© Copyright 2019

James L. Dimond

Patterns, dynamics, and potential roles of DNA methylation in reef corals and their allies

James L. Dimond

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Steven Roberts, Chair

Lorenz Hauser

Kenneth Sebens

Program Authorized to Offer Degree:

School of Aquatic and Fishery Sciences

University of Washington

Abstract

Patterns, dynamics, and potential roles of DNA methylation in reef corals and their allies

James L. Dimond

Chair of the Supervisory Committee:

Associate Professor Steven Roberts

School of Aquatic and Fishery Sciences

Epigenetic processes, which contribute to gene regulation without affecting underlying DNA sequences, are increasingly recognized as molecular mechanisms that shape phenotypes. DNA methylation is the best understood epigenetic process and has been shown to mediate environmental effects on gene expression and phenotype in a wide range of taxa. However, most of what is known about DNA methylation is based on model organisms, particularly vertebrates, while much less is known about DNA methylation in other organisms such as invertebrates. Tropical reef corals are long-lived, sessile invertebrates that are thought to be particularly reliant on physiological acclimatization and phenotypic plasticity to cope with environmental variation. The underlying basis of this plasticity could lie, at least in part, in epigenetic mechanisms like DNA methylation. The aim of this dissertation research is to examine DNA methylation in corals

with respect to its patterns, variability, response to change, and involvement in phenotypic and transcriptional plasticity.

Chapter one explores patterns of evolutionary-scale DNA methylation in corals, using CpG depletion analysis to estimate methylation levels in publicly available transcriptome data from six coral species. Consistent with what has been documented in most other invertebrates, all corals exhibited bimodal distributions of germline methylation suggestive of distinct fractions of genes with high and low levels of methylation. The hypermethylated fractions were enriched with genes with housekeeping functions, while genes with inducible functions were highly represented in the hypomethylated fractions. In three of the coral species, genes differentially expressed in response to thermal stress and ocean acidification exhibited significantly lower levels of methylation. These results support a link between gene body hypomethylation and transcriptional plasticity that may point to a role of DNA methylation in the response of corals to environmental change.

Chapter two evaluates the hypothesis that DNA methylation patterns reflect phenotypic differences among Atlantic branching corals of the genus *Porites.* Using reduced representation genome sequencing, genetic and epigenetic diversity among 27 colonies of *Porites* spp. from Belize were compared. Contrary to expectations, epigenetic patterns were inconclusive, while genetic data showed clear separation of three distinct genetic groups. One of these groups exhibited significantly thicker branches, and branch thickness was a better predictor of genetic groups than depth, habitat or symbiont type. Relationships between genetic and epigenetic patterns suggest that epigenetic patterns reflect diverse environmental histories superimposed over a relatively small heritable component. Meanwhile, the clear genetic patterns revealed by

high throughput sequencing suggest that this technique could hold promise for resolving the phylogeny of this taxonomically difficult group.

Chapter three asks whether DNA methylation in corals exhibits plasticity in response to environmental change. Colonies of the mounding coral *Porites astreoides* on the Belize Barrier Reef were experimentally transplanted to a common garden in 2015 and resampled the following year. Methylation levels in both years were determined via reduced representation genome sequencing (RADseq). Methylation status was largely stable over time, with only ~2% change in methylation over the one-year period. However, there was evidence for convergence of the methylation state of corals after a year in a common environment together. The loci that changed their methylation status were associated with mRNAs and non-coding RNAs. This work suggests that while DNA methylation levels in corals are not fixed, their response to environmental change may be subtle.

Chapter four is a functional genomics study testing the hypothesis that DNA methylation is associated with symbiotic state and alternative mRNA splicing, using symbiotic and aposymbiotic individuals of the sea anemone *Anthopleura elegantissima*. The study leveraged third-generation Oxford Nanopore sequencing for its full-length reads and ability to detect base modifications. A largely complete *A. elegantissima* draft genome was generated from aposymbiotic individuals. There was no strong evidence for significant modification of the methylome according to symbiotic state, however, three genomic regions with consistently different methylation according to symbiotic state were identified. The region exhibiting the strongest difference was associated with a DNA polymerase zeta that is noted for its role in translesion synthesis, which opens interesting questions about the biology of this symbiosis.

TABLE OF CONTENTS

List of Figures.......... iii

List of Tables.......... v

Introduction.......... 1

Chapter 1. Germline DNA methylation in reef corals: patterns and potential roles in response to environmental change.......... 19

1.1 Abstract.......... 19

1.2 Introduction.......... 20

1.3 Methods.......... 23

1.4 Results.......... 26

1.5 Discussion.......... 32

1.6 Acknowledgements.......... 37

1.7 Data Accessibility.......... 37

1.8 Literature Cited.......... 38

Chapter 2. Genetic and epigenetic insight into morphospecies in a reef coral.......... 44

2.1 Abstract.......... 44

2.2 Introduction.......... 45

2.3 Methods.......... 48

2.4 Results.......... 58

2.5 Discussion.......... 66

2.6 Acknowledgements.......... 71

2.7 Data Accessibility 71

2.8 Literature Cited 72

Chapter 3. Response of DNA methylation to environmental change in the reef coral *Porites astreoides* 80

3.1 Abstract 80

3.2 Introduction 81

3.3 Methods 83

3.4 Results 89

3.5 Discussion 96

3.6 Literature Cited 101

Chapter 4. DNA methylation profiling of a cnidarian-algal symbiosis using nanopore sequencing 107

4.1 Abstract 107

4.2 Introduction 108

4.3 Methods 110

4.4 Results 114

4.5 Discussion 121

4.6 Literature Cited 125

Appendix A 129

LIST OF FIGURES

Figure 1.1. Transcriptome-wide CpG O/E in the six coral species 27
Figure 1.2. Patterns of CpG O/E in relation to gene function (GOSlim biological processes terms) in the six coral species 29
Figure 1.3. CpG O/E vs. mean gene expression level. 30
Figure 1.4. Comparison of transcriptome-wide and differentially expressed gene CpG O/E in the acroporid corals 31
Figure 2.1. Examples of branching *Porites* spp. collected in this study...........................49
Figure 2.2. Determining methylated loci in a representative sample 57
Figure 2.3. Pairwise comparisons of SNP mismatches between all 25 individuals with both ddRADseq and EpiRADseq data...60
Figure 2.4. Genetic and phenotypic variation in *Porites* spp. 61
Figure 2.5. Examining patterns in *Porites* spp. CpG methylation....................................63
Figure 2.6. Differentially methylated CpGs among *Porites* spp. individuals 64
Figure 2.7. Association between pairwise genetic distance and pairwise epigenetic distance 66
Figure 3.1. Satellite-derived environmental conditions associated with coral collection sites and the common garden along the Belize Barrier Reef..90
Figure 3.2. Results of symbiont identification using BLAST searches of ddRADseq data against custom databases of Symbiodiniaceae...91
Figure 3.3. Pairwise analysis of SNP mismatches among the *P. astreoides* samples 93
Figure 3.4. Comparative methylation profiles of the eight *P. astreoides* samples from 2015 to 2016 94
Figure 3.5. Methylation changes from 2015 to 2016 95
Figure 4.1. Symbiotic and aposymbiotic *A. elegantissima*..111
Figure 4.2. Comparative methylation frequency among aposymbiotic and symbiotic *A. elegantissima* 117
Figure 4.3. Methylation and transcription patterns associated with contig 4024 118

Figure 4.4. Methylation and transcription patterns associated with contig 107 119
Figure 4.5. Methylation and transcription patterns associated with contig 1212 120

LIST OF TABLES

Table 1.1. Transcriptomes analyzed in this study 23
Table 1.2. Differentially expressed gene sets examined 24
Table 1.3. Transcriptome annotation results 26
Table 1.4. Results of mixture model analyses of CpG O/E in the six coral transcriptomes… 28
Table 2.1. *Porites* spp. specimen data 50
Table 3.1. BLAST hits for select differentially methylated loci 96
Table 4.1. Sequencing statistics by sample 115

ACKNOWLEDGEMENTS

This work could not have been accomplished without the support of numerous people and funding sources. First and foremost, I am truly lucky to be surrounded by a network of family and friends that have been a consistent source of encouragement over the past five years. In particular, I owe my lifelong quest for knowledge and scholarship to my parents, Allison White and the late Buz Dimond, who placed a premium on education. I also have to think that some of this is genetic (and perhaps epigenetic?). Indeed, my sister, Allison, shares my proclivity to seek detailed biological knowledge, and she also has encouraged me throughout this work.

My amazing wife, Julie, has been a steadfast supporter of this effort from the moment I first expressed interest in pursuing it. She was very forgiving of the numerous occasions where I was too absorbed in my work to give her my undivided attention, and she always expressed genuine interest in my research. She also ensured there were always plenty of diversions.

Some of my close colleagues and mentors deserve special mention. Randi Rotjan, Emily Carrington, Gisèle Muller-Parker, and Brian Bingham have all encouraged me to pursue a PhD at one time or another. Thank you all for giving me the confidence to go forward with this scholarly endeavor.

Moving on to those who were instrumental more specifically in supporting the work itself, Steven Roberts is of course at the top of that list. I met Steven years ago while working at the MBL in Woods Hole, and when it came time to find a PhD advisor, I didn't hesitate to contact him. My entry into SAFS was not a quick and easy one, but persistence paid off and I am extremely grateful to Steven for making it happen the way it did. Steven's dedication to open and reproducible science is an inspiration to me and something I hope to continue to do and pass on to others.

My dissertation committee members Lorenz Hauser, Ken Sebens, and Lauren Buckley have been very supportive throughout this process. They have asked excellent questions that have pushed me to consider my work from a variety of perspectives. And another shout-out here to Emily Carrington for agreeing to step in as GSR in Lauren's absence.

The excellent SAFS graduate and administrative staff, including Tim Essington, Amy Fox, and Sam Scherer helped me through numerous questions and procedures. I am particularly thankful for Tim and Amy's help facilitating various fellowships.

Much of the technical work in Chapters 2 and 3 could not have been accomplished without the training and lab equipment generously provided by Adam Leaché. In Adam's lab, Kevin Epperly provided expert instruction in RADseq techniques, an experience that really improved my bench skills as a molecular biologist.

Of course, much of the actual work herein was done computationally, and Sam White played a prominent role in helping me through that. Together, Sam and Steven helped develop and improve my bioinformatics acumen. Learning these kinds of new skills was exactly what I hoped to gain from my PhD work.

Besides Steven and Sam, other colleagues affiliated with the Roberts Lab have been a great source of help, advice, and encouragement over the years, including Grace Crandall, Mackenzie Gavery, Jake Heare, Megan Hintz, Hollie Putnam, Laura Spencer, Emma Timmins-Schiffman, Brent Vadopalas, and Yaamini Venkataraman.

