# CORALLIMORPHARIAN TRANSCRIPTOMES AND THEIR USE TO UNDERSTAND PHYLOGENY AND SYMBIOSIS IN THE HEXACORALLIA

PhD Thesis submitted by

**Mei-Fang LIN, MSc (NTU)**

in September 2016

*Advisors*

*Prof Dr David John Miller*
*Dr Sylvain Forêt*
*Dr Chaolun Allen Chen*
*Dr Marcelo Visentini Kitahara*
*Dr Kyall Zenger*

***For the degree of Doctor of Philosophy***

***Department of Molecular and Cell Biology***
***James Cook University***

*To Su-Hsiang Chen Lin*

*Wen-Chang Lin*

*and Yu-Ning Lin*

## *Statement of Access*

I, the undersigned author of this thesis, understand that James Cook University will make it available for use within the university library and by electronic digital format, via the Australian digital theses network, for use elsewhere.

I understand that, containing unpublished work, a thesis has significant protection under the Copyright Act. Therefore, all users consulting this work should agree with the following statement:

**"**In consulting this thesis, I agree not to copy or partially paraphrase it in whole or in part without the written consent of the author, and to make proper written acknowledgement of any assistance which I have obtained from it.**"**

Beyond this, I do not place any restriction on access to this thesis.

Mei-Fang Lin

May 2016

## *Statement of contribution of others*

**Financial Assistance**

This research project was founded by James Cook University and ARC Centre of Excellence for Coral Reef Studies as detailed below:

- James Cook University Postgraduate Research Scholarship (2/2012-2/2016)
- Robert Logan Memorial Grants (Asia Pacific Coral Reef Symposium travelling funding support)
- James Cook University Research Graduate Scheme Grant (2014, 2015)
- ARC Centre of Excellence for Coral Reef Studies

**Scientific Contributions**

I gratefully appreciate the following people and institutes for the assistance in this research project.

- Prof David John Miller (James Cook University), Dr Chaolun Allen Chen (Academia Sinica), and Dr Marcelo V. Kitahara (Universidade Federal de São Paulo) for providing critical reviews on my research and writing.
- Prof David John Miller (James Cook University), and Dr Chaolun Allen Chen (Academia Sinica) for the discussion of experimental design.
- Dr Sylvain Forêt (Australian National University), Dr Kyall Zenger (James

Cook University) for the guidance for the next-generation sequencing (NGS) data analyses.

- Dr Sylvain Forêt (Australian National University) for the assisting the NGS analyses.
- Dr Chaolun Allen Chen (Academia Sinica) and Mr Wen-Hua Chou (Academia Sinica) for providing the computational resource and assistance in the data analyses in chapter 2.
- Mr Chia-Min Hsu (Academia Sinica) and Ms Carynn Chan (Academia Sinica) for the assistance in *Rhodactis* sample collection and preparation for NGS.
- Dr Chuya Shinzato (Okinawa Institute of Science and Technology Graduate University) for providing *Porites* transcriptome data.
- Prof Eldon Ball (Australian National University), and Dr David Hayward (Australian National University) for providing the assistance in cDNA library development.
- Mr Chao-yang Kuo (ARC Centre of Excellence for Coral Reef Studies), Dr Tom Bridge (ARC Centre of Excellence for Coral Reef Studies), Prof Andrew Baird (ARC Centre of Excellence for Coral Reef Studies), Prof Morgan Pratchett (ARC Centre of Excellence for Coral Reef Studies), Prof Joshua Madin (Macquarie University), and Mr Chris Benstead (Reef HQ Aquarium) for the assistance in the corallimorpharian samples collection.
- Dr Joana Figueiredo (Nova Southeastern University), Mr Simon Wever, and Mr Ben Lawes (James Cook University), and Mr Don Booth (A1

Aquarium World) for the assistance in the tank experimental design and set-up.

- Dr Victor Beltran (The Australian Institute of Marine Science), Dr Mia Hoogenboom (James Cook University), Dr Catherine Collier (James Cook University), Prof Joe Holtum (James Cook University), Dr Rebecca Lawton (James Cook University), Dr Nicholas Paul (James Cook University), Ms Boer Boa (University of Queensland), and Dr Shunichi Takahashi (Australian National University) for the providing instrumentation for the study of *Symbiodinium* physiology.
- Ms Sue Reilly (James Cook University) for corallimorpharian histological studies.
- Ms Kellie Johns (James Cook University) for the assistance in the research proposal writing.

## *Acknowledgements*

I would like to thank Professor Fu-Shiang Chia (1931-2011) for his encouragement to do the PhD study oversea. This PhD project would not have been proceeded without the support, guidance and encouragement from many people. For each of them, I would like to express my sincere gratitude.

First and foremost I am very thankful to my supervisor Professor David Miller, who is always very supportive, for his invaluable help and guidance in this study. He gave me this opportunity to undertake this project. During my PhD study, his helpful guidance in different aspects and stages inspires me with his enthusiasm and makes me be confident to develop my own ideas. Most importantly, he is always available whenever I needed advice.

I am thankful to Dr Sylvain Forêt for leading me to the world of “command lines”. His step-by-step instructions have helped me to develop the ability to work on the large-scale genomic data that is crucial for my study. Besides, his great help for the time when I was in Australian National University for the study was deeply appreciated. Dr Yin Hua, who is Dr Forêt’s postdoc researcher, assisted me the exploring of coral genomes, and provided me her time for the discussion of animal evolution. I also thank the members in Dr Forêt’s team for their friendly company and help when I was in Canberra.

Dr Marcelo Kitahara, who is one of my supervisors of my PhD, has remained one of my best friends. Discussing the studies or chatting with him is always inspiring and fruitful. I thank the experiences he shared and suggestions he gave for my study, and his useful comments for improving my manuscripts.

Very special thanks to Dr Chaolun Allen Chen who has been my supervisor for MSc and PhD from Taiwan to Australia. His suggestions and advice have been very helpful. He has played an important role in my study. Without his help, I would not have a clear direction for my PhD study, and may not be able to study corallimorpharians. I am appreciative of members of his team for their assistance in sample collection in Taiwan and technical support for this study.

Joining the Tropical Research Network has been a highlight in my life at James Cook University. I thank Prof David Miller to give me this opportunity to join them. I also thank Prof Alan Hauquitz (JCU), and Prof Marc Tennant (The University of Western Australia) for their suggestions and advice.

I came to Australian since 2012. Throughout this time my parents have given me the strongest and warmest support. Without their encouragement, I am not sure whether I could accomplish this study. I also thank two of my dear sisters for their cheer, help and care whenever I needed. I do not have enough words or writing skill to express my gratitude towards them.

I am thankful to Mr Chao-yang Kuo, Ms Yan-xiang Ow, Ms Meen-chee Hong, Dr Greg Torda, Dr Aurelie Moya, Ms Kate Qulgley, Ms Adriana Humanes, Ms Wiebke Wessels, Mr Amin Esmail, Ms Natalia Rodriguez, Ms Catalina Aquilar, Dr Colin Wen, Dr Sen-Lin Tang, Dr Liz Tynan, Dr Bette Willis, Prof Jeffrey Warner, Mr Glen Ewels, and Ms Anne Marie Falco for their respective help.

Finally, I express my gratitude towards Mr Yu-Ning Lin, and both of our parents and families. For Yu-Ning, his company, countless warm messages, support and love help me to overcome the frustration. I deeply appreciate all his support to help me complete this thesis.

## *Abstract*

Corallimorpharia, also known as coral-like anemones, belong to the sub-order Hexacorallia, the class Anthozoa, the phylum Cnidaria. Cnidaria is a group of the simplest animals at the tissue level of organization, which reflect ancestral characteristics and hence are important for understanding the evolution of metazoan genomes and developmental mechanisms. Morphologically, corallimorpharians resemble actinarians in the absence of a calcareous skeleton, but have internal anatomy more similar to scleractinians (corals) than to actinarians (sea anemones). Corallimorpharia comprises about 46 nominal species and generally these are minor components of the benthos in a wide range of habitats, which has led to corallimorpharian taxonomy and phylogeny being largely ignored in favor of their more obvious relatives.

Corallimorpharians are typical hexacorallians in several respects, e.g. 48.2% of corallimorpharians harbor endosymbiotic dinoflagellate algae (zooxanthellae) belonging to the genus *Symbiodinium*, there are both shallow water and deep-sea species (Append A), and they have similar overall symmetry of body structure.

Phylogenetic relationships between corallimorpharians and scleractinians remain particularly controversial and, on the basis of mitochondrial genome data, two main hypotheses have been proposed for their evolution. The “naked coral” hypothesis states that scleractinians were skeleton-less in the early Triassic, a time when carbonate deposition was suppressed globally, and corallimorpharians arose

by skeleton loss from a scleractinian ancestor at a later time (during the mid-Cretaceous) when the oceans had higher $CO_2$ levels. By contrast, the "scleractinian monophyly" hypothesis has corals as a monophylic lineage, with the Corallimorpharia as a close sister clade. According to Kitahara et al. (2014), the fundamental disagreement between the phylogenies based on nucleotide and amino acid sequences for mt proteins stems from the fact that none of the available models for sequence evolution adequately account for the observed data. Comparison of corallimorpharian and scleractinian mt genome architectures has shown that gene order in one species of corallimorpharian, *Corallimorphus profundus* (Moseley, 1877), is very similar to the canonical organization in scleractinians, indicating that this organism most closely reflects the coral <-> corallimorpharian transition (Lin et al., 2014).

In this study, three corallimorpharian transcriptomes were generated, and phylogenomic analyses of these used to provide insights into evolutionary relationships between scleractinians and corallimorpharians. The results strongly support scleractinian monophyly. Moreover, surveying the corallimorpharian transcriptomes led to the identification of homolologs of some skeletal organic matrix proteins (SOMPs) that were previously considered to be restricted to scleractinians; this is particularly significant given that surprisingly few of the proteins identified in the skeletal proteome are scleractinian-specific. Comparison of the carbonic anhydrase (CA) inventories of corallimorpharians with those of corals indicates that scleractinians 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. Additionally the similar numbers and distribution of the various CA types between the non-symbiotic corallimorpharians *Corynactis* and *Ricordea*, which normally host *Symbiodinium*, suggest that, whereas an expansion of the CA repertoire has been necessary to enable calcification, it may not be a requirement to enable symbiosis. These data also indicated that the evolution of calcification in scleractinians required relatively few completely new genes.

Prior to the present work, little was known about corallimorpharian-algal symbioses. This study investigated gene expression profiles of a tropical corallimorpharian during the re-establishment of symbiosis, providing the first large-scale dataset of this kind. The comparison of corallimorpharian transcriptomes under the symbiotic and aposymbiotic states indicated the similar responses of those shown in the scleractinians and anemones. The comprehensive comparison of genomic data from symbiotic, aposymbiotic and nonsymbiotic cnidarians supports the previous idea that host genes involved in symbiosis recognition and innate immune response for *Symbiodinium* tolerance play important roles in the establishment of symbiosis. Finally, the transcriptomic data indicate that glycogen biosynthesis occurs during the re-establishment of symbiosis in corallimorpharians, and that glycogen synthesis is likely to be more active during re-infection with a homologous rather than heterologous *Symbiodinium* strains. An additional interesting finding was the identification of a suite of genes unique to symbiotic corallimorpharians that were upregulated during the establishment of symbiosis. Although the functions of these genes remain to be explored, it is tempting to speculate to interpret this as evidence

for the independent evolution of symbiosis in corallimorpharians.

# Table of Contents

## List of Tables

## List of Figures

## Appendices

# Chapter I Introduction

*This chapter provides a general introduction to the Corallimorpharia, including introductions to the species studied, and an outline of the transcriptomic approaches used in this thesis to understanding of cnidarian evolution and endosymbiosis. Three published papers mentioned herein are attached as supplementary material to this Doctorial thesis.*

## 1. Introduction to Corallimorpharia

### 1.1 General characteristics of corallimorpharians

Corallimorpharians, also known as coral-like anemones or naked corals, belong to the anthozoan subclass Hexacorallia, and are effectively the "poor cousins" of their close relatives, the hard corals (Scleractinia). Phylogenetic relationships between corallimorpharians are unclear, as are their relationships with other anthozoans. The primary division within Class Anthozoa is between the subclasses Hexacorallia and Octocorallia. Although some authors consider these to be distinct classes within Cnidaria (see, for example, Kayal et al. 2013), the consensus view is that Hexacorallia and Octocorallia are subclasses. Six orders are usually recognized within subclass Hexacorallia: Zoantharia (zoanthids), Antipatharia (black corals), Ceriantharia (tube anemones), Actiniaria (sea anemones), Scleractinia (hard corals) and Corallimorpharia (coral-like anemones) (Daly et al. 2003, 2007). Note that Ceriantharia has sometimes been considered to constitute a distinct subclass (the

Ceriatharia) within Anthozoa (Stampar et al. 2014). A common feature of the six hexacorallian orders is the possession of spirocysts, a type of cnida with a single walled capsule and a tubule composed of tiny entangling sub-threads (Daly et al. 2007). While the consensus is that each of the six hexacorallian orders is monophyletic, views about phylogenetic relationships between these have changed over time (Daly et al. 2007). This is because basic anatomical features often do not circumscribe the same taxonomic groups (reviewed in Daly et al. 2003). For examples, mesentery filament histology groups zoanthids and actiniarians, while mesentery arrangement clusters scleractinians, actiniarians and corallimorpharians (Daly et al. 2003); the calcareous exoskeleton defines scleractinians, but this is not applicable to the other orders (e.g. Veron 1995). Recent analyses based on molecular data support the monophyly of each of the hexacorallian orders, although the relationship between Corallimorpharia and Scleractinia is controversial (Fig. 1.1, and see later section). Morphology and molecular data both support the hypothesis that these are the most closely related of the hexacorallian orders, but some authors view scleractinians as paraphyletic within a larger scleractinian/corallimorpharian clade.

Order Corallimorpharia Stephenson, 1937 comprises 46 nominal species (Cha 2001, Chen 1995, Lin et al. 2014, WoRMS 2004) that are morphologically similar to corals, but lack skeletons. Corallimorpharians are generally minor components of a wide range of habitats (Chen 1995, den Hartog 1980, WoRMS 2004), and these various factors have led to corallimorpharian taxonomy and phylogeny being

largely ignored in favor of their more obvious relatives Scleractinia and Actiniaria. Systematically, the Order Corallimorpharia is classified into three (Sideractiidae, Corallimorphidae and Actinodiscidae) or four (Sideractiidae, Corallimorphidae, Ricordeidae and Discosomidae) families (reviewed in Chen 1995). These families can be separated into two major ecophysiological groups (which are without taxonomic validity) – the zooxanthellate genera of the families Ricordeidae, Discosomidae and Actinodiscidae, and azooxanthellate corallimorpharians of the families Sideractiidae and Corallimorphidae (Chen 1995). However, den Hartog (1980) identified all azooxanthellate representatives of the order, including those inhabiting deep-water, belong to the family Corallimorphidae.

## 1.2 Corallimorpharian biology

Half of the valid corallimorpharian species (48.2%) are known to harbor endosymbiotic dinoflagellates (zooxanthellae) of the genus *Symbiodinium* (Append 1, den Hartog 1980, LaJeunesse 2002). When present, zooxanthellae play a pivotal role in host nutrition via translocation of photosynthates (Kuguru et al. 2007). Thus, as in other symbiotic anthozoans, corallimorpharians can become bleached when the symbiosis breaks down (Kuguru et al. 2007). Corallimorpharians share habitats, and thus compete for resources, with other benthic organisms on reefs (Borneman 2000), and this has led to the development of competitive strategies such as chemical defense systems (Hamner and Dunn 1980, Martin 1959), and the ability to move across the substratum (Muhando et al. 2002). In addition, corallimorpharians are often fast colonizers and can occupy large areas in coral reefs by means of

asexual reproduction (Chadwick-Furman and Spiegel 2000, Chen et al. 1995, den Hartog 1980, Lin et al. 2013), but the actual process of asexual reproduction is not well understood. Asexual reproduction by marginal budding has been observed in the case of one of the most widespread corallimorpharians in the Asia-Pacific, *Ricordea yuma* (Lin et al. 2013). This mode of reproduction has only previously been seen in the Red Sea species *Rhodactis rhodostoma* although Chadwick-Furman and Spiegel (2000) speculated that it may occur in other common corallimorpharians.

Corallimorpharians are generally believed to have annual reproductive cycles (Chadwick-Furman et al. 2000, Chen et al. 1995a, b, Holts and Beauchamp 1993) like those of scleractinians, but to date there have been only three studies of sexual reproduction in corallimorpharians including two tropical species (*Rhodactis indosinensis* and *R. rhodostoma*) and one temperate species (*Corynactis californica*). Species from these two genera seem to have different reproduction characters. For examples, aggregations of the tropical corallimorpharian *R. indosinensis* have the male polyps positioned on the edge of the aggregations and larger female polyps near the centre of the aggragations; these are produced by clonal replication and follow an annual gametogenic cycle in both sexes, and the oogenesis takes about nine months, while spermatogenesis three months (Chen et al. 1995a, b). In the case of the temperate corallimopharian *C. californica*, oocytes are present in polyps during most of the year and gradually increase in size until the annual spawning in summer (Eilat, North Red Sea) (Chadwick-Furman et al. 2000).

## 2. The evolution of Corallimorpharia

As indicated above, evolutionary relationships between the six hexacorallian orders are unclear (Daly et al. 2007), the relationship between corallimorpharians and scleractinians being particularly controversial. The “naked coral” hypothesis was proposed to explain the sudden appearance of diverse scleractinian fauna in the middle Triassic (Stanley and Fautin 2001). Under this hypothesis, the skeleton (calcium carbonate) has been an ephemeral feature during scleractinian evolution, and the Scleractinia were skeleton-less in the early Triassic, a time when carbonate deposition was suppressed globally (Stanley 2003). Molecular phylogenetics based on concatenated protein sequence data from 17 anthozoan mitochondrial (mt) genomes gave support to the “naked coral” hypothesis (Medina et al. 2006). In the Medina et al. (2006) analyses, the Scleractinia were paraphyletic, corallimorpharians being closer to the Complex clade scleractinians (Complexa) than were Robust clade scleractinians (Robusta), the interpretation being that the Corallimorpharia arose by skeleton loss from a scleractinian ancestor during the mid-Cretaceous, when the oceans were more acidic than today (Medina et al. 2006). Whilst this is an attractive hypothesis, phylogenetic analyses based on mitochondrial nucleotide sequences are inconsistent with the “naked corals” scenario, instead supporting scleractininan monophyly (Stolarski et al. 2011, Kayal et al. 2013, Kitahara et al. 2014). According to Kitahara et al. (2014), the fundamental disagreement between the phylogenies based on nucleotide and amino acid sequences for mt proteins stems from the fact that none of the available models

for sequence evolution adequately account for the observed data. One possible explanation for this is that within the Robusta coral lineage, mt DNA repair processes that are believed to be ancestral within Anthozoa, have somehow become impaired (Kitahara et al. 2014, Lin et al. 2014); consistent with this, within the Robusta, mt genomes are significantly more biased in base composition and appear to be evolving faster than in the Complexa or Corallimorpharia. Although the issue of Corallimorpharia/Scleractinia relationships is equivocal, the architecture of the mt genome of the deep-sea corallimorpharian, *Corallimorphus profundus* (Corallimorphidae), is strikingly similar to the canonical scleractinian pattern, whereas there are major differences in other corallimorpharians (Lin et al. 2014). Mitochondrial genome architecture therefore implies that *C. profundus* represents the coral <-> corallimorpharian transition, which occurred either prior to the coral/corallimorpharian split or (in the context of the naked corals scenario) the point of skeleton loss (Lin et al. 2014).

Mt sequence data are clearly not appropriate for addressing the relationship between Scleractinia and Corallimorpharia (Emblem et al. 2011, Kitahara et al. 2014, Lin et al. 2014). Nuclear sequence data are therefore required; over the last few years, appropriate molecular datasets have become available for several corals (e.g. Technau et al. 2005, Shinzato et al. 2011), but no comparable resources are yet available for corallimorpharians. There is an obvious need for transcriptomic and genomic resources for corallimorpharians, from which can shed the light on the resolution of the relstionship between scleractinians and corallimorpharians.

## 3. Symbioses between hexacorallians and dinoflagellates

The diverse dinoflagellate genus collectively known as *Symbiodinium* occurs in what are viewed as mutualistic symbioses with a wide range of marine invertebrates, the best known example of which are the Scleractinia, this partnership being largely responsible for providing both the trophic and structural foundations of coral reef ecosystems (Ganot et al. 2011). A wide range of other hexacorallians, including many sea anemones, zoantharians, and corallimorpharians (LaJeunesse 2001, 2002), form similar associations with *Symbiodinium* strains.

In corallimorpharians, as in scleractinians, zooxanthellae are distributed in endodermal tissues (Kuguru et al. 2007), but symbiosis in tropical corallimorpharians has received very little attention beyond the (untested) assumption being that the relationship has the same molecular bases as in corals. In the case of scleractinians, photosynthetic carbon fixation by the zooxanthellae can supply approximately 90% of the energy requirements of the colony, although the technique and the translocation rates estimated are questionable (reviewed in Davy et al. 2012), in return for which the scleractinian provides a stable and nutrient rich environment for the symbiont (Venn et al. 2008). However, environmental perturbations such as elevated water temperature and high solar radiation can induce dysfunction of symbiosis, with the result that the zooxanthellae are expelled, a process known as bleaching (Goreau and Hayes 1994, Glynn 1996, Baird et al.

2009). Cnidarian bleaching is frequently referred to as a consequence of "a breakdown in the mutualistic symbiosis between the host cnidarians and their unicellular photosynthetic dinoflagellate symbionts" (Weis 2008).

Whilst bleaching is most obvious in scleractinians, other marine invertebrates, from sea anemones and sponges to giant clams, also undergo loss of dinoflagellate endosymbionts under stress (reviewed in Davy et al. 2012). Thus much of the blame for the declines in the state of coral reefs that have occurred over the past two decades has been attributed to "bleaching" in the wider sense (reviewed in Brown 1997, Wellington et al. 2001, Kuguru et al. 2007, Glynn et al. 1993, Vicente 1990, Leggat et al. 2003). Stressors implicated in bleaching include anomalous water temperatures, high levels of solar irradiance, and various anthropogenic disturbances, including pollution, sedimentation and fresh water runoff (Goreau and Hayes 1994, Glynn 1984, 1996, 2000, Salm et al. 2001, West et al. 2003, Hoegh-Guldberg and Fine 2004, West and Salm 2003, Evans 1977, Jackson et al. 1989, West and Salm 2003). A recent study suggested that some reefs on the Great Barrier Reef (GBR) have declined as a consequence of climate change - bleaching events driven by elevated water temperature combined with ocean acidification (Hoegh-Guldberg et al. 2007). Ultimately, the causes of coral bleaching are diverse, complex and multifactorial, and there is no doubt that in recent times mass bleaching events have increased in frequency, intensity and geographical extent. This trend will continue until the underlying molecular mechanisms are better understood, and the triggers are dealt with.

## 3.1 The mechanism of cnidarian bleaching

"Bleaching is a stress response to environmental perturbation" (Weis 2008). Although the phenomenon of coral bleaching has been recognized for some time, the intracellular mechanisms involved are still poorly understood. Temperature shocks were the first cause of zooxanthellae loss from corals to be identified (Steen and Muscatine 1987). Based on experiments carried out on tropical anemones and corals, the mechanism proposed to account for the observed bleaching was dysfunction of cell adhesion in the host (Gates et al. 1992). Subsequently, greater emphasis has been placed on the role of the symbiont in bleaching under thermal stress, on the basis that photoinhibition and then accumulation of oxidative damage at photosystem II (PSII) in the symbiont might initiate the breakdown (Iglesias-Prieto et al. 1992).

The current consensus is that thermally-induced coral bleaching is initiated by reactive oxygen species (ROS) formation in both symbiont and host cells, causing oxidative stress in the host cell and metabolic dysfunction that can lead to cell death (Lesser 2006, Baird et al. 2009). Weis (2008) concluded that in the symbiont at least three inter-related processes lead to ROS generation under heat and/or high light stress: (1) damage to the D1 protein, which is part of the water-splitting complex in photosystem II complex resident in the thylakoid membranes, leading to a backup in excitation energy and the dysfunction of photosystem II; (2) damage to ribulose bisphospate carboxylase oxygenase (Rubisco), the enzyme responsible for primary carboxylation, resulting in reduced consumption of the ATP and NADPH generated by the light reaction, again leading to dysfunction of photosystem II; and (3)

damage to the thylakoid membranes, causing uncoupling of electron transport, ultimately leading to the situation where the photosynthetic apparatus continues to generate electrons but is unable to produce ATP or NADPH.

It is clear that both host and the zooxanthellae are capable of a degree of adaptation or acclimation to environmental stress by employing protective mechanisms that include increased expression of heat shock proteins and antioxidant enzymes (Weis 2008). For instance, tropical corallimorpharians react to exposure to high levels of ultraviolet radiation by synthesizing enzymatic antioxidants against oxygen radicals or by physical migration (Kuguru et al. 2010). In parallel, the zooxanthellae in UV-irradiated corallimorpharian hosts reduced chlorophyll content, adjusting the efficiency of light absorption and utilization according to the level of irradiance (Kuguru et al. 2010).

Whilst a degree of acclimation is clearly possible, it remains unclear whether corals (including scleractinians and corallimorpharians) will be able to adjust or adapt sufficiently quickly to survive the challenges of climate change. According to the adaptive bleaching hypothesis (ABH), bleaching may enable scleractinians to adopt different zooxanthellae better suited for a new environment (Buddemeier and Fautin 1993), although species may differ in the extent to which they are able to “shuffle” symbionts (Baird et al. 2007). Laboratory experiments have verified that bleached hosts can reestablish symbiont populations from low residual tissue zooxanthellae concentrations (Kinzie III et al. 2001). In addition, scleractinians appear to be able to acquire symbionts from the environment after a bleaching episode, but this may be a temporary phenomenon and thus not helpful of enhancing the ability to

acclimatize to higher water temperatures associated with global warming (Coffroth et al. 2010). In summary, the molecular mechanisms underlying symbiont secondary acquisition are unclear and there is considerable disagreement about whether "adaptive" bleaching is a significant factor on biological time scales (Goulet 2006, Baird et al. 2007).

### 3.2 The process of symbiont acquisition

It has been proposed that the establishment and persistence of symbioses involves six phases: (1) initial host-symbiont contact, (2) symbiont engulfment, (3) dynamic intracellular sorting of the symbionts, (4) proliferation of the symbionts within the host tissues, (5) dynamic stability, and (6) dysfunction and breakdown (Davy et al. 2012). The major processes underlying these phases include recognition and phagocytosis, regulation of host-symbiont biomass, and metabolic exchange and nutrient trafficking (and calcification in corals) (Davy et al. 2012). At initial contact, signaling events presumably facilitate host-symbiont recognition, followed by engulfment of the symbiont by phagocytosis, and subsequent persistence of compatible symbionts.

Microbe associated molecular pattern (MAMP)-pattern recognition receptor (PRR) interactions are thought to be important during the onset of cnidarian-algal symbiosis (reviewed in Davy et al. 2012), and there is some evidence that specific C-type lectins may participate this process (Kvennefors et al. 2010, Vidal-Dupiol et al. 2009). In addition to lectins, cnidarians have homologs of many of the innate immunity PRRs that are present in vertebrates (Kimura et al. 2009, Lange et al.

2011, Miller et al. 2007). Several PRRs have been implicated in communication between the host and symbiont.

When symbionts enter the host, the stability of the symbiosis requires coordination between host cell growth and symbiont proliferation. It has been suggested that the host regulates symbiont growth by controlling the supply the inorganic nutrients (Rees 1986). This idea was supported by evidence that the host regulated the symbiont cell division *in hospite* (Smith and Muscatine 1999). Metabolic exchange is important to the cnidarian-algal symbiosis; much of the photosynthetically fixed carbon is translocated to the host, where it can support skeletogenesis in (reviewed in Davy et al. 2012). For this process, the enzyme carbonic anhydrate (CA), which catalyzes the interconversion of $HCO_3^-$ and $CO_2$, plays a key role in carbon assimilation (Isa and Yamazato 1984). The expression of several CAs was suppressed by elevated $pCO_2$ in *Acropora millepora*, implicating specific CA isoforms in calcification (Moya et al. 2012).

There is uncertainty around the forms in which photosynthetic carbon is translocated to the host, although candidate compounds include glycerol, glucose, succinate, and fumarate (reviewed in Davy et al. 2012). This area is controversial; some studies have concluded that glycerol is the primary form (24.8-95.09%) of fixed carbon released in the cnidarian-algal symbioses (Sutton and Hoegh-Guldberg 1990), while the application of other approaches suggest that glucose but not glycerol is translocated (Ishikura et al. 1999). Recently, the observation of carbon and nitrogen fluxes in coral-algal endosymbiosis indicated that lipid droplets and glycogen granules in the coral tissue are sinks for translocated carbon

photosynthates by dinoflagellates and confirm their key role in this association (Kopp et al. 2015). However, the controversy remains unresolved and the molecular mechanisms involved in this process remain unclear.

3.3 Genomic approaches to understanding cnidarian-algal symbiosis

Surprisingly few clear differences in transcriptomic profiles have been identified between symbiotic and aposymbiotic cnidarians, despite the application of expressed sequence tag (EST), proteomic, and microarray technologies (reviewed in Meyer and Weis 2012), although a few (17) differentially expressed genes were identified in the actiniarian *Anemonia viridis* when aposymbiotic tissues (epidermis; ectoderm) and symbiotic tissues (gastrodermis; endoderm) were compared (Ganot et al. 2011). Possible explanations for these results include (1) technical limitations associated with detecting changes in a small proportion of the total number of genes, (2) high between-individual variation, which has frequently been observed in anemones and scleractinians, effectively swamping transcriptomic profiles, and (3) the near absence of true 'symbiosis genes' (reviewed in Meyer and Weis 2012).

Although relatively few genes differ in expression between the symbotic and aposymbiotic states, a number of genes have been implicated in the establishment, maintenance and collapse of symbiosis. For example, the scavenger receptor B/CD36 (Rodriguez-Lanetty et al. 2006, Lehnert et al. 2014), Sym32 (Ganot et al. 2011), and Rab homologs (reviewed in Davy et al. 2012) have been implicated in establishment or maintenance of symbiosis, and nucleotide oligomerization domain

(NOD)-like receptors (NLRs), Bcl-2 proteins and sphingosine 1-phosphate phosphatase are thought to play roles in maintenance and/or breakdown of symbiosis (Dunn and Weis 2009, Meyer and Weis 2012, Rodriguez-Lanetty et al. 2006, Perez and Weis 2006). Genes implicated in metabolic exchange between symbiont and host include CAs and $H^+$-ATPase genes for $CO_2$ transportation (Furla et al. 2000, Ganot et al. 2011), glutamine synthase (GS) and glutamate dehydrogenase (GDH) for ammonium assimilation (Yellowlees et al. 2008), and Niemann-Pick type C proteins (NPC1 and NPC2) for sterol-trafficking (Kuo et al. 2010, Ganot et al. 2011, Dani et al. 2014).

Consistent with earlier findings (Rowan et al. 1997), it is now clear that scleractinians can change the dominant symbiotic algae depending on the environment and growth stage (Howells et al. 2013, Yuyama and Higuchi 2014). The effects on scleractinians of association with different *Symbiodinium* clades have been investigated using a variety of approaches (Klueter et al. 2015), but to date no clear general patterns have been identified. It is clear that some symbiont strains are incompatible with some hosts (e.g. Weis et al. 2001, Loram et al. 2007), but many hosts are promiscuous with respect to symbiont strains taken up, a "winnowing" phase later determining the long-term nature of the association. In corallimorpharians, as in some sea anemones, symbiosis is facultative rather than obligate and, given their close relationship with corals, exploring molecular events around the establishment of symbiosis in corallimorpharians may provide new perspectives on these phenomena more generally

### 3.4 Experimental advantages of corallimorpharians

One requirement for understanding the molecular mechanisms underlying dinoflagellate / cnidarian symbioses is the ability to deplete the host of symbionts, with the aim of providing aposymbiotic samples for comparative purposes and for experimental manipulation. Menthol treatment has been developed as a simple and relatively mild method for the generation of aposymbiotic corals (Wang et al. 2012). Based on the similarity and close evolutionary relationship between corallimorpharians and corals, the former could be good surrogates for coral research and adaptation of the menthol treatment method for corallimorpharians should also be relatively straightforward. Corallimorpharians have several advantages over corals for purposes of investigating the molecular mechanisms underlying bleaching, the infection process and the bases of symbioses – they can be propagated rapidly using simple techniques, and lack skeletons (which are an impediment to cellular isolation and analyses). If symbiotic corallimopharians can be bleached using a menthol-based treatment, this provides an opportunity to fill some of the gaps in our understanding of anthozoan biology including the symbiosis.

## 4. The species selected for study

The work described in this thesis is based on three corallimorpharian species, selected as representatives of diverse taxonomic groups within the order, i.e. *Corynactis australis* (Corallimorphidae), *Rhodactis indosinensis* (Discosomidae)

and *Ricordea yuma* (Ricordeidae). The collection sites of the species studied are illustrated in Fig. 1.2. The morphology of the two Australian species (*C. australis* and *R. yuma*) is explained in the next few pages.

*Corynactis australis* Haddon & Duerden, 1896 is an azooxanthellate corallimorpharian that has a wide distribution in temperate waters (Fautin 2011) and was collected for the work described here from the New South Wales coast. *Ricordea yuma* (Carlgren, 1900) is a tropical and zooxanthellate corallimorpharian that is distributed widely in the Indo-Pacific Ocean (Fautin 2011, den Hartog et al. 1993). *Rhodactis indosinensis* Carlgren, 1943 is distributed in shallow tropical and sub-tropical waters of eastern Asia (Chen et al. 1995a), was collected from sub-tropical water in Taiwan but, for this species, dissection was not attempted because insufficient samples were available to do so. The taxonomy of these species is summarized below.

Order Corallimorpharia Carlgren, 1940

Family Ricordeidae Watzl, 1922

Genus *Ricordea* Duchassaing & Michelotti, 1860

Species *Ricordea yuma* (Carlgren, 1900)

*Ricordea yuma*, similar to *Ri. florida*, has solitary to colonial forms, but solitary specimens are dominant in aquaria. The column is flat and spreading over the substrate. By the observation of 20 specimens in the aquarium, some specimens can reach a diameter of 12 cm when completely extended. Oral disc is circular or

elliptic. The number of oral openings of each polyp varies from 1 to 2. The marginal tentacles are usually more elongate and slender than the discal tentacles. Specimens are predominantly brown, particularly in the marginal tentacles (Fig.1.3A). Some short tentacles close to the central part are dull green or orange. The discal radial rows are brighter than the others. Under UV illumination (Fig.1.3B), *R. yuma* dissplays green and yellow fluorescence. The red color visible is due to chlorophyll fluorescence from the symbiotic zooxanthellae (Lesser et al. 2004).

Anatomy was observed from two specimens for longitudinal (Fig. 1.3C) and cross sections (Fig.1.3D). Stomodaeum distinctly ciliated, with longitudinal endodermal ridges supported by slightly thickened mesogloea (pink) (Fig. 1.3D, 1.3I) The tentacle dissection (Fig.1.3E) shows the distribution of the zooxanthellae in the tentacle. It is clear that more zooxanthellae were located at the base of the tentacle. Similar with the observation of *R. florida*, the mesogloea is thick (e.g. 1.1-3.5 um at tentacle tip) (Fig. 1.3C, 1.3G) and the endodermal circular muscle layer is present (den Hartog 1980).

Family Corallimorphidae Hertwig, 1882

Genus *Corynactis* Allman, 1846

Species *Corynactis australis* Haddon & Duerden, 1896

*Corynactis australis* has small, gregarious forms (polyp is about 1-2 cm in diameter). The column and oral disc are supple. Column variable in shape, when fully expended often wide and trumpet-shaped, mammiform when contracted. Oral disc slightly exceeds the diameter of the base. Tentacles developed with distinct,

globular acrospheres. Specimens' color is orange to orange-brown, and bright orange, rose or white for the acrospheres (Fig.1.4A). Under UV light, only green color was reflected (Fig.1.4B). Mesogloea (pink) is thickened and plaited. Similar to most *Corynactis* spp, *C. australis* has a distinct sphincter (Fig. 1.4C-E). This species is distributed in temperate water. Zooxanthellae are absent.

Family Discosomidae Verrill, 1869

Genus *Rhodactis* Milne Edwards & Haime, 1851

Species *Rhodactis indosinensis* Carlgren, 1943

*Rhodactis indosinensis* forms large, bowl-like small aggregations or smaller patches around the big aggregations on reef flats. The polyp is wide and discoidal. The oral disk is covered with short discal tentacles, surrounded by thick, split-ended marginal tentacles. Two or three mouths on one disc are often displayed during asexual reproduction. The specimens' color is brown, green and rust. Similar to the Great Barrier Reef species *R. howesii* Saville-Kent, 1893, the mesogloea is thin, and peripheral zone (between marginal tentacles and discal tentacles) indistinct.

## 5. Objectives

This Doctoral Dissertation aims to contribute to two fields of study - firstly, to understand and clarify the equivocal evolutionary relationship between corallimorpharians and corals, and secondly, to provide new perspectives on the cnidarian-dinoflagellate symbioses.

Chapter II Understanding the phylogenetic relationships of corallimorpharians and scleractinians

In this chapter, the phylogenetic relationship between corallimorpharians and scleractinians will be examined using transcriptomic approaches. Orthologous sequences identified from a range of cnidarian species are used as the basis for comprehensive phylogenetic analyses at both the amino acid and nucleotide levels in order to reconstruct the phylogeny of corallimorpharians and scleractinians.

Chapter III Comparative transcriptomics of corallimorpharians: implications for scleractinian evolution

In this chaper, corallimorpharian and scleractinian genomes and transcriptomes will be compared, focusing specifically on the distribution of genes implicated in coral calcification, with the aim of understanding the evolution of this trait.

Chapter IV Understanding the molecular mechanism of symbiont acquisition in corallimorpharians

This chapter is based on an experiment in which bleached *Ricordea yuma* individuals were reinfected with either the naturally occurring zooxanthellae strain or a strain isolated from a different corallimorpharian species. Transcriptomic data obtained at key time points post-infection provides new perspectives on the molecular interactions likely to be occurring between host and symbiont in both compatible and heterologous zooxanthellae strains.

Chapter V General discussion of the outcomes in this doctorial thesis

A summary of the results and a general discussion of this PhD study will be provided.

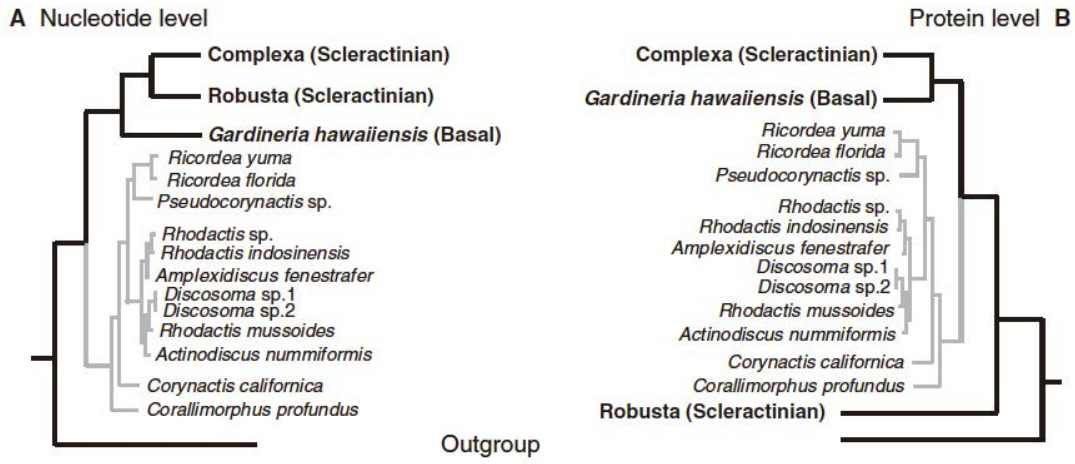



**Figure 1.1** Alternative phylogenetic hypotheses for relationships between Scleractinia and Corallimorpharia based on mt genome nucleotide sequences (A) or the amino acid sequences of the proteins that they encode (B) (Lin et al. 2014).

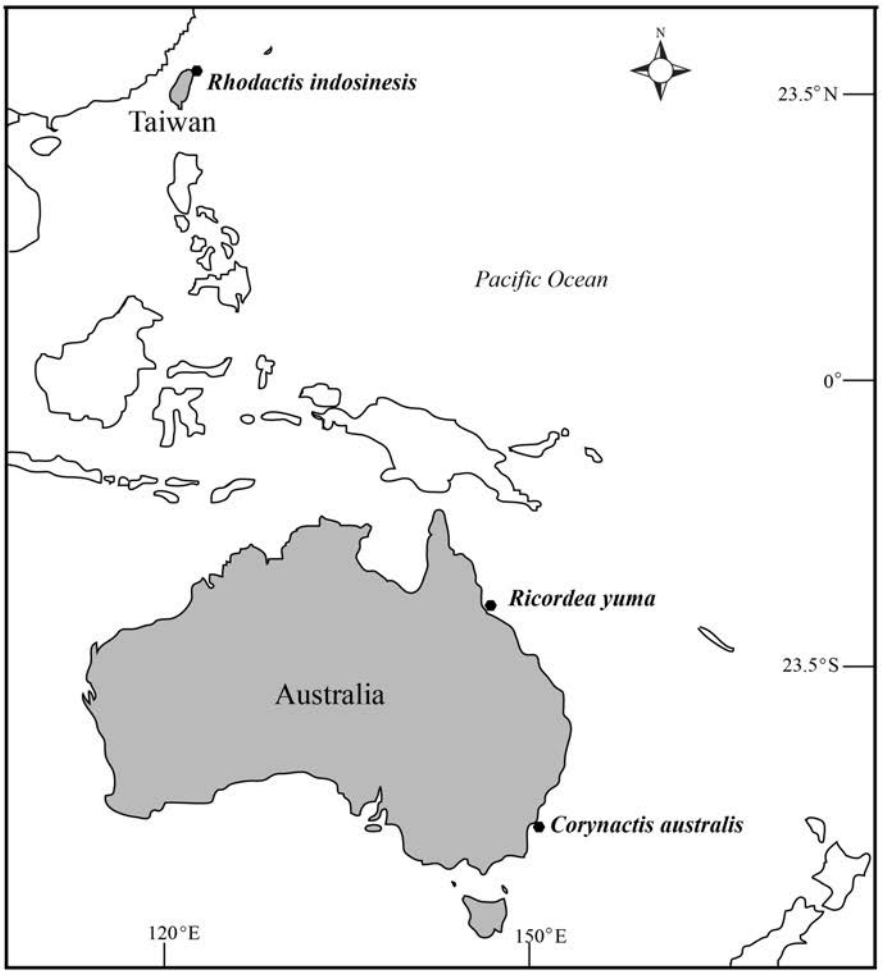



**Figure 1.2** Specimen collection sites.

*Rhodactis indosinesis*: Beitou Fishing Harbor (25°08'40"N, 121°47'23"E), 3 meters depth, *Ricordea yuma*: the Great Barrier Reef (18°25'35.20"S, 146°41'10.91"E), unknown depth, *Corynactis australis*: Jervis Bay, New South Wales (35° 4'14.11"S, 150°41'48.20"E), 1 meter depth.

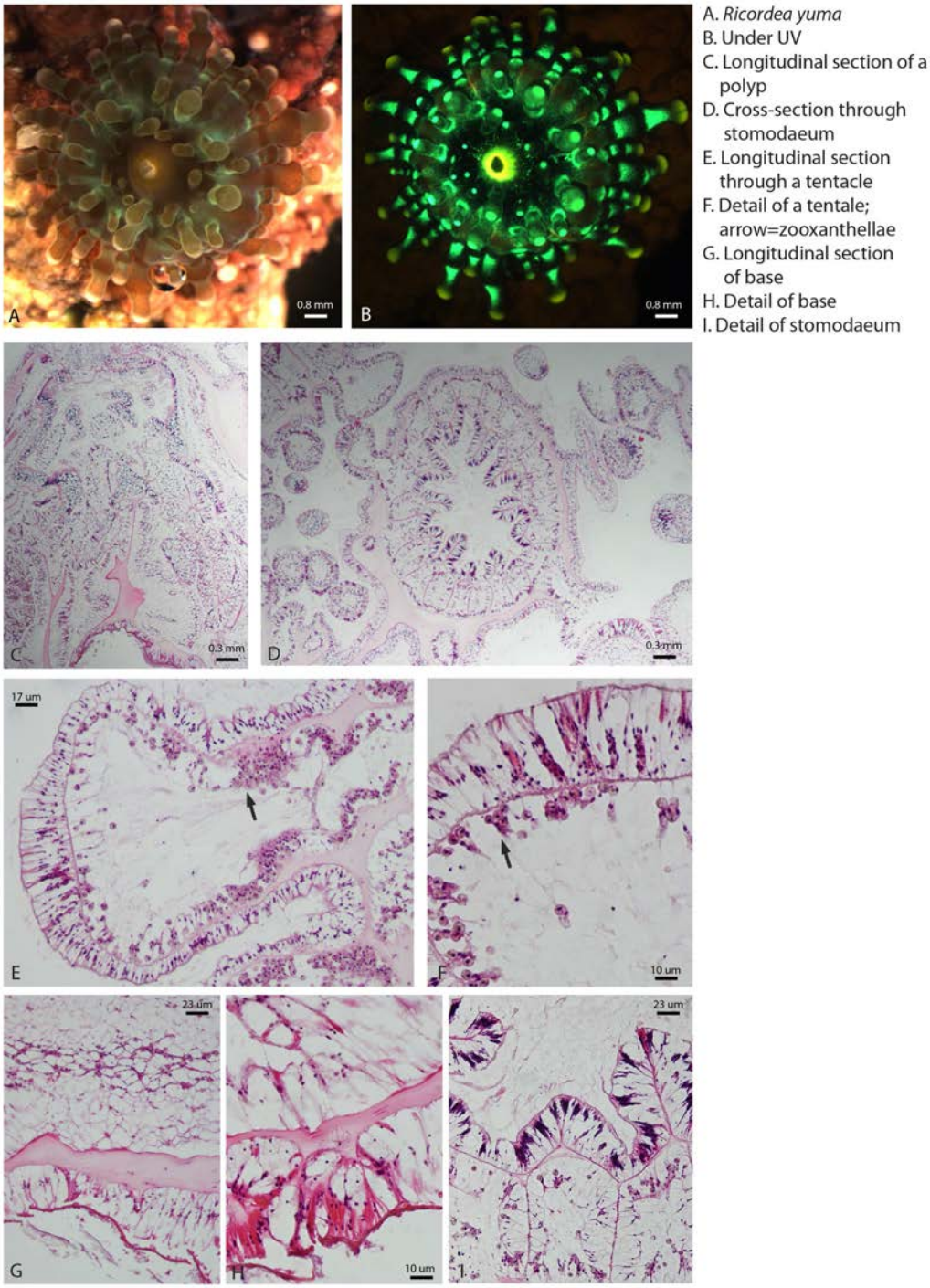
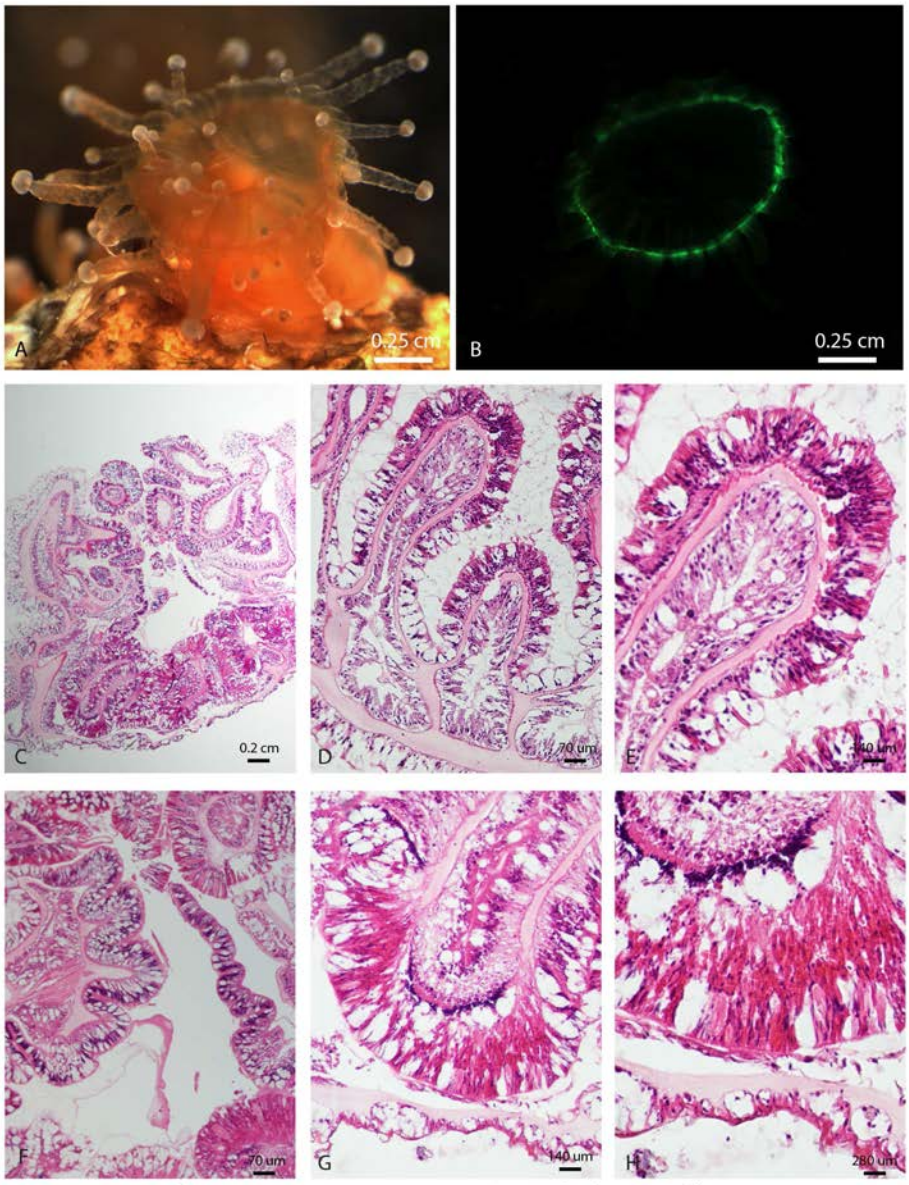



**Figure 1.3** *Ricordea yuma* specimens and dissection.




A. *Corynactis australis*
B. Under UV
C. Longitudinal section through a polyp
D. Longitudinal section through tentacles
E. Detail of tentacle
F. Longitudinal section through stomodaeum
G. Longitudinal section of base
H. Detail of base

**Figure 1.4** *Corynactis australis* specimens and dissection.

# Chapter II

# Corallimorpharians are not “naked corals”: insights into relationships between Scleractinia and Corallimorpharia from phylogenomic analyses

This chapter is published as

Mei-Fang Lin, Wen-Hwa Chou, Marcelo V. Kitahara, Chaolun Allen Chen, David J. Miller, Sylvain Forêt. Corallimorpharians are not “naked corals”: insights into relationships between Scleractinia and Corallimorpharia from phylogenomic analyses. PeerJ 4:e2463.

**Corallimorpharians are not "naked corals": insights into relationships between Scleractinia and Corallimorpharia from phylogenomic analyses**

Mei-Fang Lin[1,2], Wen-Hwa Chou[3], Marcelo V. Kitahara[4,5], Chaolun Allen Chen[3], David J. Miller[1,2,*], Sylvain Forêt[1,6*]

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

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

[3]Biodiversity Research Centre, Academia Sinica, Taipei, Taiwan

[4]Departamento de Ciências do Mar, Universidade Federal de São Paulo, Santos, São Paulo, Brazil

[5]Centro de Biologia Marinha (CEBIMar), Universidade de São Paulo, São Sebastião, Brazil

[6]Research School of Biology, Australian National University, Canberra, Australia

*Corresponding authors: david.miller@jcu.edu.au, sylvain.foret@anu.edu.au

**Abstract**

Calcification is one of the most distinctive traits of scleractinian corals. Their hard skeletons form the basis of reef ecosystems and confer on corals their remarkable diversity of shapes. The evolutionary relationship between scleractinians and corallimorpharians (non-calcifying, close relatives of corals) is key to understanding the evolution of calcification in the coral lineage. One crucial evolutionary issue is whether scleractinians are a monophyletic group, paraphyly being an alternative possibility if corallimorpharians are corals that have lost their ability to calcify, as is implied by the “naked-coral” hypothesis. Despite major efforts, relationships between scleractinians and corallimorpharians remain equivocal and controversial. Although complete mitochondrial genome sequences have been determined for a range of scleractinians and corallimorpharians, heterogeneity in composition and evolutionary rates means that mt data are inappropriate for understanding relationships between these two groups. To overcome the limitations of mitochondrial sequences, transcriptome data were generated for three representative corallimorpharians, and large numbers of orthologous nuclear sequences derived from these used to infer relationships within the coral/corallimorpharian group. The analysis presented here was based on 291 single copy nuclear protein-coding genes from 15 anthozoan taxa. Unlike mitochondrial sequences, these nuclear markers do not display any distinct compositional bias in their nucleotide or amino-acid sequences. A range of phylogenomic approaches congruently reveal a topology consistent with scleractinian monophyly and corallimorpharians as the sister clade of scleractinians.

Keywords: Anthozoa; Phylogenomics; Corallimorpharia; Scleractinia; Evolution

## Introduction

Scleractinian corals are the subject of intense scientific, public and, therefore, media interest, particularly because of the uncertain fate of coral reefs in the face of ever increasing anthropogenic challenges (Done 1999; Hughes 2003; Hughes et al. 2003; Hoegh-Guldberg et al. 2007). Due to their unique capacity to deposit continuous calcareous skeletons, the coral reef framework built by scleractinians provides one of the most complex and diverse of biological habitats (Cohen and Holcomb 2009). Despite their ecological importance and our economic dependence on them (Moberg and Folke 1999), we know remarkably little about the evolutionary history of this animal group. This lack of understanding limits our ability to predict how corals, and therefore the diverse habitats that they support, will respond to climate change and ocean acidification (OA).

Although the vast majority of scleractinian fossils post-date the sudden appearance of diverse coral families 14 My after the Permian/Triassic boundary, there is now evidence that the evolutionary origin of the group is rooted deep in the Paleozoic. In brief, molecular clock estimates calibrated using the earliest fossils that can be unambiguously assigned to extant clades, and whose unique skeletal characters can be unequivocally recognized in fossil coralla, implying that the scleractinian corals originated from a non-skeletonized ancestor in the Ordovician (Stolarski et al. 2011). When considered in conjunction with the elusive and equivocal Paleozoic fossil record of the scleractinian lineage (Scrutton and Clarkson 1991; Ezaki 1997, 2000), this suggests that either the fossil record for the period between the Ordovician and late Permian is yet to be discovered, or that skeletal formation may be an ephemeral

trait within the Scleractinia (Stanley and Fautin 2001).

The idea that the ability of corals to deposit a skeleton may be an ephemeral trait on evolutionary time scales, the presence or absence of a calcareous skeleton potentially reflecting prevailing environmental conditions, together with the anatomical similarity of Actiniaria, Corallimorpharia, and Scleractinia (Stanley and Fautin 2001; Daly et al. 2007), led Stanley (2003) to propose the naked coral hypothesis (NC). The central idea of the NC hypothesis is that "different groups of soft-bodied, unrelated anemone-like anthozoans gave rise to various calcified scleractinian-like corals through aragonitic biomineralization" (Stanley 2003), potentially explaining the sudden appearance of a diverse and differentiated range of scleractinian skeletal types in the Triassic. Under the NC hypothesis, the scleractinian skeleton is not a synapomorphy, but stands for an organization grade. Consistent with the NC hypothesis, the Scleractinia were paraphyletic in molecular phylogenetic analyses based on amino acid (aa) sequence data from mitochondrial protein-coding genes (Medina et al. 2006). In these analyses, it was estimated that corallimorpharians – anthozoans without a skeleton – diverged from the Robust scleractinian clade during the late- and mid-Cretaceous, implying that corallimorphs were descended from a coral that had undergone skeleton loss during a period of high atmospheric [$CO_2$]. Whilst this idea is superficially appealing, there is no evidence for either major OA events or "reef crises" during that time window (Kiessling and Simpson 2010). Moreover, phylogenetics based on a range of other molecular markers (Romano and Palumbi 1996; Chen et al. 2002; Fukami et al. 2008; Lin et al. 2014) did not support a sister group relationship of this sort, and it is

now clear that mitochondrial sequence data are not appropriate for the elucidation of phylogenetic relationships within the Corallimorpharia/Scleractinia clade (Kitahara et al. 2014).

The issue of coral/corallimorpharian relationships is an important one, because the idea that skeleton loss can occur as a consequence of elevated atmospheric [CO2] carries implications for the future of corals and coral reefs under climate change. A better understanding of coral evolution more broadly has further implications for coral reef futures - it is important to understand, for example, how prior OA events (Kiessling and Simpson 2010) have impacted the scleractinian lineage, and the underlying causes of previous “reef crises”.

In order to shed light on relationships between the Corallimorpharia and Scleractinia, phylogenomic analyses were carried out based on 291 nuclear protein-coding genes from a representative range of corallimorpharians (n = 3), robust (n = 6) and complex (n = 3) corals and sea anemones (n = 2), using an octocoral as outgroup. To enable the phylogenomic analysis, it was first necessary to generate comprehensive transcriptome assemblies for the three corallimorpharian species - *Rhodactis indosinensis, Ricordea yuma* and *Corynactis australis*. These molecular data constitute an important resource for this neglected animal group. Analyses were carried out on concatenated datasets at both the amino acid and nucleotide levels, and Bayesian Concordance Analyses based on statistical analysis of individual genes was applied to increase the depth of tree space. The results

provide strong support for scleractinian monophyly, allowing rejection of the hypothesis that corallimorpharians are “naked corals” - descendants of a scleractinian that had undergone skeleton loss.

## Materials and methods

*Transcriptome assembly and data matrix*

The general workflow from data collection to analysis is summarized in figure 2.1. A total of 15 anthozoan taxa were sampled, of which 14 were hexacorallians (ingroup), an octocoral was used as outgroup. Ingroup taxa comprised three "Complex" corals belonging to two families, six "Robust" corals from five families, three corallimorpharians representing three families, and two actiniarians. *Gorgonia ventalina* was set as outgroup (Table 2.1). The anthozoan expressed sequences tag (EST) and genomic data were sourced from the available databases summarized in TableS1.Transcriptome data from the zooxanthellate corallimorpharian *Ricordea yuma* and the azooxanthellate species *Corynactis australis* were obtained herein. *Ricordea yuma* samples were collected from the Great Barrier Reef (18°25'35.20"S, 146°41'10.91"E), and *Corynactis australis* colonies were collected from Jervis Bay, New South Wales (35° 4'14.11"S, 150°41'48.20"E) at a depth of 1 meter. The transcriptomes were generated from purified RNA extracted by using Trizol Reagent (Invitrogen, USA), which is based on the method described in Chomczynski and Sacchi (1987). High throughput sequencing was conducted on an Illumina HiSeq platform. The contigs were assembled with Trinity (Grabherr et al. 2011) following methods described in Lin et al. (2016).

Following the divergent GC content from *Symbiodinium* (Shinzato et al. 2014), all contigs from the transcriptomes included in the present analysis were analyzed to GC content distribution. Transcriptomic data were clustered employing

CD-HIT-EST (Li and Godzik 2006) with sequence identity threshold of 0.9. Contigs were translated in all reading frames using TransDecoder within the Trinity package (Grabherr et al. 2011).

HaMStR v13.2 (Ebersberger et al. 2009) was implemented for the core-orthologs search using the three publically available cnidarian genomes as primer taxa, i.e. *Acropora digitifera* (Shinzato et al. 2011), *Nematostella vectensis* (Putnam et al. 2007) and *Hydra magnipapillata* (Chapman et al. 2010). Using *A. digitifera* as the reference-taxon, this search resulted in 1,808 core orthologs were found. Using the same program, the ortholog search was extended to the transcriptomes of eight scleractinians, three corallimorpharians, and one actiniarian. Multiple transcripts were filtered using hmm search with an e-value cutoff of 1e-5, with settings for the highest Re-Blast scoring hit. Thus, only one ortholog in each search was retained. The HaMStR search identified 291 putative orthologous genes from the original data set (15 taxa). A BLASTp search against the NCBI non-redundant (nr) database (database downloaded on 11 August 2014) with an e-value cutoff of 1e-5 was conducted for the 291 nuclear protein-coding gene annotation used herein.

Inferred amino acid (aa) sequences from the 291 orthologous genes were aligned using MAFFT L-INSI v7.13 (Katoh and Standley 2013) and subsequently trimmed by using trimAl v1.2 implemented under the Heuristic method (Capella-Gutiérrez et al. 2009). The nucleotide (nt) sequence based alignment was generated by aligning the coding sequences from each contig to the aa sequence alignment. Orthologous gene sets were concatenated to a single matrix using custom

scripts. The nt saturation at each codon position for the concatenated alignment was estimated by DAMBE v5.3.110 (Xia et al. 2003; Xia and Lemey 2009). The aa sequence alignment comprised 122,170 amino acid positions and nt alignment had 366,510 bp.

*Phylogeny*

Supermatrix phylogeny

For the concatenated aa matrix, a single best-fit model of protein evolution was determined using ProtTest v3 (Darriba et al. 2011). The amino acid substitution model JTT+G+I (Maximum Likelihood value, lnL -1302574.56) was chosen as the best-fit model for the concatenated supermatrix. Maximum likelihood (ML) phylogenetic analyses were inferred via RAxML v7.2.6 (Stamatakis 2006). For the ML phylogeny, a rapid bootstrapping (-f a) within one step with 100 bootstrap replicates was applied. In total, ML aa analyses based analyses were calculated applying the JTT+G+I (Jones et al. 1992) substitution model initially and also conducted under the evolutionary models LG+G+I (Le and Gascuel 2008) and WAG+G+I (Whelan and Goldman 2001).

Bayesian inference (BI) for the supermatrix was inferred using PhyloBayes MPI v1.5a (Lartillot et al. 2013) running the empirical estimates under the JTT matrix (JTT+G+I). Empirical profile mixture models CAT-Possion model (Larillot and Philippe 2004), and the CAT+GTR model (Lartillot et al. 2013) were also applied for the BI analyses. Each run contained four Markov Chain Monte Carlo (MCMC) for a minimum of 10 million generations. The parameters to check the

convergence with burn-in period of 2,000 were determined by the discrepancy observed across the bipartition frequencies (the discrepancy index, maxdiff value) of all chains.

For the concatenated nt matrix, the coding sequences (CDS) were obtained using a custom script in Bioperl, and trimmed under the similar methods used for the aa alignment development. The GTR+G+I model (lnL -2966772.8659) was estimated as the best model for the nt sequences alignment by jModelTest 2 (Darriba et al. 2012). For the nt ML analyses, 100 replicates with a rapid bootstrapping were inferred using RAxML v7.2.6 (Stamatakis 2006). nt BI analysis was inferred through PhyloBayes MPI v1.5a (Lartillot et al. 2013) for 4 chains.

