SUPPLEMENTARY MATERIAL: Molecular convergence and positive selection associated with the evolution of symbiont transmission mode in stony corals

Groves B. Dixon[a] and Carly D. Kenkel[b,1]

[a]Department of Integrative Biology, The University of Texas at Austin, 1 University Station C0990, Austin, TX 78712, USA

[b]Department of Biological Sciences, University of Southern California, 3616 Trousdale Parkway, Los Angeles, CA 90089, USA

[1]Corresponding author, email: ckenkel@usc.edu; phone: 213-821-1705; fax: 213-740-8123

METHODS

Sample preparation and sequencing for Montipora aequituberculata *reference transcriptome*

Samples of *Montipora aequituberculata* were collected under the Great Barrier Reef Marine Park Authority permit G12/35236.1 and G14/37318.1. To generate a *M. aequituberculata* reference transcriptome, five replicate fragments of a single coral colony were subject to a two-week temperature stress experiment as described in [1] and snap frozen samples from control (27°C, days 4 and 17) and heat (31°C, days 2, 4 and 17) treatments were crushed in liquid nitrogen and total RNA was extracted using an Aurum Total RNA mini kit (Bio-Rad, CA). RNA quality and quantity were assessed using the NanoDrop ND-200 UV-Vis Spectrophotometer (Thermo Scientific, MA) and gel electrophoresis. RNA samples from replicate fragments were pooled in equal proportions and 1.8 µg was shipped on dry ice to the

Genome Sequencing and Analysis Facility (GSAF) at the University of Texas at Austin where Illumina TruSeq Stranded libraries were prepared and sequenced on one lane of the Illumina Hiseq 4000 to generate 2 x 150 PE reads.

Transcriptome assembly and annotation

Sequencing yielded 98 million raw PE reads. The *fastx_toolkit* (http://hannonlab.cshl.edu/fastx_toolkit) was used to discard reads < 50 bp or having a homopolymer run of 'A' ≥ 9 bases, retain reads with a PHRED quality of at least 20 over 80% of the read and to trim TruSeq sequencing adaptors. PCR duplicates were then removed using a custom perl script (https://github.com/z0on/annotatingTranscriptomes). Remaining high quality filtered reads (37.7 million paired reads; 6.7 million unpaired reads) were assembled using Trinity v 2.0.6 [2] using the default parameters and an *in silico* read normalization step at the Texas Advanced Computing Center (TACC) at the University of Texas at Austin. Since corals are 'holobionts' comprised of host, Symbiodiniaceae and other microbial components, resulting assemblies were filtered to identify the host component following the protocol described in [3].

Transcriptomic resources

Data from an additional 25 species of Scleractinia (stony corals) and 3 species of Actiniaria (anemones) were downloaded from the web (Table S1; [4]; [5]; [6]; [7]; [8]; [9]; [10]; [11]; [12]; [13]; [14]; [15]; [16]; [17]; [18]; [3]).

Table S1. Sources of reference transcriptomes used for each species. In the Citation column, 'none' indicates that the data have been made publicly available, but are not associated with any publication to date.

