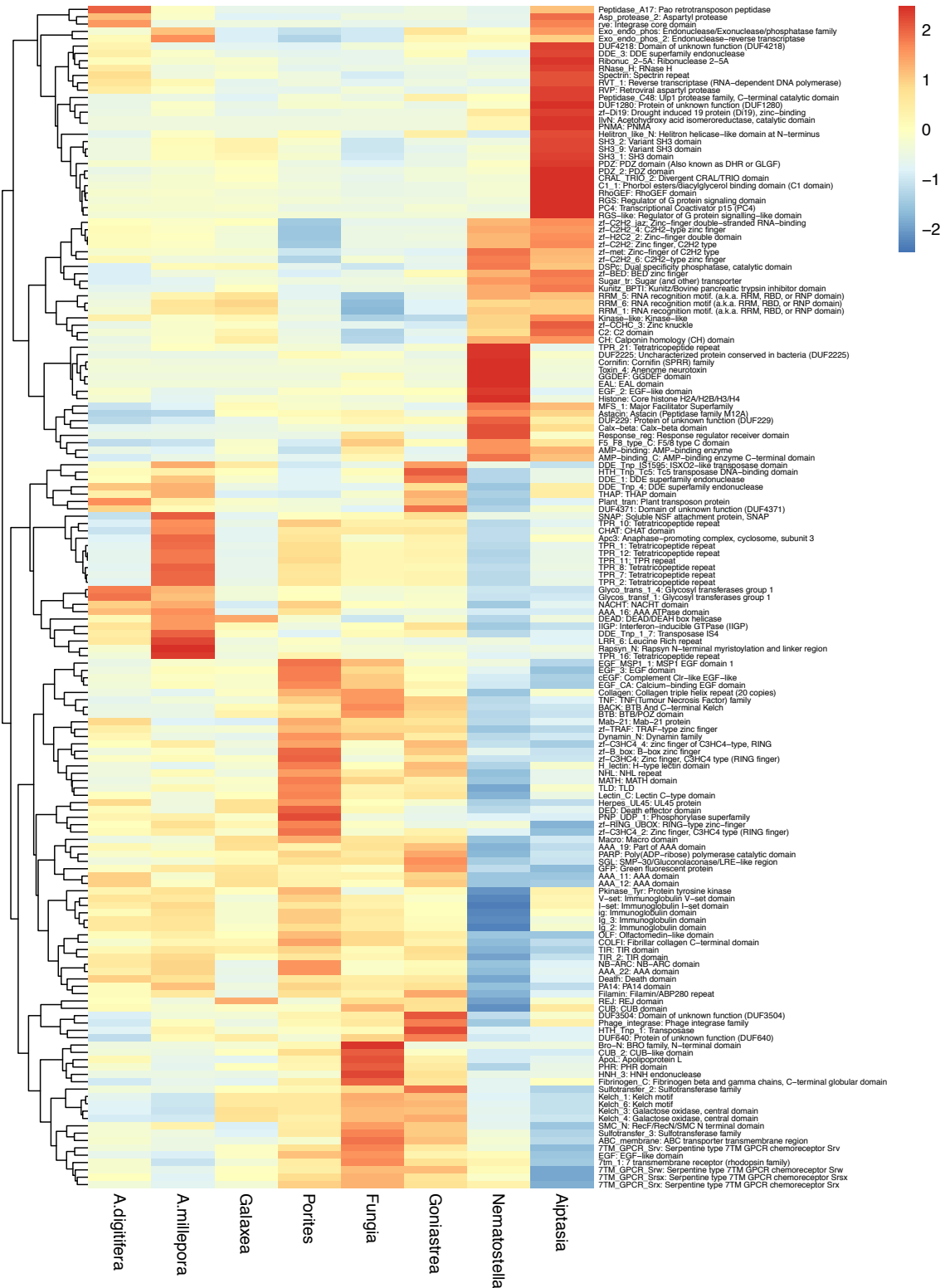
**Figure S3: Consensus phylogenetic trees estimated from GN model with node support.** The tree topology was firstly generated for *Nematostella*, *Galaxea*, *A. digitifera* and *Fungia*. *Nematostella* was used as an outgroup to indicate the root position for corals. The relative positions of *Goniastrea* (a) and *Porites* (b) to the four species were inferred separately. The percentages show node support based on the likelihood weights method described in Holland et al. 2006 [107].



