PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Dixon GB, Kenkel CD. 2019 Molecular convergence and positive selection associated with the evolution of symbiont transmission mode in stony corals. *Proc. R. Soc. B* **286**: 20190111.
http://dx.doi.org/10.1098/rspb.2019.0111

Received: 17 January 2019
Accepted: 28 March 2019

Subject Category:
Genetics and genomics

Subject Areas:
evolution, genomics

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

Author for correspondence:
Carly D. Kenkel
e-mail: ckenkel@usc.edu

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4458539.

THE ROYAL SOCIETY PUBLISHING

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

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

[1]Department of Integrative Biology, The University of Texas at Austin, 1 University Station C0990, Austin, TX 78712, USA
[2]Department of Biological Sciences, University of Southern California, 3616 Trousdale Parkway, Los Angeles, CA 90089, USA

CDK, 0000-0003-1126-4311

Heritable symbioses have been critical for the evolution of life. The genetic consequences of evolving a heritable symbiosis from the perspective of the symbiont are well established, but concomitant changes in the host remain unresolved. In stony corals, heritable, vertical transmission has evolved repeatedly, 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 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 selection in driving the evolution of convergent transmission mode phenotypes. Of 10 774 orthologous genes, 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 symbiosis including endocytosis, immune response, cytoskeletal protein binding and cytoplasmic membrane-bounded vesicles. Finally, several novel candidates were identified among 100 genes showing evidence of positive selection at the particular convergence event, highlighting the value of our approach for generating new insight into host mechanisms associated with the evolution of heritable symbioses.

1. Introduction

For organisms that engage in symbiosis, the mode of symbiont transmission 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 [1–5]. 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 [6]); although it is important to note that the mode of transmission can change over evolutionary time [7] and mixed-mode transmission is also possible [6]. 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 [7–9]. Vertical transmission has been hypothesized to play an important role in the maintenance of mutually beneficial symbioses [10], and probably facilitated major transitions in individuality, such as the evolution of the eukaryotic cell

[11,12]. From the perspective of the symbiont, the genomic consequences of evolving a heritable symbiosis include a reduction in genome size and increased host dependence owing to the loss of functionally redundant genes [4,12]. 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 [13] and theory predicts a high rate of mutation in genes responsible for the host–symbiont fitness interaction [14]. Selection on mechanisms critical for the establishment and maintenance of a horizontally transmitted symbiosis, such as cell surface molecules mediating inter-partner recognition, is probably also relaxed [6]. Among metazoan hosts, diverse behavioural, developmental and physiological mechanisms are known to facilitate the vertical transmission of microbial endosymbionts [7,13], 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 [13,15]. Convergent evolution at the phenotypic level is often the result of similar changes at the genomic level [16,17] and comparative analyses have facilitated understanding of the genetic basis of convergently evolved phenotypes in diverse taxa [16,18,19]. 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 use such a comparative approach to identify candidate genes involved in the evolution of symbiont transmission mode. The majority of coral species acquire 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 [20,21]. However there is also significant morphological, physiological and ecological trait variation across the coral phylogeny [22], 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 symbionts vertically [21]. However, the association is not perfect; some *Porites* spp. and all known *Montipora* spp. have convergently evolved to broadcast spawn eggs which contain Symbiodiniaceae [23,24]. We therefore sequenced the transcriptome of the vertically transmitting broadcast spawner, *Montipora aequituberculata*, in addition to mining other publicly available coral sequence resources (electronic supplementary material, table S1), to compile a set of transcriptomic references in which vertical transmitters could be compared with their closest horizontally transmitting relatives, while also accounting for variation in other life-history traits (figure 1). From this dataset, we inferred orthologous groups and identified genes showing both signatures of positive selection and molecular convergence (independent substitutions for the same amino acid at the same position in two or more lineages). We found that the frequency of molecular convergence tended to be higher among vertically transmitting lineages and although top genes are enriched for biological processes previously implicated in the coral-algal symbiosis, we also identify several novel candidates, generating new insight into the mechanistic basis of this relationship.

2. Results and discussion

(a) Orthologue identification

To examine molecular convergence and positive selection, we compared homologous coding sequences from transcriptomic data of 25 coral species. First, protein coding sequences were predicted from the transcriptomic data based on open reading frames (ORFs) and sequence homology to known proteins [26] and protein domains [27], and FastOrtho [28] was used to assign sequences to preliminary orthologous groups ($n = 106\,300$ groups). A subset of 1196 single-copy orthologous groups with at least 20 of the 28 taxa represented was used to construct a species tree (figure 1), which recapitulates known relationships reported in earlier studies using single-gene [20,29] and multi-gene phylogenies [30]. Next, putative single-copy orthologues (groups with only a single representative sequence from each species) were identified from the initial set of 20 563 orthologous groups for which at least seven (25%) of the species were represented. Of these, 9794 were truly single copy, whereas 10 769 had multiple sequences for one or more species. Rather than eliminate all orthologous groups with multiple sequences, we applied a filtering approach similar to that described by [31] to retain an additional 3298 of the 10 769 multiple sequence orthologues. 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 (electronic supplementary material, figure S1; see Methods). We chose this approach because monophyletic multi-sequence orthologues can be biological in origin, resulting from gene duplication events subsequent to the relevant speciation event, or transcript isoforms of the same gene [31]. 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. In this way, we identified a total of 13 092 single copy orthologues. Orthologues were then aligned using MAFFT [32] and reverse-translated into codon sequences using Pal2Nal [33].

Orthologues were further quality filtered based on the monophyly of known clades (figure 1, 1–8 and Robusta/Complexa). Individual gene trees were constructed from nucleotide alignments of each single-copy orthologue. 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 orthologue was retained (27% of orthologous groups). If more than one sequence fell outside its expected clade, the orthologue was removed (15% of orthologous groups). In total, this left 119 049 sequences (mean species per orthologous group = 10.7) comprising 11 130 orthologous groups, hereafter referred to as genes. Genes with fewer than five representative sequences were also removed, resulting in a final total of 10 774 genes which were used for the ancestral reconstruction and branch-site tests.