Partitioned phylogeny

The best-fit partitioning schemes and substitution models for aa (partitioned by genes) and nt (partitioned by genes and codons) alignments were estimated using PartitionFinder (Lanfear et al. 2012). Under relaxed clustering algorithms with top 0.1% and top 1% scheme searches (Lanfear et al. 2012), 252 and 153 gene partitions were found, respectively. Nucleotide alignment was estimated for partitions by genes and codons under related clustering algorithms with top 1% scheme searches, 75 subsets for gene partition and 106 for codon partition were found. For each partition the best-fit substitution model was calculated using PartitionFinder (Lanfear et al. 2012). Both partitioning sets were used for phylogenetic analyses. ML analysis was conducted using RAxML v7.2.6 (Stamatakis 2006) using 100 bootstrap replications. BI was inferred using MrBayes

v3.2.3 (Ronquist and Huelsenbeck 2003) consisting of 4 runs each of 2 million generations with topologies saved at each 1,000 generations. The first 2,000 saved topologies were discarded as burnin, and the remainder used for the calculation of posterior probabilities.

Comparisons of topologies were carried out by CONSEL (Shimodaira and Hasegawa 2001). The input-tree-wise likelihood scores from different phylogenetic tree reconstruction methods were input into CONSEL to perform the Kishino Hasegawa (KH) test (Kishino and Hasegawa 1989), Shimodaira Hasegawa (SH) test (Shimodaira and Hasegawa 1999), and the Approximately unbiased (AU) test (Shimodaira 2002).

Concordance factor estimation

The dominant tree which clades are true for a plurality of the genome has exceeded concordance factors (Ané et al. 2007). The concordance factors were estimated by using the individual nuclear protein coding gene topologies (291 topologies from aa and nt sequences based analyses, respectively) inferred by MrBayes v3.2.3 (Ronquist and Huelsenbeck 2003) with 2 runs by default for 2 million generations and first 2,000 topologies discarded as burnin.

Concordance factor estimation was conducted by BUCKy (Ané et al. 2007). The output files from both BI runs inference were summarized by removing first 1,000 trees (burnin). The concordance factors were then estimated with default settings (i.e. $\alpha=1$). To assess the effect of the Dirichlet process prior on the analyses of sampled nuclear genes (Weisrock 2012), additional priors of $\alpha=0.1$ was explored.

Both sample-wide and genome-wide concordance factors were estimated. The sample-wide concordance factors was referred to the estimation of the proportion of the sampled genes for which given clade is true, and for the genome-wide concordance factor is the estimates on the proportion of the genome/transcripton for which a clade is true (Ané et al. 2007). Two runs of pairwise gene dissimilarity were measured with the posterior probability from 1, and then nonmetric mutidimentional scaling (NMDS) was used to display the dissimilarity among all 291 genes by R program.

Mitochondrial genome phylogeny

The taxa used in Kitahara et al. (2014) and Lin et al. (2014) belonging to the same genus as those species included in the supermatrix were sourced in order to compare the later with the topology inferred using mitochondrial (mt) genomes. In total, nine taxa including one octocoral (*Pseudoterogorgia bipinnata*) used as outgroup were included in the mt genome phylogeny. The best-fit evolutionary model for aa mt matrix as JTT+G+F (lnL -36081.99) was determined in MEGA 6 (Tamura et al. 2013). The program MEGA6 was used for the ML analyses with a discrete Gamma (+G) distribution with 5 categories for 100 replicates.

## Results and Discussion

*Analyses of a concatenated amino acid supermatrix supports scleractinian monophyly*

The data matrix for the 15 taxa (Table 2.1) comprised 291 nuclear protein-coding genes, 263 of which have functional annotation, the other 28 coding for hypothetical proteins (supplementary table S2.2). The concatenated alignment consisted of 122,170 amino acid positions, and is thus considerably larger ( >33 fold) than the mt data alignment (3,666 aa) employed by Kitahara et al. (2014) to investigate phylogenetic relationships across the coral/corallimorpharian clade. Both ML and BI methods generated phylogenetic trees in which all nodes were strongly supported (Fig. 2.2A, Table 2.2); in ML analyses, all bootstrap values were >70%, and in BI posterior probability support at all nodes was 100%. In general, the use of different substitution models had no effect on ML tree topology (Table 2.2). However, different BI topologies were obtained based on the CAT and CAT-GTR models (Fig. S2.1). In the analyses of large datasets, both of these latter models are considered to be more robust against Long-Branch Attraction (LBA) in comparison to one-matrix or empirical models (Lartillot et al. 2009). Nevertheless, in general the CAT model fits less well than does the CAT-GTR model (Lartillot et al. 2009; Lartillot and Philip 2004) and, as CAT based runs did not reach convergence even after 10,000 cycles (Table S2.5), results under the this model were excluded from further consideration.

To enable comparisons with the mt phylogeny generated by Kitahara et al. (2014), phylogenetic analyses were conducted on mt genome data from similar taxa to those

used here in analyses based on concatenated nuclear sequences. The mitochondrial and nuclear protein coding sequences yielded very different trees in phylogenetic analyses; whereas aa data from nuclear genes strongly supported the SM hypothesis (Fig. 2.2A) notwithstanding the substitution model, analyses of mt aa data recovered the NC topology (Fig. 2.2B). The recovery of SM based on the aa matrix from nuclear genes corroborates a number of previous studies (e.g. Brugler and France 2007; Fukami et al. 2008; Kitahara et al. 2010, 2014; Stolarski et al. 2011; Park et al. 2012; Lin et al. 2014), the major implication being that corallimorpharians are not scleractinians that have undergone skeleton loss (Fig. 2.2).

Since the analyses of concatenated sequences for nuclear-encoded proteins yielded a consistent result, ML and BI phylogenies were then inferred based on the best substitution model for each partition (Table S2.3). The ML analyses based on 252 and 153 partitions of the 291 genes in total, improved support for the SM topology (Fig. 2.2, Table 2.2). Because the results inferred with these two partitioning schemes did not differ, and for computational reasons, 153 partitions were then applied in BI analysis. Both ML and BI partitioned based inference strongly supported SM topology. Thus, both unpartitioned and partitioned analyses of concatenated amino acid sequences, using different evolutionary models, consistently support the SM scenario. However, as distinct tree topologies result from analyses of mt protein-coding sequences at the amino acid and nucleotide levels (Kitahara et al. 2014), ML and BI analyses were also conducted based on the nt sequences of the nuclear protein-coding genes (see below).

*Analyses of the concatenated nt supermatrix*

The final alignment of the nucleotide sequences corresponding to the 291 nuclear protein-coding genes was 370,809 bp. was subjected to. Only sites that were fully resolved in saturation tests (DAMBE v. 5.3.110, Xia 2013) were considered in phylogenetic analyses, as gaps and unresolved sites reduce the sensitivity of the test (Xia 2013). As no significant saturation was detected, once the *Iss* index of the dataset was significantly smaller than both *Iss.c sym* and *Iss.c asym* (see Xia et al. 2003 and Table. S4), ML and BI reconstructions were inferred using GTR+G+I as the best-fit substitution model. The ML topology recovered was consistent with the nuclear aa results (Fig.2.2), all nodes again being strongly supported (Table 2.2). This result is also consistent with analyses based on the nucleotide sequences of mt protein coding genes with corallimorpharian as the sister clade of scleractinian (Fukami et al. 2008; Kitahara et al. 2010, 2014). Although BI analysis of the nt dataset has not yet converged, the current topology supports scleractinian monophyly (Fig. S2.4). In summary, analyses carried out on nuclear protein-coding genes at both the aa and nt alignments consistently recovered the SM topology with high statistical support (Table 2.4).

*Base composition*

In the case of mitochondrial protein-coding genes, significant differences in base composition were observed between corallimorpharians, and robust and complex corals, resulting in different patterns of codon use and amino acid composition

across the various lineages (Kitahara et al. 2014). In order to investigate the potential for compositional bias to affect the topology recovered for nuclear protein-coding genes, base composition was estimated for each of the 15 taxa included in the present analyses (Table 2.3). Base composition was generally similar across all the groups, although the octocoral had the highest (A+T) composition (57.96%) amongst those studied. Within the Hexacorallia, the scleractinian complex clade has a slightly higher (A+T) content (56.5%) and, consequently, a higher proportion of (A+T)-rich aa (FYMINK). The remaining groups (i.e. Actiniaria and Scleractinia [Robusta clade]) displayed an overall (A+T) content between 55.00 and 55.95% and no major differences between FYMINK and (G+C)-rich aa (GARP) (Fig. 2.3). The thymine and cytosine contents of nuclear protein coding genes of Robusta differed slightly (<1%) compared to other scleractinians; this phenomenon was observed across all three codon positions.Thus, in the case of the nuclear protein-coding genes, both the nt sequences and aa composition of the proteins that they encode were similar across the Actiniaria, Corallimorpharia, and Scleractinian (Fig. 2.3). In comparison to proteins encoded by the mitochondrial genome, nuclear-encoded proteins of anthozoans have, in general, more lysine (~7% vs ~2%), aspartic acid (~5.5% vs ~2%), and glutamic acid (~7% vs ~2.5%), but significantly less phenylalanine (~4% vs ~8%, but in the mt proteins of robust corals this value is ~13%).

Major differences in the composition of mt protein-coding genes at both the nucleotide and amino acid levels support the idea that the mt genomes of robust corals are evolving at a different rate to those of other hexacorallians (Fukami and

Knowlton 2005; Aranda et al. 2012; Kitahara et al. 2014). However, no such compositional bias appears to hold for nuclear protein-coding genes, implying that these latter are more appropriate sources of phylogenetic information than are mitochondrial data (Kitahara et al. 2014).

*Individual-gene consensus analyses refine the "true topology"*

Genes at different genomic locations may have distinct evolutionary histories, therefore divergent topologies may be recovered (reviewed in Ané et al. 2007). However, a number of approaches have been developed to extract the phylogenetic signals from each sampled gene (reviewed in Ané et al. 2007; Pisani et al. 2007; Akanni et al. 2014). To avoid inferring "wrong" phylogenetic relationships, it is necessary to apply a range of different phylogenetic reconstruction methods (von Haeseler 2012). Here, concordance factor (CF) estimation was applied to evaluate the contributions of individual genes to the resulting topology.

Bayesian approaches, which allow the simultaneous estimation of species and gene trees, were used to estimate concordance factors of individual gene phylogenies (Ané et al. 2007). According to Weisrock (2012), high CF values on branches indicate support from multiple genes. Information of the highest rank CF posterior means for each clade from the 291 BI-derived trees is summarized in the primary concordance tree (see Baum 2007). Applying different α values (1 and 0.1 in concordance analyses) and splitting runs using the same value of α resulted in no significant differences across runs. The results presented herein are those where an α=1 was used. The primary topology recovered in concordance analysis supported

scleractinian monophyly (Fig. 2.4). Both sample-wide and genome-wide CF values are similar at each clade and, although the CF at the clade of two anemones and *Gorgonia* was not available, the branches of the corallimorpharian and scleractinian groups are well-defined. This result is consistent with those inferred by unpartitioned and partitioned analyses from the concatenated sequence data. In the primary topology, most branches have high CF values, indicating support from the majority of the genes. For example, the branch leading to the Scleractinia/Corallimorpharia split has mean supporting values of 0.717 and 0.716 (full support = 1.0), respectively. This result therefore indicates broad support across the sampled nuclear genes for a coral/corallimorpharian divergence.

In order to measure the pairwise gene dissimilarities, nonmetric scaling with 2-dimentions, which explains the variation in gene-gene similarity (Ané et al. 2007), was applied. Since both aa and nt based analyses led to identical results, aa sequence based trees were used for the measurements conducted by two BI runs. Both runs resulted in similar outcomes (data not shown). Representation of gene-gene similarity revealed 3 groups of genes (Fig. S2.2). Most genes (51.72%) clustered together in the top-right corner, supporting the strict coral monophyly scenario. The genes in the top left favor a “mixture” of Complexa and Robusta representatives (i.e. the two scleractinian clades are not resolved), but Scleractinia have Corallimorpharia as sister clade. The scatter of genes towards the bottom left imply contradictory relationships. Tree examples from each main group are illustrated in figure S2.3. Examination of each individual topology inferred under BI, indicates that none support the NC scenario. In summary, most genes sampled

supported the primary concordance tree topology (the “true” tree) featuring scleractinian monophyly.

*The evolutionary relationship between corallimorpharians and corals*

The phylogenomic analyses presented here provide strong support for scleractinian monophyly, and allow rejection of the idea that the corallimorpharian lineage was derived from corals by skeleton loss. The analyses supporting this latter idea were based on amino acid sequence data from mitochondrial genomes (Medina et al., 2006), but it is now clear there are fundamental problems in using mitochondrial data to infer phylogenetic relationships amongst hexacorallians (Kitahara et al. 2014), as is also the case in beetles (Sheffield et al. 2009) and some groups of mammals (Huttley 2009).

Whereas the issue of scleractinian monophyly has been equivocal, that the Corallimorpharia are a monophyletic group has not been challenged (e.g.: Fukami et al. 2008; Kitahara et al. 2010, 2014; Kayal et al. 2013). If we accept coral monophyly, the Scleractinia and Corallimorpharia shared a soft-bodied and solitary ancestor deep in the Paleozoic. The earliest diverging of extant corals are deep water and solitary and the deep-sea corallimorpharian, *Corallimorphus profundus*, is considered the most coral-like and “primitive” of corallimorpharians (Mosely 1877; den Hartog 1980; Riemann-Zurneck and Iken 2003). Supporting its early divergence within Corallimorpharia, the architecture of the mt genome of *Corallimorphus* is strikingly coral-like (Lin et al. 2014), whereas the canonical

organization in zooxanthellate corallimorpharians differs markedly from this. Although a few colonial corallimorpharians are known (den Hartog 1980), most species are solitary with a relatively flat body plan – features that are shared with some early diverging scleractinians, and which may therefore reflect retained ancestral characteristics. Micrabaciid corals have a reduced skeleton that is invested by the fleshy polyp, resembling a corallimorpharian with a rudimentary skeleton. The striking similarity between the skeletal architecture of the extant micrabaciid *Letepsammia* and the lower Paleozoic fossil *Kilbuchophyllia* (Stolarski et al. 2011) suggests that the ancestral morphology is reflected in these genera.

After divergence, the Scleractinia and Corallimorpharia have subsequently followed very different evolutionary trajectories. In case of the Scleractinia, an explosive diversification occurred during the Triassic (Constantz 1986, MacRae et al. 1996, Cuif 2014), presumably driven by the acquisition of photosynthetic symbionts and the enhancement of calcification that these enabled – skeleton diversification may allow more efficient light absorption (Enríquez et al. 2005). Many tropical corallimorpharians also host photosynthetic symbionts, but the relationship is facultative rather than obligate, presumably established independently in this lineage, and the ability to host these did not result in a corresponding diversification event. Consequently far fewer species are recognized – around 46 (Lin et al. 2014) compared to approximately 1,400 coral species (Cairns 1999). One factor potentially biasing these numbers is that the numbers of taxonomically useful characters differ between the two groups – coral taxonomy has traditionally been

based largely on skeletal characters, as there are few other morphological features that are taxonomically informative. In the absence of a skeleton, corallimorpharian taxonomy has few characters to make use of, a corollary being that cryptic species may be the norm.

*Insights into coral evolution*

Taking into account the fossil and published molecular data (Kiessling and Simpson, 2010; Stolarski et al. 2011), the analyses above imply that the ability to secrete a skeleton was acquired early in scleractinian evolution, but was followed by multiple origins of skeleton complexity in various subclades (Romano and Cairns 2000). The Paleozoic fossil record (Scrutton and Clarkson 1991; Scrutton 1993; Ezaki 1997, 1998) and molecular data (Stolarski et al. 2011) both imply that the earliest scleractinians were solitary and inhabited deep water and therefore lacked photosynthetic symbionts. The sudden appearance of highly diversified forms of Scleractinia about 14 Ma after the end-Permian extinction (the “Great Dying” - Wells 1956; Roniewicz and Morycowa 1993; Veron 1995; Stanley 2003) might be explained by multiple independent origins from deep-water ancestors (e.g. the family Agariciidae - Kitahara et al. 2012). It thus appears likely that the acquisition of photosynthetic symbionts and the development of coloniality have both occurred independently on multiple occasions, resulting not only in a wide range of skeletal phenotypes but also in habitat expansion, which has played important roles in the formation of shallow-water reefs (Barbeitos et al. 2010).

It has been demonstrated that, when maintained under acidic conditions (pH7.3-7.6), at least some corals can survive for 12 months after undergoing skeleton loss, recovering fully after return to normal seawater (Fine and Tchernov 2007). One interpretation of these experiments is that, during evolution, the coral lineage might have been able to alternate between soft and skeletonized forms, potentially explaining the gaps in the fossil record. However, the fact that corallimorpharians are not derived from corals, and the monophyly of extant Scleractinia, implies otherwise - that skeleton-less corals are not viable on evolutionary time scales. This has important implications for the future of the coral lineage – the evolutionary resilience of the Scleractinia may have depended in the past on deep sea refugia, as most of the "reef crises" have coincided with rapid increases in both OA and sea surface temperature (Pandolfi et al. 2011). Deep-sea corals would have escaped the challenges of high SST, thus the coral lineage may have been able to re-establish itself in the shallows when more favourable conditions returned. At the present time, unprecedented rates of increase in OA and SST are occurring concurrently with massive disruption of deep-sea habitats (Guinotte et al. 2006; Ramirez-Llodra et al. 2011). Is the resilience of the Scleractinia as a lineage therefore at risk?

**Acknowledgement**

The authors thank Mr. Chris Benstead (Reef HQ Aquarium, Townsville, Australia), Dr. Joshus Madin (Macquarie University, Sydney, Australia), Mr. Chao-Yang Kuo (James Cook University, Townsville, Australia), and members of the Coral Reef Evolutionary Ecology and Genetics Laboratory (CREEG) for the assistance with sampling the corallimorpharian specimens and field logistics. We also thank Dr. Chuya Shinzato (Okinawa Institute of Science and Technology, Okinawa, Japan) to provide the *Porites* transcriptome. We gratefully thank Prof. Ingo Ebersberger (Goethe University, Frankfurt, Germany) for the helps on the methods and scripts for the data analyses. M.-F., L. appreciatively acknowledges receipt of a James Cook University Postgraduate Research Scholarship (JCUPRS).

**Table 2.1** Taxa studied for the phylogenetic analysis. Azooxanthellate species were indicated by (*).

| | Family | Species | Data type | # ortholog[a] |
|---|---|---|---|---|
| **Complex** | Acroporidae | *Acropora millepora* | Genome | 1609 |
| | Acroporidae | *Acropora digitifera* | Genome | Primer/Reference |
| | Poritidae | *Porites australiensis* | Transcriptome | 1296 |
| **Robust** | Fungiidae | *Fungia scutaria* | Transcriptome | 1511 |
| | Pocilloporidae | *Madracis auretenra* | Transcriptome | 1442 |
| | Montastraeidae | *Montastraea cavernosa* | Transcriptome | 1505 |
| | Pocilloporidae | *Pocillopora damicornis* | Transcriptome | 1203 |
| | Merulinidae | *Platygyra carnosus* | Transcriptome | 1561 |
| | Favinae | *Pseudodiploria strigosa* | Transcriptome | 1358 |
| **Corallimorpharia** | Corallimorphidae | *Corynactis australis** | Transcriptome | 1481 |
| | Discosomidae | *Rhodactis indosinensis* | Transcriptome | 1261 |
| | Ricordeidae | *Ricordea yuma* | Transcriptome | 1401 |
| **Actiniaria** | Edwardsiidae | *Nematostella vectensis** | Genome | Primer |
| | Actiniidae | *Anthopleura elegantissima* | Transcriptome | 1448 |
| **Octocorallia** | Anthothelidae | *Gorgonia ventalina* | Transcriptome | 1421 |
| **Anthomedusae** | Hydridae | *Hydra magnipapillata* | Genome | Primer |

a. 1e-5 E-value cut-off for the hmm search; retain only the top-scoring hit for RE-BLAST; Identify only one ortholog per core-ortholog group. Three cnidarian genomes were used as “primer” for the core orthologous search, and then the *A.digitifera* genome was used as “reference” for the database development.

**Table 2.2** Selected nodes and support values of ML and BI reconstruction inferred from unpartitioned and partitioned phylogenetic analyses from amino acid and nucleotide concatenated data. The best-fit substitution model for each unpartitioned dataset was indicated herein; for the partitioned phylogeny, the best-fit model for each portioning region was indicated in the supplementary table S3. The supporting values for the trees based on 252 partitions were shown in the supplementary table S2.6.

| Data type | Amino acid dataset | | | | Nucleotide dataset | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | Maximum Likelihood analyses | | Bayesian Inference | | Maximum Likelihood analysis | | | Bayesian Inference | | |
| Supporting value | Bootstrap support(%) | | Posterior Probability | | Bootstrap support(%) | | | Posterior Probability | | |
| Matrix type | Concatenated matrix | Partitions | Concatenated matrix | Partitions | Concatenated matrix | Partitions by Gene | Partition by Codon | Concatenated matrix | Partitions by Gene | Partition by Codon |
| Selected Nodes/ substition models | JTT+G+I | 153 subsets | JTT+G+I | 153 | GTR+G+I | 75 subsets | 106 subsets | GTR+G+I | 75 subsets | 106 subsets |
| ((Actiniaria,(Corallimorpharia, Scleractinia)) | - | - | - | 1 | - | - | - | - | 1 | 1 |
| (Corallimorpharia, Scleractinia) | 100 | 100 | 1 | 1 | 100 | 100 | 100 | - | 1 | 1 |
| (Complexa, Robusta) | 100 | 100 | 1 | 1 | 100 | 100 | 100 | - | 1 | 1 |
| (*Corynactis*, (*Ricordea*, *Rhodactis*)) | 100 | 100 | 1 | 1 | 100 | 100 | 100 | - | 1 | 1 |
| (*Ricordea*,*Rhodactis*) | 77 | 90 | 1 | 1 | 79 | 94 | 99 | - | 1 | 1 |
| (((*Pocillopora*, *Madracis*),(((*Pseudodiploria*, *Platygyra*), *Montastraea*),*Fungia*))) | 100 | 100 | 1 | 1 | 100 | 100 | 100 | - | 1 | 1 |
| (*Pocillopora*, *Madracis*) | 100 | 100 | 1 | 1 | 100 | 100 | 100 | - | 1 | 1 |
| (((*Pseudodiploria*, *Platygyra*), *Montastraea*),*Fungia*))) | 100 | 100 | 1 | 1 | 100 | 100 | 100 | - | 1 | 1 |
| ((*Pseudodiploria*, *Platygyra*), *Montastraea*) | 100 | 100 | 1 | 1 | 100 | 100 | 100 | - | 1 | 1 |
| (*Pseudodiploria*, *Platygyra*) | 100 | 100 | 1 | 1 | 100 | 100 | 100 | - | 1 | 1 |
| (*Porites*,(*A.digitifera*, *A.millepora*)) | - | 100 | 1 | 1 | 100 | 100 | 100 | - | 1 | 1 |
| (*A.digitifera*, *A.millepora*) | 100 | 100 | 1 | 1 | 100 | 100 | 100 | - | 1 | 1 |
| (*Nematostella*, *Anthopleura*) | 100 | 100 | - | 1 | 100 | 100 | 100 | - | 1 | 1 |

Note: Until the submission of this paper, the nucleotide concatenated matrix under the Bayesian inference has not been converged, although it has been running for more than 5,000 hours on the clusters. The current tree as shown in Fig. S4 does not support the naked-coral hypothesis.

**Table 2.3** Compositional bias of the 291 nuclear genes of the anthozoan Orders included in the present study. In order to compare with the mt genes, compositional bias based on Kitahara et al. (2014) are given in the parentheses.

| | | | **Nucleotide** | | **Protein** | | |
|---|---|---|---|---|---|---|---|
| **Sub-class** | **Order** | **Group** | G+C(%) | A+T(%) | FYMINK(%) | GARP(%) | FYMINK/ GARP |
| Octocorallia | Alcyonacea | | 42.20 | 57.80 | 27.16 | 22.13 | 1.227 |
| | | | (35.33) | (64.66) | (30.54) | (21.96) | (1.39) |
| Hexacorallia | Actiniaria | | 44.43 | 55.57 | 26.15 | 22.86 | 1.144 |
| | | | (37.95) | (62.05) | (29.92) | (22.47) | (1.33) |
| | Corallimorpharia | | 45.06 | 54.94 | 25.83 | 22.69 | 1.138 |
| | | | (37.95) | (62.05) | (28.73) | (22.15) | (1.29) |
| | Scleractinia | Complexa | 43.60 | 56.40 | 26.27 | 22.46 | 1.170 |
| | | | (37.59) | (62.41) | (29.42) | (21.81) | (1.34) |
| | | Robusta | 44.12 | 55.88 | 26.33 | 22.82 | 1.154 |
| | | | (31.2) | (68.8) | (33.71) | (19.36) | (1.74) |

**Table 2.4** Comparisons of the topologies using Kishino Hasegawa (KH), Shimodaira Hasegawa (SH), and the Approximately Unbiased (AU) tests for various tree reconstruction methods.

| | | Best ML topology | AU | KH | SH |
|---|---|---|---|---|---|
| Amino Acids | Total evidence (JTT+G+I) | SM | 0.781 | 0 | 0.992 |
| | Partition (153 subsets) | SM | 0.648 | 0 | 0.937 |
| Nucleotides | Total evidence (GTR+G+I) | SM | 0.912 | 0 | 0.981 |
| | Partition by gene (75 subsets) | SM | 0.682 | 0.537 | 0.810 |
| | Partition by codon (106 subsets) | SM | 0.701 | 0 | 0.996 |

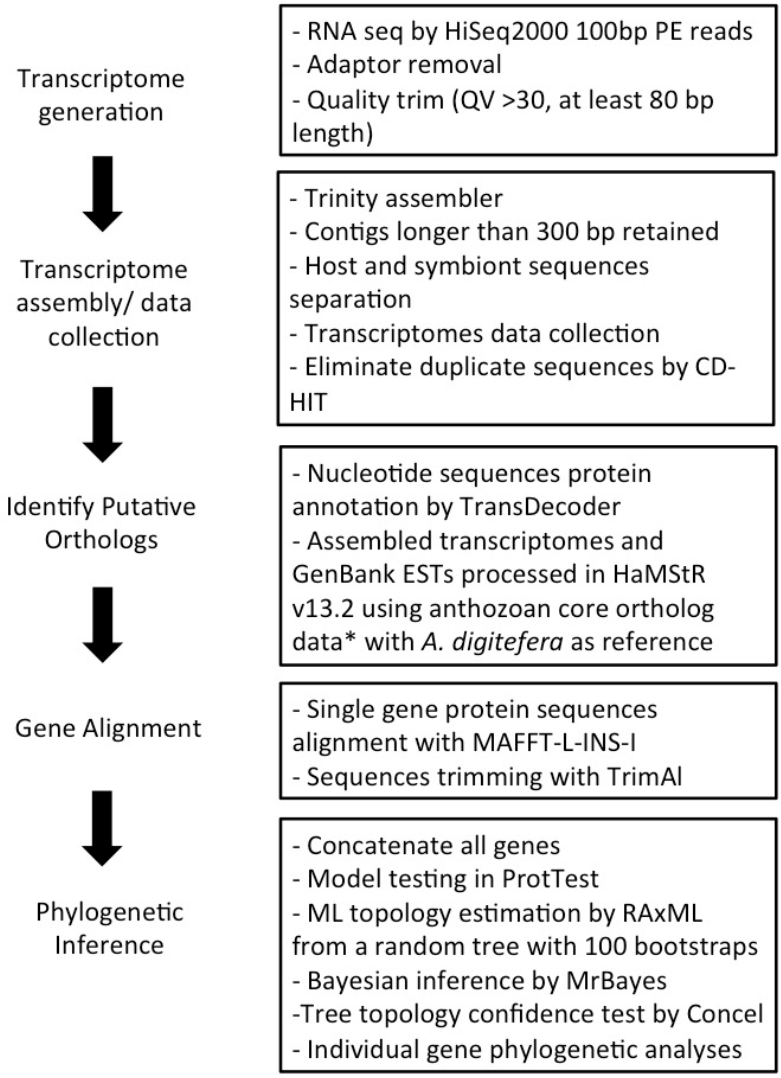



**Figure 2.1** Diagram showing general workflow from the data collection to analysis. (*) the cnidaran core orthologous database was developed by using three cnidarian genomes as indicated in Table 2.1.

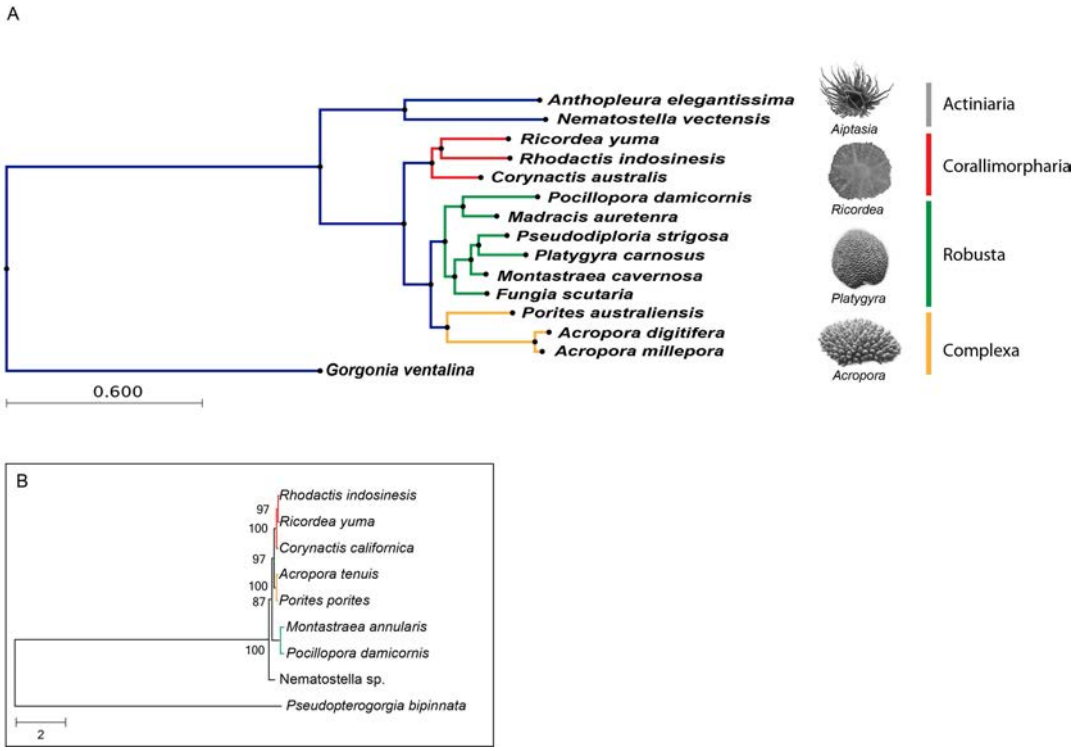



**Figure 2.2** Phylograms of the 291 nuclear genes from 15 anthzoans (A). Phylogeny of the 13 mitochondrial genes from the similar species based on Maximum likelihood analysis by using amino acid sequences (B).

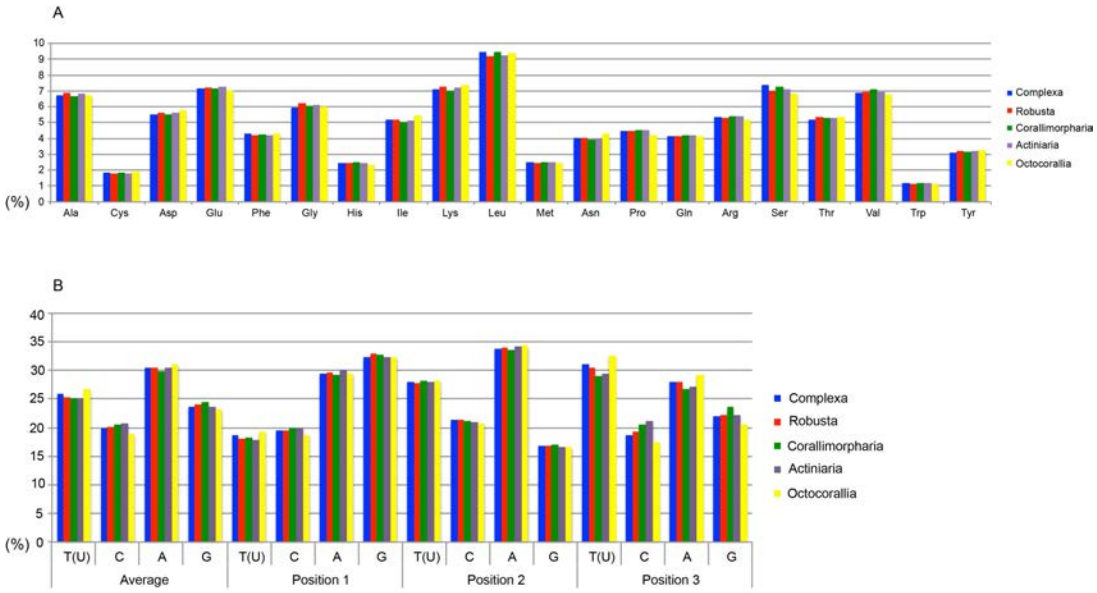



**Figure 2.3** Amino acid (upper) and nucleotide (lower) content of the 291 nuclear genes of studied species. The data shown are average across the Complexa (n=3), Robusta (n=6), Corallimorpharia (n=3), Actiniaria (n=2) and Octocorallia (n=1).

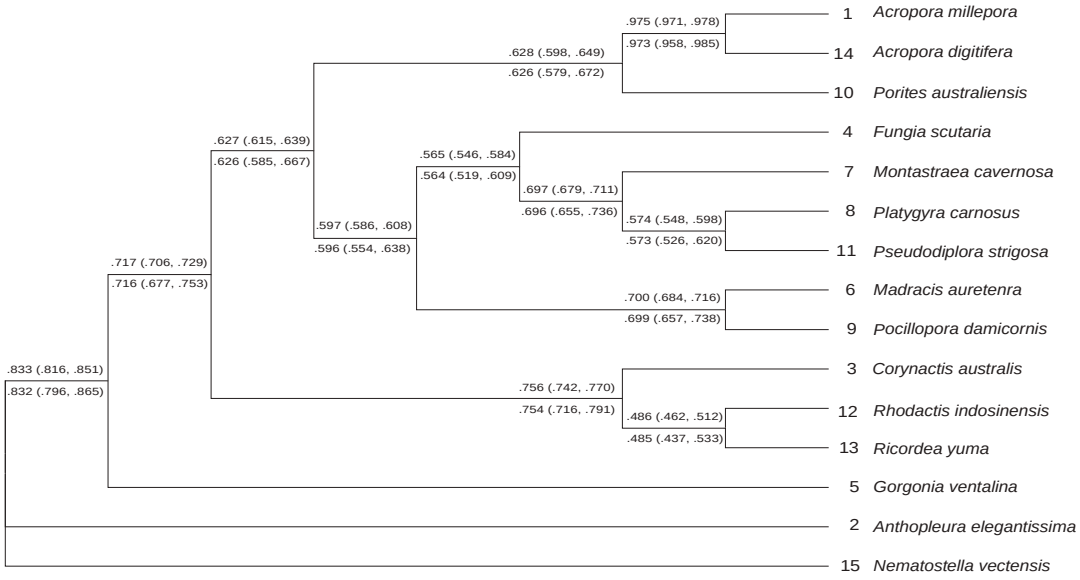



**Figure 2.4** Primary concordance tree of the 291 nuclear genes from 15 anthozoans. Numbers are posterior mean concordance factors and their 95% credibility intervals with α=1. Above edges, numbers indicate sample-wide concordance factors, and below edges, numbers indicate genome-wide concordance factors.

**Supplementary**

**Table S2.1** Data collection information

| | **Family** | **Species** | **Soruce/collection site** |
|---|---|---|---|
| **Complex** | Acroporidae | *Acropora millepora* | DM lab |
| | Acroporidae | *Acropora digitifera* | Shinzato et al. 2011, OIST[1] |
| | Poritidae | *Porites australiensis* | Shinzato et al. 2014, OIST[1] |
| **Robust** | Fungiidae | *Fungia scutaria* | Meyer and Weis lab[2] |
| | Pocilloporidae | *Madracis auretenra* | Meyer and Weis lab[2] |
| | Montastraeidae | *Montastraea cavernosa* | Meyer and Weis lab[2] |
| | Pocilloporidae | *Pocillopora damicornis* | Traylor-Knowles et al. 2011 |
| | Merulinidae | *Platygyra carnosus* | Sun et al. 2013 |
| | Favinae | *Pseudodiploria strigosa* | Meyer and Weis lab[2] |
| **Corallimorph** | Corallimorphidae | *Corynactis australis** | This study, NSW |
| | Discosomidae | *Rhodactis indosinensis* | This stusy, TW |
| | Ricordeidae | *Ricordea yuma* | This study, GBR |
| **Actiniaria** | Edwardsiidae | *Nematostella vectensis** | NCBI PRJNA19965 |
| | Actiniidae | *Anthopleura elegantissima* | Meyer and Weis lab[2] |
| **Octocorallia** | Anthothelidae | *Gorgonia ventalina* | Burge et al. 2013 |
| **Anthomedusae** | Hydridae | *Hydra magnipapillata* | Chapman et al. 2010 |

1. Okinawa Institute of Science and Technology Graduate University (OIST)

Chuya Shinzato, Eiichi Shoguchi, Takeshi Kawashima, Mayuko Hamada, Kanako Hisata, Makiko Tanaka, Manabu Fujie, Mayuki Fujiwara, Ryo Koyanagi, Tetsuro Ikuta, Asao Fujiyama, David J. Miller, Nori Satoh. 2011. Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* 476: 320–323.

Chuya Shinzato, Mayuri Inoue, Makoto Kusakabe. A snapshot of a coral "holobiont": a tanscriptome assembly of the scleractinian coral, *Porites*, captures a wide variety of genes from both the host and symbiotic zooxanthellae. *PLoS ONE* 9: e85182.

2. Meyer and Weis lab: http://people.oregonstate.edu/~meyere/data.html, last accessed 2/5/2014

3. Nikki Traylor-Knowles, Brian R. Granger, Tristan J. Lubinski, Jignesh R. Parikh, Sara Garamszegi, Yu Xia, Jarrod A. Marto, Les Kaufman, John R Finnerty. 2011. Production of a reference transcriptome and transcriptomic database (PocilloporaBase) for the cauliflower coral, *Pocillopora damicornis*. *BMC Genomics* 12: 585.

4. Jin Sun, Qian Chen, Janice C. Y. Lun, Jianliang Xu, Jian-Wen Qiu. 2013. PcarnBase: development of a transcriptomic database for the brain coral *Platygyra carnosus*. *Marine Biotechnology* 15: 244-251.

5. Colleen A. Burge, Morgan E. Mouchka, C. Drew Harvell, Steven Roberts. 2013. Immune response of the Caribbean sea fan, *Gorgonia ventalina*, exposed to an *Aplanochytrium* parasite as revealed by transcriptome sequencing. *Frontiers in Physiology* 4: 180.

**Table S2.2** Encoding proteins of the studied genes.

| GeneID | % identity | evalue | bit score | Description |
|---|---|---|---|---|
| gene11 | 59.18 | 0 | 868 | importin-5 |
| gene18 | 81.61 | 0 | 1035 | cullin-1-like |
| gene22 | 66.67 | 7.2 | 35 | ubiquitin carboxyl-terminal hydrolase 44 |
| gene27 | 62.78 | 2.00E-98 | 318 | structural maintenance of chromosomes protein 1A-like, partial |
| gene35 | 65.23 | 3.00E-106 | 336 | exportin-2-like |
| gene41 | 57 | 0 | 719 | calcium-dependent secretion activator 2 |
| gene59 | 49.46 | 0 | 841 | ubiquitin conjugation factor E4 B |
| gene61 | 72 | 0 | 743 | pre-mRNA-processing factor 19 |
| gene65 | 53.93 | 2.00E-19 | 95.5 | nuclear pore complex protein Nup205-like, partial |
| gene66 | 71.76 | 0 | 814 | 60 kDa heat shock protein, mitochondrial |
| gene68 | 92.71 | 3.00E-56 | 188 | translation elongation factor 1-alpha, partial |
| gene70 | 92.86 | 3.00E-16 | 87 | cullin-4B-like |
| gene89 | 52.62 | 0 | 696 | alpha-N-acetylglucosaminidase |
| gene99 | 52 | 6.00E-143 | 437 | leishmanolysin-like peptidase-like |
| gene103 | 73.29 | 0 | 655 | growth arrest-specific protein 8 isoform X2 |
| gene123 | 49.02 | 7.00E-06 | 52.4 | protein diaphanous homolog 1 isoform X2 |
| gene124 | 64.11 | 1.00E-125 | 394 | LOW QUALITY PROTEIN: ATP-binding cassette sub-family D member 1 |
| gene149 | 72.05 | 0 | 556 | mitogen-activated protein kinase 15-like |
| gene151 | 70 | 0 | 658 | 6-phosphogluconate dehydrogenase, decarboxylating |
| gene154 | 72.15 | 4.00E-30 | 127 | uncharacterized protein LOC102208666 |
| gene156 | 71.22 | 0 | 617 | dihydropyrimidinase |
| gene157 | 62.47 | 2.00E-157 | 487 | serine/threonine-protein kinase/endoribonuclease IRE1 |
| gene163 | 68.25 | 3.00E-21 | 95.9 | coiled-coil domain-containing protein 93-like |
| gene164 | 80.69 | 4.00E-134 | 414 | tRNA wybutosine-synthesizing protein 1 homolog |
| gene166 | 61.78 | 6.00E-72 | 258 | protein transport protein Sec24D |
| gene167 | 60.71 | 0.005 | 42.7 | protein PTHB1-like |
| gene170 | 78.59 | 0 | 556 | actin-related protein 2 |
| gene181 | 93.3 | 3.00E-159 | 453 | casein kinase I isoform alpha |
| gene182 | 63.66 | 0 | 547 | aspartate--tRNA ligase, cytoplasmic |
| gene195 | 88.24 | 6.00E-33 | 127 | BTB/POZ domain-containing protein KCTD3-like |
| gene200 | 77.78 | 3.00E-38 | 152 | F-box/WD repeat-containing protein 7-like |
| gene204 | 48.7 | 0 | 577 | collagen type IV alpha-3-binding protein |
| gene211 | 40.76 | 2.00E-36 | 150 | chromosome-associated kinesin KIF4A |
| gene216 | 69.9 | 0 | 579 | betaine--homocysteine S-methyltransferase 1-like |
| gene218 | 48.62 | 0 | 561 | alpha-(1,6)-fucosyltransferase-like |
| gene219 | 62.44 | 5.00E-83 | 276 | ubiquitin carboxyl-terminal hydrolase 15 |
| gene237 | 88.89 | 0.57 | 36.6 | dephospho-CoA kinase |
| gene247 | 78.43 | 8.00E-15 | 82.4 | macoilin |
| gene251 | 64.74 | 2.00E-148 | 478 | zinc finger SWIM domain-containing protein 5 isoform X2 |
| gene256 | 76.67 | 3.00E-07 | 56.2 | GDP-Man:Man(3)GlcNAc(2)-PP-Dol alpha-1,2-mannosyltransferase isoform X2 |
| gene257 | 69.43 | 4.00E-179 | 519 | acetoacetyl-CoA synthetase |
| gene274 | 75.41 | 3.00E-23 | 102 | uncharacterized protein LOC102159053, partial |
| gene275 | 85.33 | 2.00E-142 | 420 | casein kinase I isoform gamma-2-like |
| gene281 | 62.4 | 9.00E-160 | 469 | GTP-binding protein 2 |
| gene284 | 46.58 | 6.00E-158 | 471 | negative elongation factor B-like |
| gene286 | 61.54 | 6.00E-14 | 72.8 | importin subunit alpha-1-like, partial |
| gene291 | 67.8 | 1.00E-15 | 84 | olfactory receptor 226-like |
| gene305 | 62.26 | 5.00E-180 | 520 | putative aminopeptidase W07G4 4-like |
| gene307 | 69.8 | 3.00E-131 | 386 | hypothetical protein AMTR_s00183p00019870 |
| gene317 | 57.33 | 4.00E-141 | 418 | eukaryotic translation initiation factor 5 |
| gene320 | 76.77 | 0 | 588 | hypothetical protein D910_03931 |
| gene322 | 35.28 | 3.00E-70 | 237 | lactation elevated protein 1 |
| gene329 | 45.79 | 1.00E-107 | 337 | vang-like protein 2 |
| gene338 | 65.56 | 0 | 518 | 26S proteasome non-ATPase regulatory subunit 11 |
| gene352 | 70.08 | 2.00E-137 | 396 | 2-amino-3-carboxymuconate-6-semialdehyde decarboxylase-like |
| gene353 | 50 | 3.00E-117 | 367 | coiled-coil domain-containing protein lobo homolog |
| gene354 | 63.96 | 2.00E-95 | 294 | general transcription factor IIH subunit 4-like, partial |
| gene355 | 74.23 | 3.00E-131 | 384 | putative tRNA (cytidine(32)/guanosine(34)-2'-O)-methyltransferase-like |
| gene360 | 46.8 | 2.00E-113 | 348 | major facilitator superfamily domain-containing protein 5 |
| gene362 | 69.29 | 2.00E-138 | 409 | Phosphoenolpyruvate phosphomutase |
| gene366 | 63.47 | 6.00E-89 | 289 | frizzled-8 |
| gene372 | 67.23 | 4.00E-51 | 174 | tyrosine hydroxylase 2, partial |
| gene373 | 66.86 | 6.00E-80 | 251 | cysteine protease ATG4B-like, partial |
| gene378 | 73.1 | 1.00E-83 | 259 | BTB/POZ domain-containing protein 2-like |
| gene382 | 60.18 | 6.00E-81 | 290 | mitogen-activated protein kinase kinase kinase 1 |
| gene385 | 57.11 | 3.00E-159 | 464 | GDP-fucose protein O-fucosyltransferase 2-like |
| gene388 | 45.75 | 1.00E-35 | 139 | U3 small nucleolar RNA-associated protein 7, variant 2 |
| gene389 | 49.4 | 8.00E-130 | 390 | retinoid-inducible serine carboxypeptidase |
| gene394 | 56 | 0 | 608 | putative beta-carotenoid oxygenase |
| gene395 | 70.22 | 3.00E-137 | 404 | KDEL motif-containing protein 1 |
| gene402 | 74.9 | 7.00E-147 | 419 | 40S ribosomal protein S4-like |
| gene404 | 57 | 0 | 576 | hypothetical protein |
| gene407 | 65.92 | 5.00E-119 | 364 | nuclear respiratory factor 1 |
| gene416 | 58.69 | 4.00E-156 | 455 | glycoprotein endo-alpha-1,2-mannosidase |
| gene428 | 83.78 | 2.00E-12 | 70.1 | replication factor C subunit 2 |
| gene435 | 51.96 | 4.00E-97 | 300 | alpha-(1,3)-fucosyltransferase 10-like |
| gene444 | 55.59 | 2.00E-139 | 415 | BTB/POZ domain-containing protein 9-like, partial |
| gene445 | 57.36 | 3.00E-44 | 163 | hypothetical protein D910_00639, partial |
| gene450 | 69.23 | 6.00E-11 | 68.2 | peptidyl-prolyl cis-trans isomerase-like 2-like, partial |
| gene452 | 53.85 | 1.00E-160 | 473 | cytosol aminopeptidase-like |
| gene455 | 65.14 | 6.00E-169 | 491 | integrator complex subunit 11 |
| gene460 | 83.33 | 2.00E-144 | 413 | proteasome subunit alpha type-2 |
| gene461 | 92.17 | 4.00E-71 | 224 | proteasome subunit beta type-5 |
| gene467 | 65.41 | 5.00E-91 | 276 | protein arginine N-methyltransferase 1-like, partial |
| gene469 | 85.96 | 7.00E-25 | 101 | erythrocyte band 7 integral membrane protein |
| gene472 | 75.22 | 1.00E-49 | 179 | dihydrolipoyl dehydrogenase, mitochondrial |
| gene476 | 56.39 | 5.00E-140 | 411 | enoyl-CoA delta isomerase 2, mitochondrial |
| gene478 | 57.88 | 5.00E-157 | 467 | WDR70 WD repeat domain 70 |
| gene479 | 43.55 | 8.00E-60 | 215 | autophagy-related protein 9A |
| gene495 | 56.07 | 4.00E-30 | 134 | protein dopey-2-like, partial |
| gene496 | 64.88 | 1.00E-65 | 233 | YEATS domain-containing protein 2-like |
| gene498 | 82.17 | 5.00E-147 | 462 | hypothetical protein D910_10583 |
| gene512 | 71.57 | 4.00E-95 | 295 | LOW QUALITY PROTEIN: family with sequence similarity 46, member A |
| gene521 | 45.18 | 2.00E-140 | 424 | nuclear RNA export factor 1 |
| gene523 | 71 | 0 | 592 | uncharacterized protein |
| gene526 | 66.67 | 0.41 | 37.4 | uridine-cytidine kinase 2-b |
| gene527 | 74.12 | 1.00E-142 | 419 | N6-adenosine-methyltransferase 70 kDa subunit |
| gene531 | 79.12 | 9.00E-100 | 304 | fructose-bisphosphate aldolase C-B-like |
| gene538 | 60.82 | 3.00E-100 | 308 | bisphosphoglycerate-independent phosphoglycerate mutase |
| gene539 | 73.61 | 2.00E-28 | 108 | uncharacterized protein C4orf22 homolog |
| gene543 | 37 | 4.00E-133 | 414 | hypothetical protein |
| gene545 | 42.19 | 7.00E-85 | 269 | coiled-coil domain-containing protein 113 |
| gene551 | 53.96 | 2.00E-98 | 299 | protein YIF1B |
| gene555 | 76.86 | 5.00E-143 | 410 | mitotic checkpoint protein BUB3 |
| gene563 | 37 | 4.00E-85 | 286 | uncharacterized protein LOC100205990 |
| gene565 | 62.91 | 7.00E-113 | 334 | exosome complex component RRP4 |
| gene567 | 52 | 2.00E-152 | 448 | hypothetical protein |
| gene570 | 45.82 | 7.00E-113 | 343 | glucose-fructose oxidoreductase domain-containing protein 1 |
| gene589 | 60.69 | 8.00E-65 | 213 | soluble calcium-activated nucleotidase 1 |
| gene590 | 66.29 | 1.00E-125 | 365 | cell cycle checkpoint protein RAD1 |
| gene594 | 48.48 | 3.00E-08 | 58.2 | radial spoke head 10 homolog B-like |
| gene598 | 79 | 0 | 664 | predicted protein |
| gene599 | 47.24 | 8.00E-139 | 422 | transcription factor 25-like |
| gene605 | 80 | 1.00E-123 | 355 | proteasome subunit beta type-3-like |
| gene607 | 59.19 | 4.00E-102 | 313 | succinyl-CoA ligase [GDP-forming] subunit beta, mitochondrial-like |
| gene609 | 79.3 | 9.00E-139 | 396 | proteasome subunit alpha type-6 |
| gene611 | 47.57 | 5.00E-51 | 174 | phosphonoacetaldehyde hydrolase |
| gene616 | 73.17 | 3.00E-12 | 70.5 | AP-3 complex subunit mu-1-like |
| gene617 | 48.37 | 7.00E-150 | 444 | ADP-dependent glucokinase-like |
| gene640 | 72.47 | 2.00E-109 | 329 | radial spoke head protein 3 homolog, partial |
| gene642 | 68.87 | 6.00E-43 | 158 | glycerol-3-phosphate acyltransferase 4 |
| gene643 | 60.25 | 6.00E-67 | 213 | syntenin-2, partial |
| gene659 | 41.97 | 2.00E-117 | 357 | tetratricopeptide repeat protein 38 |
| gene662 | 81.75 | 1.00E-67 | 213 | ras-related protein Rab-18 |
| gene664 | 57.78 | 1.00E-100 | 304 | hypothetical protein D910_07068 |
| gene666 | 56.59 | 6.00E-39 | 148 | LOW QUALITY PROTEIN: nicalin, partial |
| gene682 | 56.79 | 1.00E-26 | 108 | WD repeat and FYVE domain-containing protein 2-like |
| gene697 | 62.5 | 3.00E-20 | 92 | mitochondrial import inner membrane translocase subunit TIM44-like |
| gene704 | 55 | 4.00E-91 | 299 | Pyruvate dehydrogenase [acetyl-transferring]-phosphatase 2, mitochondrial |
| gene707 | 65.71 | 3.00E-05 | 50.1 | Similar to Solute carrier family 25 member 40 |
| gene712 | 79.87 | 5.00E-91 | 271 | glucose-induced degradation protein 4 homolog |
| gene720 | 71.88 | 3.00E-08 | 56.6 | aspartyl aminopeptidase |
| gene723 | 56.69 | 1.00E-45 | 160 | G2/mitotic-specific cyclin-B1-like |
| gene728 | 65.17 | 3.00E-97 | 293 | U3 small nucleolar ribonucleoprotein protein IMP4-like |
| gene734 | 58.85 | 1.00E-78 | 244 | CCR4-NOT transcription complex subunit 8 |
| gene742 | 51.22 | 7.00E-08 | 59.3 | coiled-coil domain-containing protein 81 |
| gene744 | 56.14 | 5.00E-114 | 338 | ester hydrolase C11orf54 homolog |
| gene748 | 82.81 | 2.00E-29 | 122 | NLS-YFP-Lim4-CFP |
| gene750 | 39.69 | 4.00E-54 | 189 | prenylcysteine oxidase-like |
| gene762 | 41 | 4.00E-79 | 255 | alkylated DNA repair protein alkB homolog 1-like |
| gene763 | 73.25 | 4.00E-123 | 358 | UDP-glucose 4-epimerase |
| gene774 | 44.08 | 9.00E-118 | 362 | oxidative stress-induced growth inhibitor 2 |
| gene783 | 56.52 | 9.00E-75 | 248 | apoptosis-inducing factor 1, mitochondrial-like |
| gene786 | 73.58 | 3.00E-103 | 305 | 40S ribosomal protein S6 |
| gene796 | 56.9 | 0 | 627 | negative elongation factor D-like |
| gene800 | 66.54 | 7.00E-96 | 296 | phosphate carrier protein, mitochondrial |
| gene801 | 49.16 | 1.00E-97 | 297 | methionine adenosyltransferase 2 subunit beta |
| gene802 | 80.37 | 4.00E-96 | 285 | GTP cyclohydrolase 1 |
| gene814 | 66.01 | 6.00E-104 | 308 | WD repeat-containing protein 82 |
| gene816 | 42.89 | 8.00E-128 | 390 | sialate O-acetylesterase |
| gene820 | 55.32 | 1.00E-80 | 252 | KRR1 small subunit processome component homolog |
| gene821 | 56 | 2.00E-20 | 93.6 | V-type proton ATPase subunit C 2 |
| gene831 | 44.86 | 1.00E-86 | 272 | nucleoporin Nup43 |
| gene839 | 81.11 | 7.00E-99 | 292 | ras-related protein Ral-B |
| gene851 | 69 | 2.00E-102 | 306 | cytochrome b5 domain-containing protein 1-like |
| gene856 | 58 | 5.00E-109 | 329 | G1/S-specific cyclin-D2 |
| gene858 | 44.25 | 2.00E-93 | 291 | diphthamide biosynthesis protein 2-like |
| gene860 | 48.93 | 2.00E-65 | 213 | general transcription factor IIE subunit 2-like |
| gene865 | 55.6 | 3.00E-104 | 310 | proteasome subunit beta type-4-like |
| gene866 | 47.04 | 3.00E-94 | 297 | programmed cell death protein 4-like isoform X2 |
| gene883 | 56.2 | 4.00E-101 | 310 | N-acyl-phosphatidylethanolamine-hydrolyzing phospholipase D |
| gene885 | 80.12 | 2.00E-92 | 275 | Chain A, Crystal Structure Of Wild-type Rab11 Complexed To Fip2 |
| gene886 | 54.84 | 5.00E-19 | 87.8 | BTB/POZ domain-containing protein KCTD6 |
| gene891 | 53.24 | 2.00E-106 | 322 | nuclear inhibitor of protein phosphatase 1 |
| gene893 | 55.29 | 1.00E-91 | 284 | ribosome biogenesis regulatory protein homolog |
| gene916 | 71.52 | 2.00E-79 | 241 | ras-related protein Rab-24-like |
| gene919 | 54.29 | 8.00E-17 | 80.9 | factor VIII intron 22 protein, partial |
| gene932 | 86.82 | 3.00E-77 | 233 | ubiquitin-conjugating enzyme E2 N |
| gene940 | 59.04 | 1.00E-75 | 239 | arfaptin-2 |
| gene943 | 56.92 | 7.00E-36 | 146 | nuclear pore complex protein Nup107 |
| gene944 | 61.62 | 3.00E-81 | 252 | phytanoyl-CoA dioxygenase, peroxisomal |
| gene945 | 95.56 | 8.00E-20 | 91.3 | transcriptional enhancer factor TEF-3 |
| gene959 | 57.75 | 3.00E-21 | 95.5 | dual specificity protein phosphatase 6 |
| gene963 | 59.3 | 3.00E-113 | 338 | ribosome production factor 2 homolog |
| gene974 | 54.51 | 4.00E-99 | 297 | UPF0160 protein MYG1, mitochondrial, partial |
| gene979 | 64.67 | 2.00E-75 | 232 | GTP-binding protein Rheb |
| gene981 | 63.64 | 2.00E-94 | 285 | TP53-regulating kinase-like |
| gene985 | 72.32 | 3.00E-84 | 253 | ras-related protein Rap-2a |
| gene992 | 65.52 | 2.00E-36 | 135 | S-methyl-5'-thioadenosine phosphorylase-like |
| gene997 | 68.97 | 9.00E-05 | 46.6 | putative chaperone protein DnaJ |
| gene1003 | 77.7 | 3.00E-77 | 233 | 60s ribosomal protein l27a |
| gene1024 | 53 | 4.00E-90 | 278 | voltage-dependent anion-selective channel protein 2 |
| gene1031 | 41.95 | 7.00E-38 | 152 | targeting protein for Xklp2 |
| gene1042 | 57.3 | 6.00E-70 | 226 | tRNA methyltransferase 10 homolog A |
| gene1050 | 47.65 | 2.00E-75 | 248 | adenylyl cyclase-associated protein 1 |
| gene1054 | 56.9 | 2.00E-08 | 57.4 | dnaJ homolog subfamily B member 8 |
| gene1060 | 82.35 | 7.00E-49 | 179 | hypothetical protein D910_04853, partial |
| gene1082 | 82.3 | 8.00E-126 | 363 | eukaryotic translation initiation factor 6-like, partial |
| gene1083 | 79.63 | 7.00E-22 | 94.4 | uncharacterized protein LOC102054319 |
| gene1099 | 42.9 | 9.00E-86 | 270 | tetratricopeptide repeat protein 5 |
| gene1106 | 54.4 | 4.00E-44 | 151 | Golgi SNAP receptor complex member 1 |
| gene1112 | 42 | 5.00E-83 | 270 | ras association domain-containing protein 1-like |
| gene1125 | 68.42 | 3.00E-10 | 63.2 | glycosyltransferase-like domain-containing protein 1 |
| gene1130 | 80.77 | 2.00E-07 | 55.8 | hypothetical protein D910_02310, partial |
| gene1135 | 52.76 | 7.00E-26 | 110 | CD2 antigen cytoplasmic tail-binding protein 2-like |
| gene1141 | 82.35 | 6.00E-24 | 94.7 | 60S ribosomal protein L13a-like, partial |
| gene1146 | 91 | 6.00E-74 | 226 | gamma-aminobutyric acid receptor-associated protein-like 2-like |
| gene1147 | 54 | 5.00E-81 | 254 | cytochrome b561-like |
| gene1164 | 51.87 | 2.00E-92 | 283 | LOW QUALITY PROTEIN: lysosomal thioesterase PPT2-A-like |
| gene1175 | 64.15 | 2.00E-34 | 136 | hypothetical protein D910_03531 |
| gene1179 | 74.07 | 9.00E-21 | 87 | eukaryotic translation initiation factor 5A-1-like, partial |
| gene1191 | 42.35 | 6.00E-76 | 241 | LOW QUALITY PROTEIN: vesicle-trafficking protein SEC22a-like |
| gene1194 | 43 | 7.00E-92 | 295 | trichohyalin-like |
| gene1195 | 39 | 4.00E-67 | 231 | repulsive guidance molecule A-like |
| gene1197 | 77.78 | 1.00E-143 | 411 | proteasome subunit alpha type-1 |
| gene1211 | 46.89 | 1.00E-59 | 201 | synapse-associated protein 1 |
| gene1215 | 57.03 | 7.00E-155 | 452 | legumain |
| gene1216 | 77.78 | 3.00E-12 | 70.1 | BTB/POZ domain-containing protein KCTD20 |
| gene1219 | 76.86 | 2.00E-55 | 178 | 60S ribosomal protein L35 |
| gene1225 | 46.9 | 2.00E-30 | 117 | neuferricin-like |
| gene1234 | 64.29 | 2.00E-07 | 50.4 | thioredoxin domain-containing protein 8-like |
| gene1236 | 71.93 | 3.00E-21 | 92.8 | uncharacterized protein LOC102187703 |
| gene1240 | 45 | 1.00E-70 | 234 | uncharacterized aminotransferase C660.12c-like |
| gene1243 | 51.79 | 2.00E-80 | 254 | isoaspartyl peptidase/L-asparaginase-like isoform X1 |
| gene1254 | 60.2 | 4.00E-124 | 364 | mitochondrial carnitine/acylcarnitine carrier protein-like |
| gene1257 | 49 | 6.00E-116 | 356 | threonine synthase-like 2 isoform X3 |
| gene1259 | 42.19 | 9.00E-07 | 54.3 | Calcium-binding mitochondrial carrier protein |
| gene1261 | 43 | 1.00E-58 | 194 | transmembrane emp24 domain-containing protein 7-like |
| gene1270 | 64.41 | 1.00E-65 | 221 | muskelin isoform X9 |
| gene1273 | 69.64 | 3.00E-44 | 149 | charged multivesicular body protein 1a |
| gene1275 | 50.47 | 7.00E-30 | 113 | gamma-secretase subunit APH-1A-like, partial |
| gene1295 | 59.52 | 2.00E-09 | 58.9 | inositol hexakisphosphate kinase 2-like, partial |
| gene1305 | 47.92 | 3.00E-04 | 47 | hypothetical protein D910_12511 |
| gene1308 | 46.32 | 2.00E-19 | 96.3 | sphingomyelin phosphodiesterase 4-like |
| gene1311 | 58.54 | 1.00E-51 | 171 | receptor expression-enhancing protein 6 |
| gene1319 | 39 | 4.00E-52 | 180 | Polyglutamine-binding protein 1 |
| gene1325 | 59.04 | 7.00E-108 | 324 | LOW QUALITY PROTEIN: uroporphyrinogen decarboxylase-like |
| gene1326 | 54.23 | 2.00E-42 | 150 | microtubule-associated protein RP/EB family member 1-like |
| gene1339 | 53.62 | 1.00E-18 | 85.5 | 39S ribosomal protein L15, mitochondrial |
| gene1347 | 47.57 | 2.00E-76 | 240 | sin3 histone deacetylase corepressor complex component SDS3 |
| gene1348 | 47.17 | 3.00E-10 | 58.9 | isochorismatase domain-containing protein 2, mitochondrial |
| gene1369 | 73.33 | 431 | 24.6 | hypothetical protein P615_21255 |
| gene1376 | 81.82 | 4.00E-69 | 213 | 40S ribosomal protein S12 |
| gene1379 | 44.94 | 2.00E-58 | 197 | nucleoporin NUP53-like |
| gene1381 | 61.36 | 2.00E-11 | 62 | tumor necrosis factor alpha-induced protein 8-like protein 1 |
| gene1382 | 77.11 | 6.00E-36 | 131 | cyclic AMP-responsive element-binding protein 1 |
| gene1385 | 56.79 | 1.00E-20 | 91.3 | protein IMPACT |
| gene1392 | 76 | 7.00E-22 | 87.8 | 60S ribosomal protein L9-like, partial |
| gene1393 | 68.35 | 2.00E-27 | 112 | Similar to Malate dehydrogenase, cytoplasmic |
| gene1394 | 35 | 5.00E-42 | 155 | zinc finger CCCH domain-containing protein 10-like |
| gene1405 | 52 | 8.00E-65 | 210 | WASH complex subunit CCDC53 |
| gene1407 | 50.67 | 7.00E-11 | 67 | U3 small nucleolar RNA-associated protein 14 homolog A |
| gene1413 | 66.67 | 4.00E-44 | 151 | estradiol 17-beta-dehydrogenase 8 |
| gene1422 | 52.63 | 2.00E-15 | 72.8 | ras-related protein Rap-1b-like |
| gene1432 | 46.86 | 3.00E-52 | 174 | calcyclin-binding protein |
| gene1441 | 76.47 | 2.00E-10 | 62.4 | uncharacterized protein LOC102065257, partial |
| gene1443 | 48.13 | 4.00E-56 | 182 | RWD domain-containing protein 4 |
| gene1444 | 54.35 | 2.00E-20 | 93.6 | transcription factor AP-1-like |
| gene1450 | 94.78 | 5.00E-77 | 231 | hypothetical protein D910_01827 |
| gene1462 | 59.2 | 5.00E-33 | 122 | ATP synthase subunit delta, mitochondrial |
| gene1463 | 91.57 | 3.00E-48 | 164 | uncharacterized protein LOC102111343 |
| gene1483 | 50 | 9.00E-51 | 168 | tetratricopeptide repeat protein 36-like |
| gene1486 | 62.64 | 3.00E-30 | 112 | histone deacetylase complex subunit SAP18 |
| gene1488 | 35.67 | 8.00E-96 | 306 | LOW QUALITY PROTEIN: 26S proteasome non-ATPase regulatory subunit 5, partial |
| gene1493 | 54.48 | 5.00E-45 | 160 | ubiquinone biosynthesis protein COQ9, mitochondrial-like |
| gene1497 | 95.08 | 6.00E-33 | 127 | aminopeptidase Q-like |
| gene1501 | 70.49 | 2.00E-23 | 95.9 | density-regulated protein-like protein |
| gene1505 | 79.49 | 2.00E-15 | 73.6 | superoxide dismutase [Mn], mitochondrial |
| gene1507 | 88 | 2.00E-44 | 147 | protein BUD31 homolog |
| gene1508 | 58.87 | 1.00E-51 | 179 | E3 ubiquitin-protein ligase KCMF1-like |
| gene1525 | 37.5 | 4.00E-04 | 43.9 | apolipoprotein D-like |
| gene1527 | 53.77 | 1.00E-32 | 136 | bromodomain-containing protein 9 |
| gene1528 | 36.25 | 2.00E-46 | 166 | MOSC domain-containing protein 1, mitochondrial-like |
| gene1529 | 32.16 | 2.00E-19 | 94 | sorting nexin-14-like |
| gene1532 | 62.07 | 2.00E-20 | 89 | nanos homolog 2 |
| gene1535 | 52.5 | 5.00E-47 | 159 | sepiapterin reductase |
| gene1541 | 69.83 | 4.00E-90 | 275 | protein DDI1 homolog 2 |
| gene1545 | 82.93 | 2.00E-13 | 67.8 | nascent polypeptide-associated complex subunit alpha, muscle-specific form-like |
| gene1546 | 28.75 | 7.00E-12 | 71.2 | transmembrane protein 214 |
| gene1552 | 47.34 | 5.00E-42 | 147 | Ribosomal RNA-processing protein 7 homolog A |
| gene1553 | 63.64 | 7.00E-15 | 72.8 | tubulin polymerization-promoting protein family member 2 |
| gene1557 | 44.62 | 2.00E-08 | 57 | methyl-CpG-binding domain protein 4-like |
| gene1586 | 53.06 | 8.00E-06 | 53.9 | axin-1-like isoform X1 |
| gene1598 | 68.7 | 3.00E-46 | 156 | charged multivesicular body protein 4c-like |
| gene1619 | 51.16 | 2.00E-08 | 56.6 | UPF0472 protein C16orf72 homolog, partial |
| gene1627 | 42.69 | 1.00E-25 | 105 | RRP15-like protein |
| gene1648 | 74.89 | 1.00E-126 | 382 | bifunctional purine biosynthesis protein PURH |
| gene1672 | 63 | 7.00E-50 | 168 | predicted protein |
| gene1674 | 45.45 | 2.00E-22 | 93.2 | cytochrome c oxidase subunit 4 isoform 2, mitochondrial-like |
| gene1682 | 52.63 | 1.00E-41 | 149 | saccharopine dehydrogenase-like oxidoreductase |
| gene1685 | 42.06 | 4.00E-24 | 97.8 | FUN14 domain-containing protein 1-like |
| gene1691 | 47.75 | 3.00E-26 | 105 | cation transport regulator-like protein 1-like |
| gene1700 | 63.16 | 5.00E-23 | 93.2 | adrenodoxin, mitochondrial-like |
| gene1709 | 70.49 | 2.00E-25 | 100 | CDGSH iron-sulfur domain-containing protein 1-like |
| gene1729 | 40.56 | 1.00E-18 | 87.8 | hsp70-binding protein 1 isoform X2 |
| gene1733 | 53.92 | 6.00E-32 | 117 | transmembrane protein 208 |
| gene1737 | 44.53 | 1.00E-22 | 97.1 | androgen-dependent TFPI-regulating protein |
| gene1745 | 34 | 4.00E-25 | 105 | 40S ribosomal protein S8 |
| gene1763 | 48.48 | 1.00E-13 | 68.9 | translocon-associated protein subunit delta |
| gene1795 | 87.5 | 41026 | 20.4 | LOW QUALITY PROTEIN: zinc finger protein 850 |
| gene1805 | 45.95 | 2.00E-17 | 81.3 | transcription cofactor vestigial-like protein 4 |

