bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

Groves B. Dixon[a,1] and Carly D. Kenkel[b,2]

[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]Email: grovesdixon@gmail.com

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

KEYWORDS: intracellular symbiosis, *Scleractinia*, transcriptomes, molecular evolution, dN/dS

DATA ARCHIVAL LOCATION: Raw sequencing data generated for this study have been uploaded to NCBI's SRA: PRJNA395352. All bioinformatic scripts and input files can be found at https://github.com/grovesdixon/convergent_evo_coral.

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Heritable symbioses are thought to be important for the maintenance of mutually beneficial relationships (1), and for facilitating major transitions in individuality, such as the evolution of the eukaryotic cell (2, 3). In stony corals, vertical transmission has evolved repeatedly (4), providing a unique opportunity to investigate the genomic basis of this complex trait. We conducted a comparative analysis of 25 coral transcriptomes to identify orthologous genes exhibiting both signatures of positive selection and convergent amino acid substitutions in vertically transmitting lineages. The frequency of convergence events tends to be higher among vertically transmitting lineages, consistent with the proposed role of natural selection in driving the evolution of convergent transmission mode phenotypes (5). Of the 10,774 total orthologous genes identified, 403 exhibited at least one molecular convergence event and evidence of positive selection in at least one vertically transmitting lineage. Functional enrichments among these top candidate genes include processes previously implicated in mediating the coral-*Symbiodiniaceae* symbiosis including endocytosis, immune response, cytoskeletal protein binding and cytoplasmic membrane-bounded vesicles (6). We also identified 100 genes showing evidence of positive selection at the particular convergence event. Among these, we identified several novel candidate genes, highlighting the value of our approach for generating new insight into the mechanistic basis of the coral symbiosis, in addition to uncovering host mechanisms associated with the evolution of heritable symbioses.

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

For organisms that engage in symbiosis, the mode in which symbionts are transmitted to the next host generation is a major factor governing the ecological and evolutionary dynamics of the relationship across multiple scales of biological organization. For example, transmission mode is known to influence genome size and content, cooperative interactions between partners, holobiont ecology, and the speciation rates of both partners (7–11). Two transmission modes predominate in nature: offspring can either directly inherit symbionts, typically through the maternal line in the process of vertical transmission, or they can horizontally acquire symbionts from the environment, usually early in their development (reviewed in (12)); although it is important to note that the mode of transmission can change over evolutionary time (13) and mixed-mode transmission is also possible (12). In microbial symbioses, horizontal transmission is the basal state and repeated transitions to vertical transmission may have arisen as a means to further promote host-symbiont cooperation (13–15). Vertical transmission has been hypothesized to play an important role in the maintenance of mutually beneficial symbioses (1), and likely facilitated major evolutionary transitions in individuality, such as the evolution of the eukaryotic cell (2, 3). From the perspective of the symbiont, the genomic consequences of evolving a heritable symbiosis include a reduction in genome size and increased dependence on their hosts due to the loss of functionally redundant genes (3, 10). However, the underlying genetic architecture facilitating evolution of a heritable symbiosis from the perspective of the host remains unresolved.

The evolution of vertical transmission is predicted to be correlated with the evolution of host control mechanisms (16) and theory predicts a high rate of mutation in genes responsible for the host-symbiont fitness interaction (17). Selection on mechanisms critical for the establishment

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and maintenance of a horizontally transmitted symbiosis, such as cell surface molecules mediating inter-partner recognition, is likely also relaxed (12). Among metazoan hosts, diverse behavioral, developmental and physiological mechanisms are known to facilitate the vertical transmission of microbial endosymbionts (13, 16), yet there is also some evidence for phenotypic convergence. For example, plant-sucking stinkbugs and lice require microbial gut symbionts to facilitate digestion of sap and blood, respectively, but both have evolved additional specialized organs for housing bacteria along the female reproductive tract for the transmission of symbionts to eggs (16, 18). Convergent evolution at the phenotypic level is often the result of similar changes at the genomic level (19, 20) and comparative analyses have facilitated understanding of the genetic basis of convergently evolved phenotypes in diverse taxa (19, 21, 22). Therefore, by comparing vertically transmitting lineages with their closest horizontally transmitting relatives it may be possible to identify candidate genes involved in the evolution of convergent transmission mode phenotypes.

Reef-building corals exhibit both horizontal and vertical transmission of their obligate intracellular *Symbiodiniaceae* symbionts, offering an ideal opportunity to utilize such a comparative approach to identify candidate genes involved the evolution of symbiont transmission mode. The majority of coral species acquire their symbionts from the environment early in their development, but vertical transmission is exhibited by species in multiple different lineages, indicating that transmission mode has evolved independently at least four times (4, 23). Yet there is also significant morphological, physiological and ecological trait variation across the coral phylogeny (24), which can confound a comparative approach. In corals, transmission mode is often correlated with reproductive mode as coral species which broadcast spawn gametes tend to exhibit horizontal transmission, while species that internally brood larvae largely transmit

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

93 symbionts vertically (4). However, the association is not perfect; some *Porites* spp. and all
94 known *Montipora* spp. have convergently evolved to broadcast spawn eggs which contain
95 *Symbiodiniaceae* (25, 26). We therefore sequenced the transcriptome of the vertically
96 transmitting broadcast spawner, *Montipora aequituberculata*, in addition to mining other
97 publicly available coral sequence resources (Table S1), to compile a set of transcriptomic
98 references in which vertical transmitters could be compared with their closest horizontally
99 transmitting relatives, while also accounting for variation in other life-history traits (Fig. 1).
100 From this dataset, we inferred orthologous groups and identified genes showing both signatures
101 of positive selection and convergent amino acid substitutions (overlapping amino acid changes
102 resulting from independent amino acid substitutions at the same position in two or more
103 lineages). We found that the frequency of molecular convergence tended to be higher among
104 vertically transmitting lineages and although top genes are enriched for biological processes
105 previously implicated in the coral-algal symbiosis, we also identify several novel candidates,
106 generating new insight into the mechanistic basis of this relationship.
107
108 **Results & Discussion**
109 *Ortholog identification*
110 To examine molecular convergence and positive selection, we compared homologous
111 coding sequences from transcriptomic data of 25 coral species. First, protein coding sequences
112 were predicted from the transcriptomic data based on open reading frames and sequence
113 homology to known proteins (27) and protein domains (28), and *FastOrtho* (29) was used to
114 assign sequences to preliminary orthologous groups (N = 106,300 groups). A subset of 1,196
115 single-copy orthologous groups with at least 20 of the 28 taxa represented was used to construct

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a species tree (Fig. 1), which recapitulates known relationships reported in earlier studies using single-gene (23, 30) and multi-gene phylogenies (31). We then identified putative single-copy orthologs (groups with only a single representative sequence from each species) from the initial set of 20,563 orthologous groups for which at least 7 (25%) of the species were represented. Of these, 9,794 were truly single copy, whereas 10,769 had multiple sequences for one or more species. Two biological explanations for this observation are gene duplication events subsequent to the relevant speciation event, or transcript isoforms of the same gene (32). Transcript isoforms are more likely given the nature of the dataset, but in either case, any sequence from these monophyletic groupings can be appropriately compared to those from other species. Therefore, rather than eliminate all orthologous groups with multiple sequences, we applied a filtering approach similar to that described by (32) to retain an additional 3,298 of the 10,769 multiple sequence orthologs. Specifically, we constructed gene trees from the protein alignments and pruned away all but the longest of multiple sequences from single species that formed monophyletic clades (Fig. S1; see Methods). In this way, we identified a total of 13,092 total single copy orthologs. Orthologs were then aligned using MAFFT (33) and reverse translated into codon sequences using Pal2Nal (34).

Orthologs were further quality filtered based on monophyly of known clades. Individual gene trees were constructed from nucleotide alignments of each single-copy ortholog and checked for monophyly of known clades (Fig. 1, 1-8 and Robusta/Complexa). All species fell within their expected clades in 58% of the gene trees. If a single sequence fell outside of its expected clade or clades, that sequence was removed and the ortholog was retained (27% of orthologous groups). If more than one sequence fell outside its expected clade the ortholog was removed (15% of orthologous groups). In total, this left 119,049 sequences (mean species per

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

orthologous group = 10.7) comprising 11,130 orthologous groups, hereafter referred to as genes, which were used for the ancestral reconstruction and branch-site tests. Genes with fewer than 5 representative sequences were also removed, resulting in a final total of 10,774 genes.