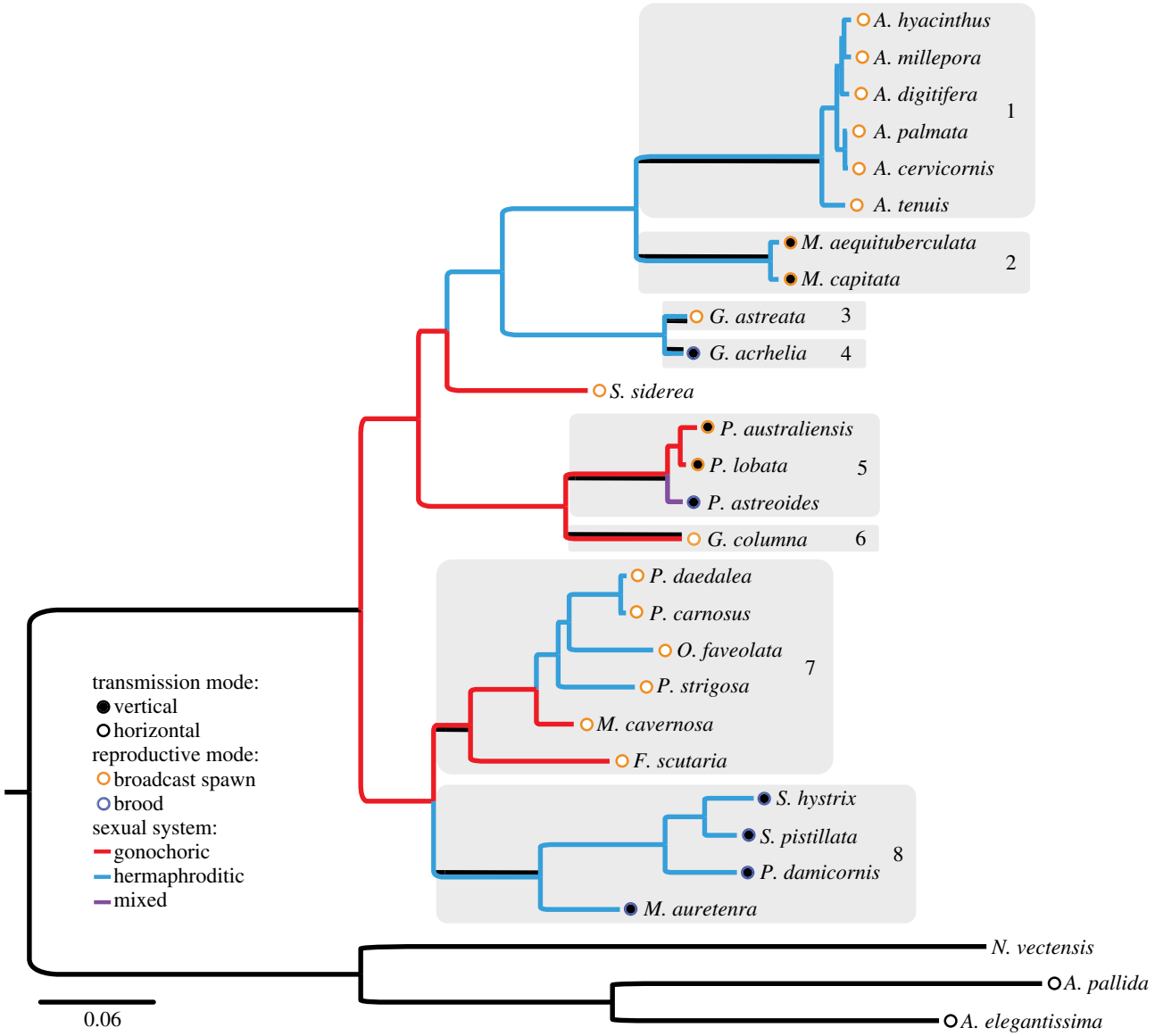

Figure 1. Species tree with phenotypic labels indicating transmission mode, reproductive mode and sexual system [21,24,25]. Vertically transmitting species are indicated by filled circles at their terminal nodes, horizontally transmitting species with open circles at their terminal nodes. For each clade (1–8), the particular branch examined for convergent substitutions and positive selection is emphasized in black. 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 and (8) Pocilloporid. (Online version in colour.)

(b) Evidence of positive selection and molecular convergence

For each orthologous nucleotide alignment, PAML [34] 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 between the branches leading directly to these clades' most recent common ancestors (figure 1). We classified these overlapping 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 the 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 [35], we consider both parallel and convergent substitutions to be indicative of molecular convergence.

Among the vertical transmitters, we identified 8952 amino acid positions exhibiting either parallel ($n = 8877$) or convergent ($n = 75$) substitutions in at least two lineages (ancestral reconstruction posterior estimate greater than 0.8, figure 2*a*; electronic supplementary material, figure S2). The convergence events occurred in 4117 out of 10 774 total genes in the dataset, with an average of 0.71 convergent sites identified per gene (median = 0; figure 2*b*). Of the four possible types of overlapping substitutions, convergent substitutions were by far the least frequent (figure 2*a*; electronic supplementary material, figure S2). The most common type was divergent substitutions. The two remaining types, parallel and 'all different' occurred with roughly similar frequency (figure 2*a*). Across the entire dataset, 11% of overlapping substitutions were classified as molecular convergence (convergent or parallel).