**Table S2.3** Best-fit aa substitution model of 252 and 153 partitioning gene regions.

252 partitioning subsets

| Subset | Best model | Partitions | Subset | Best model | Partitions | Subset | Best model | Partitions | Subset | Best model | Partitions |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | JTT+G | Gene11 | 64 | LG+G | Gene385 | 127 | LG+G | Gene744 | 190 | LG+G | Gene1259 |
| 2 | JTT+G | Gene18, Gene734 | 65 | WAG+I+G | Gene389 | 128 | DAYHOFF+G | Gene748 | 191 | VT+I+G | Gene1261 |
| 3 | JTT+G | Gene1024, Gene22 | 66 | LG+G | Gene394 | 129 | WAG+G | Gene750 | 192 | JTT+G | Gene1270 |
| 4 | JTT+G+F | Gene27 | 67 | LG+G | Gene395 | 130 | WAG+G | Gene762 | 193 | LG+G | Gene1292 |
| 5 | LG+G | Gene35 | 68 | LG+G | Gene402 | 131 | LG+G | Gene763, Gene865, Gene886 | 194 | JTT+G | Gene1305 |
| 6 | JTT+G | Gene41 | 69 | JTT+I+G | Gene404 | 132 | JTT+G | Gene774 | 195 | JTT+G | Gene1308 |
| 7 | JTT+G | Gene59 | 70 | JTT+G | Gene407 | 133 | LG+G | Gene1273, Gene783 | 196 | VT+G | Gene1311 |
| 8 | JTT+G | Gene61 | 71 | CPREV+G | Gene416 | 134 | LG+G | Gene1325, Gene786 | 197 | JTT+G | Gene1319 |
| 9 | JTT+G | Gene65 | 72 | VT+G | Gene428 | 135 | JTT+G | Gene796 | 198 | JTT+G | Gene1326 |
| 10 | LG+G | Gene467, Gene66 | 73 | LG+G | Gene435 | 136 | VT+G | Gene800 | 199 | VT+G | Gene1339 |
| 11 | LG+I+G | Gene68 | 74 | JTT+G | Gene444 | 137 | LG+I+G | Gene801 | 200 | JTT+G | Gene1347 |
| 12 | JTT+G | Gene195, Gene70 | 75 | JTT+G | Gene445 | 138 | LG+G | Gene1236, Gene802, Gene963 | 201 | LG+G | Gene1348 |
| 13 | WAG+G | Gene89 | 76 | VT+G | Gene1598, Gene450 | 139 | WAG+I+G | Gene816 | 202 | VT+I+G | Gene1369 |
| 14 | WAG+G | Gene99 | 77 | LG+G | Gene452 | 140 | JTT+G | Gene820 | 203 | VT+G | Gene1376 |
| 15 | VT+G+F | Gene103 | 78 | VT+G | Gene455 | 141 | LG+G | Gene821, Gene885 | 204 | JTT+I+G | Gene1379 |
| 16 | LG+G | Gene123 | 79 | LG+G | Gene460 | 142 | JTT+G | Gene831 | 205 | JTT+G | Gene1382 |
| 17 | LG+G | Gene1125, Gene124, Gene216 | 80 | LG+G | Gene461 | 143 | JTT+I+G | Gene839 | 206 | LG+G | Gene1385 |
| 18 | JTT+G | Gene149 | 81 | LG+I+G | Gene469 | 144 | JTT+G | Gene851 | 207 | CPREV+G | Gene1392 |
| 19 | WAG+I+G | Gene151 | 82 | LG+G | Gene472 | 145 | JTT+G | Gene856 | 208 | WAG+G | Gene1393 |
| 20 | LG+G | Gene154 | 83 | JTT+G | Gene476 | 146 | LG+G | Gene858 | 209 | DAYHOFF+G+F | Gene1394 |
| 21 | LG+G | Gene156, Gene538 | 84 | LG+G | Gene1197, Gene1381, Gene1545, Gene478 | 147 | LG+G | Gene1194, Gene860 | 210 | JTT+G+F | Gene1405 |
| 22 | JTT+G | Gene157 | 85 | LG+G | Gene479, Gene697 | 148 | LG+G | Gene866 | 211 | JTT+G | Gene1407 |
| 23 | JTT+G | Gene163 | 86 | JTT+G | Gene495 | 149 | LG+G | Gene883 | 212 | WAG+G | Gene1413 |
| 24 | JTT+G | Gene164 | 87 | VT+G | Gene496 | 150 | JTT+G | Gene891 | 213 | JTT+G | Gene1432 |
| 25 | JTT+G | Gene166 | 88 | JTT+G | Gene498 | 151 | LG+G | Gene893 | 214 | JTT+G | Gene1443 |
| 26 | JTT+G | Gene167 | 89 | JTT+G | Gene512 | 152 | JTT+G | Gene916 | 215 | JTT+G+F | Gene1444 |
| 27 | LG+G | Gene170 | 90 | JTT+G | Gene521 | 153 | JTT+G | Gene919 | 216 | VT+I+G | Gene1450 |
| 28 | JTT+G | Gene181 | 91 | JTT+G | Gene523 | 154 | LG+I+G | Gene932 | 217 | JTT+G | Gene1462 |
| 29 | JTT+G | Gene182 | 92 | JTT+G | Gene526 | 155 | JTT+G | Gene944 | 218 | LG+G | Gene1463 |
| 30 | JTT+G | Gene200 | 93 | JTT+G | Gene527 | 156 | JTT+G | Gene945 | 219 | JTT+G | Gene1483 |
| 31 | JTT+G | Gene204, Gene664 | 94 | WAG+G | Gene531 | 157 | JTT+G | Gene959 | 220 | JTT+G | Gene1486 |
| 32 | LG+G | Gene211 | 95 | JTT+G | Gene539, Gene814, Gene985 | 158 | JTT+G | Gene974 | 221 | LG+G | Gene1488 |
| 33 | LG+G | Gene218 | 96 | LG+G | Gene543 | 159 | JTT+G | Gene979 | 222 | JTT+G | Gene1493 |
| 34 | JTT+G | Gene219, Gene388 | 97 | JTT+G | Gene545 | 160 | JTT+G | Gene981 | 223 | LG+G | Gene1497 |
| 35 | VT+G | Gene237 | 98 | JTT+G | Gene551, Gene643 | 161 | WAG+G | Gene1763, Gene992 | 224 | JTT+G | Gene1501 |
| 36 | JTT+G | Gene247 | 99 | JTT+G | Gene1422, Gene555 | 162 | BLOSUM62+G | Gene1003 | 225 | VT+G | Gene1505 |
| 37 | JTT+G | Gene1557, Gene251 | 100 | JTT+G | Gene563 | 163 | LG+G | Gene1031 | 226 | BLOSUM62+G | Gene1507 |
| 38 | LG+G | Gene256 | 101 | JTT+G | Gene565, Gene943 | 164 | JTT+G | Gene1042 | 227 | JTT+G | Gene1508 |
| 39 | LG+G | Gene257 | 102 | LG+I+G | Gene567 | 165 | DCMUT+G | Gene1050 | 228 | WAG+G | Gene1525 |
| 40 | JTT+G | Gene274 | 103 | JTT+G | Gene1733, Gene570 | 166 | LG+G | Gene1082 | 229 | JTT+G | Gene1527 |
| 41 | JTT+G | Gene275 | 104 | LG+G | Gene589 | 167 | WAG+G | Gene1083 | 230 | LG+G | Gene1528 |
| 42 | LG+G | Gene281 | 105 | JTT+G | Gene590 | 168 | JTT+G | Gene1099 | 231 | JTT+G | Gene1529 |
| 43 | JTT+G | Gene284 | 106 | JTT+G | Gene594 | 169 | JTT+G | Gene1106 | 232 | JTT+G | Gene1532 |
| 44 | LG+I+G | Gene286 | 107 | WAG+G | Gene598 | 170 | JTT+G+F | Gene1112, Gene1295 | 233 | LG+G | Gene1535 |
| 45 | LG+G | Gene291 | 108 | JTT+G | Gene599 | 171 | JTT+G | Gene1135 | 234 | JTT+G | Gene1541 |
| 46 | WAG+G | Gene305 | 109 | LG+I+G | Gene605 | 172 | LG+G | Gene1141 | 235 | JTT+G | Gene1546 |
| 47 | LG+I+G | Gene307 | 110 | LG+G | Gene607 | 173 | LG+G | Gene1146 | 236 | JTT+G | Gene1552 |
| 48 | LG+G | Gene317 | 111 | JTT+G | Gene1060, Gene609 | 174 | LG+G | Gene1147 | 237 | DAYHOFF+I+G | Gene1553 |
| 49 | LG+G | Gene320 | 112 | LG+G | Gene611 | 175 | LG+G | Gene1164 | 238 | JTT+G | Gene1586 |
| 50 | LG+G | Gene322 | 113 | LG+G | Gene616 | 176 | JTT+G | Gene1175 | 239 | JTT+G | Gene1619 |
| 51 | LG+G | Gene329 | 114 | LG+G | Gene617 | 177 | LG+G | Gene1179 | 240 | JTT+G+F | Gene1627 |
| 52 | LG+G | Gene338 | 115 | JTT+G+F | Gene640 | 178 | JTT+G | Gene1191 | 241 | WAG+G | Gene1648 |
| 53 | JTT+G | Gene1441, Gene352 | 116 | LG+G | Gene642 | 179 | WAG+G | Gene1195 | 242 | LG+G | Gene1672 |
| 54 | VT+G | Gene353 | 117 | LG+G | Gene659 | 180 | JTT+G+F | Gene1211 | 243 | LG+G | Gene1674 |
| 55 | JTT+G | Gene354 | 118 | LG+G | Gene662 | 181 | WAG+I+G | Gene1215 | 244 | LG+G | Gene1682 |
| 56 | VT+G | Gene355 | 119 | LG+G | Gene1054, Gene1130, Gene666, Gene997 | 182 | JTT+G | Gene1216 | 245 | LG+G | Gene1685 |
| 57 | BLOSUM62+G | Gene360 | 120 | JTT+G | Gene682 | 183 | LG+G | Gene1219 | 246 | VT+G | Gene1691 |
| 58 | LG+G | Gene362 | 121 | VT+G | Gene704 | 184 | WAG+G | Gene1225 | 247 | VT+G | Gene1700 |
| 59 | JTT+G | Gene366, Gene940 | 122 | JTT+G | Gene712 | 185 | JTT+G | Gene1234 | 248 | LG+I+G | Gene1709 |
| 60 | LG+G | Gene372 | 123 | LG+G | Gene720 | 186 | LG+G | Gene1240 | 249 | LG+I+G | Gene1729 |
| 61 | JTT+G | Gene373 | 124 | LG+G | Gene723 | 187 | WAG+G | Gene1243 | 250 | LG+I+G | Gene1737 |
| 62 | JTT+G | Gene378 | 125 | LG+I+G | Gene728 | 188 | LG+G | Gene1254 | 251 | VT+G | Gene1745 |
| 63 | JTT+G | Gene382, Gene707 | 126 | JTT+G | Gene1805, Gene742 | 189 | LG+G | Gene1257 | 252 | JTT+G | Gene1795 |

153 partitioning subsets

| Subset | Best model | Partitions |
|---|---|---|
| 1 | JTT+G | Gene11, Gene1197, Gene445 |
| 2 | JTT+G | Gene1060, Gene18, Gene734 |
| 3 | JTT+G | Gene22 |
| 4 | LG+G | Gene1325, Gene27, Gene329, Gene786 |
| 5 | LG+G | Gene1682, Gene35 |
| 6 | JTT+G | Gene354, Gene41, Gene796, Gene856 |
| 7 | JTT+G | Gene1175, Gene59 |
| 8 | JTT+G | Gene362, Gene61 |
| 9 | JTT+G | Gene1462, Gene65 |
| 10 | LG+G | Gene467, Gene66 |
| 11 | LG+I+G | Gene68 |
| 12 | JTT+G | Gene195, Gene70, Gene885 |
| 13 | WAG+G | Gene1505, Gene89 |
| 14 | LG+G | Gene1257, Gene99 |
| 15 | LG+G | Gene103, Gene616 |
| 16 | LG+G | Gene123, Gene1292 |
| 17 | LG+G | Gene1147, Gene124, Gene216, Gene281 |
| 18 | JTT+G | Gene1441, Gene149, |
| 19 | WAG+I+G | Gene151 |
| 20 | LG+G | Gene154, Gene478 |
| 21 | LG+G | Gene156, Gene538 |
| 22 | JTT+G | Gene157 |
| 23 | JTT+G | Gene1112, Gene1486, Gene164 |
| 24 | JTT+G | Gene166, Gene682 |
| 25 | JTT+G | Gene1527, Gene167 |
| 26 | LG+G | Gene170 |
| 27 | JTT+G | Gene181 |
| 28 | JTT+I+G | Gene1508, Gene182 |
| 29 | JTT+G | Gene200, Gene461, Gene551, Gene643 |
| 30 | JTT+G | Gene1099, Gene204, Gene860 |
| 31 | LG+G | Gene1141, Gene211 |
| 32 | LG+G | Gene218 |
| 33 | JTT+G | Gene219, Gene388 |
| 34 | VT+G | Gene1311, Gene237 |
| 35 | JTT+G | Gene247, Gene545 |
| 36 | JTT+G | Gene1557, Gene251, Gene257, Gene565, Gene594, Gene943 |
| 37 | LG+G | Gene1674, Gene1709, Gene256, Gene435 |
| 38 | JTT+G | Gene274, Gene523 |
| 39 | JTT+G | Gene275, Gene407, Gene539 ,Gene748, Gene814, Gene985 |
| 40 | JTT+G | Gene1347, Gene284, Gene479, Gene599, Gene697, Gene891 |
| 41 | LG+I+G | Gene286 |
| 42 | LG+G | Gene291, Gene858 |
| 43 | WAG+G | Gene305 |
| 44 | LG+I+G | Gene307 |
| 45 | LG+G | Gene1054, Gene317, Gene607, Gene640, Gene666, Gene997 |
| 46 | LG+G | Gene1254, Gene1259, Gene320 |
| 47 | LG+G | Gene1031, Gene1432, Gene1488, Gene1672, Gene322 |
| 48 | LG+G | Gene338 |
| 49 | LG+G | Gene1381, Gene1545, Gene352, Gene821 |
| 50 | VT+G | Gene353, Gene428 |
| 51 | VT+G | Gene1236, Gene355 |
| 52 | LG+G | Gene1763, Gene360 |
| 53 | JTT+G | Gene1234, Gene1319, Gene366, Gene570, Gene820, Gene940 |
| 54 | LG+G | Gene1130, Gene1273, Gene1392, Gene372, Gene395, Gene642, Gene763, Gene783, Gene886 |
| 55 | JTT+G | Gene1394, Gene373, Gene750 |
| 56 | LG+G | Gene1082, Gene378 |
| 57 | JTT+G | Gene382, Gene707, Gene774 |
| 58 | LG+G | Gene385 |
| 59 | WAG+I+G | Gene389 |
| 60 | JTT+G | Gene1125, Gene1305, Gene1444, Gene1501, Gene1552, Gene394, Gene762 |
| 61 | LG+G | Gene402 |
| 62 | JTT+I+G | Gene404 |
| 63 | CPREV+G | Gene416 |
| 64 | JTT+G | Gene444 |
| 65 | VT+G | Gene1598, Gene450 |
| 66 | JTT+G | Gene1483, Gene452, Gene723, Gene916 |
| 67 | VT+G | Gene455, Gene800 |
| 68 | LG+G | Gene1146, Gene460 |
| 69 | LG+I+G | Gene469 |
| 70 | LG+G | Gene1024, Gene1219, Gene472 |
| 71 | JTT+G | Gene1805, Gene476, Gene742 |
| 72 | JTT+G | Gene495 |
| 73 | VT+G | Gene496 |
| 74 | JTT+G | Gene1405, Gene498, Gene589, Gene981 |
| 75 | JTT+G | Gene1532, Gene512, Gene959 |
| 76 | JTT+G | Gene521, Gene831 |
| 77 | JTT+G | Gene526 |
| 78 | JTT+G | Gene1216, Gene527, Gene919 |
| 79 | WAG+G | Gene531 |
| 80 | LG+G | Gene1535, Gene543 |
| 81 | JTT+G | Gene1083, Gene1422, Gene555 |
| 82 | JTT+G | Gene563 |
| 83 | LG+I+G | Gene567 |
| 84 | JTT+G | Gene1106, Gene590 |
| 85 | WAG+G | Gene598 |
| 86 | LG+I+G | Gene605 |
| 87 | LG+G | Gene1497, Gene609 |
| 88 | LG+G | Gene611 |
| 89 | LG+G | Gene617 |
| 90 | LG+G | Gene659 |
| 91 | LG+G | Gene662 |
| 92 | LG+G | Gene1194, Gene664, Gene865 |
| 93 | VT+G | Gene704 |
| 94 | JTT+G | Gene712 |
| 95 | LG+G | Gene1179, Gene720, Gene893 |
| 96 | LG+I+G | Gene728 |
| 97 | LG+G | Gene1733, Gene744, Gene963 |
| 98 | LG+I+G | Gene801 |
| 99 | LG+G | Gene802 |
| 100 | WAG+I+G | Gene816 |
| 101 | JTT+I+G | Gene839 |
| 102 | JTT+G | Gene851 |
| 103 | LG+G | Gene866 |
| 104 | WAG+G | Gene1215, Gene883 |
| 105 | LG+I+G | Gene932 |
| 106 | JTT+G | Gene944 |
| 107 | JTT+G | Gene945, Gene979 |
| 108 | JTT+G | Gene1191, Gene974 |
| 109 | WAG+G | Gene992 |
| 110 | BLOSUM62+G | Gene1003 |
| 111 | JTT+G | Gene1042 |
| 112 | DCMUT+G | Gene1050 |
| 113 | JTT+G | Gene1135 |
| 114 | LG+G | Gene1164 |
| 115 | WAG+G | Gene1195 |
| 116 | JTT+G+F | Gene1211 |
| 117 | WAG+G | Gene1225 |
| 118 | LG+G | Gene1240 |
| 119 | WAG+G | Gene1243 |
| 120 | VT+I+G | Gene1261 |
| 121 | JTT+G | Gene1270 |
| 122 | JTT+G | Gene1295, Gene1308, Gene1407, Gene1443, Gene1407, Gene1443 |
| 123 | JTT+G | Gene1326 |
| 124 | VT+G | Gene1339 |
| 125 | LG+G | Gene1348 |
| 126 | VT+I+G | Gene1369 |
| 127 | VT+G | Gene1376 |
| 128 | JTT+I+G | Gene1379 |
| 129 | JTT+G | Gene1382 |
| 130 | LG+G | Gene1385 |
| 131 | WAG+G | Gene1393 |
| 132 | WAG+G | Gene1413 |
| 133 | VT+I+G | Gene1450 |
| 134 | LG+G | Gene1463 |
| 135 | JTT+G | Gene1493 |
| 136 | BLOSUM62+G | Gene1507 |
| 137 | WAG+G | Gene1525 |
| 138 | LG+G | Gene1528 |
| 139 | JTT+G | Gene1529 |
| 140 | JTT+G | Gene1541 |
| 141 | JTT+G | Gene1546 |
| 142 | DAYHOFF+I+G | Gene1553 |
| 143 | JTT+G | Gene1586 |
| 144 | JTT+G | Gene1619 |
| 145 | JTT+G+F | Gene1627 |
| 146 | WAG+G | Gene1648 |
| 147 | LG+G | Gene1685 |
| 148 | VT+G | Gene1691 |
| 149 | VT+G | Gene1700 |
| 150 | LG+I+G | Gene1729 |
| 151 | LG+I+G | Gene1737 |
| 152 | VT+G | Gene1745 |
| 153 | JTT+G | Gene1795 |

**Table S2.4** Nucleotide substitution saturation test on the 291 nuclear protein coding genes. The nucleotide substitution saturation test calculates an index of substitution saturation (Iss), which is compared to a critical value which is computed for a symmetrical (Iss.cSym) or extremely asymmetrical (Iss.c asym) tree topology. Pinv: proportion of invariable site. Iss: index of substitution saturation. Iss.c sym: critical value for symmetrical tree topology. Iss.c asym: critical value for extremely assymetrical tree topology. T: T value. DF: degrees of freedom. P: probability that Iss is significantly different from the critical value (Iss.c sym or Iss.c asym). Two-tailed tests were used. Only fully resolved sites were analyzed.

| Sites | Pinv | Iss | Iss.c sym | T | DF | P | Iss.c asym | T | DF | P |
|---|---|---|---|---|---|---|---|---|---|---|
| All sites | 0.19 | 0.398 | 0.8375 | 179.0163 | 53849 | 0 | 0.7035 | 124.4593 | 53849 | 0 |
| 1st site | 0.33979 | 0.2104 | 0.8425 | 178.037 | 14633 | 0 | 0.6793 | 132.0678 | 14633 | 0 |
| 2nd site | 0.37969 | 0.7753 | 0.8425 | 27.803 | 13737 | 0 | 0.6793 | 39.7262 | 13737 | 0 |
| 3rd site | 0.35945 | 0.3494 | 0.8425 | 128.7577 | 14200 | 0 | 0.6793 | 86.1407 | 14200 | 0 |

**Table S2.5** The value of convergence index (Maxdiff) for the PhyloBayes runs on each model. If the maxdiff is smaller than 0.1, it indicates a good run while the value equal to 1, it means the run was stuck in a local maximum.

| Bayesian Run | Maxdiff |
|---|---|
| JTT+I+G | 0 |
| CAT+GTR | 0 |
| CAT+possion | 1 |

**Table S2.6** The supporting values for 252 and 153 partitioned phylogenetic trees.

| Data type | Amino acid dataset | | | |
|---|---|---|---|---|
| Method | Maximum Likelihood analyses | | Bayesian Inference | |
| Supporting value | Bootstrap support(%) | | Posterior Probability | |
| Matrix type | Partitions (#) | | Partitions (#) | |
| Selected Nodes/ substition models | 252 | 153 | 252 | 153 |
| ((Actiniaria,(Corallimorpharia, Scleractinia)) | - | - | 1 | 1 |
| (Corallimorpharia, Scleractinia) | 100 | 100 | 1 | 1 |
| (Complexa, Robusta) | 100 | 100 | 1 | 1 |
| (Corynactis, (Ricordea, Rhodactis)) | 100 | 100 | 1 | 1 |
| (Ricordea,Rhodactis) | 91 | 90 | 1 | 1 |
| (((*Pocillopora*, *Madracis*),(((*Pseudodiploria*, *Platygyra*), *Montastraea*),*Fungia*)))) | 100 | 100 | 1 | 1 |
| (*Pocillopora*, *Madracis*) | 100 | 100 | 1 | 1 |
| (((*Pseudodiploria*, *Platygyra*), *Montastraea*),*Fungia*))) | 100 | 100 | 1 | 1 |
| ((*Pseudodiploria*, *Platygyra*), *Montastraea*) | 100 | 100 | 1 | 1 |
| (*Pseudodiploria*, *Platygyra*) | 100 | 100 | 1 | 1 |
| (*Porites*,(*A.digitifera*, *A.millepora*)) | 100 | 100 | 1 | 1 |
| (*A.digitifera*, *A.millepora*) | 100 | 100 | 1 | 1 |
| (*Nematostella*, *Anthopleura*) | 100 | 100 | 1 | 1 |


**Figure S2.1** Bayesian inference with CAT model for concatenated amino acid sequences.

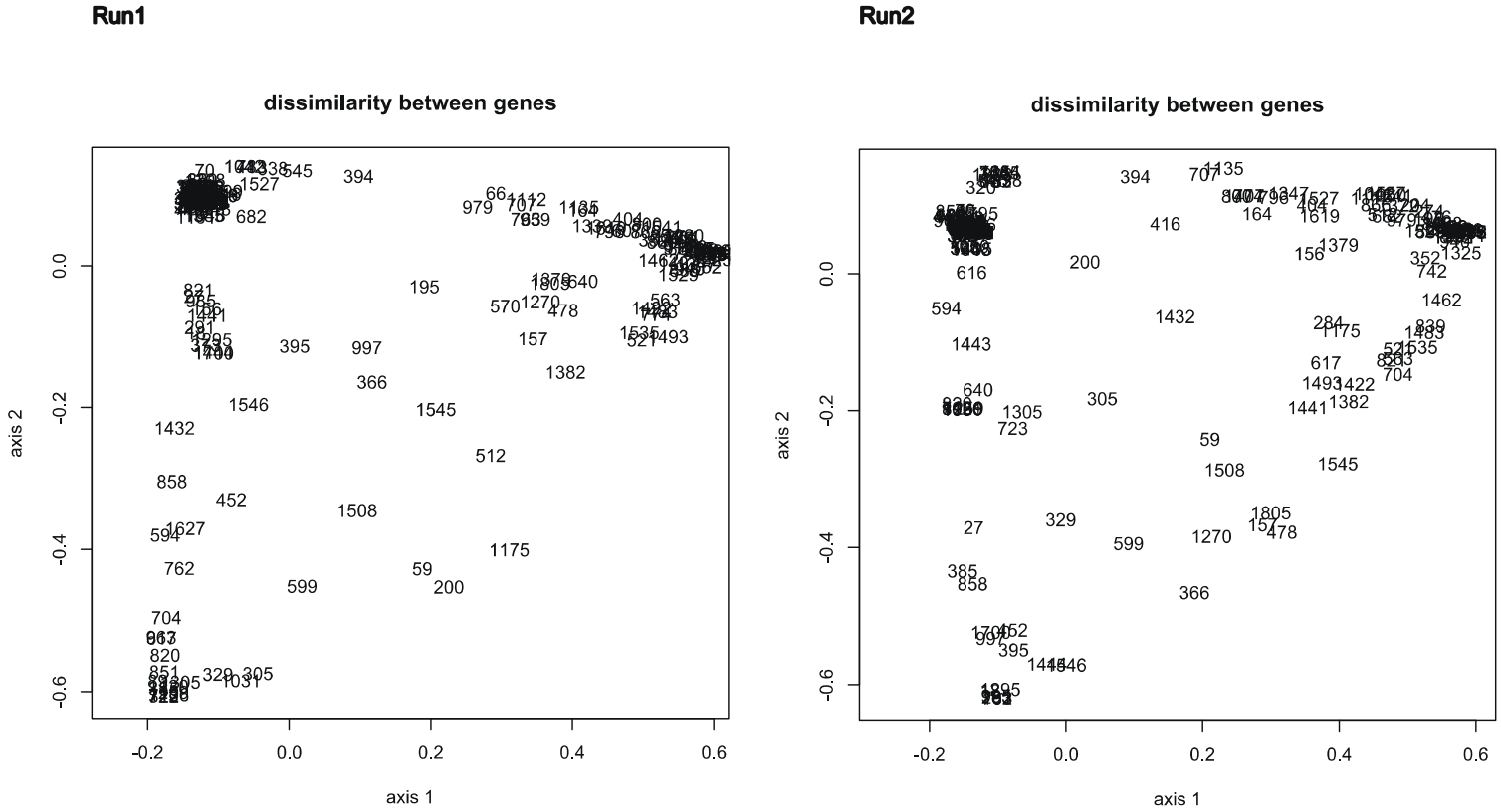



**Figure S2.2** Two-dimensional representation of the 291 anthozoan genes. Similarity between pairs of genes is measured as the posterior probability that two genes share the same tree (α=1).

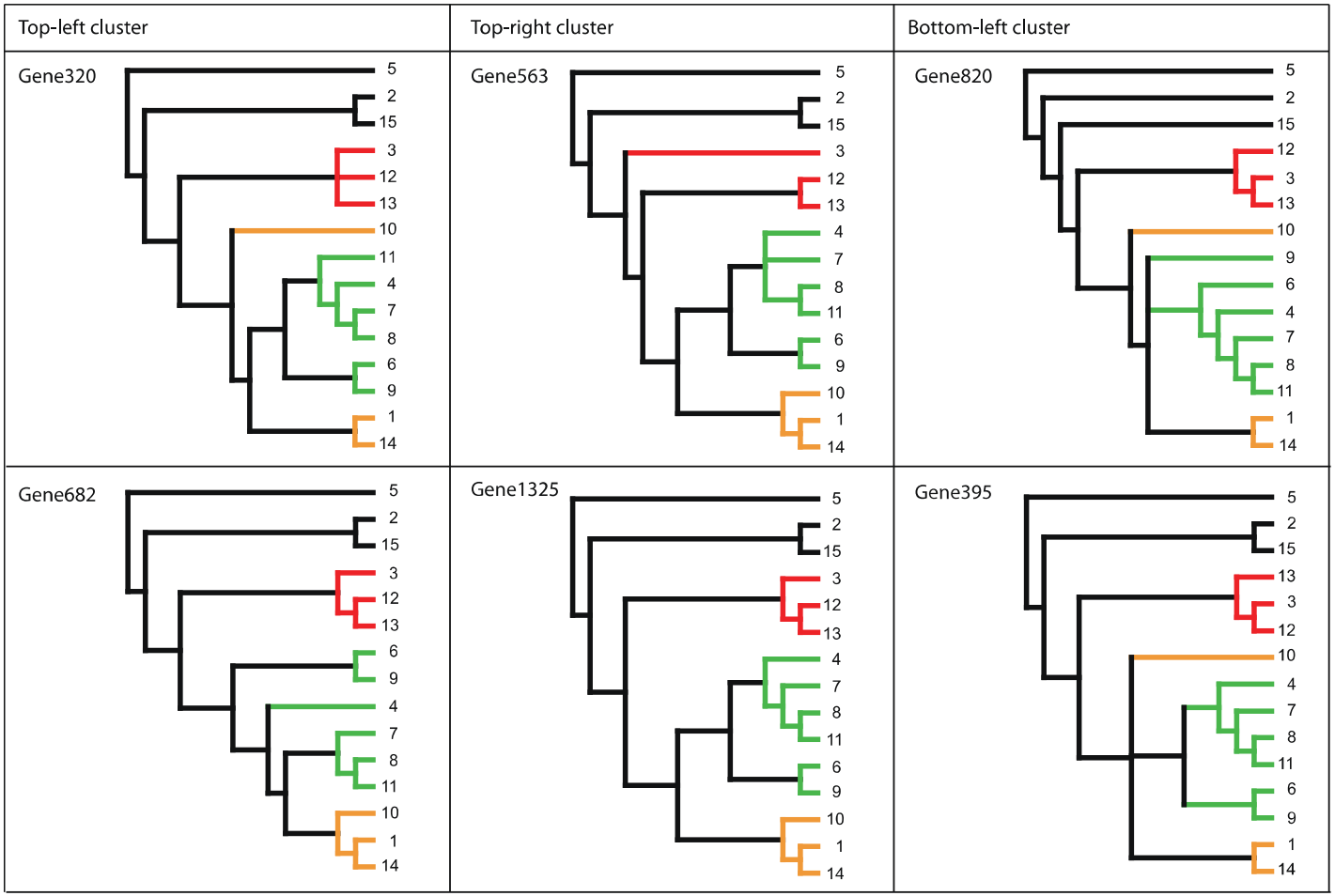



**Figure S2.3** Tree representatives from three groups based on the NMDS analysis. Species labels are as in figure 2.4.

# Chapter III

# Analyses of corallimorpharian transcriptomes provides new perspectives on the evolution of calcification in the Scleractinia (corals)

This chapter is submitted as:

Mei-Fang Lin, Aurelie Moya, Chaolun Allen Chen, Sylvain Forêt, David J. Miller. Analyses of corallimorpharian transcriptomes provides new perspectives on the evolution of calcification in the Scleractinia (corals). Genome Biology and Evolution.

**Analyses of corallimorpharian transcriptomes provides new perspectives on the evolution of calcification in the Scleractinia (corals)**

Mei-Fang Lin[1,2], Aurelie Moya[1,2], Chaolun Allen Chen[4,5], Sylvain Forêt[2,3], David J. Miller[1,2,*]

[1]Comparative Genomics Centre, Department of Molecular and Cell Biology, James Cook University, Townsville, QLD 4811, Australia

[2]ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD 4811, Australia

[3]Research School of Biology, Australian National University, Canberra, ACT 0200, Australia

[4]Biodiversity Research Centre, Academia Sinica, Nangang, Taipei 115, Taiwan

[5]Taiwan International Graduate Program (TIGP)-Biodiversity, Academia Sinica, Nangang, Taipei 115, Taiwan

*Corresponding author: david.miller@jcu.edu.au

Running title: Evolution of coral calcification

**Abstract**

Corallimorpharians (coral-like anemones) have a close phylogenetic relationship with scleractinians (corals) and can potentially provide novel perspectives on the evolution of biomineralization within the Hexacorallia. Surveying the transcriptomes of three representative corallimorpharians led to the identification of orthologs of some skeletal organic matrix proteins (SOMPs) that were previously considered to be restricted to corals; this is particularly significant given that surprisingly few of the proteins identified in the skeletal proteome are coral-specific. Whilst carbonic anhydrases (CAs) are ubiquitous proteins involved in $CO_2$ trafficking, both calcification and photo-symbiosis are assumed to place increased demands on the CA repertoire. These additional requirements have presumably driven the elaboration of complex CA repertoires that are typical of corals. Comparison of the CA inventories of corallimorpharians with those of corals indicates 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. Similar numbers of CAs, and a similar distribution across the various types, between the non-symbiotic corallimorpharian *Corynactis* and *Ricordea*, which normally hosts *Symbiodinium*, suggests that, whereas an expansion of the CA repertoire has been 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 non-symbiotic sea anemones is similar. This comparative study is consistent with

the idea that the evolution of calcification in corals required relatively few completely new genes, that are largely involved in the deposition process, but also the expansion of a specific type of carbonic anhydrases.

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 these 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), large-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 diverged. However, the close relationship between these Orders implies that corallimorpharians could be uniquely informative with respect to the evolution of the biomineralisation 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 (non-calcifying) 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 the massive calcification characteristic of corals. Although the sea anemone genome has provided some important insights into coral biology, the depth of the coral/sea anemone divergence (around 500 MY; Shinzato et al., 2011) limits the usefulness of such comparisons. The closer relationship between corals and corallimorpharians suggests that these latter may be more informative comparitors, 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 datasets (e.g. Moya et al., 2012) have been available for some time for corals.

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, but the Scleractinia are the only cnidarians to deposit skeletons composed of aragonite. Because the calcification trait 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, carbonic anhydrases (Jackson et al., 2007; Grasso et al., 2008). The non-conserved category of the calcification repertoire includes many of the genes whose products control the deposition of calcium carbonate to form the skeleton – for example, the

heterogeneous lustrin A/perlin proteins involved in mollusk calcification lack orthologs in other phyla.

Considerable progress has recently been made in characterizing the calcification repertoire of corals. By applying proteomic approaches, Ramos-Silva et al., (2013a) identified 36 SOMPs (skeletal organic matrix proteins) in the coral *Acropora millepora*. A similar study on *Stylophora pistillata*, another coral (Drake et al., 2013), implicated some of the same components, but misidentified some contaminants as SOMPs (Ramos-Silva et al., 2013b). Galaxin was the first SOMP to be identified in any coral (Fukuda et al., 2003); to date, four distinct galaxins have been identified in *Acropora* spp. (Reyes-Bermudez et al., 2009; Shinzato et al., 2011), two of which were amongst the SOMPs identified by Ramos-Silva et al., (2013a). Whilst galaxins were initially considered to be restricted to corals, related sequences are phylogenetically widespread – for example, Esgal1 from the squid *Euprymna scolopes* is involved in the establishment and maintenance of its bacterial symbiont *Vibrio fischeri* (Heath-Heckman et al., 2014). Indeed, most (28) of the 36 SOMPs identified in *Acropora* have homologs that are either widespread or are present in *Nematostella vectensis* or *Hydra* (Ramos-Silva et al., 2013a), non-calcifying cnidarians for which whole genome data are available.

Carbonic anhydrases (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 Reynold, 1999; Jackson et al., 2007; Moya et al., 2012), whereas some of the proteins that control the deposition process are taxon-specific. In symbiotic animals such as corals, carbonic anhydrases 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 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).

The involvement of solute carrier (SLC) proteins in the active transport of bicarbonate ions to the site of skeletogenesis has recently been suggested in the coral *S. pistillata*. (Zoccola et al., 2015). Zoccola and his collaborators (2015) characterized a SLC that is specifically expressed in the calicoblastic ectoderm, and suggested that this bicarbonate active transporter (BAT; SLC4γ) plays a key role in calcification by assisting the supply of inorganic carbonate to the site of calcification. 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 biomineralisation in the Scleractinia (Zoccola et al., 2015).

Although corallimorpharians lack skeletons, most (~28 in 58 valid species) of the tropical shallow-water species host the same photosynthetic symbionts as corals

(*Symbiodinium* spp.), as do two of the three species studied here (*Rhodactis* and *Ricordea*), hence their CA repertoires are of particular interest.

Protein predictions based on assembled genomes are available for the sea anemone *Nematostella* and the corals *Acropora digitifera* and *Stylophora pistillata*, and transcriptomic data have recently become available for several other anthozoan cnidarians (Moya et al., 2012; Polato et al., 2011; Traylor-Knowles et al., 2011; Shinzato et al., 2014; Lehnert et al., 2012). We recently reported (Lin et al., 2016) the assembly of large transcriptome datasets for three corallimorpharians, *Rhodactis indosinesis*, *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). A criterion applied to assess the completeness of cnidarian transcriptome assemblies was to search for close matches to the 1808 core cnidarian orthologs identified by comparing the gene predictions from *Acropora digitifera, Nematostella vectensis* and *Hydra magnipapillata* (Lin et al., 2016). Of these 1808 core orthologs, 1609 were detected in *A. millepora*, 1481 in *C. australis*, 1401 in *R. yuma* and 1261 in *R. indosinesis*.

*Searching for calcification-related genes*

Calcification-related genes, such as small cysteine-rich proteins (SCRiPs) (Sunagawa et al., 2009), galaxin (Fukuda et al., 2003; Reyes-Bermudez et al., 2009), SOMPs (Ramos-Silva et al., 2013a), carbonic anhydrases (Moya et al., 2012) and three taxonomically restricted genes (Moya et al., 2012) were searched against the corallimorpharian transcriptomes which assemblies were described in Lin et al., (submitted) with an E-value cutoff of $e \leq 10^{-5}$. 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 9 available anthozoan transcriptome datasets including 6 robust corals, 1 complex coral, 1 anemone and 1 octocoral as listed in Lin et al. (2016) as well as a recently released symbiotic anemone genome (Baumgarten et al., 2015) with the same cutoff threshold used above. An additional

BLAST search against NCBI non-redundant (nr) database (accessed on 15/10/2014) with E-value cutoff of $e \leq 10^{-6}$ was applied to see whether they are present in other organisms.

The presence and location of signal peptide cleavage sites in candidate amino acid sequences is predicted using SignalP v.4.1 (Petersen et al., 2011) (accessed on 13/12/2014). An additional tool, TargetP v.1.1 (Emanuelsson et al., 2000) (accessed on 26/03/2015), was used to predict the subcellular localization of carbonic anhydrase proteins. The InterProScan 5 platform (www.ebi.ac.uk/Tools/pfa/iprscan5; accessed on 16/01/2015) was used for functional classification of proteins and the presence of possible transmembrane domains investiged using TMHMM v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/; accessed on 14/01/2015). The compute pI/Mw tool from the ExPASy bioinformatics portal (accessed on 14/01/2015) was used to estimate the theoretical isoelectric point (pI) value for each galaxin gene.

*Phylogenetic methods*

Similarity between corallimorpharian and *A. millepora* galaxins was evaluated using BioEdit v7.0.5.3 (Hall 1999). Because the galaxin sequences recovered were diverse, as indicated in Reyes-Bermudez et al. (2009) and Moya et al. (2012), sequence saturation was evaluated at all three codon positions. The saturation test was carried out based on the transition and transversion substitutions *vs.* the Tamura-Nei (TN93) distance of three codon positions by DAMBE 5 (Xia 2013). Results indicated that galaxin sequences were saturated (substitution saturation) at all three codon

positions (Supp Figure S1), thus we did not proceed with galaxin phylogeny analysis.

The levels of nucleotide saturation were also investigated for the SOMP-1 and SOMP-2 genes. The coral and corallimorpharian SOMP-1 sequences displayed a high level of amino acid similarity (Supp Figure S2), therefore phylogenetic analyses were undertaken. ProtTest 3.4 (Darriba et al., 2011) selected JTT+G as the best-fitting model of protein sequence evolution. Maximum-Likelihood (ML) analysis was inferred by MEGA5 (Tamura et al., 2011) with 1000 bootstrap replicates.

For carbonic anhydrases, all sequences were trimmed to the conserved regions 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 analyses (BI) 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. Bayesian inference was analyzed with MrBayes v3.2.2 (Ronquist et al., 2012) for 4 chains, 2 million generations and first 2000 trees cut-off as burn-in.

Phylogeny of members of the solute carrier family 4 was carried out by following

processes. Since the sequences of some data are partial sequences, only those sites have coverage higher than 70% were used. 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). The phylogenetic trees inferred by ML conducted by PhyML 3.0 (Guindon et al., 2010) for 100 bootstraps and BI was analyzed by MrBayes v3.2.2 (Ronquist et al., 2012) for 4 chains, 2 million generation with first 2000 trees of burnin.

## Results

*Corallimorpharian counterparts of known skeletal organic matrix proteins (SOMPs)*

Thirty-six components of the SOM were recently identified in the coral *Acropora millepora* (Ramos-Silva et al 2013a). Most of them (28 of the 36) were also present in non-calcifying cnidarians (*Nematostella*, *Hydra* or both) while 8 proteins (SAP1, SAP2, SOMP-1, SOMP-2, SOMP-3, SOMP-4, SOMP-6, and cephalotoxin-like) were coral-specific (not found in anemones or other organisms), although the last of these had a surprising level of similarity to a mollusk protein (a gene found in *Sepia esculenta* salivary glands) (Ramos-Silva et al., 2013a).

As discussed above, proteins related to most (28 of the 36) components of the SOM identified in *Acropora millepora* were also present in non-calcifying cnidarians (*Nematostella*, *Hydra* or both; Ramos-Silva et al., 2013a), so it is not surprising that clear homologs of most these could also be identified in *Rhodactis* and other corallimorpharians. Using a cutoff of $e \leq 4 \times 10^{-7}$ , 21 of the 28 coral SOMPs shared with other non-calcifying cnidarians have corallimorpharian matches, 9 of these having *e*-value = 0 in BLASTP analyses (Supp Table S3.1). Levels of identity between corallimorpharian and coral orthologs were consistently higher than for coral-*Nematostella* comparisons (Ramos-Silva et al., 2013a).

The 8 SOM proteins identified by Ramos-Silva et al. as coral-specific (2013a) were searched against all three corallimorpharian transcriptomes, three sea anemones

(*Aiptasia* sp./*Exaiptasia* sp., *Anthopleura elegantissima* and *Nematostella vectensis*), and all of the available sequences data for other corals. One caveat here is that, as the available transcriptome assemblies are likely to be incomplete, presences are significant but apparent absences are less so.

SAP1, SAP2, cephalotoxin-like, and SOMP-6 had no significant counterparts in any of the databases searched ($e \leq 10^{-5}$). Clear homologs of SOMP4 were found in several other scleractinians (Supp Table S3.2), but not in other organisms, whereas a SOMP-3 homolog was detected in the sea anemones *Anthopleura* and *Aiptasia* as well as in other scleractinians (Table 3.1 and Supp table S3.2). SOMP-1 and SOMP-2 had apparent matches in the corallimorpharians ($e \leq 10^{-7}$, Table 3.1), but not in sea anemones.

The *A. millepora* SOMP-1 sequence identified by Ramos-Silva et al. (2013a) (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). Clear orthologs of SOMP-1 were found by BLASTP search in each of the three corallimorpharian species (Table 1 and Supp figure S2). An additional but incomplete SOMP-1 sequence was found in *Corynactis* (e-value $\leq 5 \times 10^{-8}$, Supp Table S3.2). This sequence is missing the N-terminal region, and matches well but differs significantly; the N-terminus of Cory_cds.comp32376 matches Cory_cds.comp95534 residue# the first from position #239. Note that the inclusion of the N-terminal region encoded in transcriptome cluster005198 led to the identification of a transmembrane region not

evident in the original analyses (Ramos-Silva et al., 2013a). Transmembrane regions were predicted in the *Acropora* sequence Cluster005198 as well as in for 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).

Orthologs of the *Acropora* SOMP-2 sequence were found in all three corallimorpharians. Searching the corallimorpharian data using the *Acropora millepora* SOMP-2 sequence resulted in matches in each of the species (Table 3.1, e-value $\leq 10^{-7}$). However, whereas in the case of SOMP-1 the corallimorpharian counterparts matches were of similar length to *Acropora* sequence, in the case of SOMP-2, 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* SOMP-2 is a cysteine-rich protein (10% of residues), with complex repeated patterns centered around a di-cysteine motif. 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.

Although the corallimorph and coral SOMP-1 sequences are clearly homologous and share some common features, the level of nucleotide saturation (Supp Figure S3.3) precludes meaningful phylogenetic analyses at the nucleotide level. Phylogenetic analysis at the amino acid level (Figure 3.1) groups the

corallimorpharian sequences together, to the exclusion of coral sequences.

The *Acropora* SOMP-2 sequence includes a predicted signal peptide, whereas the *Acropora* SOMP-1 does not (Ramos-Silva et al., 2013a), and this general pattern holds for most of the corallimorph and coral homologs of these (Supp Table S3.2). The absence of a signal peptide from SOMP-1 suggests that this protein may not be secreted. Alignment of the SOMP-1 predicted amino acid sequences indicates that corals and corallimorpharians share 281 sites (35.7 %) with similarity higher than 50%, and identified a highly conserved region between positions 242 to 394 in the alignment shown as supplementary Figure S3.2.

*Galaxin-related sequences*

Galaxin from *Galaxea fascicularis* was the first coral SOM protein to be identified (Fukuda et al., 2003). To date, five distinct galaxin-related sequences have been identified in *Acropora*; two "adult-type" galaxins (Reyes-Bermudez et al., 2009; Ramos-Silva et al., 2013a) and three divergent but related "galaxin-like" sequences (Reyes-Bermudez et al., 2009). Like SOMP-2, galaxins are cysteine-rich repetitive proteins, but whereas few di-cysteine motifs occur in SOMP-2 homologs, many more are present in galaxins. Each of the *Acropora* galaxins possesses an N-terminal signal peptide but whereas, after signal peptide cleavage, the adult-type galaxin proteins consist entirely of di-cysteine-rich repeat units, acidic domains precede the repetitive regions in the mature forms of both of the *Acropora millepora* "galaxin-like" proteins (Ramos-Silva et al., 2013a). Galaxin-related sequences have been reported from a range of other animals, but these typically have low sequence

similarity and resemble each other only in containing di-cysteine repeat motifs, and it has been suggested that galaxins *sensu stricto* may be restricted to corals (Reyes-Bermudez et al., 2009; Ramos-Silva et al., 2013).

As galaxin-related sequences have been reported from a range of non-calcifying organisms including the sea anemone *Nematostella* (EDO26732.1), it is not surprising that BLASTP searching using *Acropora* galaxins as reference sequences yielded apparent matches in two (*Corynactis* and *Ricordea*) of the three corallimorpharians. Seven sequences from *Corynactis* and three from *Ricordea* matched to *Acropora* galaxin or galaxin-like sequences (Table 3.2), but no significant hits were identified in *Rhodactis* (BLASTP cutoff $e \leq 10^{-5}$). The similarity between coral and corallimorpharian sequences largely depends on the di-cysteine motifs and repetitive structure of the proteins, hence it is difficult to interpret their evolutionary relationships. However, it is interesting to note the similarity in domain structure between two *Corynactis* sequences (Comp95728 and Comp63271) and the *Acropora* galaxin-like sequences; 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 suggests common origins i.e. that the galaxin and galaxin-like genes diverged prior to the coral-corallimorph split.

*Comparison of carbonic anhydrase repertoires*

Although corallimorpharians lack skeletons, most (~28 in 58 valid species) of the

tropical shallow-water species host the same photosynthetic symbionts as corals (*Symbiodinium* spp.), as do two of the three species studied here (*Rhodactis* and *Ricordea*), hence their CA repertoires are of particular interest.

Using a BLASTP cutoff 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 (Figure 3.2 and Supp Figure S3.4) indicate that the corallimorpharian carbonic anhydrase repertoires are considerably less complex than those of corals because few number of CAs have been identified. In particular, the secreted and membrane-associated CA types are far fewer in number in corallimorpharians – two in both *Ricordea yuma* and *Corynactis australensis*, and one in *Rhodactis indosinesis*, whereas nine have been identified in *A. millepora* and a similar number are present in *A. digitifera* (data not shown). 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 cnidarian gene set can be retrieved from the corallimorphs (Lin et al. 2016). 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 “CA-II-a” clade (labeled “Non-catalytic”) 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 *Aiptasia* (AIPGENE15469) sequences with which these corallimorpharian sequences cluster (Figure 3.2) are likewise predicted to be CARPs (Bertucci et al., 2013).

The clade of corallimorph sequences comprising CAryu2, CAcau4 and CArin1 (Figure 3.2) is likely to represent the mitochondrial 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 be due to its incompleteness. Trans-membrane regions are predicted at the C-termini on both CAryu6 and CAcau7, whereas the three members f the CAryu2, CAcau4, CArin1 clade each appear to lack trans-membrane regions, hence these two distinct clades of corallimorpharian CAs most likely represent distinct species associated with the mitochondrial membrane and matrix respectively.

*Bicarbonate transporters in corals and corallimopharians*

Searching the corallimorpharian datasets for SLC4 genes led to the identification of members of the SLC4α, β and δ types, but not the SLC4γ-type. Clear orthologs of the SLC4 β 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* (Figure 3.3). Whilst these results are consistent with a key role for SLC4γ in coral calcification, in the phylogenetic analyses summarized as Figure 3.3, 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.

*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 of the *Acropora* taxonomically restricted genes of particular interest are A036-B3, B036-D5, and C012-D9, as expression of these was suppressed under acute $CO_2$ stress (Moya et al., 2012), a condition known to repress calcification in corals (reviewed in Tambutté et al., 2011). 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 $e \leq 9 \times 10^{-31}$ and $e \leq 10^{-29}$ respectively. 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 orthologs of genes implicated in coral calcification. Several caveats apply in interpretation of the comparative data. First, although considerable bodies of corallimorpharian data are presented and the assembly statistics are good, these datasets are incomplete. Thus, presences are more significant than absences. Second, comprehensive genome and transcriptome data are as yet available only for two coral species; as these are both members of the genus *Acropora*, 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 orthologs of the skeletal organic matrix protein SOMP1 and B036-D5 in corallimorpharians means that very few of the genes known to be involved in calcification are actually unique to corals. Many of those genes that are unique to corals are cysteine-rich (SOMP-2, galaxins sensu stricto, SCRiPs) and most-likely have been recruited from structural ECM proteins (Bermudez-Reyes 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.

The transcriptome surveys clearly imply that the evolution of biomineralisation in the Scleractinia required 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 (labeled CAI in the large Figure 3.2), nine were detected in *A. millepora*. Although only a smaller number (four) could be identified in the *Pocillopora damicornis* transcriptome (Traynor-Knowles et al., 2011), this is likely to reflect the incomplete nature of the assembly (see also Lin et al., 2016), as preliminary analyses (data not shown) imply similar CA repertoire complexity in a number of other corals to that detected in *A. millepora*. Within the large clade of secreted and membrane-associated sequences, the branching pattern of coral sequences – distinct clades for *A. millepora* and *P. damicornis* - contradicted expectations. The most likely explanation for this branching pattern is that the sequences 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 Figure 3.2 imply that a similar expansion has occurred on the mollusk, *Lottia gigantea*, but 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 the non-symbiotic corallimorpharian *Corynactis* and *Ricordea*, which normally hosts *Symbiodinium*, suggests that, whereas an expansion of the CA repertoire has been 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 non-symbiotic sea anemones is similar. Conversely, on the basis of coral-sea anemone comparisons, we have previously 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 non-symbiotic corallimorpharians (this paper) and the symbiotic sea anemone *Aiptasia* (Baumgarten et al., 2015) this idea can now be more thoroughly investigated.

Are corallimorpharians simply corals that have lost their skeletons, as suggested by 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 the expansion of carbonic anhydrase repertoire, particularly the secreted and membrane-associated type, 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 the apparent concerted evolution of these genes (see above).

**Acknowledgement**

We are grateful to the funding support from the ARC Centre of Excellence for Coral Reef Studies to David Miller, and James Cook University Postgraduate Research Scholarship for Mei-Fang Lin.

**Table 3.1** SOM protein homologs detected in corallimorpharians and other anothozoans. The presence of homologs is indicated as ( ). E-value of the best hit against the *Acropora millepora* sequence is shown in parentheses. Contigs, protein structure and BLAST results are given in Table S3.2.

| Order | Corallimorpharia | | | Scleractinia (Complexa) | Scleractinia (Robusta) | | | | | | | Actiniaria | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gene (NCBI accession No.) | *Rhodactis indosinesis* | *Ricordea yuma* | *Corynactis australis* | *Porites australiensis* | *Fungia scutaria* | *Madracis auretenra* | *Montastraea cavernosa* | *Pocillopora damicornis* | *Platygyra carnosus* | *Pseudodiploria strigosa* | *Nematostella vectensis* | *Anthopleura elegantissima* | *Aiptasia* |
| SOMP-1 (B3EX00) | ✔ (3.00E-69) | ✔ (2.00E-63) | ✔ (4.00E-64) | ✔ (2.00E-67) | - | - | ✔ (7.00E-46) | - | - | - | - | - | - |
| SOMP-2 (B7WFQ1) | ✔ (3.00E-26) | ✔ (1.00E-07) | ✔ (1.00E-28) | ✔ (3.00E-67) | ✔ (4.00E-41) | ✔ (1.00E-99) | ✔ (5.00E-75) | ✔ (6.00E-38) | ✔ (1.00E-66) | ✔ (3.00E-67) | - | - | - |
| SOMP-3 (B8RJM0) | - | - | - | ✔ (6.00E-49) | - | ✔ (2.00E-31) | ✔ (2.00E-26) | ✔ (8.00E-13) | ✔ (4.00E-20) | ✔ (2.00E-09) | - | ✔ (9E-07) | ✔ (1E-07) |
| SOMP-4 (B8UU74) | - | - | - | - | - | ✔ (6.00E-83) | - | ✔ (4.00E-21) | - | - | - | - | - |

**Table 3.2** Galaxin-like genes in corallimorpharians. TM: transmembrane domain, SP: signal peptide, # di-Cys: number of di-cysteine.

| | | | Best blast hit | | | Feature | | | |
|---|---|---|---|---|---|---|---|---|---|
| Gene ID | Species | Sequence ID | *Acropora millepora* / NCBI accessio | E-value | Identity | Length (aa) | pI value | TM Helix | SP | # di-Cys |
| CauGalaxin1 | *Corynactis australis* | Cory cds.comp88726 c1 seq2\|m.27915 | Cluster013356/ADI50283 | 1.00E-35 | 36.99 | 258 | 9.18 | No | No | 17 |
| CauGalaxin2 | *Corynactis australis* | Cory_cds.comp3762_c0_seq1\|m.8827 | Cluster013356/ADI50283 | 1.00E-15 | 28.19 | 182 | 8.59 | No | No | 7 |
| CauGalaxin3 | *Corynactis australis* | Cory_cds.comp62533_c0_seq1\|m.12407 | Cluster013356/ADI50283 | 2.00E-16 | 32.85 | 191 | 9.12 | Yes | Yes | 9 |
| CauGalaxin4 | *Corynactis australis* | Cory_cds.comp94198_c0_seq1\|m.36335 | Cluster015317/ADI50283 | 5.00E-21 | 32.96 | 215 | 9 | No | Yes | 12 |
| CauGalaxin5 | *Corynactis australis* | Cory_cds.comp107187_c0_seq1\|m.997 | Cluster013356/ADI50283 | 1.00E-08 | 25.98 | 127 | 8.88 | No | No | 5 |
| CauGalaxin6 | *Corynactis australis* | Cory_cds.comp95728_c0_seq5\|m.39423 | Cluster015317/ADI50283 | 8.00E-12 | 44.09 | 235 | 4.06 | Yes | Yes | 5 |
| CauGalaxin7 | *Corynactis australis* | Cory_cds.comp63271_c0_seq1\|m.12563 | Cluster013356/ADI50283 | 2.00E-11 | 49.23 | 199 | 4.68 | No | No | 5 |
| RyuGalaxin1 | *Ricordea yuma* | Ricordea cds.comp19957 c0 seq1\|m.5654 | Cluster013356/ADI50283 | 2.00E-35 | 31.56 | 387 | 8.87 | No | Yes | 23 |
| RyuGalaxin2 | *Ricordea yuma* | Ricordea_cds.comp51202_c1_seq1\|m.10629 | Cluster013356/ADI50283 | 1.00E-24 | 31.01 | 362 | 9.19 | No | Yes | 18 |
| RyuGalaxin3 | *Ricordea yuma* | Ricordea_cds.comp75875_c0_seq1\|m.22052 | Cluster015317/ADI50283 | 1.00E-24 | 34.22 | 211 | 9.46 | No | No | 12 |

**Table 3.3** Candidate carbonic anhydrases detected in the corallimorpharian datasets. TM: transmembrane domain, SP: signal peptide, (M): mitochondrial matrix.

| Gene ID | Species | Sequences ID | Best hit Accession Number | E-value | Amino acid length | TM Helices | Signal peptide | Zinc Binding domains |
|---|---|---|---|---|---|---|---|---|
| CArin1 | *Rhodactis indosinesis* | Rhod_cds.comp92943_c1_seq3 | AAD32675 | 2.00E-82 | 272 | No | No | H1,H2,H3 |
| CArin2 | *Rhodactis indosinesis* | Rhod_cds.comp98814_c5_seq2 | ACJ64662 | 4.00E-73 | 331 | Yes | Yes | H1,H2,H3 |
| CArin3 | *Rhodactis indosinesis* | Rhod_cds.comp66236_c0_seq1 | ACJ64663 | 3.00E-117 | 265 | No | No | H1,H2,H3 |
| CArin4 | *Rhodactis indosinesis* | 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 |

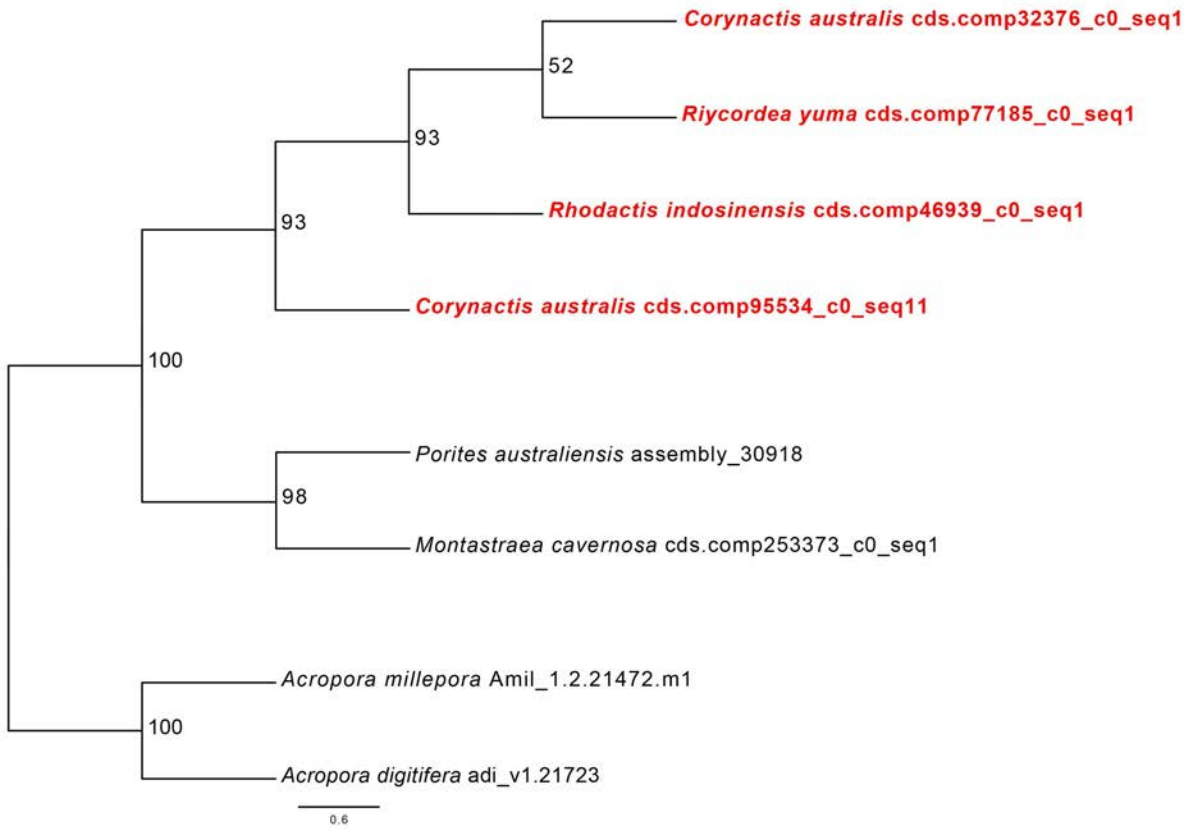



**Figure 3.1** Phylogeny of SOMP-1 amongst the cnidarians. The value indicates the supporting number of bootstraps in Maximum Likelihood analysis. The corallimorpharian sequences are highlighted in red.