For my work in Belize, I thank the Smithsonian Institution's Carrie Bow Cay Field Station, which is ably managed by Valerie Paul, Zach Foltz, and Scott Jones, not to mention many volunteers, and of course, chef extraordinaire Martha Nicholas. I hope to be able to return again sometime soon.

Finally, I feel very fortunate to have been the recipient of several fellowships that supported me financially and gave me the latitude to pursue my own ideas. The John E. Halver Fellowship at UW-SAFS was my primary source of tuition and stipend support. The Hall Conservation Genetics Research Award from the UW College of the Environment funded the bulk of the research in Chapter 2. An International Society for Reef Studies Graduate Fellowship was used to fund most of the work in Chapter 3. Last but not least, an ARCS Foundation Seattle Chapter Fellowship provided a much-appreciated source of discretionary funds that were used for aspects of all chapters, but principally Chapter 4.

DEDICATION

For Julie

INTRODUCTION

Phenotypic plasticity is the capacity of a genotype to produce a range of phenotypes in response to environmental heterogeneity. Environmental variability is thought to drive the evolution of phenotypic plasticity, as a greater phenotypic range will promote success in variable, unpredictable or rapidly changing environments. Because phenotypic plasticity permits a more rapid response to environmental change than is possible through natural selection, it is essential to the persistence of many species as anthropogenic global environmental change expands and intensifies (Charmantier et al., 2008, Chevin et al., 2010). The need to understand the underlying basis of phenotypic plasticity is therefore a fundamental biological question of particular contemporary significance.

Phenotypic change often involves modifications in gene expression. Epigenetic processes, which contribute to gene regulation without affecting underlying DNA sequences, are increasingly recognized as molecular mechanisms that contribute to phenotypic plasticity. The term epigenetics was coined by C .H. Waddington in 1942 primarily to account for the process by which an organism's phenotype changes throughout development (reprinted as Waddington 2012). Waddington recognized that a single genotype could give rise to multiple phenotypes during an organism's lifetime. The term means "above genetics", with the implication that there is some additional information aside from the genotype that can affect the phenotype. With the knowledge we have today about the existence and function of DNA, and our increasing understanding of potential pre- and post- transcriptional modifications of gene expression, the definition has evolved to be defined more narrowly as modifications of the genome that do not

involve changes to the underlying DNA sequence (Duncan et al. 2014). However, some go even further to define epigenetics as being *heritable* modifications of the genome without changes in the DNA sequence (Felsenfeld 2014).

DNA methylation is the most well-studied and best understood epigenetic process, in part because it is the most stable and easily measured process. DNA methylation involves the addition of a methyl group to a nucleotide base, which in animals occurs most often on a cytosine that occurs directly before a guanine and is termed CpG methylation. Methyltransferase enzymes are responsible for the establishment, maintenance, and erasure of methylation. DNA methylation is considered to be evolutionarily ancient and is present across the eukaryotic tree of life, with methylation of gene bodies being the most basal pattern (Zemach et al. 2010). Vertebrates tend to have highly methylated genomes (up to 80% or more), while those of plants and invertebrates tend to be less methylated (on the order of 1% to 20%) (Zemach et al. 2010). DNA methylation can be influenced by the environment, and methylation patterns can be associated with alternative phenotypes (Duncan et al. 2014). In many cases, methylation patterns appear to be stable across generations, making methylation a good candidate for transgenerational epigenetic regulation of phenotype (Miska & Ferguson-Smith 2016).

Although generally beyond the scope of this dissertation, histone modifications and small RNA molecules are two other types of epigenetic processes that deserve brief mention due to their potential interaction with DNA methylation as part of the epigenetic machinery. Histones are proteins that are closely associated with DNA and are responsible for its packing into nucleosomes, units of chromatin which collectively form chromosomes. Histones can be modified post-translationally by enzymes that affect their secondary structure, which in turn may influence the accessibility of the underlying DNA to transcription factors (Duncan et al. 2014).

Similarly to DNA methylation, histone modifications can be both created and erased by certain enzymes. Specific DNA sequences and non-coding RNAs may be involved in histone modification, and there is also evidence that DNA methylation and histone modifications work together to influence gene expression (Duncan et al. 2014). The degree to which histone modifications have a "memory" across generations is not well known, though there is some evidence that they can be carried through the germline and influence development (Miska & Ferguson-Smith 2016). Though not always considered to be a "classical" epigenetic process, small non-coding RNA molecules are also known to influence DNA methylation and chromatin structure, in addition to being involved in post-transcriptional silencing of mRNA (Duncan et al. 2014). There are several classes of small RNAs, with all being generally less than a few hundred nucleotides long, and most consisting of only 20-30 nucleotides. Aside from direct interference of gene expression through post-transcriptional silencing, there is evidence that small RNAs can be involved in an interactive epigenetic system that includes DNA methylation and histone modification. For example, DNA methylation and histone patterns may produce small RNAs that in turn can reconstruct these patterns (Miska & Ferguson-Smith 2016).

Returning to DNA methylation, a phenomenon known as gene body DNA methylation is considered the most ancestral form, occurring across all eukaryotic taxa (Zemach et al. 2010). Gene body methylation refers to patterns of genome methylation where methylation is concentrated within transcription units (genes). It is the primary form of methylation among invertebrates (Zemach et al. 2010) and is therefore the most relevant to this dissertation work. Despite its ubiquity, the precise functions of gene body methylation have remained somewhat equivocal. In contrast, most progress on the function of DNA methylation has been made on methylation of promoter regions in vertebrates, a pattern associated with silencing of

downstream genes. By contrast, gene body methylation is associated with active transcription (Duncan et al. 2014). But what does it do, if anything? Below I will discuss some hypotheses for its function.

Gene body methylation is often concentrated in housekeeping genes that perform highly evolutionarily conserved functions across taxa (Suzuki & Bird 2008; Sarda et al. 2012; Dimond & Roberts 2016). Meanwhile, less conserved genes, which are also often those associated with inducibility, tend to have lower levels of methylation (Roberts & Gavery 2012; Sarda et al. 2012; Dimond & Roberts 2016). This pattern has led to speculation that the function of gene body methylation may therefore be to promote consistent expression of genes that perform essential functions and for which aberrations would be potentially lethal (Bird 1995; Suzuki & Bird 2008; Roberts & Gavery 2012). This theory has also arisen from the fact that vertebrate genome methylation levels are much higher than those of invertebrates, and that methylation may have evolved as an essential function to reduce transcriptional noise (i.e., spurious transcription) in the more complex vertebrates with greater numbers of genes (Bird 1995). There is experimental evidence for this hypothesis; for example, Huh et al. (2013) showed a negative relationship between gene body methylation and transcriptional noise, and more recently Neri et al. (2017) showed that methylation protects the gene body from spurious entry of RNA Polymerase II. In contrast to vertebrates, Roberts & Gavery (2012) hypothesized that invertebrates may benefit from transcriptional noise to some extent because it presents more potential for transcriptional (phenotypic) plasticity, which may have greater benefit for less complex organisms that have less ability to buffer themselves from the surrounding environment.

In spite of its association with highly conserved housekeeping genes, DNA methylation is associated with increased mutation rates of methylated cytosines. However, one possible

explanation for this paradox comes from work by Dixon et al. (2016) on cnidarians. The authors found that in addition to predicting gene conservation and stable expression, there was evidence that gene body methylation drives optimal codon usage via the hypermutability of methylated cytosines. Specifically, Dixon et al. (2016) found that highly methylated genes were enriched with codons that code most efficiently during translation.

There is now substantial evidence that gene body methylation plays a role in alternative splicing in both vertebrates and invertebrates (Flores et al. 2012; Maunakea et al. 2013; Duncan et al. 2014). During transcription of a given gene, not all exons are necessarily included. Alternative splicing is the phenomenon by which transcript variants are created via splicing of exons. In the honeybee, for example, Flores et al. (2012) found higher levels of DNA methylation in exons included in transcripts than those excluded, and a higher prevalence of alternative splicing among methylated than non-methylated genes. Maunakea et al. (2013) have further elucidated a potential mechanism by which methylation enhances inclusion of exons during transcription. Using human cells, these authors found that methylation recruits a binding protein (*MeCP2*) that appears to mediate the inclusion of exons during transcription. The *MeCP2* protein appears to function by maintaining histones in a state conducive to transcription (Maunakea et al. 2013).

Perhaps related to the histone modification and alternative splicing effects induced by methylation described above, gene body methylation may also influence transcriptional elongation, which relates to the efficiency of RNA polymerase (Lorincz et al. 2004; Maunakea et al. 2013). Lorincz *et al.* (2004) found that methylation promotes a closed chromatin structure that reduces the efficiency of transcriptional elongation. Maunakea et al. (2013) speculate that this reduced efficiency could allow for more time during the transcriptional process for the

identification of alternatively spliced exons. However, the information on this process appears to be limited to vertebrates.

Promoter regions precede transcription start sites, and there may be alternative promoter and start sites within a given gene. Gene body methylation has been implicated in the regulation of alternative promoters in different cell and tissue types (Maunakea et al. 2010), which suggests a role in cell differentiation and phenotype. Differences in gene body methylation in relation to tissue type reported elsewhere may reflect this effect (Lokk et al. 2014), but again, this work has been limited to vertebrates.

It is also possible that gene body methylation does not actually direct or influence any molecular processes, and is instead simply a consequence of one or more other processes. For example, Jjingo et al. (2012) observed that it is common for methylation and gene expression levels to exhibit a bell-shaped relationship; both the highest and lowest levels of gene body methylation are associated with the lowest levels of expression, while high levels of expression are associated with intermediate levels of methylation. Based on this phenomenon, Jjingo et al. (2012) presented a model whereby it is expression levels that drive methylation levels, as opposed to the more commonly held belief that methylation influences expression. Their model suggested that low levels of expression limit access of DNA methyltransferases to the DNA due to dense nucleosome packaging, while high levels of expression limit methyltransferase access to DNA due to intense RNA polymerase traffic. Moderate levels of expression allow just enough access to the DNA to allow for methylation to occur (Jjingo et al. 2012). A clear problem with this hypothesis arises when such a relationship between methylation and expression levels is absent (e.g., Lokk et al. 2014).The overwhelming evidence suggests that gene body methylation has a function (or at least, is a cog in the epigenetic machine) and is not simply a byproduct. All

the hypotheses presented above are likely to have some validity and none are mutually exclusive. In fact, all of the hypothesis can fit into a conceptual model where the overarching effect of gene body methylation is to provide some direction to various aspects of transcription. Put as simply as possible, gene body methylation appears to help determine where transcription should start, how fast it should proceed, which exons it should include, and on which parts of the genome it should do all of this. The parts of the genome where this kind of direction appears to be most prevalent are precisely where we would expect them to be: highly conserved genes that perform essential processes with little room for error. Thus, when considered more holistically, the function of gene body methylation may not be as equivocal as it seems.