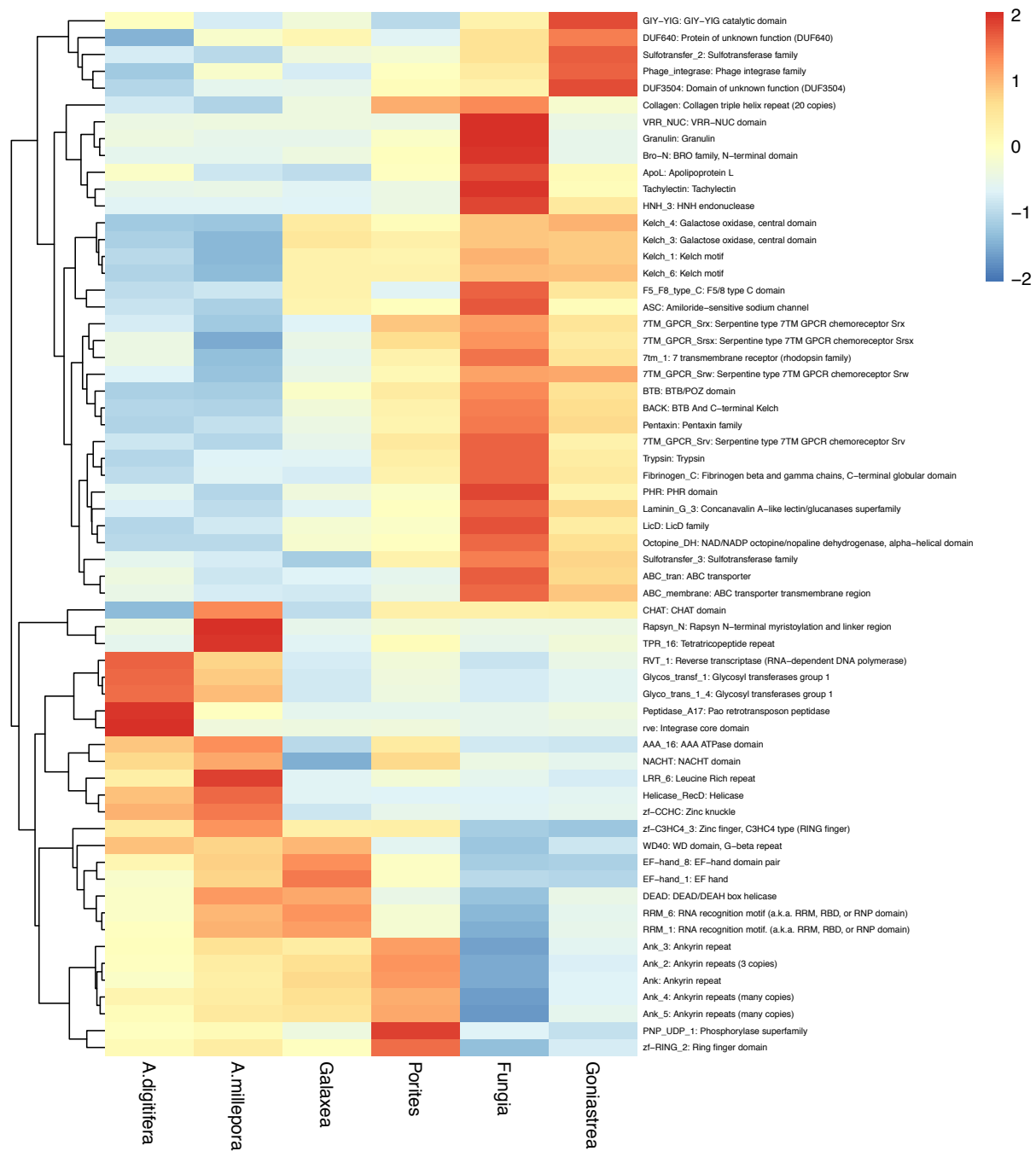
**Figure S4: Comparison of consensus phylogenetic trees for the eight cnidarian species which are the focus of comparative analyses, estimated from the GN (A) and AA (B) models.** Branch lengths of the GN-based tree were estimated from 91 protein coding sequence alignments, and node support values are shown in Figure S3. Protein alignments from the same 91 one-to-one orthologous genes were used to generate the AA-based tree. The partition model implemented in IQ-TREE was applied. Node support values (shown as percentages) were obtained from 1000 UFboot replicates.