order	family	Genus	species	Citation	URL
Actiniaria	Actiniidae	Anthopleura	elegantissima	Kitchen et al. 2015	http://people.oregonstate.edu/~meyere/data.html
Actiniaria	Aiptasiidae	Aiptasia	pallida	Lehnert et al. 2012	http://pringlelab.stanford.edu/project%20files/AposymbioticAiptasiaTranscriptomeGoodLoci.fa.gz
Actiniaria	Edwardsiidae	Nematostella	vectensis	Nordberg et al. 2014	http://genome.jgi-psf.org/Nemve1/Nemve1.download.ftp.html
Scleractinia	Acroporidae	Acropora	cervicornis	Libro et al. 2013	http://www.ncbi.nlm.nih.gov/nuccore?LinkName=bioproject_nuccore&from_uid=222758
Scleractinia	Acroporidae	Acropora	palmata	Polato et al. 2011	http://www.personal.psu.edu/ibb3/Research.htm#Data
Scleractinia	Acroporidae	Acropora	hyacinthus	Barshis et al. 2013	http://palumbi.stanford.edu/data/
Scleractinia	Acroporidae	Acropora	tenuis	none	http://www.bio.utexas.edu/research/matz_lab/matzlab/Data_files/aten_july2014.zip
Scleractinia	Acroporidae	Acropora	millepora	Moya et al. 2012	http://www.bio.utexas.edu/research/matz_lab/matzlab/Data_files/amil_july2014.zip
Scleractinia	Acroporidae	Acropora	digitifera	Shinzato et al. 2011	http://marinegenomics.oist.jp/genomes/downloads?project_id=3
Scleractinia	Astocoeniidae	Madracis	auretenra	none	http://people.oregonstate.edu/~meyere/data.html
Scleractinia	Faviidae	Platygyra	carnosus	Sun et al. 2013	http://www.comp.hkbu.edu.hk/~db/PcarnBase/
Scleractinia	Faviidae	Platygyra	daedalea	none	http://people.oregonstate.edu/~meyere/data.html
Scleractinia	Fungiidae	Fungia	scutaria	Kitchen et al. 2015	http://people.oregonstate.edu/~meyere/data.html
Scleractinia	Merulinidae	Orbicella	faveolata	Anderson et al. 2016	https://peerj.com/articles/1616/#supplemental-information
Scleractinia	Montastraeidae	Montastraea	cavernosa	Kitchen et al. 2015	http://people.oregonstate.edu/~meyere/data.html
Scleractinia	Mussidae	Pseudodiploria	strigosa	none	http://people.oregonstate.edu/~meyere/data.html
Scleractinia	Pocilloporidae	Pocillopora	damicornis	Traylor-Knowles et al. 2011	http://cnidarians.bu.edu/PocilloporaBase/cgi-bin/pdamdata.cgi
Scleractinia	Pocilloporidae	Seriatopora	hystrix	Kitchen et al. 2015	http://people.oregonstate.edu/~meyere/data.html
Scleractinia	Pocilloporidae	Stylophora	pistillata	Maor-Landaw et al. 2014	http://data.centrescientifique.mc/Data/
Scleractinia	Poritidae	Porites	astreoides	Kenkel et al. 2013	http://www.bio.utexas.edu/research/matz_lab/matzlab/Data_files/pastreoides_may2014.zip
Scleractinia	Poritidae	Porites	lobata	none	https://www.ncbi.nlm.nih.gov/bioproject/356802
Scleractinia	Poritidae	Porites	australiensis	Shinzato et al. 2014	https://www.ncbi.nlm.nih.gov/nuccore?term=236717%5BBioProject%5D
Scleractinia	Acroporidae	Montipora	aequituberculata	this study	https://www.dropbox.com/s/qvq3kus89aflyxf/Maqe.tar.gz?dl=0
Scleractinia	Acroporidae	Montipora	capitata	Frazier et al. 2017	ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE97nnn/GSE97888/suppl/GSE97888_Montiporacapitata_transcriptome.fasta.gz
Scleractinia	Oculinidae	Galaxea	acrhelia	Kenkel and Bay 2017	http://dornsife.usc.edu/labs/carlslab/data/
Scleractinia	Oculinidae	Galaxea	astreata	Kenkel and Bay 2017	http://dornsife.usc.edu/labs/carlslab/data/
Scleractinia	Poritidae	Goniopora	columna	Kenkel and Bay 2017	http://dornsife.usc.edu/labs/carlslab/data/
Scleractinia	Siderastreidae	Siderastrea	siderea	Davies et al. 2016	https://sarahwdavies.wordpress.com/data/

Figure S1: Examples of gene trees constructed for orthologous groups before and after paralog pruning. Paralog pruning was performed to remove duplicate sequences from orthologous groups if they came from a single species and formed a monophyletic clade. The figure shows gene trees for two different orthologous groups before and after pruning. Duplicated sequences from single species are shown in red. In the left orthologous group (ORTHMCL7233) a single duplicated sequence from *Galaxia acrhelia* was removed. The longer of the two sequences (Gacrhelia_16231.p1) was retained. In the right orthologous group, duplicate sequences formed monophyletic clades six species. In each of these cases, only the longest sequence was retained.

