# Analyses of Corallimorpharian Transcriptomes Provide New Perspectives on the Evolution of Calcification in the Scleractinia (Corals)

Mei-Fang Lin[1,2], Aurelie Moya[2], Hua Ying[3], Chaolun Allen Chen[4,5], Ira Cooke[1], Eldon E. Ball[2,3], Sylvain Forêt[2,3], and David J. Miller[1,2,*]

[1]Comparative Genomics Centre, Department of Molecular and Cell Biology, James Cook University, Townsville, Queensland, Australia

[2]ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland, Australia

[3]Research School of Biology, Australian National University, Canberra, Australian Capital Territory, Australia

[4]Biodiversity Research Centre, Academia Sinica, Nangang, Taipei, Taiwan

[5]Taiwan International Graduate Program (TIGP)-Biodiversity, Academia Sinica, Nangang, Taipei, Taiwan

*Corresponding author: E-mail: david.miller@jcu.edu.au.

## Abstract

Corallimorpharians (coral-like anemones) have a close phylogenetic relationship with scleractinians (hard corals) and can potentially provide novel perspectives on the evolution of biomineralization within the anthozoan subclass Hexacorallia. A survey of the transcriptomes of three representative corallimorpharians led to the identification of homologs of some skeletal organic matrix proteins (SOMPs) previously considered to be restricted to corals.

Carbonic anhydrases (CAs), which are ubiquitous proteins involved in $CO_2$ trafficking, are involved in both coral calcification and photosynthesis by endosymbiotic *Symbiodinium* (zooxanthellae). These multiple roles are assumed to place increased demands on the CA repertoire and have presumably driven the elaboration of the complex CA repertoires typical of corals (note that "corals" are defined here as reef-building Scleractinia). Comparison of the CA inventories of corallimorpharians with those of corals reveals that corals have specifically expanded the secreted and membrane-associated type CAs, whereas similar complexity is observed in the two groups with respect to other CA types.

Comparison of the CA complement of the nonsymbiotic corallimorph *Corynactis australis* with that of *Ricordea yuma*, a corallimorph which normally hosts *Symbiodinium*, reveals similar numbers and distribution of CA types and suggests that an expansion of the CA repertoire has been necessary to enable calcification but may not be a requirement to enable symbiosis. Consistent with this idea, preliminary analysis suggests that the CA complexity of zooxanthellate and nonzooxanthellate sea anemones is similar.

The comparisons above suggest that although there are relatively few new genes in the skeletal organic matrix of corals (which controls the skeleton deposition process), the evolution of calcification required an expanded repertoire of secreted and membrane-associated CAs.

**Key words:** corallimorpharian, coral calcification, carbonic anhydrase, molecular evolution, skeletal organic matrix proteins.

## Introduction

Corallimorpharia is a small and enigmatic anthozoan order closely related to the hard corals (order Scleractinia) but differing from them in that its representatives lack a skeleton. The relationship between corals and corallimorpharians has been equivocal, one factor in this being that—skeletons aside—they are essentially indistinguishable on morphological grounds (den Hartog 1980; Medina et al. 2006; Daly et al. 2007; Kitahara et al. 2014; Lin et al. 2014). Although it has been argued that the corallimorpharian ancestor was a coral that underwent skeleton loss (Medina et al. 2006), this idea has not been generally accepted (see, e.g., Budd et al. 2010; Barbeitos et al. 2010). Whole transcriptome scale phylogenomics implies that the Scleractinia and Corallimorpharia are distinct monophyletic groups (Lin et al. 2016), thus the ability

to deposit a massive aragonite skeleton evolved after the two orders diverged. However, the close relationship between these orders implies that corallimorpharians could be uniquely informative with respect to the evolution of the biomineralization process within the Hexacorallia.

One approach to understanding the evolution of taxon-specific traits is provided by comparative genomics, and this has been employed to investigate some aspects of coral biology. For example, comparisons between the coral *Acropora* and the sea anemone *Nematostella* imply that a more complex immune repertoire is mandatory for the establishment and maintenance of symbionts by the former (Shinzato et al. 2011; Hamada et al. 2013). Similar approaches indicate that the (noncalcifying) sea anemone has homologs of a number of the genes involved in skeleton deposition in corals (Ramos-Silva et al. 2013), suggesting that relatively few new genes may have been required to enable calcification on scales characteristic of reef-building Scleractinia. Although the sea anemone genome has provided some important insights into coral biology, the depth of the coral/sea anemone divergence (around 500 Myr; Shinzato et al. 2011) limits the usefulness of such comparisons. The closer relationship between corals and corallimorpharians suggests that the latter may be more informative comparators, but until recently corallimorphs have been poorly represented in terms of available molecular data, whereas whole genome sequences (Shinzato et al. 2011) and large transcriptome data sets (e.g. Moya et al. 2012) have been available for some time for corals, with at least twenty of varying quality and completeness now available (Bhattacharya et al. 2016).

Calcification has arisen independently many times during animal evolution. Within the Cnidaria, many octocorals deposit spicules composed of calcium carbonate in the form of calcite, whereas the skeletons of Scleractinia are composed exclusively of aragonite. Because calcification has arisen independently on multiple occasions, some of the components involved are unique to each lineage, but the chemistry of the process dictates that there is also a conserved component (Moya et al. 2012). The latter category of genes includes those involved in ion transport and in controlling carbonate chemistry, for example, bicarbonate transporters (Zoccola et al. 2015) and carbonic anhydrases (CAs) (Jackson et al. 2007; Grasso et al. 2008; Bertucci et al. 2013). The nonconserved category of the calcification repertoire typically includes many of the genes whose products control the deposition of calcium carbonate to form the skeleton—for example, the heterogeneous skeletal organic matrix proteins (SOMPs) involved in mollusk calcification lack orthologs in other phyla.

Considerable progress has recently been made in characterizing the calcification repertoire of corals. The first SOM protein to be identified in a coral is known as galaxin, and was originally identified by proteomic analyses of the skeleton of *Galaxea fascicularis* (Fukuda et al. 2003). To date, four distinct galaxin-related sequences have been identified in *Acropora*; two "adult-type" galaxins (Reyes-Bermudez et al. 2009; Ramos-Silva et al. 2013) and two divergent but related "galaxin-like" sequences (Reyes-Bermudez et al. 2009), which are not in the skeleton (Ramos-Silva et al. 2013) but which, from their spatial expression, may be involved in laying it down. Galaxins are cysteine-rich repetitive proteins, each repeat containing one or more di-cysteine motifs. Each of the *Acropora* galaxins possesses an N-terminal signal peptide. After signal peptide cleavage, however, the adult-type galaxin proteins consist entirely of di-cysteine-rich repeat units, whereas acidic domains precede the repetitive regions in the mature forms of both of the *Acropora millepora* "galaxin-like" proteins (Ramos-Silva et al. 2013). Galaxin-related sequences have been reported from a range of other animals, but these typically have low sequence similarity (e.g., Esgal1 from the squid *Euprymna scolopes* is involved in the establishment and maintenance of its bacterial symbiont *Vibrio fischeri*; Heath-Heckman et al. 2014) and resemble each other only in containing di-cysteine repeat motifs. So it has been suggested that galaxins *sensu stricto* may be restricted to corals (Reyes-Bermudez et al. 2009; Ramos-Silva et al. 2013). By applying proteomic approaches, Ramos-Silva et al. (2013) identified 36 SOMPs in the coral *A. millepora*. A similar study on *Stylophora pistillata*, another coral (Drake et al. 2013), implicated some of the same components, but also regarded some nonskeletal proteins as SOMPs (Ramos-Silva, Marin, et al. 2013b). Two galaxins were amongst the SOMPs identified by Ramos-Silva et al. (2013), but the most surprising aspect of this analysis was that most (28) of the 36 SOMPs identified in *Acropora* have homologs that are either widespread or are present in *Nematostella vectensis* or *Hydra* spp, noncalcifying cnidarians for which whole genome data are available. Thus only 8 (SAP1, SAP2, SOMP1, SOMP2, SOMP3, SOMP4, SOMP6, and cephalotoxin-like) of the 36 SOMPs identified in *Acropora* were coral-specific (not found in anemones or other organisms), although note that the last of these had a surprising level of similarity to a mollusk protein (Ramos-Silva et al. 2013).