Evidence of positive selection and molecular convergence

For each orthologous nucleotide alignment, PAML (35) was used to reconstruct the ancestral amino acid at each node in the species tree and identify the amino acid changes that occurred along the branches of the tree. We focused our analysis on eight clades (four with vertical transmission and four with horizontal transmission), and identified all overlapping substitutions, or independent substitutions occurring at the same position between the branches leading directly to these clades' most recent common ancestors (Fig. 1). We classified substitutions according to the type of change observed: parallel substitutions refer to the same derived amino acid evolving from the same ancestral amino acid, convergent substitutions refer to same derived amino acid evolving from different ancestral amino acids, divergent substitutions refer to different derived amino acids evolving from the same ancestral amino acid and 'all different' refer to different derived amino acids evolving from different ancestral amino acids. Following (36), we consider both parallel and convergent substitutions to be indicative of molecular convergence.

Among the vertical transmitters, we identified 8,952 amino acid positions exhibiting either parallel (n=8,877) or convergent (n=75) substitutions in at least two lineages (ancestral reconstruction posterior estimate > 0.8, Fig. 2A, Fig. S2). The convergence events occurred in 4,117 out of 10,774 total genes in the dataset, with an average of 0.71 convergent sites identified per gene (median = 0; Fig. 2B). Of the four possible types of overlapping substitutions, convergent substitutions were by far the least frequent (Fig. 2A; Fig. S2). The most common

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

type was divergent substitutions. The two remaining types, parallel and 'all different' occurred with roughly similar frequency (Fig. 2A). Across the entire dataset, 11% of overlapping substitutions were classified as molecular convergence (convergent or parallel).

In addition to quantifying molecular convergence, we also tested for evidence of positive selection in each vertically transmitting lineage and for all vertically transmitting lineages at once using the branch-site models in PAML (35). We found evidence of positive selection in 954 genes (LRT test FDR<0.1 in at least one branch-site test, Table S2) and many instances in which molecular convergence and positive selection were detected in the same gene (Fig. 3A; Fig. S3). In total, 403 genes showed at least one molecular convergence event between vertically transmitting lineages as well as positive selection in at least one of the lineages (Table S3).

Finally, we took advantage of the fact that the branch site test identifies individual amino acid positions that show evidence of positive selection (37), and identified a list of 100 genes for which the particular convergence event also showed evidence of positive selection in one or both lineages (branch site LRT p-value < 0.05 and BEB > 0.8; Table S4). No ontology enrichments were detected for this reduced group, but annotations were recovered for 66 of the 100 genes.

The frequency of molecular convergence

The probability of parallel molecular evolution in response to selection is predicted to be twice as high as that under neutrality (38). Enforcement of vertical transmission in a laboratory manipulation of an anemone-*Symbiodiniaceae* symbiosis resulted in a host growth advantage, suggesting that the evolution of vertical transmission in Cnidarian symbioses may be favored by selection (39). However, an earlier analysis of genomic convergence among phenotypically convergent marine mammal lineages revealed that convergence was actually highest in terrestrial sister taxa in which no phenotypic convergence was evident, suggesting that the options for

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

adaptive evolution may be limited by pleiotropic constraints (22). To assess the relative frequency of molecular convergence in our dataset we compared the proportion of molecular convergence in overlapping substitutions among three sets of phenotype pairs (vertical transmitters with other vertical transmitters, verticals with horizontals, and horizontals with other horizontals). This helped to control for possible confounding factors such as differences in mutation rate, and varying representation for each species based on data quality that may influence the absolute levels of molecular convergence detected (40).

We found no significant differences among phenotypic pairings in the mean proportion of molecular convergence (Fig. S4), molecular convergence and positive selection (Fig. S3), or specific convergence events in which the sites were also identified as being positively selected (Fig. S5). However, the proportion of convergence events is qualitatively different, and for each of these three data subsets, is higher among vertically transmitting pairs (Figure S3-S5). Although this pattern is tenuous, likely attributable to the small number of possible vertical-vertical comparisons, it is consistent with a proposed role of natural selection in driving the evolution of these convergent transmission mode phenotypes (5).

Functional enrichments among top candidate genes

Coral symbionts reside within host gastrodermal cells, surrounded by a host-derived membrane known as the symbiosome (41). Although the specific genes mediating the establishment and long-term maintenance of this relationship remain unresolved, a number of biological processes are thought to be involved including host-microbe signaling, regulation of the host innate immune response and cell cycle, phagocytosis, and cytoskeletal rearrangement (6). To evaluate whether any of these previously highlighted processes were enriched among the 403 genes exhibiting both signatures of selection and convergent evolution, gene annotations

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were obtained from comparisons against the UniProt Swiss-Prot database (27) and a categorical functional enrichment analysis (FDR<0.1) was performed. Top functional enrichments (FDR<0.01) among biological processes (BP) terms included regulation of developmental growth and cell morphogenesis, and biological adhesion (GO:0048638; GO:0010769; GO007155; GO0022610). Endocytosis (GO:0006897) and immune response (GO:0006955) were also significant (FDR<0.1). Among molecular functions (MF), cytoskeletal protein binding (GO:0008092) was the most significant enrichment (FDR=0.016, Fig. 3B). Extracellular region (GO:0005576) was the most significantly enriched term among cellular components (CC), however, this term was also highlighted in a comparison of horizontally transmitting sister clades (Fig. S6), suggesting that it may be under selection in all corals and not necessarily specific to the evolution of vertical transmission. Additional top CC enrichments (FDR<0.1) specific to vertically transmitting lineages include cell junctions (GO:0030054) and cytoplasmic membrane-bounded vesicles (GO:0016023).

Three individual genes, ABL proto-oncogene 1 (ABL 1, ORTHOMCL8234), filamin C (ORTHOMCL8658), and poly(rC) binding protein 2 (ORTHOMCL8545), warrant additional discussion as they are classified among significantly enriched GO terms in all three ontology categories (BP, CC and MF) and were also among the less than 1% of genes in which the particular convergence event also showed evidence of positive selection (Fig. 4; Table S4). Importantly, none of these candidates have been previously implicated in the host-symbiont relationship in earlier analyses focusing on either coral bleaching, the breakdown of the symbiosis (42–46), or on the establishment of symbiosis in horizontally-transmitting corals (47–49), highlighting the value of the present approach for identifying novel candidate genes potentially underpinning the coral symbiosis.

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

231 ABL 1 is a ubiquitously expressed nonreceptor tyrosine kinase known to be involved in
232 organismal responses to a multitude of signals, including cell adhesion, DNA damage, oxidative
233 stress and cytokines (50). This gene that has likely evolved to serve a variety of context-
234 dependent biological functions, but is known to regulate several immune response phenotypes in
235 mammals including antigen receptor signaling in lymphocytes, and bacterial adhesion to host
236 cells (51–53). Through its role in regulating actin polymerization, ABL 1 is also involved in
237 endocytosis (54), supporting the hypothesis that it may play a key role in mediating the heritable
238 transmission of symbionts. Filamins are another family of actin-binding proteins which also
239 exhibit great functional diversity in their interactions (55). While Filamin C was not identified in
240 earlier functional genomic studies, expression of Filamin A was recently reported to be modified
241 by temperature over the course of a monthly reproductive cycle in *Pocillopora damicornis*, a
242 vertically-transmitting brooding coral (56). Similarly, Filamin B was found to be differentially
243 expressed between symbiotic and aposymbiotic *Aiptasia* anemones (57). Combined, these results
244 suggest an important role for this gene family in the maintenance and transmission of symbionts.
245 Poly(C)-binding proteins also exhibit substantial functional diversity, but they are involved in
246 transcriptional and translational regulation in addition to acting as structural components in
247 DNA-protein complexes (58). Interestingly, poly(rC) binding protein 2 is a negative regulator of
248 mitochondrial antiviral signaling protein (MAVS), a critical component of innate antiviral
249 immunity, where overexpression has been shown to reduce, and knockdown to increase, cellular
250 responses to viral infection (59). MAVS interacts with RIG-I-like (RLR) pattern recognition
251 receptors, which are located in the cytoplasm, to identify foreign RNA (60). However, they have
252 also been shown to function in defense against some bacterial pathogens (60, 61), suggesting that

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

253 regulation of poly(rC) binding protein 2 could be involved in suppressing host innate immune
254 responses against intracellular symbionts.