Corals, sea anemones, and their allies, known as the Anthozoa, are prolific members of the marine invertebrate fauna in many habitats throughout the world. In shallow tropical waters, scleractinian reef corals are often the primary habitat builders and primary producers, thanks largely to their obligate associations with algal endosymbionts that provide a significant proportion of their nutrition. Flexible partnerships with different symbiont phylotypes broadens the phenotypic range of some coral species, which has been likened to an epigenetic effect (Gilbert et al. 2010). However, while the influence of symbionts on host physiology and stress tolerance has been a subject of intense research, attention has shifted more recently to the role of phenotypic plasticity in the host coral. This renewed focus on the host has been spurred by the inability of symbiont diversity to fully account for variation in coral physiology, a consensus that most corals maintain largely stable symbiotic partnerships, and by the 'omics' revolution that has expanded our ability to understand the molecular basis of organismal physiology (Baird et al. 2009, Evans and Hofmann 2012). Corals and their allies are long-lived, sessile organisms that are likely to be particularly reliant on physiological acclimatization and phenotypic plasticity to

cope with environmental variation, which is supported by classical studies on physiological and morphological plasticity as well as more recent research on gene expression plasticity. The underlying basis of this plasticity could lie, at least in part, in epigenetic mechanisms (Hofmann 2017).

Reef-building corals are well known for their inter- and intraspecific variation in growth forms, with morphologies ranging from flattened sheet-like forms to intricate branching colonies. At smaller scales within a colony, corallites can also take a variety of sizes and shapes. Intraspecific variation in coral morphology is particularly intriguing because it has lead to questions about the roles of genetic adaptation vs. phenotypic plasticity, which bears on our understanding of coral ecology, evolution, taxonomy, and persistence in the face of climate change. Variation in coral morphology within a species across different habitats could reflect either selective forces or phenotypic plasticity. Phenotypic plasticity is defined by the degree to which an organism's phenotype can change within its lifetime in response to a change in the environment. Morphological plasticity is likely to be particularly prevalent and adaptive in corals due to their immobility, longevity, modularity, and indeterminate growth (Jackson & Coates 1986; Sebens 1987; Todd 2008). In his review of 26 studies that have tested for morphological plasticity in approximately 20 different species of reef corals, Todd (2008) identified only two studies that found no evidence for plasticity, and another two that emphasized genetic effects over plasticity. The vast majority of studies have therefore found evidence for some level of plasticity either at the colony or corallite level. A number of different environmental factors have been implicated as drivers of morphological plasticity in corals, including gravity, water motion, light, particle capture, competition, and sedimentation (Todd 2008). Paramount among these factors are light and water motion (Todd 2008). Among the 26 studies on phenotypic plasticity

reviewed by Todd (2008), light was the overwhelmingly most cited factor, having been implicated in 11 of the 16 studies that identified an environmental factor. It should be noted, however, that many of these studies were conducted in the field, while for others, the effect of light was specifically tested. Nonetheless, reef corals are photosynthetic organisms (by way of their symbionts) whose distributions are limited to the photic zone, so it should not be surprising that light plays a dominant role in their morphology. Indeed, light plays a key role in structuring communities of the symbiotic dinoflagellates that reef corals rely on for up to 90% or more of their energy (Muscatine 1990), with evidence that coral depth zonation is strongly linked to that of their symbionts (Bongaerts et al. 2015). Moreover, symbiont photosynthesis is closely tied to coral calcification (Gattuso et al. 1999). As would therefore be expected, coral morphologies exhibit plasticity to deal with a range of light levels. Plastic responses to decreasing light levels (or increasing depths) include flattening of hemispherical or plating species and increased branch spacing of branching species, strategies which increase surface area available for light capture and reduce self-shading (Todd 2008). For example, Caribbean *Orbicella* spp. are massive species in shallow waters that adopt a flattened plating form on deeper reefs that is attributed to reduced light (Dustan 1975). In addition to changes in gross morphology, corallites often become smaller with less vertical depression as light decreases in order to reduce self-shading (Todd 2008).

Water motion is perhaps the second most important factor driving morphological plasticity in corals (Todd 2008). As a consequence of living in the photic zone, corals can experience a range of water motion regimes, from wave-swept intertidal and shallow subtidal zones, to mesophotic reefs where wave surge is imperceptible and water flow occurs in the form of tidal and coastal currents varying in intensity. Withstanding drag forces and potential breakage is critical for corals living in wave exposed areas, while overcoming boundary layer effects

becomes important in areas with less water motion (Wainwright & Koehl 1976). Scoffin et al. (1992) demonstrated a relationship between higher hydrodynamic forces and increased skeletal density in *Porites lutea* on Thai Reefs. Because corals are sessile and do not actively pump water, water flow is essential for gas exchange and particle delivery (Wainwright & Koehl 1976). In areas of moderate or low water motion, corals may adopt strategies that help overcome boundary layer effects and enhance prey/particle capture and gas exchange. For example, the Caribbean coral *Madracis auretenra* (= *M. mirabilis*) exhibits an inverse relationship between branch spacing and prevailing water flow speeds that appears to be an adaptive plastic response to facilitate mass transfer and feeding (Sebens et al. 1997; Bruno & Edmunds 1998).

Thermal tolerance thresholds of tropical reef corals are typically only 1° to 2°C above mean summer maximum temperatures (Hoegh-Guldberg 1999). The prevailing theory for the thermal sensitivity of reef corals is that oxidative stress originating from both the coral host cells and their symbiotic dinoflagellates leads to breakdown of the symbiosis (Weis 2008; Davy et al. 2012). Although research has historically focused on the sensitivity of the light and dark reactions of *Symbiodinium* photosynthesis to excess heat and light, there is also evidence that the host cells may be the initial site of thermal damage and oxidative stress (e.g., Hawkins et al. 2014). Regardless of the source of the stress, the outcome is bleaching via host cell apoptosis and mass expulsion of symbionts (Weis 2008). Depending on the severity of the stress, bleaching can lead to death if the coral does not regain a healthy population of symbionts.

If coral thermal tolerance thresholds do not keep pace with increasing global temperatures, they may not persist (Hoegh-Guldberg 1999). Yet, it has long been clear that thermal tolerances vary within a species, and that regional and even small-scale variability in thermal tolerance typically matches the historical thermal regime, suggesting evolutionary

processes or phenotypic plasticity (Oliver & Palumbi 2011; Riegl et al. 2011; Howells et al. 2013; Palumbi et al. 2014). The Arabian Gulf, for example, is the hottest of the world's seas, and thermal tolerance thresholds of corals there exceed those of Caribbean and Indo-Pacific corals by 1° to 3°C (Riegl et al. 2011). Variability in thermal tolerance also occurs over much smaller scales, such as in the case of corals in Samoan backreef tidepools, where individuals occurring in warmer, more variable pools exhibit higher thermal tolerance than those in cooler, less variable pools (Barshis et al. 2010; Oliver & Palumbi 2011; Palumbi et al. 2014). In both the Samoan tidepools and the Arabian Gulf, coral thermal thresholds top out at approximately 35°C for short-term exposures, while typical thermal tolerances in most of the rest of the world are in the range of 30-32°C (Riegl et al. 2011; Oliver & Palumbi 2011).

In the Samoan backreef system, Palumbi et al. (2014) were able to parse out the contribution of plastic vs. fixed (putatively adaptive) effects on thermal tolerance of the coral *Acropora hyacinthus*, finding evidence for both processes in roughly equal proportions. The maintenance of putatively adaptive effects in this system is remarkable given the close proximity of the pools and the high gene flow between them (Oliver & Palumbi 2011). The mechanism behind the fixed effects is at least partially attributed to "frontloading", or constitutively higher levels of expression, of certain genes associated with thermal tolerance (Barshis et al. 2013; Palumbi et al. 2014). Palumbi et al. (2014) provide one of the first true tests of local adaptation to thermal regimes in a coral, but the precise mechanism could be either genetic or epigenetic, and its potential heritability was not established. More recently, another study was able to demonstrate heritability of thermal tolerance across latitudes on the Great Barrier Reef in *Acropora millepora* (Dixon et al. 2015). These authors identified genomic regions associated with selection for thermal tolerance, but acknowledged that epigenetics could also be involved

on the basis of strong maternal effects. These types of studies evaluating adaptive potential and heritability of thermal tolerance could significantly change the dire predictions of modeling studies that suggest a limited scope for continued coral thermal tolerance via phenotypic plasticity alone (e.g., Ainsworth et al. 2016).

The debate surrounding coral thermal tolerance would be incomplete without considering the symbiotic dinoflagellates that play a major role in the overall physiology of the coral holobiont. Symbiotic dinoflagellates are genetically and physiologically diverse, exhibiting a range of thermal tolerances (Oppen et al. 2009). There are two primary means by which coral thermal tolerance can be augmented by their symbionts: 1) a change in the coral's dominant symbiont type, and 2) local adaptation of the symbionts themselves. There are some reports of corals gaining enhanced thermal tolerance as a result of changes in their dominant symbiont type (Berkelmans & van Oppen 2006; Silverstein et al. 2015), but coral symbioses tend to be stable, and not all corals appear to be capable of changing their symbionts (Thornhill et al. 2006). Furthermore, hosting a putatively thermally tolerant symbiont does not necessarily convey thermal tolerance. For example, in the work by Palumbi et al. (2014), corals inhabiting warmer pools tended to host more putatively thermally tolerant clade D symbionts, but this explained only a small fraction of the variance in thermal tolerance. Symbiont local adaptation is an alternative mechanism of symbiont-mediated thermal tolerance which has been demonstrated among corals on the Great Barrier Reef, as well as in the Arabian Gulf (Howells et al. 2011; Hume et al. 2015). Howells et al. (2011) demonstrated that *Symbiodinium* type C1 from two locations on the GBR exhibited significantly different thermal tolerance and physiological performance that was reflected in differential growth of juvenile corals.

