Supplementary Information: The genetic basis for adaptation in giant sea anemones to their symbiosis with anemonefish and Symbiodiniaceae

Agneesh Barua

August 2022

Figure S1: Spearmans correlations for all species and all tissues

The gene expression of jellyfish tentacles was highly distinct, with very low similarities between jellyfish and non-jellyfish samples

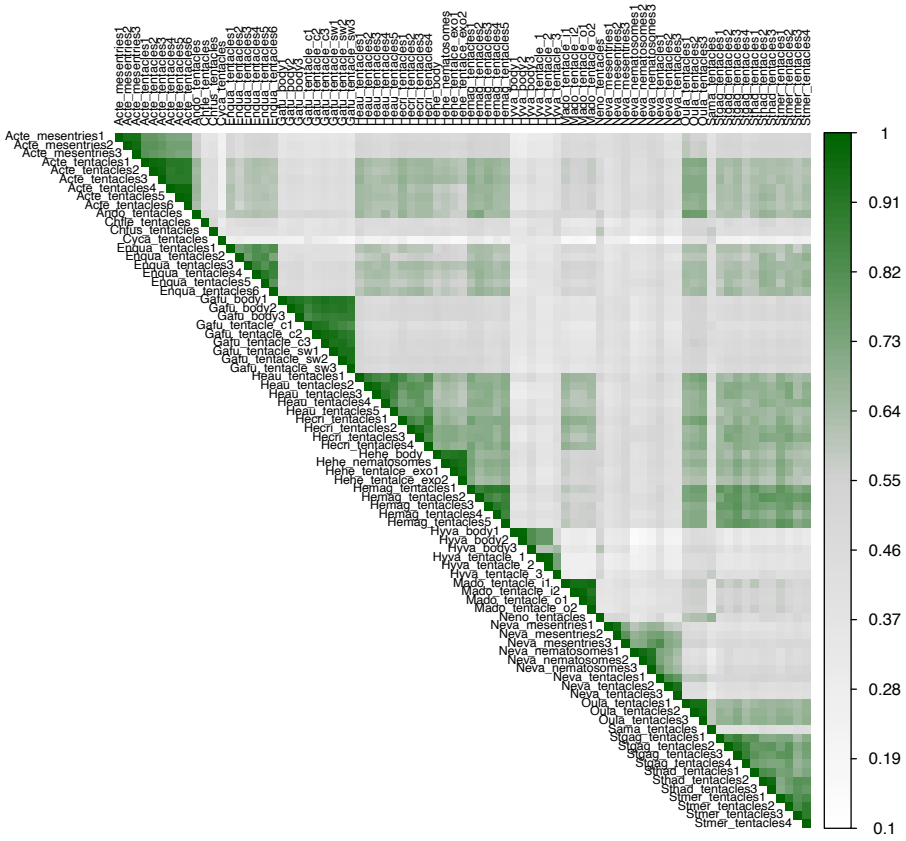


Figure S2: Spearmans correlations for non-actinaria and Heteractis-Stichodactyla

The following figure contrasts the difference in expression between jellyfish and giant sea anemones *Heteractis* and *Stichodactyla*