**Figure S5: Phylogenetic relationships among HOX-related genes in corals and anemones.** All genes are coloured according to species. Clades were identified and labelled according to the presence of sequences previously classified by Chourrout et al. 2006, Baumgarten et al. 2015 [19, 39]. Model (LG+G4) was selected as the best model. Node labels show support as a percentage of 1000 UFboot replicates. All node branch lengths less than 1E-5 were collapsed to polytomies. Sequences on which the tree was based are summarised in Additional file 2: Table S13 and provided in additional file 4.



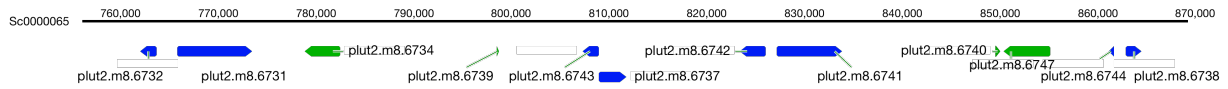
**Figure S6: Heatmap showing relative counts of genes that displayed significant differences between corals and sea anemones.** To obtain relative counts raw counts were centered and scaled for each row using the R function, scale. Included domains are those that differed significantly (FDR = 0.01 based on a Fisher's exact test) between anemones (*Nematostella* and *Aiptasia*) and corals (all other species). See Additional file 2: Table S15 for details of test statistics.



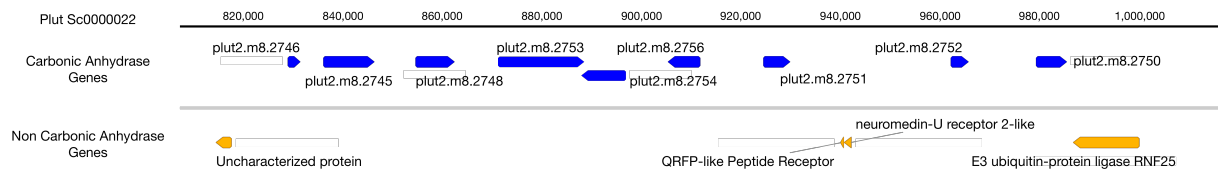
**Figure S7: Heatmap showing relative counts of genes that displayed significant differences between complex and robust corals.** To obtain relative counts raw counts were centered and scaled for each row using the R function, scale. Included domains are those that differed significantly (FDR = 0.01 based on a Fisher's exact test) between robust (*Fungia*, *Goniastrea*) and complex corals (*A. digitifera*, *A. millepora*, *Galaxea* and *Porites*). See Additional file 2: Table S16 for details of test statistics.