255
256 *Conclusions*

257 Climate change and other anthropogenic processes threaten corals because of the
258 sensitivity of the coral-dinoflagellate symbiosis to environmental stress (62, 63). Significant
259 work has gone into investigating the breakdown of this relationship in the process known as
260 'coral bleaching' over the past three decades, yet fundamental questions remain unresolved,
261 including a complete understanding of the genomic architecture underpinning the host-symbiont
262 relationship (6, 64). Here, rather than asking about molecular mechanisms correlated with the
263 breakdown of the coral symbiosis, we investigated a factor predicted to reinforce it: the evolution
264 of vertical symbiont transmission. While the genes identified here represent promising
265 candidates for further study, it is important to note that they likely represent only a fraction of the
266 molecular changes involved in the evolution of symbiont transmission mode as there are
267 alternate pathways to achieve the same phenotypic outcome that do not require changes at the
268 level of the coding sequence (65). Increasing genomic resources will facilitate a deeper
269 understanding of such alternative mechanisms, and the concurrent development of more
270 advanced genetic tools for manipulating the coral (66) and other Cnidarian model symbioses (67,
271 68) will facilitate quantification of the precise phenotypic effects of these novel genes, as well as
272 of changes in their sequence, contributing to a greater understanding of the cellular and
273 molecular mechanisms underpinning this specific relationship, and necessary for the evolution of
274 a heritable symbiosis.

275
276

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (5) 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 (69) 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

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

300 are 'holobionts' comprised of host, *Symbiodiniaceae* and other microbial components, resulting
301 assemblies were filtered to identify the host component following the protocol described in (70).
302 *Additional transcriptomic resources*
303 Transcriptomic data from 25 species of Scleractinia (stony corals) and 3 species of
304 Actiniaria (anemones) were downloaded from the web (Table S1; (71); (72); (73); (74); (75);
305 (42); (76); (77); (78); (79); (80); (81); (82); (83); (84); (70)).
306 *Protein sequence prediction*
307 To prepare sequences for protein sequence prediction, we first modified sequence
308 definition lines for each transcriptome to include the species name and an arbitrary sequence
309 number. To remove highly similar isoforms, we used cd-hit (85) to cluster sequences with a
310 sequence identity threshold of 0.98, alignment coverage for the longer sequence at least 0.3 and
311 alignment coverage of the shorter sequence at least 0.3. For each resulting cluster, we retained
312 only the longest sequence.
313 Protein coding sequences were predicted from the transcriptomic data based on open
314 reading frames and sequence homology to known proteins and protein domains. Protein
315 prediction steps were implemented with Transdecoder (86). First, the longest open reading
316 frames (ORFs) were identified using a minimum amino acid length of 100. Then protein
317 sequences were predicted from the longest ORFs based on blastp alignments against the
318 Swissprot database (27) and protein domains identified with scanHmm in HMMER version
319 3.1b2 (28). The resulting coding sequence predictions were used for all downstream analyses.
320 The predicted protein and coding sequences are available on github:
321 https://github.com/grovesdixon/transcriptomes_convergent_evo_coral.git.
322 *Ortholog assignment*

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Predicted coding sequences were assigned to orthologous groups using FastOrtho, an implementation of OrthoMCL (29) available through Pathosystems Resource Integration Center (PATRIC) web resources (87)(http://enews.patricbrc.org/fastortho/). We ran FastOrtho using reciprocal blastp results with an e-value cutoff of 1e-10, excluding hits with alignment lengths less than 75% of subject sequences.

Construction of species tree

To construct a species tree, we used a subset of 1,196 single-copy orthologous groups with at least 20 of the 28 taxa represented. The codon sequence alignments were concatenated in phylip format for input into RAxML (88). The species tree was generated with the rapid bootstrapping algorithm (100 iterations) using the GTRGAMMA model and three anemone species were used as an outgroup. Trees were visualized using Dendroscope (89) and Figtree http://tree.bio.ed.ac.uk/software/figtree/.

Paralog pruning

When putative paralogs from the same taxon were monophyletic, all but the longest sequences were removed. This was done for an initial set of 20,563 orthologous groups for which at least 7 (25%) of the species were represented. Protein sequences for these orthologs were aligned with MAFFT using localpair (33) and gene trees were constructed using FastTree (90). At this point, sequences from the three anemone species were removed, and were not used for any further analyses. We used the biopython module Phylo (91) to identify gene trees for which multiple sequences from single species formed monophyletic groups. Removal of these sequences allowed us to include many more orthologous groups as single-copy orthologs (9,794 single copy orthologs prior to pruning, 13,092 after pruning). After pruning, putative single-copy orthologs were reverse translated into codon sequences using Pal2Nal (34).

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

346 *Phylogenetic ortholog filtering*

347 Orthologous groups were further quality filtered based on monophyly of known clades.
348 Here we constructed gene trees from nucleotide alignments of each single-copy ortholog. We
349 checked these trees for monophyly of known clades (Genus *Acropora*, Genus *Montipora*, Genus
350 *Galaxia*, Genus *Porites*, favid clade with *F. scutaria* as outgroup, pocilloporid clade with *M.*
351 *auretenra* as outgroup, complex corals, robust corals), which were corroborated in our species
352 tree (Fig. 1). For 58% of gene trees, all species fell within their expected clades. If a single
353 sequence fell outside of its expected clade or clades, that sequence was removed and the ortholog
354 was retained (27% of orthologous groups). If more than one sequence fell outside its expected
355 clade, the ortholog was removed (15% of orthologous groups).

356 *Ancestral reconstruction and identification of convergent substitutions*

357 We used ancestral reconstructions to infer molecular convergence for the remaining high-
358 quality orthologous groups. For each orthologous nucleotide alignment, the ancestral amino acid
359 was identified at each node in the species tree, as well as the amino acid changes that occurred
360 along the branches of the tree. This analysis was performed with PAML (35), using the species
361 tree as a the guide. Example control files are available on the Github repository
362 (https://github.com/grovesdixon/convergent_evo_coral).

363 From the ancestral reconstruction results, we identified all substitutions that occurred at
364 the same positions in two or more selected lineages (overlapping substitutions). The selected
365 lineages included the branches leading to the common ancestor of four vertical transmitting
366 clades, and their corresponding horizontally transmitting sister clade (eight clades total, Fig. 1).
367 The horizontally transmitting sister clades were included to serve as negative controls, and for
368 normalization of GO enrichment analyses (see below). In cases where a clade was represented by

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a single species, the terminal branch was used as the lineage for that clade (e.g. the two *Galaxia* species, Fig. 1).

Following (36), we considered both parallel and convergent substitutions as molecular convergence. For a given amino acid position, parallel substitutions refer to independent changes to the same amino acid from the same ancestral amino acid. Convergent substitutions refer to independent changes to the same amino acid from different ancestral amino acids. We also recorded all other types of independent changes at the same site (i.e. changes to different amino acids from the same ancestral amino acid, and changes to different amino acids from different ancestral amino acids).

Testing for evidence of positive selection

We tested for evidence of positive selection using the branch-site test in PAML (35). Branch-site tests were performed on each ortholog using codeml with NSsites set to 2 and fix omega set to 1 for the null model, and set to 0 for the alternative model. Example command files and tree files are available on Github (https://github.com/grovesdixon/convergent_evo_coral). When labeling branches tested for evidence of positive selection for a given clade, only the branch leading to the most recent common ancestor of the clade was labeled (Fig. S7). In other words, whenever a vertically transmitting clade had more than one species, we tested for evidence of positive selection in the lineage leading to the common ancestor of the clade, rather than the terminal branches leading to each individual species. We made this choice because it seems likely that mutations enabling a vertical transmission phenotype occurred in the lineage leading to the common ancestor of the clade, in which vertical transmission was presumed to have already evolved. As with the convergence analysis, in cases where a clade was represented by a single species, the terminal branch for that species was labeled as foreground. Branch-site

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

tests were performed for each individual clade, and for all vertically transmitting clades at once. Significance was tested using likelihood ratio tests, and p-values were adjusted to control for false discovery rate using the Benjamini-Hochberg procedure (92). As with our analysis of molecular convergence, we repeated the tests for the horizontally transmitting sister clades to serve as a negative controls and normalization of GO enrichment. It should be noted that a significant result for the branch-site test does not prove that positive selection occurred, it merely provides evidence that it may have occurred. For simplicity, we refer to genes significant for these tests as "positively selected" as in (22).