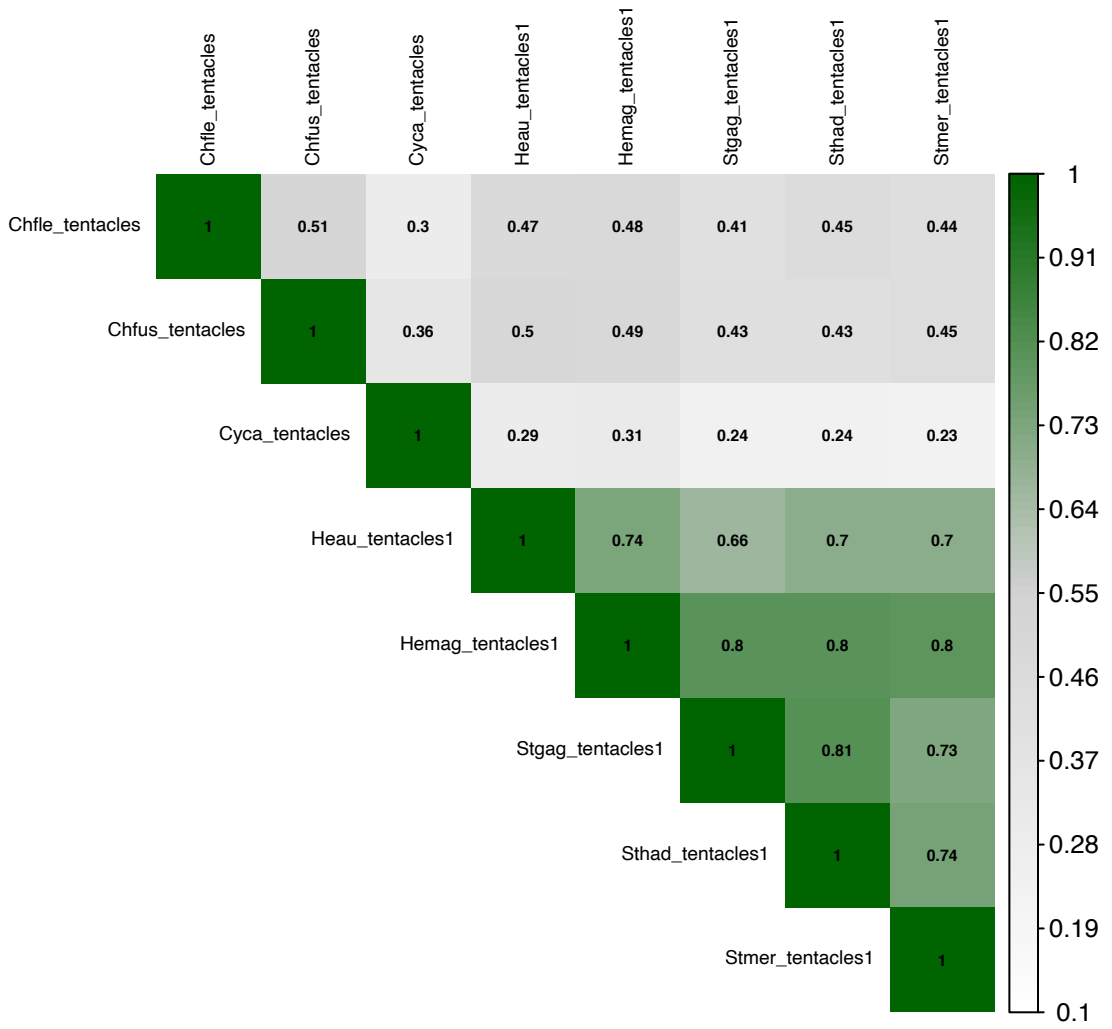


Figure S3: Spearmans correlations for all actinaria tentacles

Gene expression within actinaria was overall more similar than between actinaria and non-actinaria