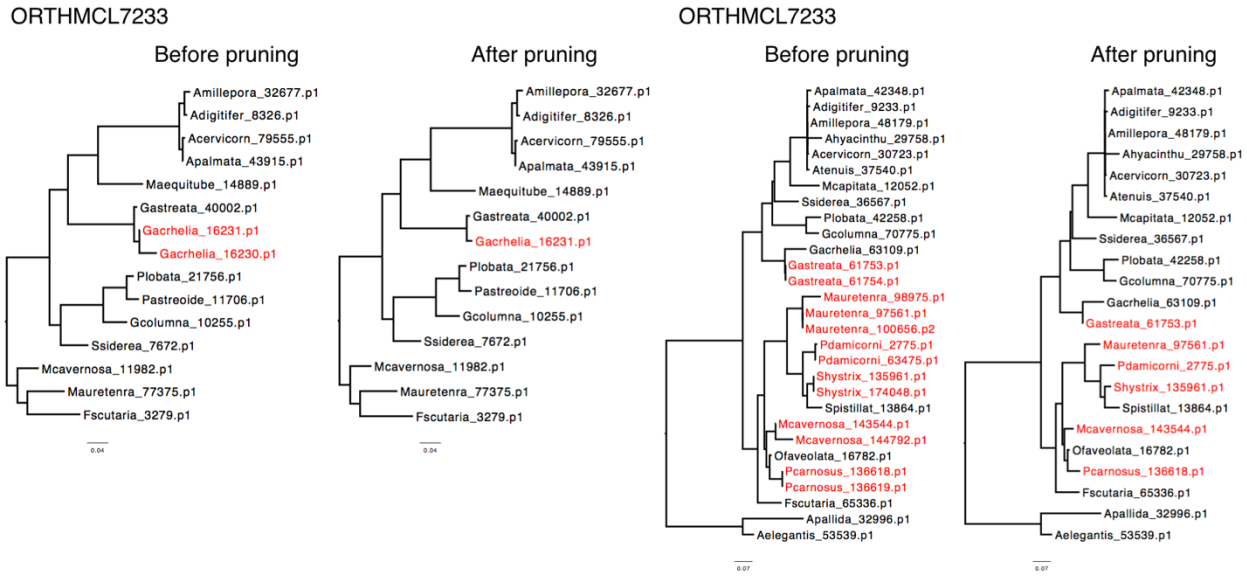


Figure S2: Categorization of all overlapping amino acid substitutions observed between all tested lineage pairs. An overlapping substitution is defined as an inferred amino acid change that occurred at the same position independently in the lineages leading to the common ancestor of the two indicated clades. To simplify comparisons, horizontal clades are labeled based on sisterhood to clades with vertical transmission (Fig. 1). Each overlapping substitution was classified into one of four categories: convergent substitutions (least frequent; salmon) are changes from different amino acids to the same amino acid; parallel substitutions (second most frequent; green) are changes from the same amino acid to the same new amino acid; divergent substitutions (most common; teal) are changes from the same amino acid to a different one; 'all different' substitutions (third most common; purple) are changes from different amino acids to different new amino acids. Examples of each type of overlapping substitution are show in in the legend using one letter amino acid codes.

