**Methods**

*Transcriptomic and computational resources:*

Transcriptomic data from 25 species of Scleractinia (stony corals) and 3 species of Actiniaria (anemones) were downloaded from the web (table S1; Polato et al. 2011; Shinzato et al. 2011; Traylor-Knowles et al. 2011; Lehnert et al. 2012; Moya et al. 2012; Barshis et al. 2013; Kenkel et al. 2013; Libro et al. 2013; Sun et al. 2013; Maor-Landaw et al. 2014; Nordberg et al. 2014; Shinzato et al. 2014; Kitchen et al. 2015; Anderson et al. 2016; Davies et al. 2016; Kenkel and Bay 2017).

*Protein sequence prediction*

To prepare sequences for protein sequence prediction, we first modified sequence definition lines for each transcriptome to include the species name and an arbitrary sequence number. To remove highly similar isoforms, we used cdhit (Li and Godzik 2006) to cluster sequences with a sequence identity threshold of 0.98, alignment coverage for the longer sequence at least 0.3 and alignment coverage of the shorter sequence at least 0.3. For each resulting cluster, we retained only the longest sequence.

Protein coding sequences were predicted from the transcriptomic data based on open reading frames and sequence homology to known proteins and protein domains. Protein prediction steps were implemented with Transdecoder (Haas et al. 2013). First, longest open reading frames (ORFs) were identified using a minimum amino acid length of 100. Then protein sequences were predicted from the longest ORFs based on blastp alignments against the Swissprot database (Uniprot 2016) and protein domains identified with scanHmm in HMMER version 3.1b2 (Eddy 2011). The resulting coding sequence predictions were used for all downstream analyses.

*Ortholog assignment*

Predicted coding sequences were assigned to orthologous groups using FastOrtho, an implementation of OrthoMCL (Li et al. 2003) available through Pathosystems Resource Integration Center (PATRIC) web resources (Wattam et al. 2014; <http://enews.patricbrc.org/fastortho/)>. We ran FastOrtho using reciprocal blastp results with an evalue 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 the 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 (Stamatakis 2014). The species tree was generated with the rapid bootstrapping algorithm (100 iterations) using the GTRGAMMA model and the three anemone species as an outgroup. Trees were visualized using Dendroscope (Huson and Scornavacca 2012) and Figtree <http://tree.bio.ed.ac.uk/software/figtree/>.

*Paralog pruning*

Paralogous sequences resulting from gene duplication can confound phylogenetic inference. One way to avoid this is to analyze orthologous groups with only a single sequence from each taxon, so called single-copy orthologs. However, eliminating all orthologous groups with multiple sequences from a single taxon would severely decrease the scope of the dataset. Also, depending on context, these putative paralogs often do not influence phylogenetic inference. For instance, we frequently observed cases where putative paralogs from the same taxon were monophyletic. Two biological explanations for such cases are gene duplication and transcript isoforms of the same gene (Kocot et al. 2013). In the case of gene duplication, monophyly of the two sequences indicates that the duplication event occurred subsequent to all relevant speciation events, so that either sequence can be appropriately compared to those from other species. The same can be said if the putative paralogs are actually sequence isoforms (Kocot et al. 2013).

With these ideas in mind, we used protein gene trees to trim putative paralogs in a manner similar to that described by Kocot et al. (2013). When putative paralogs from the same taxon were monophyletic, all but the longest sequences were removed. This was done for an initial set of 20563 orthologous groups for which at least 7 (25%) of the species were represented. Protein sequences for these orthologs were aligned with MAFFT using localpair (Katoh and Standley 2013) and gene trees were constructed using FastTree (Price et al. 2009). At this point, sequences from the three anemone species were removed, and were not used for any further analysis. We used the biopython module Phylo (Talevich et al. 2012) 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 (Suyama et al. 2006).

*Phylogenetic ortholog filtering*

Orthologous groups were further quality filtered based on monophyly of known clades. Here we constructed gene trees from nucleotide alignments of each single-copy ortholog. We checked these trees for monophyly of known clades, which were corroborated in our species tree. The clades which were checked are indicated in (Figure S1, monophyletic check groups). For 58% of gene trees all species fell within their expected clades. 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 their expected clades, the ortholog was removed (15% of of orthologous groups).

*Ancestral reconstruction and identification of convergent substitutions*

We used ancestral reconstructions to infer molecular convergence. For each orthologous nucleotide alignment, the ancestral amino acid was identified at each node in the species tree, as well as the amino acid changes that occurred along the branches of the tree. This analysis was performed with PAML (Yang 2007), using the species tree as a the guide. Example control files are available on the Github repository (<https://github.com/grovesdixon/convergent_evo_coral>).

From the ancestral reconstruction results, we identified all substitutions that occurred at the same positions in two or more selected clades. The selected clades included the four vertical transmitting clades, as well as the horizontally transmitting sister clade for each of these (eight clades total). The clades used are shown in (Figure S2 clades). The horizontally transmitting sister clades were included to serve as negative controls, and for normalization of GO enrichment analyses (see below).