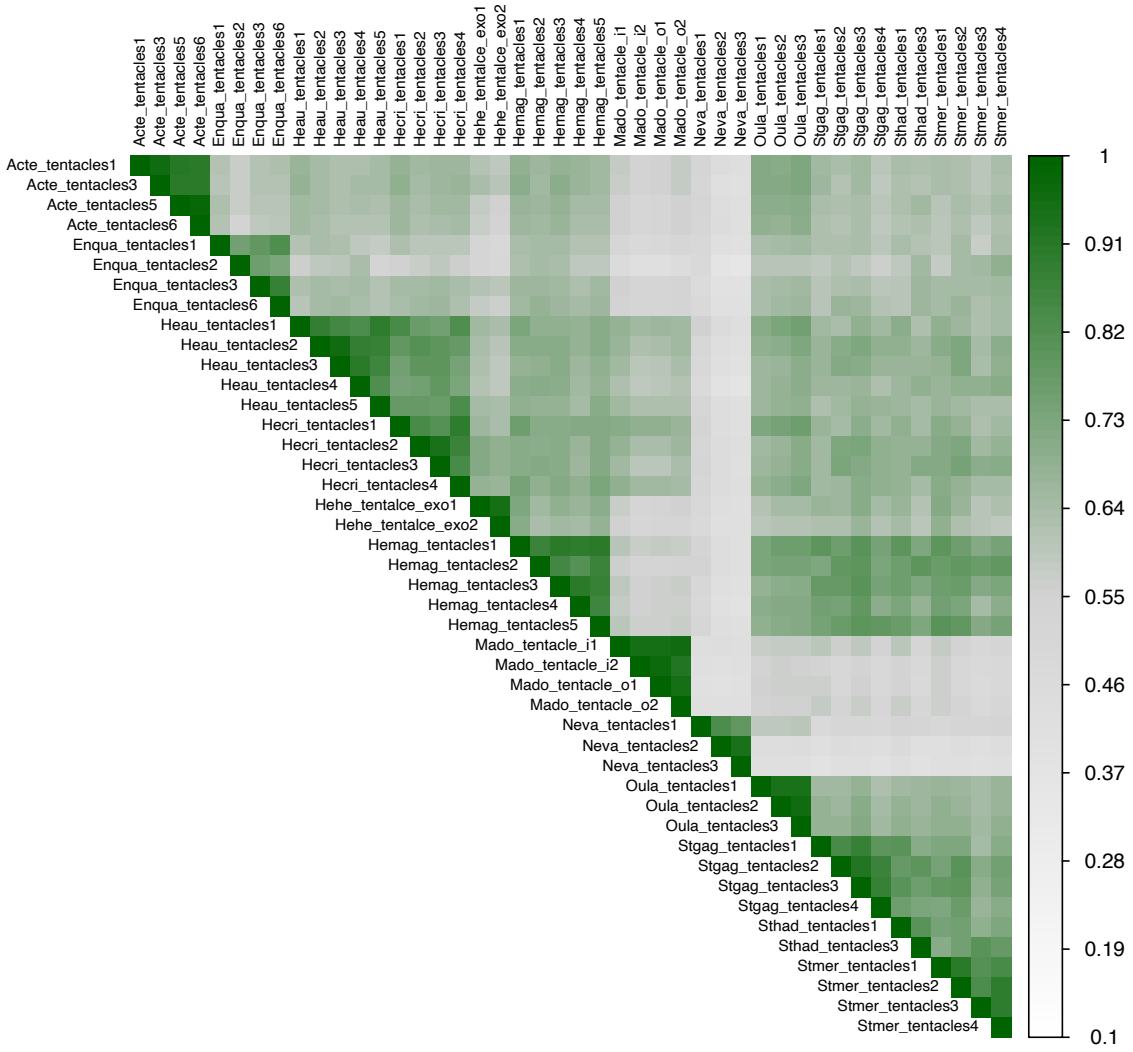


Figure S4: PCA for all cnidaria samples

The PCA with all cnidarian samples is included in the supplementary information and showed the same overall trends as ones with only anthozoans.

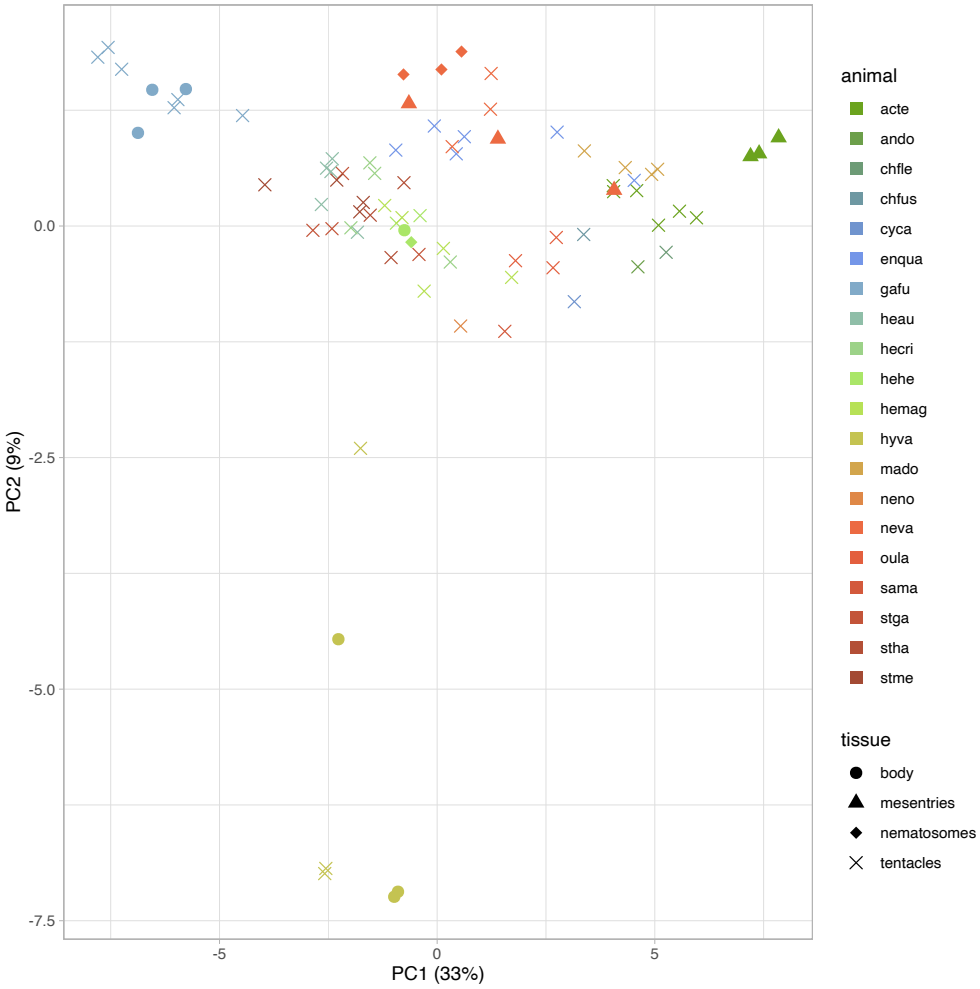


Figure S5: Spearmans correlations for random set of 47 genes

We estimated correlations using a random assortment of 47 genes to check if this was the case. A random set of 47 genes showed less noisy correlation patterns (Fig S5), suggesting that the 47 genes in cluster 1 are highly variable due to biological factors and not sampling bias.