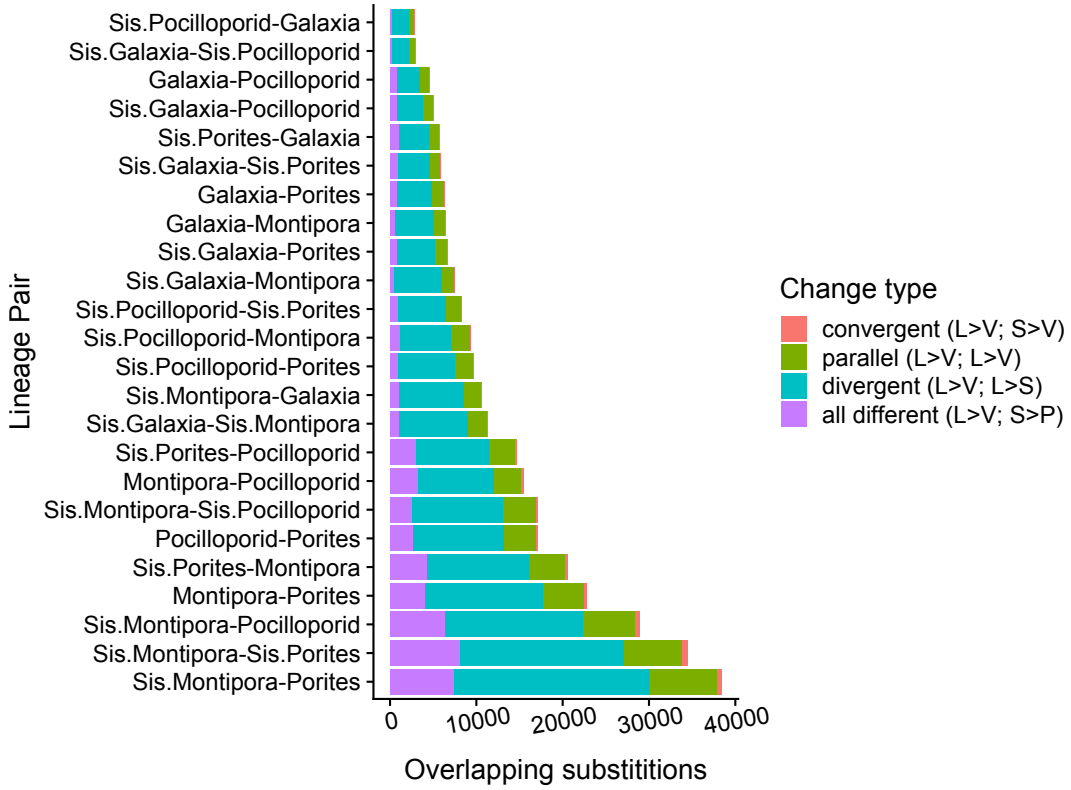


Figure S3: Comparison of frequency of convergent events among genes showing evidence of positive selection. (A) Counts of convergence events in genes showing evidence of positive selection in one or more of the indicated lineages (TRUE) as a proportion of the total gene set for each lineage pair comparison. (B) Percentage of overlapping substitutions that were convergence events in genes also showing evidence of positive selection one or more of the indicated lineages. (C) Boxplot of the percentages in (B) split by phenotype pair, VV: vertical-vertical pairs, VH: vertical-horizontal pairs, HH: horizontal-horizontal pairs.