Following Zou and Zhang (2015), we consider 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 intendent changes at the same site (ie changes from 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 for positive selection implemented in PAML (Yang 2007). 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>)](https://github.com/grovesdixon/coral_reproductive_evolution)). One aspect of this test is selection and labeling of the particular branch or branches that are being tested for evidence of positive selection, referred to as foreground branches. When testing for positive selection in a given clade, only the branch leading to the most recent common ancestor of the clade was labeled (Supplementary foreground labeling figure3). In cases where a clade was represented by a single species, that terminal branch for that species was labeled as foreground. 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 for each individual species. We made this choice because it seems likely that mutations enabling transmission to a vertical transmitting phenotype are expected to occur in the lineage leading to the common ancestor of the clade, in which is vertical transmission was presumed to have already evolved. Branch-site tests were performed 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 Benjamini-Hochberg procedure (Benjamini and Hochberg 1995). As with our analysis of molecular convergence, we repeated the tests for the horizontally transmitting sister clades to serve as negative controls. It should be noted that a significant result for the branch site test does not prove that positive selection occurred, and merely provides evidence supporting the inference of positive selection. For simplicity, we will refer to genes significant for these tests as “positively selected” as in Foote et al. (2015).

We also used the branch-site tests to identify amino acid sites likely to have experienced positive selection. As a part of each test, positions are assigned a posterior probability that they

*Overlap between evidence of positive selection and convergent substitutions*

Genes of interest were selected based on an overlap for evidence of positive selection and convergent substitutions.

*Gene annotation*

Genes were annotated based on the SwissProt and Pfam hits used for protein prediction. Gene Ontology associations were applied to each orthologous group based on the ID mappings for the SwissProt hits. Orthologous groups with multiple SwissProt hits were annotated with GO associations from both hits. 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 four sets of genes.

First, enrichment tests were performed for genes showing evidence of positive selection in at least one of the branch site tests for positive selection in a vertical transmitting clade: either one of the four tests performed for each clade individually, or the test for positive selection in all the clades. Second, we tested for enrichment for genes that had at least one molecular convergence event among the vertical transmitters. Third we tested for enrichment for genes that overlapped between the first two gene sets. For the final gene set, we took advantage of the fact that the branch site test will identify individual amino acid positions that show evidence of positive selection. For the last gene set were subset the third group by including only genes for which the particular convergent amino acid position showed evidence of positive selection. The counts for these gene sets are given in table S2. For all enrichment analyses, only large GO terms, with at least 50 annotated genes in our dataset, were tested.

**Results**

*Ortholog identification*

The final set of ortholog groups used for input into ancestral reconstruction and branch-site tests total 1,1130 groups, including 119,049 sequences (mean species per orthologous group = 10.7).

*Tests for evidence of positive selection*

We found evidence of positive selection in 1,126 genes.

*Overlap between molecular convergence and positive selection*

We identified many instances in which an amino acid position was significant for positive selection and molecular convergence (Figure 1).

*Frequency of molecular convergence*

Molecular convergence did not occur preferentially between vertical transmitters. The relative frequency of molecular convergence to all overlapping substitutions was not greater for vertical-vertical convergence events than for vertical-horizontal (Figure S4).

*Gene ontology enrichment*

Gene Ontology enrichment analysis identified several categories of genes that enriched for overlap in molecular convergence and positive selection. Notable among these were Cellular Component groups associated with vesicles, such as transport vesicle membrane, cytoplasmic vesicle, and endosome (Figure 2).

Anderson, D. A., M. E. Walz, E. Weil, P. Tonellato, and M. C. Smith. 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.

Barshis, D. J., J. T. Ladner, T. A. Oliver, F. O. Seneca, N. Traylor-Knowles, and S. R. Palumbi. 2013. Genomic basis for coral resilience to climate change. Proc. Natl. Acad. Sci. U. S. A. 110:1387–92.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. 57:289–300.

Davies, S. W., A. Marchetti, J. B. Ries, and K. D. Castillo. 2016. Thermal and pCO2 stress elicit divergent transcriptomic responses in a resilient coral. Front. Mar. Sci. 3:Article 122.

Eddy, S. R. 2011. Accelerated profile HMM searches. PLoS Comput. Biol. 7.

Foote, A. D., Y. Liu, G. W. C. Thomas, T. Vinař, J. Alföldi, J. Deng, S. Dugan, C. E. Van Elk, M. E. Hunter, V. Joshi, Z. Khan, C. Kovar, S. L. Lee, K. Lindblad-Toh, A. Mancia, R. Nielsen, X. Qin, J. Qu, B. J. Raney, N. Vijay, J. B. W. Wolf, M. W. Hahn, D. M. Muzny, K. C. Worley, M. T. P. Gilbert, and R. A. Gibbs. 2015. Convergent evolution of the genomes of marine mammals. Nat. Genet. 47:272–275. Nature Publishing Group.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. Leduc, N. Friedman, and A. Regev. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8:1494–1512.