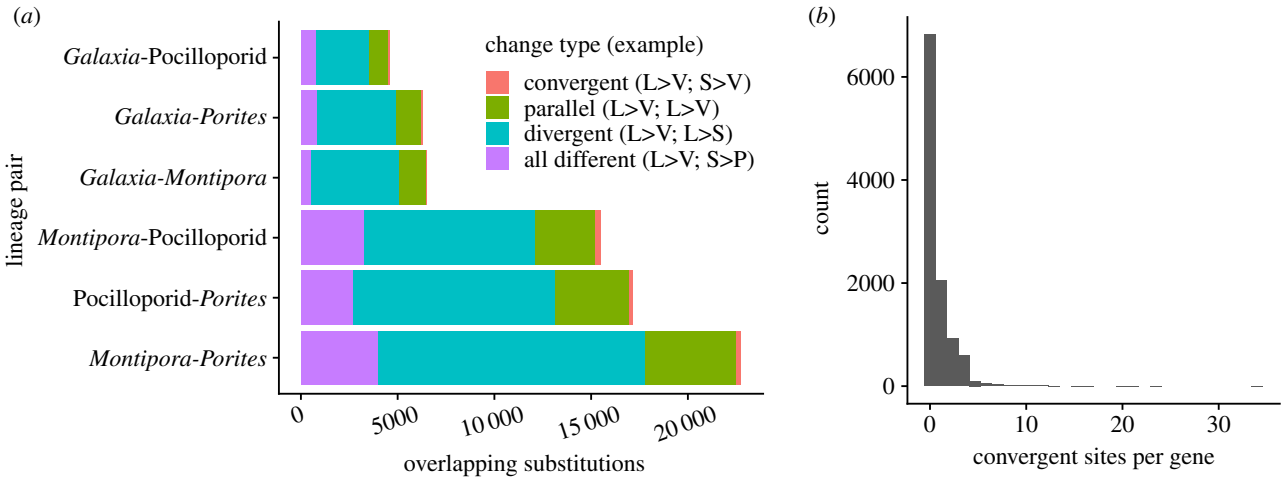

Figure 2. Frequency of convergence events. (*a*) An overlapping substitution is defined as an inferred amino acid change that occurred at the same position independently in the lineages leading to the common ancestor of the two indicated vertically transmitting clades. Each overlapping substitution was classified into one of four categories: convergent (least frequent; salmon; changes from different amino acids to the same amino acid), parallel (second most frequent; green; same amino acid to the same new amino acid), divergent (most common; teal; same amino acid to a different one), 'all different' (third most common; purple; different amino acids to different new amino acids). Examples of each type of overlapping substitution are shown in the legend using one letter amino acid codes. (*b*) Histogram of the number of sites showing molecular convergence (convergent or parallel substitutions) per tested gene (mean = 0.71; median = 0). (Online version in colour.)

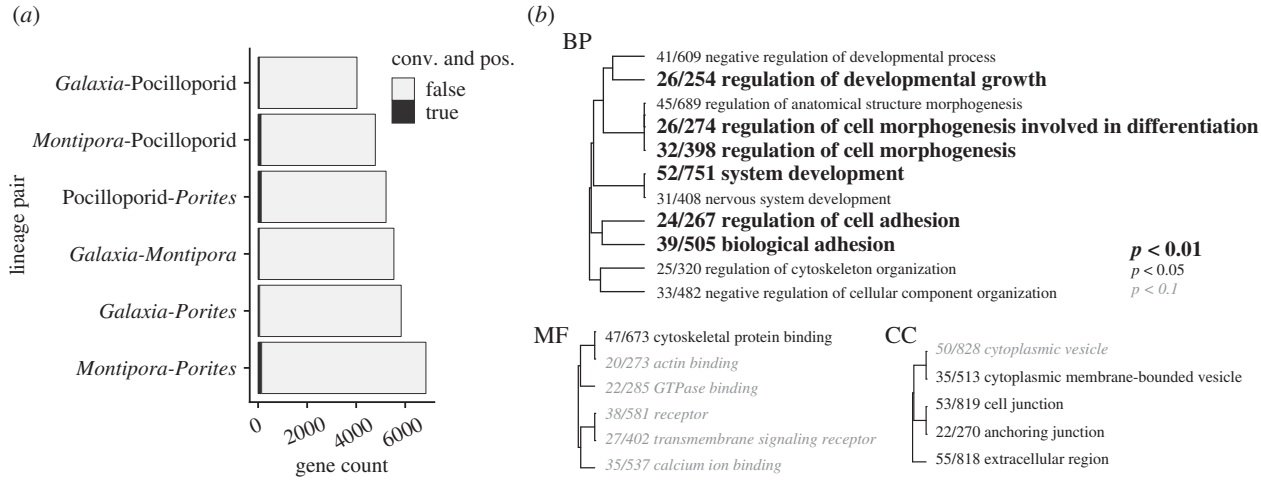

Figure 3. Frequency of genes exhibiting overlap in convergence and positive selection, and results of a categorical functional enrichment analysis of these candidates. (*a*) Frequency of genes exhibiting both signatures of convergence and positive selection per pair of vertically transmitting clades. 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 (true) relative to the total gene set (false) for each lineage pair comparison. (*b*) Gene ontology terms showing significant 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 bold text. BP, biological processes; CC, cellular component; MF, molecular function.

In addition to quantifying molecular convergence, we also tested for evidence of positive selection (amino acid sites with estimated nonsynonymous to synonymous substitution rate ratios greater than one) in each vertically transmitting lineage and for all vertically transmitting lineages at once using the branch-site models in PAML [34]. We found evidence of positive selection in 954 genes (likelihood ratio test (LRT), false discovery rate (FDR) < 0.1 in at least one branch-site test, electronic supplementary material, table S2) and many instances in which molecular convergence and positive selection were detected in the same gene (figure 3*a*; electronic supplementary material, figure 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 (electronic supplementary material, 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 [36] 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 Bayes empirical Bayes posterior probability > 0.8; electronic supplementary material, table S4). No ontology enrichments were detected for this reduced group, but annotations were recovered for 66 of the 100 genes.