Based on immunolocalization to the calicoblastic ectoderm of *S. pistillata*, it has recently been suggested that a specific solute carrier—the bicarbonate active transporter (BAT) SLC4γ—plays a key role in the deposition of the coral skeleton by facilitating movement of inorganic carbonate to the site of calcification (Zoccola et al. 2015). The presence of SLC4γ orthologs in a range of corals, but not in sea anemones, was taken as evidence that this gene played a key role in the evolution of biomineralization in the Scleractinia (Zoccola et al. 2015). For this reason, the presence or absence of SLC4γ orthologs in corallimorpharians is important in terms of understanding the origins of calcification in corals.

CAs are ubiquitous enzymes that catalyze the interconversion of $HCO_3^-$ and $CO_2$ and are involved in a wide range of functions that includes pH buffering. In calcifying organisms, CAs have important additional roles in transporting carbonate

to the site of calcification, hence these enzymes are a conserved component of the calcification repertoire (Weis and Reynolds, 1999; Jackson et al. 2007; Moya et al. 2012). In symbiotic animals such as corals, CAs also function in ensuring the supply of $CO_2$ to the photosynthetic symbionts; note that a large proportion of $CO_2$ fixed by *Symbiodinium* in corals is derived from (coral) respiration (Furla et al. 2000), and a large part of the fixed carbon may be exported back to the host (reviewed in Davy et al. 2012). These various demands have presumably driven the elaboration of complex CA repertoires that are typical of corals (see for review Bertucci et al. 2013).

We recently reported (Lin et al. 2016) the assembly of large transcriptome data sets for three corallimorpharians; *Rhodactis indosinensis*, *Ricordea yuma*, and *Corynactis australis*. To better understand the origins of the coral calcification repertoire, the transcriptomes of the corallimorpharians and those of representatives of other cnidarian groups were surveyed, focusing specifically on known components of the skeletal organic matrix, proteins associated with supplying carbonate to the site of calcification, or implicated in calcification on the basis of expression patterns in coral development. The results are consistent with the evolution of calcification requiring relatively few genomic changes in corals.

## Materials and Methods

### Corallimorpharian Transcriptomes

Full details of the methods used to generate sequence data and assemble the transcriptomes of three corallimorpharians are provided in a sister manuscript to this (Lin et al. 2016). The completeness of the corallimorpharian transcriptome assemblies was assessed using BUSCO (v1.22; Simão et al. 2015); percentages of the BUSCO metazoan transcriptome gene set that were recovered or missing are summarized in supplementary table S1, Supplementary Material online. An additional criterion applied to assess the completeness of transcriptome assemblies was to search for close matches to the 1808 core cnidarian transcripts identified by comparing the gene predictions from *Acropora digitifera*, *Nematostella vectensis* and *Hydra magnipapillata* (Lin et al. 2016). Of these 1808 core transcripts, 1609 were detected in *A. millepora*, 1481 in *C. australis*, 1401 in *R. yuma* and 1261 in *R. indosinensis*. Therefore, while some genes may be missing from individual corallimorpharian transcriptomes, the expectation is that the combination of these three species provides a representative data set for comparative analyses against Scleractinia.

### Searching for Calcification-Related Genes

Calcification-related genes, such as small cysteine-rich proteins (SCRiPs) (Sunagawa et al. 2009), galaxins (Fukuda et al. 2003; Reyes-Bermudez et al. 2009), additional SOMPs (Ramos-Silva et al. 2013), CAs (Moya et al. 2012), and three

taxonomically restricted genes (Moya et al. 2012) were searched against the corallimorpharian transcriptomes with an $E$-value cut-off of $e \leq 10^{-5}$ (Lin et al. 2016). Note, however, that more stringent cut-offs were applied when manual inspection of the results indicated that the default setting ($e \leq 10^{-5}$) was inadequate. Database accession details and sources of the reference sequences used are summarized in supplementary table S2, Supplementary Material online. To extend the knowledge of the distribution of calcification-related genes that were thought to be present only in the coral *Acropora*, the search was also applied to ten anthozoan transcriptome data sets that are publically available, including six robust corals, the complex coral *Porites australiensis*, the sea anemones *Nematostella vectensis* and *Anthopleura elegantissima* and the octocoral *Gorgonia ventalina* (Lin et al. 2016) as well as a recently released genome of the symbiotic anemone *Exaiptasia* (Baumgarten et al. 2015) with the same cut-off threshold used above. An additional BLAST search against the NCBI nonredundant (nr) database (accessed on October 15, 2014) was carried out with $E$-value cut-off of $e \leq 10^{-6}$ to evaluate the broader distribution of homologs of specific coral genes.

The presence and location of signal peptide cleavage sites in candidate amino acid sequences was predicted using SignalP v.4.1 (Petersen et al. 2011) (accessed on December 12, 2014). An additional tool, TargetP v.1.1 (Emanuelsson et al. 2000) (accessed on March 26, 2015), was used to predict the subcellular localization of carbonic anhydrase proteins. The InterProScan 5 platform (www.ebi.ac.uk/Tools/pfa/iprscan5; accessed on January 16, 2015) was used for functional classification of proteins and the presence of possible transmembrane domains investigated using TMHMM v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/; accessed on January 16, 2015). The Compute pI/Mw tool from the ExPASy bioinformatics portal (accessed on January 16, 2015) was used to estimate the theoretical isoelectric point (pI) value for each galaxin protein.

### Phylogenetic Methods

Similarity between corallimorpharian and *A. millepora* galaxins was evaluated using BioEdit v7.0.5.3 (Hall 1999). The galaxin sequences recovered were diverse, as indicated in Reyes-Bermudez et al. (2009) and Moya et al. (2012); regularity of di-cysteine repeat motifs permitted alignment of the protein sequences, but the level of variation precluded meaningful phylogenetic analyses. Sequence saturation at the nucleotide level was evaluated based on the transition and transversion substitutions versus the Tamura-Nei (TN93) distance of three codon positions using DAMBE 5 (Xia, 2013). Results indicated that galaxin sequences were saturated at all three codon positions (data not shown), thus phylogenetic analysis was not pursued.

Based on high levels of amino acid similarity (supplementary fig. S2, Supplementary Material online), the coral and corallimorpharian SOMP1 sequences were subjected to phylogenetic analyses, as follows. ProtTest 3.4 (Darriba et al. 2011) selected JTT + G as the best-fitting model of protein sequence evolution, under which maximum-likelihood (ML) analysis was inferred in MEGA5 (Tamura et al. 2011) with 1000 bootstrap replicates.