Annotation of genes of interest

Genes of interest were selected based on an overlap in both evidence of positive selection and convergent substitutions. Genes were annotated based on the SwissProt database and Pfam hits used for protein prediction (e-value < 1e-5, and default parameters for hmmscan). Gene Ontology (GO) associations were applied to each orthologous group based on all SwissProt genes used for prediction of any of its constituent sequences. The GO annotations for these genes were gathered from the Gene Ontology Annotation (GOA) Database (93) ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/UNIPROT/). For cases when sequences in an orthologous group were predicted with multiple different SwissProt hits, the orthologous group was annotated with GO associations from all included SwissProt genes. Some orthologous groups had only Pfam hits. These did not receive GO annotations.

GO enrichment

GO enrichment was performed using Fisher's exact tests on the final set of genes exhibiting overlap in evidence of positive selection in at least one of the branch site tests and had at least one molecular convergence event among the vertically transmitting lineages. A paired

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

415 control analysis was performed for genes with the same signatures among the horizontally
416 transmitting lineages (Fig. S6). To perform fewer total tests, and reduce the effect of false
417 discovery correction, only large GO terms, associated with at least 200 orthologs in our dataset,
418 were tested for enrichment.
419
420
421
422 ACKNOWLEDGEMENTS
423 This work was supported in part by an NSF International Postdoctoral Research Fellowship,
424 DBI-1401165 to CDK. Bioinformatic analyses were carried out using computational resources of
425 the Texas Advanced Computer Center (TACC).
426
427
428 STATEMENT OF AUTHORSHIP: CDK designed research and assembled new reference
429 transcriptome; GBD analyzed convergence and selection; CDK wrote the first draft of the
430 manuscript and both authors contributed to revisions.
431
432
433
434
435
436
437

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

438 REFERENCES

439 1. Frank SA (1996) Host-symbiont conflict over the mixing of symbiotic lineages. *Proc Biol Sci*
440 263(1368):339–344.

441 2. Kiers ET, West SA (2015) Evolving new organisms via symbiosis. *Science* 348(6233):392–
442 394.

443 3. Fisher RM, Henry LM, Cornwallis CK, Kiers ET, West SA (2017) The evolution of host-
444 symbiont dependence. *Nature Communications* 8:15973.

445 4. Hartmann AC, Baird AH, Knowlton N, Huang D (2017) The Paradox of Environmental
446 Symbiont Acquisition in Obligate Mutualisms. *Current Biology* 0(0).
447 doi:10.1016/j.cub.2017.10.036.

448 5. Kenkel CD, Bay LK (2018) Exploring mechanisms that affect coral cooperation: symbiont
449 transmission mode, cell density and community composition. *PeerJ*:in press.

450 6. Davy SK, Allemand D, Weis VM (2012) Cell Biology of Cnidarian-Dinoflagellate Symbiosis.
451 *Microbiol Mol Biol Rev* 76(2):229–261.

452 7. Herre EA, Knowlton N, Mueller UG, Rehner SA (1999) The evolution of mutualisms:
453 exploring the paths between conflict and cooperation. *Trends in Ecology & Evolution*
454 14(2):49–53.

455 8. Sauer C, Stackebrandt E, Gadau J, Hölldobler B, Gross R (2000) Systematic relationships
456 and cospeciation of bacterial endosymbionts and their carpenter ant host species:
457 proposal of the new taxon Candidatus Blochmannia gen. nov. *International Journal of*
458 *Systematic and Evolutionary Microbiology* 50(5):1877–1886.

459 9. Moran NA, Bennett GM (2014) The Tiniest Tiny Genomes. *Annual Review of Microbiology*
460 68(1):195–215.

461 10. Bennett GM, Moran NA (2015) Heritable symbiosis: The advantages and perils of an
462 evolutionary rabbit hole. *PNAS* 112(33):10169–10176.

463 11. Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable
464 bacterial symbionts. *Annu Rev Genet* 42:165–190.

465 12. Bright M, Bulgheresi S (2010) A complex journey: transmission of microbial symbionts.
466 *Nature Reviews Microbiology* 8(3):218–230.

467 13. Sachs JL, Skophammer RG, Regus JU (2011) Evolutionary transitions in bacterial symbiosis.
468 *PNAS* 108(Supplement 2):10800–10807.

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

14. Sachs JL, Mueller UG, Wilcox TP, Bull JJ (2004) The Evolution of Cooperation. *The Quarterly Review of Biology* 79(2):135–160.

15. West SA, Fisher RM, Gardner A, Kiers ET (2015) Major evolutionary transitions in individuality. *PNAS* 112(33):10112–10119.

16. Frank SA (1996) Host Control of Symbiont Transmission: The Separation of Symbionts Into Germ and Soma. *The American Naturalist* 148(6):1113–1124.

17. Drown DM, Zee PC, Brandvain Y, Wade MJ (2013) Evolution of transmission mode in obligate symbionts. *Evol Ecol Res* 15(1):43–59.

18. Kikuchi Y, et al. (2009) Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. *BMC Biology* 7:2.

19. Stern DL (2013) The genetic causes of convergent evolution. *Nature Reviews Genetics* 14(11):751.

20. Conte GL, Arnegard ME, Peichel CL, Schluter D (2012) The probability of genetic parallelism and convergence in natural populations. *Proc R Soc B* 279(1749):5039–5047.

21. Jones FC, et al. (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484(7392):55.

22. Foote AD, et al. (2015) Convergent evolution of the genomes of marine mammals. *Nature Genetics* 47(3):272–275.

23. Kerr AM (2005) Molecular and morphological supertree of stony corals (Anthozoa: Scleractinia) using matrix representation parsimony. *Biological reviews of the Cambridge Philosophical Society* 80(4):543–558.

24. Madin JS, et al. (2016) The Coral Trait Database, a curated database of trait information for coral species from the global oceans. *Scientific Data* 3:160017.

25. Fadlallah YH (1983) Sexual reproduction, development and larval biology in scleractinian corals. *Coral Reefs* 2(3):129–150.

26. Richmond RH, Hunter CL (1990) Reproduction and recruitment of corals: comparisons among the Caribbean, the Tropical Pacific, and the Red Sea. *Marine Ecology Progress Series* 60(1/2):185–203.

27. Uniprot CT (2016) UniProt: the universal protein knowledgebase. *Nucleic Acids Research* 45(Database issue). doi:10.1093/nar/gkw1152.

28. Eddy SR (2011) Accelerated profile HMM searches. *PLoS Computational Biology* 7(10). doi:10.1371/journal.pcbi.1002195.

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

29. Li L, Stoeckert CJJ, Roos DS (2003) OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Research* 13(9):2178–2189.

30. Kitahara M V, Cairns SD, Stolarski J, Blair D, Miller DJ (2010) A comprehensive phylogenetic analysis of the Scleractinia (Cnidaria, Anthozoa) based on mitochondrial CO1 sequence data. *PloS one* 5(7):e11490.

31. Bhattacharya D, et al. (2016) Comparative genomics explains the evolutionary success of reef-forming corals. *eLife Sciences* 5:e13288.

32. Kocot KM, Citarella MR, Moroz LL, Halanych KM (2013) PhyloTreePruner: A phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. *Evolutionary Bioinformatics* 2013(9):429–435.

33. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30(4):772–780.

34. Suyama M, Torrents D, Bork P (2006) PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* 34:609–612.

35. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24(MI):1586–1591.

36. Zou Z, Zhang J (2015) Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? *Molecular Biology and Evolution* 32(8):2085–2096.

37. Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22(4):1107–1118.

38. Orr HA (2005) The Probability of Parallel Evolution. *Evolution* 59(1):216–220.

39. Sachs JL, Wilcox TP (2006) A shift to parasitism in the jellyfish symbiont Symbiodinium microadriaticum. *Proceedings of the Royal Society of London B: Biological Sciences* 273(1585):425–429.