The aim of the research presented herein is to examine DNA methylation in corals and their allies with respect to its potential involvement in environmental change responses, phenotypic variation, and transcriptional regulation. Chapter 1 explores patterns of evolutionary-scale DNA methylation in corals, using CpG depletion analysis to estimate methylation levels in the transcriptomes and environmental response genes of several coral species. Chapter 2 compares genetic and epigenetic variation in Atlantic branching corals of the genus *Porites*, testing the hypothesis that DNA methylation patterns reflect phenotypic differences among morphotypes. Chapter 3 asks whether DNA methylation in the mounding coral *Porites astreoides* exhibits plasticity in response to experimental transplantation. Finally, Chapter 4 is a functional genomics study testing the hypothesis that differential DNA methylation is associated with symbiotic state and with alternative mRNA splicing.

Literature cited

Ainsworth TD, Heron SF, Ortiz JC et al. (2016) Climate change disables coral bleaching protection on the Great Barrier Reef. Science, 352, 338–342.

Baird AH, Bhagooli R, Ralph PJ, Takahashi S (2009) Coral bleaching: the role of the host. Trends Ecol Evol 24, 16–20.

Barshis DJ, Ladner JT, Oliver TA et al. (2013) Genomic basis for coral resilience to climate change. Proc Natl Acad Sci U S A, 110, 1387–1392.

Barshis DJ, Stillman JH, Gates RD et al. (2010) Protein expression and genetic structure of the coral *Porites lobata* in an environmentally extreme Samoan back reef: does host genotype limit phenotypic plasticity? Mol Ecol, 19, 1705–1720.

Berkelmans R, van Oppen MJH (2006) The role of zooxanthellae in the thermal tolerance of corals: a "nugget of hope" for coral reefs in an era of climate change. Proc Biol Sci, 273, 2305–2312.

Bird AP (1995) Gene number, noise reduction and biological complexity. Trends Genet, 11, 94–100.

Bongaerts P, Carmichael M, Hay KB et al. (2015) Prevalent endosymbiont zonation shapes the depth distributions of scleractinian coral species. R Soc Open Sci, 2, 140297.

Bruno J, Edmunds P (1998) Metabolic consequences of phenotypic plasticity in the coral *Madracis mirabilis* (Duchassaing and Michelotti): the effect of morphology and water flow on aggregate respiration. Journal of experimental marine biology and ecology.

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

Dimond JL, Roberts SB (2016) Germline DNA methylation in reef corals: patterns and potential roles in response to environmental change. Mol Ecol, 25, 1895–1904.

Dixon G, Davies S, Aglyamova G, Meyer E, et al. (2015) Genomic determinants of coral heat tolerance across latitudes. Science

Dixon G, Bay L, Matz M (2016) Evolutionary consequences of DNA methylation in a basal metazoan. Molecular biology and evolution.

Duncan EJ, Gluckman PD, Dearden PK (2014) Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype? J Exp Zool B Mol Dev Evol, 322, 208–220.

Dustan P (1975) Growth and form in the reef-building coral *Montastrea annularis*. Marine Biology.

Evans TG , Hofmann GE (2012) Defining the limits of physiological plasticity: how gene expression can assess and predict the consequences of ocean change. Phil Trans R Soc B, 367, 1733–1745.

Felsenfeld G (2014) A brief history of epigenetics. Cold Spring Harb Perspect Biol, 6.

Flores K, Wolschin F, Corneveaux JJ et al. (2012) Genome-wide association between DNA methylation and alternative splicing in an invertebrate. BMC Genomics, 13, 480.

Gattuso J, Allemand D, Frankignoulle M (1999) Photosynthesis and calcification at cellular, organismal and community levels in coral reefs: a review on interactions and control by carbonate chemistry. American zoologist.

Gilbert SF, McDonald E, Boyle N, Buttino N, Gyi L, Mai M, Prakash N, Robinson J (2010) Symbiosis as a source of selectable epigenetic variation: taking the heat for the big guy. Phil Trans R Soc B, 365, 671–678.

Hawkins T, Krueger T, Becker S, Fisher P, Davy S (2014) Differential nitric oxide synthesis and host apoptotic events correlate with bleaching susceptibility in reef corals. Coral Reefs.

Hoegh-Guldberg O (1999) Climate Change, coral bleaching and the future of the world' s coral reefs. Mar Freshwater Res, 50, 839–866.

Hofmann GE (2017) Ecological epigenetics in marine metazoans. Front Mar Sci 4.

Howells EJ, Beltran VH, Larsen NW et al. (2011) Coral thermal tolerance shaped by local adaptation of photosymbionts. Nat Clim Chang, 2, 116–120.

Howells E, Berkelmans R, van Oppen M, Willis B, Bay L (2013) Historical thermal regimes define limits to coral acclimatization. Ecology, 94, 1078–1088.

Huh I, Zeng J, Park T, Yi SV (2013) DNA methylation and transcriptional noise. Epigenetics Chromatin, 6, 9.

Hume BCC, D'Angelo C, Smith EG et al. (2015) *Symbiodinium thermophilum* sp. nov., a thermotolerant symbiotic alga prevalent in corals of the world's hottest sea, the Persian/Arabian Gulf. Sci Rep, 5, 8562.

Jackson JBC, Coates AG (1986) Life cycles and evolution of clonal (modular) animals. Philosophical Transactions of the Royal Society B: Biological Sciences, 313, 7–22.

Jjingo D, Conley AB, Yi SV, Lunyak VV, Jordan IK (2012) On the presence and role of human gene-body DNA methylation. Oncotarget, 3, 462–474.

Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet, 13, 484–492.

Lokk K, Modhukur V, Rajashekar B et al. (2014) DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. Genome Biol, 15, r54.

Lorincz MC, Dickerson DR, Schmitt M, Groudine M (2004) Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat Struct Mol Biol, 11, 1068–1075.

Maunakea AK, Chepelev I, Cui K, Zhao K (2013) Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res, 23, 1256–1269.

Maunakea AK, Nagarajan RP, Bilenky M et al. (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature, 466, 253–257.

Miska EA, Ferguson-Smith AC (2016) Transgenerational inheritance: Models and mechanisms of non-DNA sequence-based inheritance. Science, 354, 59–63.

Muscatine L (1990) The role of symbiotic algae in carbon and energy flux in reef corals. Ecosystems of the world.

Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, Maldotti M, Anselmi F, Oliviero S (201&) Intragenic DNA methylation prevents spurious transcription initiation. Nature, 543,72.

Oliver T, Palumbi S (2011) Do fluctuating temperature environments elevate coral thermal tolerance? Coral Reefs, 1–12.

Oppen M van, Baker A, Coffroth M, Willis B (2009) Bleaching resistance and the role of algal endosymbionts. Coral bleaching, 83–102.

Palumbi SR, Barshis DJ, Traylor-Knowles N, Bay RA (2014) Mechanisms of reef coral resistance to future climate change. Science, 344, 895–898.

Riegl BM, Purkis SJ, Al-Cibahy AS, Abdel-Moati MA, Hoegh-Guldberg O (2011) Present limits to heat-adaptability in corals and population-level responses to climate extremes. PLoS ONE, 6, e24802.

Roberts SB, Gavery MR (2012) Is there a relationship between DNA methylation and phenotypic plasticity in invertebrates? Front Physiol, 2, 116.

Sarda S, Zeng J, Hunt BG, Yi SV (2012) The evolution of invertebrate gene body methylation. Mol Biol Evol, 29, 1907–1916.Scoffin TP, Tudhope AW, Brown BE, Chansang H, Cheeney RF (1992) Patterns and possible environmental controls of skeletogenesis of *Porites lutea*, South Thailand. Coral Reefs, 11, 1–11.

Sebens KP (1987) The ecology of indeterminate growth in animals. Annu Rev Ecol Syst, 18, 371–407.

Sebens K, Witting J, Helmuth B (1997) Effects of water flow and branch spacing on particle capture by the reef coral *Madracis mirabilis* (Duchassaing and Michelotti). Journal of experimental marine biology and ecology

Silverstein RN, Cunning R, Baker AC (2015) Change in algal symbiont communities after bleaching, not prior heat exposure, increases heat tolerance of reef corals. Glob Chang Biol, 21, 236–249.

Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet, 9, 465–476.

Thornhill DJ, LaJeunesse TC, Kemp DW, Fitt WK, Schmidt GW (2006) Multi-year, seasonal genotypic surveys of coral-algal symbioses reveal prevalent stability or post-bleaching reversion. Mar Biol, 148, 711–722.

Todd PA (2008) Morphological plasticity in scleractinian corals. Biol Rev Camb Philos Soc, 83, 315–337.Waddington C (2012) The epigenotype. International journal of epidemiology.Wainwright SA, Koehl MAR (1976) The nature of flow and the reaction of benthic cnidaria to it. In: Coelenterate ecology and behavior (ed Mackie GO), pp. 5–21. Springer US, Boston, MA.

Weis VM (2008) Cellular mechanisms of Cnidarian bleaching: stress causes the collapse of symbiosis. J Exp Biol, 211, 3059–3066.Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science, 328, 916–919.

Chapter 1. GERMLINE DNA METHYLATION IN REEF CORALS: PATTERNS AND POTENTIAL ROLES IN RESPONSE TO ENVIRONMENTAL CHANGE

A version of this chapter was published as: Dimond, JL & Roberts SB (2016) Germline DNA methylation in reef corals: patterns and potential roles in response to environmental change. Molecular Ecology, 25(8), 1895-1904.

1.1 Abstract

DNA methylation is an epigenetic mark that plays an inadequately understood role in gene regulation, particularly in non-model species. Because it can be influenced by the environment and potentially transferred to subsequent generations, DNA methylation may contribute to the ability of organisms to acclimatize and adapt to environmental change. We evaluated the distribution of gene body methylation in reef-building corals, a group of organisms facing significant environmental threats. Gene body methylation in six species of corals was inferred from in silico transcriptome analysis of CpG O/E, an estimate of germline DNA methylation that is highly correlated with patterns of methylation enrichment. Consistent with what has been documented in most other invertebrates, all corals exhibited bimodal distributions of germline methylation suggestive of distinct fractions of genes with high and low levels of methylation. The hypermethylated fractions were enriched with genes with housekeeping functions, while genes with inducible functions were highly represented in the hypomethylated fractions. In three of the coral species, we found that genes differentially expressed in response to thermal stress and ocean acidification exhibited significantly lower levels of methylation. These results support a link between gene body hypomethylation and transcriptional plasticity that may point to a role of DNA methylation in the response of corals to environmental change.

