HOMOLOGY, HOMOPLASY AND THE HOLOBIONT:
A COMPARATIVE GENOMICS APPROACH

A Dissertation in
Biology
by
Bishoy Kamel

© 2016 Bishoy Kamel

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2016
The dissertation of Bishoy Kamel was reviewed and approved* by the following:

Mónica Medina  
Associate Professor of Biology  
Dissertation Advisor

Charles Fisher  
Professor of Biology  
Chair of Committee

Roberto Iglesias-Prieto  
Professor of Biology

Jennifer L. Macalady  
Associate Professor of Geosciences

Kazuhiko Kawasaki  
Research Associate of Anthropology  
Special Member

Tracy Langkilde  
Professor of Biology  
Department Head

*Signatures are on file in the Graduate School.
Abstract

Biomineralization, a process that is found across the Tree of Life, is the formation of minerals by living organisms. Animal calcification is the controlled deposition of calcium carbonate to produce supportive structures in both vertebrates and invertebrate taxa. Biominerals are composed of both inorganic minerals and proteins, which give them extra hardness and special attributes. Biomineralization proteins are also known to be associated with multiple bone disorders and are therefore of biomedical importance. As a first step towards cataloging and curating biomineralization proteins chapter one describes BioMine-DB (http://biomine.net), a biomineralization centric protein database. In chapter two I attempt to increase the phylogenetic representation of the lophotrochozoan biomineralization genetic information. This is done by sequencing transcriptomes from the biomineralizing tissue (mantle) of several molluscan species with publicly available whole genome data: *Lottia gigantea, Crassostrea gigas, Aplysia californica* and *Biomphalaria glabrata* I found that multiple protein families are shared between the four species, *A. californica* albeit having a reduced shell as an adult still retains many of the known biomineralization related proteins.

In chapter three I try to understand the tight intracellular mutualism that occurs between corals and single cell algae of the genus *Symbiodinium*. Corals survive in oligotrophic waters by nutrient recycling and carbon translocation from their algal symbionts. Using whole genome data for metabolic complementarity reconstruction, my analysis revealed *Symbiodinium*’s metabolic capacity is limited and requires support from not only the host, but also the bacterial component of the holobiont. Thus, I show that the coral-algal interaction is not always mutualistic and it is the bacteria that act as mediators to generate a successful coral-algal-bacterial symbiosis. The emergent ensemble of interactions among multiple partners confers robustness to the holobiont. These findings demonstrate the importance of multipartite interactions in the evolution of symbiotic assemblages and the utility of genomics to comprehensively study these multidimensional systems.
Table of Contents

List of Figures vi
List of Tables vii
Acknowledgments viii

Introduction 1

Chapter 1
BioMine: The Biomineralization Protein Database 11
1.1 Abstract ......................................................... 11
1.2 Introduction ...................................................... 11
1.3 Methods ......................................................... 14
   1.3.1 Biomineralization proteins list ................................ 14
   1.3.2 Domain search and protein homolog identification ............ 15
   1.3.3 BLAST Searches ............................................ 15
   1.3.4 BioMine construction ........................................ 15
1.4 Results .......................................................... 16
1.5 Discussion and conclusion ..................................... 17

Chapter 2
Comparative Analysis of Mantle Transcriptomes from Four Mollus- can Species 18
2.1 Abstract .......................................................... 18
2.2 Introduction ...................................................... 19
2.3 Methods .......................................................... 20
2.4 Results ............................................................ 21
2.5 Discussion ........................................................ 22
2.6 Conclusion ......................................................... 23
Chapter 3
Microbes stabilize coral-algal symbiosis

3.1 Abstract ........................................................................................................25
3.2 Introduction ................................................................................................26
3.3 Methods ........................................................................................................27
  3.3.1 Genome sequencing, assembly, and annotation ..................................27
    3.3.1.1 Material and DNA preparation ....................................................27
    3.3.1.2 DNA extraction .............................................................................28
    3.3.1.3 Sequencing and library Preparation ..........................................29
    3.3.1.4 Genome Assembly and scaffolding ..........................................29
    3.3.1.5 Repeat Masking ...........................................................................29
    3.3.1.6 Gene Prediction ...........................................................................29
  3.3.2 Holobiont Transcriptomes ..................................................................30
  3.3.3 Symbiodinium transcriptomes and Genomes ................................30
  3.3.4 Metabolic reconstruction ..................................................................30
  3.3.5 NetCooperate .......................................................................................31
  3.3.6 Network analysis ................................................................................31
  3.3.7 Robustness analysis ...........................................................................32
  3.3.8 Symbionts of Orbicella ......................................................................32

3.4 Results .........................................................................................................32
3.5 Discussion ...................................................................................................34

Conclusion ........................................................................................................39

Bibliography ....................................................................................................44
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The Cambrian Explosion</td>
<td>5</td>
</tr>
<tr>
<td>2.1</td>
<td>Biomineralization proteins Venn Diagram</td>
<td>24</td>
</tr>
<tr>
<td>3.1</td>
<td>BSS score</td>
<td>35</td>
</tr>
<tr>
<td>3.2</td>
<td>Analysis of Robustness</td>
<td>36</td>
</tr>
<tr>
<td>3.3</td>
<td>The metabolic network of the coral holobiont</td>
<td>37</td>
</tr>
<tr>
<td>3.4</td>
<td>Number of N atoms per residue side chain (ARSC)</td>
<td>38</td>
</tr>
<tr>
<td>5</td>
<td>The Pyramid of Complexity</td>
<td>42</td>
</tr>
<tr>
<td>6</td>
<td>Shared biomineralization domain families</td>
<td>43</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Mantle Sequencing Summary ............................................. 21
I am deeply grateful to my advisor Dr. Medina. She has provided me with unsurpassed support and guidance since the day I first met her at ICRS 2008 until the completion of this degree in 2016 I shall be forever indebted to her patience, kindness and nurturing. A PhD is an opportunity to change and be changed by others. I am grateful to a number of colleagues that I have had the luck to work with in the Medina lab over the years some are still around and some already started their own labs. You have influenced me greatly and the work presented here is truly the product of a collective-intellect formed through the many interactions I had daily with everyone in the lab over the years, including the brilliant undergraduate students that we hosted in the lab. On the personal level, your friendship is something I will always cherish, as you all made me feel home, when I was very far away from home. I truly appreciate the support of my thesis committee members both here at Penn State, and prior to arriving here at UC Merced.
Dedication

I dedicate this work to my parents, Amany and Samir and my sister Monica and my brother Michael

“What is important is to spread confusion, not eliminate it”  
—Salvador Dalí
Introduction

“Principles. You can’t say A is made of B or vice versa. All mass is interaction.”

— Richard P. Feynman

Homology, Homoplasy & the Holobiont

Homology is one of the most important concepts in biology (Owen, 1848, De Beer, 1971, Donoghue et al., 1992). In addition to being a central concept in biology, homology has been a subject of extensive theoretical reflection. Theoretical ideas about homology were first developed in the first half of the 19th century in comparative morphology and embryology (Owen, 1848). Once homology became defined to imply shared ancestry, it assumed a central role in phylogenetic systematics, such as 19th century studies in evolutionary morphology. The advent of modern phylogenetic systematics towards the end of the 20th century (Hennig et al., 1999) century improved the understanding of evolutionary relationships and aided in distinguishing between homologous and homoplastic traits (features that are similar but not due to common ancestry (Aguinaldo et al., 1997, Donoghue and Sansom, 2002, Kawasaki et al., 2005, Brigandt, 2006, Marin et al., 2008, Wray and Abouheif, 1998). In parallel, great advances have been made in recent decades in the field of evolutionary developmental biology (evo-devo). As some features of an organism evolve in a connected trajectory, other traits evolve independently in the same organism. These processes produce modular evolutionary units (genetic toolkits) that lead to overall variation (Jenner, 2006). Contrary to homology, homoplasy (Lankester, 1870) is similarity that which is not a result of common ancestry but can be due to multiple causes such as: reversals, independent evolution (conversion or parallelism). Influenced by design constrains and selective pressures homoplasy is rampant in living systems (Wake, 1991).

An important aspect of studying homoplasy is to investigate the underlying mechanism by which a trait evolved more than once. Did this trait employ the same developmental/molecular toolkit or it evolved completely de novo? Molecular data especially from the field of developmental biology have increased our understanding tremendously about how homoplasy can occur, via convergent or parallel evolution.
The distinction between parallel and convergent evolution is controversial. With parallel evolution invoked usually when novelty arises in closely related taxa, versus convergent evolution which occurs between distantly related taxa. In the view of Ardent and Reznick (2008) the distinctness of parallelism and convergence is a false dichotomy, rather they adopt a continuum view with each concept being at both extremes. The same idea of a continuum is particularly important for developmental biology, as certain traits can be turned off and on frequently in closely related taxa. Abouheif chose to call this phase of evolutionary change ‘Mesoevolution’ a term coined to avoid using parallelism and problems associated with it (Abouheif, 2008). However, the term has not found wide acceptance, and still parallel evolution is used frequently in biological literature.

Despite the current focus on mechanisms that generate variation, it is still challenging for researchers to understand the evolutionary trajectory of specific homologous and homoplastic traits. Because biological processes and traits build upon several levels of organization (e.g. tissues, cells, proteins, protein interactions, regulatory networks and genomes), homologies tend to break up at those levels. This multilayer complexity then compels us to ask which level of organization is appropriate to deem a trait homologous within a clade and at different levels of phylogenetic divergence. This has been discussed as ‘Deep homology’, in reference to traits that appear to be homoplastic on a higher biological hierarchy i.e. organs, but are homologous on the molecular or developmental level. This particular situation has been witnessed in many cases that stood as a hallmark of convergent evolution such as appendages in arthropods and vertebrate limbs (Shubin et al., 1997, 2009), vertebrate and invertebrate eyes (Piatigorsky, 2008) that were shown to use the same molecular and developmental mechanisms and share deep homology.