40. Thomas GWC, Hahn MW (2015) Determining the Null Model for Detecting Adaptive Convergence from Genomic Data: A Case Study using Echolocating Mammals. *Mol Biol Evol* 32(5):1232–1236.

41. Wakefield T, Farmer M, Kempf S (2000) Revised description of the fine structure of in situ "zooxanthellae" genus Symbiodinium. *The Biological Bulletin* 199(1):76–84.

42. Barshis DJ, et al. (2013) Genomic basis for coral resilience to climate change. *Proceedings of the National Academy of Sciences of the United States of America* 110(4):1387–92.

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

43. Desalvo MK, et al. (2008) Differential gene expression during thermal stress and bleaching in the Caribbean coral Montastraea faveolata. *Molecular Ecology* 17(17):3952–3971.

44. DeSalvo MK, Sunagawa S, Voolstra CR, Medina M (2010) Transcriptomic responses to heat stress and bleaching in the elkhorn coral Acropora palmata. *Marine Ecology Progress Series* 402:97–113.

45. Maor-Landaw K, et al. (2014) Gene expression profiles during short-term heat stress in the red sea coral Stylophora pistillata. *Global Change Biology* 20(10):3026–3035.

46. Rodriguez-Lanetty M, Harii S, Hoegh-Guldberg O (2009) Early molecular responses of coral larvae to hyperthermal stress. *Molecular Ecology* 18(24):5101–5114.

47. Voolstra CR, et al. (2009) The host transcriptome remains unaltered during the establishment of coral–algal symbioses. *Molecular Ecology* 18(9):1823–1833.

48. Schnitzler CE, Weis VM (2010) Coral larvae exhibit few measurable transcriptional changes during the onset of coral-dinoflagellate endosymbiosis. *Mar Genomics* 3(2):107–116.

49. Mohamed AR, et al. (2016) The transcriptomic response of the coral Acropora digitifera to a competent Symbiodinium strain: the symbiosome as an arrested early phagosome. *Mol Ecol* 25(13):3127–3141.

50. Wang JYJ (2014) The Capable ABL: What Is Its Biological Function? *Molecular and Cellular Biology* 34(7):1188–1197.

51. Arce KP de, et al. (2010) Synaptic Clustering of PSD-95 Is Regulated by c-Abl through Tyrosine Phosphorylation. *J Neurosci* 30(10):3728–3738.

52. Swimm A, et al. (2004) Enteropathogenic Escherichia coli Use Redundant Tyrosine Kinases to Form Actin Pedestals. *MBoC* 15(8):3520–3529.

53. Huang Y, et al. (2008) The c-Abl tyrosine kinase regulates actin remodeling at the immune synapse. *Blood* 112(1):111–119.

54. Tanos B, Pendergast AM (2006) Abl Tyrosine Kinase Regulates Endocytosis of the Epidermal Growth Factor Receptor. *J Biol Chem* 281(43):32714–32723.

55. Feng Y, Walsh CA (2004) The many faces of filamin: A versatile molecular scaffold for cell motility and signalling. *Nature Cell Biology* 6(11):1034–1038.

56. Crowder CM, Meyer E, Fan T-Y, Weis VM (2017) Impacts of temperature and lunar day on gene expression profiles during a monthly reproductive cycle in the brooding coral Pocillopora damicornis. *Molecular Ecology* 26(15):3913–3925.

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

57. Lehnert EM, et al. (2014) Extensive Differences in Gene Expression Between Symbiotic and Aposymbiotic Cnidarians. *G3: Genes, Genomes, Genetics* 4(2):277–295.

58. Makeyev AV, Liebhaber SA (2002) The poly(C)-binding proteins: a multiplicity of functions and a search for mechanisms. *RNA* 8(3):265–278.

59. You F, et al. (2009) PCBP2 mediates degradation of the adaptor MAVS via the HECT ubiquitin ligase AIP4. *Nature Immunology* 10(12):1300–1308.

60. Brubaker SW, Bonham KS, Zanoni I, Kagan JC (2015) Innate Immune Pattern Recognition: A Cell Biological Perspective. *Annu Rev Immunol* 33:257–290.

61. Chiu Y-H, Macmillan JB, Chen ZJ (2009) RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell* 138(3):576–591.

62. Hughes TP, et al. (2003) Climate Change, Human Impacts, and the Resilience of Coral Reefs. *Science* 301(5635):929–933.

63. Hoegh-Guldberg O, et al. (2007) Coral Reefs Under Rapid Climate Change and Ocean Acidification. *Science* 318(5857):1737–1742.

64. Edmunds PJ, Gates RD (2003) Has Coral Bleaching Delayed Our Understanding of Fundamental Aspects of Coral–Dinoflagellate Symbioses? *BioScience* 53(10):976–980.

65. Prud'homme B, Gompel N, Carroll SB (2007) Emerging principles of regulatory evolution. *PNAS* 104(suppl 1):8605–8612.

66. Cleves PA, Strader ME, Bay LK, Pringle JR, Matz MV (2018) CRISPR/Cas9-mediated genome editing in a reef-building coral. *PNAS* 115(20):5235–5240.

67. Jones VAS, Bucher M, Hambleton EA, Guse A (2018) Microinjection to deliver protein, mRNA, and DNA into zygotes of the cnidarian endosymbiosis model Aiptasia sp. *Scientific Reports* 8(1):16437.

68. Ohdera AH, et al. (2018) Upside-Down but Headed in the Right Direction: Review of the Highly Versatile Cassiopea xamachana System. *Front Ecol Evol* 6. doi:10.3389/fevo.2018.00035.

69. Grabherr MG, et al. (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol* 29(7):644–652.

70. Kenkel CD, Bay LK (2017) Novel transcriptome resources for three scleractinian coral species from the Indo-Pacific. *GigaScience* 6(9):1–4.

71. Polato NR, Vera JC, Baums IB (2011) Gene discovery in the threatened elkhorn coral: 454 sequencing of the acropora palmata transcriptome. *PLoS ONE* 6(12):e28634.

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

72. Shinzato C, et al. (2011) Using the Acropora digitifera genome to understand coral responses to environmental change. *Nature* 476(7360):320–3.

73. Traylor-Knowles N, et al. (2011) Production of a reference transcriptome and transcriptomic database (PocilloporaBase) for the cauliflower coral, Pocillopora damicornis. *BMC Genomics* 12(1):585.

74. 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.

75. 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(10):2440–54.

76. 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(16):4322–34.

77. 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(11):1–11.

78. 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.

79. Maor-Landaw K, et al. (2014) Gene expression profiles during short-term heat stress in the red sea coral Stylophora pistillata. *Global Change Biology* 20:3026–3035.

80. Nordberg H, et al. (2014) The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Research* 42(November 2013):26–31.

81. 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(1). doi:10.1371/journal.pone.0085182.

82. 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(11):2441–2452.

83. 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.

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

630 84. Davies SW, Marchetti A, Ries JB, Castillo KD (2016) Thermal and pCO2 stress elicit
631 divergent transcriptomic responses in a resilient coral. *Frontiers in Marine Science* 3:Article
632 122.

633 85. Li W, Godzik A (2006) Cd-hit: A fast program for clustering and comparing large sets of
634 protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659.

635 86. Haas BJ, et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the
636 Trinity platform for reference generation and analysis. *Nature Protocols* 8(8):1494–1512.

637 87. Wattam AR, et al. (2014) PATRIC, the bacterial bioinformatics database and analysis
638 resource. *Nucleic Acids Research* 42(Database issue):581–591.

639 88. Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of
640 large phylogenies. *Bioinformatics* 30:1312–1313.

641 89. Huson DH, Scornavacca C (2012) Dendroscope 3: An interactive tool for rooted
642 phylogenetic trees and networks. *Systematic Biology* 61(6):1061–1067.

643 90. Price MN, Dehal PS, Arkin AP (2009) Fasttree: Computing large minimum evolution trees
644 with profiles instead of a distance matrix. *Molecular Biology and Evolution* 26(7):1641–
645 1650.