B

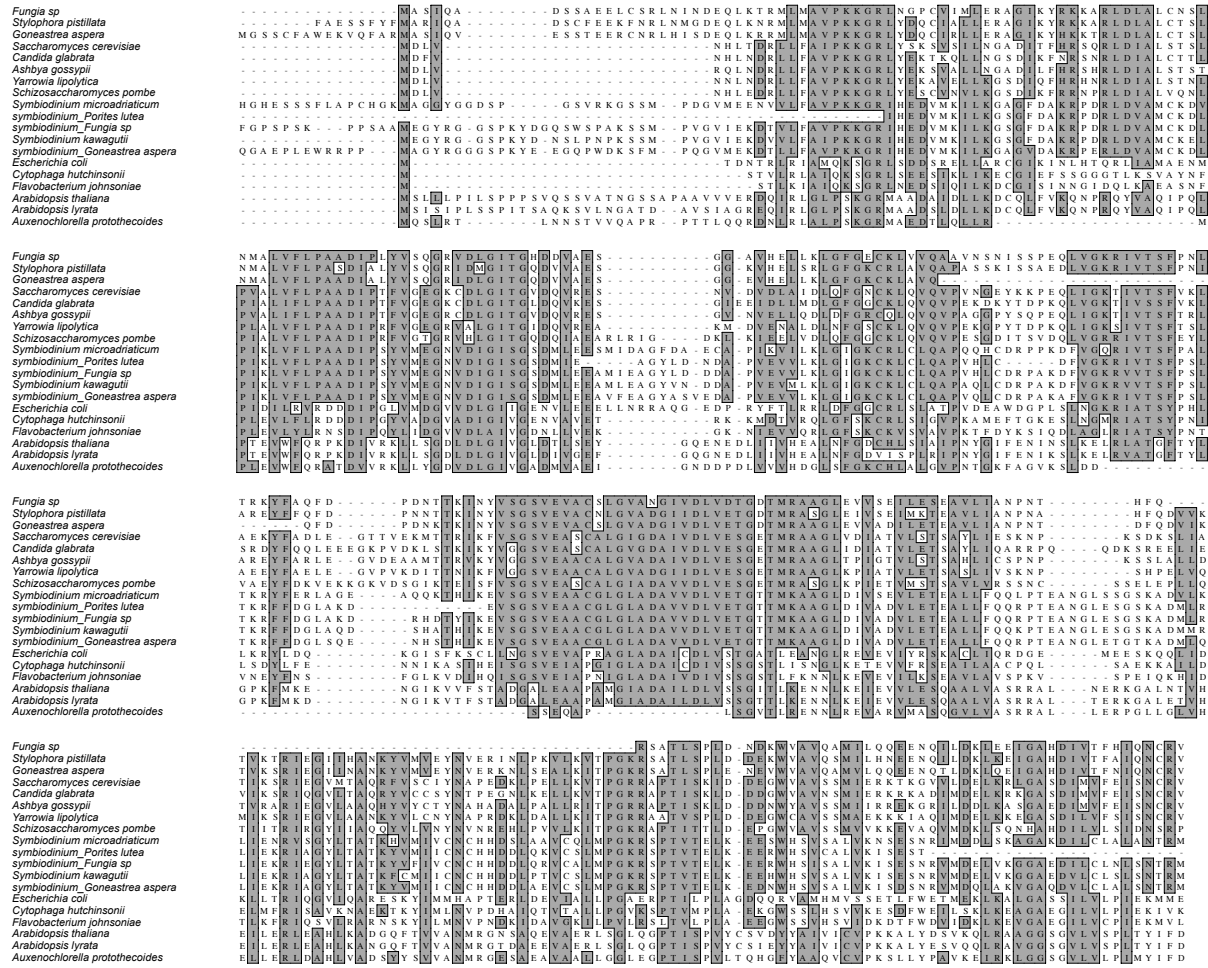


**Figure S8: Phylogenetic relationships of HSP20 genes in corals and anemones (A) and arrangement of HSP20 genes in *Porites lutea* (B).** (A) Genes are coloured according to species and corresponding scaffolds containing 2 or more HSP20 genes are shown in alternating colours. IQ-TREE was used to generate the maximum likelihood tree. Automatic model selection chose JTT+G4 as the best model and 1000 UFBoot replicates were used to generate support values shown as node labels. Sequences on which the tree was based are summarized in Additional file 2: Table S17 and provided in Additional file 5. (B) Close spacing of eight HSP20 paralogs is shown in blue with other genes shown in green.