The concept of the Holobiont

The concept of deep homology underscores another important property of biological systems, that, of ‘Hierarchy’. Issues caused by the hierarchical organization of living systems transcend the question of homology and homoplasy to other topics such as the definition of organism, colony and the concept of the holobiont (i.e. a host and its associated microbial community) and the respective hologenome. Recent advancement in the field of microbiology in particular, microbiome research fueled by the surge in DNA
sequencing technologies clearly demonstrated the importance of microbes in evolution of complex organisms. This renders a single organism as a 'set of sets' (De Landa, 2016) a multitude of organisms giving rise to the view of animals as a host-microbe ecosystem (McFall-Ngai et al., 2013). Hence, holobionts are defined assemblages that behave as individual ecological units (Margulis and Fester, 1991). This idea is not too far to the ecological species concept which states that organisms occupying the same niche should be treated as a species. The ecological species concept was first formulated by (Van Valen, 1976) which he was proposed it as “vehicle for conceptual revision and not a standing monolith”. However, while there is still controversy about the holobiont concept, it has gained wide acceptance, in many fields that deal with organisms involved in complex symbiotic relationships such as corals (Rohwer et al., 2002a).

On the other side, the hologenome concept, has stirred quite a debate, and is yet to gain wide acceptance. The main problem of the hologenome concept can be summarized elegantly by this quote from E.O Wilson “At what point does a society become so well integrated that it is no longer a society?” (Wilson, 1974). Proponents of the hologenome theory argue that the collective of the host genome and its microbial genomes encompasses a single entity the 'Hologenome'. The problem then arises, can the hologenome be a unit of selection? This is an ancient question being asked in new terms, known as multilevel selection. At the core of this multilevel selection dilemma is the problem of biological hierarchy, how does selection operate concurrently at more than one level of the biological hierarchy (genes, cells, tissues, microbes, hosts, ecosystems). More importantly, how conflict in one level does not disrupt cooperation in another level (Okasha, 2005). Whether multilevel selection does occur, or the hologenome concept is useful remains a highly debatable and active area of research (Nowak et al., 2010, Herre and Wcislo, 2011, Nowak et al., 2011, Brucker and Bordenstein, 2013, Moran and Sloan, 2015). Nevertheless, the discussion about units of selection in complex biological systems continues, and as progress is made in the areas of symbiosis and increased unification of network theory and biological systems, the field is advancing (Corominas-Murtra et al., 2013, Tëmkin and Eldredge, 2015).

**Biomineralization**

Biomineralization can be defined as a highly controlled and functional process of hard tissue formation (Simkiss and Wilbur, 1989). Biominerals are found across the Tree of
Life, ranging from metal crystals in bacteria to mineral structures in unicellular and multicellular eukaryotes. The evolutionary success of organisms with the ability to biomineralize is manifested by the wide variety and construction of mineralized structures (Lowenstam and Weiner, 1989). Such structures are found in most present-day animal (metazoan) taxa where they help in a wide range of functions: tissue support, UV protection, shelter against predation, nutrition, reproduction, gravity, light or magnetic field perceptions, storage of mineral ions, etc. (Simkiss and Wilbur, 1989, Marin et al., 2008). In particular, calcification (i.e. biologically induced calcium carbonate mineral formation) is one of the most prominent biomineralization processes found in diverse animal lineages. It is a process by which the controlled deposition of calcium carbonate salts such as calcite or aragonite (the major mineral forms in animals) produces supportive structures. Calcification is present in diverse animal taxa, including vertebrate (e.g. bones, teeth) and invertebrate species (e.g. molluscan shells, coral skeletons). The ability to calcify a supporting skeleton was a major evolutionary transition that contributed to the successful animal radiation during the Cambrian. Despite the widespread occurrence of biomineralization, controversy remains whether there is a common evolutionary origin of this trait in animals. One of the major evolutionary radiations in the history of Life on Earth took place during the late Neoproterozoic-Cambrian boundary (545 mya) with the evolution of most of the animal phyla. This event is known as the ‘Cambrian explosion’ and it was the biggest moment in the evolution of calcification, as mineralized skeletons appeared for the first time in many metazoan groups (Cloud, 1968, Lowenstam and Weiner, 1989) Figure 1 this view was challenged by Maloof et al. lately after re-examining of the fossil data using improved radiological techniques and suggested that according to the fossil data it was not a mere explosion but rather occurred gradually in at least three different pulses according to fossil evidence (Maloof et al., 2010). On the other side, according to molecular phylogenies and fossil data, diversification of metazoans occurred approximately 635–542 Mya (Knoll et al., 2004, 2006, Peterson et al., 2008), hence long before the major phyla began biomineralizing. Animal phylogenies have become increasingly more accurate and a consensus of major relationships (i.e. superphyla level) has emerged (Halanych et al., 1995, Aguinaldo et al., 1997, Medina et al., 2001, Dunn et al., 2008, Philippe et al., 2009). In more recent decades, the burgeoning field of evolutionary developmental biology has also benefited from technological breakthroughs in molecular biology and genomics. Comparative biologists have been able to gain a deeper understanding than
ever before of the ‘genetic toolkit’ of the metazoan ancestor thus tackling long-standing questions of morphological and developmental homology (Putnam et al., 2007, King et al., 2008, Srivastava et al., 2008). The Cambrian diversification of mineralized skeletons was likely part of the broader process of metazoan radiation, with increases in predation pressure favoring the evolution of a protective cover (Stanley, 1973, Bengtson, 1994).

Although some calcifying lineages have gone extinct since the Cambrian, multiple extant phyla both preserved and extended the ability to biomineralize, demonstrating the evolutionary success of this innovation. The lack of virtually no record of non-calcifying lineages makes it difficult to infer, from the record alone, the evolutionary trajectory of the biomineralization process in animals. In addition, little is known about the genetics of the ‘biomineralization toolkit’ in most of these lineages. While the timing of emergence of skeletal features in the vertebrates and corals is hundreds of millions of years apart (Zhuravlev and Riding, 2001), this does not necessarily imply...
that the molecular mechanisms that are required to enable any given eukaryotic cell
to biomineralize (e.g. the ability to transport calcium and carbonates and the ability
to inhibit calcification) evolved independently. More likely, such cellular-level mech-
anisms must have existed prior to the exhibited biomineral morphological features in
the metazoan fossil record. While there has been extensive work on the biochemistry
and genetic basis of a few hundred proteins involved in metazoan biomineralization,
the acidic nature of a vast majority of exoskeletal proteins has presented multiple chal-
lenges for fast progress with biochemical studies (Marin and Luquet, 2007). Most of the
research has come from vertebrates (Kawasaki et al., 2007), echinoderms (Wilt and Et-
tensohn, 2007), and mollusks (Sarashina and Endo, 2006), brachiopods (Luo et al., 2015)
with little contribution from the many other metazoan calcifying phyla.

Three scenarios can be hypothesized to understand how precipitation of the meta-
zoan skeletal minerals evolved. As a result of 1) independent, repeated evolutionary
innovations in the process of biomineralization (de novo), 2) unprecedented lateral ge-
netic transfer (LGT), or 3) natural selection through geologic time from an ancestral
biomineralization repertoire up to a point where homology cannot be easily inferred
(Kirschvink and Hagadorn, 2000). In favor of the first scenario, (Knoll, 2003) gave an es-
timate that carbonate skeletons must have evolved at least twenty times in metazoans.

Mineral precipitation must be carried out in a controlled fashion in specific biological
environments requiring directed transport of calcium (Ca$^{2+}$) and carbonate (CO$_3^{2-}$) ions
as well as molecules to guide mineral nucleation and growth (Westbroek and Marin,
1998). Thus, given that skeleton formation requires more than the ability to precipitate
minerals, the scenario in its fullest consequence (multiple, independent de novo evolu-
tion of biomineralization) is unlikely. The same line of argument holds for the second
scenario. If LGT played a significant role in the evolution of biomineralization, then it
must have happened multiple times or multiple genes must have been transferred at
once as skeleton formation is not only about mineral nucleation by animal tissue but
also about the regulation of it. Although common among bacteria, the movement of
genes between distantly related organisms is thought to occur only rarely among multi-
cellular eukaryotes (Daniels et al., 1990). In favor of the third scenario, we can conclude
that all cells share the ability to bind and regulate Ca$^{2+}$ concentrations (MacLennan
et al., 1997, Sanders et al., 1999), and both photoautotrophic and heterotrophic eukary-
otes regulate their internal inorganic carbon chemistry using carbonic anhydrase and
other enzymes (Aizawa and Miyachi, 1986). Thus, the biochemical supply of ions re-
quired for calcite/aragonite precipitation appears to be an ancient feature of eukaryotes, indicating that lineages with no calcifying common ancestors share the underlying capacity to synthesize and localize mineral nucleation inducing biomolecules. Because spontaneous calcification of cell and tissue surface may have been a problem in highly oversaturated Proterozoic oceans, multiple authors (Grotzinger, 1989, Knoll et al., 1993, Marin et al., 1996) reasoned that molecular inhibitors evolved early as anti-calcification defenses and were later recruited for the physiological control of skeleton growth. This prediction is known as “the anti-calcifying macromolecules that locally inhibit crystal growth hypothesis”.