646 91. Talevich E, Invergo BM, Cock PJ, Chapman B a (2012) Bio.Phylo: A unified toolkit for
647 processing, analyzing and visualizing phylogenetic trees in Biopython. *BMC Bioinformatics*
648 13:209.

649 92. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: a Practical and
650 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* 57(1):289–
651 300.

652 93. Huntley RP, et al. (2015) The GOA database: Gene Ontology annotation updates for 2015.
653 *Nucleic Acids Research* 43(D1):D1057–D1063.

654 94. Kerr AM, Baird AH, Hughes TP (2011) Correlated evolution of sex and reproductive mode
655 in corals (Anthozoa: Scleractinia). *Proceedings of the Royal Society of London B: Biological*
656 *Sciences* 278(1702):75–81.

657

658

659

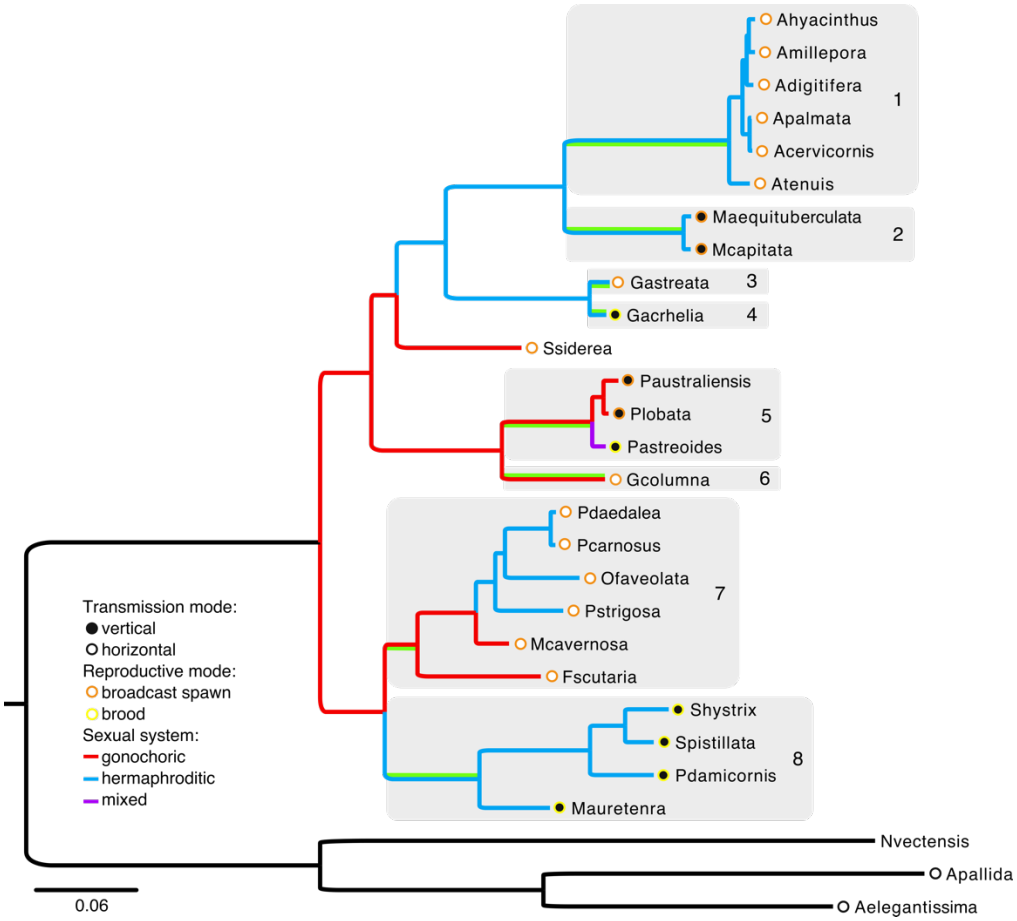
660

661

bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURES & FIGURE LEGENDS

Figure 1. Species tree with phenotypic labels indicating transmission mode, reproductive mode and sexual system (4, 26, 94). Vertically transmitting species are indicated by filled circles at their terminal nodes, horizontally transmitting species with open circles at their terminal nodes. For each clade (1-8), the particular branch examined for convergent substitutions and positive selection is indicated by a green highlight. In each case, this is the branch leading the common ancestor of the clade. Shaded clades were considered when describing overlapping convergence events, referred to as (1) Sister *Montipora*, (2) *Montipora*, (3) Sister *Galaxia*, (4) *Galaxia*, (5) *Porites*, (6) Sister *Porites*, (7) Sister Pocilloporid, (8) Pocilloporid.



bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2. Frequency of convergence events. (A) 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 vertically transmitting clades. 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. (B) Histogram of the number of sites showing molecular convergence (convergent or parallel substitutions) per tested gene (mean = 0.71; median=0).

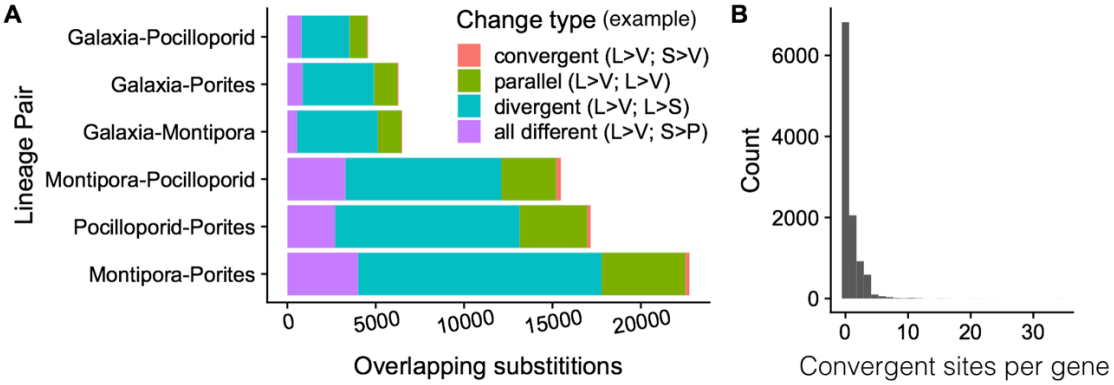
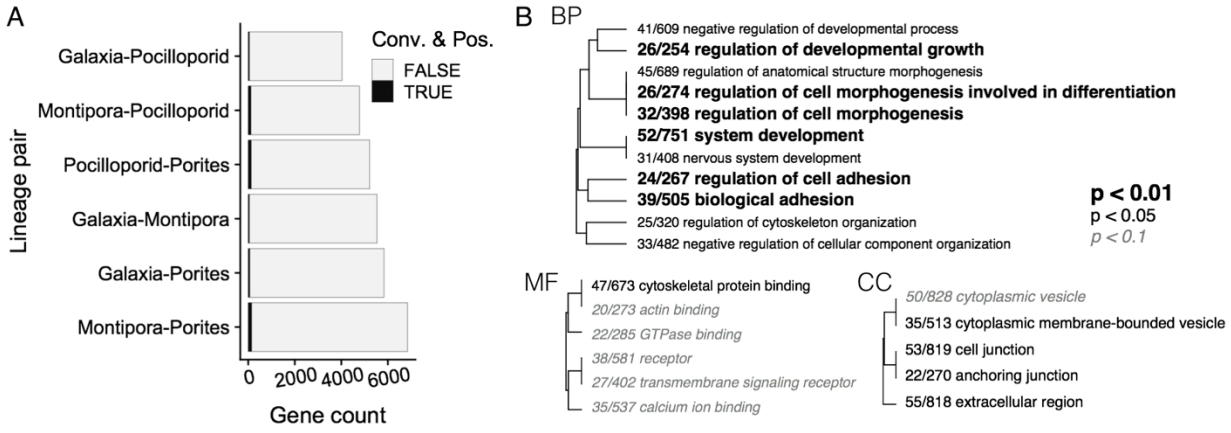
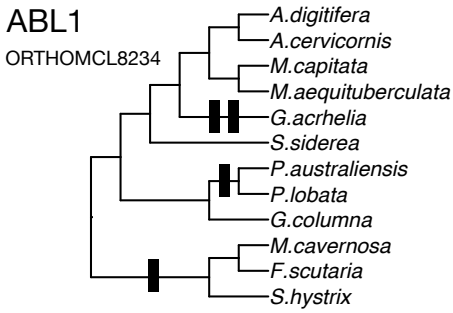


Figure 3. Frequency of genes exhibiting overlap in convergence and positive selection, and results of a categorical functional enrichment analysis of these candidates. (A) Frequency of genes exhibiting both signatures of convergence and positive selection per pair of vertically transmitting clades. Black 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. (B) Gene ontology enrichment across all convergent and positively selected genes identified for any pair of vertically transmitting clades relative to the global gene list. Significance level is indicated by bolded text. (BP) Biological Processes, (CC) Cellular Component, (MF) Molecular Function.