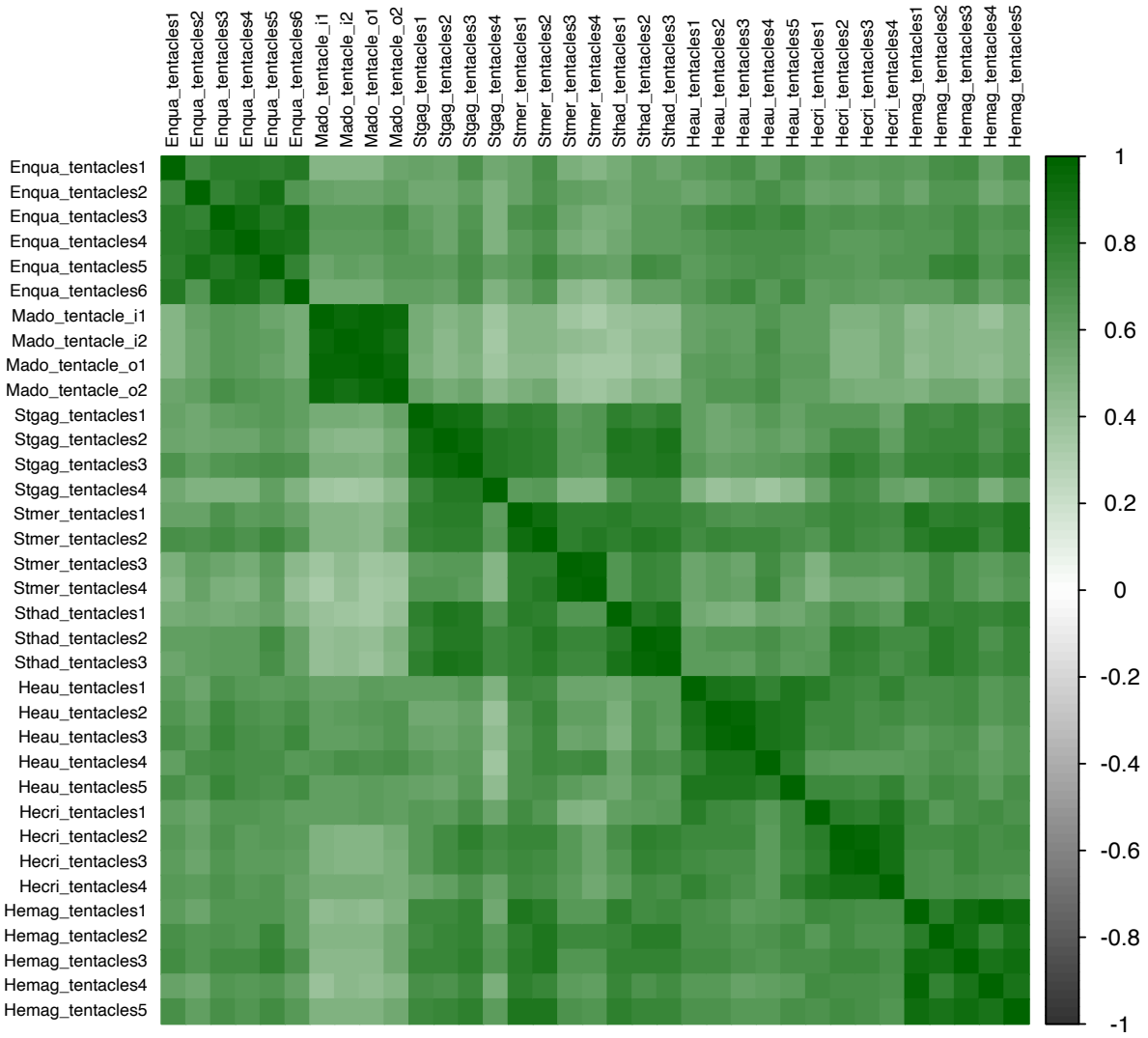


Figure S6: Coseq analysis for downregulated orthologs

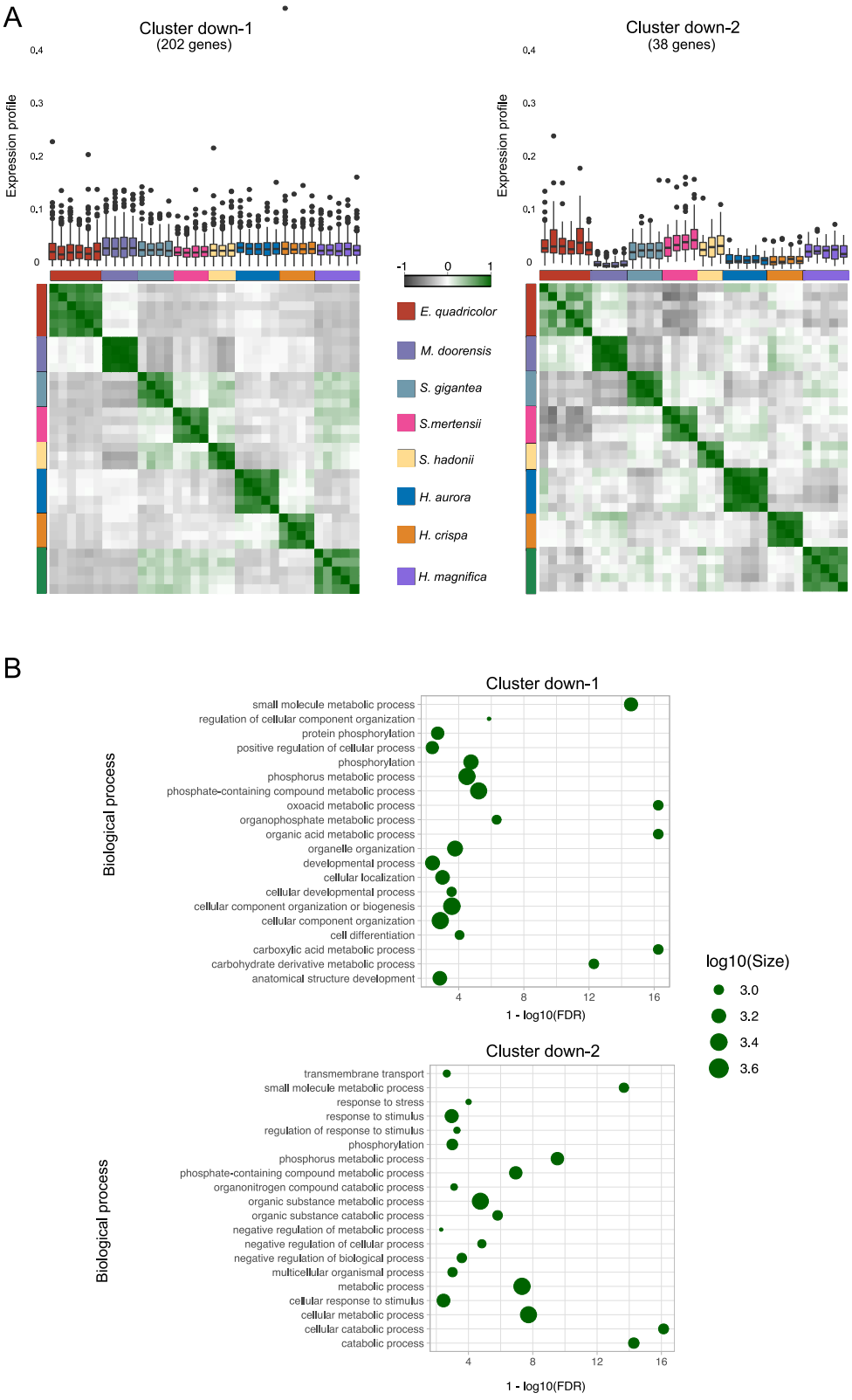


Figure S7: Symbiodiniaceae abundance

We used a pseudoalignment approach to determine the Symbiodiniaceae content in our species. The following figure shows the total number of reads pseudoaligned to the composite Symbiodiniaceae transcriptome.