(c) 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 [37]. Enforcement of vertical transmission in a

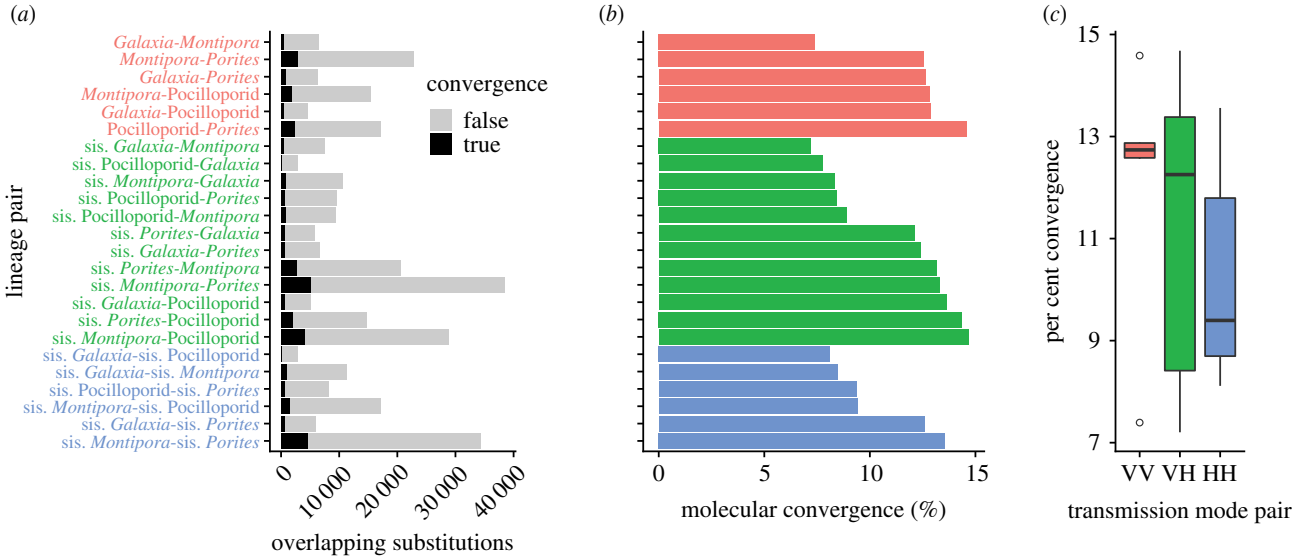

Figure 4. Comparison of the frequency of convergence events among overlapping substitutions. (*a*) Absolute counts of total overlapping substitutions and the proportion which were convergence events (true) for each species pair; see figure 1 for list of focal clade names (*b*) Percentage of overlapping 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. (Online version in colour.)

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 favoured by selection [38]. However, an earlier analysis of marine mammal lineages revealed that genomic convergence was actually highest in terrestrial sister taxa in which no phenotypic convergence was evident, suggesting that the options for adaptive evolution may be limited by pleiotropic constraints [19].

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) 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 [39]. Logistic regression indicated that the likelihood of molecular convergence was indeed higher among vertically transmitting lineage pairs ($\beta = 0.05$; $\chi^2_1 = 17.11$, $p = 0.0002$; figure 4), as was the co-occurrence of convergence and positive selection ($\beta = 0.06$; $\chi^2_1 = 12.62$, $p = 0.0018$; electronic supplementary material, figure S3). To assess the importance of other covariates, we compared mixed models including random effects for gene and lineage in addition to transmission-type pair. The effect of vertically transmitting lineage pairs on convergence remained significantly positive ($\beta = 0.04$; $\chi^2_1 = 9.29$, $p = 0.01$), but the effect for co-occurrence of convergence and positive selection did not ($\beta = -0.06$; $\chi^2_1 = 5.13$, $p = 0.08$). No trend was evident for the mean proportion of convergent sites also showing evidence of positive selection (electronic supplementary material, figure S4). Taken together, these findings appear consistent with the proposed role of natural selection in driving the evolution of convergent transmission mode phenotypes [40].

(d) 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 signalling, regulation of the host innate immune response and cell cycle, phagocytosis and cytoskeletal rearrangement [42]. 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 were obtained from comparisons against the UniProt Swiss-Prot database [26] and a categorical functional enrichment analysis (FDR < 0.1) was performed. Top enrichments (FDR < 0.01) among biological processes terms included regulation of developmental growth and cell morphogenesis, and biological adhesion (gene ontology (GO) : 0048638; GO : 0010769; GO007155; GO0022610). Endocytosis (GO : 0006897) and immune response (GO : 0006955) were also significant (FDR < 0.1). Among molecular functions, cytoskeletal protein binding (GO : 0008092) was the most significant enrichment (FDR = 0.016; figure 3*b*). Extracellular region (GO : 0005576) was the most significantly enriched term among cellular components; however, this term was also highlighted in a comparison of horizontally transmitting sister clades (electronic supplementary material, figure S5), suggesting that it may be under selection in all corals and not necessarily specific to the evolution of vertical transmission. Additional top cellular components 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 (biological processes, cellular components and molecular functions) and were also among the less than 1% of genes in which the particular convergence event also showed evidence of positive selection (electronic supplementary material, figure S6 and table S4). Importantly, none of these candidates have been previously implicated in the host-symbiont relationship in

analyses focusing on either coral bleaching, the breakdown of the symbiosis [43–47] or on the establishment of symbiosis in horizontally transmitting corals [48–50], highlighting the value of the present approach for identifying novel candidate genes.