In the case of the CAs, sequences were trimmed to the conserved regions (pfam domains) based on the conserved domains search in Web CD-Search Tool (http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi, last accessed on 15 March 2015) and aligned using ClustalW. Both ML and Bayesian inference (BI) analyses were conducted. For the ML method, ProtTest 3.4 selected WAG + G as best-fit model and analyses were conducted using PhyML 3.0 (Guindon et al. 2010) with aLRT (approximate likelihood-ratio test) branch support search based on a Shimodaira–Hasegawa-like procedure. BI was analyzed with MrBayes v3.2.2 (Ronquist et al. 2012) for four chains, 2 million generations, checking for convergence as the average split frequency and discarding the first 25% of trees as burn-in. For the carbonic anhydrase tree, the split frequency after 2 million generations was 0.027671 (0.01–0.05 is acceptable). Trees were sampled every 1000 generations for 2 runs, hence consensus trees were based on 3000 trees.

Phylogenetic analyses of members of the solute carrier family (SLC) 4 were carried out as follows. As some of the matches recovered were incomplete, only those sites with coverage of greater than 70% were included in the analyses. The alignment on which analyses were conducted consisted of 598 amino acid positions, and the best-fit model applied was JTT + G estimated by MEGA 5 (Tamura et al. 2011). ML phylogenetic analyses were conducted using PhyML 3.0 (Guindon et al. 2010) for 1000 bootstraps and BI was carried out in MrBayes v3.2.2 (Ronquist et al. 2012) as described above; in this case, the split frequency value after 2 million generations was 0.000373 (<0.001 is confident).

Sequence alignments used in phylogenetic analyses are provided as supplementary materials S3—S8, Supplementary Material online.

## Results

### Corallimorpharian Counterparts of Known SOMPs

Using a cut-off of $e \leq 4 \times 10^{-7}$, 21 of the 28 coral SOMPs shared with other noncalcifying cnidarians have corallimorpharian matches, 10 of these having e-value = 0 in BLASTP analyses (table 1 and supplementary table S3, Supplementary Material online). Levels of identity between corallimorpharian and coral homologs were consistently higher than for coral-*Nematostella* comparisons (Ramos-Silva et al. 2013).

Of the 8 SOM proteins identified by Ramos-Silva et al. (2013) as coral-specific, SOMP1 and SOMP2 had apparent matches in the corallimorpharians ($e \leq 10^{-7}$, table 1), but not in sea anemones; possible homologs of SOMP3 were identified in two sea anemones, but not in *Nematostella* or corallimorpharians (supplementary table S3, Supplementary Material online). The *A. millepora* SOMP1 sequence identified by Ramos-Silva et al. (2013) (accession B3EX00.1) is incomplete; that sequence matches well the C-terminal region of *A. millepora* transcriptome Cluster005198 (residue#251 in transcriptome Cluster005198 corresponds to residue#1 in B3EX00.1; see supplementary fig. S1A, Supplementary Material online). Clear homologs of SOMP1 were found by BLASTP search in each of the corallimorpharian species (table 1, supplementary table S3 and fig. S2, Supplementary Material online). An additional but incomplete SOMP1 sequence was found in *Corynactis* (e-value $\leq 5 \times 10^{-8}$, see supplementary table S3, Supplementary Material online). This sequence is missing the N-terminal region, and matches well but differs significantly in places; the N-terminus of Cory_cds.comp32376 matches Cory_cds.comp95534 from position #239. Note that the inclusion of the N-terminal region encoded in transcriptome Cluster005198 led to the identification of a transmembrane region (see supplementary fig. S1B, Supplementary Material online) not evident in the original analyses (Ramos-Silva et al. 2013). Transmembrane regions were predicted in the *Acropora* sequence Cluster005198 as well as in the related proteins from each of the three corallimorpharians, in each case close to the N-terminus (Cory_cds.comp32376 is missing the N-terminus, and therefore does not contain the predicted TM domain in this region). Alignment of the SOMP1 predicted amino acid sequences (see supplementary fig. S1, Supplementary Material online) identified a region that is highly conserved between coral and the corallimorpharians, corresponding to positions 198–283 in the *Acropora* sequence. Phylogenetic analysis at the amino acid level (fig. 1) grouped the corallimorpharian sequences together, to the exclusion of coral sequences.

Sequences matching *Acropora* SOMP2 were identified in all three corallimorpharians (table 1 and supplementary table S3, Supplementary Material online). However, whereas in the case of SOMP1 the corallimorpharian matches were of similar length to the *Acropora* sequence, in the case of SOMP2, the corallimorpharian matches were much shorter than the *Acropora* reference, the longest being less than half and the majority less than one third that of the *Acropora* sequence. *Acropora millepora* SOMP2 is a cysteine-rich protein (10% of residues), with complex repeated patterns centered around di-cysteine motifs. The corallimorpharian sequences are likewise cysteine-rich (7.3–10.8%), and the significance of the matches is due in large part to similarities in the cysteine arrangement patterns.

**Table 1**

Matches to the *Acropora millepora* SOMPs Reported by Ramos-Silva et al. (2013) in the Corallimorpharian Transcriptome Data Sets