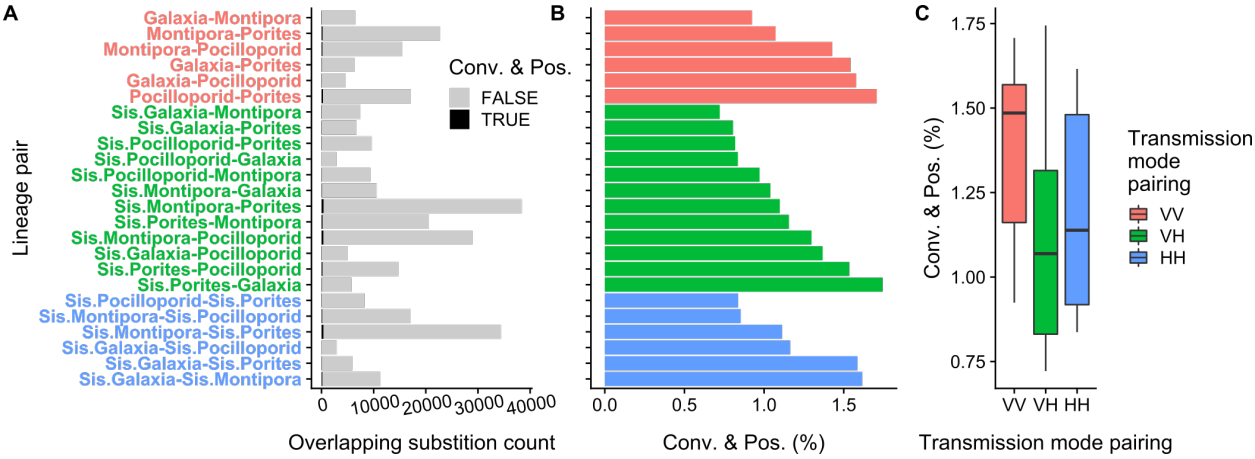


Figure S4: Comparison of the frequency of specific convergence events that were also identified as being positively selected. (A) Counts of convergence events and the proportion which were also the sites exhibiting positive selection in one or more of the indicated lineages (branch site LRT FDR < 0.1 and Bayes Empirical Bayes posterior probability, BEB > 0.8). (B) Percentage of overlapping substitutions that were convergence events in which the specific change was also the site of positive selection in one or more of the indicated lineages. (C) Boxplot of the percentages in (B) split by phenotype pair, VV: vertical-vertical pairs, VH: vertical-horizontal pairs, HH: horizontal-horizontal pairs.

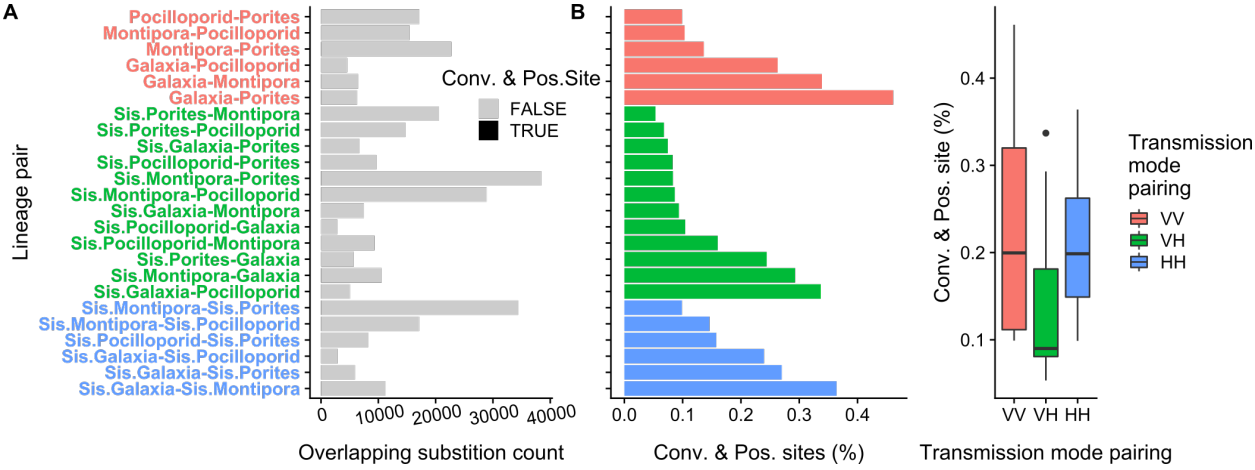


Figure S5: Functional enrichment for genes with convergence events and evidence of positive selection among horizontally transmitting sister clades. (A) Frequency of tested genes showing convergence and positive selection per pair of horizontally transmitting clades. Teal shading indicates the set of genes with at least one convergence event and evidence of positive selection (FDR < 0.1) in at least one of the indicated lineages (TRUE) relative to the total gene set. (B) Gene ontology enrichment across all convergent and positively selected genes identified for any pair of horizontally transmitting clades relative to the global gene list. Significance level is indicated by bolded text. (CC) Cellular Component, (MF) Molecular Function. No ontology terms for Biological Process were significant.