1.2 INTRODUCTION

As human influence on the planet expands, many organisms must acclimatize and adapt to rapid environmental change. Phenotypic plasticity facilitates a more rapid response to environmental change than is possible through natural selection, and will likely be critical to the persistence of many species (Charmantier et al. 2008, Chevin et al. 2010). Phenotypic change often involves modifications in gene expression. Epigenetic mechanisms, involving alterations to the genome that do not affect the underlying DNA sequence, are increasingly recognized as some of the principal mediators of gene expression (Duncan et al. 2014). The most researched and best understood epigenetic process is DNA methylation, which most commonly involves the addition of a methyl group to a cytosine in a CpG dinucleotide pair. The role of DNA methylation is best understood in mammals, where methylation in promoter regions has a repressive effect on gene expression (Jones and Takai 2001). In plants and invertebrates, methylation of gene bodies prevails, and is thought to be the ancestral pattern (Zemach et al. 2010). Gene body methylation appears to have a range of functions, including regulating alternative splicing, repressing intragenic promoter activity, and reducing the efficiency of transcriptional elongation (Duncan et al. 2014). Methylation of gene bodies also varies according to gene function, and studies on invertebrates indicate that highly conserved genes with housekeeping functions tend to be more heavily methylated than those with inducible functions (Roberts and Gavery 2012, Sarda et al. 2012, Dixon et al. 2014, Gavery and Roberts 2014). This has led to speculation that gene body methylation may promote predictable expression of essential genes for basic biological processes, while an absence of methylation could allow for

stochastic transcriptional opportunities in genes involved in phenotypic plasticity (Roberts and Gavery 2012, Dixon et al. 2014, Gavery and Roberts 2014).

Direct relationships between DNA methylation and phenotypic plasticity are increasingly being established. Some examples include caste structure in honeybees and ants (Kucharski et al. 2008, Bonasio et al. 2012), expression of the agouti gene in mice (Wolff et al. 1998), and the influence of prenatal maternal mood on newborn stress levels in humans (Oberlander et al. 2008). In many cases, changes in methylation patterns can be attributed to external cues such as temperature, stress, or nutrition. A prime example is the honeybee *Apis mellifera*, where larval consumption of royal jelly induces changes in methylation that ultimately determine the developmental fate of an individual into a queen or a worker (Kucharski et al. 2008). Thus, DNA methylation has been established as a key link between environment and phenotype.

Reef-building corals, the organisms that form the trophic and structural foundation of coral reef ecosystems, are known to display a significant degree of phenotypic plasticity (Todd 2008, Forsman et al. 2009, Granados-Cifuentes et al. 2013). As long-lived, sessile organisms, corals are thought to be particularly reliant on phenotypic plasticity to cope with environmental heterogeneity, because they must be able to withstand whatever nature imposes on them over long periods of time (Bruno and Edmunds 1997). As phenotypically flexible as they may be, corals' longevity and immobility may also contribute to their vulnerability in a changing environment. Reef corals worldwide are experiencing severe declines due to a variety of anthropogenic effects, including climate change, ocean acidification, and a host of local stressors (Hoegh-Guldberg et al. 2007). This has raised doubt concerning the ability of corals to survive coming decades. Yet there are also signs that, at least in some cases, corals possess sufficient resiliency to overcome their numerous challenges (Palumbi et al. 2014). Recent studies on gene

expression variation, for example, support the view that phenotypic plasticity in corals is robust and may provide resilience in the face of ocean warming (Barshis et al. 2013, Granados-Cifuentes et al. 2013, Palumbi et al. 2014). However, the underlying basis of gene expression variation, and indeed phenotypic plasticity, remain largely unknown.

Evaluation of epigenetic processes therefore represents a logical next step in understanding coral gene expression and phenotypic variation. While recent annotation of the *Acropora digitifera* genome revealed a broad repertoire of genes involved in DNA methylation and other epigenetic processes (Dunlap et al. 2013), to date, only one study has investigated possible roles of epigenetic processes in corals (Dixon et al. 2014). Germline DNA methylation patterns in the transcriptome of *Acropora millepora* corroborated findings reported in studies of other invertebrate species (Dixon et al. 2014). Most interestingly, genes that were differentially expressed in response to a common garden transplantation experiment were among the genes exhibiting lower levels of germline methylation (Dixon et al. 2014), suggesting a link between hypomethylation and gene expression plasticity.

Coral gene expression studies continue to expand, providing rich datasets to further probe the relationship between DNA methylation and gene function. In this study, we performed a comprehensive evaluation of germline methylation patterns in reef corals by examining the transcriptomes of six scleractinian coral species. Germline methylation levels in these data were inferred based on the hypermutability of methylated cytosines, which leads to a reduction in CpG dinucleotides over evolutionary time (Sved and Bird 1990). These data were then matched with gene ontology information, permitting evaluation of methylation patterns associated with broad categories of biological processes. Lastly, in three of the six species, we evaluated germline methylation patterns in genes involved in response to thermal stress and ocean acidification.

1.3 METHODS

Transcriptome data sources

The transcriptomes of six scleractinian coral species were evaluated to determine germline methylation patterns in relation to gene function and activity. Species examined included *Acropora hyacinthus*, *A. millepora*, *A. palmata*, *Pocillopora damicornis*, *Porites astreoides*, and *Stylophora pistillata* (Table 1.1).

Table 1.1. Transcriptomes analyzed in this study.

Organism	Life history stage	Method	No. contigs	Reference
Acropora hyacinthus	Adult	sequencing	33,496	Barshis et al. (2013)
Acropora millepora	Embryo to adult	sequencing	52,963	Moya et al. (2012)
Acropora palmata	Larval	sequencing	88,020	Polato et al. (2011)
Pocillopora damicornis	Adult	sequencing	72,890	Vidal-Dupiol et al. (2013)
Porites astreoides	Adult	sequencing	30,740	Kenkel et al. (2013)
Stylophora pistillata	Adult	sequencing	15,052	Karako-Lampert et al. (2014)

These transcriptomes reflect a range of life history stages. Some transcriptomes were developed from life history stages that had not yet been infected with symbiotic dinoflagellates (*Symbiodinium* spp.), while others used bioinformatic techniques to filter out putative *Symbiodinium* sequences. However, two of the transcriptomes (*P. damicornis* and *S. pistillata*) were developed from adult corals and did not remove putative symbiont sequences. We therefore applied a filtering step to these transcriptomes by comparing them to *Symbiodinium* clade A and B transcriptomes from Bayer et al. (2012) using blastn (version 2.2.29). An evalue threshold of 10-5 was used for these queries, and all matched sequences were removed from further analyses.

Details of the blastn query and filtering procedures are provided online (https://github.com/jldimond/Coral-CpG-ratio-MS).

Differentially expressed gene datasets

In addition to analyzing whole transcriptomes, we also examined genes differentially expressed in response to environmental stressors for the three acroporid species. For *A. hyacinthus* and *A. millepora* these gene sets were derived from the same studies that developed the reference transcriptomes (Barshis et al. 2013, Moya et al. 2012), and for *A. palmata* differentially expressed genes sets were reported in Polato et al. (2013) (Table 1.2).

Table 1.2. Differentially expressed gene sets examined.

Organism	Life history stage	Method	No. contigs	Environmental factor	Reference
Acropora hyacinthus	Adult	sequencing	484	Thermal stress	Barshis et al. (2013)
Acropora millepora	Juvenile	sequencing	234	Ocean acidification	Moya et al. (2012)
Acropora palmata	Larval	microarray	2002	Thermal stress	Polato et al. (2013)

Annotation

To maintain consistency in comparing datasets, all transcriptomes and differentially expressed gene sets were compared to the UniProt/Swiss-Prot protein database (version 2/17/2015) using Blastx (version 2.2.29) with an evalue threshold of 10-5. To further annotate genes with functional categories, corresponding Gene Ontology Slim (GOSlim) biological

process were identified. Details of annotation are provided in jupyter notebooks in an online repository (https://github.com/jldimond/Coral-CpG-ratio-MS).

Predicted germline methylation

Germline methylation levels were inferred based on the hypermutability of methylated cytosines, which tend towards conversion to thymines over evolutionary time. This results in a reduction in CpG dinucleotides, meaning that heavily methylated genomic regions are associated with reduced numbers of CpGs. Thus, methylation patterns that have been inherited through the germline over evolutionary time can be estimated using the ratio of observed to expected CpG, known as CpG O/E. Germline DNA methylation estimated by analysis of CpG O/E is highly correlated with direct assays of methylation (Suzuki et al. 2007, Sarda et al. 2012, Gavery and Roberts 2013). CpG O/E was defined as:

$$\text{CpG O/E} = (\text{number of CpG} / \text{number of C} \times \text{number of G}) \times (l2/l-1)$$

where l is the number of nucleotides in the contig. Only annotated sequences were used for calculation of CpG O/E to increase the likelihood that sequences were oriented in the 5’ to 3’ direction. For subsequent analyses, we set minimum and maximum limits for CpG O/E at 0.001 and 1.5, respectively, to exclude outliers. Details of germline methylation prediction methods are provided in jupyter notebooks (https://github.com/jldimond/Coral-CpG-ratio-MS/).

Statistical analyses

Transcriptome CpG O/E patterns were fitted with the normalmixEM function in the mixtools package in the R statistical platform. Mixture models were evaluated against the null single component model by comparison of log-likelihood statistics. High and low CpG O/E components were delineated in mixture models using the intersection point of component density curves. For each GOSlim term, enrichment in the high and low CpG O/E components identified in mixture models was evaluated with Fisher's exact test. Whole transcriptome and differentially expressed gene CpG O/E distributions were compared with the Kolmogorov–Smirnov test. To compare representation of GOSlim terms in the differentially expressed genes relative to whole transcriptomes, the relative abundance of each GOSlim term in each dataset was calculated as a percentage, and these values for whole transcriptomes were subtracted from those for differentially expressed genes. Details and specific code used are available online (https://github.com/jldimond/Coral-CpG-ratio-MS/).

1.4 RESULTS

Annotation

For *P. damicornis* and *S. pistillata* transcriptomes, 2892 and 138 putative *Symbiodinium* sequences were removed from the transcriptomes, respectively. Comparisons of the six coral transcriptomes to the UniProt/Swiss-Prot database resulted in annotation of 26% to 47% of the contigs in each transcriptome (Table 1.3).

Table 1.3. Transcriptome annotation results.

Organism	No. contigs	Contigs annotated (UniProt/Swiss-Prot)	Contigs annotated (GOslim)
Acropora hyacinthus	33,496	11,593	9,935
Acropora millepora	52,963	21,026	17,274

Acropora palmata	88,020	35,303	29,450
Pocillopora damicornis	72,890	19,133	16,150
Porites astreoides	30,740	13,788	11,621
Stylophora pistillata	15,052	7,061	5,812

Predicted gene methylation

Whole transcriptome patterns of predicted germline DNA methylation were similar for all coral species, suggesting bimodal distributions of CpG O/E with relatively large high-CpG O/E components (Figure 1.1).