| *Acropora millepora* | *Rhodactis indosinensis* | | *Ricordea yuma* | | *Corynactis australis* | |
|---|---|---|---|---|---|---|
| (Ramos-Silva et al. 2013) | E-value | Best match | E-value | Best match | E-value | Best match |
| SOMP1 (Cluster005198) | 3.00E−69 | Rhod_cds.comp46939 | 2.00E−63 | Riy_cds.comp77185 | 4.00E−64 | Cory_cds.comp95534 |
| SOMP2 (B7WFQ1) | 3.00E−26 | Rhod_cds.comp62430 | 1.00E−07 | Riy_cds.comp51189 | 1.00E−28 | Cory_cds.comp32915 |
| Hephaestin (B3EWZ9) | 0.00E+00 | Rhod_cds.comp98964 | 0.00E+00 | Riy_cds.comp73702 | 0.00E+00 | Cory_cds.comp93706 |
| CUB_and_peptidase_domain-containing_protein_1 | 2.00E−79 | Rhod_cds.comp102548 | | | 4.00E−82 | Cory_cds.comp95946 |
| SAARP1 (B3EWY6) | 3.00E−35 | Rhod_cds.comp74830 | | | 4.00E−31 | Cory_cds.comp88349 |
| SAARP2 (B3EWY8) | | | 6.00E−27 | Riy_cds.comp81594 | 8.00E−37 | Cory_cds.comp30507 |
| Mucin-like (B3EWY9) | 0.00E+00 | Rhod_cds.comp97279 | 0.00E+00 | Riy_cds.comp80209 | 0.00E+00 | Cory_cds.comp97042 |
| Coadhesin (B3EWZ3) | 0.00E+00 | Rhod_cds.comp101126 | 8.00E−74 | Riy_cds.comp77785 | 0.00E+00 | Cory_cds.comp94649 |
| SOMP8 (B3EWZ2) | | | 4.00E−07 | Riy_cds.comp92829 | 1.00E−13 | Cory_cds.comp52096 |
| MAM and LDL-receptor 1 (B3EWZ5) | 6.00E−141 | Rhod_cds.comp94084 | 1.00E−175 | Riy_cds.comp6460 | 0.00E+00 | Cory_cds.comp95654 |
| MAM and LDL-receptor 2 (B3EWZ6) | 0.00E+00 | Rhod_cds.comp101967 | 0.00E+00 | Riy_cds.comp19364 | 0.00E+00 | Cory_cds.comp95423 |
| Ectin (B3EWZ8) | | | 1.00E−28 | Riy_cds.comp21629 | | |
| MAM and fibronectin dcps (B3EX02) | 6.00E−47 | Rhod_cds.comp97180 | | | 5.00E−51 | Cory_cds.comp95861 |
| PKD1-related protein (B8UU59) | 0.00E+00 | Rhod_cds.comp99245 | 0.00E+00 | Riy_cds.comp79701 | 0.00E+00 | Cory_cds.comp97301 |
| ZP domain-containing protein (G8HTB6) | 4.00E−74 | Rhod_cds.comp93760 | 4.00E−142 | Riy_cds.comp76140 | 9.00E−155 | Cory_cds.comp88110 |
| EGF and laminin G dcp (B8UU78) | 0.00E+00 | Rhod_cds.comp98800 | 0.00E+00 | Riy_cds.comp80367 | 0.00E+00 | Cory_cds.comp86524 |
| Protocadherin-like (B8V7Q1) | 0.00E+00 | Rhod_cds.comp87389 | 0.00E+00 | Riy_cds.comp384406 | 0.00E+00 | Cory_cds.comp84737 |
| Collagen (B8V7R6) | 0.00E+00 | Rhod_cds.comp102025 | 0.00E+00 | Riy_cds.comp70659 | 0.00E+00 | Cory_cds.comp90994 |
| SOMP5 (B8VIU6) | 5.00E−63 | Rhod_cds.comp101521 | 2.00E−61 | Riy_cds.comp80304 | 4.00E−56 | Cory_cds.comp91462 |
| Neuroglian-like (B8VIW9) | 0.00E+00 | Rhod_cds.comp29308 | 0.00E+00 | Riy_cds.comp73158 | 0.00E+00 | Cory_cds.comp94627 |
| SOMP7 (B8WI85) | 1.00E−54 | Rhod_cds.comp87301 | 9.00E−136 | Riy_cds.comp77961 | 2.00E−143 | Cory_cds.comp91615 |

Note.—SOMP1 and SOMP2 (first two lines of the table) were previously thought to be coral-specific, whereas homologs of the other proteins listed have previously been identified in noncalcifying cnidarians. Data for corallimorpharian matches to galaxins and CAs are listed in tables 2 and 3 respectively. Details of sequences from corallimorpharians and other cnidarians matching SOMP1, SOMP2, SOMP3 and SOMP4 from *Acropora millepora* are listed in supplementary table S3, Supplementary Material online.
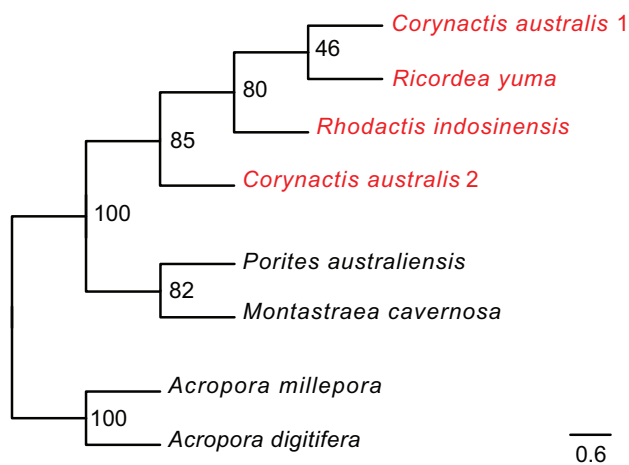


FIG. 1.—Phylogenetic analysis of cnidarian SOMP1 protein sequences. The values on nodes indicate bootstrap support in Maximum Likelihood analysis (for details, see Materials and Methods). Corallimorpharian sequences are in red, and are listed in supplementary table S3, Supplementary Material online. As indicated in the text, two SOMP1 sequences were identified in *Corynactis*. The coral sequences used in the phylogenetic analysis were *P. australiensis* (assembly_30918), *M. cavernosa* (cds.comp253373_c0_seq1), *A. millepora* (Amil_1.2.21472.m1) and *A. digitifera* (adi_v1.21723).

The *Acropora* SOMP2 sequence includes a predicted signal peptide, whereas the *Acropora* SOMP1 does not (Ramos-Silva et al. 2013), and this general pattern holds for most of the corallimorph and coral homologs of these (see supplementary table S3, Supplementary Material online). The absence of a signal peptide from SOMP1 suggests that this protein may not be secreted.

### Galaxin-Related Sequences

Seven sequences from *Corynactis* and three from *Ricordea* matched to *Acropora* galaxin or galaxin-like sequences (table 2), but no significant hits were identified in *Rhodactis* (BLASTP cut-off $e \leq 10^{-5}$). The apparent similarity between coral and corallimorpharian sequences is due largely to the repetitive structure and presence of di-cysteine motifs in these proteins and it is difficult to interpret the evolutionary significance of these data. However, the similarity in domain structure between two *Corynactis* sequences (Comp95728 and Comp63271) and the *Acropora* galaxin-like sequences is interesting; in each case an acidic domain follows the signal peptide and precedes the cysteine-rich region. The similarity in domain structure across this group of four sequences

**Table 2**

Galaxin-Related Sequences Identified in the Corallimorpharian Transcriptome Data Sets

| Gene ID | Species | Sequence ID | Best Database Match | | | Feature | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A. millepora/ NCBI Accession | E-Value | Identity | Amino Acid Residues | pI Value | TM Helix | Signal Peptide | # di-Cys |
| CauGalaxin1 | Corynactis australis | Cory cds.comp88726 c1 seq2 | Cluster013356/ADI50283 | 1.00E−35 | 36.99 | 258 | 9.18 | No | No | 17 |
| CauGalaxin2 | Corynactis australis | Cory_cds.comp3762_c0_seq1 | Cluster013356/ADI50283 | 1.00E−15 | 28.19 | 182 | 8.59 | No | No | 7 |
| CauGalaxin3 | Corynactis australis | Cory_cds.comp62533_c0_seq1 | Cluster013356/ADI50283 | 2.00E−16 | 32.85 | 191 | 9.12 | Yes | Yes | 9 |
| CauGalaxin4 | Corynactis australis | Cory_cds.comp94198_c0_seq1 | Cluster015317/ADI50283 | 5.00E−21 | 32.96 | 215 | 9 | No | Yes | 12 |
| CauGalaxin5 | Corynactis australis | Cory_cds.comp107187_c0_seq1 | Cluster013356/ADI50283 | 1.00E−08 | 25.98 | 127 | 8.88 | No | No | 5 |
| CauGalaxin6 | Corynactis australis | Cory_cds.comp95728_c0_seq5 | Cluster015317/ADI50283 | 8.00E−12 | 44.09 | 235 | 4.06 | Yes | Yes | 5 |
| CauGalaxin7 | Corynactis australis | Cory_cds.comp63271_c0_seq1 | Cluster013356/ADI50283 | 2.00E−11 | 49.23 | 199 | 4.68 | No | No | 5 |
| RyuGalaxin1 | Ricordea yuma | Ricordea cds.comp19957 c0 seq1 | Cluster013356/ADI50283 | 2.00E−35 | 31.56 | 387 | 8.87 | No | Yes | 23 |
| RyuGalaxin2 | Ricordea yuma | Ricordea_cds.comp51202_c1_seq1 | Cluster013356/ADI50283 | 1.00E−24 | 31.01 | 362 | 9.19 | No | Yes | 18 |
| RyuGalaxin3 | Ricordea yuma | Ricordea_cds.comp75875_c0_seq1 | Cluster015317/ADI50283 | 1.00E−24 | 34.22 | 211 | 9.46 | No | No | 12 |

suggests common origins, that is, that the galaxin and galaxin-like genes may have diverged prior to the coral-corallimorpharian split.