Indeed, biochemical similarities of anti-calcifying molecules in mollusks and cnidarians support the hypothesis that this biochemistry already existed in the last common ancestor of cnidarians and bilaterian animals, if not before that (Marin et al., 1996, 2000). Similarly, proteins related to biomineralization are highly conserved across a variety of vertebrate species, and some invertebrate taxa have protein domains analogous to those present in the vertebrate mineralized tissues (Weiss et al., 2001). Moreover, immunological similarity has been shown between the macromolecules involved in hydroxyapatite formation in vertebrates and those connected to the aragonite present in molluscan nacre (Atlan et al., 1997). It is also known that freshly ground mollusk nacre fails to elicit an immune response in humans, and in fact stimulates bone regeneration (Lopez et al., 1992). Such a result is highly unlikely if both biominerals had evolved through separate pathways, and it argues for common ancestry or deep homology. Another example supporting evidence of an ancestral biomineralization toolkit is the alpha type carbonic anhydrase family, which has been implicated in CaCO$_3$ deposition from sponges to vertebrates (Jackson et al., 2007a). Most animals have multiple carbonic anhydrases, and distinct subfamilies are recognized (Hewett-Emmett and Tashian, 1996, Tambutte et al., 2007) each of which are widely distributed phylogenetically. In addition, some calcifying animals have atypical carbonic anhydrases that may represent lineage specific adaptations to facilitate CaCO$_3$ deposition (Miyamoto et al., 1996, Moya et al., 2008). Although the process of biomineralization is distinct in different species and tissue types, there are some clear homologies across taxa. In the view of Westbroek and Marin 1998, the many origins of calcareous skeletons reflect multiple events of independent co-option of molecular and physiological processes that are widely shared among eukaryotes. The idea of deploying, duplicating, and linking in different ways, readily available “off-the-shelf” biochemistry, for a new role, structure
of functions (a phenomenon coined exaptation by (Gould and Vrba, 1982) can be invoked as a likely mechanism for the evolution of eukaryotic and, in particular, animal biomineralization. The overall pattern, then, would be one of early eukaryotic clade differentiation, with later, and in some cases much later, evolution of biomineralized skeletons within clades. This is supported by the fact that evolution of mineralized skeletons in Cambrian organisms was not instantaneous. Only a few taxa occur in lowermost Cambrian rocks. Skeletal abundance and diversity increased successively over the succeeding 25 million years. Skeletons evolved as principal components of body plans, shaping the subsequent course of evolution in these clades. The emerging scenario would then be that skeletons that are not homologous as structures, seem to share underlying physiological pathways that are. This powerful idea implies that there is at least a minimal common ancestral molecular toolkit that animals deployed in order to allow skeletal evolution with subsequent specialization within lineages.

**The *Orbicella faveolata* genome**

Corals harbor dinoflagellate endosymbionts from the photosynthetic genus *Symbiodinium*. *Symbiodinium* derives its carbon from the seawater bicarbonate and metabolically regenerated carbon. The carbon fixed by *Symbiodinium* is used to generate energy, or is translocated to the host. Some of the carbon acquired by *Symbiodinium* ends up in the coral skeleton (Pearse, 1970). Corals reside in oligotrophic (i.e. nutrient poor) environments. Survival is thus only possible via tight nutrient recycling and photosymbiosis (carbon fixation), which enables corals to fill their highly productive and important ecological role. Coral reefs known as Rainforests of the Sea harbor the highest tropical marine biodiversity. In addition to hosting the main driver in coral symbiosis (*Symbiodinium*) the coral holobiont is home to a plethora of organisms encompassing viruses, archaea, bacteria, fungi and protists (Rohwer et al., 2002a). The complex microbial community living inside the coral takes active roles in welfare as well as warfare, and the conflicting interests of the host against symbionts and other organisms external to the holobiont (Reshef et al., 2006). Therefore, these interactions set the stage to a dynamic landscape within the holobiont comprising gradient that ranges from altruism, through mutualism and parasitism.

Recent advancements in genome sequencing has illuminated the nature of symbiotic relations between multiple organisms and gave rise to the hologenome theory of
evolution (McFall-Ngai et al., 2013, Moran and Sloan, 2015). Corals can host a variety of *Symbiodinium* species and also switch *Symbiodinium* species depending on environmental conditions (e.g. light and temperature) (Ritchie, 2012) especially during bleaching conditions according to the adaptive bleaching hypothesis (Kinzie et al., 2001). This phenomenon has been well documented in the coral *Orcella favoeolata* (Rowan et al., 1997). Although Symbiodinium cultures isolated from coral tissue have been around for several decades (Trench and Blank, 1987), most symbionts are extremely difficult to obtain in pure culture. Multiple attempts using physical and antibiotic treatments have not produced conclusive results with the presence of contaminating bacteria (Andersen and Kawachi, 2005).

The first coral genome to be sequenced was that of *Acropora digitifera* (Shinzato et al., 2011), and more recently the genome of the symbiotic anemone *Aiptasia pallida* was also elucidated (Baumgarten et al., 2015). Symbiodinium genomes on the other hand are much larger than that of the coral and contain a large amount of repeat elements making them harder to sequence (LaJeunesse et al., 2005a). Yet, a partial genome was obtained for *Symbiodinium minutum*, and recently *Symbiodinium kawaguti* (Shoguchi et al., 2013, Lin et al., 2015). While the coral hosts a wide array of other symbionts (Cumbo et al., 2013), bacterial symbionts play an important role in the life history, disease and survival (Thompson et al., 2015). The nature of interactions among all the partners of the holobiont, is still understudied. The nature of interactions among all the partners of the holobiont, is still understudied. While we have an idea about nutrient recycling in the holobiont, the complexity of coral morphology, and inability to replicate its natural environment in the laboratory makes it hard to conduct experiments on the *in situ* biology of the symbionts.

With the accumulation of genomic data from the multiple members of the holobiont mentioned earlier, one is now able to understand the environment where these members live through the methods of the new field of ‘Reverse Ecology’. Reverse ecology is a term describing various approaches that utilize genomic information to directly understand ecological context (Levy and Borenstein, 2012). When these methods are combined with systems biology approaches one is able to infer global patterns of evolution such as detecting the lifestyle of microbes as obligate parasites or free-living based on their metabolic capacities (Borenstein et al., 2008). By representing the metabolic capacity of the different players in the holobiont as a network (undirected graph) we can then investigate the general properties of these networks, and the network of networks.
(the holobiont) and its behavior. This allows us to make predictions about the lifestyle of the host versus the symbionts and gain better understanding of the metabolic recycling inside the coral *Orbicella faveolata* and other corals.

To highlight the importance of accounting for biological hierarchy when investigating biological processes, I chose to investigate the evolutionary history of the process of biomineralization. Chapter 1 aims to gather and catalog most of the proteins involved in this process across metazoans to enable comparing across taxa in a dynamic searchable database. Chapter 2 deals with molluscan biomineralization comparing 4 different species of mollusks highlighting the retention of biomineralization proteins even after shell-loss in the opisthobranch mollusc *Aplysia californica*. Chapter 3 investigates the complexity of metabolic interactions in the holobiont amongst the ensemble of symbionts in the Caribbean coral *Orbicella faveolata* using network theory and reverse ecology methods.
Chapter 1  |  BioMine: The Biomineralization Protein Database

Information: the negative reciprocal value of probability.

—Claude Shannon

1.1 Abstract

Biomineralization is the process by which living organisms construct hard skeletons creating complex structures that range from specialized tissues such as bone or teeth to ecosystems such as coral reefs. Biominerals are composed of both inorganic minerals and proteins, which give them extra hardness and special attributes. Biomineralization proteins are also known to be associated with multiple bone disorders and are therefore of biomedical importance. Herein we describe BioMine, a biomineralization centric protein database. Availability and implementation: BioMine can be accessed at http://biomine.net, SQL dump, FASTA files and source code are available for download as well.

1.2 Introduction

Biomineralization is a process in which minerals form inside or outside the cells of a variety of organisms (Lowenstam and Weiner, 1989, Simkiss and Wilbur, 1989). In animals, these minerals are primarily calcium carbonates and calcium phosphates (Knoll,
The majority of biominerals formed in bones, shells, skeletons and spicules are composed of mineral crystals, however all biominerals contain various amounts of other proteins that give these minerals extraordinary properties. The cell orchestrates the mineral formation process through the expression and translocation of proteins that nucleate the crystals either intracellularly or extracellularly. More importantly, the cell has to inhibit mineral formation and crystal growth in unwanted sites (Marin et al., 1996, Kawasaki et al., 2009). Both nucleation and inhibition can be achieved through multiple cellular mechanisms. For example, the cell will produce enzymes that modify proteins by breaking them into smaller peptides (Qin et al., 2004) thus changing their function. The cell is able to tightly regulate the biomineralization process by molecular modification (e.g. adding sugars or other moieties) and regulation of ion transport across membrane (Saavedra, 1994, Sarashina and Endo, 2006). Such modifications to biomineralization-associated proteins determine how they interact with other proteins, other cells, and with the biomineral in general. Biominerals are essential to the survival of a broad range of animal taxa because they deliver protection against predation, act as energy storage, provide support and unique optical properties (Addadi et al., 2006). In particular, biomineralization plays a pivotal role in multiple human diseases and other pathological phenomena such as coronary artery calcification (Lopez et al., 1992, Atlan et al., 1997, Westbroek and Marin, 1998, Collette et al., 2010, Salih et al., 1996, Fisher et al., 2001, Wallin et al., 2001). A growing interest in bio-inspired materials has generated a large body of work that uses proteins and other biological scaffolds for in vitro mineralization and synthetic materials (Perry et al., 2009, Chiu et al., 2012). The process of biomineralization is ubiquitous throughout the animal tree. Such distribution has generated speculation about the origin of metazoan biomineralization and its evolutionary history. Biomineralization is a complex process that relies on multiple cellular pathways (Marin et al., 1996, Knoll, 2003). Many of the studied biomineralization proteins are part of other important processes such as cell adhesion, extracellular matrix organization and immune functions (Bryden et al., 1999, Clendennon et al., 2009). This evidence favors the idea that biomineralization independently evolved in multiple phyla using pre-existing pathways in the early eumetazoan ancestor. It can also be argued that biomineralization was present in the early eumetazoan ancestor yet various parts of the pathway were lost in several animal lineages. According to fossil evidence and when mapped onto a phylogeny, carbonate skeletons seem to have evolved at least 20 different times in metazoans (Knoll, 2003). If biomineral-
ization evolved multiple times, it is relevant to understand which components of the process exactly underwent innovations. Since biomineralization is an active process, it requires 1) targeted localization of calcium and carbonate, 2) an organic matrix as a template for the mineral nucleation, 3) growth, and 3) efficient inhibitors in order to stop undesired calcification or even formation of the mineral (Marin et al., 1996, Jackson et al., 2010). When all these different requirements are taken into account, it seems unlikely that such diverse biochemical processes involved in metazoan biomineralization evolved independently more than 20 times. A process such as transport is quite conserved across animal lineages and it shows a clear history of gene duplication events (Dean et al., 2001, Saier et al., 2009). Such complexity presents us with a conundrum. While the biomineralization process, with different minerals and methods of calcification and clear evolutionary novelties, is found across multiple animal phyla (Marin et al., 1996, 2000, Marin and Luquet, 2004, Jackson et al., 2006, 2007a,b), it also remains that many parts of biomineralization pathways must be conserved.