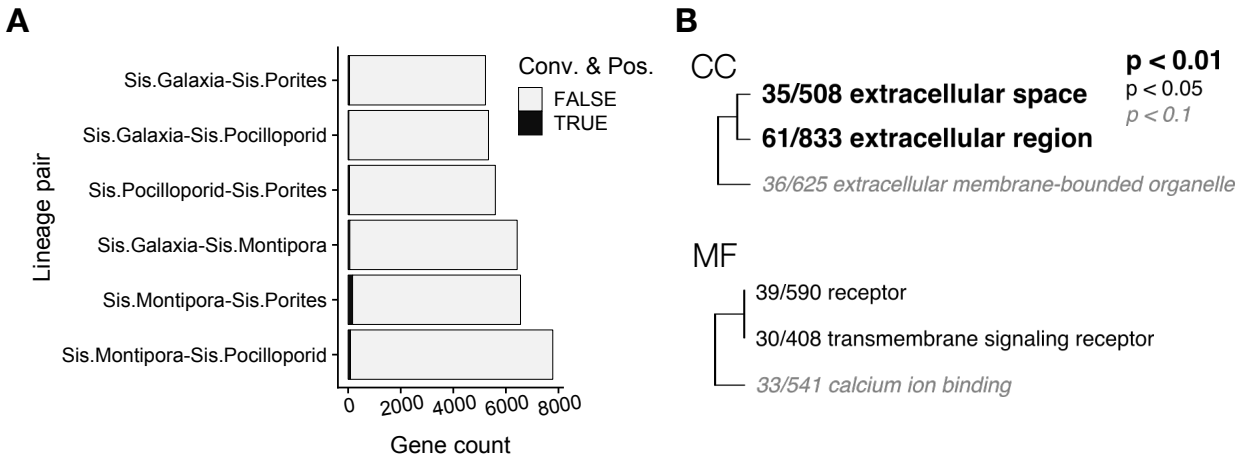
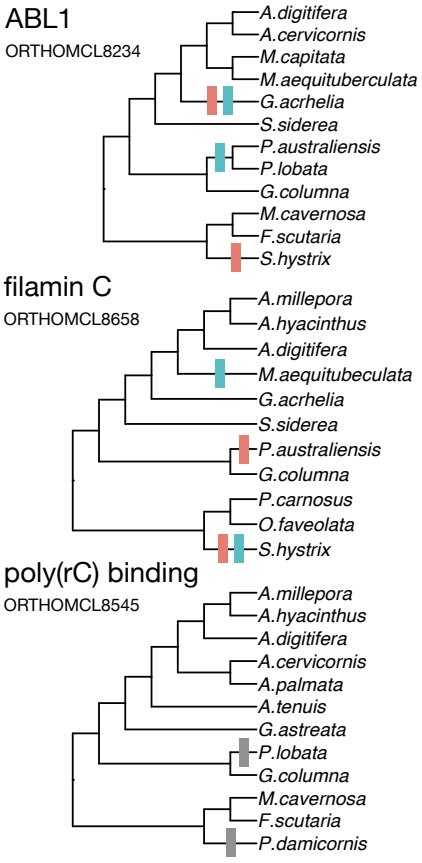