ABL 1 is a ubiquitously expressed nonreceptor tyrosine kinase known to be involved in organismal responses to a multitude of signals, including cell adhesion, DNA damage, oxidative stress and cytokines [51]. This gene has probably evolved to serve a variety of context-dependent biological functions, but is known to regulate several immune response phenotypes in mammals including antigen receptor signalling in lymphocytes, and bacterial adhesion to host cells [52–54]. Through its role in regulating actin polymerization, ABL 1 is also involved in endocytosis [55], supporting the hypothesis that it may play a key role in mediating the heritable transmission of symbionts. Filamins are another family of actin-binding proteins which also exhibit great functional diversity in their interactions [56]. While Filamin C was not identified in earlier functional genomic studies, expression of Filamin A was recently reported to be modified by temperature over the course of a monthly reproductive cycle in *Pocillopora damicornis*, a vertically transmitting brooding coral [57]. Similarly, Filamin B was found to be differentially expressed between symbiotic and aposymbiotic *Aiptasia* anemones [58]. Combined, these results suggest an important role for this gene family in the maintenance and transmission of symbionts. Poly(C)-binding proteins also exhibit substantial functional diversity, but they are involved in transcriptional and translational regulation in addition to acting as structural components in DNA-protein complexes [59]. Interestingly, poly(rC) binding protein 2 is a negative regulator of mitochondrial antiviral signalling protein (MAVS), a critical component of innate antiviral immunity, where overexpression has been shown to reduce, and knockdown to increase cellular responses to viral infection [60]. MAVS interacts with RIG-I-like pattern recognition receptors, which are located in the cytoplasm, to identify foreign RNA [61]. However, they have also been shown to function in defence against some bacterial pathogens [61,62], suggesting that regulation of poly(rC) binding protein 2 could be involved in suppressing host innate immune responses against intracellular symbionts.

(e) Conclusions

Climate change and other anthropogenic processes threaten corals because of the sensitivity of the coral-dinoflagellate symbiosis to environmental stress [63,64]. Significant work has gone into investigating the breakdown of this relationship in the process known as 'coral bleaching' over the past three decades, yet fundamental questions remain unresolved, including a complete understanding of the genomic architecture underpinning the host–symbiont relationship [42,65]. Here, rather than asking about molecular mechanisms correlated with the breakdown of the coral symbiosis, we investigated a factor predicted to reinforce it: the evolution of vertical symbiont transmission. While the genes identified here represent promising candidates for further study, it is important to note that they probably represent only a fraction of the molecular changes involved in the evolution of symbiont transmission mode as there are alternate pathways to achieve the same phenotypic outcome that do not require changes at the level of the coding sequence [66]. Increasing genomic resources will facilitate a deeper understanding of such alternative mechanisms, and the concurrent development of more advanced genetic tools for manipulating the coral [67] and other cnidarian model symbioses [68,69] will facilitate quantification of the precise phenotypic effects of these novel genes, as well as of changes in their sequence, contributing to a greater understanding of the cellular and molecular mechanisms underpinning this specific relationship, and necessary for the evolution of a heritable symbiosis.

3. Methods

(a) Transcriptomic resources

A reference transcriptome was sequenced and assembled for *M. aequituberculata* and transcriptomic data from 25 species of Scleractinia (stony corals) and three species of Actiniaria (anemones) were downloaded from the web (electronic supplementary material, table S1).

(b) Protein sequence prediction

Sequence definition lines for each transcriptome were modified to include the species name and an arbitrary sequence number. To remove highly similar isoforms, we used cd-hit [70] 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, retaining only the longest sequence.

Protein coding sequences were then predicted based on ORFs and sequence homology to known proteins and protein domains as implemented in TransDecoder [71]. The longest ORFs were identified using a minimum amino acid length of 100 and protein sequences were predicted based on blastp alignments against the Swissprot database [26] and protein domains identified with scanHmm in HMMER (v.3.1b2 [27]).

(c) Orthologue assignment

Predicted coding sequences were assigned to orthologous groups using FastOrtho, an implementation of OrthoMCL [28] available through Pathosystems Resource Integration Center (PATRIC) web resources [72] (http://enews.patricbrc.org/fastortho/). We ran FastOrtho using reciprocal blastp results with an e-value cut-off of 1×10^{-10}, excluding hits with alignment lengths less than 75% of subject sequences.

(d) Construction of species tree

To construct a species tree, we used a subset of 1196 single-copy orthologous groups with at least 20 of the 28 taxa represented. Codon sequence alignments were concatenated in phylip format for input into RAxML [73]. 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 [74] and FigTree http://tree.bio.ed.ac.uk/software/figtree/.

(e) Paralogue pruning

When putative paralogues 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 seven (25%) of the species were represented. Protein sequences for these orthologues were aligned with MAFFT using local pair [32] and gene trees were constructed using FastTree [75]. 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 [76] 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 orthologues (9794 single-copy orthologues prior to pruning, 13 092 after pruning). After pruning, putative single-copy orthologues were reverse-translated into codon sequences using PAL2NAL [33].

(f) Phylogenetic orthologue filtering

Orthologous groups were further quality filtered by checking gene trees constructed from nucleotide alignments of each single-copy orthologue for monophyly of known clades (genus *Acropora*, genus *Montipora*, genus *Galaxia*, genus *Porites*, favid clade with *Fungia scutaria* as outgroup, pocilloporid clade with *Madracis auretenra* as outgroup, complex corals, robust corals), which were corroborated in our species tree (figure 1).

(g) Ancestral reconstruction and identification of convergent substitutions

For each orthologous nucleotide alignment, the ancestral amino acid was identified at each node in the species tree with PAML [34], as well as the amino acid changes that occurred along the branches of the tree using the species tree as a guide.

From the ancestral reconstruction results, we identified all substitutions that occurred at the same positions in two or more selected lineages (overlapping substitutions). The selected lineages included the branches leading to the common ancestor of four vertical transmitting clades, and their corresponding horizontally transmitting sister clade (eight clades total, figure 1). The horizontally transmitting sister clades were included to serve as negative controls, and for normalization of GO enrichment analyses (see below). In cases where a clade was represented by a single species, the terminal branch was used as the lineage for that clade (e.g. the two *Galaxia* species, figure 1).

(h) Testing for evidence of positive selection

Branch-site tests were performed on each orthologue in PAML [34] 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. 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 (electronic supplementary material, figure S7). In cases where a clade was represented by a single species, the terminal branch for that species was labelled as foreground. Branch-site tests were performed for each individual clade, and for all vertically transmitting clades at once. Significance was tested using LRTs, and *p*-values were adjusted to control for FDR using the Benjamini–Hochberg procedure [77]. 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 significant genes as 'positively selected' as in [19].