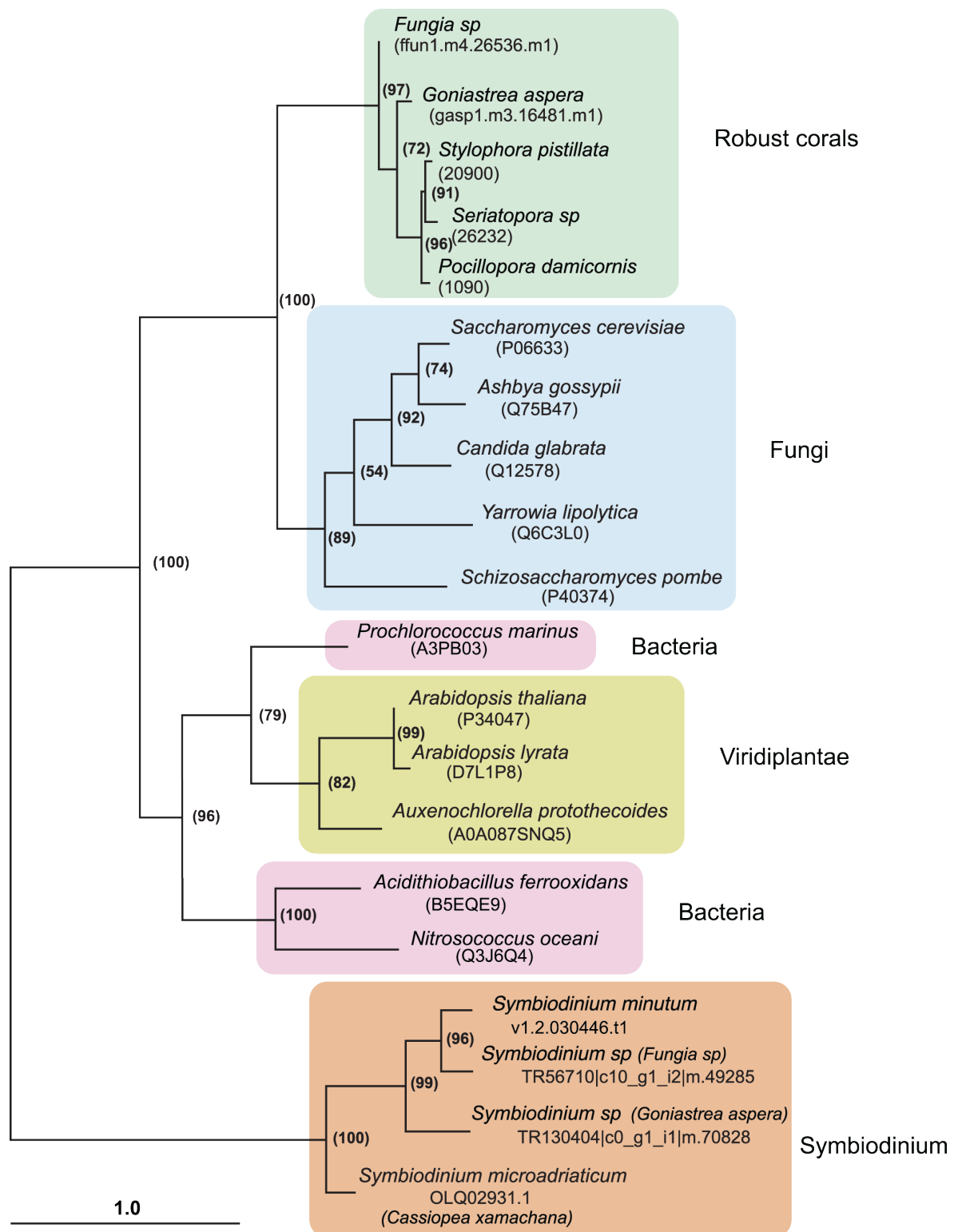


**Figure S9: Organisation of membrane associated carbonic anhydrase genes in *Porites lutea*.** Genes are shown as coloured blocks with the point designating the direction of transcription. Blue track shows the position and orientation of genes coding for carbonic anhydrases. The blue region spans ~160kb and contains just two non-carbonic anhydrase genes (shown in yellow).