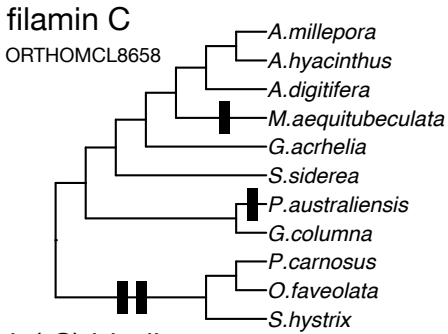


bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

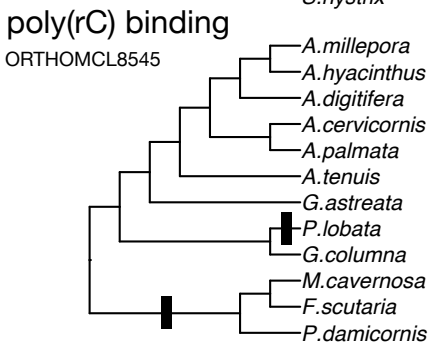
Figure 4. Select genes showing molecular convergence and positive selection at the same site. Left panels show gene trees constructed from nucleotides for each gene. Molecular convergence events that also showed evidence of positive selection are indicated with vertical bars. Tables show details of the 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; (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

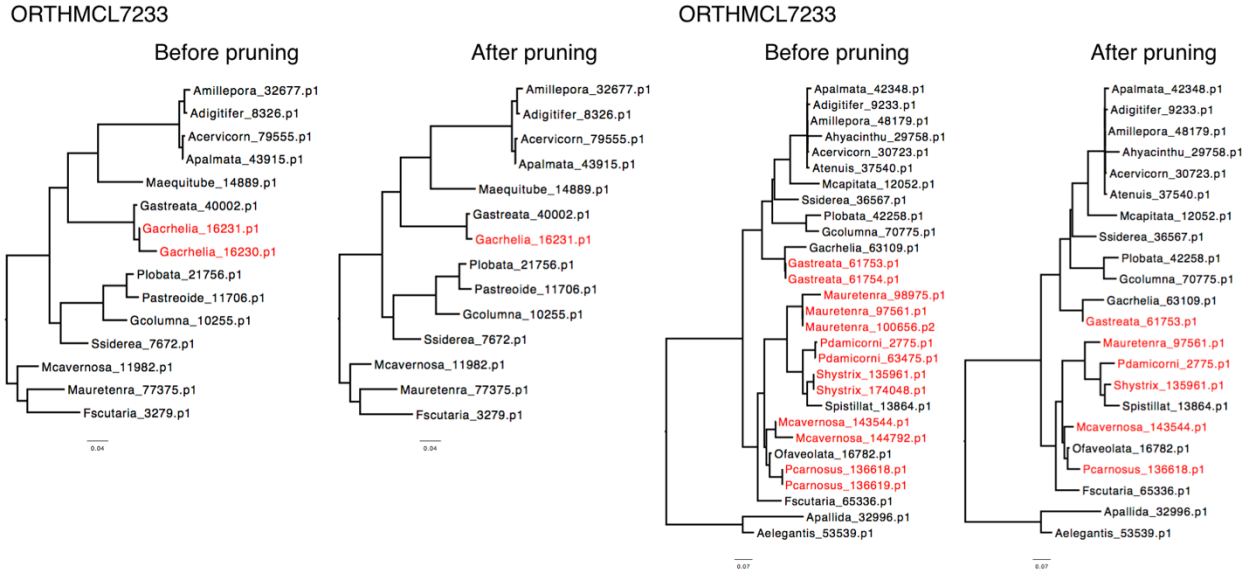
bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SUPPLEMENTARY MATERIAL

Table S1. Sources of reference transcriptomes used for each species.

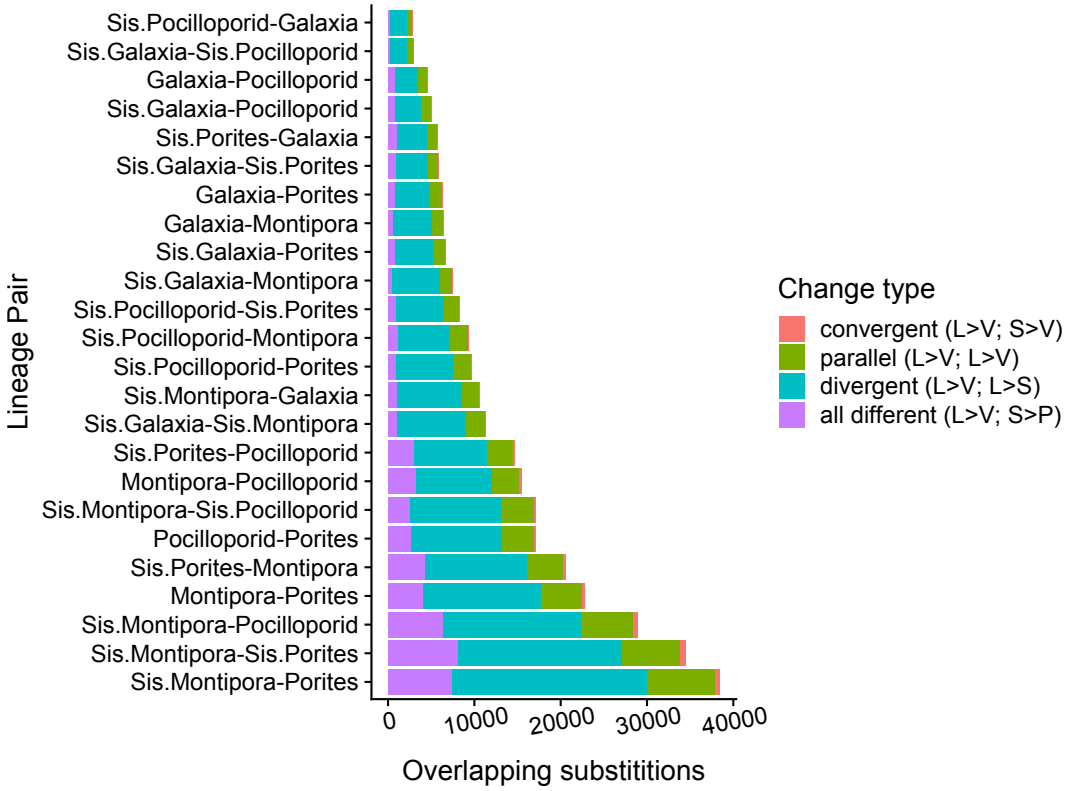
order	family	Genus	species	Citation	URL
Actiniaria	Actiniidae	*Anthopleura*	*elegantissima*	Kitchen et al. 2015	http://people.oregonstate.edu/~meyere/data.html
Actiniaria	Aiptasiidae	*Aiptasia*	*pallida*	Lehnert et al. 2012	http://pringlelab.stanford.edu/project%20files/AposymbioticAiptasiaTranscriptomeGoodLoci.fa.gz
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*	none	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.



bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.



bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. (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.