As a first step in tackling such questions in the evolution of animal biomineralization, we have created a database to accumulate, annotate and curate biomineralization proteins and protein-coding sequences. The database aims to serve the community by bridging the gap between the few identified biomineralization proteins, and the unannotated plethora of Expressed Sequence Tags (ESTs), draft genome gene models and transcriptomes generated using next-gen sequencing technologies. We employed various bioinformatics techniques using domain-based searches to collect and identify novel biomineralization proteins in metazoans. We hope that due to the increasing surge of sequence information along with broad phylogenetic representation in the public domain, a clearer picture of the evolutionary history of biomineralization proteins will emerge, rendering BioMine as a dynamic platform to answer not only fundamental questions in animal evolution but also about the process of biomineralization in particular lineages.

Currently there are various approaches used to provide solutions for biological data integration such as navigational, warehousing, and mediator-based methods, each of which has pros and cons. Navigational: this is the most widely used data integration tool in life sciences, and it relies on flat file storage. Data are extracted from public data sources and assembled into one or more flat files. These flat files are indexed or cross-referenced according to table or specific field locations. This kind of solution has been implemented in public resources, including SRS (Etzold et al., 1996) BioNavigation
Data warehousing, also known as “data translation”, is accomplished by the initial collection of data from multiple resources into a local warehouse and subsequent execution of all queries on the warehoused data in a relational manner. Data warehousing requires the use of extraction and transformation tools to both load and map these data to a global schema. The price of creating the extraction and transformation tools to import data into a common schema is compensated through efficient and complex query execution, query optimization, more control over the data, and data consistency and redundancy. The major downside of warehousing is scalability because frequently updated databases need to be always in synchrony with the global schema warehoused. Thus the process of updating and re-indexing the databases can be challenging.

This method works by leaving the data in their original schema at their production environments. During execution, an application translates the user queries, optimizes, and maps them to remote or local databases. The most time consuming part in this approach is constructing the tools to communicate with the remote data resources, and to map and optimize queries across them. This solution has particular drawbacks, including 1) lack of control over the data, 2) absence of permission to directly query any given remote resource, and 3) sudden loss of service due to physical or logical change in the provider site. Implementations in this category include Tambis (Baker et al., 1998) and DiscoveryLink (Haas et al., 2001). In principle, BioMine falls under the category of warehousing. We choose this method because it enables us to have more control over the data since one of the strengths of BioMine is the human-curated-annotations of biomineralization proteins.

1.3 Methods

1.3.1 Biomineralization proteins list

We carried out a wide primary literature and database survey in order to compile a list of proteins that are functionally implicated biomineralization in animals. Specifically, we included data from scleractinian corals, calcareous sponges, gastropod and
bivalve mollusks, crustaceans, echinoderms, and vertebrates. Additional sequences were collected from the AMIGO Gene Ontology database (Carbon et al., 2009). The complete biomineralization gene list is accessible through the BioMine web application (http://biomine.net/). These already annotated sequences were used as a seed to search for related biomineralization proteins in undocumented taxa or new sequence databases, and were stored in a pre-computed BLAST search database. After building the list of candidate proteins, each protein could be traced back to an original publication where it was described.

1.3.2 Domain search and protein homolog identification

To further improve the search strategy, using the Pfam database, we scanned for conserved protein domains in the proteins we gathered from primary literature (Finn et al., 2008). The identified domains in the already known biomineralization proteins were scanned against 6-frame translations of ESTs and protein sequences from dbEST and nr databases (NCBI) of the taxa Cnidaria, Mollusca, Brachiopoda, Arthropoda, Echinodermata and Vertebrata using the HMMER 3 package (Eddy et al., 2009). The tool used for the translation was sixpack from the EMBOSS package (Rice et al., 2000).

1.3.3 BLAST Searches

In addition to the domain searches, we conducted BLASTp searches for all the proteins in the seed database against nr and against the 6-frame translations of the dbEST for the selected taxa. For filtering the results, we only considered hits that match e-value < 0.000001 and bitscore > 150 to be significant.

1.3.4 BioMine construction

In order, to organize all the data in a searchable platform, we constructed a web application that enables us to search the results and to submit new sequences into the database. BioMine is written in PHP and Perl, and relies on MySQL for relational information. Source code for BioMine is under GPL v3 at https://github.com/bishoyh/biomineDB. The MySQL database contains the results of all the HMMER results in addition to the BLAST results and FASTA files of all sequences can be downloaded from the website.
1.4 Results

After assembling the initial list of biomineralization-related proteins, we identified putative homologs of given candidate genes in calcifying lineages (mollusks, brachiopods, cnidarians, arthropods and echinoderms). A protein domain search was initiated on our candidate list of 472 proteins, based on Pfam models using HMMER 3 (Eddy et al., 2009). In the Pfam search 198 domain families were found to be linked to biomineralization. The search results were stored in a relational database linking the detected domains with the species to detected orthologs. Below we describe two potential scenarios for the use of BioMine by the scientific community.

Use case 1:

1. A user prepares a list of proteins from a newly sequenced organism.

2. The user submits the protein list to BioMine through the web interface.

3. BioMine generates potential matching biomineralization proteins in the submitted dataset, together with the publications in which these similar proteins have been described. In addition, a predicted protein-protein interaction network will be generated if the submitted protein list can be decomposed successfully to Pfam domains. These domains are obtained from the Pfam database, which contains curated conserved domains of various functions.

Use case 2:

1) A user is already working with a known biomineralization protein and is doing functional work, i.e. the user is doing whole mount in situ gene expression research in a given organism and finds it hard to explain the observed expression pattern. Thus, the user thinks there could be other proteins involved.

2) The user submits his protein (as sequence or as a gi identifier or uniprot id) to BioMine.

3) The user gets back a list of potential interacting proteins and possible paralogs restricted to a taxon of his choice if needed.
1.5 Discussion and conclusion

Discussion and Conclusion By combining thorough literature scrutiny with similarity searches we were able to construct a large dataset of biomineralization-related proteins. BioMine proved useful in annotating sequence data from non-model organisms involved in the particular process of biomineralization. The ability to always link back to the primary literature provides a unique opportunity to the investigator to directly examine the experimental evidence that deemed a particular protein as biomineralization-associated. We believe this should fast-forward research in non-model systems by knowledge transfer from model species in biomineralization research. By providing a BLAST interface and downloadable versions, we are certain that biomineralization researchers can benefit from BioMine. BioMine is a dedicated database containing manually curated biomineralization proteins from all mineralizing animal taxa that is presented through a user-friendly fully searchable interface.
Chapter 2  
Comparative Analysis of Mantle Transcriptomes from Four Molluscan Species

“Absence of evidence is not evidence of absence.”  
—Carl Sagan

2.1 Abstract

Biomineralization, a process that is found across the Tree of Life, is the formation of minerals by living organisms. Animal calcification in particular is the controlled deposition of calcium carbonate to produce supportive structures in both vertebrates and invertebrate taxa. The ability to calcify was an evolutionary innovation that is thought to be greatly responsible for what has been coined as the "Cambrian explosion", a major adaptive radiation event that took place at the late Neoproterozoic-Cambrian boundary. Genomic approaches are opening a new window of opportunity to engage in questions regarding homology and evolution of this important biological innovation. In order to increase the phylogenetic representation of the lophotrochozoan biomineralization genetic information we undertook the sequencing of transcriptomes from the biomineralizing tissue (mantle) of several molluscan species with publicly available whole genome data: *Lottia gigantea, Crassostrea gigas, Aplysia californica* and *Biomphalaria glabrata*. The transcriptomes were compared against each other after filtering through
the BioMine pipeline to determine which biomineralization protein families are shared between these taxa. We found that multiple protein families are shared between the four species, in particular *A. californica*.

### 2.2 Introduction

Molluscs are among the most successful metazoan groups. Harnessing the utility of their calcified shells since the Cambrian (Lowenstam and Weiner, 1989, Weiner and Dove, 2003). Genomic and post-genomic approaches are opening a new window of opportunity to engage in understanding molluscan shell structure, as they provide a means to analyze the overall composition of the animal biomineralization secretome (at the mRNA level) by high throughput sequencing (Wilt et al., 2008), and the proteomic level (Marie et al., 2010, 2011, Mann et al., 2012, Zhang et al., 2012). As new genomic data emerge, we are able to address evolutionary questions regarding the homology and origin of this molecular machinery involved in this important Cambrian innovation (Kocot et al., 2016).