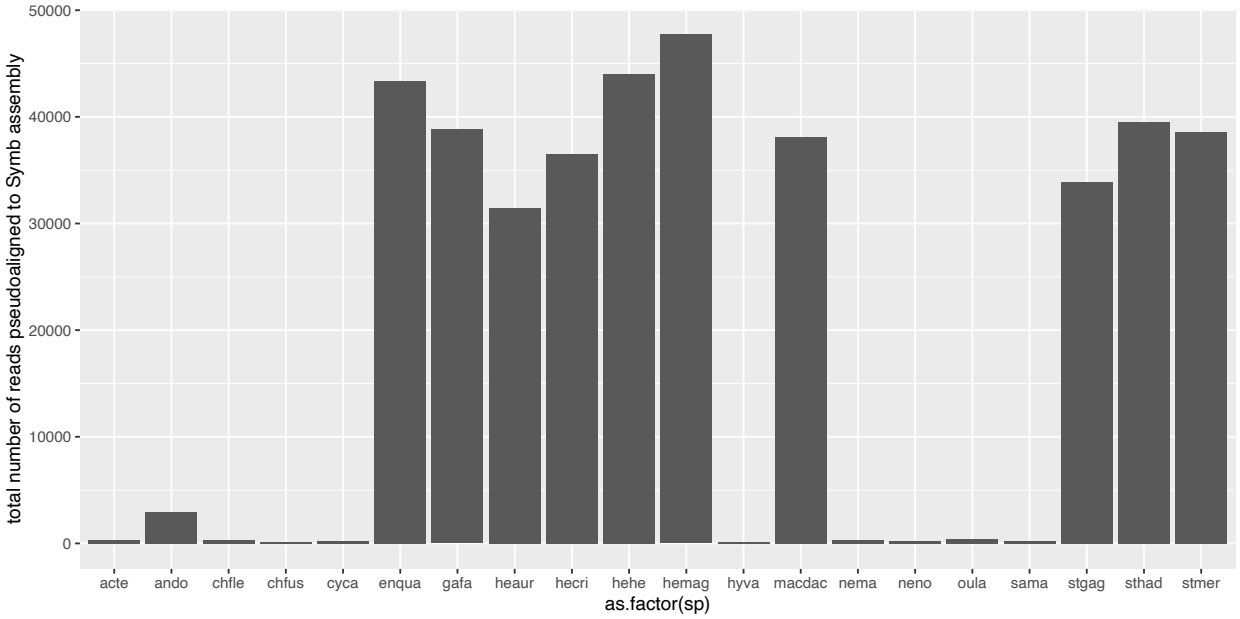
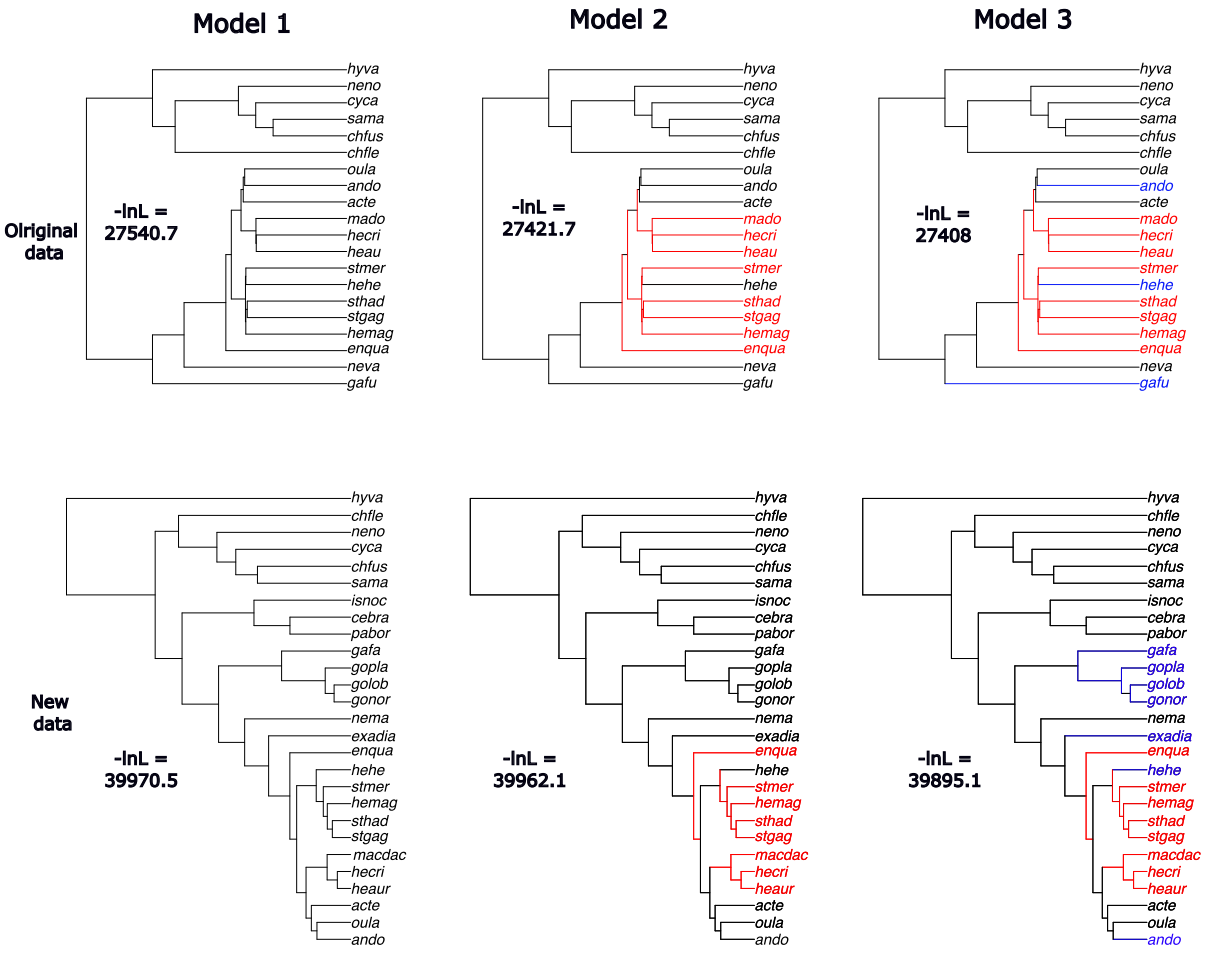