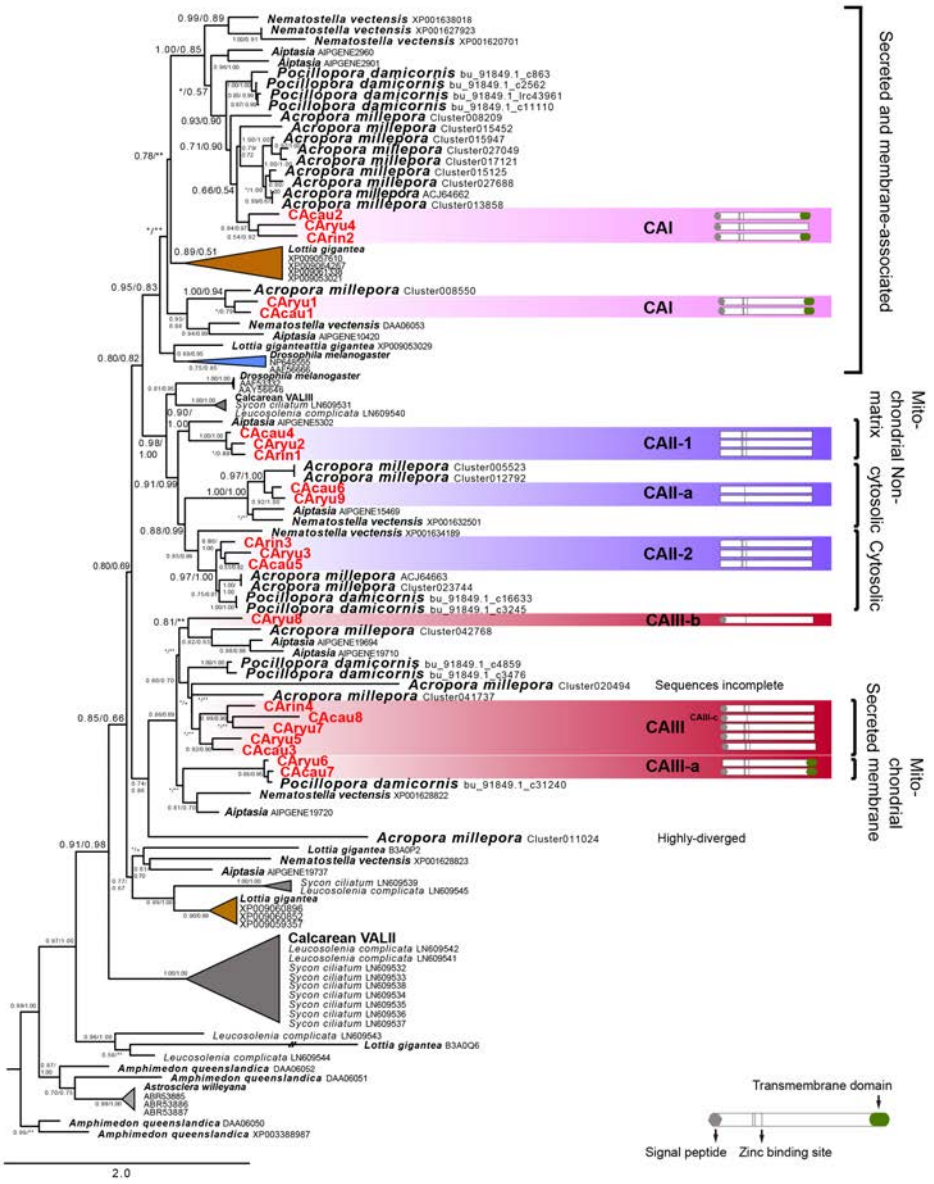



**Figure 3.2** Phylogenetic analysis of carbonic anhydrase sequences inferred from Maximum Likelihood (ML) and Bayesian Inference (BI) analyses. Both ML and BI analyses recovered nearly identical tree topology. The ML aLRT branch support values and BI posterior probabilities are indicated as ML/BI in the tree. * indicates the aLRT value < 0.5, ** indicates the presence of polytomies in the Bayesian phylogeny, + indicates discrepancies between ML and BI trees. The corallimorpharian sequences are highlighted in red.

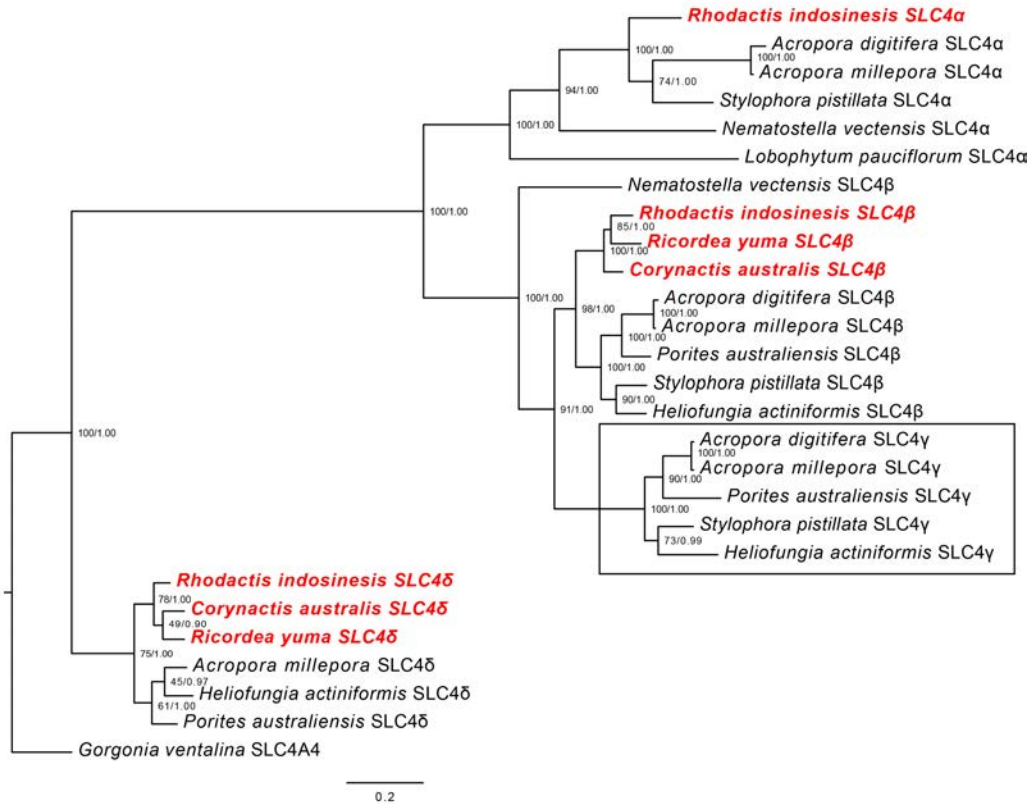



**Figure 3.3** The phylogenetic tree of cnidarian SLC4 inferred from Maximum likelihood (ML) and Bayesian analyses (BI). The bootstrap values of ML and posterior probabilities of BI are indicated as ML/BI. Corrallimorpharian sequences are in red. The group of SLC4γ was marked in square.

**Supporting material**

Table S3.1 SOMPs genes (present in two *Acropora*, *Nematostella* and *Hydra*) candidates in corallimorpharians. SOMPs (SAP1, SAP2, SOMP1, SOMP2, SOMP4, SOMP6, Cephalotoxin-like) that present in *Acropora* only are not shown here. Genes of galaxin and carbonic anhydrates were discussed separately.

*The identification of each protein group was followed the description in Ramos-Silva et al. (2013). dcp: domain containing protein.

| Groups* | SOMPs present in corals and noncalcifying cnidarians | Species | Sequences ID | NCBI Accession Number | E-value | No. TM Helices | No. Cleavages |
|---|---|---|---|---|---|---|---|
| Enzymes | Hephaestin | *Corynactis australis* | Cory_cds.comp93706_c0_seq1lm.35425 | B3EWZ9 | 0 | 0 | 43 |
| | | *Corynactis australis* | Cory_cds.comp53721_c0_seq1lm.10906 | B3EWZ9 | 0 | 1(N) | 82 |
| | | *Corynactis australis* | Cory_cds.comp78841_c0_seq1lm.18548 | B3EWZ9 | 2.00E-93 | 0 | 13 |
| | | *Corynactis australis* | Cory_cds.comp55762_c0_seq1lm.11285 | B3EWZ9 | 5.00E-58 | 0 | 10 |
| | | *Corynactis australis* | Cory_cds.comp145295_c0_seq1lm.5519 | B3EWZ9 | 9.00E-29 | 0 | 9 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp98964_c1_seq9lm.33038 | B3EWZ9 | 0 | 0 | 67 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp93492_c0_seq2lm.22165 | B3EWZ9 | 0 | 0 | 76 |
| | | *Ricordea yuma* | Riy_cds.comp73702_c0_seq4lm.19905 | B3EWZ9 | 0 | 1(C) | 66 |
| | | *Ricordea yuma* | Riy_cds.comp136850_c0_seq1lm.4362 | B3EWZ9 | 6.00E-37 | 0 | 10 |
| | | *Ricordea yuma* | Riy_cds.comp70698_c0_seq1lm.17616 | B3EWZ9 | 7.00E-50 | 0 | 11 |
| | CUB_and_peptidase_domain-containing_protein_1 | *Corynactis australis* | Cory_cds.comp95946_c0_seq1lm.39897 | B8V7S0 | 4.00E-82 | 0 | 91 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp102548_c0_seq1lm.5991 | B8V7S0 | 2.00E-79 | 0 | 26 |
| Acidic proteins | SAARP1 | *Corynactis australis* | Cory_cds.comp88349_c0_seq5lm.27414 | B3EWY6 | 4.00E-31 | NA | NA |
| | | *Rhodactis indosinesis* | Rhod_cds.comp74830_c0_seq1lm.13832 | B3EWY6 | 3.00E-35 | NA | NA |
| | SAARP2 | *Corynactis australis* | Cory_cds.comp42625_c0_seq2lm.9474 | B3EWY8 | 3.00E-31 | 1(C) | 15 |
| | | *Corynactis australis* | Cory_cds.comp30507_c0_seq1lm.7930 | B3EWY8 | 8.00E-37 | 0 | 16 |
| | | *Ricordea yuma* | Riy_cds.comp81594_c0_seq1lm.29973 | B3EWY8 | 6.00E-27 | 0 | 14 |
| Extracellular matrix/ Cell adhesion | Mucin-like | *Corynactis australis* | Cory_cds.comp97042_c0_seq3lm.42471 | B3EWY9 | 0.00E+00 | 1(C) | 94 |
| | | *Corynactis australis* | Cory_cds.comp97408_c0_seq1lm.43569 | B3EWY9 | 0.00E+00 | 1(C) | 139 |
| | | *Corynactis australis* | Cory_cds.comp87884_c0_seq2lm.26780 | B3EWY9 | 0.00E+00 | 1(C) | 77 |
| | | *Corynactis australis* | Cory_cds.comp69847_c0_seq1lm.14437 | B3EWY9 | 2.00E-80 | 1(C) | 30 |
| | | *Corynactis australis* | Cory_cds.comp92505_c0_seq1lm.33434 | B3EWY9 | 2.00E-76 | 0 | 36 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp97279_c0_seq2lm.29000 | B3EWY9 | 0.00E+00 | 3(C) | 102 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp101244_c2_seq1lm.3408 | B3EWY9 | 0.00E+00 | 0 | 40 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp101988_c0_seq2lm.5586 | B3EWY9 | 0.00E+00 | 1(C) | 184 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp101807_c0_seq8lm.5008 | B3EWY9 | 1.00E-154 | 1(C) | 41 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp96741_c1_seq1lm.27869 | B3EWY9 | 1.00E-106 | 0 | 55 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp96741_c0_seq1lm.27868 | B3EWY9 | 4.00E-84 | 1(C) | 20 |
| | | *Ricordea yuma* | Riy_cds.comp80209_c0_seq2lm.28751 | B3EWY9 | 0.00E+00 | 1(C) | 72 |
| | | *Ricordea yuma* | Riy_cds.comp69801_c0_seq1lm.17040 | B3EWY9 | 0.00E+00 | 1(C) | 60 |
| | | *Ricordea yuma* | Riy_cds.comp45544_c0_seq1lm.9402 | B3EWY9 | 2.00E-66 | 0 | 9 |
| Extracellular matrix/ Cell adhesion | Coadhesin | *Corynactis australis* | Cory_cds.comp94649_c0_seq2lm.37258 | B3EWZ3 | 2.00E-121 | 0 | 43 |
| | | *Corynactis australis* | Cory_cds.comp90368_c2_seq2lm.30169 | B3EWZ3 | 0.00E+00 | 1(C) | 61 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp101126_c0_seq2lm.3069 | B3EWZ3 | 0.00E+00 | 1(C) | 85 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp101399_c0_seq1lm.3864 | B3EWZ3 | 7.00E-63 | 0 | 26 |
| | | *Ricordea yuma* | Riy_cds.comp77785_c1_seq1lm.24567 | B3EWZ3 | 8.00E-74 | 1(N) | 23 |
| | | *Ricordea yuma* | Riy_cds.comp133506_c0_seq1lm.4125 | B3EWZ3 | 2.00E-49 | 0 | 9 |
| | | *Ricordea yuma* | Riy_cds.comp71898_c0_seq1lm.18465 | B3EWZ3 | 2.00E-41 | 0 | 20 |
| Uncharacterized proteins | SOMP8 | *Corynactis australis* | Cory_cds.comp52096_c0_seq1lm.10596 | B3EWZ2 | 1.00E-13 | NA | NA |
| | | *Corynactis australis* | Cory_cds.comp88704_c0_seq1lm.27881 | B3EWZ2 | 6.00E-09 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp92829_c0_seq1lm.31176 | B3EWZ2 | 4.00E-07 | NA | NA |
| Extracellular matrix/ Cell adhesion | MAM and LDL-receptor 1 | *Corynactis australis* | Cory_cds.comp95654_c0_seq1lm.39252 | B3EWZ5 | 0.00E+00 | NA | NA |
| | | *Corynactis australis* | Cory_cds.comp75667_c0_seq1lm.16827 | B3EWZ5 | 3.00E-146 | NA | NA |
| | | *Corynactis australis* | Cory_cds.comp51690_c0_seq2lm.10527 | B3EWZ5 | 4.00E-114 | NA | NA |
| | | *Corynactis australis* | Cory_cds.comp4985_c0_seq1lm.10294 | B3EWZ5 | 9.00E-81 | NA | NA |
| | | *Corynactis australis* | Cory_cds.comp77728_c0_seq1lm.17912 | B3EWZ5 | 1.00E-87 | NA | NA |
| | | *Rhodactis indosinesis* | Rhod_cds.comp94084_c0_seq2lm.22996 | B3EWZ5 | 6.00E-141 | NA | NA |
| | | *Rhodactis indosinesis* | Rhod_cds.comp95252_c0_seq72lm.24967 | B3EWZ5 | 7.00E-39 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp6460_c0_seq1lm.14464 | B3EWZ5 | 1.00E-175 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp107067_c0_seq1lm.860 | B3EWZ5 | 5.00E-150 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp30412_c0_seq1lm.6866 | B3EWZ5 | 2.00E-83 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp128880_c0_seq1lm.3745 | B3EWZ5 | 2.00E-78 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp134975_c0_seq1lm.4233 | B3EWZ5 | 1.00E-69 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp108656_c0_seq1lm.1074 | B3EWZ5 | 8.00E-66 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp114855_c0_seq1lm.2016 | B3EWZ5 | 2.00E-61 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp56452_c0_seq1lm.11793 | B3EWZ5 | 3.00E-24 | NA | NA |
| | MAM and LDL-receptor 2 | *Corynactis australis* | Cory_cds.comp95423_c0_seq1lm.38735 | B3EWZ6 | 0.00E+00 | NA | NA |
| | | *Corynactis australis* | Cory_cds.comp86257_c0_seq1lm.24852 | B3EWZ6 | 0.00E+00 | NA | NA |
| | | *Corynactis australis* | Cory_cds.comp5329_c0_seq1lm.10825 | B3EWZ6 | 3.00E-144 | NA | NA |
| | | *Corynactis australis* | Cory_cds.comp91517_c0_seq3lm.31883 | B3EWZ6 | 4.00E-112 | NA | NA |
| | | *Corynactis australis* | Cory_cds.comp42799_c0_seq1lm.9508 | B3EWZ6 | 5.00E-80 | NA | NA |
| | | *Rhodactis indosinesis* | Rhod_cds.comp101967_c0_seq2lm.5503 | B3EWZ6 | 0.00E+00 | NA | NA |
| | | *Rhodactis indosinesis* | Rhod_cds.comp92604_c0_seq4lm.21141 | B3EWZ6 | 5.00E-103 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp19364_c0_seq1lm.5575 | B3EWZ6 | 0.00E+00 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp124055_c0_seq1lm.3218 | B3EWZ6 | 7.00E-71 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp122027_c0_seq1lm.2981 | B3EWZ6 | 2.00E-55 | NA | NA |
| Extracellular matrix/ Cell adhesion | Ectin | *Ricordea yuma* | Riy_cds.comp21629_c0_seq1lm.5839 | B3EWZ8 | 1.00E-28 | NA | NA |
| Extracellular matrix/ Cell adhesion | MAM and fibronectin dcps | *Corynactis australis* | Cory_cds.comp95861_c0_seq32lm.39714 | B3EX02 | 5.00E-51 | NA | NA |
| | | *Corynactis australis* | Cory_cds.comp28237_c0_seq1lm.7637 | B3EX02 | 1.00E-31 | NA | NA |
| | | *Corynactis australis* | Cory_cds.comp75489_c0_seq1lm.16741 | B3EX02 | 4.00E-18 | NA | NA |
| | | *Rhodactis indosinesis* | Rhod_cds.comp97180_c1_seq1lm.28780 | B3EX02 | 6.00E-47 | NA | NA |
| Extracellular matrix/ Cell adhesion | PKD1-related protein | *Corynactis australis* | Cory_cds.comp97301_c0_seq1lm.43192 | B8UU59 | 0.00E+00 | 11 | 88 |
| | | *Corynactis australis* | Cory_cds.comp95142_c0_seq1lm.38183 | B8UU59 | 5.00E-53 | 0 | 13 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp99245_c3_seq2lm.33791 | B8UU59 | 0.00E+00 | 12 | 105 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp103352_c0_seq1lm.6116 | B8UU59 | 1.00E-143 | 0 | 46 |
| | | *Ricordea yuma* | Riy_cds.comp79701_c0_seq1lm.27778 | B8UU59 | 0.00E+00 | 12 | 89 |
| | | *Ricordea yuma* | Riy_cds.comp62856_c0_seq2lm.13701 | B8UU59 | 5.00E-62 | 0 | 21 |
| | | *Ricordea yuma* | Riy_cds.comp32272_c0_seq2lm.7185 | B8UU59 | 1.00E-55 | 0 | 7 |
| Extracellular matrix/ Cell adhesion | ZP domain-containing protein | *Corynactis australis* | Cory_cds.comp88110_c0_seq1lm.27083 | G8HTB6 | 9.00E-155 | 1(C) | 21 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp93760_c1_seq1lm.22539 | G8HTB6 | 4.00E-74 | 1(C) | 14 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp93760_c0_seq1lm.22538 | G8HTB6 | 1.00E-62 | 0 | 9 |
| | | *Ricordea yuma* | Riy_cds.comp76140_c0_seq2lm.22363 | G8HTB6 | 4.00E-142 | 1(C) | 22 |
| Extracellular matrix/ Cell adhesion | EGF and laminin G dcp | *Corynactis australis* | Cory_cds.comp86524_c0_seq1lm.25146 | B8UU78 | 0.00E+00 | 1(C) | 62 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp98800_c0_seq3lm.32578 | B8UU78 | 0.00E+00 | 1(C) | 71 |
| | | *Ricordea yuma* | Riy_cds.comp80367_c0_seq4lm.29037 | B8UU78 | 0.00E+00 | 1(C) | 66 |
| Extracellular matrix/ Cell adhesion | Protocadherin-like | *Corynactis australis* | Cory_cds.comp84737_c0_seq1lm.23264 | B8V7Q1 | 0.00E+00 | 0 | 197 |
| | | *Corynactis australis* | Cory_cds.comp94776_c0_seq1lm.37504 | B8V7Q1 | 0.00E+00 | 0 | 105 |
| | | *Corynactis australis* | Cory_cds.comp96816_c0_seq2lm.41884 | AGC70195 | 0.00E+00 | 2(N,C) | 146 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp87389_c0_seq1lm.17437 | B8V7Q1 | 0.00E+00 | 1(C) | 215 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp101677_c1_seq1lm.4616 | AGC70195 | 0.00E+00 | 0 | 130 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp99216_c1_seq12lm.33722 | B8V7Q1 | 3.00E-141 | 0 | 52 |
| | | *Ricordea yuma* | Riy_cds.comp38446_c0_seq1lm.8319 | B8V7Q1 | 0.00E+00 | 0 | 123 |
| | | *Ricordea yuma* | Riy_cds.comp42751_c0_seq1lm.9004 | B8V7Q1 | 0.00E+00 | 0 | 30 |
| | | *Ricordea yuma* | Riy_cds.comp102679_c0_seq1lm.300 | B8V7Q1 | 0.00E+00 | 0 | 34 |
| | | *Ricordea yuma* | Riy_cds.comp80692_c0_seq1lm.29596 | B8V7Q1 | 7.00E-153 | 1(C) | 20 |
| | | *Ricordea yuma* | Riy_cds.comp74313_c0_seq1lm.20443 | AGC70195 | 0.00E+00 | 1(C) | 70 |
| Extracellular matrix/ Cell adhesion | Collagen | *Corynactis australis* | Cory_cds.comp90994_c0_seq1lm.31102 | B8V7R6 | 0.00E+00 | NA | NA |
| | | *Rhodactis indosinesis* | Rhod_cds.comp102025_c0_seq1lm.5658 | B8V7R6 | 0.00E+00 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp70659_c0_seq1lm.17587 | B8V7R6 | 0.00E+00 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp18548_c0_seq1lm.5267 | B8V7R6 | 2.00E-77 | NA | NA |
| Uncharacterized proteins | SOMP5 | *Corynactis australis* | Cory_cds.comp91462_c0_seq8lm.31802 | B8VIU6 | 4.00E-56 | NA | NA |
| | | *Corynactis australis* | Cory_cds.comp87432_c0_seq8lm.26226 | B8VIU6 | 5.00E-32 | NA | NA |
| | | *Corynactis australis* | Cory_cds.comp84150_c1_seq1lm.22677 | B8VIU6 | 3.00E-27 | NA | NA |
| | | *Corynactis australis* | Cory_cds.comp41428_c0_seq1lm.9316 | B8VIU6 | 7.00E-15 | NA | NA |
| | | *Rhodactis indosinesis* | Rhod_cds.comp101521_c0_seq1lm.4196 | B8VIU6 | 5.00E-63 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp80304_c0_seq3lm.28915 | B8VIU6 | 2.00E-61 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp104424_c0_seq1lm.483 | B8VIU6 | 2.00E-22 | NA | NA |
| Extracellular matrix/ Cell adhesion | Neuroglian-like | *Corynactis australis* | Cory_cds.comp94627_c0_seq1lm.37209 | B8VIW9 | 0.00E+00 | 1(C) | 44 |
| | | *Corynactis australis* | Cory_cds.comp94627_c0_seq1lm.37210 | B8VIW9 | 1.00E-113 | 0 | 35 |
| | | *Corynactis australis* | Cory_cds.comp96391_c0_seq1lm.40870 | B8VIW9 | 3.00E-63 | 1(C) | 39 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp29308_c0_seq1lm.9447 | B8VIW9 | 0.00E+00 | 1(C) | 77 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp101049_c0_seq61lm.2876 | B8VIW9 | 2.00E-56 | 0 | 27 |
| | | *Rhodactis indosinesis* | Rhod_cds.comp101457_c1_seq40lm.4034 | B8VIW9 | 3.00E-41 | 0 | 68 |
| | | *Ricordea yuma* | Riy_cds.comp73158_c0_seq1lm.19405 | B8VIW9 | 0.00E+00 | 1(C) | 76 |
| | | *Ricordea yuma* | Riy_cds.comp56926_c0_seq1lm.11887 | B8VIW9 | 4.00E-58 | 1(C) | 40 |
| Uncharacterized proteins | SOMP7 | *Corynactis australis* | Cory_cds.comp91615_c0_seq3lm.32051 | B8WI85 | 2.00E-143 | NA | NA |
| | | *Rhodactis indosinesis* | Rhod_cds.comp96387_c0_seq2lm.27114 | B8WI85 | 1.00E-46 | NA | NA |
| | | *Rhodactis indosinesis* | Rhod_cds.comp87301_c0_seq1lm.17396 | B8WI85 | 1.00E-54 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp77961_c0_seq1lm.24841 | B8WI85 | 9.00E-136 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp107956_c0_seq1lm.980 | B8WI85 | 3.00E-25 | NA | NA |
| | | *Ricordea yuma* | Riy_cds.comp132166_c0_seq1lm.4034 | B8WI85 | 2.00E-12 | NA | NA |

Table S3.2 Using SOMP sequences identified in Ramos-Silva et al. (2013a), the table shown the novel SOMPs in corallimorpharians and other anthozoans and their best blast hit against NCBI nr databse. TM: transmembrane.

| SOMPs present in corals and corallimorpharians | Species | Sequences ID | Blast Hit NCBI Accession Number | E-value | Signal peptide | TM domain |
|---|---|---|---|---|---|---|
| SOMP1 | *Corynactis australis* | Cory_cds.comp32376_c0_seq1lm.8145 | B3EX00 | 5.00E-11 | No | No |
| | *Corynactis australis* | Cory_cds.comp95534_c0_seq11lm.38993 | B3EX00 | 5.00E-08 | No | Yes |
| | *Rhodactis indosinesis* | Rhod_cds.comp46939_c0_seq1lm.10823 | B3EX00 | 3.00E-14 | No | Yes |
| | *Ricordea yuma* | Riy_cds.comp77185_c0_seq1lm.23716 | B3EX00 | 1.00E-09 | No | Yes |
| | *Porites australiensis* | Pau-assembly_30918 | B3EX00 | 8.00E-42 | No | Yes |
| | *Montastraea cavernosa* | Monta_cds.comp253373_c0_seq1lm.32600 | B3EX00 | 3.00E-13 | No | No |
| SOMP2 | *Corynactis australis* | Cory_cds.comp32915_c0_seq1lm.8277 | B7WFQ1 | 3.00E-25 | Yes | No |
| | *Corynactis australis* | Cory_cds.comp77745_c0_seq3lm.17921 | B7WFQ1 | 2.00E-15 | Yes | No |
| | *Corynactis australis* | Cory_cds.comp78158_c0_seq1lm.18153 | B7WFQ1 | 4.00E-18 | Yes | No |
| | *Corynactis australis* | Cory_cds.comp83818_c0_seq1lm.22387 | B7WFQ1 | 4.00E-07 | Yes | Yes |
| | *Corynactis australis* | Cory_cds.comp89749_c0_seq5lm.29283 | B7WFQ1 | 3.00E-06 | Yes | No |
| | *Corynactis australis* | Cory_cds.comp74546_c0_seq2lm.16265 | B7WFQ1 | 3.00E-05 | No | No |
| | *Rhodactis indosinesis* | Rhod_cds.comp62430_c0_seq1lm.12202 | B7WFQ1 | 2.00E-23 | Yes | No |
| | *Rhodactis indosinesis* | Rhod_cds.comp91789_c1_seq3lm.20330 | B7WFQ1 | 6.00E-19 | Yes | Yes |
| | *Rhodactis indosinesis* | Rhod_cds.comp90813_c1_seq4lm.19542 | B7WFQ1 | 4.00E-05 | Yes | Yes |
| | *Ricordea yuma* | Riy_cds.comp51189_c0_seq1lm.10625 | B7WFQ1 | 3.00E-05 | Yes | Yes |
| | *Fungia scutaria* | Fung_cds.comp36970_c0_seq2lm.33530 | B7WFQ1 | 4.00E-38 | Yes | No |
| | *Fungia scutaria* | Fung_cds.comp36970_c0_seq1lm.31687 | B7WFQ1 | 5.00E-27 | Yes | Yes |
| | *Fungia scutaria* | Fung_cds.comp315761_c0_seq1lm.5654 | B7WFQ1 | 3.00E-11 | No | No |
| | *Madracis auretenra* | Madra_cds.comp124813_c0_seq1lm.32588 | B7WFQ1 | 3.00E-57 | Yes | No |
| | *Madracis auretenra* | Madra_cds.comp985927_c0_seq1lm.4565 | B7WFQ1 | 4.00E-25 | No | No |
| | *Madracis auretenra* | Madra_cds.comp991849_c0_seq1lm.5826 | B7WFQ1 | 5.00E-26 | No | No |
| | *Madracis auretenra* | Madra_cds.comp9566_c0_seq5lm.15974 | B7WFQ1 | 1.00E-33 | Yes | No |
| | *Montastraea cavernosa* | Monta_cds.comp75555_c0_seq1lm.14108 | B7WFQ1 | 2.00E-89 | No | Yes |
| | *Montastraea cavernosa* | Monta_cds.comp23009_c0_seq1lm.67127 | B7WFQ1 | 7.00E-47 | No | Yes |
| | *Montastraea cavernosa* | Monta_cds.comp198620_c0_seq1lm.28837 | B7WFQ1 | 6.00E-25 | Yes | No |
| | *Montastraea cavernosa* | Monta_cds.comp67183_c0_seq2lm.8213 | B7WFQ1 | 5.00E-06 | Yes | No |
| | *Montastraea cavernosa* | Monta_cds.comp5764_c0_seq1lm.49088 | B7WFQ1 | 1.00E-06 | Yes | No |
| | *Pocillopora damicornis* | Pdam_bu_91849.1_c33874 | B7WFQ1 | 1.00E-34 | No | No |
| | *Pocillopora damicornis* | Pdam_bu_91849.1_c44462 | B7WFQ1 | 6.00E-14 | No | No |
| | *Pocillopora damicornis* | Pdam_bu_91849.1_c44398 | B7WFQ1 | 4.00E-14 | No | Yes |
| | *Platygyra carnosus* | Platy_CL5629.Contig1_Mix_Normal_N | B7WFQ1 | 3.00E-63 | No | No |
| | *Platygyra carnosus* | Platy_Unigene983_Mix_Normal_N | B7WFQ1 | 2.00E-56 | No | No |
| | *Platygyra carnosus* | Platy_CL518.Contig1_Mix_Normal_N | B7WFQ1 | 2.00E-30 | No | No |
| | *Porites australiensis* | Pau-assembly_38125 | B7WFQ1 | 2.00E-64 | No | Yes |
| | *Pseudodiploria strigosa* | Pseudo_cds.comp49762_c0_seq1lm.16184 | B7WFQ1 | 4.00E-70 | No | No |
| | *Pseudodiploria strigosa* | Pseudo_cds.comp24617_c0_seq3lm.10806 | B7WFQ1 | 4.00E-53 | Yes | No |
| | *Pseudodiploria strigosa* | Pseudo_cds.comp273065_c0_seq1lm.26012 | B7WFQ1 | 1.00E-39 | Yes | No |
| | *Pseudodiploria strigosa* | Pseudo_cds.comp24617_c0_seq1lm.11453 | B7WFQ1 | 3.00E-29 | Yes | No |
| | *Pseudodiploria strigosa* | Pseudo_cds.comp385574_c0_seq1lm.32061 | B7WFQ1 | 4.00E-22 | No | No |
| **SOMPs present in corals only** | **Species** | **Sequences ID** | **Blast Hit NCBI Accession Number** | **E-value** | **Signal peptide** | **TM domain** |
| SOMP3 | *Anthopleura elegantissima* | Aele_cds.comp155444_c0_seq1lm.37210 | B8RJM0 | 8.00E-04 | NA | NA |
| | *Aiptasia* | AIPGENE15989 | B8RJM0 | 6.00E-08 | NA | NA |
| | *Madracis auretenra* | Madra_cds.comp44921_c0_seq2lm.23782 | B8RJM0 | 1.00E-50 | NA | NA |
| | *Madracis auretenra* | Madra_cds.comp276735_c0_seq1lm.46969 | B8RJM0 | 8.00E-10 | NA | NA |
| | *Madracis auretenra* | Madra_cds.comp44921_c0_seq1lm.26686 | B8RJM0 | 8.00E-53 | NA | NA |
| | *Montastraea cavernosa* | Monta_cds.comp7530_c0_seq11lm.51052 | B8RJM0 | 3.00E-16 | NA | NA |
| | *Montastraea cavernosa* | Monta_cds.comp7530_c0_seq14lm.47919 | B8RJM0 | 3.00E-12 | NA | NA |
| | *Montastraea cavernosa* | Monta_cds.comp7530_c0_seq1lm.49045 | B8RJM0 | 7.00E-37 | NA | NA |
| | *Pocillopora damicornis* | Pdam_bu_91849.1_c23421 | B8RJM0 | 3.00E-10 | NA | NA |
| | *Platygyra carnosus* | Platy_Unigene15970_Mix_Normal_N | B8RJM0 | 2.00E-17 | NA | NA |
| | *Platygyra carnosus* | Platy_Unigene56832_Mix_Normal_N | B8RJM0 | 1.00E-17 | NA | NA |
| | *Porites australiensis* | Pau-assembly_27669 | B8RJM0 | 5.00E-73 | NA | NA |
| | *Pseudodiploria strigosa* | Pseudo_cds.comp386429_c0_seq1lm.31936 | B8RJM0 | 1.00E-11 | NA | NA |
| SOMP4 | *Madracis auretenra* | Madra_cds.comp105232_c1_seq1lm.30550 | B8UU74 | 5.00E-17 | NA | NA |
| | *Pocillopora damicornis* | Pdam_bu_91849.1_c70018 | B8UU74 | 1.00E-07 | NA | NA |

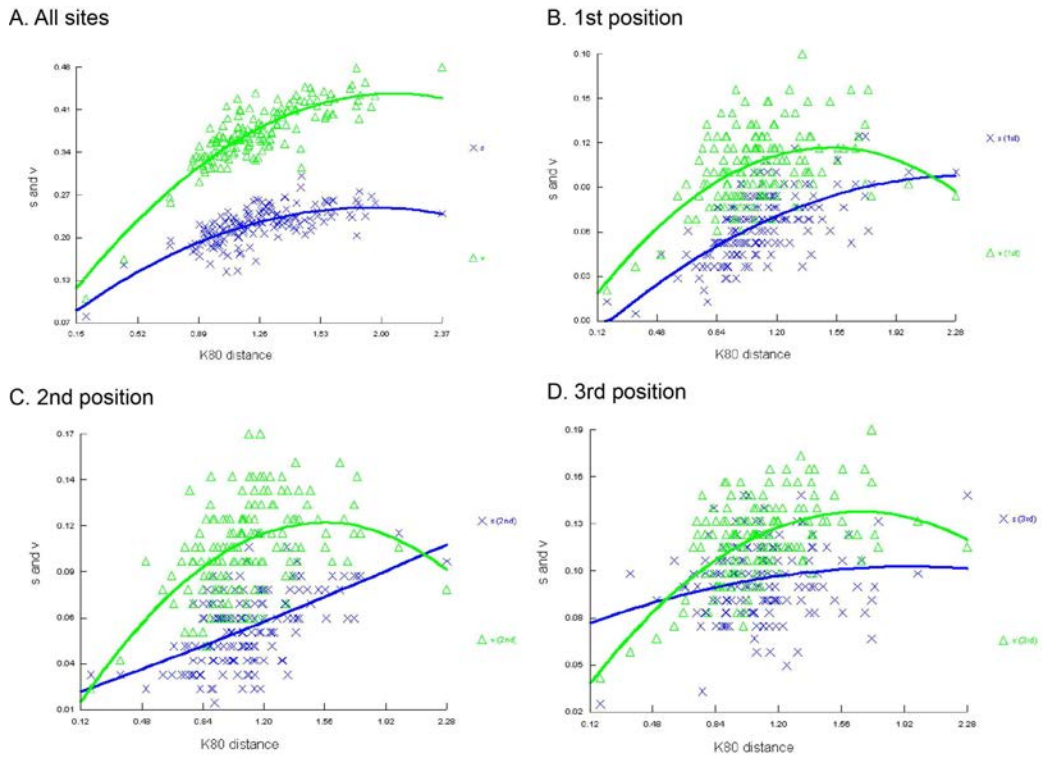



Figure S3.1 Galaxin saturation curves of different codon position transitions (x) and transversions (Δ) under K80 distance. The analyses were conducted by DMABE5 (Xia 2013) for galaxin sequences from coral (*Acropora*, *Galaxea*), corallimorpharians (*Corynactis*, *Ricordea*) and actiniarian (*Nematostella*). The substitution model was estimated by MEGA5 (Tamura et al. 2011). Only fully resolved sites were analyzed.

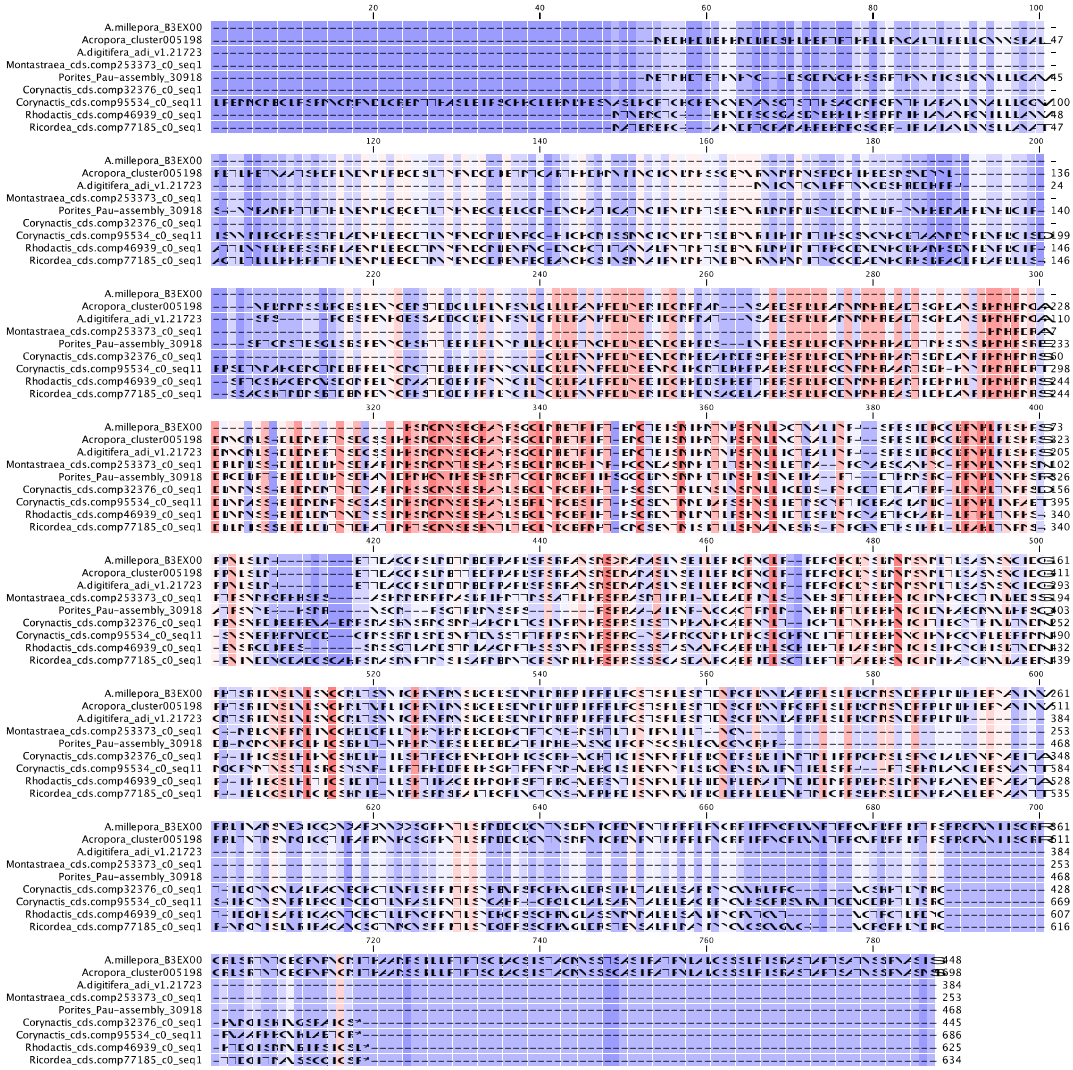



Figure S3.2 SOMP-1 sequences alignment. The site with similarity higher than 50% were highlighted in red, lower than 50% in blue.

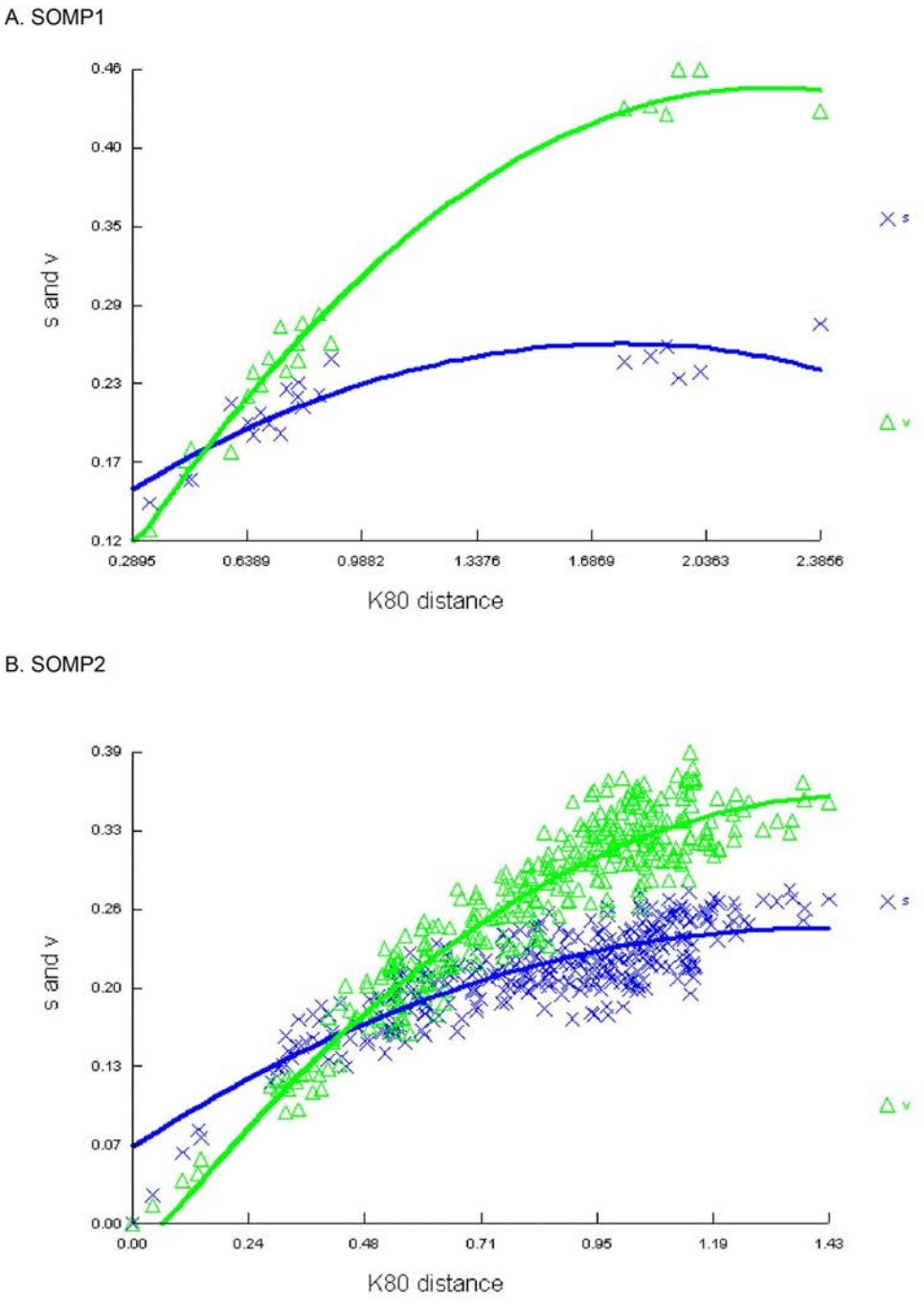



Figure S3.3 SOMP-1 (A) and SOMP-2 (B) saturation curves of different codon position transitions (x) and transversions (Δ) under K80 distance. The best-fit model was estimated by MEGA5 (Tamura et al. 2011). Only fully resolved sites were analyzed.

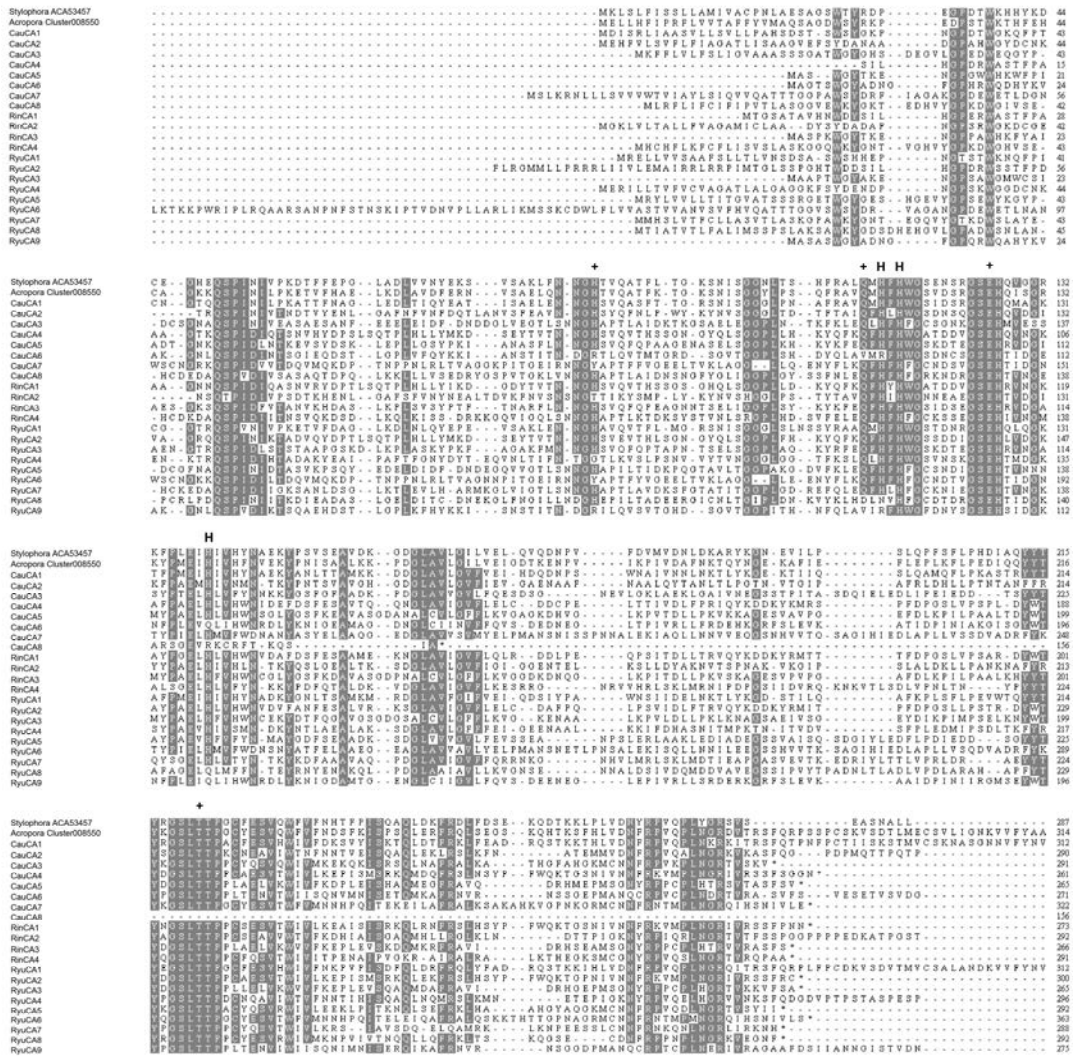


Figure S3.4 Amino acid alignment of α-CA sequences of three corallimorpharians. Conserved domains with 70% identity were highlighted. Histidine residues essential for zinc binding are denoted with H. + marks the seven active sites.

# Chapter IV

## Transcriptomic profiles of the tropical corallimorpharian *Ricordea yuma* (Ricordeidae, Corallimorpharia) after re-infection with homologous and heterologous *Symbiodinium* clades

**Transcriptomic profiles of the tropical corallimorpharian *Ricordea yuma* (Ricordeidae, Corallimorpharia) after re-infection with homologous and heterologous *Symbiodinium* clades**

Mei-Fang Lin[1,2], Sylvain Foret[2,3], David J. Miller[1,2*]

[1]Comparative Genomics Centre and Department of Molecular and Cell Biology, James Cook University, Townsville, QLD 4811, Australia

[2]ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD 4811, Australia

[3]Research School of Biology, Australian National University, Canberra, ACT 0200, Australia

*Corresponding author: david.miller@jcu.edu.au

**Abstract**

Many cnidarians host strains of the photosynthetic dinoflagellate *Symbiodinium*, the best-known example being the tropical shallow water scleractinians on which coral reef ecosystems are based. The molecular bases of this interaction remain largely unknown. *Symbiodinium* is a highly diverse taxon, and this is reflected in great metabolic diversity, but the interaction with cnidarian hosts appears to be specific, presumably reflecting appropriate physiological characteristics on both sides of the relationship. The presence of different symbiont types is known to affect expression levels of specific host genes, but effects on the transcriptome more broadly have not previously been investigated. To better understand the impact of different symbiont types on host gene expression, in the present study transcriptome profiling was used to investigate the biological processes occurring after reinfection of bleached samples of the tropical corallimorpharian, *Ricordea yuma*, with either the natural ("homologous") clade C1 symbiont or a compatible but heterologous *Symbiodinium* strain (D1a) isolated from a different corallimorpharian host (*Rhodactis indosinesis).*

Transcriptomic analyses indicate that higher levels of glycogen biosynthesis pathway enzymes were induced in the host by the homologous symbiont, and genes involved in ammonia assimilation showed altered expression in the two types of infection. Some of the most highly up-regulated genes in symbiosis were unique to symbiotic corallimorpharians, suggesting that this trait may have evolved independently in corallimorpharians.

Key words: corallimorpharian, reinfection, symbiosis

## Introduction

Symbiosis is central to the physiology and health of many cnidarians (Weis & Allemand 2012), and cnidarian-dinoflagellate symbioses are particularly significant in the case of tropical reef-building corals (Davy et al. 2012). Photosynthesis in the dinoflagellate symbiont is a primary feature of this relationship; photosynthetic products such as sugar, lipids and oxygen are transferred to the host, supporting growth, respiration, reproduction, and in the case of stony corals, calcification. In return, within the host, the symbiont has access to inorganic carbon, and potentially also to higher levels of inorganic nitrogen and phosphorus than are present in the environment (Davy et al. 2012).

Whilst some (around 15%) of corals maternally transmit *Symbiodinium* to the offspring, in the majority of corals the symbionts are acquired from the environment during the early stages of each generation (Baird et al. 2009). Many aspects of the interaction are unclear, including the process of infection and establishment of symbiosis; the establishment of a stable cnidarian-dinoflagellate relationship is thought to involve a complex series of processes including recognition, engulfment, signal transduction, cell proliferation and the suppression of apoptosis, and ultimately metabolite trafficking (Davy et al. 2012). In the coral *Acropora*, compatible symbionts are phagocytosed, but are able to block the maturation of phagosomes at an early stage and thus escape digestion (Mohamed et al. 2016); data from two symbiotic sea anemones suggest that this model applies more generally,

and the mechanism resembles those by which some parasites and microbes evade elimination by their hosts (Mohamed et al. 2016). Whether the same mechanism of symbiont infection applies to corallimorpharians is unknown, but given their close relationships with reef-building corals, this is likely to be the case. Whilst gene expression data are beginning to yield insights into the initial interaction between cnidarian host and *Symbiodinium*, less is known about molecular events beyond that point. Symbiont uptake appears to be a relatively promiscuous process (Cumbo et al. 2013), but later there are likely to be several "winnowing" (Dunn & Weis 2009), or selective, stages which are opportunities for the host to reject strains of *Symbiodinium* that do not fit its physiological requirements. The literature implies that many symbiotic anthozoans, including corallimorpharians, corals, and anemones, can form stable associations with only one or a few specific *Symbiodinium* types, but the simultaneous presence of numerous symbiont types has been demonstrated in several host species (Baker 2003, Kuguru et al. 2008, Wilkinson et al. 2015). The dogma is that the homologous symbiont enables the coral holobiont to function most efficiently (Weis et al. 2001), and that post-phagocytic recognition mechanisms remove all but the optimal (homologous) symbiont type (Dunn & Weis 2009).

Whilst the consensus is that compatible symbionts enter their hosts essentially by stealth (Schnitzler & Weis 2010, DeSalvo et al. 2010, Mohamed et al. 2016), the response of a cnidarian host to incompatible symbiont strains is thought to be more extensive. In both *Acropora palmata* and *Orbicella* (formerly *Montastraea*)

*faveolata*, the very few gene expression changes were associated with infection by a compatible strain, whereas an incompatible strain triggered major transcription changes at 6d after infection (Voolstra et al. 2009), which presumably reflect rejection. However, the Voolstra et al. (2009) study is a rather extreme case – the incompatible symbiont strain EL1 was isolated from the jellyfish *Cassiopea*, and the relationship failed over a period of days. Other studies have examined gene expression changes when corals were infected with different clades of *Symbiodinum*. Yuyama et al. (2011) used the high-coverage gene expression profiling (HiCEP) method to examine the response of aposymbiotic juveniles of *Acropora tenuis* to two different symbiont types; although 765 genes were found to be differentially expressed between the two groups, only 33 (some of which may be involved in lipid metabolism) could be annotated and validated.

*Symbiodinium* is a highly diverse taxon, and this is reflected in great metabolic diversity (e.g. Klueter et al. 2015). Differences in holobiont thermotolerance have been attributed to the presence of particular *Symbiodinium* strains (Baird et al. 2009), and growth rates of *Acropora tenuis* primary polyps differ depending on whether they are infected with C1 or D-type *Symbiodinium* (Yuyama & Higuchi 2014). Whilst the presence of different symbiont types has been shown to affect expression levels of specific host genes (Yuyama & Higuchi 2014), to our knowledge, effects on the transcriptome more broadly have not previously been investigated.

In the present study, transcriptome profiling was used to investigate the biological

processes occurring after reinfection of a bleached cnidarian host, the tropical corallimorpharian, *Ricordea yuma,* with either the natural ("homologous") clade C1 symbiont or a compatible but heterologous *Symbiodinium* strain (D1a) isolated from a different corallimorpharian host (*Rhodactis indosinesis)*. Although nothing has been reported from *R. yuma* symbiosis, the preliminary test of this experiment indicated that bleached *R. yuma* can be survived after the reinfection with *Symbiodinium* clade D1a from *Rhodactis*. While both strains of symbiont formed what were assumed on the basis of observation to be stable associations, despite slower initial growth *in hospite*, over time the homologous symbiont (growing from a low background in bleached animals) displaced the heterologous strain in most cases. Transcriptomic data imply that the homologous symbiont releases more glucose to the host than does the heterologous strain, as higher levels of glycogen biosynthesis enzymes were induced in the host by the former, and a number of genes involved in ammonia assimilation showed altered expression in the two types of infection. Some of the most highly up-regulated genes in symbiosis involving either strain of symbiont were unique to symbiotic corallimorpharians, the suggestion being that symbiosis evolved independently in corallimorpharians and these are consequently taxonomically restricted genes required for this.

**Materials and methods**

Sample collection

*Ricordea yuma* polyps originally collected from the Great Barrier Reef (18°25'35.20"S, 146°41'10.91"E) were maintained at Reef HQ aquaria (Townsville, Australia) for several years prior to the start of the work described here. On 23 August 2012, twelve polyps of *Ricordea* were transferred to MARFU at James Cook University and, within an hour of dispatch, were placed in a 1000 l tank that received a constant flow of seawater (3000 l $h^{-1}$) at an average water temperature of 26.15 ± 0.016°C (as recorded by HOBO Light/Temperature Data Loggers, Onset Corp.). All the samples were exposed to the same shaded ambient light condition. Prior to the experiment, six samples were split to produce six pairs of genetically identical samples. After two months, the split samples had fully recovered, resulting in complete polyps with similar diameters (Fig. S4.1). The remaining six *Ricordea* polyps were used to construct the reference transcriptome (described below).

Chlorophyll fluorescence (maximum quantum yield of PSII) (*Fv/Fm*) (Maxwell & Johnson 2000) was detected using a mini PAM (Walz, Germany). At weekly intervals, PAM measurements were conducted at one hour after sunset, following the method of Kuguru et al. (2007). Chlorophyll autofluorescence under blue light was also monitored microscopically to determine approximate zooxanthellae content of tissues (Schwarz et al. 2002).

Sample acclimation and tank experiment

To buffer against fluctuations and ensure temperature consistency across the treatments, the three (one control and two experimental) 13.8 l aquaria were placed under shaded ambient light in the same larger (1000 l) aquarium, which was subjected to constant water flow and aeration (300 $lh^{-1}$). The samples were uniformly distributed across the three tanks, each tank sharing two duplicates with the other tanks (Fig. S4.1). During the acclimation period, all 12 polyps gave similar *Fv/Fm* readings (0.56-0.81) (Fig. 4.1).

Samples were separated into two groups for infection experiments (group C, D) and one control group. From 14 September 2013 to 24 April 2014, all samples were kept at a depth of 12 cm in three aquaria (each with 4 polyps) built with 300 $lh^{-1}$ flow rate and consistent temperature of 26 °C with ambient light. Prior to reinfection with zooxanthellae, all samples were bleached using the menthol method described in Wang et al. (2012). After the 11 days of menthol treatment, samples were bleached with an average *Fv/Fm* $\leq$0.005 (3 replicates for each measurement) and allowed to recover for 7 days before reinfection was conducted.

Samples in the experimental group C were reinfected with the *Symbiodinium* clade C1, which is the natural symbiont of *R. yuma*, and those in the experimental group D were reinfected with the *Symbiodinium* clade D1a, which was isolated from another tropical corallimorpharian, *Rhodactis indosinensis*. Because antibiotic treatment may act as a selective force during *Symbiodinium* culturing (Santos et al.

2001), and it has been suggested that the stress susceptibilities of *Symbiodinium in hospite* and differ from those of freshly isolated cells (Bhagooli & Hidaka 2003). In the present experiment, freshly isolated zooxanthellae were used for the reinfection. During the isolation process, zooxanthellae were washed at least four times by resuspension in 8ml 1 micron filtered seawater (FSW) followed by centrifugation (860 x g for 3 minutes). This procedure was repeated until the resulting pellets were clear and homogeneous. Isolated zooxanthellae were counted by using hemacytometer and diluted to approximatelt 800 cells $l^{-1}$ (FSW) before being used for reinfection. The reinfection process consisted of injecting 1 ml aliquots of the zooxanthellae suspension into the mouth of the each polyp at 3-4 day intervals, always at 5pm. The water in each tank was changed before the next reinfection, and the water in the control aquarium was changed on the same days as the experimental aquaria. During the course of the experiment,, two individuals were lost (one from the control tank and one from C group) probably because of the weakness of samples after the bleaching stress, thus, at the end of the experiment was complete, a total of 10 samples were available for analysis. The reinfection was terminated when the sign of recovery was visualized and determined by the *Fv/Fm* value (Fig. 4.1), and each sample recorded photographically (Fig. S4.1).