## Comparison of Carbonic Anhydrase Repertoires

Using a BLASTP cut-off of $e \leq 10^{-18}$, nine members of the alpha CA superfamily were identified in *R. yuma*, eight in *C. australis*, and four in *R. indosinensis* (table 3). Analyses of the features and phylogenetic relations of these (fig. 2 and see supplementary fig. S3, Supplementary Material online) indicate that the corallimorpharian carbonic anhydrase repertoires are considerably less complex than those of corals. In particular, the secreted and membrane-associated CA types are far fewer in number in corallimorpharians—two in both *R. yuma* and *Corynactis australensis*, and one in *R. indosinensis*, whereas $\geq 9$ have been identified in *A. millepora* and *P. lutea*. It is unlikely that this difference in numbers is due to the quality of the assemblies, as representation in other parts of the tree is comparable, and a high proportion of the core metazoan gene set can be retrieved from the corallimorphs (see supplementary table S1, Supplementary Material online). Rather, this difference is consistent with the expansion of this particular CA type being a requirement for calcification.

The phylogenetic analysis groups some corallimorpharian CA sequences with coral sequences having similar properties. For example, the clade labeled "Noncatalytic" in figure 2 comprises single sequences from *Corynactis* and *Ricordea* (CAcau6 and CAryu9 respectively) that both lack zinc-binding histidine residues critical for activity and are thus predicted to be inactive CAs (or CARPs). The *Acropora* (Cluster005523), *Nematostella* (XP001632501), and *Exaiptasia* (AIPGENE15469) sequences with which these corallimorpharian sequences cluster (fig. 2) are likewise predicted to be CARPs (Bertucci et al. 2013).

The clade of corallimorph sequences comprising CAryu2, CAcau4, and CArin1 (fig. 2) is likely to represent the mitochondrial matrix CAs of each species. Although a clear mitochondrial targeting sequence is predicted only in the former sequence (TargetP v1.1 prediction confidence 0.93), alignment of the three sequences indicates that both CAcau4 and CArin1 are missing the N-terminal regions that are required for mt localization. CAryu6 may also be targeted to mitochondria (TargetP v1.1 prediction confidence 0.8) but may have an incomplete N-terminus as there is no upstream methionine residue; the CAcau7 sequence match begins 43 amino acid residues into the CAryu6 sequence, hence the apparent absence of a targeting sequence in CAcau7 could also be due to its incompleteness. Transmembrane regions are predicted at the C-termini on both CAryu6 and CAcau7, whereas the three members of the CAryu2, CAcau4, CArin1 clade each appear to lack transmembrane regions, hence these two distinct clades of CAs most likely represent distinct types associated with the mitochondrial membrane and matrix respectively.

## Bicarbonate Transporters in Corals and Corallimorpharians

Searching the corallimorpharian data sets for SLC4 genes led to the identification of members of the SLC4α, β, and δ types, but not the SLC4γ-type. Clear homologs of the SLC4β and δ types were identified in each of the three corallimorpharians, but a clear counterpart of coral SLC4α proteins could only be identified in *Rhodactis* (fig. 3). Whilst these results are consistent with a key role for SLC4γ in coral calcification, in the phylogenetic analyses summarized as figure 3, the *Nematostella* SLC4β sequence is the nearest neighbor of the clade comprising the coral SLC4γ and SLC4β sequences, suggesting that the former diverged from an ancestral SLC4β type within the coral/corallimorph clade.

**Table 3**

Carbonic Anhydrase Sequences Identified in the Corallimorpharian Transcriptome Data Sets

| Gene ID | Species | Sequence ID | Best Database Match Accession Details | E-Value | Amino Acid Residues | TM Helices | Signal Peptide | Histidine Residues Present |
|---|---|---|---|---|---|---|---|---|
| CArin1 | *Rhodactis indosinensis* | Rhod_cds.comp92943_c1_seq3 | AAD32675 | 2.00E−82 | 272 | No | No | H1,H2,H3 |
| CArin2 | *Rhodactis indosinensis* | Rhod_cds.comp98814_c5_seq2 | ACJ64662 | 4.00E−73 | 331 | Yes | Yes | H1,H2,H3 |
| CArin3 | *Rhodactis indosinensis* | Rhod_cds.comp66236_c0_seq1 | ACJ64663 | 3.00E−117 | 265 | No | No | H1,H2,H3 |
| CArin4 | *Rhodactis indosinensis* | Rhod_cds.comp60899_c0_seq1 | ACE95141 | 3.00E−55 | 290 | No | Yes | H1,H2,H3 |
| CAcau1 | *Corynactis australis* | Cory_cds.comp91300_c1_seq1 | ACA53457 | 8.00E−110 | 608 | Yes | Yes | H1,H2,H3 |
| CAcau2 | *Corynactis australis* | Cory_cds.comp77787_c0_seq2 | ACJ64662 | 1.00E−83 | 327 | Yes | Yes | H1,H2,H3 |
| CAcau3 | *Corynactis australis* | Cory_cds.comp79250_c0_seq13 | ACE95141 | 1.00E−80 | 290 | No | Yes | H1,H2,H3 |
| CAcau4 | *Corynactis australis* | Cory_cds.comp85489_c0_seq1 | AAD32675 | 4.00E−89 | 260 | No | No | H1,H2,H3 |
| CAcau5 | *Corynactis australis* | Cory_cds.comp91311_c1_seq1 | ACJ64663 | 4.00E−119 | 264 | No | No | H1,H2,H3 |
| CAcau6 | *Corynactis australis* | Cory_cds.comp31183_c0_seq1 | XP_001632501 | 2.00E−88 | 281 | No | No | H2 |
| CAcau7 | *Corynactis australis* | Cory_cds.comp86633_c0_seq2 | XP_002154788 | 8.00E−63 | 321 | Yes | Yes | H1,H2,H3 |
| CAcau8 | *Corynactis australis* | Cory_cds.comp84345_c0_seq3 | ACE95141 | 4.00E−18 | 155 | No | Yes | H1,H2 |
| CAryu1 | *Ricordea yuma* | Riy_cds.comp78514_c0_seq1 | ACA53457 | 2.00E−102 | 614 | Yes | Yes | H1,H2,H3 |
| CAryu2 | *Ricordea yuma* | Riy_cds.comp66760_c0_seq1 | AAD32675 | 7.00E−85 | 299 | No | No (M) | H1,H2,H3 |
| CAryu3 | *Ricordea yuma* | Riy_cds.comp80554_c0_seq1 | ACJ64663 | 2.00E−117 | 264 | No | No | H1,H2,H3 |
| CAryu4 | *Ricordea yuma* | Riy_cds.comp77213_c0_seq2 | ACJ64662 | 5.00E−80 | 320 | No | Yes | H1,H2,H3 |
| CAryu5 | *Ricordea yuma* | Riy_cds.comp35028_c0_seq1 | ACE95141 | 1.00E−73 | 291 | No | Yes | H1,H2,H3 |
| CAryu6 | *Ricordea yuma* | Riy_cds.comp72038_c0_seq1 | ACE95141 | 5.00E−41 | 362 | Yes | No (M) | H1,H2,H3 |
| CAryu7 | *Ricordea yuma* | Riy_cds.comp66488_c0_seq1 | 4HBA_A | 8.00E−56 | 287 | No | Yes | H1,H2,H3 |
| CAryu8 | *Ricordea yuma* | Riy_cds.comp48510_c0_seq2 | ACE95141 | 2.00E−54 | 291 | No | Yes | H2 |
| CAryu9 | *Ricordea yuma* | Riy_cds.comp75888_c0_seq1 | XP_001632501 | 6.00E−83 | 313 | No | No | H2 |