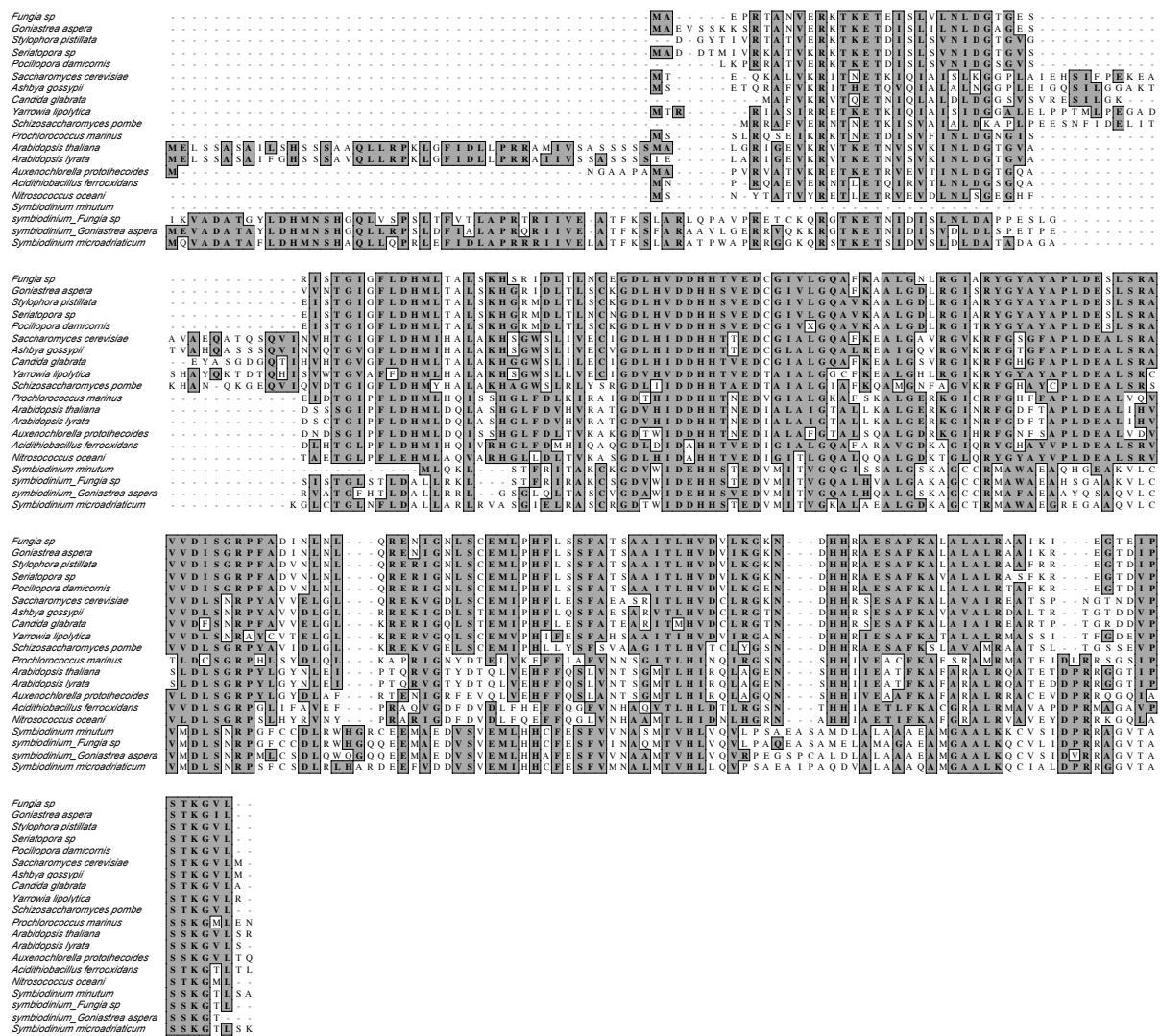




**Figure S10: Alignment of ATP phosphoribosyltransferase (K00765) proteins.** Poorly aligned columns with less than 20% alignable sequences were removed. The alignment shown was used to generate the phylogenetic tree Fig. 6. In the case of *Symbiodinium* strains, the host species are indicated after the underscore. Sequences are summarised in additional file 2: Table S20. Full length annotated protein sequences are provided in Additional file 6.



**Figure S11: Phylogenetic tree of Imidazoleglycerol-phosphate dehydratase (K01693) proteins.** IQ-TREE was used to generate the unrooted tree shown. Automatic model selection chose LG+R4 as the best model and 1000 UFBoot replicates were used to generate support values shown as node labels. The tree is displayed as rooted at the midpoint. Sequences on which the analyses are based are summarised in additional file 2:Table S21 and provided in Additional file 7. In general, the *Symbiodinium* proteins are significantly longer than other sequences, and typically contain two IGPD domains.



**Figure S12: Alignment of Imidazoleglycerol-phosphate dehydratase (K01693) protein sequences.** Poorly aligned columns with less than 20% alignable sequences were removed. The alignment shown was used to generate the phylogenetic tree Figure S11. In the case of *Symbiodinium* strains, the host species are indicated after the underscore. Sequences are summarised in Additional file 2: Table S21. Full length of annotated protein sequences are provided in Additional file 7.