Reference transcriptome samples

The reference transcriptome was constructed from six different individuals, under a range of treatments, giving rise to nine individual samples, as summarised in Table

S4.1. *Ricordea* has the ability to regenerate new polyps from fragments within two months, and frequently reproduces asexually by marginal budding (Lin et al. 2013). To increase the range of transcriptional states sampled, one polyp (sample E) was bisected, and an individual produced by marginal budding also sampled (sample A). Dark bleached samples were from polyps that had been maintained in the dark for three months under the same water flow and temperature conditions as other treatments. The menthol bleaching method employed was identical to that used for *Rhodactis* (Lin et al. submitted). After 8-11 days, menthol-treated *Ricordea* became bleached with PSII activity of (Fv/Fm <0.13) and had zooxanthellae abundance in tentacles of <200 cell/cm.

*Symbiodinium* identification and cell size measurement

*Ricordea* tissue specimens were cut and preserved in 70% ethanol for genomic DNA extraction followed the method described in Chen and Yu (2000). For the identification of symbionts present in tissue samples, Polymerase Chain Reaction (PCR) was used to amplify the internal transcribed spacer 2 region (ITS2) with the modified primers ITSintfor2 5'-GAATTGCAGAACTCCGTG-3' and ITSrev 5'-GGGATCCATATGCTTAAGTTCAGCGGGT -3' (LaJeunesse & Trench 2000). Following the supplier's (MyTaq; Bioline, Australia) recommended protocol, samples were denatured for 1 min at 95°C, and then subjected to 30 PCR cycles of 15 s at 95°C, 15 s at 50°C, and 10 s at 72°C. The PCR products were directly sequenced, and then assembled using DNAStar (Lasergene, USA) for

*Symbiodinium* genotyping.

In order to measure the size of *Symbiodinium* cells from each clade to see if the cell size affects the infection effiency, two measurements of diameter (vertical and horizontal) were made for each of more than 400 individual cells, using the CellSens software (Olympus, Japan) in conjuction with an Olympus XB53 microscope. The size distributions *Symbiodinium* cells are shown as a box plot included as Fig. S4.2.

Tissue sampling and RNA extraction

Tissues samples (~300 mg) for RNA extraction were cut with sterile scissors at three time points: R1, the first sign of recovery after reinfection, R2, the point at which zooxanthellae appeared to be evenly distributed in each polyp, and R3, upon full recovery after infection. Reinfection stages were determined by the value of *Fv/Fm* (Fig. 4.1). Tissues samples were immediately snap-frozen in liquid nitrogen and suspended in Trizol for RNA extraction. Total RNA was extracted using TRI Reagent (Ambion) protocol, which is based on the Chomczynski & Sacchi (1987) method, and then dissolved in RNase-free water. RNA quality and quantity were assessed using a NanoDrop ND-1000 spectrometer and denaturing gel electrophoresis using standard methods (Sambrook and Russell 2001). In order to minimise circadian effects on gene expression, sampling was performed at 8 am (about 2-2.5 hours after sunrise) in all cases (as suggested in Ganot et al. 2011).

cDNA libraries development, sequencing and transcriptome assembly

cDNA libraries were generated using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina RNAseq (New England BioLabs, Inc. USA). To generate the reference transcritome, a total of 769 M raw reads were obtained using the Illumina HiSeq2000 sequencing platform. One of the *Ricordea* (symbiotic) libraries was deeply sequenced (a full lane of 100bp PE reads), the remaining nine samples spread across a further two lanes and 100 bp single-end (SE) data collected. For the experimental reinfection samples, libraries were sequenced by the use of two lanes (100bp PE reads). The libngs program (https://github.com/sylvainforet/libngs) was used for quality trimming. After removal of data with quality scores lower than 30 or shorter than 70 bases, a total of 214 M reads remained, from which the reference transcriptome was assembled using Trinity (Haas et al. 2013). Symbiont and host sequences were separated and predicted proteins annotated as described in Lin et al. (submitted). The *Ricordea* reference transcriptome has 94,579 contigs with mean size of 1,175 bases and N50 of 1,679 bases.

Differential expression analysis

Read data from experimental reinfection samples were mapped to the reference transcriptome using the Bowtie v2.1.0 software (Langmead & Salzberg 2012). Differential gene expression was inferred by mapping counts using the EdgeR

package (Robinson et al. 2010) with an expression level cutoff of 5 counts in more than 10 samples and the GLM approach. An MDS plot was used to investigate the relative similarities of the samples (Fig. 4.2). For gene ontology analyses, GOseq v1.16.1 (Young et al. 2010) was employed, with differential expression threshold of FDR<0.05.

Sequence analysis and function prediction

To investigate the potential functions of genes that have no clear homolog in the NR database, PROSCAN (Combet et al. 2000) was used to scan for protein signatures and functional prediction. Conserved domain searching was based on the CDD search tool at the NCBI conserved domain database (Marchler-Bauer A et al. 2015). The TargetP 1.1 (Emanuelsson et al. 2000) and TMHMM v. 2.0 (Krogh et al. 2001) servers were used to predict intracellular location and scan for transmembrane domains, respectively.

**Results and Discussion**

Infection of corallimorpharian host with clade C1 or clade D1 *Symbiodinium*

Prior to bleaching the *Ricordea* samples by menthol treatment (Wang et al. 2012), the taxonomic affinities of the *Symbiodinium* types present were investigated by PCR of the ITS2 region (LaJeunesse & Trench 2000). In each case, sequence comparisons indicated that the natural symbiont is a clade C1-type *Symbiodinium* which is sensus those identified in LaJeuness (2001) (Suppl material TableS4.2). In the re-infection experiment, the samples were distributed across three treatment groups, one group being reinfected with clade C1 *Symbiodinium* (C group), one group reinfected clade D1a type *Symbiodinium* (D group), and a control group which were not exposed to *Symbiodinium* (control group).

At the conclusion of the re-infection experiment (sampling point 3; see Fig. 4.1), *Symbiodinium* clade typing was again carried out using the same primers. Clade C1 *Symbiodinium* dominated all of the samples in the C infection group but, surprisingly, clade C1 also dominated three of the four samples in the D treatment group, clade D1a being the dominant type in only one of the four samples. With the exception of a few ambiguous sites, the ITS2 sequence from that one sample D infection group was identical to that of the D1a strain used to infect *Ricordea* samples (Fig. S4.3). The dominance of *Symbiodinium* clade C1 sequences in the D treatment group at the final sampling point implies that the menthol bleaching

treatment does not remove all of the zooxanthellae present. Consistent with this, a small number of clade C-type ITS2 sequences were detected in control group samples at the first sampling point (data not shown), despite the very low Fv/Fm at that time (Fig. 4.1).

The first indicators of successful infection were increased Fv/Fm values, which were first detected in the D group after two weeks of reinfection and in the C group somewhat later (Fig. 4.1). Approximately seven days later, zooxanthellae could be visualized as yellow/brown spots in the infected animals. Interestingly, the rate reinfection was faster in the D group than in the C group. Measurements of the diameter of individual cells indicate that clade D1 *Symbiodinium* are on average smaller than are clade C1 cells, which may underlie the slower infection rate of the latter (Fig. S4.2). However, further investigation is required to confirm this.

Differential gene expression during the re-establishment of symbiosis

The differential expression (DE) dispersion of samples in the three treatment groups is shown as Fig. 4.2. Multidimensional scaling analysis (MDS) indicates that samples are resolved both by treatment group and by time in the reinfection process, implying that the differential expression observed is largely a consequence of the experimental treatments, rather than variation among the individuals. In the MDS plot (Fig. 4.2), samples from the control group are distantly related to most of the experimental samples, the proximity of some control samples and two reinfection

group C samples perhaps being a consequence of the low residual population of type C1 *Symbiodinium* in some of the control samples. At the first infection stage (R1), samples from the three treatment groups were not as well resolved as at the later sampling points (R2 and R3), indicating limited variation in gene expression patterns at the early time point. Later, the differential expression patterns in the C and D treatment groups differed significantly both from each other and from the control group. Changes in expression patterns over time were most consistent for the C treatment group. In the D treatment group, some overlap was observed for samples at the R2 and R3 stages (Fig. 4.2), which is consistent with the similar Fv/Fm values observed in these cases (Fig. 4.1). In summary, the MDS analysis indicated that reinfection of the corallimorpharian samples with clade C1 (the natural symbiont) or clade D (a compatible, but heterologous strain) *Symbiodinium* consistently gave rise to differing patterns of gene expression.

Genes differentially regulated during the establishment of symbiosis

Dispersion patterns of differentially expressed transcripts at each stage in each treatment group are shown as Figure 4.3. Using a significance cutoff of FDR < 0.05, 88 transcripts were differentially expressed across the three treatment groups. Among these, 40 transcripts have open reading frames encoding more than 100 amino acid residues, 25 of which correspond to annotated proteins and 15 of which are uncharacterized proteins (Table S4.4). Of these 40 protein-coding genes, 23 were differentially expressed in the C1 and D1a treatment groups only (Fig. 4.5).

For most of genes, differential expression was much more extensive at the late (R3) stage, probably as a consequence of the recovery of clade C1 zooxanthallae, as indicated by Fv/Fm.

## Gene ontology

In general, there are more overrepresented genes at the late stage than those of early stage (Fig. 4.3). Gene ontology (GO) analyses indicated that two GO terms were significantly over-represented in two experimental treatment groups, but not in the control group, during the reinfection processes. The term glycogen biosynthetic process (GO:0005978) was over-represented amongst up-regulated genes in the R2 and R3 stages of C group samples and R3 stage D group samples, and oxidation-reduction process (GO:0055114) was over-represented amongst down-regulated genes at both the R2 and R3 stages in D group samples. The genes responsible for over-representation of these GO terms were examined further, as described below.

### *Glycogen biosynthesis*

Most animals synthesise glycogen from glucose as a short-term energy and carbon storage system. A model of the glycogen biosynthesis pathway inferred by KAAS (KEGG Automatic Annotation Server) is shown as Fig. 4.4. In cnidarian-dinoflagellate symbioses, glycogen is an important energy storage

compound (Kopp et al. 2015), and glucose is thought to be the major photosynthetic product transferred from symbiont to host (Burriesci et al. 2012). Two genes involved in glycogen biosynthesis, glycogenin-1 and 1,4-alpha-glucan-branching enzyme, were highly up-regulated in later stage C and D treatment group samples, as detailed above (Fig. 4.4). Also belonging to the same GO term, glycogen synthase and UTP-glucose-1-phosphate uridylyltransferase were significantly differentially expressed only in the C treatment group samples. Although the GO term glycogen biosynthesis was not over-represented in early (R2) stage of D treatment group samples, some individual genes in this category were significantly up-regulated, including glycogen synthase, phosphoglucomutase-1 and 1,4-alpha-glucan-branching enzyme. Overall, levels of expression of genes involved in glycogen biosynthesis in the D treatment group samples were approximately half of those in C treatment group samples.

In addition to up-regulation of genes encoding the catalytic activities involved in glycogen biosynthesis, two regulatory proteins were also up-regulated. Protein phosphatase 1 regulatory subunits 3B and 3E (PP1R3B and PP1R3E) both contain carbohydrate-binding domains and act to facilitate glycogen synthesis, by supressing the rate at which PP1 inactivates (dephosphorylates) glycogen phosphorylase, and stimulating the rate at which it activates (phosphorylates) this enzyme, effectively driving the reaction in favour of glycogen biosynthesis. Note that levels of induction of both regulatory proteins were higher in the clade C1 infection than in the Clade D treatment (Fig. 4.5).

The up-regulation of genes involved in glycogen biosynthesis observed following infection with *Symbiodinium* is consistent with translocation of photosynthetically-derived glucose from symbiont to the host, and active storage of glucose as glycogen in the corallimorpharian host. These data are consistent with reports of higher levels of expression of glucose- and glycogen-related genes in coral at noon in comparison with those at night (Ruiz-Jones & Palumbi 2015). A recent study demonstrated carbon translocation from symbiont to coral host by visualizing external glucose incorporation into glycogen granules in oral epidermal cells, providing the first direct evidence for storage of photosynthetically derived carbon in coral tissue (Kopp et al. 2015). The up-regulation of glycogen biosynthesis during recovery from bleaching supports a central metabolic role for glucose in corallimorpharian-dinoflagellate symbioses, as in corals

The lower levels of expression of glycogen biosynthetic enzymes observed in the corallimorpharians infected with clade D1a rather than clade C1 *Symbiodinium*, and the faster growth of the former strain in hospite, are consistent with the idea that less glucose is transferred to the host by clade D1a symbionts than by clade C1 symbionts. It has been suggested that corals hosting different clades of *Symbiodinium* have distinct bleaching characteristics, and host transcriptomic patterns are also affected by the nature of the *Symbiodinium* strain hosted (DeSalvo et al. 2010).

*Symbiosis results in altered expression of genes involved in ammonia assimilation*

The GO term oxidation-reduction process (GO:0055114) was over-represented in both R2 and R3 stage samples from the D1a treatment group; some of the genes captured under this GO term were also differentially expressed in the C1 treatment group (Table S4.3). Two genes involved in nitrogen assimilation were down-regulated in both clade C1 and clade D1a samples: glutamate dehydrogenase (GDH), which is responsible for the incorporation of ammonium into -ketoglutarate and delta-1-pyrroline-5-carboxylate dehydrogenase (ALDH4A1), a mitochondrial enzyme responsible for the production of glutamate from glutamate semialdehyde during breakdown of either proline or ornithene. Based on sequence comparisons, the GDH that is differentially expressed in symbiosis is assumed to be the NADP+-specific variant (Catmull et al. 1987), which is restricted to cnidarians. Both GDH and glutamine synthetase (GS) are thought to play important roles in regulating host cytoplasmic ammonium levels in the coral-dinoflagellate symbiosis (Yellowlees et al. 1994). Both of these enzymes are able to function in ammonium assimilation, but do so with very different kinetics – in the case of animal GS proteins, Km values for ammonium are typically around 10 $\mu$ M, whereas the corresponding figure for GDH is about two orders of magnitude higher; in *Acropora*, the Km of the NADP-specific GDH for ammonium has been estimated as 9.2mM (Catmull et al. 1987). During the reestablishment of symbiosis with the homologous C1 clade, not only was GDH down-regulated, but also GS was significantly up-regulated (Table 4.1). However, GS was apparently not up-regulated in the clade D1a treatment group. Given the lower Km of GS for ammonium, these results imply

that the host cytoplasmic ammonium concentrations will be lower in the presence of the homologous (clade C1) symbiont than a compatible but heterologous strain (clade D1a). One interpretation of the observed activation of host GS in the symbiotic state is that decreasing cytoplasmic ammonium levels imposes nutrient limitation on the symbiont, enabling the host to restrict *Symbiodinium* biomass (Davy et al. 2012), but another is that the host must maintain low cytoplasmic ammonium levels to prevent collapse of the pH gradient across the symbiosome membrane (Miller and Yellowlees 1988) Whatever its physiological role, at this stage it is not clear why GS expression is activated by the homologous symbiont, but not by a compatible but heterologous strain.

The observed up-regulation of a homolog of the human Rh type C protein in the symbiotic state has a precedent in the symbiotic sea anemone, *Aiptasia* (Lehnert et al. 2014), and may also be relevant to ammonia metabolism in corallimorpharians because the mammalian Rh proteins are ammonia channels. However, the algal Rh proteins are thought to be channels for $CO_2$ rather than ammonium (Huang and Peng 2005), and the cellular location of the cnidarian Rh protein is unknown, therefore its role in symbioses is unclear.

*Other metabolic aspects of the establishment of symbiosis*

The transition from aposymbiotic to symbiotic stages involves many metabolic adjustments beyond those documented above. Prior to symbionts being introduced,

the corallimorpharians were essentially starved – to induce and maintain the aposymbiotic state, they were maintained in the shaded ambient light in filtered artificial seawater, and were not fed. Consequently, the up-regulation of some genes observed during the course of the infection experiment simply reflects the down-regulation of catabolic processes. Examples of this are dimethylglycine dehydrogenase (DMGDH), glycine dehydrogenase [decarboxylating](GLDC), and phenylalanine hydroxylase (PAH) (Table S4.2).

*NGFR and NOS - indicators of coral health?*

Nitric oxide is used as a signalling molecule across the animal kingdom, and diverse roles have been demonstrated in different cnidarians (Colasanti et al., 2010). Nitric oxide synthase (NOS) has been implicated in bleaching (the loss of symbionts) in several cnidarians, but the mechanism is unclear. It has been suggested that, in *Aiptasia*, nitric oxide production essentially serves as an eviction notice to symbionts, NO potentially combining with superoxide generated under stress to produce peroxynitrite, which induces cell death and bleaching (Perez and Weis 2006). However, it is unclear whether NO generated during the bleaching response is a product of the host (Perez and Weis 2006) or *Symbiodinium* (Trapido-Rosenthal et al., 2005). In the present case, a nitric oxide synthase (NOS) that has high similarity to that of *Discosoma striata* (NCBI accession NO. AAK61379) was down-regulated significantly in samples from both later stages of C1 group individuals and the early stage of the D1a treatment group (Table S4.2). In *Acropora,*

NOS activity has been localized to the endoderm, which is also the location of the symbionts (Safavi-Hemami et al. 2010). During temperature-induced bleaching coral and anemone elevated levels of NOS are correlated with the up-regulation of host caspase-like enzyme activity (Hawkins et al. 2013, 2014). The down-regulation of this gene observed in the C infection group suggests that low NOS activity could be an indicator of coral health.

A homolog of tumor necrosis factor receptor superfamily member 16 (NGFR) was the only gene down-regulated in both reinfection groups during the infection process. Members of the mammalian tumor necrosis factor (TNF) family were characterized based on their ability to induce apoptosis and in the context of immune responses (Pfeffer 2003), and NGFR has been shown to mediate apoptosis in neural cells (Frade et al. 1996). The coral genome encodes a large number of TNF family members (Quistad et al. 2014), two of which were up-regulated in response to heat stress in *Acropora hyacinthus* (Barshis et al. 2013), suggesting involvement in the regulation of apoptosis. In the present case, the down-regulation of NGFR observed in both reinfection groups could be interpreted in terms of a requirement to suppress cell death processes during symbiosis (Dunn & Weis 2009).

Active nutrient allocation

In addition to the implied importance of glucose export and glycogen storage in the symbiotic cnidarians, data from our infection study are also consistent with lipid

and/or sterol translocation from symbiont to host. Annotated genes found to be up-regulated during the establishment of symbiosis included some implicated in lipid/sterol translocation in other symbiotic cnidarians, for instance, NPC2 (Kuo et al. 2010, Ganot et al. 2011, Dani et al. 2014, Lehnert et al. 2014) and the lipid storage droplet surface-binding protein 2 (Lehnert et al. 2014).

Seven of these common DE genes in groups C and D are membrane-bound proteins, some of which are transport proteins. In addition to the Rh type C (RHCG) protein discussed above, which is involved in transport of ammonium or possibly $CO_2$, a $Cl^-/HCO_3^-$ transport protein known as prestin (solute carrier family 26 member 5, SLC26A5) was up-regulated in the symbiotic state. Prestin belongs to the sulfate permease (SulP) protein family, and like several other SLC26 proteins, it contains both sulfate transporter and STAS (Sulphate Transporter and Anti-Sigma factor antagonist) domains. Three SLC26 proteins (SLC26α, β, γ) have previously been indentified in the coral *Stylophora pistillata*, but ubiquitous expression of SLC26β suggests that that it does not function in symbiosis or calcification (Zoccola et al. 2015). Phylogenetic analyses indicated that prestin has close relationship with SLC26β (Fig. S4.4). The fact that expression of this potential HCO3- transporter is up-regulated during *Symbiodinium* infection suggests a role in symbiosis that deserves further exploration.

*Novel genes: are they corallimorpharian specific symbiosis genes?*

During infection with either clade of symbiont, eleven of the most highly

up-regulated genes were taxonomically-restricted (Fig. 4.5), encoding proteins with no significant matches (*e* value cutoff < 1*e*-5) with sequences in the NCBI nr database (accessed on 10 July 2015). To verify that these were taxonomically restricted genes, the available resources for *Acropora*, *Nematostella* and all other cnidarians were searched using BlastP with *e* value cutoff of <1 x $e^{-5}$ and, to exclude the possibility that these were of symbiont origin, the *Symbiodinium* B genome was searched at the same stringency. All but one (Novel6) of these 11 taxonomically restricted genes lacked conserved domains. Novel6 is an Arg-rich protein and contains a Ribosomal protein S1-like RNA-binding domain, suggesting that it may be involved in regulation of transcription. Subcellular location prediction (conducted using TargetP v1.1; Emanuelsson et al. 2000) suggested that three of these genes encode secreted proteins, and two others carry mitochondrial (mt) targeting sequences (Table 4.1). Clear homologs of six of these *Ricordea* genes (Novel2, 4, 6, 8, 10, 11) were identified in the transcriptome of *Rhodactis*, another *Symbiodinium*-hosting corallimorpharian, but no corresponding genes could be identified in the temperate corallimorpharian, *Corynactis*, in any other cnidarian or in *Symbiodinium* B. The up-regulation of these genes in the presence of *Symbiodinium*, together with their phylogenetic distribution suggests that these may be unique to corallimorpharian symbioses; given the deep divergence of corals and corallimorpharians, symbioses with *Symbiodinum* may have evolved independently in the two lineages.

Many aspects of the cnidarian-*Symbiodinium* relationship remain unclear. The

symbiosome is highly acidic (pH ~ 4, Barott et al. 2015), and in both sea anemones and corals resembles an arrested early phagosome (Mohamed et al. 2016), but its structure is unknown. Many of the symbiosome proteins (31 of 48) identified in proteomic analyses could not be annotated (Peng et al. 2010), but most were acidic (Peng et al. 2010, Barott et al. 2015), as were three of the novel genes identified here (Novel1, 5, 11 have pI values ranging from 4.53-4.58). However, with limited information for these novel genes, their functions in symbiosis require further investigation.

**Conclusions**

Monitoring Fv/Fm changes over time indicated that the natural (homologous) *Symbiodinium* clade C1 (group C) had a slower growth rate when in hospite with *Ricordea* than did the compatible but heterologous clade D1a *Symbiodinium* strain. Transcriptomic analyses suggest that the clade C1 symbiont grows more slowly in hospite than does the clade D1a strain because in the former more of the photosynthetically-fixed carbon is translocated to the host, as evidenced by higher expression of glycogen biosynthetic pathway genes. The two *Symbiodinium* strains also differed with respect to impact on host pathways of ammonium assimilation, only the Clade C1 strain causing up-regulation of the high-affinity ammonium assimilation enzyme, glutamine synthase. These results suggest that, at least in *Ricordea*, glucose is the main form in which photosynthetically-fixed carbon is translocated from symbiont to host (Burriesci et al. 2012), and are consistent with the idea that symbioses with different *Symbiodinium* strains are not functionally equivalent (Loram et al. 2007, DeSalvo et al. 2010).

Eleven "orphan" genes that are associated with the symbiotic state were identified in *Ricordea*. Homologs of six of these genes were identified in *Rhodactis*, a second corallimorpharian that hosts *Symbiodinium*, but these had no matches in the temperate and non-symbiotic corallimorpharian *Corynactis* or any other organism. The phylogenetic distribution of these symbiont-induced genes is consistent with symbiosis having evolved independently in the corallimorpharian and coral lineages;

these two orders diverged deep in evolutionary time (Lin et al. 2016) and it is likely that symbiosis evolved independently in the two lineages.

**Acknowledgement**

The authors thank Dr David Hayward and Dr Eldon Ball (Australian National University) for the assistance with RNAseq library preparation, and thank Dr Victor Beltran (The Australian Institute of Marine Science, Townsville, Australia), Dr Catherine Collier and Dr Joe Holtum (James Cook University) for periodic loans of a mini PAM apparatus. The assistance of Ms Boer Bao (The University of Queensland) and Dr Rebecca Lawton (James Cook University) in measuring the diameter of zooxanthellae is also gratefully acknowledged. We also thank the Reef HQ Great Barrier Reef Aquarium (Townsville, Australia) for providing *Ricordea* samples, and the Marine & Aquaculture Research Facilities Unit (James Cook University) for assistance in developing the tank maintenance system. MFL gratefully acknowledges receipt of a James Cook University Postgraduate Research Scholarship.

**Table 4.1** The expression level of glutamine synthetase and glutamate dehydrogenase in *Ricordea* samples under the symbiotic and aposymbiotic states. FDR < 0.05 was indicated in bold.

| Gene | Contig | Symbiotic state | | | | Aposymbiotic state |
|---|---|---|---|---|---|---|
| | | CR1 | CR2 | DR1 | DR2 | |
| Glutamine synthetase (GS) | comp83744_c0_seq1 | **1.91** | **3.05** | 1.06 | 1.01 | **-3.59** |
| Glutamate dehydrogenase (GDH) | comp50565_c0_seq1 | -0.37 | **-1.7** | **-1.84** | **-1.47** | **1.59** |

**Table 4.2** Genes unique to *Ricordea* and *Rhoadactis* that were up-regulated in the symbiotic state.

| | Contig | Gene | Domain/Family | TM | Location | pI | Blast NR Hit | *Rhodactis*[1] | *Corynactis*[2] | *Acropora*[3] | *Porites*[4] | *Pocillopora*[5] | *Nematostella*[6] | *Anthoplera*[7] | *Hydra*[8] | *SymbB*[9] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Novel1 | comp80835_c0_seq1 | Hypothetical protein | N/A | No | Secretory pathway | 4.53 | N/A | - | - | - | - | - | - | - | - | - |
| Novel2 | comp17315_c0_seq1 | Hypothetical protein | N/A | Yes (1) | Secretory pathway | 6.06 | N/A | comp28825_c0_seq1 | - | - | - | - | - | - | - | - |
| Novel3 | comp81403_c0_seq1 | Hypothetical protein | N/A | No | Any other location | 8.25 | N/A | - | - | - | - | - | - | - | - | - |
| Novel4 | comp82357_c0_seq1 | Hypothetical protein | N/A | No | Any other location | 10.27 | N/A | comp102925_c0_seq1 | - | - | - | - | - | - | - | - |
| Novel5 | comp21509_c0_seq1 | Hypothetical protein | N/A | No | Mitochondrion | 4.58 | N/A | - | - | - | - | - | - | - | - | - |
| Novel6 | comp83997_c0_seq1 | Hypothetical protein | S1-like cold-shock domain | No | Mitochondrion | 11.27 | N/A | comp104425_c0_seq1 | - | - | - | - | - | - | - | - |
| Novel7 | comp63712_c0_seq1 | Hypothetical protein | N/A | No | Any other location | 9.9 | N/A | - | - | - | - | - | - | - | - | - |
| Novel8 | comp37834_c0_seq1 | Hypothetical protein | N/A | No | Any other location | 9.71 | N/A | comp105783_c0_seq1 | - | - | - | - | - | - | - | - |
| Novel9 | comp17700_c0_seq1 | Hypothetical protein | N/A | No | Any other location | 11.82 | N/A | - | - | - | - | - | - | - | - | - |
| Novel10 | comp93444_c0_seq1 | Hypothetical protein | N/A | No | Any other location | 8.7 | N/A | comp103016_c0_seq1 | - | - | - | - | - | - | - | - |
| Novel11 | comp30818_c0_seq1 | Hypothetical protein | N/A | No | Secretory pathway | 4.58 | N/A | comp45856_c0_seq1 | - | - | - | - | - | - | - | - |

Abbreviation: TM, presence/absence of a transmembrane domain, with number of TM in parentheses; Location, predicted intracellular location; pI, predicted isoelectric point.

Species full name, life stage, source:

1. *Rhodactis indosinesis*, symbiotic adult, Lin et al. 2016.

2. *Corynactis australis*, nonsymbiotic adult, Lin et al. 2016.

3. *Acropora millepora*, larvae, David Miller Lab.

4. *Porites australiensis*, symbiotic adult, Shinzato et al. 2014. PLoS ONE 9: e85182.

5. *Pocillopora damicornis*, symbiotic adult, Traylor-Knowles et al. 2011. BMC Genomics 12: 585.

6. *Nematostella vectensis*, nonsymbiotic adult, Putnam et al. 2007. Science 317: 86-94.

7. *Anthopleura elegantissima*, symbiotic adult, Meyer and Weis Lab.

8. *Hydra magnipapillata*, nonsymbiotic adult, Chapman et al. 2010. Nature 464: 592-596.

9. *Symbiodinium minutum*, Shoguchi et al. 2013. Current Biology 23: 1399-1408.

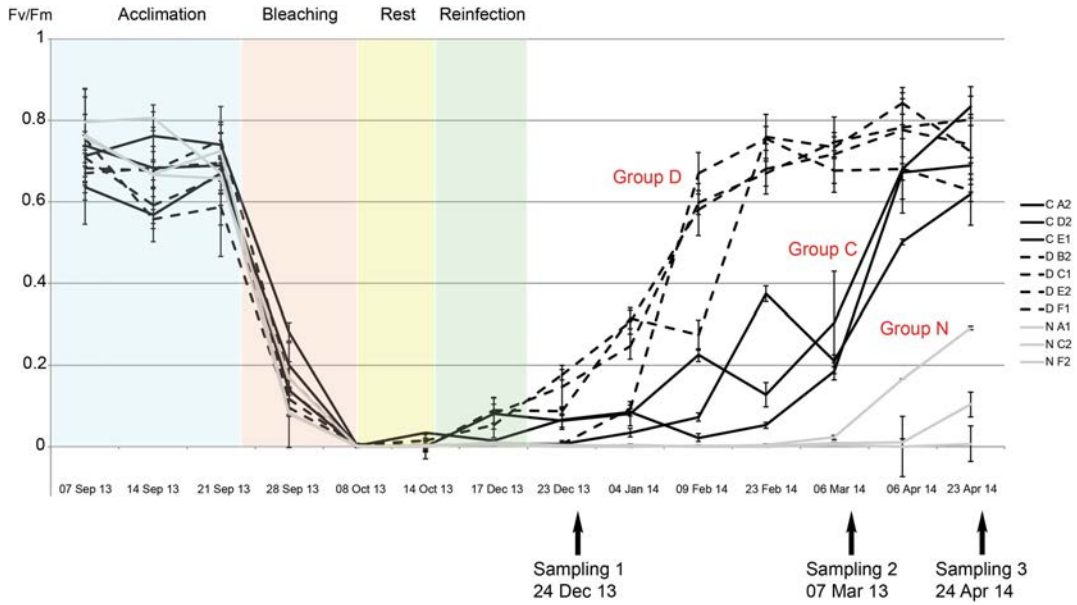



**Figure 4.1** Symbiont density during the reinfection experiment. The points indicate values of nocturnal maximum quantum yield of photosystem II (Fv/Fm) for polyps in the control (N) group (grey lines), C group (solid black lines), and D group (black dash line) throughout the experiment. The control group polyps were not experimentally reinfected. Polyps in the C and D treatment groups were reinfected with *Symbiodinium* C1 or *Symbiodinium* D1a, respectively. The labels of samples are referred to Figure S4.1. The arrows indicate the points at which sampling for RNA extraction was carried out.

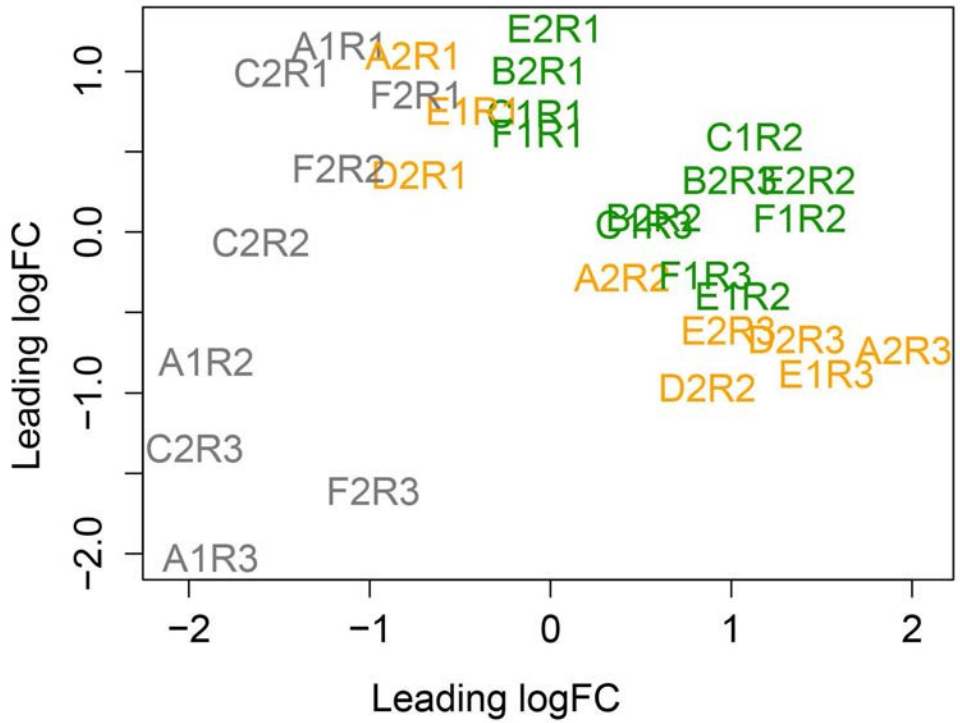



**Figure 4.2** Relationships between samples based on multidimentional scaling performed using the edgeR package. Distances on the plot represent the biological coefficient of variation (BCV) between samples. BCV = 0.2994. R1, R2 and R3 correspond to the sampling points as indicated in figure 4.1. The labels of samples are referred to Figure S4.1. Samples in the control groups are marked in grey, samples in C group are in yellow, and samples in D group are in green.

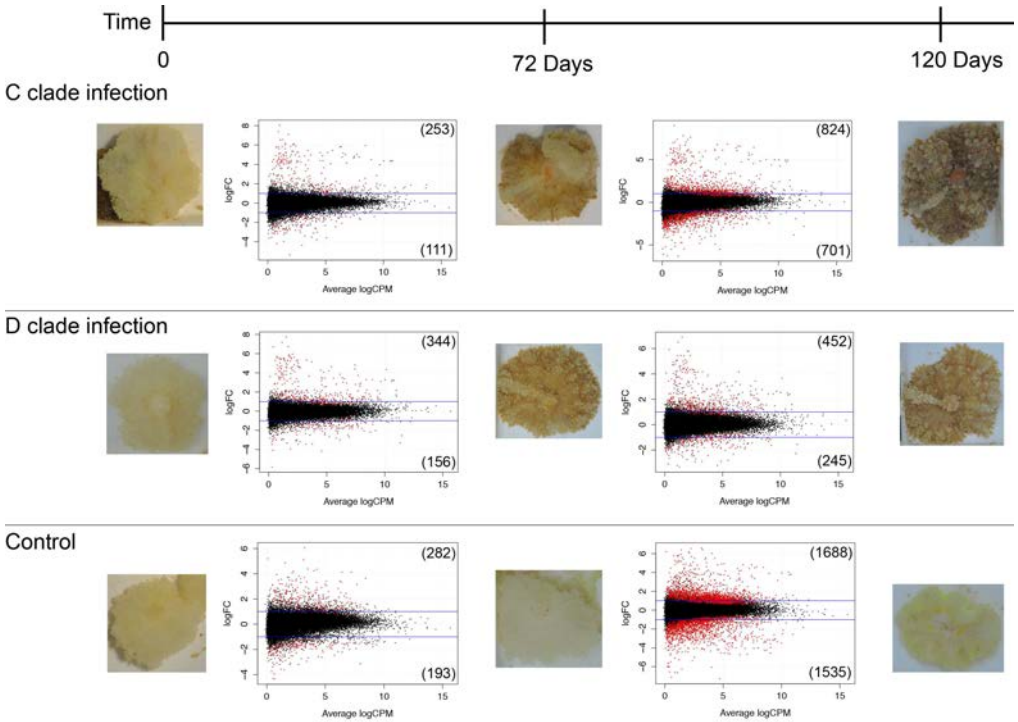



**Figure 4.3** Dot plots of log-fold-change versus log-cmp (counts per million) at each stage in the three treatment groups, with differentially expressed genes highlighted (5% FDR). In each case, the blue line indicates 2-fold change. In each case the comparisons were based on the sampling point 2 vs sampling point 1, and sampling 3 vs sampling point 1. Numbers of differential expressed genes are indicated in parentheses above and below the blue lines.

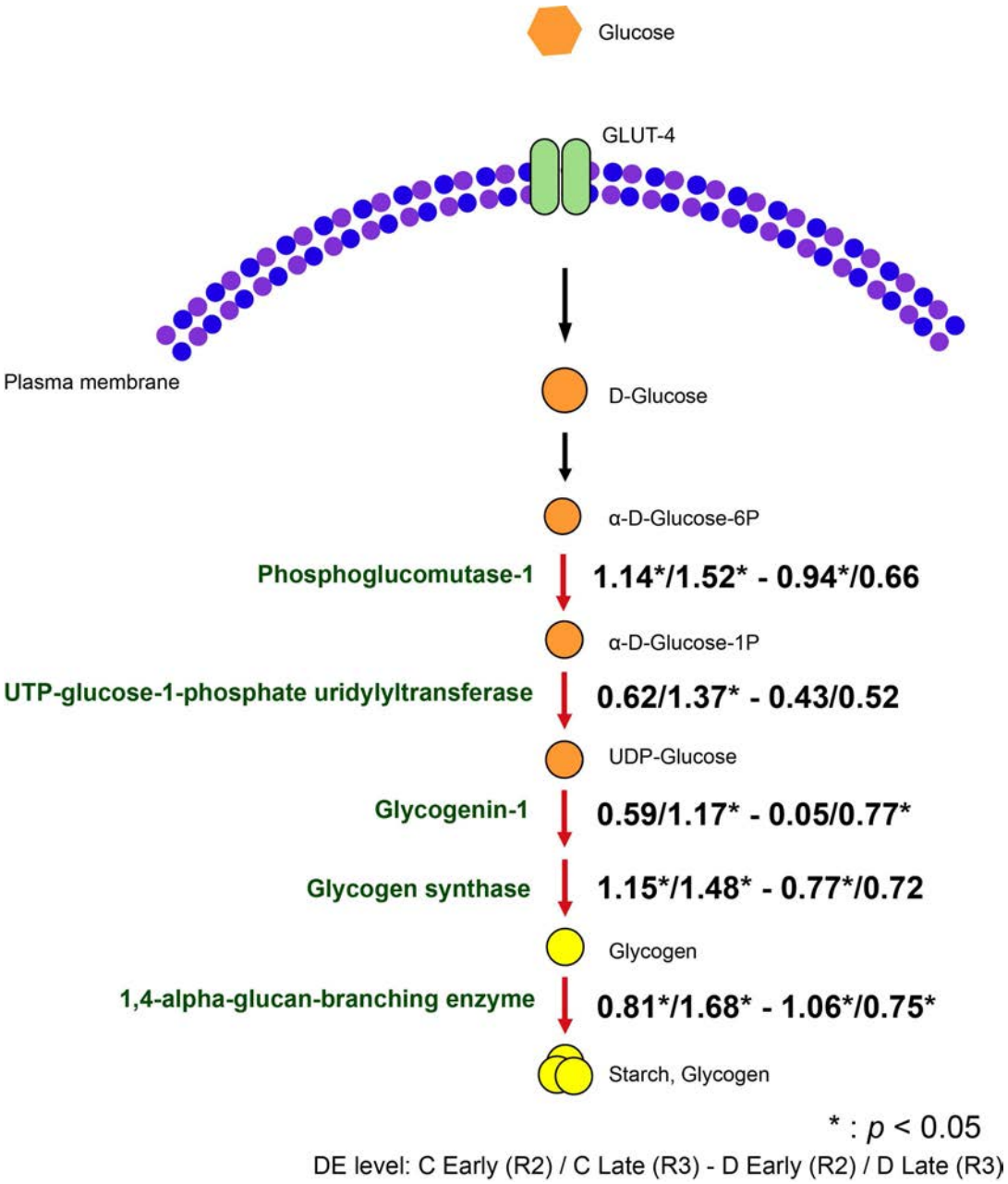



**Figure 4.4** Differential expression of host genes involved in glycogen biosynthesis during reinfection with clade C1 or D1a symbionts. Arrows highlighted in red indicate up-regulation of expression relative to levels at the first sampling point. The values represent the logFC of expression at early stage (R2) and late stage (R3) of both groups. (*) indicates the FDR < 0.05.

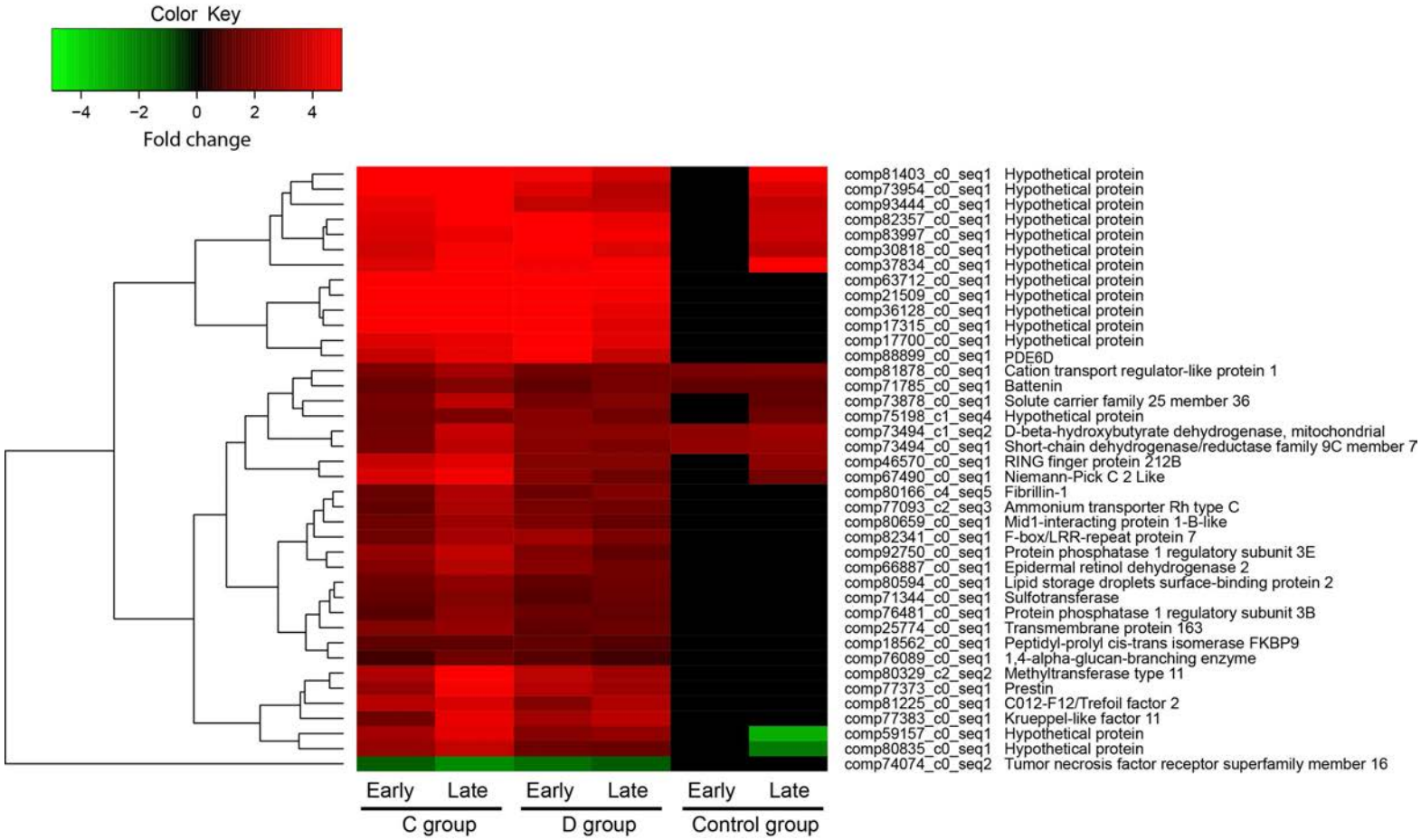



**Figure 4.5** Genes differentially expressed genes in the three treatment groups. Further information about the genes is given in Table S4.2

**Supplementary**

**Table S4.1** Experimental design for the *Ricordea* reference transcriptome development.

| Samples treatment | | | | | | |
|---|---|---|---|---|---|---|
| Symbiotic | **SampleA** | **SamplB** | **SampleC** | | **SampleE1** | |
| Darkness bleached | | | SampleCd | **SampleDd** | | |
| Menthol bleached | SampleA1m | | | | SampleE2m | **SampleFm** |

Sample relations:

The samples marked in bold present the mother polyps.

For the daughter polyps, SampleA1m was duplicated from SampleA via asexual reproduction budding, SampleCd is the same polyp as SampleC but with darkness treatment after sampling for symbiotic condition, and SampleE2m was duplicated from SampleE1 via cutting.

**Table S4.2** *Symbiodinium* ITS 2 identification in each sample before and after the reinfection.

| | Group control | | | Group C | | | Group D | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Sample | A1 | C2 | F2 | A2 | D2 | E1 | B2 | C1 | E2 | F1 |
| Before reinfection | Clade C1 | Clade C1 | Clade C1 | Clade C1 | Clade C1 | Clade C1 | Clade C1 | Clade C1 | Clade C1 | Clade C1 |
| After reinfection | - | - | - | Clade C1 | Clade C1 | Clade C1 | Clade C1 | CladeD1a | Clade C1 | Clade C1 |

**Table S4.3** Differentially expressed genes captured by the GO term oxidation-reduction process (insert GO term). Asterisks indicate where FDR < 0.05.

| Gene | Contig | Early C | Late C | Early D | Late D |
|---|---|---|---|---|---|
| Glutamate dehydrogenase | comp50565_c0_seq1 | -0.37 | -1.70* | -1.84* | -1.47* |
| Delta-1-pyrroline-5-carboxylate mitochondrial | comp70669_c0_seq1 | -0.10 | -0.90* | -0.96* | -0.73* |
| Glycine dehydrogenase [decarboxylating], mitochondrial | comp76067_c0_seq1 | -0.22 | -0.75 | -0.60 | -0.78* |
| Dimethylglycine mitochondrial | comp82635_c0_seq1 | -0.22 | -1.25* | -2.21* | -1.20* |
| Phenylalanine hydroxylase | comp22749_c0_seq2 | 0.18 | -0.66* | -0.84* | -0.78* |
| Nitric oxide synthase | comp17588_c0_seq1 | -2.45* | -1.25* | -1.77* | 0.10 |

**Table S4.4** Protein-coding genes common to the infection groups.

| Contig | Gene | Domain/Family | TM | SP | Blast Hit | Accession numbe | Evalue | Predicted function |
|---|---|---|---|---|---|---|---|---|
| comp67490_c0_seq1 | Niemann-Pick C 2 Like | ML (MD-2-related lipid-recognition) domain | Yes | No | *Anemonia viridis* | CDJ55918 | 7.00E-21 | Cholesterol transport |
| comp66887_c0_seq1 | Epidermal retinol dehydrogenase 2 | NADB domain | No | No | *Nematostella vectensis* | XP_001623156 | 1.00E-139 | Oxidoreductase |
| comp46570_c0_seq1 | RING finger protein 212B | N/A | No | No | *Nematostella vectensis* | XP_001638276 | 8.00E-29 | Meiotic recombination |
| comp80329_c2_seq2 | Methyltransferase type 11 | Methyltransferase domain | No | No | *Nematostella vectensis* | XP_001637968 | 4.00E-41 | Methyltransferase |
| comp92750_c0_seq1 | Protein phosphatase 1 regulatory subunit 3E | Carbohydrate binding domain | No | No | *Lepisosteus oculatus* | XP_006641290 | 1.00E-31 | Glycogen metabolizm |
| comp81225_c0_seq1 | C012-F12/Trefoil factor 2 | Trefoil domain | No | Yes | *Acropora millepora* | AET09734 | 1.00E-17 | N/A |
| comp36128_c0_seq1 | Hypothetical protein | N/A | No | No | *Aureococcus anophagefferens* | XP_009037845 | 6.00E-26 | N/A |
| comp80659_c0_seq1 | Mid1-interacting protein 1-B-like | Thyroid hormone-inducible hepatic protein Spot 14 | No | No | *Nematostella vectensis* | XP_001641632 | 8.00E-40 | Induction of hepatic lipogenesis |
| comp77373_c0_seq1 | Prestin | SulP family | Yes (12) | No | *Nematostella vectensis* | XP_001641754 | 0.00E+00 | Bicarbonate transport |
| comp81878_c0_seq1 | Cation transport regulator-like protein 1 | GGCT-like domain | No | No | *Nematostella vectensis* | XP_001632355 | 1.00E-76 | Inorganic ion transport and metabolism |
| comp80166_c4_seq5 | Fibrillin-1 | C-type lectin-like domain, H-type lectin domain, Kazal type serine protease inhibitors and follistatin-like domains | No | Yes | *Nematostella vectensis* | XP_001641556 | 2.00E-147 | Microfibrils formation |
| comp80835_c0_seq1 | Hypothetical protein | N/A | No | Yes | N/A | N/A | N/A | N/A |
| comp25774_c0_seq1 | Transmembrane protein 163 | Cation efflux family | Yes (6) | No | *Seriatopora hystrix* | ADJ94114 | 2.00E-28 | Cation efflux |
| comp59157_c0_seq1 | Hypothetical protein | N/A | No | Yes | *Cedecea neteri* | WP_039290972 | 9.00E-36 | N/A |
| comp73954_c0_seq1 | Hypothetical protein | N/A | No | No | *Necator americanus* | ETN69442 | 3.00E-14 | N/A |
| comp73878_c0_seq1 | Solute carrier family 25 member 36 | Mitochondrial carrier protein | No | No | *Nematostella vectensis* | XP_001641850 | 4.00E-48 | Pyrimidine transport |
| comp71785_c0_seq1 | Battenin | Nucleoside transporters | Yes (9) | No | *Nematostella vectensis* | XP_001640256 | 5.00E-66 | Nucleotide transport |
| comp80594_c0_seq1 | Lipid storage droplets surface-binding protein 2 | Perilipin family | No | No | *Nematostella vectensis* | XP_001623054 | 2.00E-123 | Lipid storage |
| comp77093_c2_seq3 | Ammonium transporter Rh type C | Ammonium Transporter Family | Yes (12) | Yes | *Branchiostoma floridae* | BAJ10273 | 6.00E-166 | Ammonium transporter |
| comp17315_c0_seq1 | Hypothetical protein | N/A | Yes (1) | No | N/A | N/A | N/A | N/A |
| comp81403_c0_seq1 | Hypothetical protein | N/A | No | No | N/A | N/A | N/A | N/A |
| comp88899_c0_seq1 | Retinal rod rhodopsin-sensitive cGMP 3',5'-cyclic phosphodiesterase subunit delta (PDE6D) | GMP-PDE delta subunit | No | No | *Danio rerio* | NP_001002708 | 6.00E-53 | Prenylated target proteins release |
| comp82357_c0_seq1 | Hypothetical protein | N/A | No | No | N/A | N/A | N/A | N/A |
| comp21509_c0_seq1 | Hypothetical protein | N/A | No | No | N/A | N/A | N/A | N/A |
| comp83997_c0_seq1 | Hypothetical protein | S1-like cold-shock domain | No | No | N/A | N/A | N/A | N/A |
| comp73494_c0_seq1 | Short-chain dehydrogenase/reductase family 9C member 7 | Rossmann-fold NAD(P)H/NAD(P)(+) binding (NADB) domain | No | No | *Nematostella vectensis* | XP_001626221 | 9.00E-26 | Retinol dehydrogenase activity |
| comp63712_c0_seq1 | Hypothetical protein | N/A | No | No | N/A | N/A | N/A | N/A |
| comp37834_c0_seq1 | Hypothetical protein | N/A | No | No | N/A | N/A | N/A | N/A |
| comp76481_c0_seq1 | Protein phosphatase 1 regulatory subunit 3B | Carbohydrate binding domain | No | No | *Nematostella vectensis* | XP_001629778 | 1.00E-44 | Glycogen metabolizm |
| comp82341_c0_seq1 | F-box/LRR-repeat protein 7 | F-box domain | No | No | *Nematostella vectensis* | XP_001635637 | 2.00E-88 | Signal transduction regulation |
| comp17700_c0_seq1 | Hypothetical protein | N/A | No | No | N/A | N/A | N/A | N/A |
| comp93444_c0_seq1 | Hypothetical protein | N/A | No | No | N/A | N/A | N/A | N/A |
| comp75198_c1_seq4 | Hypothetical protein | BRICHOS domain | No | Yes | *Branchiostoma floridae* | XP_002602912 | 3.00E-12 | N/A |
| comp18562_c0_seq1 | Peptidyl-prolyl cis-trans isomerase FKBP9 | EF-hand domain pair | No | Yes | *Nematostella vectensis* | XP_001636824 | 5.00E-69 | Calcium ion binding, protein folding |
| comp71344_c0_seq1 | Sulfotransferase | Sulfotransferase domain | Yes (1) | No | *Nematostella vectensis* | XP_001623762 | 2.00E-54 | Sulphate transport |
| comp30818_c0_seq1 | Hypothetical protein | N/A | No | No | N/A | N/A | N/A | N/A |
| comp76089_c0_seq1 | 1,4-alpha-glucan-branching enzyme | 1,4-alpha-glucan branching enzyme | No | No | *Nematostella vectensis* | XP_001634433 | 0 | Carbohydrate transport and metabolism |
| comp77383_c0_seq1 | Krueppel-like factor 11 | Zinc-finger double domain | No | No | *Anolis carolinensis* | XP_008115786 | 9.00E-62 | DNA binding |
| comp73494_c1_seq2 | Short-chain dehydrogenase/reductase family 9C member 7 | Rossmann-fold NAD(P)H/NAD(P)(+) binding (NADB) domain | No | No | *Nematostella vectensis* | XP_001626221 | 9.00E-26 | Retinol dehydrogenase activity |
| comp74074_c0_seq2 | Tumor necrosis factor receptor superfamily member 16 | TNFR superfamily | Yes | No | *Nematostella vectensis* | XP_001637775 | 1.00E-13 | Apoptosis |

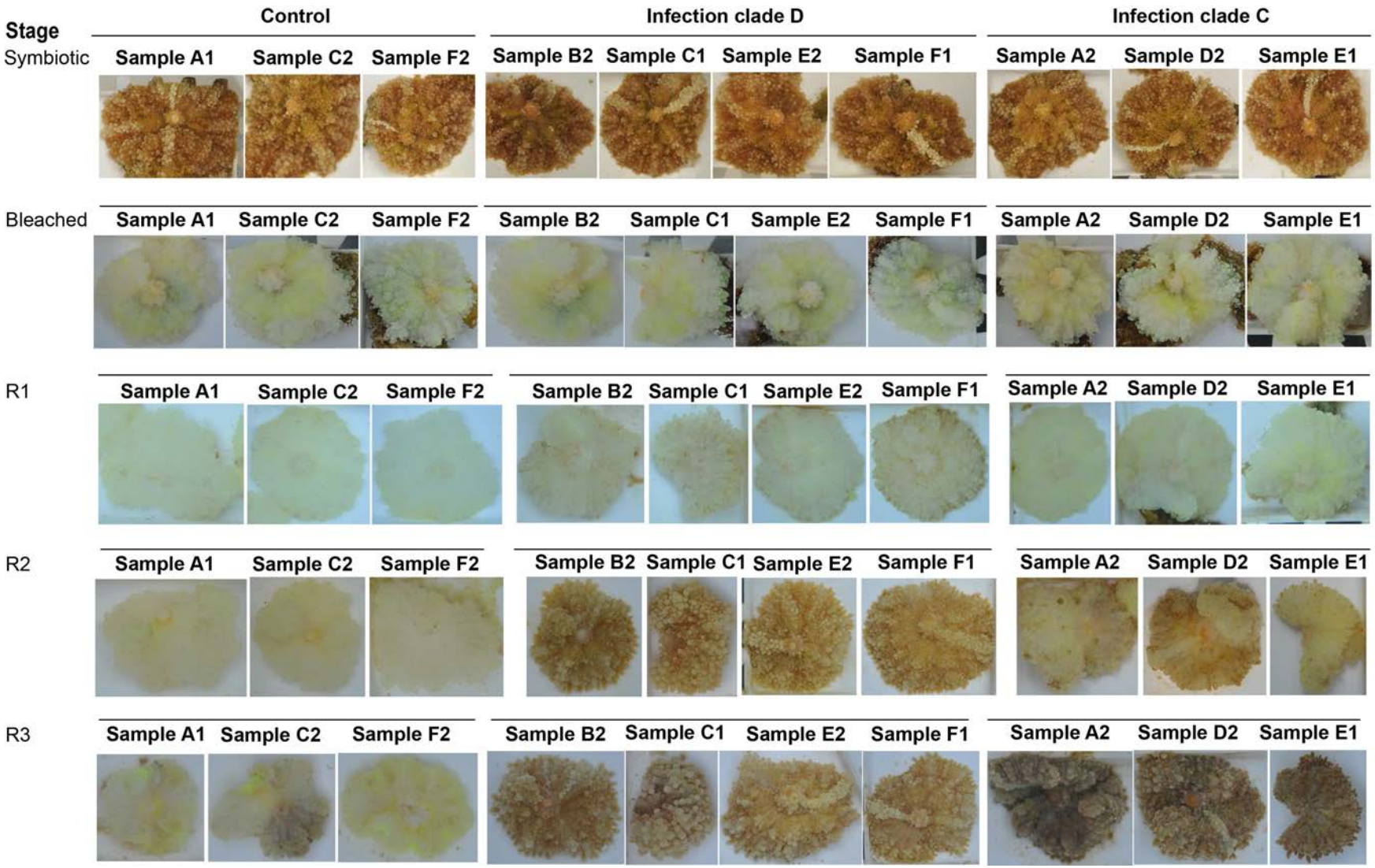



**Figure S4.1** The samples in each group throughout the experiment.

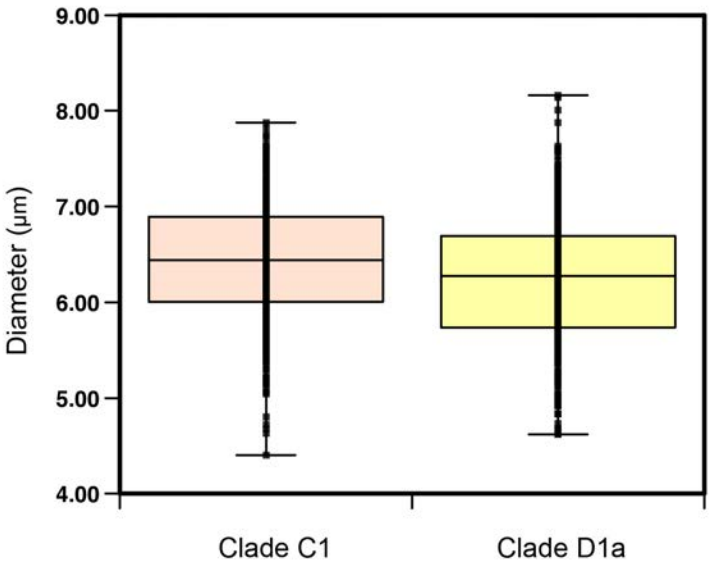



**Figure S4.2** The box plot of zooxanthellae diameter of two different *Symbiodinium* clades. Number of examinated samples from Clade C1 and Clade D1a are 426 and 418, respectively.


**Figure S4.3** Alignment of ITS2 sequences from *Symbiodinium* D1a in *Ricorde*a (D group) and *Rhodactis* (donor). (-) indicates ambiguous sites, (−) indicates polymorphic sites.

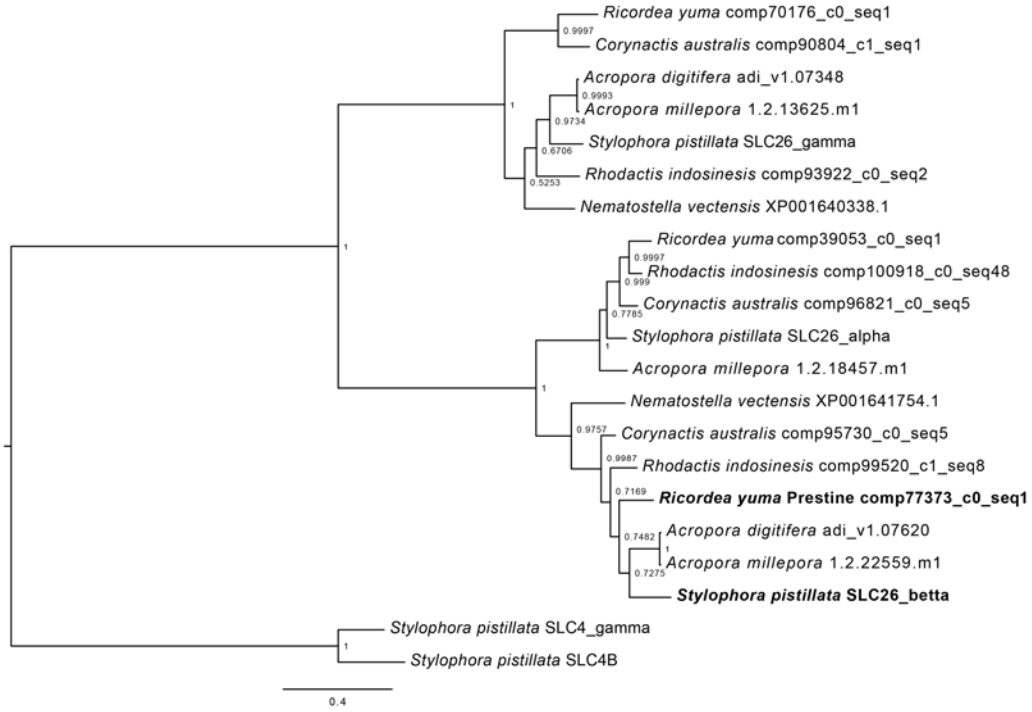



**Figure S4.4** Phylogenetic tree of SLC26 sequences inferred by Bayesian analysis with best-fit model of WAG+G for 2 million generations with first 2000 trees as burnin. The *Ricordea* Pristine (SLC26) and *Stylophora* SLC26β sequences are indicated in bold.