## Other Genes Implicated in Calcification

Several nominally coral-specific genes have been implicated in calcification on the basis of temporal expression and spatial localization in *Acropora* (Grasso et al. 2008; Hayward et al. 2011; Moya et al. 2012). Three genes unique to *Acropora* (A036-B3, B036-D5, and C012-D9) are of particular interest because their expression was suppressed under acute $CO_2$ stress (Moya et al. 2012), a condition known to repress calcification in corals (see, e.g., Iguchi et al. 2014). No significant matches to A036-B3 and C012-D9 were found on BLASTP analyses of the corallimorpharian data. However, sequences matching B036-D5 were identified in both *Corynactis* and *Ricordea* with *E*-values of $e \le 9 \times 10^{-31}$ and $e \le 10^{-29}$, respectively (see supplementary material S1, Supplementary Material online). As in the case of *Acropora* B036-D5, (using InterProScan 5) no conserved motifs or sequence features could be identified in the corallimorpharian predictions.

## Discussion

To better understand how the ability to secrete an aragonite skeleton arose within the Scleractinia, we searched the transcriptomes of three representative corallimorpharians for homologs of genes implicated in coral calcification.

Several caveats apply in interpretation of the comparative data. First, although considerable bodies of data from three corallimorpharian species are presented and the assembly statistics are good (Lin et al. 2016; see also below), these data sets are incomplete. Thus, presences are more significant than absences. Second, comprehensive genome and transcriptome data are as yet available only for a very small number of coral species and it is unclear how well these reflect corals in general.

One relatively robust conclusion from the comparative analyses is that corallimorpharian genomes encode clear homologs of some genes previously considered to be coral-specific. The identification of homologs of SOMP1 and B036-D5 in corallimorpharians means that surprisingly few of the genes known or suspected to be involved in the deposition of the skeleton are actually unique to corals. Many of those genes that are unique to corals are cysteine-rich (SOMP2, galaxins *sensu stricto*, SCRiPs) and are likely to have been recruited from structural ECM proteins (Reyes-Bermudez et al. 2009). Subject to the caveats above, the apparent differences between corals and corallimorpharians in terms of the machinery involved in transport of inorganic carbon across membranes have important evolutionary implications.