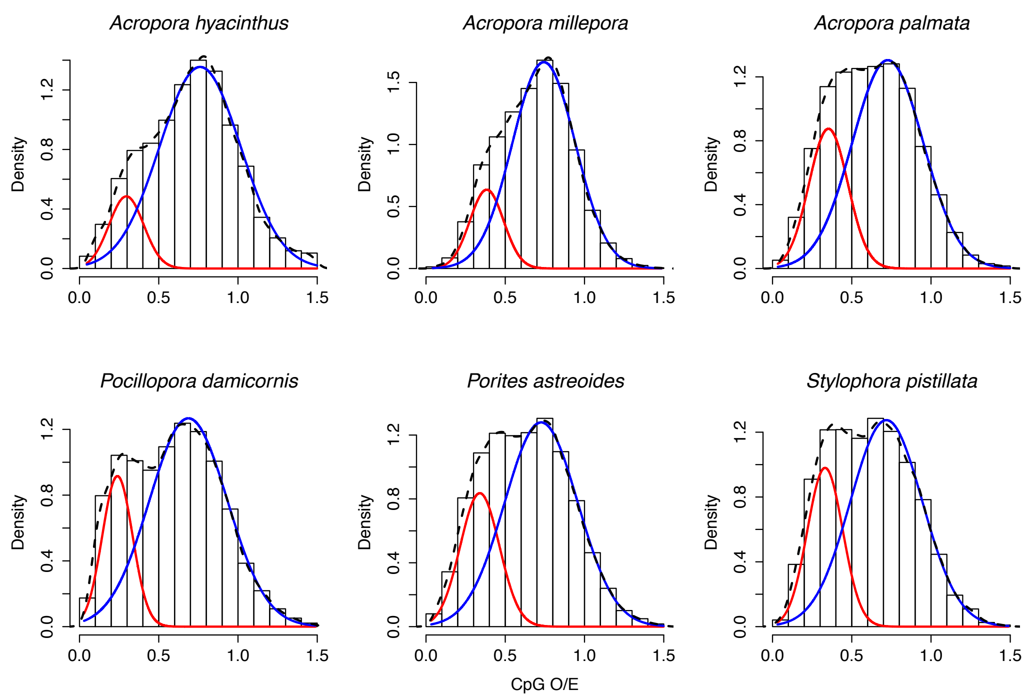

Figure 1.1. Transcriptome-wide CpG O/E in the six coral species. The component density curves of two-component mixture models are superimposed over histograms and density curves.

This was confirmed by mixture model analyses indicating that a 2-component Gaussian model provided better fit than a single-component model in all cases (Table 1.4). While there was some variability in models between species, all models were characterized by a low-CpG

O/E component with a mean of 0.24 to 0.38 and weighted at 14-28% of the distribution. Conversely, high-CpG O/E components had means of 0.69 to 0.75 and weights of 72-86%.

Table 1.4. Results of mixture model analyses of CpG O/E in the six coral transcriptomes. Model statistics are reported for a 2-component mixture model, which provided better fit than a single component model as indicated by log-liklihood statistics. Numbers separated by commas represent statistics for the first and second components, respectively.

Organism	lambda	mu	sigma	log-likelihood (k = 1)	log-likelihood (k = 2)
Acropora hyacinthus	0.14, 0.86	0.30, 0.75	0.11, 0.26	-1950	-1771
Acropora millepora	0.17, 0.83	0.38, 0.74	0.11, 0.20	1023	1307
Acropora palmata	0.27, 0.73	0.35, 0.73	0.12, 0.22	-2664	-1824
Pocillopora damicornis	0.22, 0.78	0.24, 0.69	0.10, 0.25	-3324	-2464
Porites astreoides	0.26, 0.74	0.34, 0.73	0.12, 0.23	-1450	-1115
Stylophora pistillata	0.28, 0.72	0.33, 0.72	0.11, 0.23	-660	-413

When CpG O/E was evaluated according to gene function, we observed consistent patterns among species in relation to broad classes of biological processes. For all six species, the top four biological processes with the highest mean CpG O/E were cell-cell signaling, cell adhesion, signal transduction, and developmental processes (Figure 1.2).

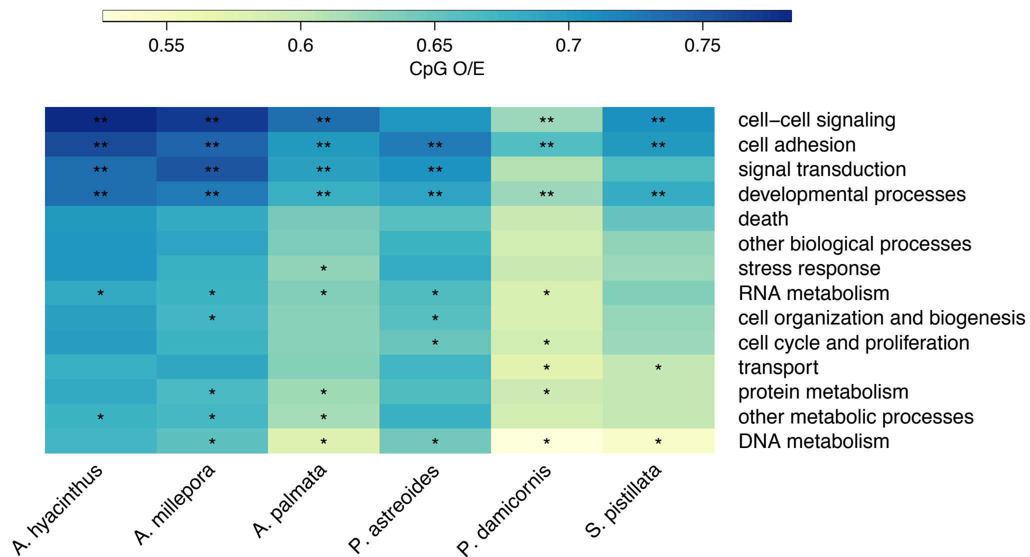

Figure 1.2. Patterns of CpG O/E in relation to gene function (GOSlim biological processes terms) in the six coral species. Mean CpG O/E for each term is shown, with terms ordered by the grand mean across species. Asterisks denote results of significant Fisher's exact tests ($\alpha = 0.05$); double asterisks indicate terms that were significantly enriched in the high-CpG O/E component, and single asterisks indicate terms significantly enriched in the low-CpG O/E component.

Low-ranked biological processes with the lowest mean CpG O/E were more variable between species. However, DNA metabolism was consistently ranked lowest, while protein metabolism and other metabolic processes were also typically among the lowest categories in terms of CpG O/E. Relatively high and low ranked biological processes were also more likely to be significantly enriched in the high and low CpG O/E components identified in the mixture model.

Relationship between gene expression levels and CpG O/E

In *Acropora hyacinthus*, *A. millepora*, and *A. palmata,* we observed similar patterns of gene expression with respect to CpG O/E (Figure 1.3). Gene expression in all species was characterized by bell-shaped distributions, with the highest levels of gene expression occurring at

intermediate levels CpG O/E. These curves were skewed to the left in *A. hyacinthus* and to the right in *A. millepora* and *A. palmata.*

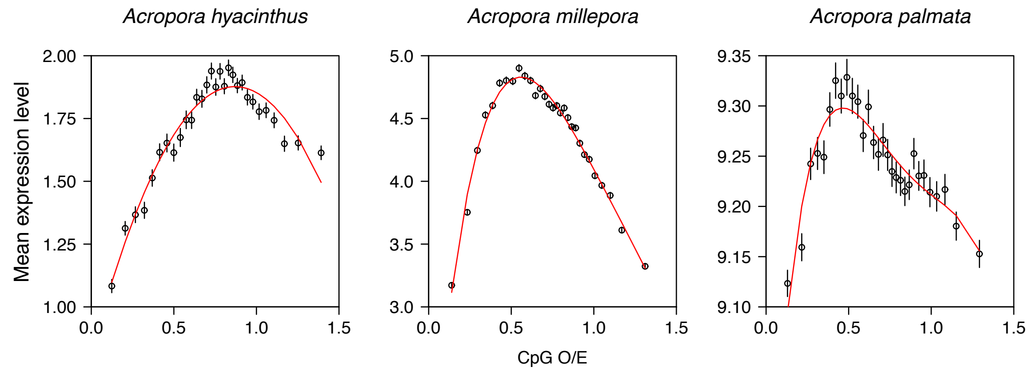

Figure 1.3. CpG O/E vs. mean gene expression level. Means (±SE) for 30 bins of equal sample size are shown. Data were fitted with 2nd (A. *hyacinthus*) and 4th (A. *millepora* and A. *palmata*) order polynomial functions (red lines).

Methylation of differentially expressed genes

In *Acropora hyacinthus*, *A. millepora*, and *A. palmata*, genes expressed differentially in response to environmental stress showed distinct CpG O/E distributions from those of the whole transcriptomes (Figure 1.4, upper panel).

In all cases, mean CpG O/E of differentially expressed genes was higher than that of the whole transcriptome. This was especially true for differentially expressed genes in response to thermal stress in *A. hyacinthus* and *A. palmata* (both $p < 0.001$), but CpG O/E distributions of ocean acidification differentially expressed genes and the whole transcriptome of *A. millepora* were also significantly different ($p = 0.005$).

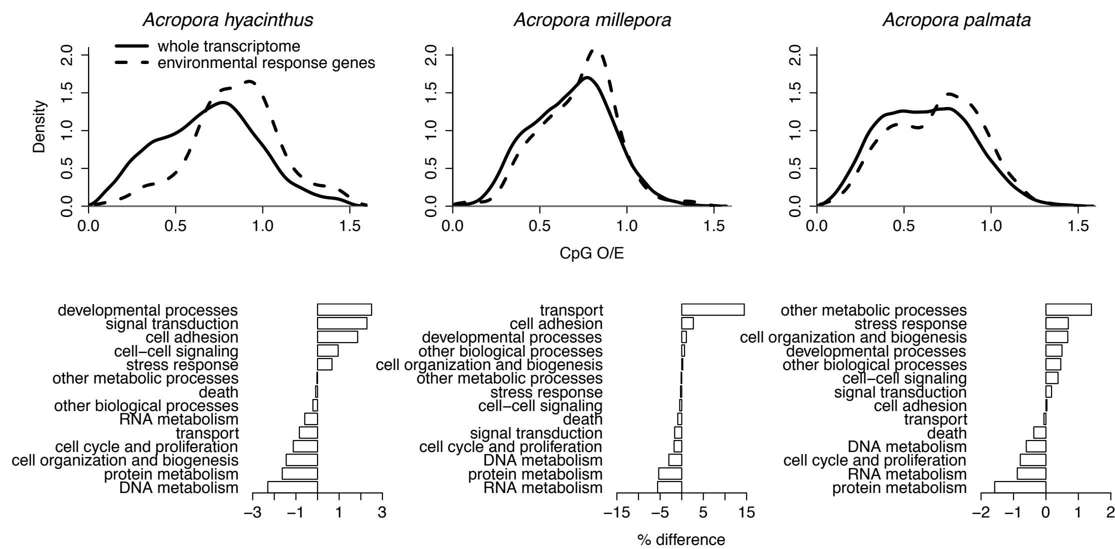

Figure 1.4. Comparison of transcriptome-wide and differentially expressed gene CpG O/E in the acroporid corals. Upper panel: density curves of whole transcriptomes (solid lines) and differentially expressed genes (dashed lines). Lower panel: representation of different gene classes in differentially expressed genes relative to whole transcriptome levels. These values were obtained by determining the percent representation of each of the 14 GOSlim classes in whole transcriptomes and differentially expressed genes, then taking the difference between the two.