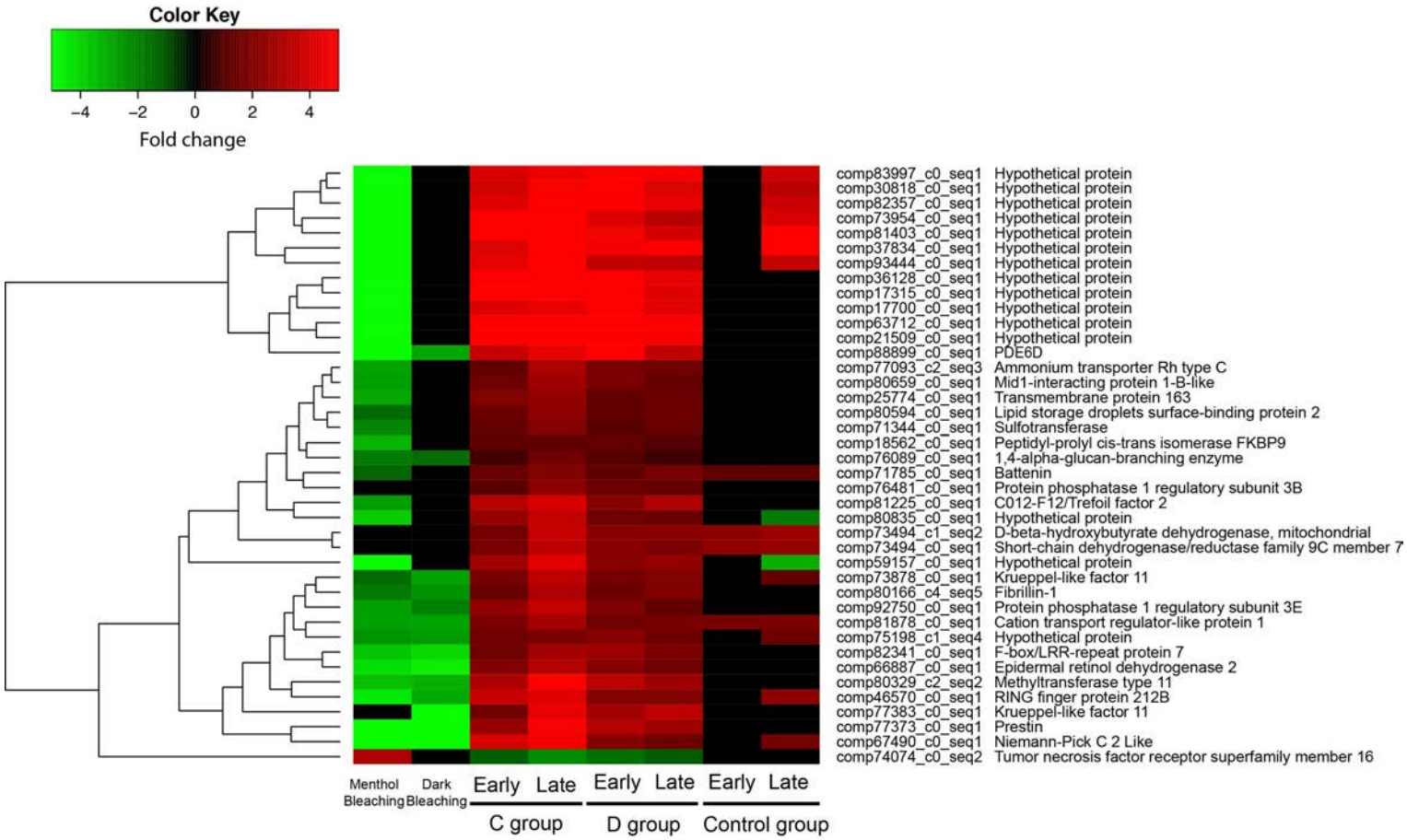



**Figure S4.5** Heatmap summarising differential expression of 40 common genes present in the C and D treatment groups and bleaching states induced by menthol and darkness treatments (data shown in Lin et al. in prep.).

# Chapter V General discussion and summary of conclusions

*In this chapter, the major findings of each chapter are summarised. A general discussion, conclusions and directions for future research are also included herein.*

>Summary of major findings

Evolutionary relationships between corals and corallimorpharians - from the mitochondrial genome to the large-scale nuclear data

In this PhD dissertation, phylogenetic relationships between scleractinians and corallimorpharians were clarified using phylogenomic methods (Chapter 2). This approach resolved the relationship between the two orders, which was previously equivocal as a consequence of being primarily based on mitochondrial (mt) genome data (Medina et al. 2006, Kitahara et al. 2014, Lin et al. 2014).

The dataset used in phylogenomic analyses comprised a total of 291 orthologous genes from 15 anthozoans, and included scleractinians, corallimorpharians, actiniarians and an octocoral. The results of the application of a wide range of phylogenetic approaches provide strong support for the scenario of scleractinian monophyly, and allow the rejection of the alternative “naked corals” hypothesis. This study indicates that scleractinians and corallimorpharians share a common ancestor, and the coral skeleton was gained in the scleractinian lineage after the scleractinian and corallimorpharian split. Thus corallimorpharians are not

descended from scleractinians by skeleton loss, as was proposed by Medina et al. (2006).

Corallimorpharian transcriptomes and perspectives on the evolution of coral calcification

Three corallimorpharian transcriptomes - two from symbiotic species and one from a nonsymbiotic species - were generated and analyzed in this doctoral dissertation (Chapter 3). These transcriptome assemblies are comparable in quality to those presently available for other anthozoans, and have provided important insights into the evolution of coral calcification. The survey conducted indicated that the corallimorpharians contain many genes present in both calcifying and non-calcifying anthozoans. The most surprising finding was that corallimorpharians also contain several genes that were previously considered to be coral-specific. These results indicate that the evolution of coral calcification required the acquisition of only a few new genes. Also, analysis of the repertoires of carbonic anhydrases (CAs), which are important for carbonate transport, suggests that the evolution of symbiosis may not have required novel CAs.

The application of transcriptomics to understanding cnidarian-algae symbioses

Chapter 4 provides transcriptomic data on the re-establishment of symbiosis in an aposymbiotic corallimorpharian. Some genes previously implicated in the symbiotic

state in corals and sea anemones were up-regulated during the establishment of symbiosis in the corallimorpharian, including NPC2, ammonium transporter Rh type C, and lipid storage droplet surface-binding protein. Moreover, these analyses highlight the likely importance of glycogen biosynthesis during the establishment of symbiosis. Comparison of gene expression profiles between hosts infected with homologous or heterologous *Symbiodinium* strains showed that up-regulation of glycogen biosynthesis in the host was significantly higher in the case of infection with the homologous strain. Additionally, the data suggest that immune responses were suppressed and genes involved in translocation of nutrients such as glycogen, lipid, cholesterol and ammonia up-regulated at the onset of symbiosis. Intriguingly, several up-regulated genes (hypothetical proteins) were restricted to the symbiotic corallimorpharians, suggesting that the symbiotic state may have evolved independently in this lineage.

>Major conclusions

The results presented indicate that corallimorpharians are the sister clade of scleractinians on the basis of phylogenomic analyses and transcriptomic data. These findings have two major implications. First, corallimorpharians are not scleractinians that have undergone skeleton loss. This indicates that calcification arose in the coral lineage after the divergence between corallimorpharians and scleractinians. Hence, although at least some corals can survive and fully recover from skeleton loss induced by incubation in acidic seawater (Fine and Tchernov,

2007), the monophyly of extant Scleractinia indicates that skeleton-less corals are unlikely to be viable on evolutionary time scales. Since the soft-bodied form has not been present in evolutionary history, the scleractinian lineage may therefore be more vulnerable to climate change and ocean acidification than predicted under the "naked corals" hypothesis. Second, finding "scleractinian-specific" genes in corallimorpharians not only supports the close relationships of these lineages but also implies that relatively few new genes are required to enable calcification.

> Other questions and directions for future research

1. Did symbiosis evolve independently in corallimorpharians?

Cnidarian-algal symbioses are thought to be predominantly managed by pre-existing repertoires of genes rather than by "new" genes uniquely required for symbiosis (Meyer and Weis 2012). The discovery of corallimorpharian genes that appear to be unique to symbiotic members of this Order, and which are up-regulated in symbiosis, directly challenges this statement. The corallimorpharian "symbiosis-specific" genes encode uncharacterized proteins, whose functions remain to be explored. Broader scale sampling of corallimorpharians and their relatives should be carried out to validate the link implied by the work presented here. Ultimately, some form of gene knock-down or interference should be applied to investigate the roles of these genes in symbiosis.

2. What are the functions of corallimorpharian homologs of coral

calcification-related genes?

At last some genes implicated in coral calcification, either by proteomics having demonstrated that they are in the skeleton, or *in situ* hybridization patterns consistent with such a role, have clear homologs in corallimorpharians. It would be of interest to examine exactly where and when these genes are expressed in corallimorpharians, in order to better understand how genes have been coopted to calcification. A corallimorpharian genome sequencing project might also reveal additional "coral-specific" genes, and would add new perspectives on genome evolution in "lower" animals.

3. Developmental biology of corallimorpharians

There appears to be a correlation between phylogeny and developmental mode in the Scleractinia, and it is not clear which type of development is ancestral. A series of papers have described the early development of *Acropora* (Miller and Ball 2000, Grasso et al. 2008, Hayward et al. 2011), which features a characteristic pre-gastrulation "prawn chip" stage that is essentially a flattened bilayer of cells. This kind of development appears to be typical of complex corals, whereas no "prawn chip' stage is seen in representatives of the robust clade – the other major division within the Scleractinia (Okubu et al. 2016). Description of the early development of corallimorpharians should provide new perspectives on the evolution of developmental mechanisms in corals and their relatives, as well as being important for understanding many aspects of corallimorpharian biology.

## References

Akanni W. A., C. J. Creevey, et al. (2014). L.U.St: a tool for approximated maximum likelihood supertree reconstruction. BMC Bioinformatics 15(1): 183.

Allman G. J. (1846). Description of a new genus of Helianthoid Zoophytes. The Annals and Magazine of Natural History 17: 417-419.

Ane C., B. Larget, et al. (2006). Bayesian estimation of concordance among gene trees. Molecular Biology and Evolution 24(2): 412-426.

Aranda M., M. K. DeSalvo, et al. (2012). Evolutionary insights into scleractinian corals using comparative genomic hybridizations. BMC Genomics 13: 501.

Arthur C. P. and M. H. B. Stowell (2007). Structure of Synaptophysin: a hexameric MARVEL-domain channel protein. Structure 15(6): 707-714.

Babbey C. M., N. Ahktar, et al. (2006). Rab10 regulates membrane transport through early endosomes of polarized Madin-Darby canine kidney cells. Molecular Biology of the Cell 17(7): 3156-3175.

Baird A. H., V. R. Cumbo, et al. (2007). Fidelity and flexibility in coral symbioses. Marine Ecology Progress Series 347: 307-309.

Baird A. H., R. Bhagooli, et al. (2009). Coral bleaching: the role of the host. Trends in Ecology and Evolution 24(1): 16-20.

Baker A. C. (2003). Flexibility and specificity in coral-algal symbiosis: diversity, ecology, and biogeography of *Symbiodinium*. Annual Review of Ecology, Evolution, and Systematics 34(1): 661-689.

Barbeitos M. S., S. L. Romano, et al. (2010). Repeated loss of coloniality and symbiosis in scleractinian corals. Proceedings of the National Academy of Sciences 107(26): 11877-11882.

Barott K. L., A. A. Venn, et al. (2015). Coral host cells acidify symbiotic algal microenvironment to promote photosynthesis. Proceedings of the National Academy of Sciences 112(2): 607-612.

Barshis D. J., J. T. Ladner, et al. (2013). Genomic basis for coral resilience to climate change. Proceedings of the National Academy of Sciences of the United States of America 110: 1387-1392.

Baum D. A. (2007). Concordance trees, concordance factors, and the exploration of reticulate genealogy. Taxon 56(2): 417-426.

Baumgarten S., O. Simakov, et al. (2015). The genome of *Aiptasia*, a sea anemone model for coral symbiosis. Proceedings of the National Academy of Sciences 112: 11893-11898.

Bayer T., M. Aranda, et al. (2012). *Symbiodinium* transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals. PLoS One 7(4): e35269.

Bazinet A. L., M. P. Cummings, et al. (2013). Can RNA-Seq resolve the rapid radiation of advanced moths and butterflies (Hexapoda: Lepidoptera: Apoditrysia)? an exploratory study. PLoS One 8(12): e82615.

Bertucci A., S. Forêt, et al. (2015). Transcriptomic differences between day and night in *Acropora millepora* provide new insights into metabolite exchange and light-enhanced calcification in corals. Molecular Ecology 24(17): 4489-4504.

Bertucci A., A. Moya, et al. (2013). Carbonic anhydrases in anthozoan corals-a review. Bioorganic and Medicinal Chemistry 21(6): 1437-1450.

Bertucci A., E. Tambutte, et al. (2009). Symbiosis-dependent gene expression in coral-dinoflagellate association: cloning and characterization of a P-type H+-ATPase gene. Proceedings of the Royal Society B: Biological Sciences 277(1678): 87-95.

Bremer J. (1983). Carnitine-metabolism and functions. Physiological Reviews 63:

1420-1480.

Budd A. F., S. L. Romano, et al. (2010). Rethinking the phylogeny of scleractinian corals: a review of morphological and molecular data. Integrative and Comparative Biology 50(3): 411-427.

Buddemeier R. W. and Fautin D. G. (1993). Coral bleaching as an adaptive mechanism. BioScience 43(5): 320-326.

Burriesci M. S., T. K. Raab, et al. (2012). Evidence that glucose is the major transferred metabolite in dinoflagellate-cnidarian symbiosis. Journal of Experimental Biology 215(19): 3467-3477.

Cairns, S. D. (1999). Species richness of recent Scleractinia. Atoll Research Bullentin 459: 1-46.

Calnan B. J., S. Biancalana, et al. (1990). Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition. Genes and Development 5: 201-210.

Capella-Gutierrez S., J. M. Silla-Martinez, et al. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15): 1972-1973.

Carlgren O. (1900). Ostafrikanische Actinien. Jahrbuch Der Hamburgischen Wissenschaftlichen Anstalten 17(2): 21-144.

Carlgren O. (1943). East-Asiatic Corallimorpharia and Actiniaria. Kungl. Svenska vetenskapsakademiens handlingar 20(6): 1-43.

Chadwick-Furman N. E. and M. Spiegel (2000). Abundance and clonal replication in theh tropical corallimorpharian *Rhodactis rhodostoma*. Invertebrate Biology 119(4): 351-360.

Chapman J. A., E. F. Kirkness, et al. (2010). The dynamic genome of *Hydra*. Nature

464(7288): 592-596.

Chen C. and J.-K. Yu (2000). Universal primers for amplification of mitochondrial small subunit ribosomal RNA-encoding gene in scleractinian corals. Marine Biotechnology 2: 146-153.

Chen C. A., C. C. Wallace, et al. (2002). Analysis of the mitochondrial 12S rRNA gene supports a two-clade hypothesis of the evolutionary history of scleractinian corals. Molecular Phylogenetics and Evolution 23(2): 137-149.

Chen C. C. H., P. J. Schweinsberg, et al. (2006). RAB-10 is required for endocytic recycling in the *Caenorhabditis elegans* intestine. Molecular Biology of the Cell 17(3): 1286-1297.

Chera S., W. Buzgariu, et al. (2009). Autophagy in *Hydra*: A response to starvation and stress in early animal evolution. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1793(9): 1432-1443.

Chomczynski P. and N. Sacchi (1987). Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate Phenol Chloroform Extraction. Analytical Biochemistry 162(1): 156-159.

Coffroth M. A., D. M. Poland, et al. (2010). Environmental symbiont acquisition may not be the solution to warming seas for reef-building corals. PLoS One 5(10): e13258.

Cohen A. L. and M. Holcomb (2009). Why corals care about ocean acidification-uncovering the mechanism. Oceanography 22: 118-128.

Combet C., C. Blanchet, et al. (2000). NPS@: network protein sequence analysis. Trends in Biochemicl Sciences 25: 147-150.

Constantz B. R. (1986). Coral skeleton construction: a physiochemically dominated process. PALAIOS 1: 152-157.

Coon T. A., J. R. Glasser, et al. (2012). Novel E3 ligase component FBXL7 ubiquitinates and degrades Aurora A, causing mitotic arrest. Cell Cycle 11(4): 721-729.

Cuif J.-P. (2014). The Rugosa-Scleractinia gap re-examined through microstructural and biochemical evidence: a tribute to H.C. Wang. Palaeoworld 23(1): 1-14.

Cumbo V. R., A. H. Baird, et al. (2013). *Chromera velia* is endosymbiotic in larvae of the reef corals *Acropora digitifera* and *A. tenuis*. Protist 164: 237-244.

Dale C. (2003). Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration. Molecular Biology and Evolution 20(8): 1188-1194.

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

Daly M., D. G. Fautin, et al. (2003). Systematics of the Hexacorallia (Cnidaria : Anthozoa). Zoological Journal of the Linnean Society 139(3): 419-437.

Dani V., P. Ganot, et al. (2014). Are Niemann-Pick type C proteins key players in cnidarian-dinoflagellate endosymbioses? Molecular Ecology 23(18): 4527-4540.

Darriba D., G. L. Taboada, et al. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27: 1164-1165.

Darriba D., G. L. Taboada, et al. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9(8): 772.

Davy S. K., D. Allemand, et al. (2012). Cell biology of cnidarian-dinoflagellate symbiosis. Microbiology and Molecular Biology Reviews 76(2): 229-261.

den Hartog J. C. (1980). Caribbean shallow water Corallimorpharia. Zoologische Verhandelingen 176: 1-83.

den Hartog J. C., O. Ocaña, et al. (1993). Corallimorpharia collected during the CANCAP expeditions (1976-1986) in the south-eastern part of the North Atlantic. Zoologische Verhandelingen 282: 1-58.

DeSalvo M. K., A. Estrada, et al. (2012). Transcriptomic responses to darkness stress point to common coral bleaching mechanisms. Coral Reefs 31(1): 215-228.

DeSalvo M. K., S. Sunagawa, et al. (2010). Coral host transcriptomic states are correlated with *Symbiodinium* genotypes. Molecular Ecology 19(6): 1174-1186.

DeSalvo M. K., C. R. Voolstra, et al. (2008). Differential gene expression during thermal stress and bleaching in the Caribbean coral *Montastraea faveolata*. Molecular Ecology 17(17): 3952-3971.

Detournay O. and V. M. Weis (2011). Role of the Sphingosine Rheostat in the regulation of Cnidarian-Dinoflagellate symbioses. Biological Bulletin 221: 261-269.

Dimond J. L., B. J. Holzman, et al. (2012). Thicker host tissues moderate light stress in a cnidarian endosymbiont. Journal of Experimental Biology 215(13): 2247-2254.

Done T. J. (1999). Coral community adaptability to environmental change at the scales of regions, reefs and reef zones. American Zoologist 39(1): 66-79.

Downs C. A., K. E. McDougall, et al. (2013). Heat-stress and light-stress induce different cellular pathologies in the symbiotic dinoflagellate during coral bleaching. PLoS One 8(12): e77173.

Drake J. L., T. Mass, et al. (2013). Proteomic analysis of skeletal organic matrix from the stony coral Stylophora pistillata. Proceedings of the National Academy of Sciences 110: 3788–3793.

Duchassaing P. and J. Michelotti (1860). Memoire sur les Coralliaires des Antilles. Memorie della Reale academia delle scienze di Torino 19:279-365.

Dudler D. and D. Miller (1988). Characterization of two glutamate dehydrogenases

from the symbiotic microalga *Symbiodinium microadriaticum* isolated from the coral *Acropora formosa*. Marine Biology 97: 427-430.

Dujon B., D. Sherman, et al. (2004). Genome evolution in yeasts. Nature 430: 35-44.

Dunlap W. C., A. Starcevic, et al. (2013). KEGG orthology-based annotation of the predicted proteome of *Acropora digitifera*: ZoophyteBase - an open access and searchable database of a coral genome. BMC Genomics 14: 509.

Dunn S. R. and V. M. Weis (2009). Apoptosis as a post-phagocytic winnowing mechanism in a coral-dinoflagellate mutualism. Environmental Microbiology 11(1): 268-276.

Ebersberger I., S. Strauss, et al. (2009). HaMStR: Profile hidden markov model based search for orthologs in ESTs. BMC Evolutionary Biology 9(1): 157.

Edgar G. J. (2009). Australian marine life: the plants and animals of temperate waters. New Holland Publishers.

Emanuelsson O., H. Nielsen, et al. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of Molecular Biology 300(4): 1005-1016.

Enríquez S., E. R. Méndez, et al. (2005). Multiple scattering on coral skeletons enhances light absorption by symbiotic algae. Limnology and Oceanography 50: 1025–1032.

Evans E. C. (1977). Microcosm responses to environmental perturbants - extension of baseline field survey. Helgolander Wissenschaftliche Meeresuntersuchungen 30(1-4): 178-191.

Exposito J. Y., C. Larroux, et al. (2008). Demosponge and sea anemone fibrillar collagen diversity reveals the early emergence of A/C Clades and the maintenance of the modular structure of type V/XI collagens from sponge to human. Journal of

Biological Chemistry 283(42): 28226-28235.

Ezaki Y. (1997). The Permian coral *Numidiaphyllum*: new insights into anthozoan phylogeny and Triassic scleractinian origins. Palaeontology 40: 1-40.

Ezaki Y. (1998). Paleozoic Scleratinia: progenitors or extinct experiments? Paleobiology 24: 227-234.

Ezaki Y. (2000). Palaeoecological and phylogentic implications of a new scleractinian genus from Permian sponge reefs, South China. Paleontology 78(1): 84-97.

Fautin D. G. (2011). *Corallimorphus niwa* new species (Cnidaria: Anthozoa), New Zealand members of *Corallimorphus*, and redefinition of Corallimorphidae and its members. Zootaxa 2775: 37-49.

Febbraio M., D. P. Hajjar, et al. (2001). CD36: a class B scavenger receptor involved in angiogenesis, atherosclerosis, inflammation, and lipid metabolism. Journal of Clinical Investigation 108(6): 785-791.

Fine M. and D. Tchernov (2007). Scleractinian coral species survive and recover from decalcification. Science 316(5820): 1811.

Fitzpatrick D. A., M. E. Logue, et al. (2006). A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. BMC Evolutionary Biology 6(1): 99.

Fleischmann T. T., L. B. Scharff, et al. (2011). Nonessential plastid-encoded Ribosomal proteins in Tobacco: a developmental role for plastid translation and implications for reductive genome evolution. The Plant Cell 23(9): 3137-3155.

Frade J. M., A. Rodríguez-Tébar, et al. (1996). Induction of cell death by endogenous nerve growth factor through its p75 receptor. Nature 383: 166-168.

Fukami H., C. A. Chen, et al. (2008). Mitochondrial and nuclear genes suggest that

stony corals are monophyletic but most families of stony corals are not (Order Scleractinia, Class Anthozoa, Phylum Cnidaria). PLoS One 3(9): e3222.

Fukami H. and N. Knowlton (2005). Analysis of complete mitochondrial DNA sequences of three members of the *Montastraea annularis* coral species complex (Cnidaria, Anthozoa, Scleractinia). Coral Reefs 24(3): 410-417.

Fukuda I., S. Ooki, et al. (2003). Molecular cloning of a cDNA encoding a soluble protein in the coral exoskeleton. Biochemical and Biophysical Research Communications 304(1): 11-17.

Furla P., D. Allemand et al. (2005). The symbiotic anthozoan: a physiological chimera between alga and animal. Integrative and Comparative Biology 45(4): 595-604.

Furla P., I. Galgani, et al. (2000). Sources and mechanisms of inorganic carbon transport for coral calcification and photosynthesis. The Journal of Experimental Biology 203: 3445-3457.

Ganot P., A. Moya, et al. (2011). Adaptations to endosymbiosis in a cnidarian-dinoflagellate association: differential gene expression and specific gene Duplications. PLoS Genetics 7(7): e1002187.

Gates R. D., G. Baghdasarian, et al. (1992). Temperature stress causes host-cell detachment in symbiotic cnidarians - implications for coral bleaching. Biological Bulletin 182(3): 324-332.

Glick D., S. Barth, et al. (2010). Autophagy: cellular and molecular mechanisms. The Journal of Pathology 221(1): 3-12.

Glynn P. W. (1993). Coral reef bleaching: ecological perspectives. Coral Reefs 12: 1-17.

Glynn P. W. (1996). Coral reef bleaching: facts, hypotheses and implications. Global Change Biology 2(6): 495-509.

Glynn P. W. (2000). El Niño-Southern Oscillation mass mortalities of reef corals: a model of high temperature marine extinctions? Carbonate platform systems: components and interactions. P. P. W. Skelton, T. J. London, Geological Society of London: 117-133.

Glynn P. W., L. S. Howard, et al. (1984). The occurrence and toxicity of herbicides in reef building corals. Marine Pollution Bulletin 15(10): 370-374.

Goulet T. L. (2006). Most corals may not change their symbionts. Marine Ecology Progress Series 321:1-7.

Goreau, T. J. and R. L. Hayes (1994). Coral bleaching and ocean hot-spots. Ambio 23(3): 176-180.

Grabherr M. G., B. J. Haas, et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29(7): 644-652.

Grasso L. C., J. Maindonald, et al. (2008). Microarray analysis identifies candidate genes for key roles in coral development. BMC Genomics 9: 540.

Guindon S., J. Dufayard, et al. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology 59(3): 307-321.

Guo D. F., Y. L. Sun, et al. (2001). The angiotensin II type 1 receptor and receptor-associated proteins. Cell Research 11: 165-180.

Haas B. J., A. Papanicolaou, et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8: 1494-1512.

Haddon A. C. and Duerden J. E. (1896). On some Actiniaria from Australia and other districts. Scientific Proceedings of The Royal Dublin Society 6(16): 139-164.

Hahn S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nature Structural and Molecular Biology 11(5): 394-403.

Hall T. A. (1999). BioEdit: a user-friendly biological sequence alignment program for Windows 95/98/NT. Nucleic Acids Symposium Seriers 41: 95-98.

Hamada M., E. Shoguchi, et al. (2013). The complex NOD-Like receptor repertoire of the coral *Acropora digitifera* includes novel domain combinations. Molecular Biology and Evolution 30(1): 167-176.

Hamner W. M. and D. F. Dunn (1980). Tropical Corallimorpharia (Coelenterata: Anthozoa): feeding by envelopment. Micronesica 16: 37-41.

Hawkins T. D., B. J. Bradley, et al. (2013). Nitric oxide mediates coral bleaching through an apoptotic-like cell death pathway: evidence from a model sea anemone-dinoflagellate symbiosis. The FASEB Journal 27(12): 4790-4798.

Hawkins T. D., T. Krueger, et al. (2014). Differential nitric oxide synthesis and host apoptotic events correlate with bleaching susceptibility in reef corals. Coral Reefs 33(1): 141-153.

Hayward D. C., S. Hetherington, et al. (2011). Differential gene expression at coral settlement and metamorphosis - a subtractive hybridization study. PLoS One 6(10): e26411.

Heath-Heckman E. A., A. A. Gillette, et al. (2014). Shaping the microenvironment: evidence for the influence of a host galaxin on symbiont acquisition and maintenance in the squid-vibrio symbiosis. Environmental Microbiology 16(12): 3669-3682.

Hedin M., J. Starrett, et al. (2010). Phylogenomic resolution of Paleozoic divergences in Harvestmen (Arachnida, Opiliones) via analysis of next-generation transcriptome data. PLoS One 7(8): e42888.

Hikima S., J. Hikima, et al. (2003). Characterization and function of kuruma shrimp

lysozyme possessing lytic activity against Vibrio species. Gene 316: 187-195.

Hittinger C. T., M. Johnston et al. (2010). Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. Proceedings of the National Academy of Sciences 107(4): 1476-1481.

Hoegh-Guldberg O. (1999). Climate change, coral bleaching and the future of the world's coral reefs. Marine and Freshwater Research 50(8): 839-866.

Hoegh-Guldberg O., P. J. Mumby et al. (2007). Coral reefs under rapid climate change and ocean acidification. Science 318(5857): 1737-1742.

Hoffman G. E., M. V. Puerta, et al. (2011). Evolution of light-harvesting complex proteins from Chl c-containing algae. BMC Evolutionary Biology 11(1): 101.

Holland B. S., M. N. Dawson, et al. (2004). Global phylogeography of *Cassiopea* (Scyphozoa: Rhizostomeae): molecular evidence for cryptic species and multiple invasions of the Hawaiian Islands. Marine Biology 145(6): 1119-1128.

Hong M.-C., Y.-S. Huang, et al. (2009). ApRab3, a biosynthetic Rab protein, accumulates on the maturing phagosomes and symbiosomes in the tropical sea anemone, *Aiptasia pulchella*. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 152(3): 249-259.

Howells E., R. Berkelmans, et al. (2013). Historical thermal regimes define limits to coral acclimation. Ecology 94(5): 1078-1088.

Hughes L. (2003). Climate change and Australia: trends, projections and impacts. Austral Ecology 28(4): 423-443.

Hughes T. P., A. H. Baird, et al. (2003). Climate change, human impacts, and the resilience of coral reefs. Science 301(5635): 929-933.

Iglesias-Prieto R., J. L. Matta, et al. (1992). Photosynthetic response to elevated-temperature in the symbiotic dinoflagellate

*Symbiodinium-Microadriaticum* in culture. Proceedings of the National Academy of Sciences of the United States of America 89(21): 10302-10305.

Isa Y. and K. Yamazato (1984). The distribution of carbonic anhydrase in a staghorn coral *Acropora hebes* (Dana). Galaxea 3: 25-36.

Ishikura M., K. Adachi, et al. (1999). Zooxanthellae release glucose in the tissue of a giant clam, *Tridacna crocea*. Marine Biology 133: 665-673.

Jackson D. J., L. Macis, et al. (2007). Sponge paleogenomics reveals an ancient role for carbonic anhydrase in skeletogenesis. Science 316 (5833): 1893-1895.

Jackson J. B. C., J. D. Cubit, et al. (1989). Ecological effects of a major oil spill on panamanian coastal marine communities. Science 243 (4887): 37-44.

Jahren H. A., C. N. Arens, et al. (2001). Terrestrial record of methane hydrate dissociation in the Early Cretaceous. Geology 29: 159-162.

Jensen S. A., G. Aspinall, et al. (2014). C-terminal propeptide is required for fibrillin-1 secretion and blocks premature assembly through linkage to domains cbEGF41-43. Proceedings of the National Academy of Sciences 111(28): 10155-10160.

Jones R. (2005). The ecotoxicological effects of Photosystem II herbicides on corals. Marine Pollution Bulletin 51(5-7): 495-506.

Kang S. K., Y. K. Chae, et al. (2008). Role of human Aquaporin 5 in colorectal carcinogenesis. The American Journal of Pathology 173(2): 518-525.

Katoh K. and D. M. Standley (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution 30(4): 772-780.

Kayal E., B. Roure, et al. (2013). Cnidarian phylogenetic relationships as revealed by mitogenomics. BMC Evolutionary Biology 13(5): 13:15.

Klueter A., J. B. Crandall, et al. (2015). Taxonomic and environmental variation of metabolite profiles in marine dinoflagellates of the genus *Symbiodinium*. Metabolites 5: 74-99.

Kiessling W., C. Simpson, et al. (2010). Reefs as cradles of evolution and sources of biodiversity in the Phanerozoic. Science 327(5962): 196-198.

Kimura A., E. Sakaguchi, et al. (2009). Multi-component complement system of Cnidaria: C3, Bf, and MASP genes expressed in the endodermal tissues of a sea anemone, *Nematostella vectensis*. Immunobiology 214:165-178.

Kinzie III R. A., M. Takayama, et al. (2001). The adaptive bleaching hypothesis: experimental tests of critical assumptions. Biological Bulletin 200: 51-58.

Kishino H. and M. Hasegawa (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. Journal of Molecular Evolution 29 (2): 170-179.

Kita A., M. Jimbo, et al. (2015). Crystal structure of a symbiosis-related lectin from octocoral. Glycobiology 25(9): 1016-1023.

Kitahara M. V., S. D. Cairns, et al. (2010). A comprehensive phylogenetic analysis of the Scleractinia (Cnidaria, Anthozoa) based on mitochondrial CO1 sequence data. PLoS One 5(7): e11490.

Kitahara M. V., M.-F. Lin, et al. (2014). The ‘‘Naked Coral’’ hypothesis revisited – evidence for and against scleractinian monophyly. PLoS One 9(4): e94774.

Kitahara M. V., J. Stolarski, et al. (2012). The first modern solitary Agariciidae (Anthozoa, Scleractinia) revealed by molecular and microstructural analysis. Invertebrate Systematics 26(3): 303.

Kopp C., I. Domart-Coulon, et al. (2015). Subcellular investigation of photosynthesis-driven carbon assimilation in the symbiotic reef coral *Pocillopora*

*damicornis*. mBio 6(1): e02299-02214.

Krogh A., B. Larsson, et al. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. Journal of Molecular Biology 305(3): 567-580.

Krueger T., P. L. Fisher, et al. (2015). Transcriptomic characterization of the enzymatic antioxidants FeSOD, MnSOD, APX and KatG in the dinoflagellate genus Symbiodinium. BMC Evolutionary Biology 15(1): 48.

Kuguru B., Y. Achituv, et al. (2010). Photoacclimation mechanisms of corallimorpharians on coral reefs: photosynthetic parameters of zooxanthellae and host cellular responses to variation in irradiance. Journal of Experimental Marine Biology and Ecology 394: 53-62.

Kuguru B., N. E. Chadwick, et al. (2008). Mechanisms of habitat segregation between corallimorpharians: photosynthetic parameters and *Symbiodinium* types. Marine Ecology Progress Series 369: 115-129.

Kuguru B., G. Winters, et al. (2007). Adaptation strategies of the corallimorpharian Rhodactis rhodostoma to irradiance and temperature. Marine Biology 151(4): 1287-1298.

Kuo J., M. C. Chen, et al. (2004). Comparative gene expression in the symbiotic and aposymbiotic Aiptasia pulchella by expressed sequence tag analysis. Biochemical and Biophysical Research Communications 318(1): 176-186.

Kuwahara H., Y. Takaki, et al. (2011). Loss of genes for DNA recombination and repair in the reductive genome evolution of thioautotrophic symbionts of Calyptogena clams. BMC Evolutionary Biology 11(1): 285.

Kuwahara H., T. Yoshida, et al. (2007). Reduced genome of the Thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. Current Biology 17(10): 881-886.

Kvennefors E. C. E., W. Leggat, et al. (2008). An ancient and variable mannose-binding lectin from the coral *Acropora millepora* binds both pathogens and symbionts. Developmental and Comparative Immunology 32(12): 1582-1592.

LaJeunesse, T. C. (2001). Investigating the biodiversity, ecology, and phylogeny of endosymbiotic dinoflagellates in the genus *Symbiodinium* using the ITS region: in search of a "species" level marker. Journal of Phycology 37: 866-880.

LaJeunesse, T. C. (2002). Diversity and community structure of symbiotic dinoflagellates from Caribbean coral reefs. Marine Biology 141: 387-400.

LaJeunesse, T. C. and R. K. Trench (2000). Biogeography of two species of *Symbiodinium* (Freudenthal) inhabiting the intertidal sea anemone *Anthopleura elegantissima* (Brandt). Biological Bulletin 199(2): 126-134.

Lanfear R., B. Calcott, et al. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29(6): 1695-1701.

Lange C. G. Hemmrich, et al. 2011. Defining the origins of the NOD-like receptor system at the base of animal evolution. Molecular Biology and Evolution 28 (5):1687-1702.

Langmead B. and S. L. Salzberg (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357-359.

Lartillot N. and H. Philippe (2004). A bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Molecular Biology and Evolution 21(6): 1095-1109.

Lartillot N., N. Rodrigue, et al. (2013). PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Systematic Biology 62(4): 611-615.

Le Q. S. and O. Gascuel (2008). An improved general amino acid replacement

matrix. Molecular Biology and Evolution 25(7): 1307-1320.

Leckie M. R., T. J. Bralower, et al. (2002). Oceanic anoxic events and plankton evolution: Biotic response to tectonic forcing during the mid-Cretaceous. Paleoceanography 17: 13.

Leggat W., F. Seneca, et al. (2011). Differential responses of the coral host and their algal symbiont to thermal stress. PLoS One 6(10): e26687.

Lehnert E. M., M. S. Burriesci, et al. (2012). Developing the anemone *Aiptasia* as a tractable model for cnidarian-dinoflagellate symbiosis: the transcriptome of aposymbiotic *A. pallida*. BMC Genomics 13(1): 271.

Lehnert E. M., M. E. Mouchka, et al. (2014). Extensive differences in gene expression between symbiotic and aposymbiotic cnidarians. G3 (Bethesda) 4(2): 277-295.

Lesser M. P. (2006). Oxidative stress in marine environments: biochemistry and physiological ecology. Annual Review of Physiology 68: 253-278.

Li G., J. Zhang, et al. (2009). The evolutionarily dynamic IFN-Inducible GTPase proteins play conserved immune functions in vertebrates and cephalochordates. Molecular Biology and Evolution 26(7): 1619-1630.

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

Lin M.-F., C. A. Chen, et al. (2013). Asexual reproduction by marginal budding in the tropical corallimorpharian, *Ricordea yuma* (Corallimorpharia; Ricordeidae). Galaxea, Journal of Coral Reef Studies 15: 41-42.

Lin M.-F., M. V. Kitahara, et al. (2012). Novel organization of the mitochondrial genome in the deep-sea coral, *Madrepora oculata* (Hexacorallia, Scleractinia, Oculinidae) and its taxonomic implications. Molecular Phylogenetics and Evolution 65(1): 323-328.

Lin M.-F., K. S. Luzon, et al. (2011). Seventy-four universal primers for chracterizing the complete mitochondrial genome of scleractinian corals (Cnidaria; Anthozoa). Zoological Studies 40(4): 513-524.

Lin M. F., M. V. Kitahara, et al. (2014). Mitochondrial genome rearrangements in the Scleractinia/Corallimorpharia complex: implications for coral phylogeny. Genome Biology and Evolution 6(5): 1086-1095.

Loram J. E., H. G. Trapido-Rosenthal, et al. (2007). Functional significance of genetically different symbiotic algae *Symbiodinium* in a coral reef symbiosis. Molecular Ecology 16(22): 4849-4857.

Luo W., M. S. Friedman, et al. (2009). GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics 10(1): 161.

MacMicking J. D. (2004). IFN-inducible GTPases and immunity to intracellular pathogens. Trends in Immunology 25(11): 601-609.

MacRae R. A., R. A. Fensome, et al. (1996). Fossil dinoflagellate diversity, originations, and extinctions and their significance. Canadian Journal of Botany 74: 1687-1694.

Marchler-Bauer A., M. Derbyshire, et al. (2015). CDD: NCBI's conserved domain database. Nucleic Acids Research 43(D1): D222-D226.

Martin E. J. (1960). Observations on the toxic sea anemone, *Rhodactis howesii* (Coelenterata). Pacific Science XIV: 403-407.

Maxwell K. and G. N. Johnson (2000). Chlorophyll fluorescence - a practical guide. Journal of Experimental Botany 51(345): 659-668.

Mayr E. and P. D. Ashlock (1991). Principles of systematic zoology. New York, McGraw-Hill.

Medina M., A. G. Collins, et al. (2006). Naked corals: Skeleton loss in Scleractinia. Proceedings of the National Academy of Sciences of the United States of America 103(24): 9096-9100.

Meggio F. and L. A. Pinna (2003). One-thousand-andone substrates of protein kinase CK2? The FASEB Journal 17: 349-368.

Meints R. and Pardy R. L. (1980). Quantitative demonstration of cell surface involvement in a plant-animal symbiosis: lectin inhibition of reassociation. Journal of Cell Science 43: 239-251.

Mendonça A. G., R. J. Alves, et al. (2011). Loss of genetic redundancy in reductive genome evolution. PLoS Computational Biology 7(2): e1001082.

Meusemann K., B. M. von Reumont, et al. (2010). A phylogenomic approach to resolve the Arthropod tree of life. Molecular Biology and Evolution 27(11): 2451-2464.

Meyer E. and V. M. Weis (2012). Study of cnidarian-algal symbiosis in the omics age. Biological Bulletin 223: 44-65.

Miller D. J., G. Hemmrich, et al. (2007). The innate immune repertoire in Cnidaria-ancestral complexity and stochastic gene loss. Genome Biolology 8(4): R59.

Miller D. and E. Ball (2000). The coral *Acropora*: what it can contribute to our knowledge of metazoan evolution and the evolution of developmental processes. BioEssays 22: 291-296.

Milne Edwards H. and J. Haime (1851). Monographie des Polypiers fossiles des terrains Palaeozoiques, précédée d'un tableau general de la classification des Polypes. Archives de Muséum d'Histoire Naturelle, Paris 5: 1-502.

Moberg F. and C. Folke (1999). Ecological goods and services of coral reef ecosystems. Ecological Economics 29(2): 215-233.

Mohamed A. R., V. Cumbo, et al. (2016). The transcriptomic response of the coral *Acropora digitifera* to a competent *Symbiodinium* strain: the symbiosome as an arrested early phagosome. Molecular Ecology. In press.

Moya A., P. Ganot, et al. (2012). The transcriptomic response to thermal stress is immediate, transient and potentiated by ultraviolet radiation in the sea anemone *Anemonia viridis*. Molecular Ecology 21(5): 1158-1174.

Moya A., L. Huisman, et al. (2012). Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to $CO_2$-driven acidification during the initiation of calcification. Molecular Ecology 21(10): 2440-2454.

Muhando C. A., B. L. Kuguru, et al. (2002). Environmental effects on the distribution of corallimorpharians in Tanzania. AMBIO: A Journal of the Human Environment 31(7): 558-561.

Nakayama K., A. Yamashita, et al. (2008). The whole-genome sequencing of the obligate intracellular bacterium *Orientia tsutsugamushi* revealed massive gene amplification during reductive genome evolution. DNA Research 15(4): 185-199.

Negri A., C. Vollhardt, et al. (2005). Effects of the herbicide diuron on the early life history stages of coral. Marine Pollution Bulletin 51(1-4): 370-383.

Niyogi K. K. (1999). Photoprotection revisited: genetic and molecular approaches. Annual Review of Plant Physiology and Plant Molecular Biology 50: 333-359.

ÓBrien K. P., M. Remm, et al. (2005). Inparanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Research 33: D476-D480.

Okubo N., D. Hayward, et al. (2016). A comparative view of early development in the corals *Favia lizardensis*, *Ctenactis echinata*, and *Acropora millepora* – morphology, transcriptome, and developmental gene expression. BMC Evolutionary Biology 18:48.

Parchman T. L., K. S. Geist, et al. (2010). Transcriptome sequencing in an

ecologically important tree species: assembly, annotation, and marker discovery. BMC Genomics 11.

Pawlowski K. (2008). Uncharacterized/hypothetical proteins in biomedical 'omics' experiments: is novelty being swept under the carpet? Briefings in Functional Genomics and Proteomics 7(4): 283-290.

Peng S.-E., Y.-B. Wang, et al. (2010). Proteomic analysis of symbiosome membranes in Cnidaria-dinoflagellate endosymbiosis. Proteomics 10: 1002-1016.

Perez S. and Weis V. (2006). Nitric oxide and cnidarian bleaching: an eviction notice mediates breakdown of a symbiosis. The Journal of Experimental Biology 209: 2804-2810.

Pernice M., A. Meibom, et al. (2012). A single-cell view of ammonium assimilation in coral-dinoflagellate symbiosis. The ISME Journal 6(7): 1314-1324.

Petersen T. N., S. Brunak, et al. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 8: 785-786.

Pfeffer K. (2003). Biological functions of tumor necrosis factor cytokines and their receptors. Cytokine and Growth Factor Reviews 14(3-4): 185-191.

Pinzón J. H., B. Kamel, et al. (2015). Whole transcriptome analysis reveals changes in expression of immune-related genes during and after bleaching in a reef-building coral. Royal Society Open Science 2(4): 140214.

Pisani D., J. A. Cotton, et al. (2007). Supertrees disentangle the chimerical origin of eukaryotic genomes. Molecular Biology and Evolution 24(8): 1752-1760.

Polato N. R., J. C. Vera, et al. (2011). Gene discovery in the threatened elkhorn coral: 454 sequencing of the *Acropora palmata* transcriptome. PLoS One 6(12): e28634.

Pont-Kingdon G., N. Okada et al. (1998). Mitochondrial DNA of the coral

*Sarcophyton glaucum* contains a gene for a homologue of bacterial MutS: a possible case of gene transfer from the nucleus to the mitochondrion. Journal of Molecular Evolution 46(4): 419-431.

Porter J. W., W. K. Fitt, et al. (1989). Bleaching in reef corals: physiological and stable isotopic responses. Proceedings of the National Academy of Sciences 86: 9342-9346.

Putnam N. H., M. Srivastava, et al. (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317(5834): 86-94.

Quistad S. D., A. Stotland, et al. (2014). Evolution of TNF-induced apoptosis reveals 550 My of functional conservation. Proceedings of the National Academy of Sciences 111(26): 9567-9572.

Rambold A. S., S. Cohen, et al. (2015). Fatty acid trafficking in starved cells: regulation by lipid droplet lipolysis, autophagy, and mitochondrial fusion dynamics. Developmental Cell 32(6): 678-692.

Ramirez-Llodra E, P. A. Tyler, et al. (2011) Man and the last great wilderness: human impact on the deep sea. PLoS One 6(8): e22588.

Ramos-Silva P., J. Kaandorp, et al. (2013). The skeletal proteome of the coral *Acropora millepora*: the evolution of calcification by co-option and domain shuffling. Molecular Biology and Evolution 30(9): 2099-2112.

Ramos-Silva P., F. Marin, et al. (2013). Biomineralization toolkit: the importance of sample cleaning prior to the characterization of biomineral proteomes. Proceedings of the National Academy of Sciences 110(24): E2144-E2146.

Rees T. A.V. (1986). The green hydra symbiosis and ammonium. I. The role of the host in ammonium assimilation and its possible regulatory significance. Proceedings of the Royal Society of London. Series B 229: 299-314.

Reyes-Bermudez A., Z. Y. Lin, et al. (2009). Differential expression of three

galaxin-related genes during settlement and metamorphosis in the scleractinian coral *Acropora millepora*. BMC Evolutionary Biology 9.

Rivera C., J. Voipio, et al. (2005). Two developmental switches in GABAergic signalling: the K+-Cl−cotransporter KCC2 and carbonic anhydrase CAVII. The Journal of Physiology 562(1): 27-36.

Robinson M. D., D. J. McCarthy, et al. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1): 139-140.

Rodriguez-Lanetty M., W. S. Phillips, et al. (2006). Transcriptome analysis of a cnidarian-dinoflagellate mutualism reveals complex modulation of host gene expression. BMC Genomics 7(1): 23.

Romano S. and S. Palumbi (1996). Evolution of scleractinian corals inferred from molecular systematics. Science 271: 640-642.

Romano S. L. and S. D. Cairns (2000). Molecular phylogenetic hypotheses for the evolution of scleractinian corals. Bulletin of Marine Science 67(3): 1043-1068.

Roniewicz E. and E. Morycowa (1993). Evolution of the Scleractinia in the light of microstructural data. Courier Forschungsinstitut Senckenberg 164: 233-240.

Ronquist F. and J. P. Huelsenbeck (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572-1574.

Ronquist F., M. Teslenko, et al. (2012). MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61(3): 539-542.

Roth M. S. (2014). The engine of the reef: photobiology of the coral-algal symbiosis. Frontiers in Microbiology 5: 422.

Rowan R., Knowlton N., et al. (1997). Landscape ecology of algal symbionts creates variation in episodes of coral bleaching. Nature: 388: 265-269.

Ruiz-Jones L. and S. R. Palumbi (2015). Transcriptome-wide changes in coral gene expression at noon and midnight under field conditions. The Biological Bulletin 228: 227-241.

Safavi-Hemami H., N. D. Young, et al. (2010). Characterisation of nitric oxide synthase in three cnidarian-dinoflagellate symbioses. PLoS One 5(4): e10379.

Sambrook J. and D. W. Russell (2001). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sánchez-Pulido L., F. Martín-Belmonte, et al. (2002). MARVEL: a conserved domain involved in membrane apposition events. TRENDS in Biochemical Sciences 27: 599-601.

Sano H., L. Eguez, et al. (2007). Rab10, a target of the AS160 Rab GAP, is required for insulin-stimulated translocation of GLUT4 to the adipocyte plasma membrane. Cell Metabolism 5(4): 293-303.

Schnitzler C. E. and V. M. Weis (2010). Coral larvae exhibit few measurable transcriptional changes during the onset of coral-dinoflagellate endosymbiosis. Marine Genomics 3(2): 107-116.

Schwartz S. L., C. Cao, et al. (2008). Rab GTPases at a glance. Journal of Cell Science 120: 3905-3910.

Schwarz J. A., V. M. Weis, et al. (2002). Feeding behavior and acquisition of zooxanthellae by planula larvae of the sea anemone *Anthopleura elegantissima*. Marine Biology 140(3): 471-478.

Scrutton C. T. (1993). New kilbuchophyllid corals from the Ordovician of the Southern Uplands, Scotland. Courier Forschungsinstitut Senckenberg. 164: 153-158.

Scrutton C. T. and E. N. K. Clarkson (1991). A new scleractinian-like coral from the Ordovician of the Southern Uplands, Scotland. Palaeontology 34: 179 -194.

Shimodaira H. (2002). An approximately unbiased test of phylogenetic tree selection. Systematic Biology 51(3): 492-508.

Shimodaira H. and M. Hasegawa (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Molecular Biology and Evolution 16(8): 1114-1116.

Shimodaira H. and M. Hasegawa (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17: 1246-1247.

Shinzato C., M. Inoue, et al. (2014). A snapshot of a coral ‘‘holobiont’’: a transcriptome assembly of the scleractinian coral, *Porites*, captures a wide variety of genes from both the host and symbiotic zooxanthellae. PLoS One 9(1): e85182.

Shinzato C., E. Shoguchi, et al. (2011). Using the *Acropora digitifera* genome to understand coral responses to environmental change. Nature 476(7360): 320-323.

Shoguchi E., C. Shinzato, et al. (2013). Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. Current Biology 23: 1-10.

Smith G. J., L. Muscatine (1999). Cell cycle of symbiotic dinoflagellates: variation in G(1) phase-duration with anemone nutritional status and macronutrient supply in the *Aiptasia pulchella-Symbiodinium pulchrorum* symbiosis. Marine Biology 134: 405- 418.

Stamatakis A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22(21): 2688-2690.

Stampar S. N., M. M. Maronna, et al. (2014). Fast-evolution mitochondrial DNA in

Ceriantharia: a reflection of Hexacorallia paraphyly? PLoS One 9(1): e86612.

Stanley G. D. J. and D. G. Fautin (2001). The origins of modern corals. Science 291: 1913-1914.

Stanley G. D. J. and P. K. Swart (1995). Evolution of the coral-zooxanthellae symbiosis during the Triassic: a geochemical approach. Paleontological Society 21: 179-199.

Stanley J. G. D. (2003). The evolution of modern corals and their early history. Earth-Science Reviews 60: 195-225.

Stampar

Steen R. G., and L. Muscatine (1987). Low temperature evokes rapid exocytosis of symbiotic algae by a sea anemone. Biological Bullutin 172: 246 -263.

Stolarski J., M. V. Kitahara, et al. (2011). The ancient evolutionary origins of Scleractinia revealed by azooxanthellate corals. BMC Evolutionary Biology 11(1): 316.

Sunagawa S., M. K. DeSalvo, et al. (2009). Identification and gene expression analysis of a taxonomically restricted Cysteine-riched protein family in reef-building corals. PLoS One 4(3): e4865.

Sunagawa S., E. C. Wilson, et al. (2009). Generation and analysis of transcriptomic resources for a model system on the rise: the sea anemone *Aiptasia pallida* and its dinoflagellate endosymbiont. BMC Genomics 10(1): 258.

Sutton DC, Hoegh-Guldberg O. 1990. Host-zooxanthella interactions in 4 temperate marine invertebrate symbioses: assessment of effect of host extracts on symbionts. Biological Bulletin. 178:175-186.

Tamura K., D. Peterson, et al. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum

parsimony methods. Molecular Biology and Evolution 28: 2731-2739.

Tarduno J., W. Sliter, et al. (1991). Rapid formation of ontong java plateau by aptian mantle plume volcanism. Science 254(5030): 399-403.

Trapido-Rosenthal H. G., K. H. Sharp, et al. (2001). Nitric oxide and cnidarian-dinoflagellate symbioses: pieces of a puzzle. American Zoologist 41: 247-257.

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

Uehlein N., H. Sperling, et al. (2012). The *Arabidopsis* aquaporin PIP1;2 rules cellular $CO_2$ uptake. Plant, Cell and Environment 35(6): 1077-1083.

Venn A. A., J. E. Loram, et al. (2008). Photosynthetic symbioses in animals. Journal of Experimental Botany 59(5): 1069-1080.

Veron J. E. N. (1995). Corals in space and time: the biogeography and evolution of the Scleractinia. Sydney, University of New South Wales Press.

Vidal-Dupiol J., M. Adjeroud, et al. (2009). Coral bleaching under thermal stress: putative involvement of host/symbiont recognition mechanisms. BMC Physiology 9(1): 14.

Voigt O., M. Adamski, et al. (2014). Calcareous sponge genomes reveal complex evolution of α-carbonic anhydrases and two key biomineralization enzymes. BMC Evolutionary Biology 14: 230.

von Haeseler A. (2012). Do we still need supertrees? BMC Biology 10(1): 13.

Voolstra C. R., J. A. Schwarz, et al. (2009). The host transcriptome remains unaltered during the establishment of coral-algal symbioses. Molecular Ecology 18(9): 1823-1833.

Wang J.-T., Y.-Y. Chen, et al. (2012). Physiological and biochemical performances of menthol-induced aposymbiotic corals. PLoS One 7(9): e46406.

Wang J. -T. and A. E. Douglas (1999). Essential amino acid synthesis and nitrogen recycling in an alga-invertebrate symbiosis. Marine Biology 135(2): 219-222.

Weiner I. D. and J. W. Verlander (2010). Molecular physiology of the Rh ammonia transport proteins. Current Opinion in Nephrology and Hypertension 19(5): 471-477.

Weis V. M. and A. Denis (2009). What Determines Coral Health? Science 324 (5931): 1153-1155.

Weis V. M. and W. S. Reynolds (1999). Carbonic anhydrase expression and synthesis in the sea anemone *Anthopleura elegantissima* are enhanced by the presence of dinoflagellate symbionts. Physiological and Biochemical Zoology 72: 307-316.

Weis V. M., W. S. Reynolds, et al. (2001). Host-symbiont specificity during onset of symbiosis between the dinoflagellates *Symbiodinium* spp. and planula larvae of the scleractinian coral *Fungia scutaria*. Coral Reefs 20: 301-308.

Weisrock D. W. (2012). Concordance analysis in mitogenomic phylogeneitcs. Molecular Phylogenetics and Evolution 65(1): 194-202.

Wells J. (1956). Scleractinia. Lawrence, University of Kansas Press.

West J. M. and R. V. Salm (2003). Resistance and resilience to coral bleaching: Implications for coral reef conservation and management. Conservation Biology 17(4): 956-967.

Whelan S. and N. Goldman (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Molecular Biology and Evolution 18(5): 691-699.

Whitehead L. F. (2003). Metabolite comparisons and the identity of nutrients translocated from symbiotic algae to an animal host. Journal of Experimental Biology 206(18): 3149-3157.

Wilkinson S. P., P. L. Fisher, et al. (2015). Intra-genomic variation in symbiotic dinoflagellates: recent divergence or recombination between lineages? BMC Evolutionary Biology 15:46.

Woo J., J. Lee, et al. (2008). The effect of aquaporin 5 overexpression on the Ras signaling pathway. Biochemical and Biophysical Research Communications 367(2): 291-298.

Xia X. (2013). DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. Molecular Biology and Evolution 30: 1720-1728.

Xia X. and P. Lemey (2009). Assessing substitution saturation with DAMBE. The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. P. Lemey, M. Salemi and V. Anne-Mieke. New York, Kluwer Publishing House Academic Publishers.

Xia X., Z. Xie, et al. (2003). An index of substitution saturation and its application. Molecular Phylogenetics and Evolution 26(1): 1-7.

Yellowlees D., T. Rees, et al. (1994). Effect of Ammonium-supplemented Seawater on Glutamine Synthetase and Glutamate Dehydrogenase Activities in Host Tissue and Zooxanthellae of Pocillopora damicornis and on Ammonium Uptake Rates of the Zooxanthellae. Pacific Science 48: 291-295.

Yellowlees D., T. A. V. Rees, et al. (2008). Metabolic interactions between algal symbionts and invertebrate hosts. Plant, Cell and Environment 31(5): 679-694.

Young M. D., M. J. Wakefield, et al. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biology 11(2): R14.

Yuyama I. and Higuchi T. (2014). Comparing the effects of symbiotic algae (*Symbiodinium*) clades C1 and D on early growth stages of *Acropora tenuis*. PLoS One 9(6): e98999.

Yuyama I., T. Watanabe, et al. (2011). Profiling differential gene expression of symbiotic and aposymbiotic corals using a high coverage gene expression profiling (HiCEP) analysis. Marine Biotechnology 13(1): 32-40.

Zoccola D., P. Ganot et al. (2015). Bicarbonate transporters in corals point towards a key step in the evolution of cnidarian calcification. Scientific Reports 5: 9983.

**Appendix A**

Corallimorpharian taxonomy. Species examined in this study are indicated in bold. Asterisk (*) indicates the azooxanthellate species, and a (+) indicate deep-water species.

| | | | |
|---|---|---|---|
| | | Phylum Cnidaria | |
| | | Class Anthozoa | |
| | | Subclass Hexacorallia | |
| | | Order Corallimorpharia | |
| Family | Genus | Specie | Synonym |
| Discosomidae | *Actinodiscus* | *A. dawydoffi* | *D. dawydoffi* |
| | | *A. fungiforme* | *D. fungiforme* |
| | | ***A. nummiforme*** | *D. nummiforme* |
| | | *A. rubraoris* | *D. rubraoris* |
| | | *A. unguja* | *D. unguja* |
| | *Amplexidiscus* | ***A. fenestrafer*** | |
| | *Discosoma* | *D. album* | |
| | | *D. carlgreni* | |
| | | *D. dawydoffi* | *A. dawydoffi* |
| | | *D. fowleri* | |
| | | *D. fungiforme* | *A. fungiforme* |
| | | *D. molle* | |
| | | *D. neglecta* | |
| | | *D. nummiforme* | *A. nummiforme* |
| | | *D. rubraoris* | *A. rubraoris* |
| | | *D. unguja* | *A. unguja* |
| | | *D. viridescens* | |
| | *Metrahodactis* | *M. boninensis* | |
| | *Rhodactis* | *R. bryoides* | |
| | | *R. howesii* | |
| | | *R. inchoata* | |
| | | ***R. indosinensis*** | |
| | | *R. musciformis* | |
| | | *R. osculifera* | |
| | | *R. rhodostoma* | |

| | | | |
|---|---|---|---|
| | | ***R. mussoides*** | *Platyzoanthus mussoides* |
| Ricordeidae | *Ricordea* | *R. florida* | |
| | | ***R. yuma*** | |
| Corallimorphidae | *Corallimorphus* | *C. denhartogi*+* | |
| | | *C. ingens*+* | |
| | | *C. niwa*+* | |
| | | *C. pilatus*+* | |
| | | ***C. profundus*+*** | |
| | | *C. rigidus*+* | |
| | *Corynactis* | *C. annulata** | |
| | | *C. australis** | |
| | | *C. caboverdensis** | *P. caboverdensis* |
| | | ***C. californica**** | |
| | | *C. caribbeorum** | *P. caribbeorum* |
| | | *C. carnea** | |
| | | *C. chilensis** | |
| | | *C. delawarei** | |
| | | *C. dengartogi** | |
| | | *C. denticulosa** | |
| | | *C. globulifera** | *P. globulifera*[2] |
| | | *C. hoplites** | |
| | | *C. mediterranea** | |
| | | *C. parvula** | |
| | | *C. sanmatiensis** | |
| | | *C. viridis** | |
| | *Pseudocorynactis* | *P. caribbeorum** | *C. caribbeorum* |
| | | *P. caboverdensis** | *C. caboverdensis* |
| | | | *cf. C. parvula* |
| | | ***P.* sp.**[1] | |
| | | *P. globulifera*[2]* | |
| | | *P. tuberculata*[2]* | |
| | *Paracorynactis*[2] | *P. hoplites** | |
| Sideractiidae | *Nectactis* | *N. singularis*+* | |

| | | |
|---|---|---|
| | *Sideractis* | *S. glacialis*+* |

[1] *Pseudocorynactis* sp. examined/sequenced herein are similar in morphology to *Corynactis* with bulb-end tentacle, but contains zooxanthellae. It is tentatively accommodated in *Pseudocorynactis*, even though anatomical information and more material from the North Pacific Ocean are required.

[2] Ocaña et al. 2010.

**Appendix B**

Asexual reproduction by marginal budding in the tropical corallimorpharian, *Ricordea yuma* (Corallimorpharia; Ricordeidae).

Mei-Fang Lin, Chaolun Allen Chen, David J. Miller.

Galaxea, Journal of Coral Reef Studies 15: 41-42, 2013

Photogallery

# Asexual reproduction by marginal budding in the tropical corallimorpharian, *Ricordea yuma* (Corallimorpharia; Ricordeidae)

**Mei-Fang LIN[1], Chaolun Allen CHEN[2,*], and David J. MILLER[1,*]**

[1] ARC Centre of Excellence Coral Reef Studies, James Cook University, Townsville, 4811 Australia
[2] Biodiversity Research Centre/Taiwan International Graduate Program-Biodiversity, Academia Sinica, Taipei, 115 Taiwan

* Corresponding author: A. Chen, D.J. Miller
E-mail: cac@gate.sinica.edu.tw, david.miller@jcu.edu.au

Communicated by Hiroya Yamano (Environment and Conservation Editor)

**Keywords** "False corals", Corallimorpharia, Anthozoa, reproduction, marginal budding

## Appendix C

The ‘‘Naked Coral’’ Hypothesis Revisited – Evidence for and Against Scleractinian Monophyly.

Marcelo V. Kitahara, Mei-Fang Lin, Sylvain Forêt, Gavin Huttley, David J. Miller, Chaolun Allen Chen.

PLoS One 9: e94744, 2014

OPEN ACCESS Freely available online

PLOS ONE

# The "Naked Coral" Hypothesis Revisited – Evidence for and Against Scleractinian Monophyly

**Marcelo V. Kitahara[1,2☯], Mei-Fang Lin[3,4,5☯], Sylvain Forêt[5,6], Gavin Huttley[7], David J. Miller[3,5*], Chaolun Allen Chen[4,8,9*]**

**1** Departamento de Ciências do Mar, Universidade Federal de São Paulo, Santos, São Paulo, Brazil, **2** Centro de Biologia Marinha (CEBIMar), Universidade de São Paulo, São Sebastião, São Paulo, Brazil, **3** School of Pharmacy and Molecular Sciences, James Cook University, Townsville, Queensland, Australia, **4** Biodiversity Research Centre, Academia Sinica, Taipei, Taiwan, **5** ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland, Australia, **6** Research School of Biology, Australian National University, Canberra, Australian Capital Territory, Australia, **7** John Curtin School of Medical Research, Australian National University, Canberra, Australian Capital Territory, Australia, **8** Institute of Oceanography, National Taiwan University, Taipei, Taiwan, **9** Taiwan International Graduate Program (TIGP)-Biodiversity, Academia Sinica, Taipei, Taiwan

## Abstract

The relationship between Scleractinia and Corallimorpharia, Orders within Anthozoa distinguished by the presence of an aragonite skeleton in the former, is controversial. Although classically considered distinct groups, some phylogenetic analyses have placed the Corallimorpharia within a larger Scleractinia/Corallimorpharia clade, leading to the suggestion that the Corallimorpharia are "naked corals" that arose via skeleton loss during the Cretaceous from a Scleractinian ancestor. Scleractinian paraphyly is, however, contradicted by a number of recent phylogenetic studies based on mt nucleotide (nt) sequence data. Whereas the "naked coral" hypothesis was based on analysis of the sequences of proteins encoded by a relatively small number of mt genomes, here a much-expanded dataset was used to reinvestigate hexacorallian phylogeny. The initial observation was that, whereas analyses based on nt data support scleractinian monophyly, those based on amino acid (aa) data support the "naked coral" hypothesis, irrespective of the method and with very strong support. To better understand the bases of these contrasting results, the effects of systematic errors were examined. Compared to other hexacorallians, the mt genomes of "Robust" corals have a higher (A+T) content, codon usage is far more constrained, and the proteins that they encode have a markedly higher phenylalanine content, leading us to suggest that mt DNA repair may be impaired in this lineage. Thus the "naked coral" topology could be caused by high levels of saturation in these mitochondrial sequences, long-branch effects or model violations. The equivocal results of these extensive analyses highlight the fundamental problems of basing coral phylogeny on mitochondrial sequence data.