(i) Comparing frequency of molecular convergence

We compared the prevalence of molecular convergence between transmission modes by testing for differences in the frequency of convergence among all overlapping substitutions. We first applied simple logistic regression to test for significant effects of transmission mode pair on the log-odds of molecular convergence among overlapping substitutions. The binomial trial in this case is a single pair of independent substitutions at the same position. Those showing molecular convergence (parallel or convergent substitutions) were scored as success, and those that did not were scored as failure. Each overlapping substitution could occur between three different transmission mode pairs: vertical–vertical, vertical–horizontal or horizontal–horizontal. The same was done for the logistic models for the co-occurrence of convergence and positive selection, and the co-occurrence of convergence and identification of the particular site as positively selected. The logistic models including transmission mode pair were compared to the basic intercept models using χ^2 tests. To assess the effects of other covariates, we compared logistic mixed-effects models including random effects for gene and the lineage included [78]. Here, overlapping substitution was encoded as two binomial trials for which molecular convergence was counted as success. Based on model comparisons using ANOVA, a final model was chosen including a fixed effect for transmission mode pairing (three factors: vertical–vertical, vertical–horizontal or horizontal–horizontal), and random effects for orthologue (9933 factors) and individual lineage (eight factors; figure 1).

(j) 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 $< 1 \times 10^{-5}$, and default parameters for hmmscan). 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 [79] 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.

(k) Gene ontology 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 control analysis was performed for genes with the same signatures among the horizontally transmitting lineages (electronic supplementary material, figure S5). To perform fewer total tests, and reduce the effect of false discovery correction, only large GO terms, associated with at least 200 orthologues in our dataset, were tested for enrichment.

Data accessibility. 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/groves-dixon/convergent_evo_coral.

Authors' contributions. C.D.K. designed research, obtained funding and assembled new reference transcriptome; G.B.D. analysed convergence and selection; C.D.K. wrote the first draft of the manuscript and both authors contributed to revisions.

Competing interests. We declare we have no competing interests.

Funding. This work was supported, in part, by an NSF International Postdoctoral Research Fellowship, DBI-1401165 to C.D.K. Bioinformatic analyses were carried out using computational resources of the Texas Advanced Computer Center (TACC).