Although corallimorpharians lack skeletons, most of the tropical shallow-water species (28 of the ~34 valid species) host the same photosynthetic symbionts as corals (*Symbiodinium* spp.), as do 2 of the 3 species studied here (*Rhodactis* and *Ricordea*), hence their CA repertoires are of particular interest. The transcriptome surveys clearly imply that the evolution of biomineralization in the Scleractinia required
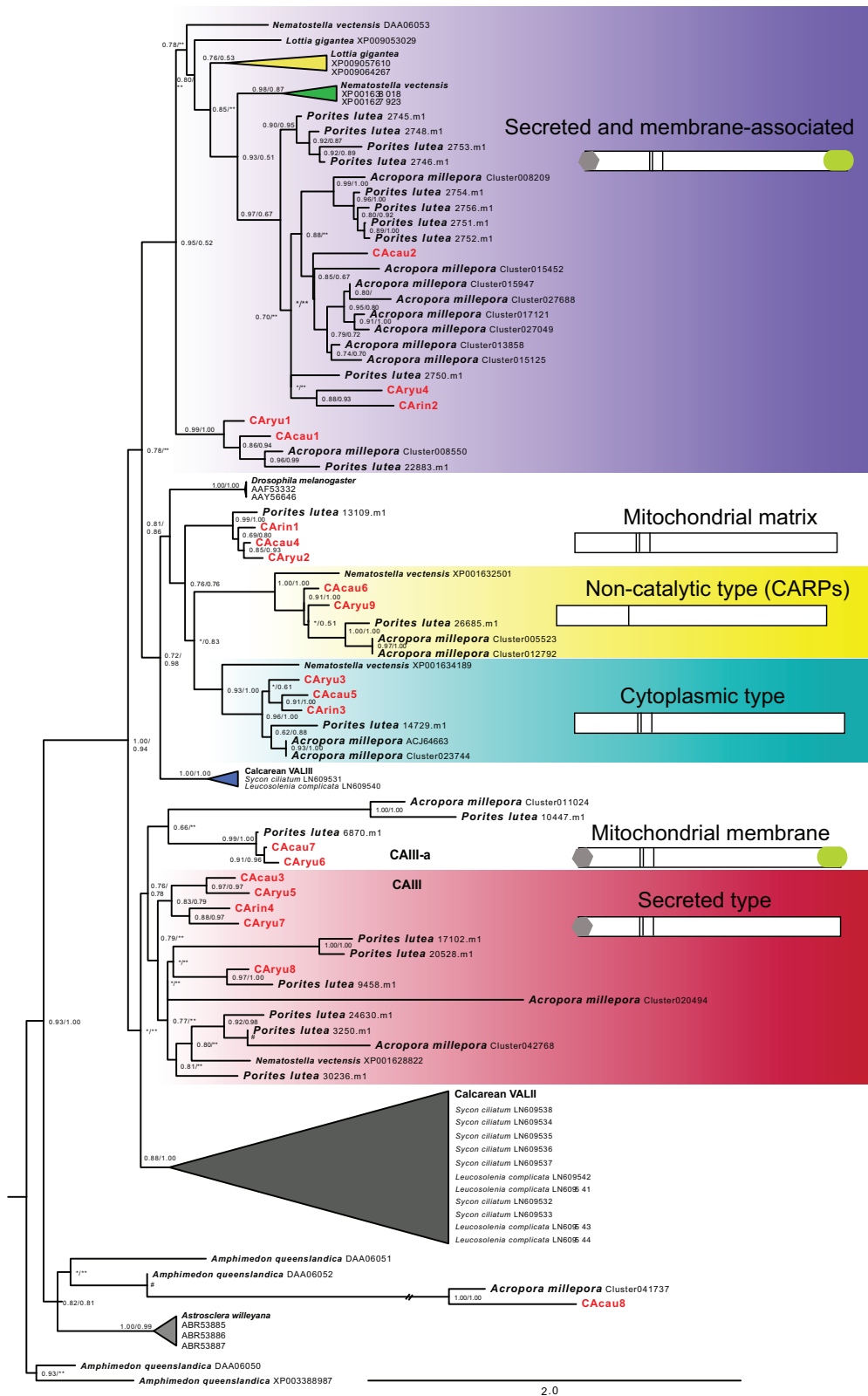
**Fig. 2.**——Phylogenetic relationships of cnidarian carbonic anhydrase sequences inferred from Maximum Likelihood (ML) and BI analyses. Note that ML and BI analyses recovered near identical tree topologies. The ML aLRT branch support values and BI posterior probabilities are indicated as ML/BI on the tree. *The aLRT value < 0.5; **The presence of polytomies in the Bayesian phylogeny, #Discrepancies between ML and BI trees. Corallimorpharian sequences are
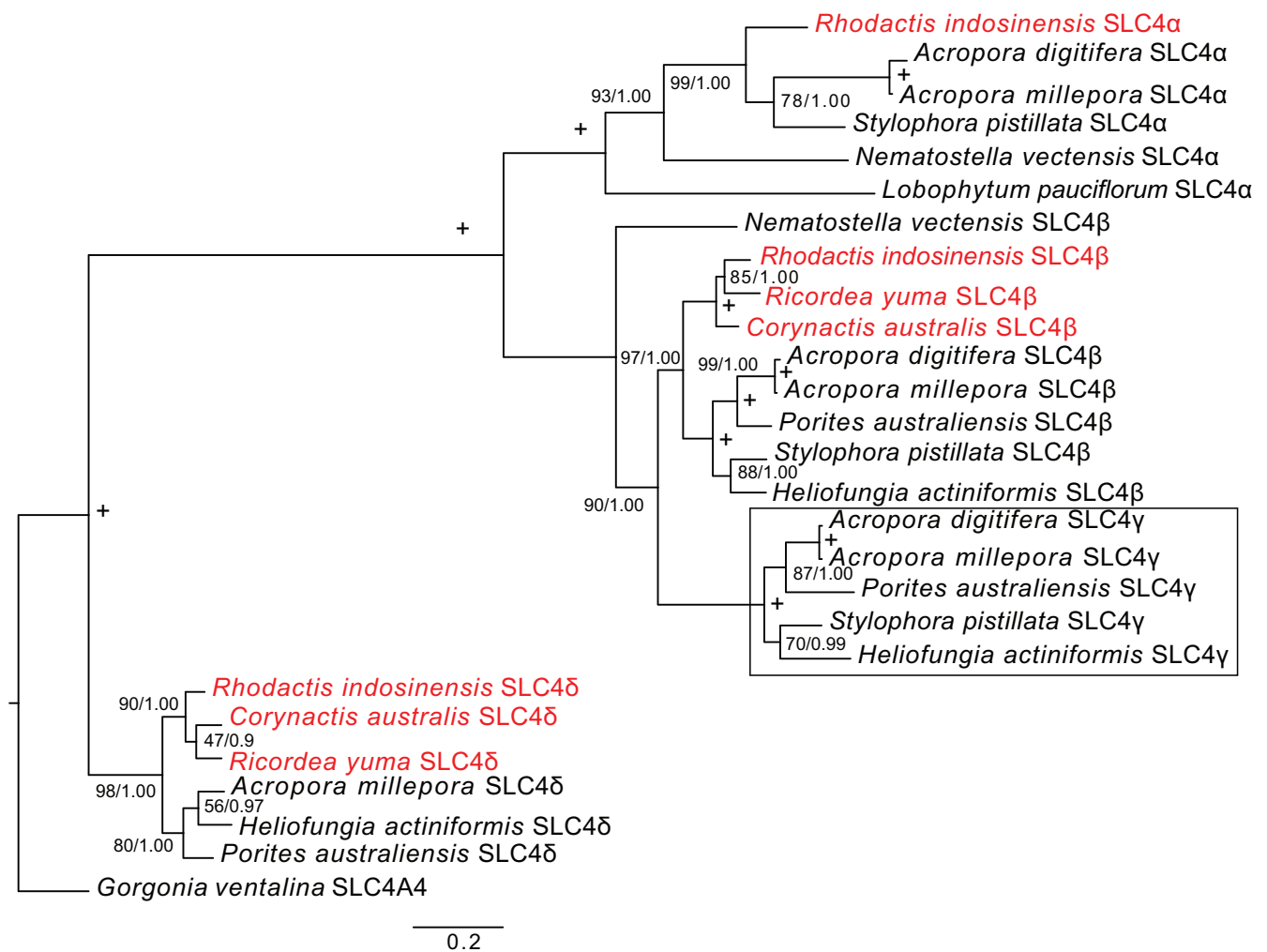
Fig. 3.—Phylogenetic tree of cnidarian SLC4 sequences inferred from Maximum likelihood (ML) and Bayesian analyses (BI). Bootstrap values under ML and posterior probabilities under BI are indicated as ML/BI at or close to nodes. For clarity in this figure values of 100/1.00 are represented by "+". Corallimorpharian sequences are in red, and the SLC4$\gamma$ clade is boxed.

expansion of the carbonic anhydrase repertoire, particularly of the secreted and membrane-associated type. Whereas a maximum of two sequences of this type was detected in the corallimorpharians surveyed (fig. 2), $\geq$ 9 were detected in *A. millepora* and *Porites lutea* (see supplementary material S2, Supplementary Material online). Although only a smaller number (four) could be identified in the *Pocillopora damicornis* transcriptome (see supplementary fig. S3, Supplementary Material online; Traylor-Knowles et al. 2011), this almost certainly reflects the incomplete nature of the assembly. Searching

the *P. damicornis* transcriptome assembly revealed that 390 of the 753 genes comprising the BUSCO core metazoan set (i.e. 46%) were missing, whereas far fewer were missing from the corallimorpharian data sets (see supplementary table S1, Supplementary Material online). Within the large clade of secreted and membrane-associated sequences, the branching pattern of coral sequences—distinct clades for *A. millepora*, *P. lutea* (and, in supplementary fig. S3, Supplementary Material online, also for *P. damicornis*)—contradicted expectations. The most likely explanation for this branching pattern is that the sequences

Fig. 2.—Continued

in red. The schematics associated with some clades summarize the main features characteristic of members of that clade; grey rectangles at the left end indicate the presence of a signal peptide, vertical bars indicate the presence of (the three) histidine residues involved in zinc binding (essential for catalytic activity) and green symbols at the right end indicate the presence of a membrane anchor. Likely cellular location is indicated above the schematics, but note that several coral members of the clade labeled "secreted type" are incomplete and in these cases the assignments are tentative. Structural features of each of the corallimorpharian CA sequences are shown in supplementary fig. S3, Supplementary Material online.

have undergone concerted evolution in each species, but alternative interpretations, including independent expansion of CA repertoires, cannot be rejected.

As in the case of the Scleractinia, the repertoires of secreted and membrane-associated CAs have likewise been independently expanded in other calcifying invertebrates; this phenomenon has been documented in the case of calcisponges (Voigt et al. 2014). The analyses presented as supplementary figure S3, Supplementary Material online, imply that a similar expansion has occurred in the mollusk, *Lottia gigantea*, but the incomplete nature of many of the sequences means that it is unclear whether signal peptides and transmembrane domains are present in the sequences that group with the secreted and membrane-associated CAs from Cnidaria (both features are predicted in the case of one member of this clade, *Lottia* XP_009053021).