**Citation:** Kitahara MV, Lin M-F, Forêt S, Huttley G, Miller DJ, et al. (2014) The "Naked Coral" Hypothesis Revisited – Evidence for and Against Scleractinian Monophyly. PLoS ONE 9(4): e94774. doi:10.1371/journal.pone.0094774

**Editor:** John Murray Roberts, Heriot-Watt University, United Kingdom

**Received** October 31, 2013; **Accepted** March 20, 2014; **Published** April 16, 2014



**Funding:** This study was supported by grants from the National Science Council (NSC) and Academia Sinica (Thematic Grants 2005-2010) to CAC, from the Australian Research Council to DJM and SF, and from São Paulo Funding Agency (FAPESP grant 2012/21583-1) to MVK. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: cac@gate.sinica.edu.tw (CAC); david.miller@jcu.edu.au (DJM)

☯ These authors contributed equally to this work.

## Introduction

The order Scleractinia, comprising the anthozoan cnidarians that produce a continuous external aragonitic skeleton [1], are not only the architects of some of the most complex habitats (i.e. coral reefs) but are also near ubiquitous in distribution. Despite their global significance [2–7], several key aspects of scleractinian evolution are as yet poorly understood. Most coral families are first identifiable in the Triassic, by which time much of the extant morphological diversity is represented. Molecular data implies a deep split of extant corals between two large clades, the "Complex" and "Robust" [8–17], but many families defined by morphology are not monophyletic by molecular criteria and some are split between "Complex" and "Robust" clades [9–11,14,17–20]. One hypothesis to explain the sudden appearance of a highly diverse Middle Triassic coral fauna is that skeletonisation has been an ephemeral trait during the evolution of the Scleractinia [21]. Under this scenario, scleractinian lineages may have undergone skeleton loss in the face of global environmental instability [21], which would severely compromise fossil preservation. Consistent with this idea, some corals have been shown to undergo complete (but reversible) skeleton loss under acid conditions [22], whereas other species are apparently much less susceptible to skeleton dissolution [23].

The "naked coral" hypothesis [24] is a topical extension of the idea of skeleton ephemerality in corals. Corallimorpharians, anthozoans that lack skeletons, have a close but unclear relationship to the Scleractinia. Corallimorpharians and scleractinians are very similar both in terms of anatomy and histology (see [92]), and these characteristics have in the past been used to argue for merging the orders [21,26]. Medina et al. [24] conducted a phylogenetic analysis based on the proteins encoded by 17 complete mitochondrial (mt) genomes, which suggested that scleractinians are paraphyletic, corallimorpharians being more closely related to "Complex" than are "Robust" corals (Figure 1).

The authors hypothesize that the Corallimorpharia (''naked corals'') may have arisen during the Cretaceous (110~132 Mya) from a scleractinian ancestor that had undergone skeleton loss as a consequence of ocean acidification. A recent study [27] using complete mitochondrial genomes from a broad range of representative cnidarians also failed to unambiguously reject the ''naked coral'' hypothesis.

Here we applied a wide variety of analytical methods to a more comprehensive dataset of complete mitochondrial genome sequences (50 representative anthozoans) to better understand hexacorallian phylogeny. Whilst phylogenetic analyses based on amino acid (aa) data were for the most part consistent with the ''naked corals'' scenario (i.e. paraphyletic Scleractinia), it remains possible that the ''naked corals'' topology is an artefact caused by saturation, compositional biases or other violations of model assumptions. With the possible exception of cerianthiids [28], anthozoan cnidarians are thought to carry out mt DNA repair and thus differ from bilaterians *sensu stricto* in this respect. The main lines of evidence for repair are the extremely slow rate of evolution of the mt genomes of Anthozoa [29,30] and the presence of a *MutS*-like gene in the octocoral mt genome [25,31]. We speculate that reduced efficiency of mt DNA repair in the ''Robust'' coral lineage could account for the observed anomalies in codon use and composition.

## Material and Methods

### DNA extraction and Polymerase Chain Reaction

Total genomic DNA was extracted using a Qiagen Qiamp or DNeasy Tissue Kit (QIAGEN). DNA concentrations were determined using a Nanodrop 1000 (Thermo Scientific) and an aliquot of each species total genomic DNA was diluted or concentrated to a final concentration of 40 ng/µl. Previously described primer sequences [32] were used to amplify the complete mt genomes of the following scleractinians: *Alveopora* sp.; *Astreopora explanata; A. myriophthalma; Isopora palifera*; and *I. togianensis*.

Two to three fragments (3~9 kb) covering the entire mt genome of all but *G. hawaiiensis* were amplified by long Polymerase Chain Reaction (long-PCR) technique [33] from partial fragments of *rns*, *rnl* and *cox1* genes obtained from coral-specific primers and invertebrate universal primer [34,35,36]. Long-PCR were carried out using Takara La Taq using a slightly modified conditions from those recommended by the manufacturer as follows: 94°C for 1 min, then 30 cycles of 10 s at 98°C, 45 s at 62~63°C, 14.25 min at 68°C for, and 10 min at 72°C. For *G. hawaiiensis* portions of *rnl, rns* and *cox1* were first amplified with the universal primers mentioned above, followed by the scleractinian universal primers CS-1 to CS-21 [37] that covered the entire mt genome. To obtain sequence from regions that did not yield product using these primers, nineteen specific primers were developed based on the sequences retrieved from *G. hawaiiensis* (Table S1). Polymerase Chain Reaction (PCR) were carried out using the Advantage2 polymerase kit (Clontech) under the conditions recommended by the manufacturer. PCR conditions were: 95°C for 5 min, then 30 cycles of 30 s at 94°C, 60 to 90 s at 54~60°C, 90 s at 72°C, followed by 5 min at 72°C.

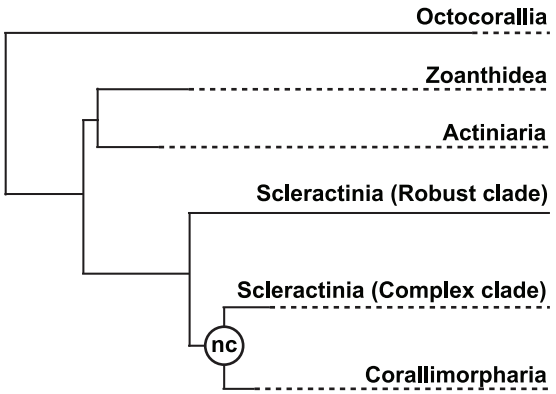



**Figure 1. Phylogenetic relationships within the cnidarian Class Anthozoa according to the ''naked corals'' (nc) hypothesis (modified from** [24]**).**
doi:10.1371/journal.pone.0094774.g001

### Phylogenetic analysis

Resulting sequences were verified and assembled using Sequencher 4.8 (Gene Codes Corporation) and then analyzed in Vector NTI 9.0 (InforMax, Invitrogen life science software). Examination of open reading frames (ORFs) and codon usage, as well as other DNA statistics were performed using Dual Organelle Genome Annotator [38], Sequence Manipulation Suite v.2 [39], and MEGA5 [40]. In addition to the 6 new complete scleractinian mt genome sequences determined for this study, those of 25 other scleractinians, 12 corallimorpharians, 2 sea anemones, single antipatharian and zoanthid species, and 3 octocorals were obtained from public databases (Table 1).

In order to make the analyses based on nucleotide and amino acid sequence data strictly comparable, the rRNAs, IGS, and tRNAs coding sequences were excluded from consideration. Therefore, for each species included in the present analysis, the data set included all protein-coding genes. The predicted amino acid sequences encoded by each of the 50 mt genomes were aligned using MAFFT v.5 [41]. These alignments were reverse translated to generate nucleotide sequence alignments, and phylogenetic inferences carried out on the concatenated amino acid and nucleotide alignments, removing all positions containing more than 50% gaps. The most appropriate model of nucleotide substitution was determined for the final alignment (totaling 11,802 bp) by the hierarchical likelihood ratio test implemented in MEGA5 as GTR+I+G (lnL -133020.1). Maximum Likelihood fits of 48 different amino acid substitution models using only positions that did not contain any gaps or missing data were calculated in MEGA5 [40]. There were a total of 3,666 positions (from the 3,934 aa) in this final dataset and JTT+G+I+F (lnL -51687.5) was chosen as the best evolutionary model.

Phylogenetic analyses were performed using PhyML [42] for Maximum Likelihood (ML) and MrBayes version 3.1.2 [43] for Bayesian Inference (BI). ML analyses were performed under the GTR model for nt alignments and JTT for aa alignments. For the BI, 2 runs of 4 chains were calculated for 10 million generations with topologies saved every 1,000 generations. One million generations were discarded as burn-in to ensure that the likelihood had plateaued and that the two runs had converged to less than 0.002 average standard deviation of split frequencies.

Given concerns for the influence of the long branch of the ''Robust'' scleractinian clade, ML phylogenetic analyses were repeated under several different scenarios for the nt data matrices as follow: i) different substitution model categories following Bayesian Information Criterion [BIC] and ML [lnL] recommendations; ii) systematically codon exclusion ($1^{st}$, $2^{nd}$, and $3^{rd}$); iii) different nucleotide divergence rates across frames; iv) coding nucleotide data as purines and pyrimidies (RY-coding) (see [44,45]) RY-coding excluding the $3^{rd}$ codon. Likewise, additional ML phylogenetic analyses of the aa final alignment included: i)

**Table 1.** Mitochondrial genome sequence data included in the present analyses

| Order | Species | size (bp) | GenBank accession # | Reference |
|---|---|---|---|---|
| Actiniaria | | | | |
| | *Metridium senile* | 17,443 | NC000933 | [78] |
| | *Nematostella* sp. | 16,389 | DQ643835 | [24] |
| Alcyonacea | | | | |
| | *Briareum asbestinum* | 18,632 | DQ640649 | [24] |
| | *Pseudopterogorgia bipinnata* | 18,733 | DQ640646 | [24] |
| | *Sarcophyton glaucum* | Incomplete | AF064823 | [86] |
| Antipatharia | | | | |
| | *Chrysopathes formosa* | 18,398 | NC008411 | [87] |
| Corallimorpharia | | | | |
| | *Actinodiscus nummiformis* | 20,922 | | Lin et al. (submitted) |
| | *Amplexidiscus fenestrafer* | 20,188 | | Lin et al. (submitted) |
| | *Corallimorphus profundus* | 20,488 | | Lin et al. (submitted) |
| | *Corynactis californica* | 20,632 | | Lin et al. (submitted) |
| | *Discosoma* sp.1 | 20,908 | DQ643965 | [24] |
| | *Discosoma* sp.2 | 20,912 | DQ643966 | [24] |
| | *Pseudocorynactis* sp. | 21,239 | | Lin et al. (submitted) |
| | *Rhodactis indosinesis* | 20,092 | | Lin et al. (submitted) |
| | *Rhodactis mussoides* | 20,826 | | Lin et al. (submitted) |
| | *Rhodactis* sp. | 20,093 | DQ640647 | [24] |
| | *Ricordea florida* | 21,376 | DQ640648 | [24] |
| | *Ricordea yuma* | 22,015 | | Lin et al. (submitted) |
| Scleractinia | | | | |
| | *Acropora tenuis* | 18,338 | AF338425 | [64] |
| | *Agaricia humilis* | 18,735 | DQ643831 | [24] |
| | *Anacropora matthai* | 17,888 | AY903295 | [32] |
| | *Alveopora* sp. | 18,146 | | KJ634271 |
| | *Astrangia* sp. | 14,853* | DQ643832 | [24] |
| | *Astreopora explanata* | 18,106 | | KJ634269 |
| | *Astreopora myriophthalma* | 18,106 | | KJ634272 |
| | *Colpophyllia natans* | 16,906* | DQ643833 | [24] |
| | *Euphyllia ancora* | 18,875 | JF825139 | [37] |
| | *Fungiacyathus stephanus* | 19,381 | JF825138 | [37] |
| | *Gardineria hawaiiensis* | 19,429 | | Lin et al. (submitted) |
| | *Goniopora columna* | 18,766 | JF825141 | [37] |
| | *Isopora palifera* | 18,725 | | KJ634270 |
| | *Isopora togianensis* | 18,637 | | KJ634268 |
| | *Madracis mirabilis* | 16,951* | EU400212 | [61] |
| | *Madrepora oculata* | 15,839* | | [65] |
| | *Montastraea annularis* | 16,138* | AP008974 | [62] |
| | *Montastraea faveolata* | 16,138* | AP008978 | [62] |
| | *Montastraea franksi* | 16,137* | AP008976 | [62] |
| | *Montipora cactus* | 17,887* | AY903296 | [32] |
| | *Mussa angulosa* | 17,245* | DQ643834 | [24] |
| | *Pavona clavus* | 18,315 | DQ643836 | [24] |
| | *Pocillopora damicornis* | 17,425* | EU400213 | [61] |
| | *Pocillopora eydouxi* | 17,422* | EF526303 | [88] |
| | *Polycyathus* sp. | 15,357* | JF825140 | [37] |
| | *Porites okinawensis* | 18,647 | JF825142 | [37] |

**Table 1.** Cont.

| Order | Species | size (bp) | GenBank accession # | Reference |
|---|---|---|---|---|
| | *Porites porites* | 18,648 | DQ643837 | [24] |
| | *Seriatopora caliendrum* | 17,010* | NC010245 | [59] |
| | *Seriatopora hystrix* | 17,059* | EF633600 | [59] |
| | *Siderastrea radians* | 19,387 | DQ643838 | [24] |
| | *Stylophora pistillata* | 17,177* | EU400214 | [61] |
| Zoanthidea | | | | |
| | *Savalia savaglia* | 20,764 | NC008827 | [89] |

doi:10.1371/journal.pone.0094774.t001

coding aa using the common six groups that usually replace one another [46,47], where MVIL were recoded as 1, FYW as 2, ASTGP as 3, DNEQ as 4, and RKH as 5, and C as 6; ii) to allow general-time-reversible (GTR) matrix to be used, the aa dataset was recoded to four categories instead of six. In this case, following [47] the aromatic (FYW) and hydrophobic (MVIL) amino acids were combined and the rare cysteine was considered as missing data. The four amino acid categories were named A, T, G, and C, respectively; iii) exclusion of all Phe, Ala, Thr and Tyr from the alignment, once the percentage of occurrence of these aa, especially of Phe and Ala, in the "Robust" scleractinian clade are significantly different once compared to all other hexacorallians included in the present analysis; iv) phylogenetic reconstructions using different evolutionary models as retrieved from results of the Bayesian Information Criterion [BIC] and ML [lnL] recommendations (JTT+G+I+F, cpREV+G+I+F, WAG+G+I+F, and Dayhoff+G+I+F); v) exclusion of Octocorallia sequences as outgroups; and vi) systematically exclusion of fast evolving sites. To find such sites, the mean (relative) evolutionary rate was estimated under the JTT+G+F in MEGA5, and a discrete Gamma (+G) distribution with 5 categories was used to model evolutionary rate differences among sites. Subsequently, 3 minimum evolution ML phylogenies were reconstructed systematically excluding all fast evolving sites that had means of >2.99, >1.99, and >1.49 respectively.

Trees with non-stationary, non-homogeneous models were computed using nhPhyML [48] with 5 categories of (G+C) content. Quartet puzzling with the Barry and Hartigan model was implemented using the PyCogent library [49], as was the Goldman [50] test.

Codon-based ML trees were inferred using CodonPhyml [51]. The results presented here used the Yap et al. model [52], but similar results were observed with other models. Codon-based trees were also inferred using MrBayes with a GTR substitution model and three categories of non-synonymous/synonymous ratios (M3 model).

For amino acids, phylogenies based on the CAT-GTR, CAT-Poisson and GTR models were inferred using PhyloBayes [53]. For each inference, the program was run until the effective size was greater than 300 and until the largest discrepancy across bipartition between runs was less than 0.1. Majority rule posterior consensus trees were built after deleting 1000 burn-in samples and taking every 10 generations.

Comparisons of topologies were carried out using the Approximately Unbiased, Kishino-Hasegawa and Shimodaira-Hasegawa tests implemented in the program Consel [54]. These tests compare the significance of the difference in likelihood of two competing topologies under the same model.

### Base Frequencies Distance Trees

The homogeneity of base frequencies among taxa is a major assumption of many molecular phylogenetic methods [55]. Therefore, changes in base composition between lineages can lead to errors in phylogenetic inference, particularly in the case of mt genome data (see [56]). In an attempt to assess the potential for compositional bias affecting the anthozoan phylogenetic inference, minimum evolution Base Frequencies distance trees (BF) were estimated using MEGA5 from matrices of pairwise BF distances.

Following Phillips et al. [45], BF distance was calculated for each taxon pair for each nucleotide category (i.e. $BF\ distance = ([\{Ai - Aj\} + \{Ti - Tj\} + \{Ci - Cj\} + \{Gi - Gj\}]/2)$, where $i$ and $j$ are the frequencies of each corresponding nucleotide for the $i$th and $j$th taxa, respectively.

### tRNA and rRNA trees

For each of the 50 anthozoan species studied, the four mitochondrial genes encoding stable RNAs (i.e., 12S rRNA, 16S rRNA, *trnM*, and *trnW*) were retrieved, but the octocoral and actiniarian data were excluded from phylogenetic analyses because of the difficulty of generating unambiguous alignments when they were included. For these analyses, the scleractinian and corallimorpharian sequences were aligned and the antipatharian *Chrysopathes formosa* used as the outgroup. Each stable RNA sequence was aligned using essentially the same approach as for the protein-coding genes, the final alignments being 1,039 bp for 12S rDNA, 1,866 bp for 16S rDNA, 72 bp for *trnM*, and 70 bp for the *trnW*. Phylogenetic inferences were based on concatenated alignments and the most appropriate model of nucleotide substitution as determined by the hierarchical likelihood ratio test was GTR+G. ML (SH-like and 100 bootstrap) were performed using PhyML [42], and BI using MrBayes version 3.1.2 [43]. BI and ML analyses were performed using the GTR model as described above.

## Results and Discussion

Whereas previous analyses were based on a limited range of scleractinian and corallimorpharian mitochondrial genomes [24,27], taxon sampling was increased in the present study to a total of 50 mt genomes, which included 12 corallimorpharians and 31 scleractinians (Table 1). Two taxa included in the present study are of particular evolutionary significance: *Corallimorphus profundus* and *Gardineria hawaiiensis*. Both anatomical [57] and molecular (this study) data suggest that *C. profundus* represents a deep-diverging corallimorpharian clade. *G. hawaiiensis* represents a lineage of scleractinians that is thought to have diverged prior to the "Complex"/"Robust" split [11,17,58].

**Table 2.** Compositional biases of the mitochondrial genomes of the anthozoan Orders included in the present analysis.

| | | | Nucleotide | | Protein | | |
|---|---|---|---|---|---|---|---|
| **Sub-class** | **Order** | **Group** | **G+C (%)** | **A+T (%)** | **FYMINK (%)** | **GARP (%)** | **FYMINK/GARP** |
| Octocorallia | Alcyonacea | | 35.33 | 64.66 | 30.54 | 21.96 | 1.39 |
| Hexacorallia | Antipatharia | | 38 | 62 | 29.78 | 22.05 | 1.35 |
| | Actiniaria | | 37.95 | 62.05 | 29.92 | 22.47 | 1.33 |
| | Zoanthidea | | 43.9 | 56.1 | 29.4 | 23.09 | 1.27 |
| | Corallimorpharia | | 37.95 | 62.05 | 28.73 | 22.15 | 1.29 |
| | Scleractinia | Basal* | 38.8 | 61.2 | 28.51 | 22.19 | 1.28 |
| | | Complex* | 37.59 | 62.41 | 29.42 | 21.81 | 1.34 |
| | | Robust* | 31.2 | 68.8 | 33.71 | 19.36 | 1.74 |

The figures shown are averages across the range of species included. For proteins, the comparisons are made between the (A+T)-rich amino acids FYMINK (Phe, Tyr, Met, Ile, Asn, and Lys), and the (G+C)-rich amino acids GARP (Gly, Ala, Arg, and Pro). Asterisks indicate groupings based on molecular data but whose taxonomic validity remains to be established.

doi:10.1371/journal.pone.0094774.t002

## General characteristics of the mt genomes of hexacorallians

All of the hexacorallian mt genomes sequenced to date contain 13 protein-coding genes (*atp6* and *8*, *cox1-3*, *cob*, *nad1-6*, and *nad4L*), 2 genes encoding ribosomal RNAs (*rns* and *rnl*), and 2 encoding tRNAs (*trnW* and *trnM*), although members of the scleractinian genus *Seriatopora* have a duplication of *trnW* and thus have a total of 3 tRNA genes [59]. Whilst Hexacorallia in general display little variation in size of the mt genome, members of the Scleractinia are exceptional in having mt genomes ranging from >19.4 Kb in the "Basal" coral *Gardineria*, to <15 Kb in some "Robust" corals, those of "Complex" corals being intermediate in size (~18.9–19.4 Kb) [37]. The size of each mt gene is also relatively stable across the range of hexacorallians, exceptions being *rnl* and *rns*, which vary by almost 500 and 700 bp respectively. As in Octocorallia [60], intergenic regions and introns (*cox1* and *nad5*) account for most of the observed variation in mt genome size. Some differences were apparent across hexacorallian orders, but gene organisation was remarkably uniform across the full range of Scleractinia, the only deviations from the canonical gene map [24,32,59,61–64] being two azooxanthellate corals, *Lophelia pertusa* [90] and *Madrepora* spp. [65]. Likewise the majority (10 of 12) of corallimorpharian mt genomes conform to a distinct gene order, the exceptions being the azooxanthellate species *Corallimorphus profundus* and *Corynactis californica*.

The nucleotide composition of the mt protein-coding genes of hexacorallians has a clear (A+T)-bias, ranging from around 56% in the zoanthid *Savalia* to an average of 69% in "Robust" corals (Table 2 and Figure 2). The coding sequences of "Robust" corals have a high thymine and low cytosine content compared to other scleractinians (Figure 2). Surprisingly, this T-enrichment over other scleractinians is not restricted to silent codon positions, but is also clear at the first (5%) and second (3%) codon positions (Figure 2), resulting in over 400 aa substitutions in "Robust" corals relative to other hexacorallians (see also Table 2).

## Contradictory phylogenies based on nucleotide or amino acid sequence data

Based on the final nt alignment (11,298 bp) consisting of all 13 protein-coding genes from 50 anthozoan mt genomes (Table 1), ML and BI recovered identical topologies and indicated that all anthozoan orders included in the analysis are monophyletic (Figure 3a). The topology shown was strongly supported, with only few nodes not having 100% support in both ML and BI. Scleractinians and corallimorpharians appear as sister groups. Whilst these results based on nt data flatly contradict the "naked coral" hypothesis [24], the application of the same phylogenetic methods to the corresponding amino acid sequence data (3,934 aa residues) consistently placed corallimorpharians as the sister group to the "Complex" Scleractinia, within the scleractinian clade (Figure 3b). In addition, the protein-based phylogenies differ in the positions of Actiniaria, Zoanthidea, and Antipatharia, and also in placing *G. hawaiiensis* as a member of the "Complex" corals instead of forming a basal scleractinian lineage [11,17]. Codon-based phylogenies also strongly support the grouping of corallimorpharians with "Complex" corals (Figure 4). The significance of the difference in likelihood between to the competing topologies is shown on table (Table 3). The preference for the "naked coral" topology is highly significant for the trees based on amino acids and codons, whereas the significance of the difference is weaker for nucleotide-based trees.

To better understand the basis of these contrasting results, we examined the potential for artifacts to arise as a result of the analytical methods or biases in the datasets.

## Use of different substitution models and removal of rapidly evolving sites

In the case of both nt and aa analyses, changing the outgroup had no effect on topology, and neither did the use of different substitution model categories. For nucleotides, the models validated included: (i) parametric GTR [93] with gamma distribution of rates among sites; (ii) TN93 [66] with gamma distribution and invariable sites; (iii) and HKY [67]. Using amino acid data, the JTT+G+I+F [68], cpREV+G+I+F [91], WAG+–G+I+F [69], and Dayhoff+G+I+F [70]. Furthermore, in order to verify differences in evolutionary models selection, the same analyses were extended with the exclusion of all sequences from "Robust" corals from the dataset. However, these new analyses recovered similar results as described above.

Next, standard rate effects were examined. Potential saturation effects in the nt data were examined by systematically excluding the 1$^{st}$, 2$^{nd}$, and 3$^{rd}$ codon positions from analyses, but the ML topology retrieved and statistical support for nodes did not differ significantly from those shown in Figure 3a. In fact, exclusion of the 3$^{rd}$ codon position improved support for some nodes (Figure

S1). Removing the most rapidly evolving sites in the aa alignment [47] also had no effect on the topology. This approach involved estimation of the mean (relative) evolutionary rate (ER) for each site under the JTT+G+F [68] model and then excluding those sites with ERs of ≥2.99, ≥1.99, or ≥1.49 (note that sites displaying ER >1 are evolving faster). The percentages of sites excluded in these cases were 18.3%, 23.1%, and 32.4% respectively; 1,275 of the 3,934 sites could therefore be excluded from the analyses without influencing the overall topology (Figure S2).

## Compositional bias effects: nucleotides

Having investigated potential artifacts arising from standard rate effects, the effects of compositional heterogeneity in the nucleotide and amino acid data were examined. In the case of the nt alignment, this involved RY coding [44,45], with or without exclusion of the third codon position from the resulting alignment (Table S2), which also resulted in scleractinian monophyly (Figure S3).

As noted above, the mt genomes of "Robust" corals do differ significantly from those of all other hexacorallians in terms of nucleotide composition, and this has consequences for both codon use and amino acid composition in the proteins that it encodes. Figure 2, Table 2 and Figure S4 show the overall base composition of mt protein-coding genes of the anthozoans included in the present analysis, and also the percentage of each base occurring in the 1st, 2nd, and 3rd codon positions.

Whereas most hexacorallians have (A+T) contents around 62% (hence (A+T)/(G+C) of around 1.63), "Robust" corals have a significantly higher (A+T content ((A+T)/(G+C) = 2.20). Consequently, the (A+T)-skew is >6% higher in "Robust" corals than in all other hexacorallians (Figure S4). This bias is asymmetrically distributed, such that in "Robust" corals the coding strand is anomalously high in thymine and low in cytosine. Such heterogeneities in base composition are a potential source of error in phylogenetic analyses [71].

In order to take into account this variability in nucleotide composition, we used the GG98 non-stationary, non-homogeneous model [72] implemented in the nhPhyML software [48]. In this approach the "naked coral" topology has the highest likelihood, but the difference in likelihood of the two competing topologies is not statistically significant (Table 3).

We further explored the effect of compositional heterogeneity using the Barry and Hartigan model [4] implemented in the PyCogent library [49]. The Barry and Hartigan model is the most general (makes the fewest assumptions) substitution model for nucleotides. It allows for non-reversible and non-stationary processes on every branch of a phylogeny and does not assume the process is time-homogeneous within or between branches. The complexity of this model precluded tree inference; instead, 1,000 quartets, each comprising a "Robust" coral, a "Complex" coral, a corallimorpharian and an outgroup, were sampled. The majority

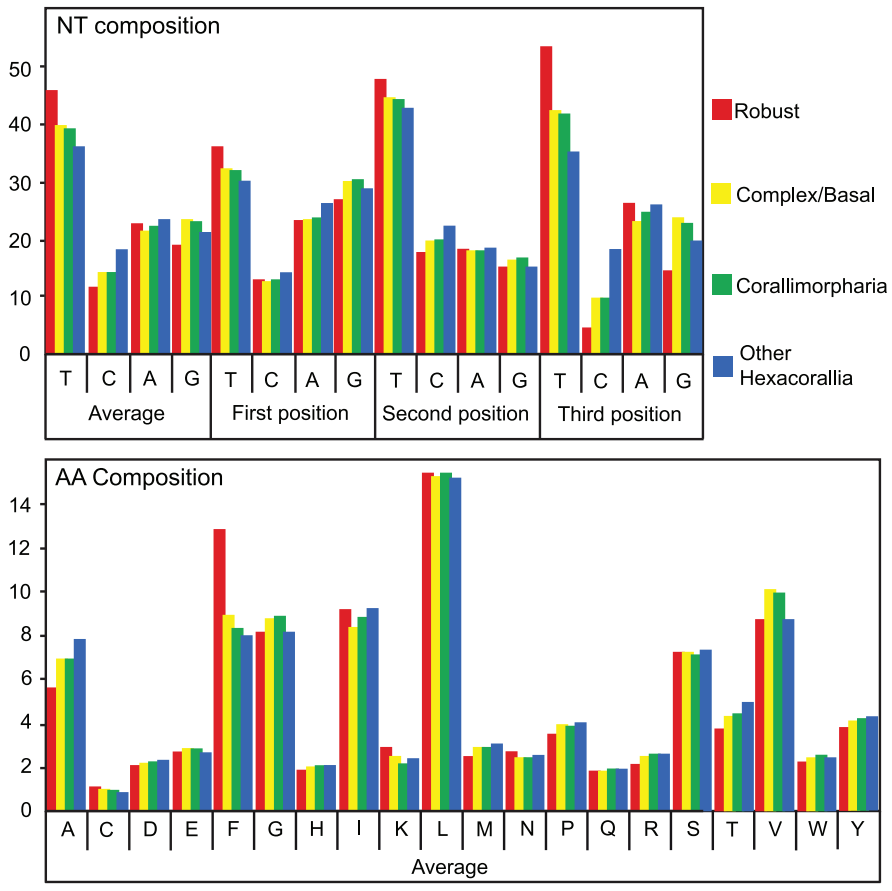



**Figure 2. Nucleotide (upper) and amino acid (lower) content of the protein coding genes in the mitochondrial genomes of hexacorallians.** The data shown are averages across the "Robust" corals (n = 14), basal and "Complex" corals (n = 17), corallimorpharians (n = 12) and other hexacorallians (n = 4).
doi:10.1371/journal.pone.0094774.g002

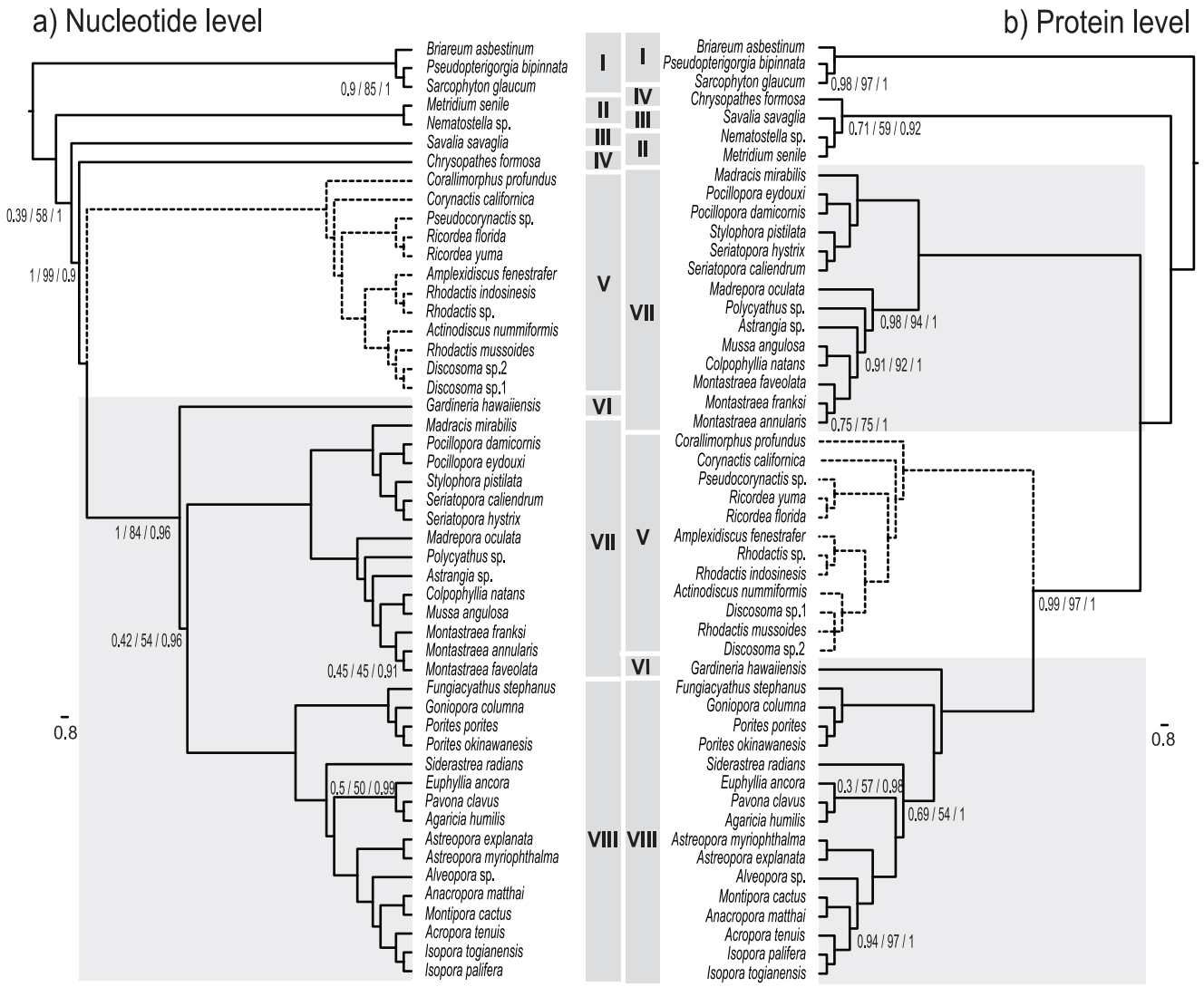



**Figure 3. Phylogenetic analyses based on (a) the nucleotide sequences of the mitochondrial genes encoding proteins and (b) the amino acid sequences of the proteins encoded by the mitochondrial genomes.** Values on the nodes indicate the non-parametric SH test and bootstrap replicates (ML), and posterior probability (BI) support respectively. Where no values are shown on a node, that edge was fully supported under all analyses. Dashed lines indicate the corallimorpharian clade. Light grey blocks identify the scleractinian clades. (I) Octocorallia used as outgroup; (II) Actiniaria; (III) Zoanthidea; (IV) Antipatharia; (V) Corallimorpharia; (VI) "Basal" Scleractinia; (VII) "Robust" Scleractinia; (VIII) "Complex" Scleractinia.
doi:10.1371/journal.pone.0094774.g003

(94%) of these quartets grouped complex corals with corallimorphs and "Robust" corals with the outgroup.

Taken together, these results based on models that do not assume compositional homogeneity or time reversibility suggest that the strong support of nt-based phylogenies for scleractinian monophyly might be an artifact of sequence composition. However, phylogenetic analyses carried out on a concatenated rRNA and tRNA alignment recovered a monophyletic Scleractinia clade with high statistical support irrespective of the method of analysis (Figure 5 and Table 3). Using this alignment, quartet puzzling with a Barry and Hartigan model also favored scleractinian monophyly in 99.9% of cases.

## Compositional bias effects: codons and amino acids

Clear biases in codon usage are seen throughout the Hexacorallia, but in "Robust" corals, this bias is more extreme, as evidenced by consistently lower effective number of codon (NC) scores and higher codon adaptation indices (calculated using CodonW [73]) than other scleractinians or corallimorpharians (Figure 6). For some amino acids, codon usage in "Robust" corals differed markedly from that in the other hexacorallians for which data are available (Figure 2). This pattern was also seen in the AT skew analyses (Figure S4).

In both aa and codon based phylogenies (Figures 3b and 4 respectively), the branch leading to the "Robust" coral clade is disproportionately long due to the presence of a large number of aa substitutions that are shared across most or all members of this clade but are not seen in other hexacorallians. Phenylalanine is the second most abundant aa in the mt-encoded proteins of "Robust" corals, and is approximately 1.5-fold more abundant in members of this clade, compared to other hexacorallians (Figure 2). The observed increase in abundance of (TTT-encoded) Phe residues in

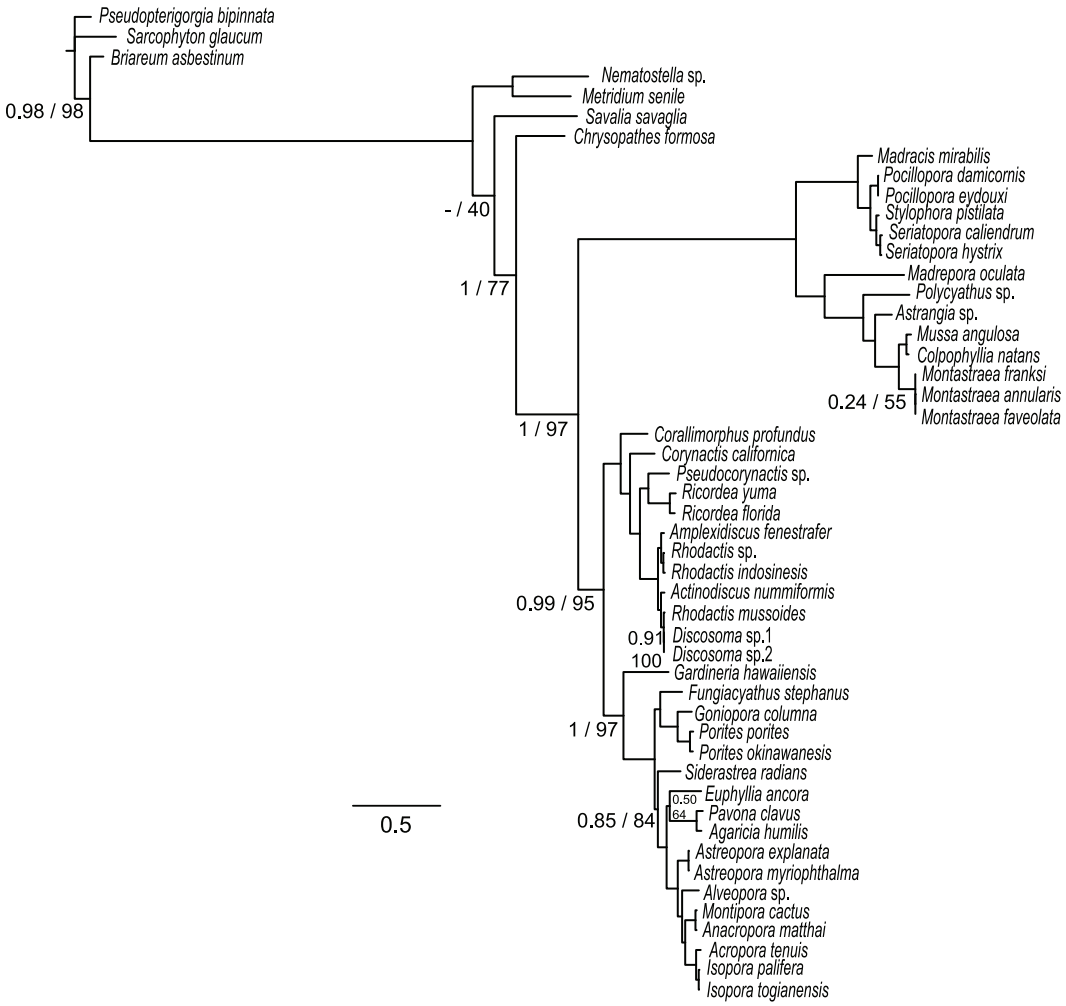



**Figure 4. Codon-based phylogenetic analyses inferred using CodonPhyml – Yap et al. model** [52] **for Maximum Likelihood – and MrBayes – GTR substitution model and three categories of non-synonymous/synonymous ratios (M3 model).** Values on the nodes indicate the non-parametric SH test and bootstrap replicates (ML), and posterior probability (BI) support respectively. Where no values are shown on a node, that edge was fully supported under all analyses.
doi:10.1371/journal.pone.0094774.g004

**Table 3.** Comparison of the two competing topologies (scleractinian monophyly, SM, and "naked coral", NC) using the Approximately Unbiased (AU), Kishino-Hasegawa (KH) and Shimodaira-Hasegawa (SH) tests for a variety of tree reconstruction methods.

| | Best ML topology | AU | KH | SH |
|---|---|---|---|---|
| Nucleotides (GTR+G+I+F) | SM | 0.10 | 0.11 | 0.11 |
| Nucleotides (GG98) | NC | 0.33 | 0.32 | 0.32 |
| Codons (YAP+W+K+F) | NC | 0.03 | 0.03 | 0.03 |
| Amino Acids (JTT+G+I+F) | NC | 0.007 | 0.008 | 0.008 |
| Nucleotides (rRNA, tRNA, GTR+G+I+F) | SM | 0.009 | 0.013 | 0.013 |
| Nucleotides (rRNA, tRNA, nhPhyML) | SM | 0.086 | 0.094 | 0.094 |

The p-values denote the probability that the best ML topology is equivalent to the alternative topologies. Unless otherwise indicated, the trees were based on the alignment of protein coding genes.
doi:10.1371/journal.pone.0094774.t003

"Robust" corals suggests that shifts in nt abundance may have forced the large number (>18%) of changes at the aa level that are unique to and shared within the "Robust" clade. (A+T)-rich amino acids (FYMICK) are more abundant, and (G+C)-rich amino acids (GARP) less abundant in the proteins encoded by the mt genomes of "Robust" corals than in other hexacorallians (Table 2).

To further investigate the possibility of long-branch artifacts on the amino-acid-based phylogenies, the PhyloBayes program was employed to infer trees using the CAT-Poisson, CAT-GTR and GTR models [74]. The CAT model allows different positions to evolve using a distinct substitution process and to have a different equilibrium frequency. It has been shown that this type of model is less sensitive to saturation and can reduce long-branch artifacts [75]. Table 4 summarizes the topologies and posterior probabilities obtained with this approach. The CAT-Poisson and CAT-GTR models tend to support scleractinian monophyly, while the GTR model is consistent with the naked coral hypothesis. These results are consistent with a recent study [27], where the monophyly of scleractinians received a modest support from the

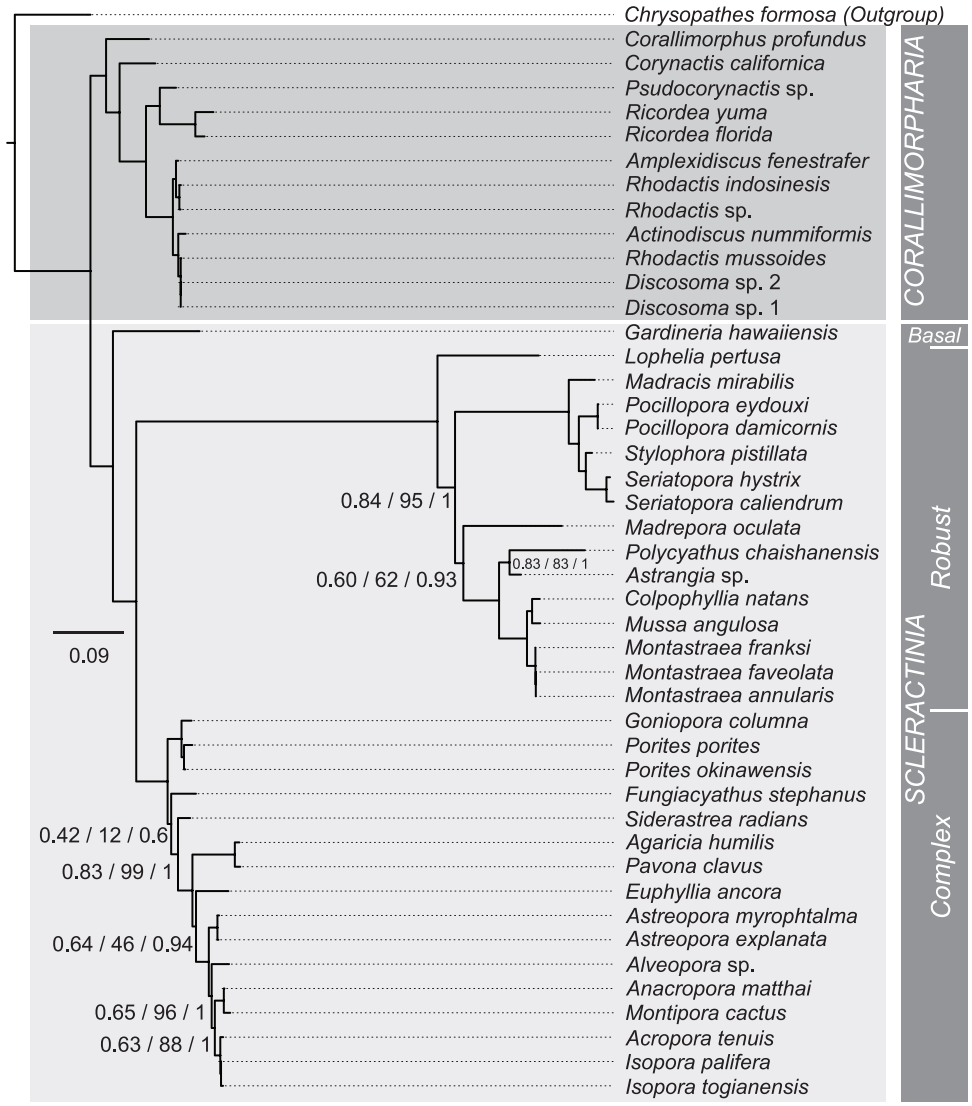



**Figure 5. Phylogenetic analyses of the mitochondrial sequences encoding stable RNAs (12S rRNA, 16S rRNA, *trnM*, and *trnW*).** Values on the nodes indicate the non-parametric SH test and bootstrap replicates (ML), and posterior probability (BI) support respectively. Where no values are shown on a node, that edge was fully supported under all analyses. Large boxes indicate the Corallimorpharia (dark-gray) and Scleractinia (light-gray) clades. Note that deep-water azooxanthellate species (*Corallimorphus profundus* and *Gardineria hawaiiensis*) represent the earliest diverging branches for Corallimorpharia and Scleractinia respectively.
doi:10.1371/journal.pone.0094774.g005

CAT-GTR model, but was not supported by the GTR model or by any nucleotide-based phylogenies.

## Hypothesis: did impaired mt DNA repair and constraints on tRNA uptake result in the altered amino acid usage in "Robust" corals?

If mt DNA repair is an ancestral property within Anthozoa [76], then the faster rate of mt genome evolution and differences in base composition that characterize "Robust" corals may reflect decreased efficiency of the repair process in this clade (see also [77]); if the resulting mismatches were not repaired, spontaneous deamination of cytosine to uracil could account for the atypical base composition. A consequence of the atypical base composition in the mt genomes of the "Robust" corals (higher T and lower C when compared to other hexacorallians) is a shift in the amino acid composition of mt proteins towards those encoded by T-rich codons.

The mt genomes of Anthozoa differ from those of the Bilateria in encoding only two tRNAs – tRNAMet and tRNATrp [78], hence in anthozoans most of the tRNAs required for mt translation must be imported. The mt tRNA uptake systems of anthozoans clearly have specificity, as codon use differs between mitochondrial and nuclear genes in *Acropora* despite similar overall base composition; for example, TTT being the strongly favoured Phe codon in mt-genes but bias being much less apparent in the case of nuclear genes [79,80].

"Robust" corals consistently display higher %(A+T) (around 5 to 6% higher than in "Complex" corals, for example) than either corallimorpharians or "Complex" corals, the most obvious difference being an increased frequency of thymine at third codon positions on the coding strand. In organisms that must import

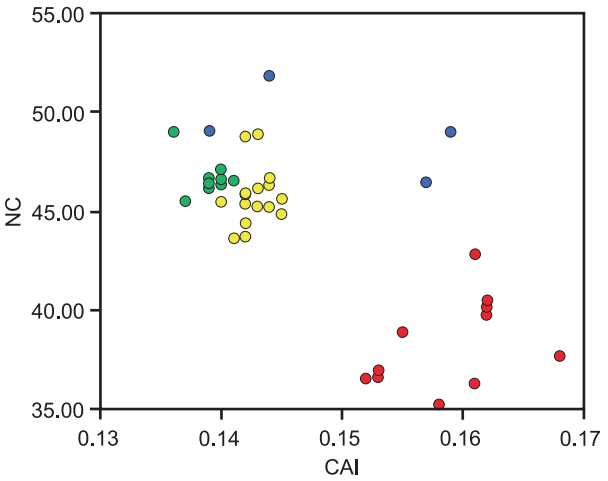



**Figure 6. Codon usage in hexacorallian anthozoans.** CodonW [73] was used to estimate codon usage biases; default settings were used in calculation of the codon adaptation index. NC: effective codon number. CAI: codon adaptation index. Colour coding and taxon choice is as shown in Figure 2.
doi:10.1371/journal.pone.0094774.g006

most tRNAs into mitochondria, changes in the base composition of the mt genome may lead to changes at the amino acid level in the proteins that they encode; the higher % (A+T) in the mt genomes of "Robust" corals not only drives protein coding sequences towards (A+T)-rich codons but may also force non-silent changes towards amino acid residues that are encoded by (A+T)-rich codons [81]. Such a mechanism could account for the higher abundance of phenylalanine residues in proteins encoded by the mt genomes of "Robust" corals, due to the increased frequency of TTT codons. Hence many of the amino acid substitutions unique to the proteins encoded by the mt genomes of "Robust" corals likely reflect the compound effects of base composition changes and the constraints under which tRNA uptake operates. We hypothesize that these amino acid substitutions bias phylogenetic analyses based on mitochondrial amino acid sequence data, obscuring relationships amongst the major scleractinian clades and corallimorpharians.

Consistent with compositional biases affecting analyses based on amino acid sequence data, phylogenetic analyses based on the mitochondrial rRNA and tRNA sequences consistently resulted in monophyletic Scleractinia (Figure 5). Furthermore, BF distance topologies inferred to assess the potential for compositional bias to affect phylogenetic inference suggested that overall, the aa data slightly favors the "naked coral" hypothesis, whereas nt based BF topology appears to be more homogeneous (Figure S5). Although the issue remains equivocal, molecular support for the "naked coral" hypothesis may therefore be an artifact resulting from compositional bias and saturation between the two major scleractinian clades. Note that these results do not challenge the issue of skeleton ephemerality sensu Stanley and Fautin [21] in Scleractinia, but imply that corallimorpharians are not descendants of a scleractinian that had undergone skeletal loss.

Changes in DNA repair mechanisms in some clades would result, in an evolutionary history, in violation of the assumptions of the models used for inference. With the exception of that of Barry and Hartigan, all models of substitution assume time-homogeneity both within and between branches [82]. All models of recoded sequences, including the aa substitution models, are non-Markovian, which results in a non-linear relationship between the true

**Table 4.** Posterior probability of the topologies recovered by different models using Phylobayes (SM: scleractinian monophyly, NC: naked coral).

| Model | Topology | Posterior probability |
|---|---|---|
| CAT GTR | SM | 0.56 |
| CAT Poisson | SM | 0.94 |
| GTR | NC | 1 |

doi:10.1371/journal.pone.0094774.t004

substitution dynamics operating on the nucleotide sequences and what is inferred using these models [83]. Thus, the ambiguous results outlined above could be a consequence of a poor fit between the models (despite these being selected as fitting the data best from the collection of models tested) and the evolutionary process. We evaluated how well the models fit compared to the best-possible likelihood, as proposed by Goldman [50] and implemented in the PyCogent library [49]. In brief, this test compares the difference in likelihood between the fitted model and the best-possible likelihood (calculated without assuming any phylogenetic relationship between the sequences) to the distribution of difference between these two likelihoods that one would expect if the data were generated according the fitted model (Figure S6). These tests can only be carried out in the maximum likelihood framework, thus for nt data the GTR and Barry-Hartigan models were tested and for aa data the JTT model. P-values were computed based on 200 Monte Carlo simulations, and for all models the fitted likelihoods were vastly inferior compared to the best possible, confirming a poor agreement between the data and the models used, even the most general one (with the fewest assumptions).

## Conclusion

The hypothesis outlined here – that, for hexacorallians, analyses based on mitochondrial sequences may be intrinsically biased - can and should be tested when appropriate nuclear sequence data are available for a wide range of corals and corallimorpharians. Molecular phylogenetics has led to radical revisions in thinking about coral evolution, but such analyses have largely been based on mt sequence data. Similar problems with mitochondrial sequences have been highlighted for a number of other animal groups including mammals [84] and beetles [85]. Given the above concerns, it is important that the bias towards mt data is redressed, and coral phylogenetics more broadly be based on a wide range of nuclear loci.

## Supporting Information

**Figure S1 Phylogenetic analyses based on the nucleotide sequences of the mitochondrial genes encoding proteins with the exclusion of the 3$^{rd}$ codon position.** Values on the nodes indicate the posterior probability (BI) support. Where no values are shown on a node, that edge was fully supported under all analyses.
(EPS)

**Figure S2 Phylogenetic analyses based on amino acid sequences of the mitochondrial genes encoding proteins removing the most rapidly evolving sites based on the mean (relative) evolutionary rate (ER) for each site under the JTT+G+F model.** A, B, and C indicate the topologies recovered excluding all those sites with ERs of ≥2.99,

≥1.99, or ≥1.49 respectively. For each ER reconstruction the topology with the highest log likelihood is shown. Values on the nodes indicate the ML bootstrap (100 replicates) support.
(EPS)

**Figure S3 Phylogenetic analyses based on the nucleotide sequences of the mitochondrial genes encoding proteins re-coded as purines and pyrimidies (RY-coding see** [44,45]**) with the exclusion of the third codon position from the resulting alignment.** Values on the nodes indicate the posterior probability (BI) support. Where no values are shown on a node, that edge was fully supported under all analyses.
(EPS)

**Figure S4 Graphical representation of (G+C)- (red line) and (A+T)- (blue line) skew calculated on the whole mitochondrial genome of all species included in the present analysis.** The (A+T)-skew is >6% higher in "Robust" corals than in all other hexacorallians included in the present analyses (highlighted in yellow).
(EPS)

**Figure S5 Minimum evolution tree on BF distances from the complete mt protein coding DNA sequences.** Topologies are based on nucleotide BF distances (left topology) and aa BF distances (right topology). AA compositional bias slightly favors the "naked coral" hypothesis (yellow box) whereas nt based BF topology appears to be more homogeneous.
(EPS)

**Figure S6 Empirical distribution of the difference between the likelihood of the fitted model and the best possible likelihood (the product of column pattern frequencies).** In each case the arrow indicates the observed value of that difference. The distributions are shown for the nucleotide alignment of the protein coding sequences with the GTR model (A) and the Barry and Hartigan model (B), for the amino acid alignment with the JTT model (C), and for the RNA alignment with the GTR model (D).
(EPS)

**Table S1 Primer names and sequences used for the amplification/sequence of the mitochondrial genome of *Gardineria hawaiiensis*.** The position and amplicons length of primers designed in the present study or the reference for previously published primers are provided.
(DOC)

**Table S2 Alignment of the nucleotide sequences from the mitochondrial genes encoding proteins re-coded as purines and pyrimidines (RY-coding see** [44,45]**).**
(TXT)

## Acknowledgments

The authors thank members of the Coral Reef Evolutionary Ecology and Genetics Laboratory (CREEG), Biodiversity Research Center, Academia Sinica (BRCAS) for assistance with sampling and field logistics. MVK acknowledges the support of the São Paulo Research Foundation (FAPESP) and São Paulo University Marine Biology Centre (CEBIMar). MFL gratefully acknowledges receipt of a James Cook University Postgraduate Research Scholarship (JCUPRS). This research was undertaken with the assistance of resources provided to SF at the NCI National Facility systems at the Australian National University through the National Computational Merit Allocation Scheme supported by the Australian Government. Coral samples were collected under appropriate permits. This is the CREEG, RCBAS contribution no. 95.

## Author Contributions

Conceived and designed the experiments: MK MFL DJM CAC. Performed the experiments: MK MFL. Analyzed the data: MK MFL SF GH. Contributed reagents/materials/analysis tools: DJM CAC. Wrote the paper: MK MFL SF DJM CAC.

## References

1. Roberts JM, Wheeler AJ, Freiwald A, Cairns SD (2009) Cold-water corals: the biology and geology of deep-sea coral habitats. Cambridge: Cambridge University Press. xvi+334 p.
2. Dower JF, Perry RI (2001) High abundance of larval rockfish over Cobb Seamount, an isolated seamount in the Northeast Pacific. Fish Oceanogr 10: 268–374.
3. Fischer E, Jones G (2012) Atmospheric dimethysulphide production from corals in the Great Barrier Reef and links to solar radiation, climate and coral bleaching. Biogeochemistry 110: 31–46.
4. Kleypas JA, Danabasoglu G, Laugh JM (2008) Potential role of the ocean thermostat in determining regional differences in coral reef bleaching events. Geophys Res Lett 35: L03613.
5. Reaka-Kudla ML (1997) Biodiversity II: understanding and protecting our biological resources. In: Reaka-Kudla ML, Wilson DE, Wilson EO editors. Washington DC: Joseph Henry Press. pp. 83–108.
6. Reed JK (2002) Deep-water *Oculina* coral reefs of Florida: biology, impacts, and management. Hydrobiologia 471: 43–55.
7. Rogers AD (1999) The biology of *Lophelia pertusa* (Linnaeus, 1758) and other deep-water reef-forming corals and impacts from human activities. Hydrobiologia 84: 315–406.
8. Barbeitos M, Romano S, Lasker HR (2010) Repeated loss of coloniality and symbiosis in scleractinian corals. Proc Natl Acad Sci USA 26: 11877–11882.
9. Fukami H, Chen CA, Budd AF, Collins A, Wallace C, et al. (2008) Mitochondrial and nuclear genes suggest that stony corals are monophyletic but most families of stony corals are not (Order Scleractinia, Class Anthozoa, Phylum Cnidaria). PLoS One 3: e3222.
10. Kitahara MV, Cairns SD, Miller DJ (2010a) Monophyletic origin of the *Caryophyllia* (Scleractinia; Caryophylliidae), with description of six new species. Syst Biodivers 8: 91–118.
11. Kitahara MV, Cairns SD, Stolarski J, Blair D, Miller DJ (2010b) A comprehensive phylogenetic analysis of the Scleractinia (Cnidaria, Anthozoa) based on mitochondrial CO1 sequence data. PloS One 5: e11490.
12. Kitahara MV, Cairns SD, Stolarski J, Miller DJ (2012a) Deltocyathiidae, an early-diverging family of Robust corals (Anthozoa, Scleractinia). Zool Scr 42: 201–212.
13. Kitahara MV, Stolarski J, Cairns SD, Benzoni F, Stake JL, et al. (2012b) The first modern solitary Agariciidae (Anthozoa, Scleractinia) revealed by molecular and microstructural analysis. Invertebr Syst 26: 303–305.
14. Le Goff-Vitry MC, Rogers AD, Baglow D (2004) A deep-sea slant on the molecular phylogeny of the Scleractinia. Mol Phylogenet Evol 30: 167–177.
15. Romano SL, Cairns SD (2000) Molecular phylogenetic hypotheses for the evolution of scleractinian corals. Bull Mar Sci 67: 1043–1068.
16. Romano SL, Palumbi SR (1996) Evolution of scleractinian corals inferred from molecular systematics. Science 271: 640–642.
17. Stolarski J, Kitahara MV, Miller DJ, Cairns SD, Mazur M, et al. (2011) The ancient evolutionary origins of Scleractinia revealed by azooxanthellate corals. BMC Evol Biol 11: 2–15.
18. Fukami H, Budd AF, Paulay G, Sole-Cava A, Chen CA, et al. (2004) Conventional taxonomy obscures deep divergence between Pacific and Atlantic corals. Nature 427: 832–835.
19. Huang D, Licuanan WY, Baird AH, Fukami H (2011) Cleaning up the "Bigmessidae": molecular phylogeny of scleractinian corals from Faviidae, Merulinidae, Pectiniidae and Trachyphylliidae. BMC Evol Biol 11: 37.
20. Huang D (2012) Threatened reef corals of the world. PLoS ONE 7: e34459.
21. Stanley GD Jr, Fautin DG (2001) The origin of modern corals. Science 291: 1913–1914.
22. Fine M, Tchernov D (2007) Scleractinian coral species survive and recover from decalcification. Science 315: 1811.
23. Rodolfo-Metalpa R, Martin S, Ferrier-Pages C, Gattuso JP (2011) Response of the temperate coral *Cladocora caespitosa* to mid- and long-term exposure to pCO sub and temperature levels projected for the year 2100 AD. Biogeosciences 7: 289–300.
24. Medina M, Collins AG, Takaoka TL, Kuehl JV, Boore JL (2006) Naked corals: skeleton loss in Scleractinia. Proc Natl Acad Sci USA 103: 96–100.
25. Culligan KM, Meyer-Gauen G, Lyons-Weiler J, Hays JB (2000) Evolutionary origin, diversification and specialization of eukaryotic MutS homolog mismatch repair proteins. Nucl Acids Res 28: 463–471.
26. Stephenson TA (1921) On the classification of Actiniaria. Part II. Consideration of the whole group and its relationships, with special reference to forms not treated in part I. Quart J Micr Sci 65: 493–576.