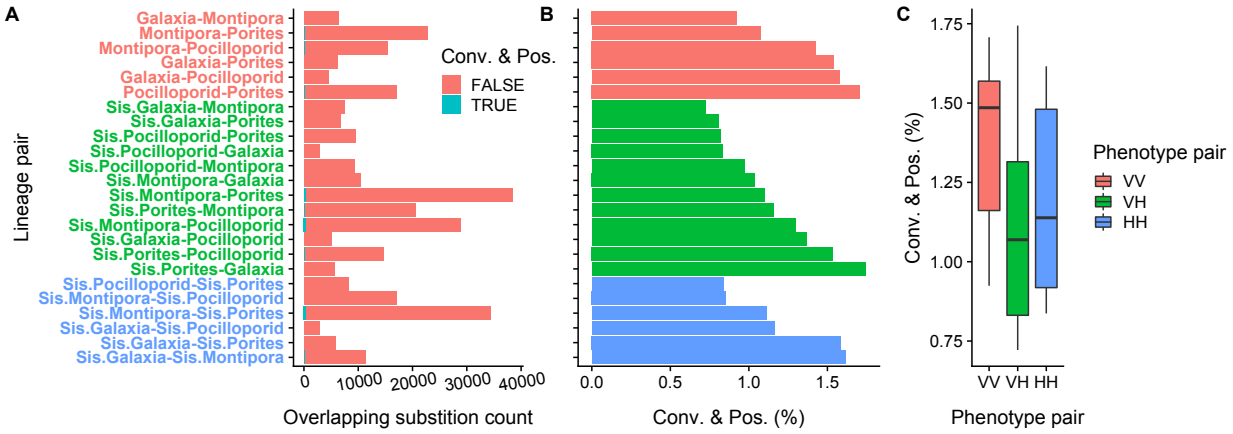
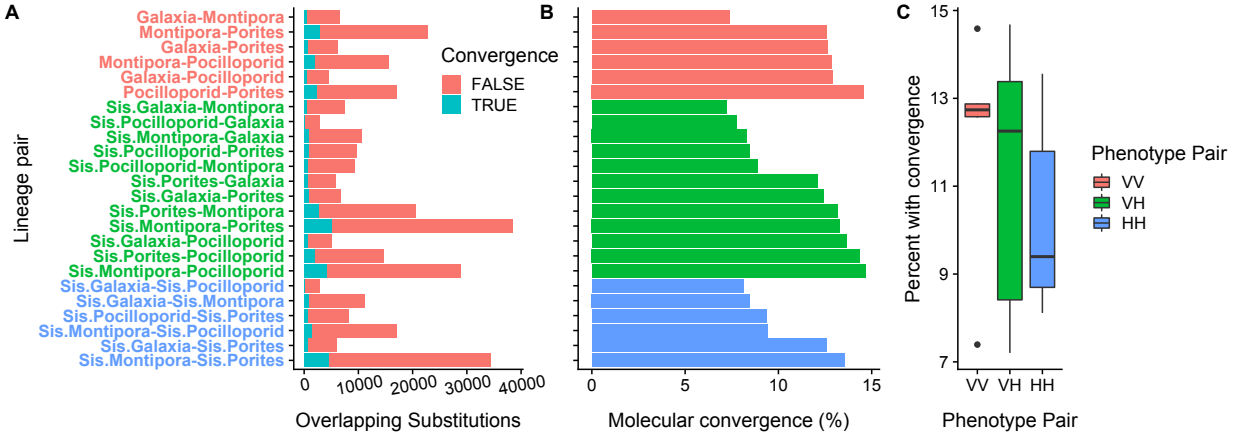


Figure S4: Comparison of the frequency of convergence events among overlapping substitutions. (A) Absolute counts of overlapping substitutions and convergence events for each species pair. (B) Percentage of overlapping substitutions that were convergence events. (C) Boxplot of the percentages in (B) split by phenotype pair, VV: vertical-vertical pairs, VH: vertical-horizontal pairs, HH: horizontal-horizontal pairs.



bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S5: Comparison of the frequency of specific convergence events that were also identified as being positively selected. (A) Counts of convergence events which were also the sites exhibiting positive selection in one or more of the indicated lineages (Branch site test FDR < 0.1 for gene). (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. Note that eight pairs have values of zero. (C) Boxplot of the percentages in (B) split by phenotype pair, VV: vertical-vertical pairs, VH: vertical-horizontal pairs, HH: horizontal-horizontal pairs.

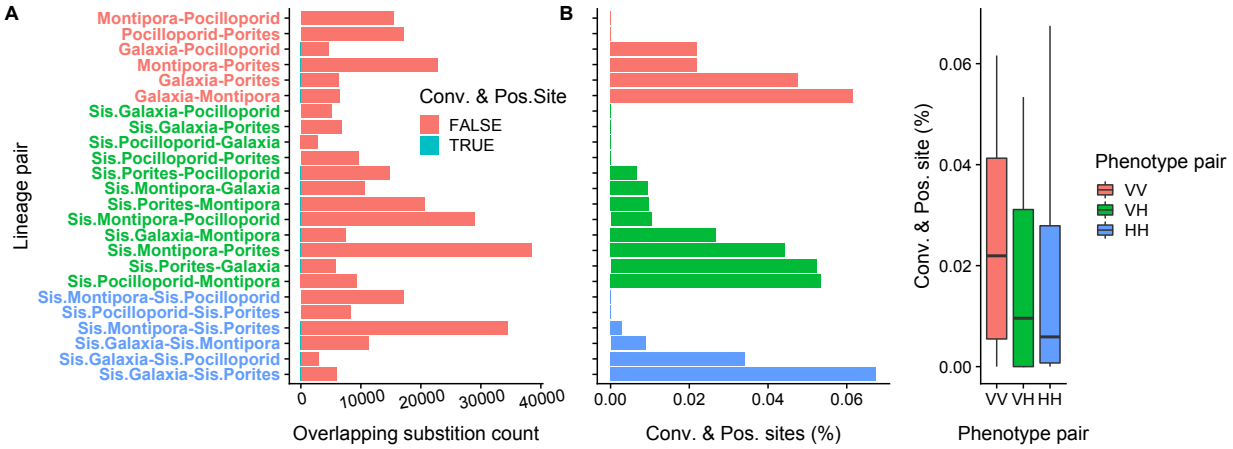
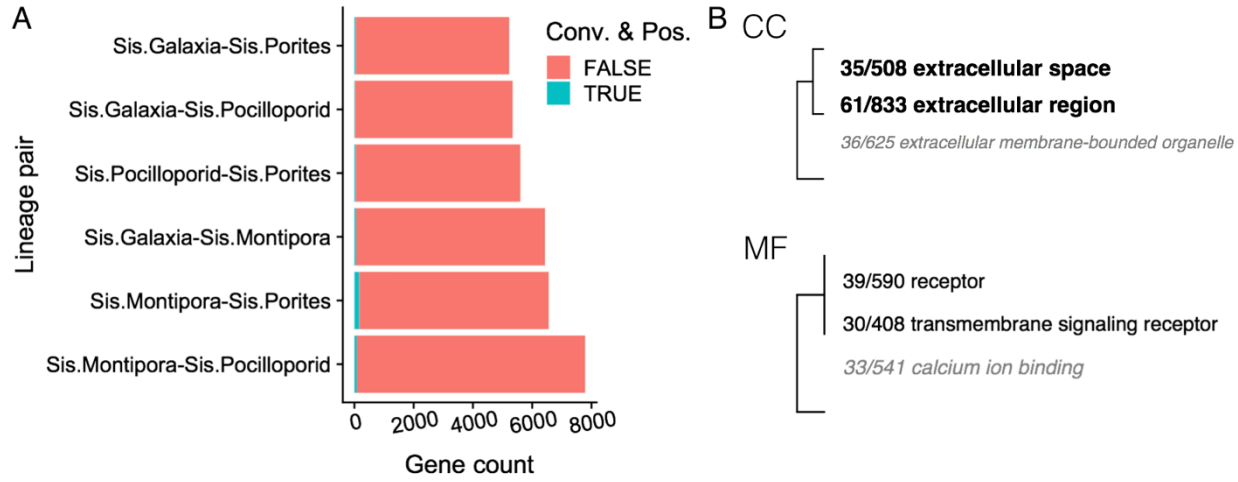


Figure S6: Functional enrichment for genes with convergence events and evidence of positive selection among horizontally transmitting sister clades. (A) Frequency of tested genes showing convergence and positive selection per pair of horizontally transmitting clades. Teal shading indicates the set of genes with at least one convergence event and evidence of positive selection (FDR < 0.1) in at least one of the indicated lineages. (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. Fractions preceding ontology terms indicate … (BP) Biological Processes, (CC) Cellular Component, (MF) Molecular Function. No ontology terms for Biological Process were significant.



bioRxiv preprint doi: https://doi.org/10.1101/489492; this version posted December 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.