Huson, D. H., and C. Scornavacca. 2012. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. Syst. Biol. 61:1061–1067.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Kenkel, C. D., and L. K. Bay. 2017. Novel transcriptome resources for three scleractinian coral species from the Indo-Pacific. Gigascience 6:1–4.

Kenkel, C. D., E. Meyer, and M. V Matz. 2013. Gene expression under chronic heat stress in populations of the mustard hill coral (Porites astreoides) from different thermal environments. Mol. Ecol. 22:4322–34.

Kitchen, S. A., C. M. Crowder, A. Z. Poole, V. M. Weis, and E. Meyer. 2015. De novo assembly and characterization of four anthozoan (phylum Cnidaria) transcriptomes. G3 Genes Genomes Genet. 5:2441–2452.

Kocot, K. M., M. R. Citarella, L. L. Moroz, and K. M. Halanych. 2013. PhyloTreePruner: A phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. Evol. Bioinforma. 2013:429–435.

Lehnert, E. M., M. S. Burriesci, and J. R. Pringle. 2012. Developing the anemone Aiptasia as a tractable model for cnidarian-dinoflagellate symbiosis: the transcriptome of aposymbiotic A. pallida. BMC Genomics 13:271.

Li, L., C. J. J. Stoeckert, and D. S. Roos. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. Genome Res. 13:2178–2189.

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

Libro, S., S. T. Kaluziak, and S. V. Vollmer. 2013. RNA-seq profiles of immune related genes in the staghorn coral Acropora cervicornis Infected with white band disease. PLoS One 8:1–11.

Maor-Landaw, K., S. Karako-Lampert, H. W. Ben-Asher, S. Goffredo, G. Falini, Z. Dubinsky, and O. Levy. 2014. Gene expression profiles during short-term heat stress in the red sea coral Stylophora pistillata. Glob. Chang. Biol. 20:3026–3035.

Moya, A., L. Huisman, E. E. Ball, D. C. Hayward, L. C. Grasso, C. M. Chua, H. N. Woo, J.-P. Gattuso, S. Forêt, and D. J. Miller. 2012. Whole transcriptome analysis of the coral Acropora millepora reveals complex responses to CO₂-driven acidification during the initiation of calcification. Mol. Ecol. 21:2440–54.

Nordberg, H., M. Cantor, S. Dusheyko, S. Hua, A. Poliakov, I. Shabalov, T. Smirnova, I. V. Grigoriev, and I. Dubchak. 2014. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res. 42:26–31.

Polato, N. R., J. C. Vera, and I. B. Baums. 2011. Gene discovery in the threatened elkhorn coral: 454 sequencing of the acropora palmata transcriptome. PLoS One 6:e28634.

Price, M. N., P. S. Dehal, and A. P. Arkin. 2009. Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. 26:1641–1650.

Shinzato, C., M. Inoue, and M. Kusakabe. 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.

Shinzato, C., E. Shoguchi, T. Kawashima, M. Hamada, K. Hisata, M. Tanaka, M. Fujie, M. Fujiwara, R. Koyanagi, T. Ikuta, A. Fujiyama, D. J. Miller, and N. Satoh. 2011. Using the Acropora digitifera genome to understand coral responses to environmental change. Nature 476:320–3. Nature Publishing Group.

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Sun, J., Q. Chen, J. C. Y. Lun, J. Xu, and J. W. Qiu. 2013. PcarnBase: Development of a Transcriptomic Database for the Brain Coral Platygyra carnosus. Mar. Biotechnol. 15:244–251.

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

Talevich, E., B. M. Invergo, P. J. Cock, and B. a Chapman. 2012. Bio.Phylo: A unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. BMC Bioinformatics 13:209.

Traylor-Knowles, N., B. R. Granger, T. J. Lubinski, J. R. Parikh, S. Garamszegi, Y. Xia, J. A. Marto, L. Kaufman, and J. R. Finnerty. 2011. Production of a reference transcriptome and transcriptomic database (PocilloporaBase) for the cauliflower coral, Pocillopora damicornis. BMC Genomics 12:585.

Uniprot, C. T. 2016. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45.

Wattam, A. R., D. Abraham, O. Dalay, T. L. Disz, T. Driscoll, J. L. Gabbard, J. J. Gillespie, R. Gough, D. Hix, R. Kenyon, D. MacHi, C. Mao, E. K. Nordberg, R. Olson, R. Overbeek, G. D. Pusch, M. Shukla, J. Schulman, R. L. Stevens, D. E. Sullivan, V. Vonstein, A. Warren, R. Will, M. J. C. Wilson, H. S. Yoo, C. Zhang, Y. Zhang, and B. W. Sobral. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res. 42:581–591.

Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

Zou, Z., and J. Zhang. 2015. Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? Mol. Biol. Evol. 32:2085–2096.