Interestingly, many lineages of mollusks lost their shell either completely or partially (Morton, 1960, Kröger et al., 2011) By losing their shells, the potential to develop into multiple forms was unlocked and this enabled mollusks to diversify and occupy new niches. By losing the protective power of the shell, some lineages resorted to nematocyst sequestration and multiple other defense mechanisms (Rudman, 1981, Rudman and Avern, 1989, Vermeij, 2013). Molluscan shells are built from calcium carbonate minerals and a complex organic matrix consisting mainly of proteins and polysaccharides. The organic matrix is mainly responsible for the crystal nucleation and controlling the crystal and mineral form of the shell. However other intracellular proteins are also important in regulation of shell formation. Therefore, it is important to understand both organic matrix proteins, in addition to other intracellular proteins involved in the process of biomineralization.

By using the combined strategy of computational analysis and increasing the current phylogenetic coverage we can gain better insights in the evolutionary questions regarding biomineralization. Since most of the shell making occurs by the mantle tissue in mollusks specifically the mantle edge epithelia (Marin and Luquet, 2004) we chose four mollusks (*Biomphalaria glabrata, Aplysia californica, Lottia gigantea* and *Crasostrea gigas*) to increase our knowledge repertoire of biomineralization transcripts.
We mainly chose species that have already genomes sequences, and where there was at least some understanding of the shell formation and mineralization (Marxen et al., 2003, Marie et al., 2013b). Aplysia in particular was of a great interest because of the reduced shell that is retained in the adult and its relevance to shell loss question in opisthobranches (Medina et al., 2011, Vue et al., 2014) Biomphalaria glabrata is well known as a carrier of the Schistosoma parasite that is currently infecting more than 200 million people 85% of which live in Africa, while exposing 700 million rural workers to risk of infection. In addition B. glabrata is an invasive species in many regions around the world increasing the risk of the spread of parasitic schistosomes and making it an important species of interest for various investigations (Pointier et al., 2005) especially under fears of range expansion due to climate change (Poulin, 2006, Zhou et al., 2008).

2.3 Methods

Collection: B. glabrata snails were a kind gift from Dr. Coen Adema (CETI UNM), C. gigas oysters were a kind gift from Dr. Dennis Hedgecock (USC), flash frozen L. gigantea was a kind gift from the Joint Genome Institute (JGI) and A. californica was obtained from the National Resource for Aplysia (University of Miami). Live snails were relaxed using MgCl₂ and mantle tissue was dissected and snap frozen in liquid nitrogen and preserved for RNA extraction and cDNA preparation.

cDNA library construction and sequencing: Total RNA is isolated by using Trizol (Cat no. 15596018 Thermo-Fisher) reagent method with the following modification: the upper phase was extracted twice with chloroform rather than once. The RNA was further cleaned using RNAeasy with the DNase treatment step. RNA quality was verified using NanoDrop 2000 (Thermo-fisher) and Bioanalyzer 2100 (Agilent Genomics). RNA was shipped on Dry Ice to KAUST genomic facility and sequenced there. Libraries were constructed for from every mantle tissue RNA and pooled with equimolar ratios and sequenced on Illumina HiSeq 2000 2 × 100 bp paired-end sequencing, according to the manufacturer’s protocols.

Assembly and annotation: fastq files were quality filtered and trimmed using Trimmomatic (Bolger et al., 2014), then were assembled using Megahit (Li et al., 2015). Expression counts were performed using Kallisto (Bray et al., 2015). Prior to quantification reads matching to the mitochondrial genomes of each species were removed using
<table>
<thead>
<tr>
<th>Species Name</th>
<th>Number of PE reads</th>
<th>Assembled transcripts</th>
<th>Biomineralization related transcripts</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. californica</td>
<td>22,949,835</td>
<td>42,170</td>
<td>111</td>
</tr>
<tr>
<td>B. glabrata</td>
<td>75,470,630</td>
<td>86,222</td>
<td>507</td>
</tr>
<tr>
<td>L. gigantea</td>
<td>59,916,759</td>
<td>5,074</td>
<td>30</td>
</tr>
<tr>
<td>C. gigas</td>
<td>81,523,842</td>
<td>85635</td>
<td>569</td>
</tr>
</tbody>
</table>

Table 2.1. Summary of sequencing reads and assembled transcripts. Lists of hits to known biomineralization proteins are in supplementary file 1

BBmap’s bbduk pipeline (Bushnell, 2016) except for L. gigantea where no mitochondrial genome is publicly available, the mitochondrial genome of L. digitalis was used instead. Transcripts were annotated using dammit (Scott, 2016). For every species, we attempted to identify the most expressed transcripts, and identify biomineralization related protein-coding transcripts. The transcriptomes were searched against the BioMine database (Hanna et al.). Venn diagram was constructed using Venny (Oliveros, 2007).

2.4 Results

The sequencing yielded more than 20 million reads per sample. Except for L. gigantea, the assembly was successful in reconstructing more than 40 thousand transcripts Table 2.1. Using the combination of dammit and BioMine we were able to detect biomineralization related proteins even in low-yield sample as L. gigantea. L. gigantea contained 2,168,831, B. glabrata 64,924,257, A. californica 26,542,545, C. gigas 60,322,268 unique 25-mers. The unique k-mer counts corresponds nicely with the total number of assembled transcripts. Unfortunately, L. gigantea had a large portion of ribosomal RNA content so recovered assembled mRNA sequences were not comparable to the rest of the samples. This highlights the fact the complexity of the transcriptome is an important factor rather than mere depth of sequencing in achieving adequate transcriptome assemblies. A. californica contained a number of biomineralization related sequences similar to other shell-forming taxa. In particular, we were able to detect multiple splice variants of secreted chitinases and alpha carbonic anhydrases in the A. californica, B. glabrata and C. gigas Figure 2.4.
2.5 Discussion

Currently three studies have reported on the shell proteomes of *L. gigantea* (Mann et al., 2012, Marie et al., 2013a, Mann and Edsinger, 2014), while we were not able to recover a comprehensive transcriptome from the sample we had, we were still able to recover at least some of the proteins found in the shell proteome such as PIF. It was also reported that there was more shared shell matrix proteins (SMPs) between *L. gigantea* and other bivalve species rather than gastropods (Marie et al., 2013a). We see a similar trend between *B. glabrata* and *C. gigas* in this study, however this could be due to the fact that compared A. californica, both *B. glabrata* and *C. gigas* harbor a complete developed shell. Additionally, it is conceivable that this result can be due to the taxonomic bias in available public data from molluscan taxa. We detected multiple copies of the granulin/epithelin in *B. glabrata*, *A. californica* and *C. gigas*. This particular protein was recently found to have a new crucial role in biomineralization in the pearl oyster *Pinctada fucata* using knock-down experiment (Zhao et al., 2016). In mammals this protein is involved in skeletal formation and controlling the mineral content of enamel in addition to bone regeneration (Xu et al., 2007, Cao et al., 2010). Such pattern of shared function between proteins involved in bone formation and shell formation have been proposed before (Westbroek and Marin, 1998). Another protein that is also found in vertebrates is Asporin. Multiple transcripts from the 4 taxa were annotated as Asporin. Asporin is an extracellular protein that binds collagen in the presence of calcium and is involved in bone formation and considered an osteoblast marker (Yamada et al., 2007, Kalamajski et al., 2009). We speculate that Asporin might be involved in shell formation in mollusks. The top most expressed biomineralization transcript in all the samples was Ubiquitin. While Ubiquitin is common protein that is considered a house keeping protein that regulates many other cellular processes it has been particular found play an important role in the regulation CaCO$_3$ mineralization in prismatic layer of *P.fucata* (Fang et al., 2012). The second most abundant biomineralization related sequence coded for Ferritin containing proteins in all the samples. Ferritin is an important protein in mollusks as it regulates iron concentration in the hemolymph and on the cellular level by storing iron(Salinas-Clarot et al., 2011). In addition to its iron homeostasis role, ferritin was found in high concentrations in the mantle fold responsible for shell formation in *P.fucata* (Zhang et al., 2003). However, the exact function of ferritin in shell formation still remains to be determined. The
mantle tissue is a complex organ that plays takes on important roles in the biology of mollusks. In addition to shell formation, the mantle is a sensory organ and also serves an immune function (Jones and Saleuddin, 1978, Loker et al., 2004). Thus while transcriptomic analysis reveals many biomineralization related proteins, it is possible that many of these proteins are likely to have more than one function, for example proteases and protease inhibitors are known to be involved in biomineralization and also immune defense (Loker, 2010, Marie et al., 2010).

2.6 Conclusion

In conclusion, we sequenced 4 mantle transcriptomes, our results highlight the conserved repertoire of biomineralization related transcripts even in lineages that underwent shell-loss. The data presented in this study serves as steppingstone to explore further the evolutionary relationships among mollusks. Especially among groups extremely different morphology such as shell-loss and adaptations to completely different habitats such as freshwater versus marine environment and the implications of that for shell formation. Furthering our understanding of the molecular mechanisms that underlie shell-formation in different environments and taxa is vitally important as many mollusks and marine organisms face global challenges due to ocean acidification (Moya et al., 2016).
Figure 2.1. Shared biomineralization proteins between the different mantle tissues compared. The largest shared group is between *A. californica*, *B. glabrata* and *C. gigas* while the 2nd largest group is between *B. glabrata* and *C. gigas*. Breakdown of shared families are in supplementary file 1. The Venn diagram was constructed from unique protein names, treating multiple transcripts that are annotated the same as one protein.
Chapter 3
Microbes stabilize coral-algal symbiosis

"While the part change, the whole always remains the same. For every thief who departs this world, a new one is born. And every decent person who passes away is replaced by a new one. In this way not only does nothing remain the same but also nothing ever really changes.