We also evaluated the contribution of different biological processes (GOslim terms) to differentially expressed gene profiles by comparing their abundances in differentially expressed genes relative to whole transcriptomes (Figure 1.4, lower panel). In *A. hyacinthus*, differentially expressed genes were overrepresented by biological processes associated with the high-CpG O/E components identified in the mixture model, and underrepresented by biological processes associated with the low CpG O/E component. In *A. millepora*, differentially expressed genes were characterized by a striking increase in transport processes, which were not significantly enriched in either the high or low CpG O/E components of the mixture model. To a lesser extent, cell adhesion and developmental processes were overrepresented in differentially expressed

genes, and these processes were enriched in the high CpG O/E component. Biological processes underrepresented in differentially expressed genes tended to be associated with the low CpG O/E component. Finally, in *A. palmata*, a mixture of processes enriched in both the high and low CpG O/E components were overrepresented in the differentially expressed genes. Processes underrepresented in differentially expressed genes were largely associated with the low CpG O/E component.

1.5 Discussion

Research on DNA methylation to date has revealed that its extent and function varies considerably among organisms. While several studies have compared DNA methylation patterns between distantly related taxa (Tweedie et al. 1997, Zemach et al. 2010, Sarda et al. 2012), our analysis focused on a single taxonomic group, the reef-building scleractinian corals. Within this group, we took a comparative approach of six species to provide a comprehensive evaluation of germline methylation patterns. In three of these species, representing three independent studies, we found that genes expressed differentially in response to environmental stressors exhibited significantly lower levels of methylation. This work adds to a small but growing body of evidence supporting an association between hypomethylation and gene expression plasticity (Roberts and Gavery 2012, Dixon et al. 2014).

As in most other invertebrate taxa surveyed (Gavery and Roberts 2010, Sarda et al. 2012), we observed patterns of CpG O/E that were indicative of bimodal distributions in all of the coral transcriptomes. All distributions were dominated by a relatively high CpG O/E fraction, suggesting that the majority of genes in reef corals experience relatively low levels of methylation. A similar pattern was observed in a genome-wide analysis of the sea anemone *Nematostella vectensis* (Zemach et al. 2010, Sarda et al. 2012), and in an analysis of exons in *A.*

millepora (Dixon et al. 2014). In contrast, CpG O/E profiles in other invertebrates, such as the oyster *Crassostrea gigas* and the sea squirt *Ciona intestinalis*, suggest larger low-CpG O/E fractions (Gavery and Roberts 2010, Sarda et al. 2012). This is consistent with higher genome-wide levels of CpG methylation in *C. intestinalis* and *C. gigas* than in *N. vectensis* (21.6%, 15%, and 9.4%, respectively; Zemach et al. 2010, Olson and Roberts 2014a). Levels of DNA methylation broadly reflect evolutionary relationships (Tweedie et al. 1997, Zemach et al. 2010), so it could be speculated that coral methylation is similar to that of *N. vectensis*. However, phylogenies derived from methylation patterns may differ considerably from those derived from protein sequences, suggesting lineage-specific changes in methylation (Sarda et al. 2012). Lineage-specific changes in methylation could reflect differences in life history strategies (Sarda et al. 2012).

Across reef coral species, ranking of biological processes according to mean CpG O/E was largely consistent, corroborating a trend reported in other studies of invertebrate gene body methylation (Gavery and Roberts 2010, Sarda et al. 2012, Dixon et al. 2014). With little variation between coral species, the biological processes enriched in high CpG O/E values (predicted to have low CpG DNA methylation) reflect genes involved in inducible functions, while processes associated with the low CpG O/E values reflect genes for essential housekeeping functions. Housekeeping genes are ubiquitously expressed across tissue types, and tend to be evolutionarily conserved. Indeed, higher levels of germline DNA methylation are associated with gene orthology among invertebrate taxa (Sarda et al. 2012, Park et al. 2011, Suzuki et al. 2007), suggesting a protective effect that contrasts with the tendency for methylated cytosines to experience higher mutation rates than nonmethylated nucleotides. However, it may be that DNA methylation has an overall protective influence despite the mutagenic effect on CpGs, or that

CpG mutations have largely run their course over time in heavily methylated genes, with fewer methylated CpGs left for substitutions (Park et al. 2011). One hypothesis for the role of gene body methylation is that it may facilitate consistent expression of ubiquitously expressed core genes needed for essential biological functions (Bird 1995, Roberts and Gavery 2012, Gavery and Roberts 2014). While there have been some reports of negative associations between gene body methylation and gene expression in invertebrates (Suzuki et al. 2007, Riviere et al. 2013), positive or bell-shaped relationships have been reported in several taxa, including corals (Zemach et al. 2010, Gavery and Roberts 2013, Dixon et al. 2014, Jjingo et al. 2012). In *A. millepora*, Dixon et al. (2014) reported that the most highly expressed genes in a reciprocal transplant experiment were those exhibiting higher levels of gene body methylation. These genes were enriched for housekeeping functions. A similar phenomenon of increased expression among hypermethylated genes was reported for *C. gigas*, in addition to an inverse relationship between DNA methylation and variation in expression between tissues (Gavery and Roberts 2013). High transcript abundances among highly methylated genes may reflect their widespread expression across cell or tissue types (Gavery and Roberts 2013, Dixon et al. 2014).

In contrast to hypermethylation and consistent expression, our finding of relatively low methylation among differentially expressed genes in response to thermal stress and ocean acidification in the three acroporid corals supports the hypothesis that hypomethylation is associated with transcriptional plasticity (Roberts and Gavery 2012). Sparse methylation of gene bodies, and the tendency for genes involved in inducible functions to have lower methylation levels, is thought to leave these genes open to a greater variety of transcriptional opportunities (Roberts and Gavery 2012). Potential sources of transcriptional variation could include access to alternative start sites, increased sequence mutations, exon skipping, and transient methylation

(Roberts and Gavery 2012). Such transcriptional variation could contribute to phenotypic plasticity, and it might be particularly beneficial among species that experience variable environments that require constant tuning of gene expression, such as corals and other sessile organisms. Dixon et al. (2014) found support for this theory in *A. millepora*, showing that genes differentially expressed in response to transplantation to new environments were significantly enriched in the low methylation (high CpG O/E) component.

In our analysis, lower methylation among differentially expressed genes was at least partially attributable to increased representation of genes involved in inducible functions relative to those with housekeeping functions. This was especially true in *A. hyacinthus*. While unsurprising, this highlights an important caveat to our analysis; low methylation among differentially expressed genes could simply reflect the fact that environmental stressors elicit higher expression of inducible genes relative to housekeeping genes, and that these genes simply happen to exhibit different methylation patterns. Furthermore, it remains unclear whether gene body DNA methylation actively regulates genes, as opposed to the alternative hypothesis that it is simply a byproduct of an open chromatin state (Jjingo et al. 2012). For example, gene body methylation may reflect exposure of DNA to DNA methyltransferases (DNMTs) as a consequence of unpacked chromatin during transcription (Jjingo et al. 2012). Further studies will be needed to evaluate causative relationships between DNA methylation and transcriptional activity. Additionally, despite the utility of CpG O/E to infer germline methylation patterns, experimental work on transient methylation in somatic cells and how it might influence transcription is needed. However, experimental studies on social insects have already provided compelling evidence for links between differential DNA methylation, transcription, and phenotypic plasticity (Kucharski et al. 2008, Elango et al. 2009). Studies of DNA methylation in

plants have gone even further, having already incorporated epigenetics into ecological research (Bossdorf et al. 2008).

There is a great deal of interest in the adaptive potential of corals in the face of continued environmental change. Perhaps the most intriguing aspect of DNA methylation is that its effects may extend beyond an individual organism's lifetime and be transferred to successive generations. Although the epigenomes of some organisms, such as mammals, are extensively reprogrammed during meiosis and embryogenesis, in some cases DNA methylation can be passed on to offspring (Duncan et al. 2014). This is especially true in plants (Hauser et al. 2011, Heard and Martienssen 2014), but there is also evidence for inheritance of DNA methylation in oysters (Olson and Roberts 2014b). Transgenerational epigenetic inheritance is a potentially transformative biological concept, but its extent and significance is only just beginning to be understood, and remains largely unstudied in the vast majority of organisms (Grossniklaus et al. 2013, Heard and Martienssen 2014).

Our results illustrate patterns of gene body DNA methylation that are similar across coral species and largely consistent with what has been found in other invertebrates. This work serves as a basis with which to further characterize and compare DNA methylation in corals and related taxa. Within the context of environmental challenges faced by corals, our analysis found broad support for an inverse relationship between gene body methylation and differential gene expression. This suggests that the potential role of DNA methylation in the response of corals to environmental change warrants closer investigation.

1.6 Acknowledgements

This study was funded in part by the John E. Halver Fellowship at the UW School of Aquatic and Fishery Sciences, and by an ARCS Seattle Chapter Fellowship, both awarded to JLD. Three anonymous reviewers provided comments that helped improve the manuscript.

1.7 Data Accessibility

This study made use of publicly available transcriptome data. Data files derived from the original data and generated for this study are available at Dryad under doi:10.5061/dryad.pq827. The analyses can be reproduced using a respository with the complete workflow at https://github.com/jldimond/Coral-CpG.