27. Kayal E, Roure B, Philippe H, Collins AG, Lavrov DV (2013) Cnidarian phylogenetic relationships as revealed by mitogenomics. BMC Evol Biol 13: 5.
28. Stampar SN, Maronna MM, Kitahara MV, Reimer JD, Morandini AC (2014) Fast-evolving mitochondrial DNA in Ceriantharia: a reflection of Hexacorallia paraphyly? PLoS One 9: e86612.
29. Huang D, Meier R, Todd PA, Chou LM (2008) Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. J Mol Evol 66: 167–174.
30. Shearer TL, Coffroth MA (2008) DNA BARCODING: Barcoding corals: limited by interspecific divergence, not intraspecific variation. Mol Ecol Resour 8: 247–255.
31. Pont-Kingdon GA, Okada NA, Macfarlane JL, Beagley CT, Wolstenholme DR, et al. (1995) A coral mitochondrial MutS gene. Nature 375: 109–111.
32. Tseng CC, Wallace CC, Chen CA (2005) Mitogenomic analysis of *Montipora cactus* and *Anacropora matthai* (Cnidaria; Scleractinia; Acroporidae) indicates an unequal rate of mitochondrial evolution among Acroporidae corals. Coral Reefs 24: 502–508.
33. Cheng S, Higuchi R, Stoneking M (1994) Complete mitochondrial genome amplification. Nature Genet 7: 350–351.
34. Chen CA, Yu JK (2000) Universal primers for amplification of mitochondrial small subunit ribosomal RNA-encoding gene in scleractinian corals. Mar Biotechnol 2: 146–503.
35. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol 3: 294–299.
36. Romano SL, Palumbi SR (1997) Molecular evolution of a portion of the mitochondrial 16S ribosomal gene region in scleractinian corals. J Mol Evol 45: 397–411.
37. Lin MF, Luzon KS, Licuanan WY, Ablan-Lagman CM, Chen AC (2011) Seventy-four universal primers for characterizing the complete mitochondrial genomes of scleractinian corals (Cnidaria; Anthozoa). Zool Stud 50: 513–524.
38. Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20: 3252–3255.
39. Stothard P (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28: 1102–1104.
40. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.
41. Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: Improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33: 511–518.
42. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696–704.
43. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics Applications Note 17: 754–755.
44. Harrison GL, McLenachan PA, Phillips MJ, Slack KE, Cooper A, et al. (2004) Four new avian mitochondrial genomes help get to basic evolutionary questions in the late Cretaceous. Mol Biol Evol 21: 974–983.
45. Phillips MJ, Gibb GC, Crimp EA, Penny D (2010) Tinamous and moa flock together: mitochondrial genome sequence analysis reveals independent losses of flight among ratites. Syst Biol 59: 90–107.
46. Hrdy I, Hirt RP, Dolezal P, Bardonova L, Foster PG, et al. (2004). *Trichomonas* hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. Nature 432: 618–622.
47. Rodríguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang F, et al. (2007) Detecting and overcoming systematic errors in genome-scale phylogenies. Syst Biol 56: 389–399.
48. Boussau B, Gouy M (2006) Efficient Likelihood Computations with Non-Reversible Models of Evolution. Syst Biol 55: 756–68.
49. Knight R, Maxwell P, Birmingham A, Carnes J, Caporaso JG, et al. (2007) PyCogent: a toolkit for making sense from sequence. Genome Biol 8: R171.
50. Goldman N (1993) Statistical tests of models of DNA substitution. J Mol Evol 36: 182–198.
51. Gil M, Zanetti MS, Zoller S, Anisimova M (2013) CodonPhyML: fast maximum likelihood phylogeny estimation under codon substitution models. Mol Biol Evol 30: 1270–1280.
52. Yap VB, Lindsay H, Easteal S, Huttley G (2010) Estimates of the effect of natural selection on protein-coding content. Mol Biol Evol 27: 726–734.
53. Lartillot N, Lepage T, Blanquart S (2009) PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25: 2286–2288.
54. Shimodaira H, Hasegawa M (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17: 1246–1247.
55. Rosenberg MS, Kumar S (2003) Heterogeneity of nucleotide frequencies among evolutionary lineages and phylogenetic inference. Mol Biol Evol 20: 610–621.
56. Phillips MJ, Penny D (2003) The root of the mammalian tree inferred from whole mitochondrial genomes. Mol Phylogenet Evol 28: 171–185.
57. Riemann-Zürneck K, Iken K (2003) *Corallimorphus profundus* in shallow Antarctic habitats: bionomics, histology, and systematics (Cnidaria: Hexacorallia). Zool Verh Leiden 345: 367–386.
58. Stolarski J (1995) Ontogenetic development of the thecal structures in Caryophylliinae scleractinian corals. Acta Palaeontol Pol 40: 19–44.
59. Chen C, Dai CF, Plathong S, Chiou CY, Chen CA (2008a) The complete mitochondrial genomes of needle corals, *Seriatopora* spp. (Scleractinia: Pocilloporidae): an idiosyncratic *atp8*, duplicated *trnW* gene, and hypervariable regions used to determine species phylogenies and recently diverged populations. Mol Phylogenet Evol 46: 19–33.
60. Park E, Song J, Won Y (2011) The complete mitochondrial genome of *Calicogorgia granulosa* (Anthozoa:Octocorallia): potential gene novelty in unidentified ORFs formed by repeat expansion and segmental duplication. Gene 486: 81–87.
61. Chen C, Chiou CY, Dai CF, Chen CA (2008b) Unique mitogenomic features in the scleractinian family Pocilloporidae (Scleractinia: Astrocoeniina). Mar Biotechnol 10: 538–553.
62. Fukami H, Knowlton N (2005) Analysis of complete mitochondrial DNA sequences of three members of the *Montastraea annularis* species complex (Cnidaria, Anthozoa, Scleractinia). Coral Reefs 24: 410–417.
63. Park E, Hwang DS, Lee JS, Song JI, Seo TK, et al. (2012) Estimation of divergence times in cnidarian evolution based on mitochondrial protein-coding genes and the fossil record. Mol Phylogenet Evol 62: 329–345.
64. van Oppen MJ, Catmull J, McDonald BJ, Hislop NR, Hagerman PJ, et al. (2000) The mitochondrial genome of *Acropora tenuis* (Cnidaria; Scleractinia) contains a large group I intron and a candidate control region. J Mol Evol 55: 1–13.
65. Lin MF, Kitahara MV, Tachikawa H, Fukami H, Miller DJ, et al. (2012) Novel organization of the mitochondrial genome in the deep-sea coral, *Madrepora oculata* (Hexacorallia, Scleractinia, Oculinidae) and its taxonomic implications. Mol Phylogenet Evol 65: 323–328.
66. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10: 512–526.
67. Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22: 160–174.
68. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8: 275–282.
69. Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18: 691–699.
70. Dayhoff M, Schwartz R, Orcutt B (1978) A model of evolutionary change in protein. In: Dayhoff MO editor. Atlas of protein sequences and structure. Washington, DC: National Biomedical Research Foundation. pp. 345–351.
71. van Den Bussche RA, Baker RJ, Huelsenbeck JP, Hillis DM (1998) Base compositional bias and phylogenetic analyses: a test of the "flying DNA" hypothesis. Mol Phylogenet Evol 13: 408–416.
72. Galtier N, Gouy M (1998) Inferring pattern and process: Maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol Biol Evol 15: 871–879.
73. Peden J (2005) CodonW version 1.4.2. Available: http://codonw.sourceforge.net/ culong.html#License
74. Lartillot N. Philippe H (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol 21: 1095–1109.
75. Lartillot N, Brinkmann H, Philippe H (2007) Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol (Suppl 1): S4.
76. Shearer TL, van Oppen MJH, Romano SL, Worheide G (2002) Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). Mol Ecol 11: 2475–2487.
77. Aranda M, DeSalvo MK, Bayer T, Medina M, Voolstra C (2012) Evolutionary insights into scleractinian corals using comparative genomic hybridizations. BMC Genomics 13: 501.
78. Beagley CT, Okimoto R, Wolstenholme DR (1998) The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): introns, a paucity of tRNA genes, a near-standard genetic code. Genetics 148: 1091–1108.
79. Shinzato C, Shoguchi E, Kawashima T, Hamada M, Hisata K, et al. (2011) Using the *Acropora digitifera* genome to understand coral responses to environmental change. Nature 476: 320–323.
80. van Oppen MJH, Hislop NR, Hagerman PJ, Miller DJ (1999) Gene content and organization in a segment of the mitochondrial genome of the scleractinian coral *Acropora tenuis*: major differences in gene order within the anthozoan subclass Zoantharia. Mol Biol Evol 16: 1812–1215.
81. Knight RD, Freeland SJ, Landweber LF (2001) A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. Genome Biol 2: 4.
82. Barry D, Hartigan JA (1987) Statistical analysis of Hominoid molecular evolution. Stat Sci 2: 191–207.
83. Kosiol C (2006) Markov models for protein sequence evolution. [Ph.D. thesis]. EMBL-European Bioinformatics Institute. Cambridge: University of Cambridge.
84. Huttley G (2009) Do genomic datasets resolve the correct relationship among the placental, marsupial and monotreme lineages? Austral J Zool 57: 167–174.
85. Sheffield NC, Song H, Cameron SL, Whiting MF (2009) Nonstationary evolution and compositional heterogeneity in beetle mitochondrial phylogenomics. Syst Biol 58: 381–394.
86. Beaton MJ, Roger AJ, Cavalier-Smith T (1998) Sequence analysis of the mitochondrial genome of *Sarcophyton glaucum*: conserved gene order among octocorals. J Mol Evol 47: 697–708.

87. Brugler MR, France SC (2007) The complete mitochondrial genome of the black coral *Chrysopathes formosa* (Cnidaria: Anthozoa: Antipatharia) supports classification of antipatharians within the subclass Hexacorallia. Mol Phylogenet Evol 42: 776–788.
88. Flot JF, Tillier S (2007) The mitochondrial genome of *Pocillopora* (Cnidaria: Scleractinia) contains two variable regions: the putative D-loop and a novel ORF of unknown function. Gene 401: 80–87.
89. Sinniger F, Chevaldonné P, Pawlowski J (2007) Mitochondrial genome of *Savalia savaglia* (Cnidaria, Hexacorallia) and early metazoan phylogeny. J Mol Evol 64: 196–203.
90. Emblem A, Karlsen BO, Evertsen J, Johansen SD (2011) Mitogenome rearrangement in the cold-water scleractinian coral *Lophelia pertusa* (Cnidaria, Anthozoa) involves a long-term evolving group I intron. Mol Phylogenet Evol 61: 495–503.
91. Adachi J, Waddell PJ, Martin W, Hasegawa M (2000) Plastid genome phylogeny and a model of amino acid substitution for protein encoded by chloroplast DNA. J Mol Evol 50: 348–358.
92. Daly M, Fautin DG, Cappola VA (2003) Systematics of the Hexacorallia (Cnidaria: Anthozoa). Zool J Linn Soc-Lond 139: 419–437.
93. Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures Math Life Sci 17: 57–86.

## Appendix D

Mitochondrial Genome Rearrangements in the Scleractinia/Corallimorpharia Complex: Implications for Coral Phylogeny.

Mei-Fang Lin, Marcelo Visentini Kitahara, Haiwei Luo, Dianne Tracey, Jonathan Geller, Hironobu Fukami, David John Miller, Chaolun Allen Chen.

Genome Biology and Evolution 6:1086-1095, 2014

GBE

Genome Biology and Evolution

SMBE

# Mitochondrial Genome Rearrangements in the Scleractinia/Corallimorpharia Complex: Implications for Coral Phylogeny

Mei-Fang Lin[1,2,3,†], Marcelo Visentini Kitahara[4,†], Haiwei Luo[5], Dianne Tracey[6], Jonathan Geller[7], Hironobu Fukami[8], David John Miller[2,3,*], and Chaolun Allen Chen[1,9,10,*]

[1]Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

[2]School of Pharmacy and Molecular Sciences, James Cook University, Townsville, Australia

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

[4]Centro de Biologia Marinha, Universidade de São Paulo, São Sebastião, Brazil

[5]Department of Marine Sciences, University of Georgia

[6]New Zealand Institute of Water and Atmospheric Research, Wellington, New Zealand

[7]Moss Landing Marine Laboratories, California State University, Moss Landing

[8]Department of Marine Biology and Environmental Science, Faculty of Agriculture, The University of Miyazaki, Japan

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

[10]Institute of Oceanography, National Taiwan University, Taipei, Taiwan

*Corresponding author: E-mail: david.miller@jcu.edu.au; cac@gate.sinica.edu.tw.

[†]These authors contributed equally to this work.

**Accepted:** April 18, 2014

## Abstract

Corallimorpharia is a small Order of skeleton-less animals that is closely related to the reef-building corals (Scleractinia) and of fundamental interest in the context of understanding the potential impacts of climate change in the future on coral reefs. The relationship between the nominal Orders Corallimorpharia and Scleractinia is controversial—the former is either the closest outgroup to the Scleractinia or alternatively is derived from corals via skeleton loss. This latter scenario, the "naked coral" hypothesis, is strongly supported by analyses based on mitochondrial (mt) protein sequences, whereas the former is equally strongly supported by analyses of mt nucleotide sequences. The "naked coral" hypothesis seeks to link skeleton loss in the putative ancestor of corallimorpharians with a period of elevated oceanic $CO_2$ during the Cretaceous, leading to the idea that these skeleton-less animals may be harbingers for the fate of coral reefs under global climate change. In an attempt to better understand their evolutionary relationships, we examined mt genome organization in a representative range (12 species, representing 3 of the 4 extant families) of corallimorpharians and compared these patterns with other Hexacorallia. The most surprising finding was that mt genome organization in *Corallimorphus profundus*, a deep-water species that is the most scleractinian-like of all corallimorpharians on the basis of morphology, was much more similar to the common scleractinian pattern than to those of other corallimorpharians. This finding is consistent with the idea that *C. profundus* represents a key position in the coral <-> corallimorpharian transition.

**Key words:** naked coral hypothesis, gene order, mitochondrial genome, coral evolution.

## Introduction

Understanding the evolutionary history of the Scleractinia and relationships between corals and other members of the anthozoan subclass Hexacorallia should enable a better understanding of how it has been influenced by climate in the past and thus enable better predictions of the likely impacts of climate change (Romano and Palumbi 1996). Of the six Orders of hexacorals, only members of the Scleractinia develop continuous external calcified skeletons (Daly et al. 2003). The Scleractinia suddenly appear in the fossil record in the middle Triassic, about 240 Ma, but the range of morphological variation seen in the Middle Triassic fossils is comparable to that of extant scleractinians (Romano and

Palumbi 1996). Molecular phylogenies based on both mitochondrial (mt) and nuclear (nucl) genes imply a deeper divergence (~300 Ma—in the Late Carboniferous) of extant scleractinians into two major clades, the "Complexa" and the "Robusta" (Romano and Palumbi 1996; Romano and Cairns 2000; Chen et al. 2002; Le Goff-Vitry et al. 2004; Fukami et al. 2008; Barbeitos et al. 2010; Kitahara, Cairns, and Miller 2010; Kitahara, Cairns, Stolarski, et al. 2010; Kitahara, Cairns, et al. 2012; Kitahara et al. 2012; Kayal et al. 2013). By adding deep-water species to existing molecular data sets and applying an appropriately calibrated molecular clock, Stolarski et al. (2011) demonstrated that two exclusively deep-sea families, the Gardineriidae and Micrabaciidae, form a "basal" clade that diverged at around 425 Ma, prior to the Complexa/Robusta split, pushing the evolutionary origin of scleractinians deep into the Paleozoic. These results support the scenario that scleractinians are the descendants of soft-bodied (corallimorpharian-like) ancestors that survived the mass extinction at the Permian/Triassic boundary and subsequently gained the ability to deposit calcified skeletons (Stolarski et al. 2011).

The "naked coral" hypothesis, first put forward by Stanley and Fautin (2001) to explain the sudden appearance of diverse scleractinian fauna in the middle Triassic, is based on the idea that the skeleton has been an ephemeral trait during coral evolution. Under this hypothesis, the Scleractinia were skeleton-less in the early Triassic, a time when carbonate deposition was suppressed globally (Stanley 2003). Consistent with the idea of skeleton ephemerality, some coral species can undergo reversible skeleton loss under acid conditions (Fine and Tchernov 2007). Strong phylogenetic support for the "naked coral" hypothesis came from analyses based on the alignment of concatenated proteins encoded by 17 complete mt genomes from hexacorallians (Medina et al. 2006); in their analysis, scleractinians were paraphyletic, corallimorpharians being more closely related to the Complexa than are Robusta, the interpretation being that the Corallimorpharia arose by skeleton loss from a scleractinian ancestor at a time (during the mid-Cretaceous) of high oceanic $CO_2$ levels (Medina et al. 2006).

Although the "naked coral" scenario is supported by analyses of protein sequence data, phylogenetics based on mt nucleotide sequences instead strongly support scleractinian monophyly (Stolarski et al. 2011; Kayal et al. 2013; Kitahara et al. 2014). The fundamental disagreement between phylogenies based on nucleotide (fig. 1*A*) or amino acid (fig. 1*B*) sequence data for mt proteins stems from the fact that none of the available models for sequence evolution adequately account for the observed data (Kitahara et al. 2014). One possible explanation for this is the occurrence of a "catastrophic" event—a major and unpredictable change, such as sudden impairment of mt DNA repair processes (which are believed to be an ancestral trait within Anthozoa (Pont-Kingdon et al. 1998; Shearer et al. 2002; Brockman and McFadden 2012).

Given the intractability of coral/corallimorph relationships using conventional molecular phylogenetics, we explored the informativeness of mt genome architecture in this context. mt gene rearrangements occur relatively infrequently and have proven useful in resolving evolutionary relationships, both shallow and deep, across a broad range of organisms (e.g., Gai et al. 2008; Brockman and McFadden 2012; Kilpert et al. 2012). This study is based on the complete mt genomes of a total of 12 corallimorpharians (8 of which are novel), representing 3 of 4 currently described families (Daly et al. 2007; Fautin et al. 2007), and 32 scleractinians, and includes both the early diverging coral *Gardineria hawaiiensis* (Stolarski et al. 2011), and corallimorpharian, *Corallimorphus profundus*, which is considered to be the most coral-like of corallimorpharians based on morphological grounds (Moseley 1877; den Hartog 1980; Riemann-Zürneck and Iken 2003). The results indicate that, by contrast with the Scleractinia, extensive rearrangements of the mt genome have occurred within Corallimorpharia. The most surprising finding, however, was that the mt genome of *C. profundus* is scleractinian-like, and is organized very differently to those of all other corallimorpharians for which data are available. Both nucleotide and amino acid sequenced-based phylogenetics unequivocally place *C. profundus* as an early diverging corallimorpharian, indicating that this organism most closely reflects the coral <-> corallimorpharian transition.

## Materials and Methods

### DNA Extraction, Polymerase Chain Reaction, Long Polymerase Chain Reaction, Cloning, and Sequencing

Genomic DNA was extracted from corallimorpharian samples that had been preserved in 95% (V/W) ethanol following Chen et al. (2002)—sampling information is summarized in table 1. Long-range polymerase chain reaction (L-PCR; Cheng et al. 1994) was used to amplify large (6–9 kb) and overlapping fragments covering the entire mt genomes of corallimorpharians and corals. For each species, either two- or three-specific primer pairs were designed on the basis of previously available partial sequence data for of *rns*, *rnl*, and *COI* (Folmer et al. 1994; Romano and Palumbi 1997; Chen and Yu 2000; Lin et al. 2011) (supplementary table S1, Supplementary Material online). Reactions were set up in a total volume of 50 µl: 10× LA PCR buffer, 2.5 mM $MgCl_2$, 2.5 mM of each dNTP, 2.5 units of *TaKaRa La Taq*, 0.5 µm of each primer, and approximately 0.5 µg of genomic DNA. The L-PCR conditions were slightly modified from those recommended by the polymerase manufacturer as follows: 94 °C for 1 min, then 30 cycles of 10 s at 98 °C, 45 s at 62–63 °C, 14.25 min at 68 °C, and 10 min at 72 °C. PCR products were recovered from the agarose gel using the TOPO XL gel purification

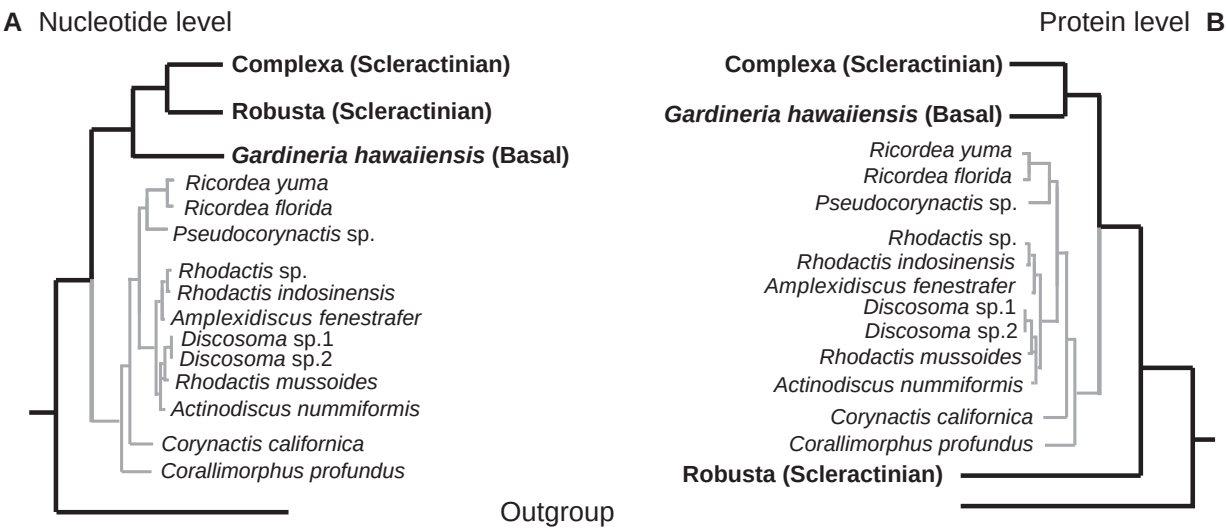



**Fig. 1.**—Alternative phylogenetic hypotheses for relationships between Scleractinia and Corallimorpharia based on mt genome nucleotide sequences (*A*) or the amino acid sequences of the proteins that they encode (*B*). The trees were modified from Kitahara et al. (2014). Note that, for both (*A*) and (*B*) scenarios, support for the node separating Corallimorpharia from Scleractinia (the root of the gray part of the tree) was over 97% under both maximum-likelihood analysis and Bayesian inference.

method, cloned into a pCR-XL-PCR vector system using topoisomerase I (Invitrogen), and transformed into *Escherichia coli* (Top10) by electroporation. The nucleotide sequences were determined for complementary strains of two to six clones from each sample using primer walking on the same PCR product by an ABI 377 Genetic Analyzer (Applied Biosystems). The M13 forward and reverse primers were used to obtain the initial sequences from the ends of each insertion. The consensus sequences from three sequenced clones were present for each species.

### Genome Annotation and Sequence Analysis

Sequences were verified and assembled using SeqManII (DNAstar v5.0) or Sequencher v4.8 (Gene Codes Corporation) and then analyzed in Vector NTI v9.0 (InforMax). Open-reading frames (ORFs) of length more than 50 (amino acids) were translated using National Center for Biotechnology Information translation table 4 and compared with the databases using BlastX (Gish and States 1993). No novel ORFs were identified on this basis. MEGA v5.0 (Tamura et al. 2011) with a weighted matrix of Clustal W (Thompson et al. 1994) was used to align the identical putative ORFs and rRNA genes with previously published data. The 5′- and 3′-ends of the rRNA genes were predicted using the program SINA on the Silva ribosomal RNA database site (www.arb-silva.de/, last accessed February 1, 2014) using the default settings (Pruesse et al. 2012). tRNAs were predicted using tRNAscan-SE search server v1.21 (Lowe and Eddy 1997). rRNA loci were identified on the basis of sequence similarity. Finally, Vector NTI v9.0 was used to generate maps of the mt genomes based on the assembled sequence data.

### Gene Order Phylogeny

The double cut and join (DCJ) distance metric (Yancopoulos et al. 2005), implemented in GRAPPA (Moret et al. 2002; Zhang et al. 2009), was used to calculate the pairwise DCJ and breakpoint distances (BPDs) from the gene order data and to generate pairwise distance matrixes. Gene order phylogenies (DCJ and BPD) were estimated with FastME (Desper and Gascuel 2002).

Because gene order is a single character with multiple states (Shi et al. 2010), bootstrapping is not applicable, hence the reliability of each branch was estimated by applying a jackknife resampling technique that in each iteration randomly removed 25% of the initial orthologous gene sets. Note that, because the data set consisted of only 13 protein-coding genes, higher removal rates (e.g., 50%) are unable to resolve the tree branching order. Jackknifing was used to generate 1,000 matrices, which were imported into FastME and used to obtain 1,000 DCJ- and BPD-based trees. Finally, the CONSENSE program in the PHYLIP software package (Felsenstein 1989) was used to calculate majority-rule consensus trees with percent values at each node. Each value represents the percentage of trees supporting a clade defined by a node.

## Results

### Characteristics of mt Genomes of Corallimorpharians and *Gardineria hawaiiensis*

The molecular characteristics of the mt genomes of a representative range (8) of corallimorpharians and the "basal" scleractinian *G. hawaiiensis* are summarized in table 1, along

**Table 1**

Characteristics of the mt Genomes of Corallimorpharians, Scleractinians, and Other Anthozoans

| Order | Scleractinian Clades | Species | Total Length (bp) | Nucleotide (%) | | Gene Size (bp) | | | | | | | | | | | | | | | | | | Species Collection Site and GenBank No. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | A+T | C+G | atp6 | atp8 | cob | COI | COI intron | COII | COIII | nd1 | nd2 | nd3 | nd4 | nd4l | nd5 | nd6 | rnl | rns | trnM | trnW | IGS (length) | |
| Corallimorpharia | | *Actinodiscus nummiformis* | 20,922 | 60.9 | 39 | 699 | 210 | 1,161 | 1,581 | 1,209 | 756 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,839 | 612 | 2,350 | 1,304 | 71 | 70 | 3,032 | Wanlitung, Taiwan |
| | | *Amplexidiscus fenestrafer* | 20,188 | 61 | 39 | 699 | 210 | 1,161 | 1,561 | 1,207 | 756 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,839 | 612 | 2,349 | 1,304 | 71 | 70 | 2,729 | Taioshi, Taiwan |
| | | *Corallimorphus profundus*[a] | 20,488 | 60.3 | 39.6 | 699 | 237 | 1,149 | 1,734 | 1,183 | 744 | 789 | 984 | 1,098 | 357 | 1,521 | 300 | 1,839 | 630 | 2,434 | 1,253 | 71 | 70 | 3,396 | Southern Ocean |
| | | *Corynactis californica*[a] | 20,632 | 60.2 | 39.8 | 699 | 219 | 1,179 | 1,602 | 1,266 | 765 | 789 | 984 | 1,098 | 357 | 1,479 | 300 | 1,839 | 612 | 2,552 | 1,256 | 71 | 69 | 2,906 | California, USA |
| | | *Discosoma* sp. 1 | 20,908 | 61 | 38.9 | 699 | 210 | 1,161 | 1,581 | 1,208 | 756 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,839 | 615 | 2,340 | 1,224 | 71 | 70 | 4,284 | NC_008071 |
| | | *Discosoma* sp. 2 | 20,912 | 61 | 38.9 | 699 | 210 | 1,161 | 1,581 | 1,207 | 756 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,839 | 615 | 2,342 | 1,068 | 71 | 70 | 4,289 | NC_008072 |
| | | *Pseudocorynactis* sp. | 21,239 | 60.9 | 39 | 699 | 213 | 1,230 | 1,575 | 1,178 | 756 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,839 | 612 | 2,537 | 1,223 | 71 | 70 | 3,177 | Birch Aquarium at SIO |
| | | *Rhodactis indosinesis* | 20,092 | 60.9 | 39.1 | 699 | 210 | 1,161 | 1,581 | 1,205 | 756 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,839 | 612 | 2,350 | 1,303 | 71 | 71 | 2,624 | Wanlitung, Taiwan |
| | | *R. mussoides* | 20,826 | 61 | 39 | 699 | 210 | 1,161 | 1,581 | 1,207 | 756 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,839 | 615 | 2,355 | 1,304 | 71 | 70 | 3,007 | Taioshi, Taiwan |
| | | *Rhodactis* sp. | 20,093 | 61 | 39 | 699 | 210 | 1,161 | 1,581 | 1,207 | 756 | 789 | 984 | 1,098 | 357 | 1,413 | 300 | 1,839 | 612 | 2,348 | 1,240 | 71 | 70 | 3,358 | NC_008158 |
| | | *Ricordea florida* | 21,376 | 62.1 | 37.9 | 699 | 210 | 1,140 | 1,623 | 1,180 | 756 | 789 | 987 | 1,098 | 357 | 1,476 | 300 | 1,839 | 606 | 2,447 | 1,218 | 71 | 70 | 4,510 | NC_008159 |
| | | *Ri. yuma* | 22,015 | 62.4 | 37.6 | 699 | 213 | 1,140 | 1,599 | 1,199 | 756 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,839 | 606 | 2,444 | 1,262 | 71 | 70 | 5,148 | Wanlitung, Taiwan |
| Scleractinia | Basal | *Gardineria hawaiiensis*[a] | 19,429 | 60.3 | 39.7 | 699 | 264 | 1,197 | 1,584 | 1,136 | 738 | 789 | 981 | 1,098 | 357 | 1,452 | 300 | 1,836 | 615 | 2,400 | 1,159 | 71 | 70 | 2,278 | New Caledonia |
| | Complex | *Acropora tenuis* | 18,338 | 62.1 | 37.9 | 699 | 219 | 1,155 | 1,602 | | 744 | 780 | 984 | 1,098 | 357 | 1,476 | 300 | 1,836 | 594 | 2,261 | 1,176 | 71 | 70 | 3,615 | NC_003522 |
| | | *Agaricia humilis* | 18,735 | 59.6 | 40.4 | 699 | 196 | 1,152 | 1,581 | | 744 | 789 | 984 | 1,098 | 357 | 1,479 | 300 | 1,836 | 594 | 1,577 | 1,136 | 70 | 69 | 4,773 | NC_008160 |
| | | *Alveopora* sp. | 18,146 | 62.2 | 37.8 | 699 | 237 | 1,158 | 1,602 | | 744 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,836 | 594 | 2,261 | 1,125 | 71 | 70 | 3,444 | KJ634271 |
| | | *Anacropora matthai* | 17,888 | 61.6 | 38.4 | 699 | 219 | 1,158 | 1,602 | | 744 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,836 | 594 | 2,261 | 1,174 | 71 | 70 | 3,155 | NC_006898 |
| | | *Astreopora explanata* | 18,106 | 62.2 | 37.8 | 699 | 219 | 1,146 | 1,587 | | 744 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,836 | 594 | 2,243 | 1,176 | 71 | 70 | 3,416 | KJ634269 |
| | | *Astreopora myriophthalma* | 18,106 | 62.1 | 37.8 | 699 | 219 | 1,146 | 1,587 | | 744 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,836 | 594 | 2,244 | 1,176 | 71 | 70 | 3,415 | KJ634272 |
| | | *Euphyllia ancora* | 18,875 | 62.3 | 37.8 | 699 | 219 | 1,155 | 2,301 | | 744 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,863 | 594 | 2,308 | 1,177 | 71 | 70 | 3,369 | NC_015641 |
| | | *Fungiacyathus stephanus*[a] | 19,381 | 62.2 | 37.8 | 699 | 231 | 1,161 | 1,629 | 962 | 744 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,839 | 594 | 2,366 | 1,114 | 71 | 70 | 3,596 | NC_015640 |
| | | *Goniopora columna* | 18,766 | 62.9 | 37.1 | 699 | 216 | 1,164 | 1,652 | 947 | 744 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,836 | 594 | 2,227 | 1,029 | 69 | 70 | 3,214 | NC_015643 |
| | | *Isopora palifera* | 18,725 | 61.7 | 38.2 | 699 | 219 | 1,158 | 1,602 | | 744 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,836 | 594 | 2,259 | 1,175 | 71 | 70 | 3,993 | KJ634270 |
| | | *Isopora togianensis* | 18,637 | 61.8 | 38.2 | 699 | 219 | 1,158 | 1,602 | | 744 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,836 | 594 | 2,259 | 1,177 | 71 | 70 | 3,903 | KJ634268 |
| | | *Montipora cactus* | 17,887 | 61.6 | 38.4 | 699 | 219 | 1,158 | 1,602 | | 744 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,836 | 594 | 2,266 | 1,172 | 71 | 70 | 3,151 | NC_006902 |
| | | *Pavona clavus* | 18,315 | 59.5 | 40.5 | 699 | 219 | 1,152 | 1,581 | | 744 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,836 | 606 | 2,299 | 1,169 | 70 | 69 | 3,566 | NC_008165 |
| | | *Porites okinawanesis* | 18,647 | 63.8 | 36.2 | 699 | 216 | 1,161 | 1,531 | 966 | 744 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,836 | 594 | 2,301 | 1,029 | 71 | 70 | 3,124 | NC_15644 |
| | | *Porites porites* | 18,648 | 63.7 | 36.2 | 699 | 216 | 1,161 | 1,578 | 966 | 744 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,836 | 594 | 2,271 | 1,060 | 71 | 70 | 3,077 | NC_008166 |
| | | *Siderastrea radians* | 19,387 | 63.1 | 36.9 | 699 | 234 | 1,155 | 1,584 | 989 | 744 | 789 | 984 | 1,098 | 357 | 1,476 | 300 | 1,836 | 594 | 2,242 | 1,296 | 71 | 70 | 3,568 | NC_008167 |
| | Robust | *Astrangia* sp. | 14,853 | 68.1 | 31.9 | 678 | 198 | 1,140 | 1,551 | | 685 | 780 | 948 | 1,092 | 342 | 1,440 | 300 | 1,812 | 561 | 1,178 | 532 | 72 | 70 | 1,474 | NC_008161 |
| | | *Colpophyllia natans* | 16,906 | 66.4 | 33.5 | 678 | 198 | 1,140 | 1,566 | | 685 | 780 | 948 | 1,104 | 342 | 1,440 | 300 | 1,815 | 561 | 1,885 | 1,012 | 72 | 70 | 2,310 | NC_008162 |
| | | *Lophelia pertusa*[a] | 16,150 | 65.1 | 34.9 | 699 | 159 | 1,161 | 1,566 | | 618 | 780 | 948 | 1,092 | 345 | 1,446 | 300 | 1,836 | 507 | 1,829 | 907 | 71 | 70 | 1,816 | NC_015143 |
| | | *Madracis mirabilis* | 16,951 | 68.4 | 31.7 | 678 | 224 | 1,140 | 1,587 | | 759 | 780 | 978 | 1,092 | 345 | 1,446 | 300 | 1,815 | 564 | 1,937 | 910 | 71 | 70 | 2,255 | NC_011160 |
| | | *Madrepora oculata*[a] | 15,839 | 69.6 | 30.3 | 681 | 198 | 1,140 | 1,560 | | 792 | 780 | 948 | 1,092 | 345 | 1,446 | 300 | 1,815 | 567 | 1,998 | 1,163 | 71 | 70 | 873 | JX_236041 |
| | | *Montastraea annularis* | 16,138 | 66.4 | 33.5 | 678 | 198 | 1,140 | 1,578 | | 708 | 780 | 948 | 1,287 | 342 | 1,440 | 300 | 1,815 | 561 | 1,973 | 903 | 73 | 69 | 1,345 | NC_007224 |
| | | *Montastraea faveolata* | 16,138 | 66.4 | 33.6 | 678 | 198 | 1,140 | 1,578 | | 708 | 780 | 948 | 1,287 | 342 | 1,440 | 300 | 1,815 | 561 | 1,973 | 903 | 72 | 69 | 1,346 | NC_007226 |
| | | *Montastraea franksi* | 16,137 | 66.4 | 33.6 | 678 | 198 | 1,140 | 1,578 | | 708 | 780 | 948 | 1,287 | 342 | 1,440 | 300 | 1,815 | 561 | 1,973 | 903 | 72 | 69 | 1,345 | NC_007225 |
| | | *Mussa angulosa* | 17,245 | 66.3 | 33.7 | 678 | 198 | 1,140 | 1,575 | | 685 | 780 | 948 | 1,104 | 342 | 1,440 | 300 | 1,815 | 561 | 550 | 695 | 72 | 70 | 4,292 | NC_008163 |
| | | *Pocillopora eydouxi* | 17,422 | 69.8 | 30.1 | 678 | 213 | 1,140 | 1,550 | | 801 | 780 | 978 | 1,308 | 345 | 1,491 | 300 | 1,839 | 564 | 1,917 | 909 | 71 | 70 | 2,468 | NC_009798 |
| | | *Polycyathus chaishanensis* | 15,357 | 70.9 | 29.1 | 678 | 198 | 1,140 | 1,574 | | 708 | 780 | 948 | 1,092 | 342 | 1,440 | 300 | 1,812 | 561 | 1,893 | 905 | 72 | 70 | 844 | NC_015642 |
| | | *Seriatopora caliendrum* | 17,010 | 69.7 | 30.3 | 678 | 237 | 1,140 | 1,548 | | 759 | 780 | 978 | 1,092 | 345 | 1,446 | 300 | 1,839 | 564 | 1,902 | 916 | 71 | 70 | 2,345 | NC_010245 |
| | | *S. hystrix* | 17,059 | 69.9 | 30.2 | 678 | 237 | 1,140 | 1,548 | | 759 | 780 | 978 | 1,092 | 345 | 1,446 | 300 | 1,839 | 564 | 1,904 | 916 | 71 | 70 | 2,392 | NC_010244 |
| | | *Stylophora pistilata* | 17,177 | 70.2 | 29.9 | 678 | 249 | 1,140 | 1,548 | | 837 | 780 | 978 | 1,092 | 345 | 1,446 | 300 | 1,839 | 564 | 1,936 | 914 | 71 | 70 | 2,390 | NC_011162 |

(continued)

**Table 1** Continued

| Order | Scleractinian Clades | Species | Total Length (bp) | Nucleotide (%) | | Gene Size (bp) | | | | | | | | | | | | | | | | | Species Collection Site and GenBank No. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | A + T | C + G | atp6 | atp8 | cob | COI | COI intron | COII | COIII | nd1 | nd2 | nd3 | nd4 | nd4l | nd5 | nd6 | rnl | rns | tmM | trnW | IGS (length) | |
| Other Anthozoa | | *Chrysopathes formosa* | 18,398 | 60.5 | 39.6 | 714 | 213 | 1,143 | 1,593 | | 750 | 750 | 984 | 1,146 | 357 | 1,476 | 300 | 1,851 | 633 | 2,588 | 1,168 | 71 | 70 | 2,591 | NC_008411 |
| | | *Savalia savaglia* | 20,764 | 51.7 | 48.3 | 699 | 219 | 1,161 | 1,521 | 1,239 | 753 | 789 | 990 | 1,158 | 357 | 1,515 | 300 | 1,848 | 666 | 2,644 | 1,197 | 71 | | 3,637 | NC_008827 |
| | | *Nematostella* sp.[a] | 16,389 | 60.9 | 39.1 | 699 | 231 | 1,179 | 1,587 | | 744 | 789 | 984 | 1,110 | 357 | 1,476 | 300 | 1,816 | 600 | 602 | 693 | 71 | 70 | 3,081 | NC_008164 |
| | | *Metridium senile*[a] | 17,443 | 61.8 | 38.1 | 690 | 219 | 1,182 | 1,593 | 853 | 747 | 789 | 1,005 | 1,158 | 357 | 1,476 | 300 | 1,803 | 609 | 2,189 | 1,082 | 71 | 70 | 2,103 | NC_000933 |
| | | *Briareum asbestinum* | 18,632 | 62.9 | 37.1 | 708 | 218 | 1,143 | 1,582 | | 762 | 786 | 972 | 1,164 | 354 | 1,449 | 294 | 1,818 | 558 | 2,224 | 581 | 71 | | 882 | DQ_640649 |
| | | *Pseudopterogorgia bipinnata* | 18,733 | 62.7 | 37.3 | 708 | 219 | 1,144 | 1,597 | | 762 | 786 | 972 | 1,093 | 354 | 1,449 | 294 | 1,818 | 558 | 2,211 | 924 | 71 | | 815 | DQ_640646 |

Note.—Sources of publically available data and collection sites in the case new sequences are also listed. IGS here refers to the total length (bp) of IGSs in each of the mt genomes. Data for the octocoral mtMuts gene are not included. NC_008158 rns gene is 1,239 bp based on the analyses in this study.

[a]Azooxanthellate species.

with the publically available data for hexacorallians (42 species). All the corallimorpharian and scleractinian mt genomes, both those determined in this study and previous work, encode 13 protein-coding genes, 2 tRNA genes (*trnM* and *trnW*; but note that *Seriatopora* spp. have a duplicated *trnW*), the small (*rns*) and large (*rnl*) subunit ribosomal DNA genes, and a *COI* group I intron. Corallimorpharian mt genomes range in size from 20,093 bp in *Rhodactis* sp. to 22,015 bp in *Ricordea yuma* and are significantly larger than those of both Complexa and Robusta corals due not only to the presence of *COI* group I intron (table 1) but also to differences in size of the intergenic spacers (IGSs) between the three lineages (supplementary fig. S1, Supplementary Material online). In fact, the mt genome architectures of the Corallimorpharia are less dense than those of Scleractinia; mt genome size correlates with the total size of the IGS ($r^2 = 0.5371$, $P < 0.001$; supplementary fig. S2, Supplementary Material online). Corallimorpharian mt genomes are characterized by the genes being discrete (i.e., nonoverlapping), whereas this is quite rare in the Scleractinia, where this in shown by only 2 (the complex corals, *Siderastrea* sp. and *Fungiacyathus stephanus*) of the 29 species for which data are available.

The mt genomes of scleractinians are smaller than those of corallimorpharians, but the size (19,429 bp) reported here for that of *G. hawaiiensis* is the largest known for a scleractinian. Two cases of gene overlap were observed in the *G. hawaiiensis* mt genome; *ND4* and *rns* loci overlap by 1 bp, and *ATP8* and *COI* overlap by 18 bp.

## Gene Order and Rearrangements

The organization of the mt genomes of hexacorallian anthozoans is summarized as linear maps in figure 2 and potential rearrangement mechanisms discussed below. As in the Scleractinia, there is a canonical corallimorpharian gene arrangement (CII), but these two patterns are clearly distinct. Ten of 12 corallimorpharian mt genomes exhibited an identical gene arrangement (referred to as Type CII in fig. 3), the exceptions being those of *Corynactis californica* (Type CI) and *C. profundus* (Type CIII). In the Scleractinia, 27 of the 29 complete mt genomes have identical gene order, but again two cases of rearrangement are known (fig. 2). However, although noncanonical gene arrangements have been observed in both Corallimorpharia and Scleractinia, those in the latter involve relatively small changes (i.e., can be explained by single rearrangement events), the rearrangements within Corallimorpharia are much more extensive (fig. 2). At least four rearrangement events are required for the transition between Type CII and Type CI, up to six rearrangement events were identified between Type CII and Type CIII. In the case of scleractinians, far fewer rearrangement events can explain the two deviations from the canonical pattern (Type SII), which *G. hawaiiensis* shares with most of the


**Fig. 2.**—Linear maps showing mt genome architecture in Corallimorpharia, Scleractinia, and other members of the anthozoan subclass Hexacorallia. Names of each Order are indicated in bold. The arrow indicates the direction of transcription. The positions of the 5′- and 3′-ends of the *ND5* intron are indicated by black squares. Corresponding blocks of genes are marked with color; for clarity, lines showing how genes or gene blocks differ in organization between the mt genomes are shown for only the Scleractinia. Note the relatively small number of rearrangements required to account for genome organization between the scleractinians and *Corallimorphus* compared with the large number of rearrangements that appear to have occurred in the corallimorpharians.

Scleractinia. *Madrepora oculata* (Type SIII) differs from the SII pattern only in having the order of the *COII–COIII* genes changed, whereas in *Lophelia pertusa* (Type SI), a block of genes (*COB-ND2-ND6*) has been rearranged (Type SI). The most surprising finding was that, in terms of gene organization, the mt genome of the deep sea corallimorph *C. profundus* (Type CIII) was more similar to the canonical scleractinian organization (Type SII) than it was to other corallimorpharians. Only two rearrangements of blocks of genes are required to explain the SII–CIII transition (fig. 2). Thus, although *Corallimorphus* is unquestionably a corallimorpharian in terms of the sequences of mt genes, the organization of those genes is scleractinian-like, implying that it might represent a key transitional state.

Among metazoans, one unique characteristic of the mt genomes of hexacorallians is the presence of a self-splicing intron within the *ND5* gene that contains a number of complete genes. In the case of the Zoanthidea, Antipatharia, and Actiniaria for which data are available, only two genes, *ND1* and *ND3*, are contained in the *ND5* intron, whereas in the Type CII, all of the genes (including *trnM*, but excluding *trnW*) are contained in the *ND5* intron. In the Type CI pattern, nine protein-encoding genes are located in the *ND5* intron, whereas in Types CIII, SII, and SIII, the same ten protein-coding genes and rns are contained in the *ND5* intron. In Type SI, the number of genes within the *ND5* intron is reduced to 8 due to a rearrangement event between Type SI and these two types of mt genomes in the scleractinians (fig. 2).

## Discussion

The most surprising finding of this study was that the mt genome of the deep-sea corallimorpharian, *C. profundus*, more closely resembles scleractinians in gene organization than it does other corallimorpharians (fig. 3*A* and *B*). Although molecular phylogenetic analyses based on nucleotide or amino acid sequence data for mt proteins yield

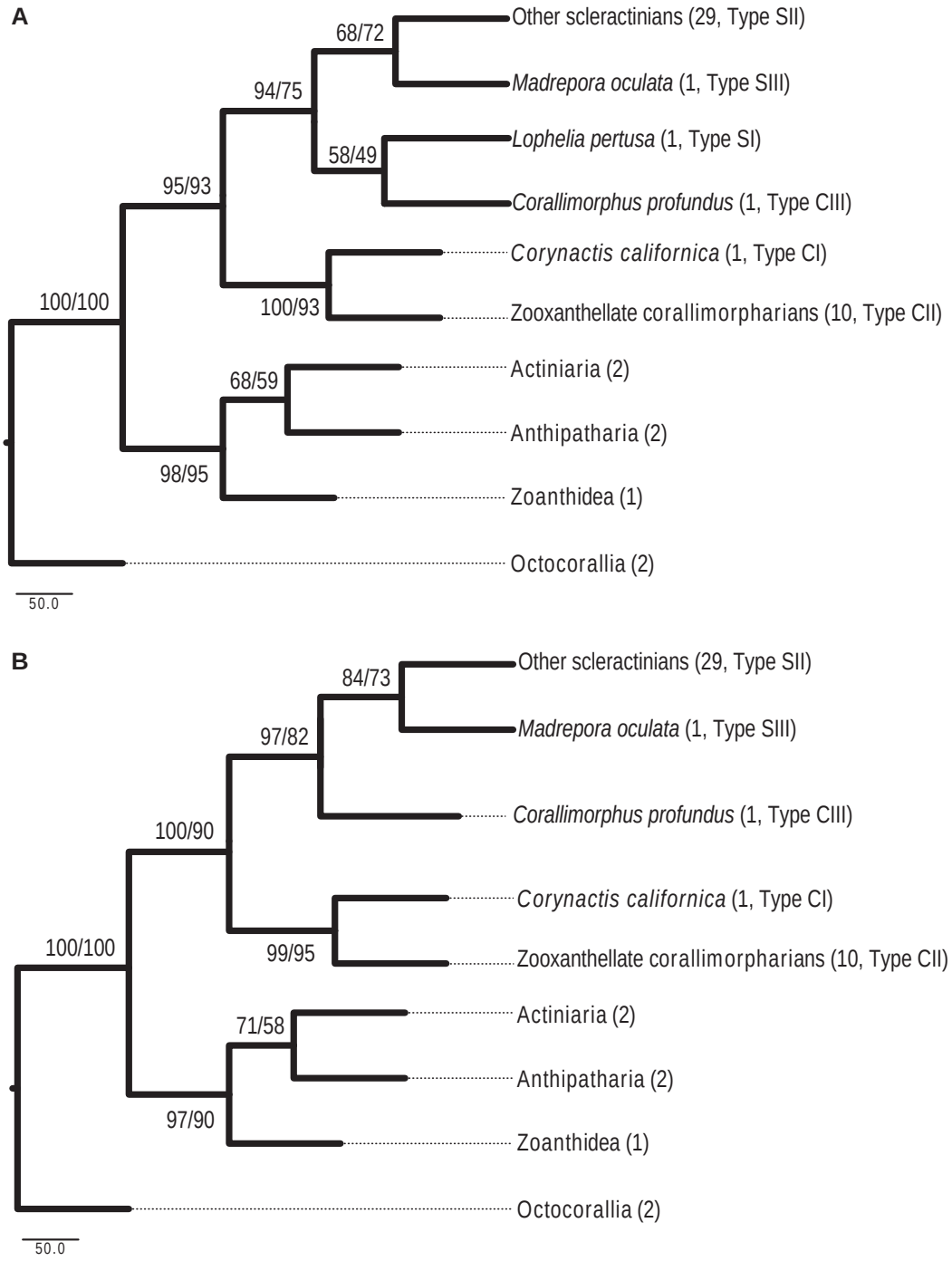



**Fig. 3.**—mt gene order phylogeny of anthozoans. The trees shown are majority-rule cladograms generated using the CONSENSE program in PHYLIP (Felsenstein 1989). The numbers shown at the nodes indicate the percentages of 1,000 jackknife analyses supporting the topology shown in breakpoint and DCJ analyses, respectively. Numbers of species exhibiting the gene arrangement shown are indicated in parentheses. (*A*) Gene order phylogeny with *Lophelia* included. (*B*) Gene order phylogeny with *Lophelia* excluded. Note the weak support for the *Lophelia*/*Corallimorphus* clade in (*A*).

fundamentally different results with respect to the relationship between the "complex" and "robust" scleractinian clades, there is no disagreement concerning the monophyly of the Corallimorpharia nor about the early divergence of *Corallimorphus* within that clade (fig. 1; Kitahara et al. 2014). On morphological grounds, *Corallimorphus* is also considered the most coral like of corallimorpharians (Moseley 1877; den Hartog 1980; Riemann-Zürneck and Iken 2003).

Several authors (den Hartog 1980; Owens 1984; Cairns 1989, 1990; Fautin and Lowenstein 1992) have pointed out the level of similarity between *Corallimorphus* and members of the scleractinian family Micrabaciidae, which are characterized by a reduced skeleton, the fleshy polyp totally investing the rudimentary corallum. Molecular clock estimates imply that the micrabaciids and gardineriids diverged from the scleractinian lineage in the mid-Paleozoic, well prior to the Robusta/Complexa split (Stolarski et al. 2011). The similarity between the earliest diverging members of both the Scleractinia and Corallimorpharia in terms of both morphology and mt genome architecture (fig. 2) implies that *Corallimorphus* occupies a key position in the corallimorpharian <-> scleractinian transition. *Corallimorphus* therefore diverged either close to the point of the scleractinian/corallimorpharian divergence (under scleractinian monophyly) or at the point of skeleton loss (under the "naked coral" scenario).

If we accept that the organization of the mt genome in *Corallimorphus* most closely reflects the ancestral pattern (figs. 1 and 4), then extensive reorganizations are required to generate the consensus corallimorpharian architecture (CII in fig. 2) and that seen in *Corynactis*; in contrast, the rearrangements documented to date within Scleractinia require far fewer steps. In the case of *Lophelia*, the presence of a 67 bp direct repeat comprising the 3′-end of the *ND1* and 5′-end of *COB* genes (Emblem et al. 2011) implies that the likely mechanism of reorganization was tandem duplication and random loss (Moritz et al. 1987; Zhang 2003), which may also account for the *COII–COIII* inversion seen in *Madrepora* (Lin et al. 2012). We were unable to identify signatures of duplication-mediated rearrangement in corallimorpharians; however, neither are there obvious examples of inversion of segments of the mt genome in this Order. Rather, extensive segmental reorganization without inversion has occurred within Corallimorpharia, possibly facilitated by the less compact nature of the mt genomes (reviewed in Boore and Brown 1998). This contrasts markedly with the situation in octocorals, where many successive inversion events explain the observed diversity of mt gene organization (Brockman and McFadden 2012).

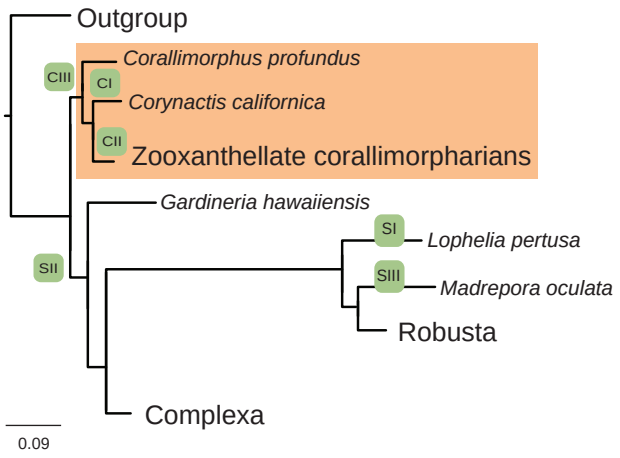



**Fig. 4.**—Hypothetical scheme for the evolution of mt genome architecture in the Scleractinia and Corallimorpharia. The scheme is based on the phylogenetic tree shown as figure 5 in Kitahara et al. (2014), with patterns of gene organization (numbered as in fig. 2) indicated in green boxes.

Can comparisons of mt genome organization resolve the question of coral monophyly? Although the data presented here are consistent with monophyly of the Scleractinia, they do not exclude the possibility of an origin for corallimorpharians within the coral clade. Phylogenetic analyses based on gene order (fig. 3*A* and *B*) were ambiguous. Although both AA- and nt-based molecular phylogenetic analyses unambiguously support monophyly of the Corallimorpharia, the gene order analysis (fig. 3*A* and *B*) did not. We interpret the grouping of *Lophelia* and *Corallimorphus* in this analysis as an artifact resulting from superficial similarities in gene organization in these two organisms; although gene order is similar, the sequences of those genes are highly divergent. The idea that the grouping of *L. pertusa* with *C. profundus* is artifactual is supported by the relatively low DCJ and BPD confidence values (58/49) associated with this node (i.e., well below the 85% confidence interval recommended by Shi et al. 2010). When *L. pertusa* was removed from the analysis, the overall DCJ and BPD statistic performances at the nodes of Corallimorpharia and Scleractinia increased, particularly for the node of *C. profundus* and Scleractinia/*M. oculata*, where support increased from 94/75 to 97/82 (fig. 3).

The mt genomes of the Robusta differ from both corallimorpharians and all other corals in several characteristics. First, within the larger Scleractinia/Corallimorpharia clade, the Robusta have the most compact mt genomes (size range 14,853–17,422 bp) as a consequence of having in general shorter intergenic regions and the largest number of overlapping gene pairs (three to six cases of overlaps). In contrast, corallimorpharians have the largest mt genomes (size range 20,092–22,015 bp), longer intergenic regions, and no cases of overlapping genes, with complex corals intermediate in these characteristics (genome sizes 17,887–19,387 bp; 0–2 overlapping gene pairs—most frequently a single case of overlapping genes). Second, the Robusta differ in structural comparisons of the *ND5* group I intron (Emblem et al. 2011) as well as in molecular phylogenetics based on this feature. A group I intron interrupts the *ND5* gene of all hexacorallians examined to date; these introns typically come and go during evolution but that in hexacorallians contains a variable number of genes and has become an essential feature. The hexacorallian *ND5* intron has been "captured" in the sense that it is now dependent on host-derived factors for splicing, as indicated by the substitution of the ωG (the last nucleotide of the intron) by ωA (reviewed in Nielsen and Johansen 2009; Emblem et al. 2011). Although these characteristics are common across the

coral-corallimorpharian clade, the *ND5* introns of robust corals have a more compact core and overlapping intron and *ND5*-coding sequences (Emblem et al. 2011). In some robust corals, ωA is replaced by ωC, indicating a higher level of dependency on host factors for processing and thus greater integration of intron and host. These qualitative factors, as well as molecular phylogenetics of the *ND5* intron sequences, are most parsimoniously accommodated by scleractinian monophyly (Emblem et al. 2011). Third, of the three lineages, the mt genomes of Robusta have the highest (A+T) content and most constrained codon usage, one obvious consequence of which is that phenylalanine is overrepresented in the proteins that they encode, suggesting that mt DNA repair may be reduced in the Robusta (Kitahara et al. 2014).

The features outlined above, in which the Robusta differ from complex corals and corallimorphs, are derived characteristics—they serve to resolve the robust corals but do not unambiguously identify the sister group. Scleractinian monophyly explains all of the data most parsimoniously, but the alternative cannot yet be ruled out. The mt genome has been exhaustively mined for answers, but these must likely wait for the availability of appropriate nuclear markers.

## Supplementary Material

Supplementary table S1 and figures S1 and S2 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

## Acknowledgments

The authors thank members of the Coral Reef Evolutionary Ecology and Genetics Laboratory (CREEG), Biodiversity Research Center, Academia Sinica (BRCAS), and New Zealand Institute of Water and Atmospheric Research for assistance with sampling and field logistics. We also thank Philippe Bouchet (MNHN) who generously loaned the specimen of *Gardineria hawaiiensis* from the Paris Museum, and Bertrand Richer de Forges and IRD- Nouméa staff and collaborators for their great effort in collecting and preserving the New Caledonian specimen examined in the present study. The corresponding authors also thank two anonymous reviewers for their comments, which were very helpful in revising the manuscript. M.-F.L. gratefully acknowledges receipt of a James Cook University Postgraduate Research Scholarship (JCUPRS). M.V.K. acknowledges the support of the São Paulo Research Foundation (FAPESP) and São Paulo University Marine Biology Centre (CEBIMar). This study was supported by grants from the National Science Council (NSC) and Academia Sinica (Thematic Grants 2005–2010) to C.A.C., and from the Australian Research Council to D.J.M. Coral samples were collected under appropriate permits. This is the CREEG, RCBAS contribution no. 108.

## Literature Cited

Barbeitos M, Romano S, Lasker HR. 2010. Repeated loss of coloniality and symbiosis in scleractinian corals. Proc Natl Acad Sci U S A. 26: 11877–11882.

Boore JL, Brown WM. 1998. Big tree from little genomes: mitochondrial gene order as a phylogenetic tool. Curr Opin Genet Dev. 8:668–674.

Brockman SA, McFadden CS. 2012. The mitochondrial genome of *Paraminabea aldersladei* (Cnidaria: Anthozoa: Octocorallia) supports intramolecular recombination as the primary mechanism of gene rearrangement in Octocoral mitochondrial genomes. Genome Biol Evol. 4:882–894.

Cairns SD. 1989. A revision of the ahermatypic Scleractinia of the Philippine Islands and adjacent waters, Part 1: Fungiacyathidae, Micrabaciidae, Turbinoliinae, Guyniidae, and Flabellidae. Smith Contrib Zool. 486:136.

Cairns SD. 1990. Antarctic Scleractinia. In: Wägele JW, Sieg J, editors. Synopses of the Antarctic Benthos VI. Koenigstein: Koeltz Scientific Books. p. 1–78.

Chen CA, Wallace CC, Wolstenholme J. 2002. Analysis of the mitochondrial *12S* rDNA gene supports a two-clade hypothesis of the evolutionary history of scleractinian corals. Mol Phylogenet Evol. 23: 137–149.

Chen CA, Yu JK. 2000. Universal primers for amplification of mitochondrial small subunit ribosomal RNA-encoding gene in scleractinian corals. Mar Biotechnol. 2:146–503.

Cheng S, Higuchi R, Stoneking M. 1994. Complete mitochondrial genome amplification. Nat Genet. 7:350–351.

Daly M, Fautin DG, Cappola VA. 2003. Systematics of the Hexacorallia (Cnidaria: Anthozoa). Zool J Linn Soc. 139:419–437.

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

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

Desper R, Gascuel O. 2002. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. J Comput Biol. 9:687–705.

Emblem A, Karlsen BO, Evertsen J, Johansen SD. 2011. Mitogenome rearrangement in the cold-water scleractinian coral *Lophelia pertusa* (Cnidaria, Anthozoa) involves a long-term evolving group I intron. Mol Phylogenet Evol. 61:495–503.

Fautin DG, Lowenstein JM. 1992. Phylogenetic relationships among scleractinians, actinians, and corallimorpharians (Coelenterata: Anthozoa). Proceedings of the 7th International Coral Reef Symposium. UOG Station (Guam): University of Guam Press. Vol. 2. p. 665–670.

Fautin DG, Zelenchuk T, Raveendran D. 2007. Genera of orders Actiniaria and Corallimorpharia (Cnidarian, Anthozoa, Hexacorallia), and their type species. Zootaxa 1668:183–244.

Felsenstein J. 1989. PHYLIP—phylogeny inference package. Cladistics 5: 164–166.

Fine M, Tchernov D. 2007. Scleractinian coral species survive and recover from decalcification. Science 315:1811.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol. 3:294–299.

Fukami H, et al. 2008. Mitochondrial and nuclear genes suggest that stony corals are monophyletic but most families of stony corals are not (Order Scleractinia, Class Anthozoa, Phylum Cnidaria). PLoS One 3: e3222.

Gai Y, Song D, Sun H, Yang Q, Zhou K. 2008. The complete mitochondrial genome of *Symphylella* sp. (Myriapoda: Symphyla): extensive gene order rearrangement and evidence in favor of Progoneata. Mol Phylogenet Evol. 49:574–585.

Gish W, States DJ. 1993. Identification of protein coding regions by database similarity search. Nat Genet. 3:266–272.
Kayal E, Roure B, Philippe H, Collins AG, Lavrov DV. 2013. Cnidarian phylogenetic relationships as revealed by mitogenomics. BMC Evol Biol. 13:5.
Kilpert F, Held C, Podsiadlowski L. 2012. Multiple rearrangements in mitochondrial genomes of Isopoda and phylogenetic implications. Mol Phylogenet Evol. 64:106–117.
Kitahara MV, Cairns SD, Miller DJ. 2010. Monophyletic origin of the Caryophyllia (Scleractinia; Caryophylliidae), with description of six new species. Syst Biodivers. 8:91–118.
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.
Kitahara MV, Cairns SD, Stolarski J, Miller DJ. 2012. Deltocyathiidae, an early-diverging family of Robust corals (Anthozoa, Scleractinia). Zool Scr. 42:201–212.
Kitahara MV, et al. 2012. The first modern solitary Agariciidae (Anthozoa, Scleractinia) revealed by molecular and microstructural analysis. Invertebr Syst. 26:303–305.
Kitahara MV, et al. 2014. The "naked coral" hypothesis revisited—evidence for and against scleractinian monophyly. PLoS One 9:e94774.
Le Goff-Vitry MC, Rogers AD, Baglow D. 2004. A deep-sea slant on the molecular phylogeny of the Scleractinia. Mol Phylogenet Evol. 30: 167–177.
Lin MF, Luzon KS, Licuanan WY, Ablan-Lagman CM, Chen AC. 2011. Seventy-four universal primers for characterizing the complete mitochondrial genomes of scleractinian corals (Cnidaria; Anthozoa). Zool Stud. 50:513–524.
Lin MF, et al. 2012. Novel organization of the mitochondrial genome in the deep-sea coral, *Madrepora oculata* (Hexacorallia, Scleractinia, Oculinidae) and its taxonomic implications. Mol Phylogenet Evol. 65: 323–328.
Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25: 955–964.
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.
Moret BME, Tang J, Wang LS, Warnow T. 2002. Steps toward accurate reconstruction of phylogenies from gene-order data. J Comput Syst Sci. 65:508–525.
Moritz C, Dowling TE, Brown WM. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Ann Rev Ecol Syst. 18:269–292.
Moseley HM. 1877. On new forms of Actiniaria dredged in the deep sea; with a description of certain pelagic surface-swimming species. Trans Linn Soc Lond. 1:295–305.
Nielsen H, Johansen SD. 2009. Group I introns: moving in new directions. RNA Biol. 6:375–383.
Owens JM. 1984. Evolutionary trends in the Micrabaciidae: an argument in favor of preadaptation. Geologos 11:87–93.
Pont-Kingdon, et al. 1998. Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial *MutS*: a possible case of gene transfer from the nucleus to the mitochondrion. J Mol Evol. 46:419–431.
Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28:1823–1829.
Riemann-Zürneck K, Iken K. 2003. *Corallimorphus profundus* in shallow Antarctic habitats: bionomics, histology, and systematics (Cnidaria: Hexacorallia). Zool Verh Leiden. 345:367–386.
Romano SL, Cairns SD. 2000. Molecular phylogenetic hypotheses for the evolution of scleractinian corals. Bull Mar Sci. 67:1043–1068.
Romano SL, Palumbi SR. 1996. Evolution of scleractinian corals inferred from molecular systematics. Science 271:640–642.
Romano SL, Palumbi SR. 1997. Molecular evolution of a portion of the mitochondrial *16S* ribosomal gene region in scleractinian corals. J Mol Evol. 45:397–411.
Shearer TL, van Oppen MJH, Romano SL, Worheide G. 2002. Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). Mol Ecol. 11:2475–2487.
Shi J, Zhang Y, Luo H, Tang J. 2010. Using jackknife to assess the quality of gene order phylogenies. BMC Bioinformatics 11:168.
Stanley GD Jr. 2003. The evolution of modern corals and their early history. Earth Sci Rev. 60:195–225.
Stanley GD Jr, Fautin DG. 2001. The origin of modern corals. Science 291: 1913–1914.
Stolarski J, et al. 2011. The ancient evolutionary origins of Scleractinia revealed by azooxanthellate corals. BMC Evol Biol. 11:2–15.
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.
Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.
Yancopoulos S, Attie O, Friedberg R. 2005. Efficient sorting of genomic permutations by translocation, inversion and block interchange. Bioinformatics 21:3340–3346.
Zhang J. 2003. Evolution by gene duplication: an update. Trends Ecol Evol. 18:292–298.
Zhang M, Arndt W, Tang J. 2009. An exact solver for the DCJ median problem. Pac Symp Biocomput. 14:138–149.

**Associate editor:** Richard Cordaux