—Law 39 Shams Al Deen Tabrizi’s 40 laws of love

3.1 Abstract

Symbionts would cooperate or defect conditional on the context in which the symbiosis is established. Symbiotic interactions are dynamic and most mutualisms are facultative and evolved from conflicting interactions. Here we studied the dynamics of the symbiosis among members of the coral holobiont. The symbiosis generates a nutrient efficient recycling mechanism in an oligotrophic system that generates one of the most diverse marine ecosystems on earth. To understand the dynamic nature of the symbiosis among corals, algae and bacteria; we modeled the metabolic capacity of each partner from genomic data. The *Orcicella faveolata* holobiont is a robust metabolic system with high redundancy in all essential enzymatic reactions. Surprisingly, algae become parasitic to corals in the absence of bacteria. The coral-algal symbiosis is a dynamic interaction of conflict and cooperation, stable and mutualistic only through bacterial metabolic contribution.
3.2 Introduction

A holobiont is defined as an organismal host that harbors a complex symbiotic microbiome (Gilbert, Bosch et al. 2015). Symbionts cooperate or defect to enhance their performance resulting in interactions that range from mutualistic to parasitic. A tight intracellular mutualism occurs between corals and single cell algae of the genus Symbiodinium. Corals survive in oligotrophic waters by nutrient recycling and carbon translocation from their algal symbionts (Falkowski, Dubinsky et al. 1984, Pernice, Meibom et al. 2012, Tremblay, Grover et al. 2012). While the nature of the coral-algal symbiosis has received a lot of attention, the study of other coral associated microbes has only occurred at the dawn of the 21st century (Rohwer, Seguritan et al. 2002). Using whole genome data for metabolic complementarity reconstruction, our analysis revealed Symbiodinium’s metabolic capacity is limited and requires support from not only the host, but also the bacterial component of the holobiont. Thus, we showed that the coral-algal interaction is not always mutualistic and it is the bacteria that act as mediators to generate a successful coral-algal-bacterial symbiosis. The emergent ensemble of interactions among multiple partners confers robustness to the holobiont. Our results demonstrate the importance of multipartite interactions in the evolution of symbiotic assemblages and the utility of genomics to comprehensively study these multidimensional systems.

Advances in genome sequencing have uncovered a vast microbial diversity in coral holobionts, raising questions about the potential role of these microbial communities and their placement in the spectrum from facultative to obligate symbioses. Even under strictly axenic conditions, Symbiodinium cultures maintain characteristic bacterial associations, which suggests a dependence of Symbiodinium on bacteria (Ritchie 2011, Shoguchi, Shinzato et al. 2013). While we are beginning to appreciate the microbial complexity associated with the coral holobiont, we have yet to understand the dynamic and metabolic interactions underlying these multipartite relationships. It has been suggested that Symbiodinium may have shifted from parasitic to mutualistic associations with coral hosts multiple times independently (Stat, Morris et al. 2008). Although the definition of parasitism is not agreed upon, most definitions state that a parasite must cause harm to the host (Zelmer 1998). Here, we use the term ‘parasitism’ to strictly describe asymmetrical dependency between partners leading to ‘metabolic parasitism’, stressing the fact that the nature of the interaction is dynamic and can change between
these various states.

To understand the metabolic potential of the coral, Symbiodinium and bacterial partners in a holobiont, we sequenced genomes of the Caribbean coral *O. faveolata* and analyzed them in simultaneously with available genomic data for Symbiodinium spp. and associated bacteria. Reverse ecology methods (Li, Costello et al. 2008), can be used to quantify the ability of a host species to supply the nutritional requirements of its microbial symbionts. In particular, the Biosynthetic Support Score (BSS) has been a powerful tool for this purpose. BSS reflects the ratio of metabolic reactions in one organism that can be supplied with reactions from another. BSS analyses have revealed that parasitic bacterial species have a high BSS relative to free living species (Borenstein and Feldman 2009). We constructed metabolic networks for both *O. faveolata* and one of its putative algal symbionts Symbiodinium minutum (Clade B) using the publicly available genome and transcriptome data. In this coral holobiont we observed that the fraction of Orbicella reactions supported by Symbiodinium is strikingly low (BSS = 0.2) while the fraction of Symbiodinium reactions supported by Orbicella is comparatively high (BSS = 0.7). These numbers reflect a more parasitic Symbiodinium life style (Figure 3.1). To validate our findings, we compiled all available genomic and transcriptomic data from both symbiotic and non-symbiotic cnidarians, and Symbiodinium. Our analysis exposed a similar pattern to what we observed in the *O. faveolata* holobiont – that is, Symbiodinium BSSs resemble a parasitic lifestyle in all cnidarian hosts.

### 3.3 Methods

#### 3.3.1 Genome sequencing, assembly, and annotation

**3.3.1.1 Material and DNA preparation**

Multiple colonies of Orbicella were used in this analysis DNA was extracted from sperm collected from Colony 452 from Puerto Morelos (20.874898°N, -86.851228°W) during coral spawning season of 2009, Sperm from colony 33 from the Florida Keys (24.812697° N, Florida 80.66925° W) collected in Sep 2012 and from Bocas Del Toro Panama (9.325572°N, -82.197692°W) (coordinates are taken on a WGS 1984 projection) from the spawning season of 2013.
Colonies were netted prior to spawning as described in (Voolstra et al., 2009) and bundles were collected prior to breaking up. The bundles were first filtered using a plastic sieve 1 mm pore size, to eliminate large sand particles and other foreign organism trapped inside the collection tubes. Bundles were left to separate on boat or at the lab and the sperm was separated from the eggs using a 100 µm cell strainer (Cat no. 08-771-19, Fisher Scientific), this was followed by passing through a 70 µm cell strainer (Cat no. 08-771-2, Fisher Scientific) in sterile polypropylene 50 ml centrifuge tubes (Cat no. 430290, Corning). Then the procedure is repeated for the sperm through a 20 µm cell strainer and a 5 µm cell strainer constructed from mesh and glued on PVC pipes. At this stage, the sperm is clean and mostly free from eggs and Symbiodinium cells.

Depending on the amount of spawn obtained, the sperm is then split into 50 ml centrifuge tubes and mixed with a 0.22 µm filtered sea water. The tubes are centrifuged at 1000 x g at 4 ºC. After this initial centrifugation, the pellet is separated from the supernatant into separate tubes. The pellet usually contains few pigmented cells that fluoresce under blue light. These cells are likely symbiodinium or other algae. Thus the supernatant is collected and redistributed and washed again using 0.22 µm FSW, and pelleted (the pellet is retained in the next steps). This is repeated at least 10 times and every time a sample is taken and examined to make sure it is free of visible contaminants. Approximately, 300 ml of bundles, yield 1 ml of concentrated clean sperm. The sperm is aliquoted in 300 µl cryogenic vials using a wide bore pipette tip, and instantly frozen in liquid nitrogen, then transferred to -80 ºC until further processing.

### 3.3.1.2 DNA extraction

DNA was extracted using QIAGEN genomic-tip (cat no. 10243) method. Briefly, the sperm pellet around 300 µl was incubated in the lysis buffer G2 with RNAase, and excess proteinase K for 8 hours in a water bath at 55 º C. After the lysis, the sperm was completely digested, the mixture was then diluted with buffer QBT 1:1 ratio, then loaded onto the genomic tip, after the washing steps, the DNA was spooled using a sterile glass shepherd’s crook and dissolved in 500 µl of TE buffer pH 8.0 (Sigma Aldrich cat no. 93283-100ML) and stored at 4 º C. The DNA quality and quantity was assessed using NanoDrop, Qubit systems and agarose gel electrophoresis.
3.3.1.3 Sequencing and library Preparation

Sequencing was conducted by Alpha Hudson Inc. 2 Hi-Seq lanes were sequenced from the Mexican Colony using pair-end 100 bp TruSeq library, for the Florida colony which became the main reference genome we sequenced 4 Illumina hi-seq libraries using unamplified NuGen standard method, 2 mate pair libraries sized at 6-9 kb, MiSeq run of an 800 bp sized NuGEN library, and additional MiSeq for an additional small mate pair library 2-3 kb size. The Panama sample was ran using 2x250 PCRfree libraries twice on the MiSeq.

3.3.1.4 Genome Assembly and scaffolding

Raw FASTQ files were trimmed for adapters and quality using Trimmomatic (Bolger et al., 2014) followed by fastq-mcf https://code.google.com/archive/p/ea-utils/. The sequences were reexamined using FastQC http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ to ensure the quality of the sequences going into the assembly. Reads were decontaminated for PhiX and other common lab contaminants using Bbmap (Bushnell) The reads we error corrected using RACER (Ilie and Molnar, 2013) Abyss (Simpson et al., 2009) version 1.5.3 was used for the assembly, the optimal K-mer used for assembly was determined by Kmergenie (Chikhi and Medvedev, 2013). The assembly was then scaffolded with multiple rounds of SSPACE (Hunt et al., 2014) . Then gaps were filled using GapCloser (http://soap.genomics.org.cn/). To improve the scaffolding, we sequenced additional libraries using the Chicago method, used in conjunction with the Hi-Rise scaffolding software (Dovetail Genomics LLC).

3.3.1.5 Repeat Masking

Repeat masking was conducted using RepeatMasker in addition to constructing custom repeat libraries using RepARK (Koch et al., 2014) and re-running them with RepeatMasker (Smit et al., 2014).

3.3.1.6 Gene Prediction

We mainly used AUGUSTUS (Stanke and Waack, 2003) for gene prediction, we started to train a model by using manually aligned transcripts from the assembled transcriptomes to the genome, we used 100 non-redundant genes, that produced high quality
alignment where we could identify acceptor and donor sites in addition to UTRs. These data were used to train the gene prediction model according to the instructions provided with the software. We used the parameter file created by AUGUSTS training to predict genes in the various genomes.

3.3.2 Holobiont Transcriptomes

We used raw data from this study (Pinzón et al., 2015), in addition to libraries that were sequenced in our lab from a coral bleaching experiment (medina lab unpublished), in addition to one library from a control fragment in a tank experiment at normal conditions (Medina lab unpublished). These libraries were prepared using the RiboZero method which depletes ribosomal RNA, to generate metatranscriptomes. In addition, we sequenced multiple pooled developmental stages using Roche 454 technology. The transcriptomes were assembled using Trinity (Grabherr et al., 2011) and de-replicated using CD-HIT (Fu et al., 2012).