References

1. Herre EA, Knowlton N, Mueller UG, Rehner SA. 1999 The evolution of mutualisms: exploring the paths between conflict and cooperation. *Trends Ecol. Evol.* **14**, 49–53. (doi:10.1016/S0169-5347(98)01529-8)
2. Sauer C, Stackebrandt E, Gadau J, Hölldobler B, Gross R. 2000 Systematic relationships and cospeciation of bacterial endosymbionts and their carpenter ant host species: proposal of the new taxon *Candidatus Blochmannia* gen. nov. *Int. J. Syst. Evol. Microbiol.* **50**, 1877–1886. (doi:10.1099/00207713-50-5-1877)
3. Moran NA, Bennett GM. 2014 The tiniest tiny genomes. *Annu. Rev. Microbiol.* **68**, 195–215. (doi:10.1146/annurev-micro-091213-112901)
4. Bennett GM, Moran NA. 2015 Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. *Proc. Natl Acad. Sci. USA* **112**, 10 169–10 176. (doi:10.1073/pnas.1421388112)
5. Moran NA, McCutcheon JP, Nakabachi A. 2008 Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* **42**, 165–190. (doi:10.1146/annurev.genet.41.110306.130119)
6. Bright M, Bulgheresi S. 2010 A complex journey: transmission of microbial symbionts. *Nat. Rev. Microbiol.* **8**, 218–230. (doi:10.1038/nrmicro2262)
7. Sachs JL, Skophammer RG, Regus JU. 2011 Evolutionary transitions in bacterial symbiosis. *Proc. Natl Acad. Sci. USA* **108**, 10 800–10 807. (doi:10.1073/pnas.1100304108)
8. Sachs JL, Mueller UG, Wilcox TP, Bull JJ. 2004 The evolution of cooperation. *Q. Rev. Biol.* **79**, 135–160. (doi:10.1086/383541)
9. West SA, Fisher RM, Gardner A, Kiers ET. 2015 Major evolutionary transitions in individuality. *Proc. Natl Acad. Sci. USA* **112**, 10 112–10 119. (doi:10.1073/pnas.1421402112)
10. Frank SA. 1996 Host-symbiont conflict over the mixing of symbiotic lineages. *Proc. R. Soc. B* **263**, 339–344. (doi:10.1098/rspb.1996.0052)
11. Kiers ET, West SA. 2015 Evolving new organisms via symbiosis. *Science* **348**, 392–394. (doi:10.1126/science.aaa9605)
12. Fisher RM, Henry LM, Cornwallis CK, Kiers ET, West SA. 2017 The evolution of host-symbiont dependence. *Nat. Commun.* **8**, 15973. (doi:10.1038/ncomms15973)
13. Frank SA. 1996 Host control of symbiont transmission: the separation of symbionts into germ and soma. *Am. Nat.* **148**, 1113–1124. (doi:10.1086/285974)
14. Drown DM, Zee PC, Brandvain Y, Wade MJ. 2013 Evolution of transmission mode in obligate symbionts. *Evol. Ecol. Res.* **15**, 43–59.
15. Kikuchi Y, Hosokawa T, Nikoh N, Meng X-Y, Kamagata Y, Fukatsu T. 2009 Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. *BMC Biol.* **7**, 2. (doi:10.1186/1741-7007-7-2)
16. Stern DL. 2013 The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764. (doi:10.1038/nrg3483)
17. Conte GL, Arnegard ME, Peichel CL, Schluter D. 2012 The probability of genetic parallelism and convergence in natural populations. *Proc. R. Soc. B* **279**, 5039–5047. (doi:10.1098/rspb.2012.2146)
18. Jones FC *et al.* 2012 The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61. (doi:10.1038/nature10944)
19. Foote AD *et al.* 2015 Convergent evolution of the genomes of marine mammals. *Nat. Genet.* **47**, 272–275. (doi:10.1038/ng.3198)
20. Kerr AM. 2005 Molecular and morphological supertree of stony corals (Anthozoa: Scleractinia) using matrix representation parsimony. *Biol. Rev. Camb. Philos. Soc.* **80**, 543–558. (doi:10.1017/S1464793105006780)
21. Hartmann AC, Baird AH, Knowlton N, Huang D. 2017 The paradox of environmental symbiont acquisition in obligate mutualisms. *Curr. Biol.* **27**, P3711–P3716. (doi:10.1016/j.cub.2017.10.036)
22. Madin JS *et al.* 2016 The coral trait database, a curated database of trait information for coral species from the global oceans. *Sci. Data* **3**, 160017. (doi:10.1038/sdata.2016.17)
23. Fadlallah YH. 1983 Sexual reproduction, development and larval biology in scleractinian corals. *Coral Reefs* **2**, 129–150. (doi:10.1007/BF00336720)
24. Richmond RH, Hunter CL. 1990 Reproduction and recruitment of corals: comparisons among the Caribbean, the Tropical Pacific, and the Red Sea. *Marine Ecol. Prog. Ser.* **60**, 185–203. (doi:10.3354/meps060185)
25. Kerr AM, Baird AH, Hughes TP. 2011 Correlated evolution of sex and reproductive mode in corals (Anthozoa: Scleractinia). *Proc. R. Soc. B* **278**, 75–81. (doi:10.1098/rspb.2010.1196)
26. Uniprot CT. 2016 UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, 115. (doi:10.1093/nar/gkw1152)
27. Eddy SR. 2011 Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195. (doi:10.1371/journal.pcbi.1002195)
28. Li L, Stoeckert CJJ, Roos DS. 2003 OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189. (doi:10.1101/gr.1224503.candidates)
29. Kitahara MV, 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**, e11490. (doi:10.1371/journal.pone.0011490)
30. Bhattacharya D *et al.* 2016 Comparative genomics explains the evolutionary success of reef-forming corals. *eLife Sci.* **5**, e13288. (doi:10.7554/eLife.13288)
31. Kocot KM, Citarella MR, Moroz LL, Halanych KM. 2013 PhyloTreePruner: a phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. *Evol. Bioinform.* **2013**, 429–435. (doi:10.4137/EBO.S12813)
32. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)
33. Suyama M, Torrents D, Bork P. 2006 PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, 609–612. (doi:10.1093/nar/gkl315)
34. Yang Z. 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)
35. Zou Z, Zhang J. 2015 Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? *Mol. Biol. Evol.* **32**, 2085–2096. (doi:10.1093/molbev/msv091)
36. Yang Z, Wong WSW, Nielsen R. 2005 Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107–1118. (doi:10.1093/molbev/msi097)
37. Orr HA. 2005 The probability of parallel evolution. *Evolution* **59**, 216–220. (doi:10.1111/j.0014-3820.2005.tb00907.x)
38. Sachs JL, Wilcox TP. 2006 A shift to parasitism in the jellyfish symbiont *Symbiodinium microadriaticum*. *Proc. R. Soc. B* **273**, 425–429. (doi:10.1098/rspb.2005.3346)
39. 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**, 1232–1236. (doi:10.1093/molbev/msv013)
40. Kenkel CD, Bay LK. 2018 Exploring mechanisms that affect coral cooperation: symbiont transmission mode, cell density and community composition. *PeerJ* **6**, e6047. (doi:10.7717/peerj.6047)
41. Wakefield T, Farmer M, Kempf S. 2000 Revised description of the fine structure of *in situ* 'zooxanthellae' genus *Symbiodinium*. *Biol. Bull.* **199**, 76–84. (doi:10.2307/1542709)
42. Davy SK, Allemand D, Weis VM. 2012 Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiol. Mol. Biol. Rev.* **76**, 229–261. (doi:10.1128/MMBR.05014-11)
43. Barshis DJ, Ladner JT, Oliver TA, Seneca FO, Traylor-Knowles N, Palumbi SR. 2013 Genomic basis for coral resilience to climate change. *Proc. Natl Acad. Sci. USA* **110**, 1387–1392. (doi:10.1073/pnas.1210224110)
44. Desalvo MK, Voolstra CR, Sunagawa S, Schwarz JA, Stillman JH, Coffroth MA, Szmant AM, Medina M. 2008 Differential gene expression during thermal stress and bleaching in the Caribbean coral *Montastraea faveolata*. *Mol. Ecol.* **17**, 3952–3971. (doi:10.1111/j.1365-294X.2008.03879.x)
45. DeSalvo MK, Sunagawa S, Voolstra CR, Medina M. 2010 Transcriptomic responses to heat stress and bleaching in the elkhorn coral *Acropora palmata*. *Marine Ecol. Prog. Ser.* **402**, 97–113. (doi:10.3354/meps08372)
46. Maor-Landaw K, Karako-Lampert S, Ben-Asher HW, Goffredo S, Falini G, Dubinsky Z, Levy O. 2014 Gene

expression profiles during short-term heat stress in the red sea coral *Stylophora pistillata*. *Glob. Change Biol.* **20**, 3026–3035. (doi:10.1111/gcb.12592)