Figure S8: Additional CAGEE analysis

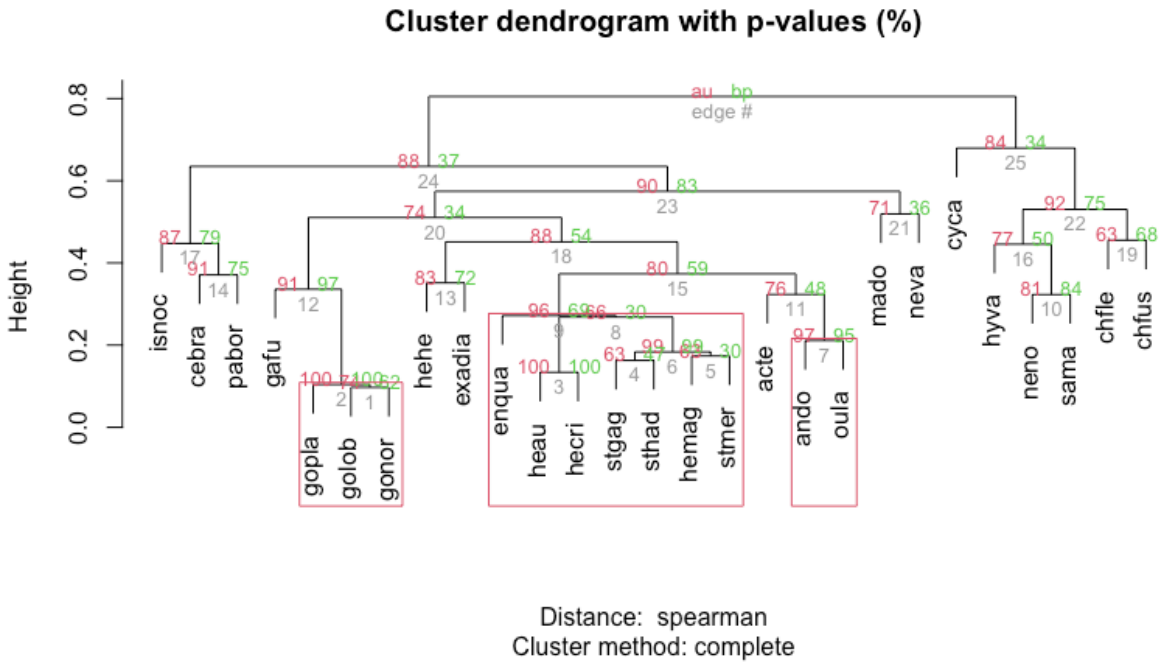
In our original analysis, the multi-rate model has significantly better fit. However an argument against our results could be the fact that we sample few species that host only Symbiodiniaceae. To address this, we carried the CAGEE analysis using tentacle transcriptomes from additional three coral species and one anemone, all of which harbour only Symbiodiniaceae. Furthermore, we also include three species of anemones that do not host either anemonefish or Symbiodiniaceae. Using this dataset, we not only confirm our previous trend - where the multi-rate model has the best fit - but we also observe that the multi-rate model has substantially better fit than any of the alternative models.
Hence, the inclusion of additional data not only corroborates our initial results, but it also provides more robust evidence supporting the evolutionary rate heterogeneity in gene expression within our dataset.