3.3.3 Symbiodinium transcriptomes and Genomes

_Symbiodinium_ transcriptome assemblies were downloaded from NCBI TSA archives [http://www.ncbi.nlm.nih.gov/genbank/tsa/] accession numbers GBGW01, GAKY01, GBRZ01, GAFO01, GBSC01, GAFP01, GBRR01, in addition to the 2 other transcriptomes from (Bayer, Aranda et al. 2012). The _Symbiodinium minutum_ genome was obtained from [http://marinegenomics.oist.jp/]?

3.3.4 Metabolic reconstruction

Metabolic reconstruction was done by scanning the predicted gene models, in addition to the actual genomes for enzymatic proteins using the PRIAM database (Claudel-Renard et al., 2003) in combination with KAAS webserver (Moriya et al., 2007). Resulting SBML files were then manipulated using Python and Perl Scripts available here [https://github.com/bishoyh/OrbicellaGenomePaper]. MetaNetX (Ganter et al., 2013) was used for identification of dead-end metabolites and cleaning up of the metabolic models and conducting logical operations on the metabolic models. We also used OptFlux (Rocha et al., 2010) and F.A.M.E (Boele et al., 2012) to conduct flux balance analysis, identify problems with the models and visualize them.
In addition, to computational predictions and annotation of metabolic pathways, we have manually curated the metabolic pathways in the genome using a printed poster of the Roche Metabolic pathways chart (http://biochemical-pathways.com), http://web.expasy.org/pathways/. The chart was divided into quadrants, within each quadrant the reactions where identified and the EC number for each enzyme was looked up at http://enzyme.expasy.org/, candidate proteins for the enzyme were selected from the most closely related available taxa. The protein was then blasted against the predicted protein models using BLASTP from the genome and the actual genomic sequencing using TBLASTN. The identified homologs in the *Orbicella* were then annotated using HMMER3 phmmer https://www.ebi.ac.uk/Tools/hmmer/search/phmmer to ensure the identified protein was in fact coding for the particular enzyme. A label with the protein ID and the genomic scaffold coordinates was placed on the printed map to present the existence of such metabolic reaction in the genome.

### 3.3.5 NetCooperate

In order to calculate the BSS scores, the metabolic models derived earlier were parsed into a network format compatible for the input of NetCooperate (Levy et al., 2015), in which every reaction is broken down into nodes representing metabolites, in this case encoded as KEGG compound ids. This was achieved using a Perl script that converts SBML v3 format available here https://github.com/bishoyh/OrbicellaGenomePaper NetCooperate was obtained from http://elbo.gs.washington.edu/software_ netcooperate/NetCooperatePython.zip

### 3.3.6 Network analysis

At multiple stages, of the analysis we utilized Cytoscape (Shannon et al., 2003), Gephi (Bastian et al., 2009), Pajek (Batagelj and Mrvar, 1998) and a local modified version of Gource http://gource.io/ to work with biochemical networks available in the github repository.
3.3.7 Robustness analysis

GML files were exported using Cytoscape and the simulations were ran using the algorithm described in (Iyer et al., 2013) a modified version of the implementation https://github.com/bwallace/robustness, to mainly fix problems with GML format and streamline the analysis for our purposes the modified code is also posted in the github repository https://github.com/bishoyh/OrbicellaGenomePaper.

3.3.8 Symbionts of Orbicella

While Orbicella can harbor many different species of Symbiodinium (Rowan et al., 1997), mostly clade B symbionts have been identified in Orbicella (Green et al., 2014). Symbiodinium minutum was likely a co-isolate when the initial cultures were being isolated from Orbicella, however it is not likely a dominant symbiont of Orbicella in the wild, and it is most likely another uncultivable clade B symbiont (M.A.C, TC LaJeunesse personal communication). However, it is notable that Symbiodinium minutum can infect and establish sustainable symbiosis during lab experiments (Voolstra et al., 2009). Mostly we chose to use it due to it being the only available genome, that is likely to resemble a Clade-B symbiont present in Orbicella.

3.4 Results

The metabolic reconstruction results presented here bring into question the prevailing paradigm that the coral-algal interaction is mutualistic, especially given that coral-algal symbiosis is widely considered to be the main driver of coral reef ecosystem maintenance. Based on this working hypothesis it became apparent that we must consider the role of bacteria within the coral holobiont. Specifically, we adapted our model to include the bacterium Parvibaculum sp., an inextricable bacterial symbiont of Symbiodinium minutum (Shoguchi et al., 2013). When the metabolic capacities of Symbiodinium minutum and Parvibaculum sp. are combined, the BSSs of the merged network supporting coral reactions doubles from 0.2 to 0.4, while the BSS of Orbicella reactions decreases from 0.7 to 0.5 (it is also worth noting that BSS of S. minutum, and Parvibaculum sp. is 0.79 and 0.3 respectively). The similarity of those BSS values (0.4 and 0.5) indicates a mutualistic relationship between Symbiodinium and its coral host.
This shift from parasitism to mutualism is also seen when the bacterial metatranscriptome is merged with the Symbiodinium metabolic model (Figure 3.1 and Figure 3.3). By using metatranscriptomes in the metabolic model we confirm that many of the metabolic genes predicted from genomes are in fact expressed within the holobiont in the wild. Symbiodinium biosynthetic capacity dramatically increases when acting in concert with other microbial members. This clearly demonstrates of the complexity of the coral holobiont and its dependence on diverse microbial metabolic services. Our analysis underscores the important role of bacteria in stabilizing the coral-algal interaction.

In addition to catalyzing the coral-Symbiodinium mutualism, microbes also increase metabolic robustness of the coral holobiont (Figure 3.2). While it is difficult to assess metabolic network robustness given its intrinsic dynamic nature (Wagner, 2013), we can consider how the structure of the network changes as it is degraded through the removal of metabolites and reactions (Schneider et al., 2011). We used percolation theory to study the metabolic network robustness (Callaway et al., 2000). If an attack on the network removes enough vertices, the largest connected component of the network shrinks and eventually collapses. In such degraded network, processes are disrupted (Holme et al., 2002). We compared the fraction of vertices in the largest component ($\sigma$) vs. the fraction of vertices randomly removed ($\rho$) to estimate robustness in the *Orbicella* holobiont. Vulnerability $V$ is a normalized robustness measure that allows comparing networks of different sizes (Iyer et al., 2013). We observed that the combined metabolic network (constructed both metatranscriptome data and the combination of the metabolic models for each taxon) is at least two times more robust relative to any individual member of the holobiont.

This highlights the strong interdependence among all members. As a consequence, these findings shed light on the adaptive advantages conferred by the coral’s symbiotic network when compared to aposymbiotic cnidarians such as *Nematostella vectensis*. *N. vectensis* forms associations with bacteria but not *Symbiodinum* and follows a sedentary carnivorous lifestyle (Har et al., 2015). Corals are subject to a wide range of environmental stressors. A robust metabolic network provides greater resilience to the heterogeneous metabolic demands that corals face on a daily basis that includes complete hypoxia at night (Revsbech et al., 1995), UV radiation and high concentrations of reactive oxygen species during the day (Revsbech et al., 1995, Lesser and Farrell, 2004) as well as high rates of microbial turnover (Rohwer et al., 2002b). By analyzing the num-

33
ber of nitrogen atoms per residue side chain (ARSC) as described in (Baudouin-Cornu et al., 2001) we tested for genomic signatures that members of the holobiont might have underwent biosynthetic cost minimization. One method marine microorganisms minimize the cost of proteins in nitrogen limited environment is by selectively substituting amino acids with less nitrogen atoms in the side chains of proteins (Grzymski and Dussaq, 2012). We found that Symbiodinium proteins reflect a nitrogen rich environment relative to most bacteria found in corals and the host. This is an unexpected result since, corals live in oligotrophic waters. However, this trend might reflect an adaptation to the endosymbiotic lifestyle. It is also worth noting that different clades of Symbiodinium differ in their N ARSC numbers (Figure 3.4). Such differences between clades highlight the importance of understanding the role of various Symbiodinium strains in establishing unique host-symbiont interactions with different metabolic demands.

3.5 Discussion

Taken together, our results highlight the roles multipartite associations have had in the evolution of a robust and successful coral holobiont. In brief, we have uncovered a novel role for the underappreciated microbial community that thrives within and its contribution to the metabolic robustness of the holobiont. The metabolic network robustness conferred by the interaction with diverse microbes allows corals to effectively function under wide ranging metabolic demands (Rohwer et al., 2002a). More importantly, the synchronous multipartite association among corals and their microbes shown here brings new light to the study of coral bleaching. The majority of bacteria are likely to have high replication rates than Symbiodinium due to the larger (i.e. over 1,000-fold) genome sizes of Symbiodinium spp. (LaJeunesse et al., 2005b). Under nutritional constraint these dynamics will induce the holobiont to become less hospitable for Symbiodinium. We hypothesize that under heat stress, the metabolic demands of the holobiont cannot be met and Symbiodinium becomes more of a burden to the system leading to the breakdown of symbiosis (Lesser, 1997). It is thus likely that as a result of increased sea surface temperatures, the metabolic balance among holobiont partners will be affected by shifts in the microbial community. Our comprehensive approach to studying multipartite associations between a coral host and its microbial consortium opens the possibility of developing experimental studies to better understand how coral holobionts will fare under the threat of a changing climate and polluted
Figure 3.1. BSS scores for a variety of host-symbiont interactions. BSS scores are normalized to 1. Combining the metabolic networks of both *Symbiodinium minutum* and *Parvibaculum sp.*, changes the interaction between parties to resemble mutualistic interactions. We evaluated other eukaryotic parasites (*Toxoplasma gondii* and *Plasmodium falciparum*) in their native hosts (*Mus musculus* and *Aedes aegypti* respectively), as well as free-living photosynthetic algae (*Chlorella NC64A* and *Guillardia theta*). Our method can differentiate between parasitic and non-parasitic eukaryotes.
Figure 3.2. Comparison of metabolic network reconstructions of each coral holobiont member along with the overall coral holobiont metabolic network based on genomic and metatranscriptome data. The lines show the relationship between $\rho$ and $\sigma$, which reflect the sensitivity of each metabolic network to random attacks. The more vulnerable (V) a system the less robust it is (numbers in box). The Orbicella holobiont metabolic network is less vulnerable to random attacks than each member alone highlighting the interdependence of the three major partners and the important role of bacteria.
Figure 3.3. The metabolic network of the coral holobiont based on the integration of microbial meta-transcriptomes and three coral genomes generated in this study with genomic information from the algal symbiont. The interdependence of the three major partners and the important role of bacteria is highlighted by the number of connections between different metabolic reactions. Each node is a metabolite; each edge designates a metabolic reaction that transforms into another metabolite. Colors of edges represent the same metabolic reaction connecting the different metabolites (nodes).
### Figure 3.4