Similar numbers of CAs, and a similar distribution across the various types, between *Corynactis*, a nonsymbiotic corallimorphian, and *Ricordea*, which normally hosts *Symbiodinium*, suggests that, whereas an expansion of the CA repertoire was necessary to enable calcification, it may not be a requirement to enable symbiosis. Consistent with this idea, preliminary analysis suggests that the CA complexity of symbiotic and nonsymbiotic sea anemones is similar. Conversely, on the basis of coral-sea anemone comparisons, it has previously been suggested that the recognition and maintenance of appropriate symbionts may require a more sophisticated innate immune repertoire (Shinzato et al. 2011). With the availability of data for symbiotic and nonsymbiotic corallimorpharians (this paper) and the symbiotic sea anemone *Exaiptasia* (Baumgarten et al. 2015) this idea can now be more thoroughly investigated.

Are corallimorpharians simply corals that have lost their skeletons, as has been suggested (Medina et al. 2006), or did calcification evolve after the Scleractinia diverged from Corallimorpharia? Data presented here and elsewhere imply that the evolution of calcification required at least one novel bicarbonate transport protein (SLC4γ; Zoccola et al. 2015) and an expansion of the carbonic anhydrase repertoire, particularly the secreted and membrane-associated types, as well as the recruitment of some ECM-derived genes to control the deposition process. If the corallimorpharian ancestor lost the ability to calcify, those genes—including a large number of carbonic anhydrase isoforms—have been lost, which is a less parsimonious explanation than if calcification post-dates the coral-corallimorpharian divergence. However, fewer loss events may be required if coral CAs are encoded by linked loci, and linkage seems likely given their apparent concerted evolution.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Literature Cited

Barbeitos MS, Romano SL, Lasker HR. 2010. Repeated loss of coloniality and symbiosis in scleractinian corals. Proc Natl Acad Sci U S A. 107:11877–11882.

Baumgarten S, et al. 2015. The genome of *Aiptasia*, a sea anemone model for coral symbiosis. Proc Natl Acad Sci U S A. 112:11893–11898.

Bertucci A, et al. 2013. Carbonic anhydrases in anthozoan corals-a review. Bioorgan Med Chem. 21:1437–1450.

Bhattacharya D, et al. 2016. Comparative genomics explains the evolutionary success of reef-forming corals. eLIFE 5:e13288.

Budd AF, Romano SL, Smith ND, Barbeitos MS. 2010. Rethinking the phylogeny of scleractinian corals: a review of morphological and molecular data. Int Comp Biol. 20:411–427.

Daly M, et al. 2007. The phylum Cnidaria: a review of phylogenetic patterns and diversity 300 years after Linnaeus. Zootaxa 1668:127–182.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165.

Davy SK, Allemand D, Weis VM. 2012. Cell biology of cnidarian-dinoflagellate symbiosis. Microbiol Mol Biol R. 76:229–261.

den Hartog JC. 1980. Caribbean shallow water Corallimorpharia. Zool Verh. 176:1–83.

Drake JL, et al. 2013. Proteomic analysis of skeletal organic matrix from the stony coral *Stylophora pistillata*. Proc Natl Acad Sci U S A. 110:3788–3793.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol. 300:1005–1016.

Fukuda I, et al. 2003. Molecular cloning of a cDNA encoding a soluble protein in the coral exoskeleton. Biochem Biophys Res Commun. 304:11–17.

Furla P, Galgani I, Durand I, Allemand D. 2000. Sources and mechanisms of inorganic carbon transport for coral calcification and photosynthesis. J Exp Biol. 203:3445–3457.

Grasso LC, et al. 2008. Microarray analysis identifies candidate genes for key roles in coral development. BMC Genomics 9:540.

Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59:307–321.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment program for Windows 95/98/NT. Nucleic Acids Symp Ser. 41:95–98.

Hamada M, et al. 2013. The complex NOD-like receptor repertoire of the coral *Acropora digitifera* includes novel domain combinations. Mol Biol Evol. 30:167–176.

Hayward DC, et al. 2011. Differential gene expression at coral settlement and metamorphosis - a subtractive hybridization study. PLoS One 6:e26411.

Heath-Heckman EAC, et al. 2014. Shaping the microenvironment: evidence for the influence of a host galaxin on symbiont acquisition and maintenance in the squid-vibrio symbiosis. Environ Microbiol. 16:3669–3682.

Iguchi A, et al. 2014. Responses of calcification of massive and encrusting corals to past, present and near-future ocean carbon dioxide concentrations. Mar Pollut Bull. 89:348–355.

Jackson DJ, Macis L, Reitner J, Degnan BM, Wörheide G. 2007. Sponge paleogenomics reveals an ancient role for carbonic anhydrase in skeletogenesis. Science 316:1893–1895.

Kitahara MV, et al. 2014. The "naked coral" hypothesis revisited-evidence for and against scleractinian monophyly. PLoS One 9:e94774.

Lin M-F, et al. 2014. Mitochondrial genome rearrangements in the Scleractinia/Corallimorpharia complex: implications for coral phylogeny. Genome Biol Evol. 6:1086–1095.

Lin M-F, et al. 2016. Corallimorpharians are not "naked corals": insights into relationships between Scleractinia and Corallimorpharia from phylogenomic analyses. PeerJ Preprints 4:e2151v1. https://doi.org/10.7717/peerj.2463.

Medina M, Collins AG, Takaoka TL, Kuehl JV, Boore JL. 2006. Naked corals: skeleton loss in Scleractinia. Proc Natl Acad Sci U S A. 103:96–100.

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. Mol Ecol. 21:2440–2454.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786.

Ramos-Silva P, et al. 2013. The skeletal proteome of the coral *Acropora millepora*: the evolution of calcification by co-option and domain shuffling. Mol Biol Evol. 30:2099–2112.

Ramos-Silva P, Marin F, Kaandorp J, Marie B. 2013. Biomineralization toolkit: the importance of sample cleaning prior to the characterization of biomineral proteomes. Proc Natl Acad Sci U S A. 110:E2144–E2146.

Reyes-Bermudez A, Lin ZY, Hayward DC, Miller DJ, Ball EE. 2009. Differential expression of three galaxin-related genes during settlement and metamorphosis in the scleractinian coral *Acropora millepora*. BMC Evol Biol. 9:178.

Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61:539–542.

Shinzato C, et al. 2011. Using the *Acropora digitifera* genome to understand coral responses to environmental change. Nature 476:320–323.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single- copy orthologs. Bioinformatics 31:3210–3212.

Sunagawa S, DeSalvo MK, Voolstra CR, Reyes-Bermudez A, Medina M. 2009. Identification and gene expression analysis of a taxonomically restricted cysteine-rich protein family in reef-building corals. PLoS One 4:e4865.

Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731–2739.

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

Voigt O, Adamski M, Sluzek K, Adamska M. 2014. Calcareous sponge genomes reveal complex evolution of α-carbonic anhydrases and two key biomineralization enzymes. BMC Evol Biol. 14:230.

Weis VM, Reynolds WS. 1999. Carbonic anhydrase expression and synthesis in the sea anemone *Anthopleura elegantissima* are enhanced by the presence of dinoflagellate symbionts. Physiol Biochem Zool. 72:307–316.

Xia X. 2013. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. Mol Biol Evol. 30:1720–1728.

Zoccola D, et al. 2015. Bicarbonate transporters in corals point towards a key step in the evolution of cnidarian calcification. Sci Rep. 5:9983.

**Associate editor**: Maria Costantini