Figure S6. Select genes showing molecular convergence and positive selection at the same site. Left panels show gene trees constructed from nucleotides for each gene. Molecular convergence events that also showed evidence of positive selection are indicated with vertical bars, color-coded to match amino acid positions in the corresponding table. Tables show details of the molecular convergence events and evidence of positive selection: (Pos) amino acid position of convergence event; (Vert.1) first vertical lineage; (Vert. 2) second vertical lineage (see Fig. 1 for list of focal clades); (Anc.1) Ancestral amino acid for first vertical lineage; (Anc.2) Ancestral amino acid for second vertical lineage; (Sub.1) derived amino acid for first vertical lineage; (Sub.2) derived amino acid for second vertical lineage; (BEB all) Bayes Empirical Bayes posterior probability for positive selection at the position for the branch site test including all vertical transmitting lineages as foreground. Derived amino acids with BEB posteriors > 0.8 for tests using individual lineages as foreground are indicated with asterisks.

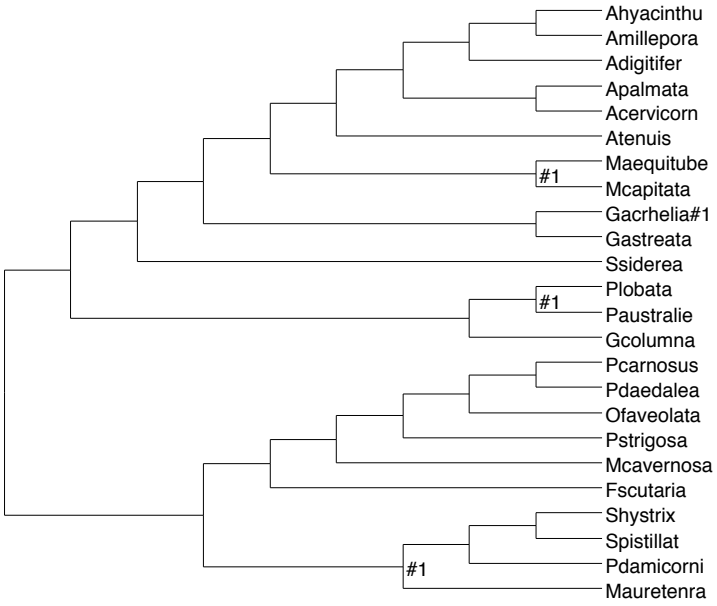


Pos	Vert.1	Vert.2	Anc.1	Anc.2	Sub.1	Sub.2	BEB all
479	Galaxia	Pocilloporid	P	P	S	S	0.98
500	Galaxia	Porites	L	L	V	V	0.84

Pos	Vert.1	Vert.2	Anc.1	Anc.2	Sub.1	Sub.2	BEB all
1518	Porites	Pocilloporid	D	D	P*	P*	0.99
1973	Montipora	Pocilloporid	K	K	S	S*	0.96

Pos	Vert.1	Vert.2	Anc.1	Anc.2	Sub.1	Sub.2	BEB all
193	Porites	Pocilloporid	K	K	S	S	0.94

Figure S7: Labeling of branches for branch site tests. When performing the branch site test, the branch or branches being tested for evidence of positive selection are labeled with "#1". When testing for evidence of positive selection in a clade, we labeled only the branch leading to the common ancestor of that clade. In cases when a clade had only a single species, for example *Galaxia acrhelia*, the branch for that species was labeled.



REFERENCES

1. Kenkel CD, Bay LK. 2018 Exploring mechanisms that affect coral cooperation: symbiont transmission mode, cell density and community composition. *PeerJ* **6**, e6047.

2. Grabherr MG *et al.* 2011 Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol* **29**, 644–652. (doi:10.1038/nbt.1883)

3. Kenkel CD, Bay LK. 2017 Novel transcriptome resources for three scleractinian coral species from the Indo-Pacific. *GigaScience* **6**, 1–4. (doi:10.1093/gigascience/gix074)

4. Polato NR, Vera JC, Baums IB. 2011 Gene discovery in the threatened elkhorn coral: 454 sequencing of the acropora palmata transcriptome. *PLoS ONE* **6**, e28634. (doi:10.1371/journal.pone.0028634)