The nitrogen content using the metric of number of N atoms per residue side chain (ARSC) calculated for every protein in the available proteomes for the respective proteomes. *Symbiodinium* species have a high relative amount N atom per side chain than other organisms in the holobiont.
Conclusion

“When one has finished building one’s house, one suddenly realizes that in the process one has learned something that one really needed to know in the worst way - before one began.”

—Friedrich Nietzsche

Evolution of Biomineralization

A great amount of work has been devoted to discover if the majority of biological networks form a random topology and universal connectivity structure (Oliveros, 2007). Most of the networks investigated in biological systems exhibit a toboggan shape, or a hyperbole that can be expressed in terms of exponential laws (power laws). This critical observation implies that a small number of nodes can possess a very large number of connections (hubs) pointing to more order and regularity than expected in these networks. Biological networks such as protein-protein interaction networks are subjected to the “small-world” phenomenon (Barabasi and Oltvai, 2004), which allows the network to rapidly link two nodes taken randomly with the aid of few intermediate nodes. In addition, when a new node appears, it tends to connect itself to the existing nodes that have the most number of connections. The small set of hubs thus emerges by attaching the new incoming nodes preferentially to existing hubs (i.e. preferential attachment). Moreover, the networks do more than just grow; nodes equally leave the network when they do not satisfy the criteria that allow them to stay within the network. At a certain point the whole network acts as a selector. Therefore, a node is not admitted to the network forever and can disappear just after being recruited (Oltvai and Barabasi 2002, Barabasi and Oltvai 2004). Biomineralization is an excellent example of such a system, since it is based on multiple processes such as molecular transport, extra cellular matrix regulation, inhibition of calcification, etc.

The process of biomineralization can be regarded as a large-scale organized network (the top of the pyramid in Figure 5) where an underlying protein network that has different functional modules gives rise to its properties. The process is conserved at this top level (universality), yet exhibits organismal specificity on the downstream level where taxonomically restricted proteins exist (Figure 5). To answer the whether
biomineralization evolved once or multiple times in Metazoans, one concludes that the biomineralization capacity present in most metazoans must be a case of deep homology i.e. a trait that appears homoplastic at one level yet can be homologous at the developmental or molecular levels. One of the important outcomes of my analyses was the identification of a core set of candidate biomineralization protein domains that are shared across all metazoans (Figure 6). The shared domain toolkit I uncovered supports the various pre-requisites for metazoan biomineralization. Construction of a network that integrates the set of shared domains shows that the biomineralization network might be comprised of discrete modules, which facilitates integration of new genes, interactions, and regulatory processes, and hence adaptation. It remains, however, to be determined if this shared toolkit shares a conserved regulatory pattern or if divergence of regulatory pattern would explain the diversity of biomineralization schemes observed in various taxa. My computational pipeline BioMine represents an important tool to sift through computational data from various layers of biological organization (e.g. proteins, genes, transcripts, interactions, etc.), synthesizing information, and making it available in a quantitative and systematic way. In the future, I anticipate our BioMine web database to be a host for incorporation of new data and analyses. This pipeline can be easily extended by inclusion of newly discovered proteins yielding further insights into the set of shared protein domains in a biomineralizing network. Taxonomic coverage is critical to this kind of analysis, as the integration of more genomic data will allow for more robust predictions about the intricacies of biomineralization as a process.

**Orbicella Genome**

I believe that by making the *Orbicella* genome the “Open Source Coral Genome” would enable the Caribbean coral community to engage in deeper scientific experimentation and discovery of these reef builders. Using Reverse ecology methods, I found out that due to the Symbiodinium’s limited metabolic capacity and in absence of other holobiont members it would rely on the coral for obtaining most its amino acids and other important metabolites. This creates a metabolic dependency on the coral that is akin to the pattern of BSS scores witnessed in parasitic organisms.

When this relationship is examined in presence of symbiotic bacteria or other holobiont partners other than the coral and *Symbiodinium* (non-binary) many of the amino
acids are delivered by more than one member of the holobiont thus reliving the stress on the host and rendering the binary interaction of the host and Symbiodinium mutualistic rather than parasitic in absence of bacteria. By appreciating the importance of bacterial interactions and the services they provide for the coral animal and the Symbiodinium, better studies can be designed to incorporate this multipartite symbiosis rather than focusing on one side or studying the system in a binary fashion. The inconsistent metabolic patchwork created by the loss of metabolic function, makes more difficult to generalize patterns of evolution across all species of coral, for example Acropora digitifera cannot synthesize the amino acid cysteine, however this is not the case for Orbicella or other acroporids (Shoguchi et al., 2013) which only increases our appreciation of their diversity and most crucially to protect and conserve them in pristine conditions.

Many questions remain: Do bacteria play a role in the compatibility of the host with Symbiodinium species. Bacteria confer robustness to the metabolic network. But are the associations with bacteria permanent or transient? Are some of the bacteria obligate symbionts? Does Symbiodinium co-infect with strains of bacteria that are beneficial for its establishment? The results presented in this study establish clearly the role of microbes as the stabilizer of the algal-coral symbiosis. Stemming from that, a more fine-grain understanding of the coral-microbial interactions is urgently needed.

To be able to understand the role of microbes in the coral holobiont we need to go beyond metagenomics surveys using markers such as 16S rRNA. Methods such as single cell genomics and exploring new methods to culture or to co-culture taxa that we can’t grow in the lab yet will enable us to understand the physiology of these microbes and understand their actual role in the holobiont. More importantly is the question of symbiont fidelity, many of the current microbial studies of corals detect thousands of bacterial taxa or OTUs, however it is unclear if there exists a group of bacterial symbionts that are unique to corals, or taxa that co-evolve with corals. This question is important, because if bacterial taxa associated with corals are just transient, how can robust interactions form among the multiple symbionts.

This question goes beyond coral biology, and is in fact a general question about cooperation and symbiosis, as any kind of strong symbiotic interaction first starts as a short transient interaction. Computer modeling demonstrated that fluctuations in populations size and the ability to partition populations of cells plays an important role in microbial communities crossing the threshold from transient to stable coopera-
tion (Cremer et al., 2012). Corals are primed for both features, as they possess a very complex morphology entrenched within multiple microenvironments, spanning different nutrient gradients, in addition multiple secondary metabolites that control cell division/antibacterial/cytotoxic molecules have been isolated from multiple coral taxa (Harder et al., 2003, Geffen and Rosenberg, 2005). Therefore, corals can be a hotbed for evolving microbial symbiotic interactions. This hotbed is likely to harbor more non-bacterial symbiotic interactions than those I investigated, for example, fungal-coral and fungal-bacterial interactions are still at their infancy in coral research (Yarden, 2014), albeit being heavily investigated in other settings such as soil, plants and the biomedical field (Wargo and Hogan, 2006, Frey-Klett et al., 2011). More studies that integrate both knowledge about bacterial and fungal symbionts of the holobiont are likely to yield novel insights into the biology of the coral holobiont. Taken together, these findings highlight the importance of approaching the coral holobiont as a many-to-many interaction network, rather than focusing on the individual interactions between Symbiodinium and the host.

**Figure 5.** The Pyramid of complexity. Different levels of conservation in biomineralization processes based on preliminary data. The three hierarchical layers are represented as 1) The protein interaction network as the most conserved portion of the organization, 2) The universal biomineralization proteins in the middle, and 3) The species-specific proteins at the bottom.
Figure 6. Distribution of shared Pfam families across biomineralizing taxa


Eric S Loker, Coen M Adema, Si-Ming Zhang, and Thomas B Kepler. Invertebrate immune systems—not homogeneous, not simple, not well understood. *Immunological reviews*, 198(1):10–24, 2004. ISSN 1600-065X.


Vita
Bishoy Kamel

Biology Department
208 Muller Laboratory
University Park, PA 16802 U.S.A.

Phone: 209-761-2681
email: bishoyh@gmail.com
URL: http://bishoy.kamel.space
Born: February 12, 1986—Cairo, Egypt

Current position

PhD Candidate, The Pennsylvania State University

Education

2009  MSc in Biotechnology, American University in Cairo
2007  BSc in Biology, American University in Cairo

Grants, honors and awards

2014  Microsoft Azure Cloud Computing Research Award
2011  EDEN-NSF Research Exchange grant
2010  NRT fellowship, University of California, Merced
2010  Travel Award, World Congress of Malacology
2009  Travel Award, OIST Winter Course "Evolution of Complex Systems" (OWECS)
2008  Travel Award, 11th International Coral Reef Symposium 2008
2003  ACS Student Award, Academic Computing conference 2003