47. Rodriguez-Lanetty M, Harii S, Hoegh-Guldberg O. 2009 Early molecular responses of coral larvae to hyperthermal stress. *Mol. Ecol.* **18**, 5101–5114. (doi:10.1111/j.1365-294X.2009.04419.x)
48. Voolstra CR, Schwarz JA, Schnetzer J, Sunagawa S, Desalvo MK, Szmant AM, Coffroth MA, Medina M. 2009 The host transcriptome remains unaltered during the establishment of coral–algal symbioses. *Mol. Ecol.* **18**, 1823–1833. (doi:10.1111/j.1365-294X.2009.04167.x)
49. Schnitzler CE, Weis VM. 2010 Coral larvae exhibit few measurable transcriptional changes during the onset of coral-dinoflagellate endosymbiosis. *Mar. Genomics* **3**, 107–116. (doi:10.1016/j.margen.2010.08.002)
50. 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**, 3127–3141. (doi:10.1111/mec.13659)
51. Wang JYJ. 2014 The capable ABL: what is its biological function? *Mol. Cell. Biol.* **34**, 1188–1197. (doi:10.1128/MCB.01454-13)
52. Arce KP, Varela-Nallar L, Farias O, Cifuentes A, Bull P, Couch BA, Koleske AJ, Inestrosa NC, Alvarez AR. 2010 Synaptic clustering of PSD-95 is regulated by c-Abl through tyrosine phosphorylation. *J. Neurosci.* **30**, 3728–3738. (doi:10.1523/JNEUROSCI.2024-09.2010)
53. Swimm A, Bommarius B, Li Y, Cheng D, Reeves P, Sherman M, Veach D, Bornmann W, Kalman D. 2004 Enteropathogenic *Escherichia coli* use redundant tyrosine kinases to form actin pedestals. *MBoC* **15**, 3520–3529. (doi:10.1091/mbc.e04-02-0093)
54. Huang Y, Comiskey EO, Dupree RS, Li S, Koleske AJ, Burkhardt JK. 2008 The c-Abl tyrosine kinase regulates actin remodeling at the immune synapse. *Blood* **112**, 111–119. (doi:10.1182/blood-2007-10-118232)
55. Tanos B, Pendergast AM. 2006 Abl tyrosine kinase regulates endocytosis of the epidermal growth factor receptor. *J. Biol. Chem.* **281**, 32 714–32 723. (doi:10.1074/jbc.M603126200)
56. Feng Y, Walsh CA. 2004 The many faces of filamin: a versatile molecular scaffold for cell motility and signalling. *Nat. Cell Biol.* **6**, 1034–1038. (doi:10.1038/ncb1104-1034)
57. 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*. *Mol. Ecol.* **26**, 3913–3925. (doi:10.1111/mec.14162)
58. Lehnert EM, Mouchka ME, Burriesci MS, Gallo ND, Schwarz JA, Pringle JR. 2014 Extensive differences in gene expression between symbiotic and aposymbiotic cnidarians. *G3: Genes, Genomes, Genetics* **4**, 277–295. (doi:10.1534/g3.113.009084)
59. Makeyev AV, Liebhaber SA. 2002 The poly(C)-binding proteins: a multiplicity of functions and a search for mechanisms. *RNA* **8**, 265–278. (doi:10.1017/S1355838202024627)
60. You F, Sun H, Zhou X, Sun W, Liang S, Zhai Z, Jiang Z. 2009 PCBP2 mediates degradation of the adaptor MAVS via the HECT ubiquitin ligase AIP4. *Nat. Immunol.* **10**, 1300–1308. (doi:10.1038/ni.1815)
61. Brubaker SW, Bonham KS, Zanoni I, Kagan JC. 2015 Innate immune pattern recognition: a cell biological perspective. *Annu. Rev. Immunol.* **33**, 257–290. (doi:10.1146/annurev-immunol-032414-112240)
62. 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**, 576–591. (doi:10.1016/j.cell.2009.06.015)
63. Hughes TP *et al.* 2003 Climate change, human impacts, and the resilience of coral reefs. *Science* **301**, 929–933. (doi:10.1126/science.1085046)
64. Hoegh-Guldberg O *et al.* 2007 Coral reefs under rapid climate change and ocean acidification. *Science* **318**, 1737–1742. (doi:10.1126/science.1152509)
65. Edmunds PJ, Gates RD. 2003 Has coral bleaching delayed our understanding of fundamental aspects of coral–dinoflagellate symbioses? *BioScience* **53**, 976–980. (doi:10.1641/0006-3568(2003)053[0976:HCBDOU]2.0.CO;2)
66. Prud'homme B, Gompel N, Carroll SB. 2007 Emerging principles of regulatory evolution. *Proc. Natl Acad. Sci. USA* **104**, 8605–8612. (doi:10.1073/pnas.0700488104)
67. Cleves PA, Strader ME, Bay LK, Pringle JR, Matz MV. 2018 CRISPR/Cas9-mediated genome editing in a reef-building coral. *Proc. Natl Acad. Sci. USA* **115**, 5235–5240. (doi:10.1073/pnas.1722151115)
68. 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. *Sci. Rep.* **8**, 16437. (doi:10.1038/s41598-018-34773-1)
69. 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**, 35. (doi:10.3389/fevo.2018.00035)
70. Li W, Godzik A. 2006 Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659. (doi:10.1093/bioinformatics/btl158)
71. Haas BJ *et al.* 2013 De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512. (doi:10.1038/nprot.2013.084)
72. Wattam AR *et al.* 2014 PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* **42**, 581–591. (doi:10.1093/nar/gkt1099)
73. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. (doi:10.1093/bioinformatics/btu033)
74. Huson DH, Scornavacca C. 2012 Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* **61**, 1061–1067. (doi:10.1093/sysbio/sys062)
75. Price MN, Dehal PS, Arkin AP. 2009 Fasttree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650. (doi:10.1093/molbev/msp077)
76. Talevich E, Invergo BM, Cock PJ, Chapman BA. 2012 Bio.Phylo: a unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. *BMC Bioinf.* **13**, 209. (doi:10.1186/1471-2105-13-209)
77. Benjamini Y, Hochberg Y. 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300.
78. Bates D, Mächler M, Bolker B, Walker S. 2015 Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48. (doi:10.18637/jss.v067.i01)
79. Huntley RP, Sawford T, Mutowo-Meullenet P, Shypitsyna A, Bonilla C, Martin MJ, O'Donovan C. 2015 The GOA database: gene ontology annotation updates for 2015. *Nucleic Acids Res.* **43**, D1057–D1063. (doi:10.1093/nar/gku1113)