5. Shinzato C *et al.* 2011 Using the Acropora digitifera genome to understand coral responses to environmental change. *Nature* **476**, 320–3. (doi:10.1038/nature10249)

6. Traylor-Knowles N, Granger BR, Lubinski TJ, Parikh JR, Garamszegi S, Xia Y, Marto JA, Kaufman L, Finnerty JR. 2011 Production of a reference transcriptome and transcriptomic database (PocilloporaBase) for the cauliflower coral, Pocillopora damicornis. *BMC Genomics* **12**, 585. (doi:10.1186/1471-2164-12-585)

7. Lehnert EM, Burriesci MS, Pringle JR. 2012 Developing the anemone Aiptasia as a tractable model for cnidarian-dinoflagellate symbiosis: the transcriptome of aposymbiotic A. pallida. *BMC genomics* **13**, 271. (doi:10.1186/1471-2164-13-271)

8. Moya A *et al.* 2012 Whole transcriptome analysis of the coral Acropora millepora reveals complex responses to CO_2-driven acidification during the initiation of calcification. *Molecular ecology* **21**, 2440–54. (doi:10.1111/j.1365-294X.2012.05554.x)

9. Barshis DJ, Ladner JT, Oliver TA, Seneca FO, Traylor-Knowles N, Palumbi SR. 2013 Genomic basis for coral resilience to climate change. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 1387–92. (doi:10.1073/pnas.1210224110)

10. Kenkel CD, Meyer E, Matz M V. 2013 Gene expression under chronic heat stress in populations of the mustard hill coral (Porites astreoides) from different thermal environments. *Molecular ecology* **22**, 4322–34. (doi:10.1111/mec.12390)

11. Libro S, Kaluziak ST, Vollmer S V. 2013 RNA-seq profiles of immune related genes in the staghorn coral Acropora cervicornis Infected with white band disease. *PLoS ONE* **8**, 1–11. (doi:10.1371/journal.pone.0081821)

12. Sun J, Chen Q, Lun JCY, Xu J, Qiu JW. 2013 PcarnBase: Development of a Transcriptomic Database for the Brain Coral Platygyra carnosus. *Marine Biotechnology* **15**, 244–251. (doi:10.1007/s10126-012-9482-z)

13. Maor-Landaw K, Karako-Lampert S, Ben-Asher HW, Goffredo S, Falini G, Dubinsky Z, Levy O. 2014 Gene expression profiles during short-term heat stress in the red sea coral Stylophora pistillata. *Global Change Biology* **20**, 3026–3035. (doi:10.1111/gcb.12592)

14. Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, Smirnova T, Grigoriev I V., Dubchak I. 2014 The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Research* **42**, 26–31. (doi:10.1093/nar/gkt1069)

15. Shinzato C, Inoue M, Kusakabe M. 2014 A snapshot of a coral 'holobiont': A transcriptome assembly of the scleractinian coral, Porites, captures a wide variety of genes from both the host and symbiotic zooxanthellae. *PLoS ONE* **9**. (doi:10.1371/journal.pone.0085182)

16. Kitchen SA, Crowder CM, Poole AZ, Weis VM, Meyer E. 2015 De novo assembly and characterization of four anthozoan (phylum Cnidaria) transcriptomes. *G3 Genes Genomes Genetics* **5**, 2441–2452. (doi:10.1534/g3.115.020164)

17. Anderson DA, Walz ME, Weil E, Tonellato P, Smith MC. 2016 RNA-Seq of the Caribbean reef-building coral *Orbicella faveolata* (Scleractinia-Merulinidae) under bleaching and disease stress expands models of coral innate immunity. *PeerJ* **4**, e1616. (doi:10.7717/peerj.1616)

18. Davies SW, Marchetti A, Ries JB, Castillo KD. 2016 Thermal and pCO2 stress elicit divergent transcriptomic responses in a resilient coral. *Frontiers in Marine Science* **3**, Article 122.