Additional species names:
pabor - *Pachycerianthus borealis.* isnoc - *Isarachnathus nocturnus.* cebra - *Ceriantheomorphe brasiliensis.* exadia - *Exaiptasia diaphana.* golob - *Goniophora lobata.* gopla - *Goniopora planulata.* gonor - *Goniopora norfolkensis*

Figure S9: Additional hclust analysis

Using the additional data, we also checked how the tentacle transcriptome clustered. We used the pvclust R package, which uses bootstrapping to identify groups that form significant clusters. The figure below shows that the giant sea anemones form a combined significant cluster (red) with corals *golob*, *gopla*, and *gonor* forming one cluster, *ando* and *oula* forming the other. Interestingly, *hehe* and *exadia* which are sea anemones which host Symbiodiniaceae but not anemonefish, do not cluster with the giant sea anemones suggesting the expression profile between anemones hosting anemonefish and Symbiodiniaceae is distinct from those hosting Symbiodiniaceae alone.



The standard error plot below shows that as the significance threshold (AU p-value) increases, the standard error becomes close to zero. This demonstrates the power of the current dataset.

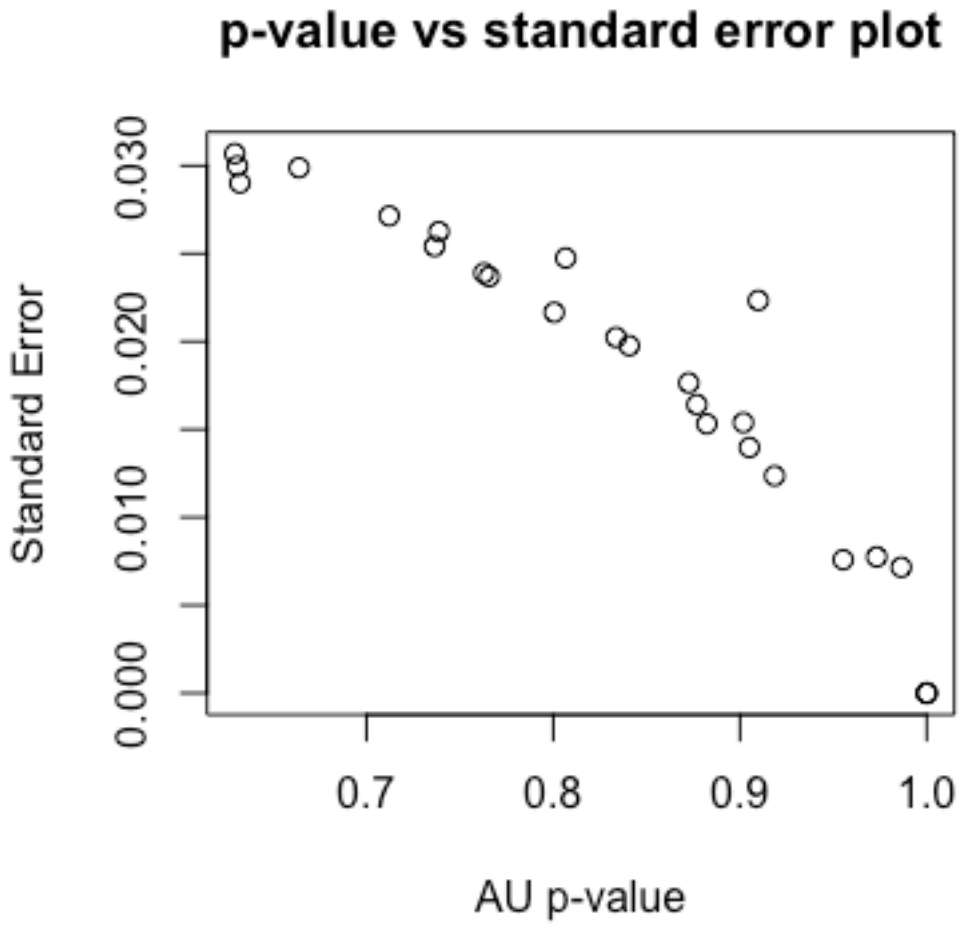


Figure S10: Additional PCA

Our dataset includes more samples of tentacles than other tissues. While the idea of selecting other tissues was to observe how distinct the tentacle expression pattern might be, the large number of tentacle samples might make interpretation of the PCA difficult. To improve the interpretation, we use mean values of all tissues such that design of the data matrix is unform and any variation can be attributed to the level of gene expression in each tissue. Using this design, we see that most tentacles tightly group together, with *Hydra*, *Nematostella*, and *Actinia tenebrosa* forming their own groups.