Original datasets were obtained from several sources. *Acropora hyacinthus* data were obtained from Barshis et al. (2013), available at http://palumbi.stanford.edu/data/33496_A*hyacinthus*_CoralContigs.fasta.zip (transcriptome assembly) and http://palumbi.stanford.edu/data/Barshis_etal_33496contigs_deduped_all_scaledounts_Nov2012.txt (gene expression counts). *Acropora millepora* data were obtained from Moya et al. (2012), available at the NCBI Transcriptome Shotgun Assembly Database under accession numbers JR970414–JR999999 and JT000001–JT023377 (transcriptome assembly) and the NCBI Gene Expression Omnibus (GEO) database under accession number GSE33016 (gene expression counts). *Acropora palmata* transcriptome data were obtained from Polato et al. (2011) and Polato et al. (2013), located at https://usegalaxy.org/datasets/cb51c4a06d7ae94e/display?to_ext=fasta (transcriptome assembly) and NCBI Gene Expression Omnibus database (GEO) accession

number GSE36983 (gene expression counts). *Pocillopora damicornis* data were obtained from Vidal-Dupiol et al. (2013) at http://2ei.univ-perp.fr/telechargement/transcriptomes/blast2go_fasta_Pdamv2.zip (transcriptome assembly). *Porites astreoides* data were obtained from Kenkel et al. (2013) and downloaded from http://www.bio.utexas.edu.offcampus.lib.washington.edu/research/matz_lab/matzlab/Data.html (transcriptome assembly). *Stylophora pistillata* data were obtained from Karako-Lampert et al. (2014) and downloaded from http://data.centrescientifique.mc/Data/454Isotigs.fas.zip (transcriptome assembly).

1.8 LITERATURE CITED

Barshis D, Ladner J, Oliver T, Seneca F, Traylor-Knowles N, Palumbi S (2013) Genomic basis for coral resilience to climate change *Proc Natl Acad Sci USA,* 110, 1387–92

Bayer T, Aranda M, Sunagawa S, Yum L, Desalvo M, Lindquist E, Coffroth M, Voolstra C, Medina M (2012) *Symbiodinium* transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals *PLoS One* 7, e35269

Bird A (1995) Gene number, noise reduction and biological complexity *Trends Genet* 11, 94–100

Bonasio R, Li Q, Lian J, Mutti N, Jin L, Zhao H, Zhang P, Wen P, Xiang H, Ding Y et al (2012) Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator. Curr Biol* 22, 1755–64

Bossdorf, O, Richards, C and Pigliucci, M (2008) Epigenetics for ecologists *Ecol Lett* 11, 106–15

Bruno JF, Edmunds PJ (1997) Clonal variation for phenotypic plasticity in the coral *Madracis mirabilis*. *Ecology* 78, 2177–2190

Charmantier A, McCleery R, Cole L, Perrins C, Kruuk L, Sheldon B (2008) Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science* 320, 800–3

Chevin L, Lande R, Mace G (2010) Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory *PLoS Biol* 8, e1000357

Dixon G, Bay L, Matz M (2014) Bimodal signatures of germline methylation are linked with gene expression plasticity in the coral *Acropora millepora. BMC Genomics* 15, 1109

Duncan E, Gluckman P, Dearden P (2014) Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype? *J Exp Zool B Mol Dev Evol* 322, 208–20

Dunlap W, Starcevic A, Baranasic D, Diminic J, Zucko J, Gacesa R, van O M, Hranueli D, Cullum J, Long P (2013) KEGG orthology-based annotation of the predicted proteome of *Acropora digitifera*: ZoophyteBase - an open access and searchable database of a coral genome *BMC Genomics* 14, 509

Elango N, Hunt B, Goodisman M, Yi S (2009) DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera Proc Natl Acad Sci U S A* 106, 11206–11

Forsman Z, Barshis D, Hunter C, Toonen R (2009) Shape-shifting corals: molecular markers show morphology is evolutionarily plastic in *Porites BMC Evol Biol* 9, 45

Gavery M, Roberts, S (2010) DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*) *BMC Genomics* 11, 483

Gavery, M and Roberts, S (2013) Predominant intragenic methylation is associated with gene expression characteristics in a bivalve mollusc *PeerJ* 1, e215

Gavery, M and Roberts, S (2014) A context dependent role for DNA methylation in bivalves *Brief Funct Genomics* 13, 217–22

Granados-Cifuentes, C, Bellantuono, A, Ridgway, T, Hoegh-Guldberg, O and Rodriguez-Lanetty, M (2013) High natural gene expression variation in the reef-building coral *Acropora millepora*: potential for acclimative and adaptive plasticity *BMC Genomics* 14, 228

Grossniklaus, U, Kelly, W, Ferguson-Smith, A, Pembrey, M and Lindquist, S (2013) Transgenerational epigenetic inheritance: how important is it? *Nat Rev Genet* 14, 228–35

Hauser, M, Aufsatz, W, Jonak, C and Luschnig, C (2011) Transgenerational epigenetic inheritance in plants *Biochim Biophys Acta* 1809, 459–68

Heard, E and Martienssen, R (2014) Transgenerational epigenetic inheritance: myths and mechanisms *Cell* 157, 95–109

Hoegh-Guldberg, O, Mumby, P, Hooten, A, Steneck, R, Greenfield, P, Gomez, E, Harvell, C, Sale, P, Edwards, A, Caldeira, K et al (2007) Coral reefs under rapid climate change and ocean acidification *Science* 318, 1737–42

Jjingo, D, Conley, A, Yi, S, Lunyak, V and Jordan, I (2012) On the presence and role of human gene-body DNA methylation *Oncotarget* 3, 462–74

Jones, P and Takai, D (2001) The role of DNA methylation in mammalian epigenetics *Science* 293, 1068–70

Karako-Lampert, S, Zoccola, D, Salmon-Divon, M, Katzenellenbogen, M, Tambutté, S, Bertucci, A, Hoegh-Guldberg, O, Deleury, E, Allemand, D and Levy, O (2014) Transcriptome analysis of the scleractinian coral *Stylophora pistillata PLoS One* 9, e88615

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

Kucharski, R, Maleszka, J, Foret, S and Maleszka, R (2008) Nutritional control of reproductive status in honeybees via DNA methylation *Science* 319, 1827–30

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

Oberlander, T, Weinberg, J, Papsdorf, M, Grunau, R, Misri, S and Devlin, A (2008) Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (NR3C1) and infant cortisol stress responses *Epigenetics* 3, 97–106

Olson, C and Roberts, S (2014a) Genome-wide profiling of DNA methylation and gene expression in *Crassostrea gigas* male gametes *Front Physiol* 5, 224

Olson, C E and Roberts, S B (2014b) Indication of family-specific DNA methylation patterns in developing oysters. http://biorxiv.org/content/early/2014/12/16/012831

Palumbi, S, Barshis, D, Traylor-Knowles, N and Bay, R (2014) Mechanisms of reef coral resistance to future climate change *Science* 344, 895–8

Park, J, Peng, Z, Zeng, J, Elango, N, Park, T, Wheeler, D, Werren, J and Yi, S (2011) Comparative analyses of DNA methylation and sequence evolution using *Nasonia* genomes *Mol Biol Evol* 28, 3345–54

Polato, N, Altman, N and Baums, I (2013) Variation in the transcriptional response of threatened coral larvae to elevated temperatures *Mol Ecol* 22, 1366–82

Polato, N, Vera, J and Baums, I (2011) Gene discovery in the threatened elkhorn coral: 454 sequencing of the *Acropora palmata* transcriptome *PLoS One* 6, e28634

Polato N, Baums, I, Altman, N (2013) Data from: Variation in the transcriptional response of threatened coral larvae to elevated temperatures. *Dryad Digital Repository*. http://dx.doi.org/10.5061/dryad.t3pr6

Roberts, S and Gavery, M (2012) Is there a relationship between DNA methylation and phenotypic plasticity in invertebrates? *Front Physiol* 2, 116

Sarda, S, Zeng, J, Hunt, B and Yi, S (2012) The evolution of invertebrate gene body methylation *Mol Biol Evol* 29, 1907–16

Suzuki, M, Kerr, A, De, S D and Bird, A (2007) CpG methylation is targeted to transcription units in an invertebrate genome *Genome Res* 17, 625–31

Sved, J and Bird, A (1990) The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model *Proc Natl Acad Sci U S A* 87, 4692–6

Todd, P (2008) Morphological plasticity in scleractinian corals *Biol Rev Camb Philos Soc* 83, 315–37

Tweedie, S, Charlton, J, Clark, V and Bird, A (1997) Methylation of genomes and genes at the invertebrate-vertebrate boundary *Mol Cell Biol* 17, 1469–75

Vidal-Dupiol, J, Zoccola, D, Tambutté, E, Grunau, C, Cosseau, C, Smith, K, Freitag, M, Dheilly, N, Allemand, D and Tambutté, S (2013) Genes related to ion-transport and energy production are upregulated in response to CO2-driven pH decrease in corals: new insights from transcriptome analysis *PLoS One* 8, e58652

Wolff, G, Kodell, R, Moore, S and Cooney, C (1998 Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice *FASEB J* 12, 949–57

Zemach, A, McDaniel, I, Silva, P and Zilberman, D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation *Science* 328, 916–9

Chapter 2. GENETIC AND EPIGENETIC INSIGHT INTO MORPHOSPECIES IN A REEF CORAL

A version of this chapter was published as: Dimond JL, Gamblewood SK, & Roberts SB (2017) Genetic and epigenetic insight into morphospecies in a reef coral. Molecular Ecology, 26(19), 5031-5042.

2.1 ABSTRACT

Incongruence between conventional and molecular systematics has left the delineation of many species unresolved. Reef-building corals are no exception, with phenotypic plasticity among the most plausible explanations for alternative morphospecies. As potential molecular signatures of phenotypic plasticity, epigenetic processes may contribute to our understanding of morphospecies. We compared genetic and epigenetic variation in Caribbean branching *Porites* spp., testing the hypothesis that epigenetics—specifically, differential patterns of DNA methylation—play a role in alternative morphotypes of a group whose taxonomic status has been questioned. We used reduced representation genome sequencing to analyze over 1,000 single nucleotide polymorphisms and CpG sites in 27 samples of *Porites* spp. exhibiting a range of morphotypes from a variety of habitats in Belize. We found stronger evidence for genetic rather than epigenetic structuring, identifying three well-defined genetic groups. One of these groups exhibited significantly thicker branches, and branch thickness was a better predictor of genetic groups than depth, habitat, or symbiont type. In contrast, no clear epigenetic patterns emerged with respect to phenotypic or habitat variables. While there was a weak positive correlation between pairwise genetic and epigenetic distance, two pairs of putative clones exhibited substantial epigenetic differences, suggesting a strong environmental effect. We speculate that epigenetic patterns are a complex mosaic reflecting diverse environmental histories

superimposed over a relatively small heritable component. Given the role of genetics in branching *Porites* spp. morphospecies we were able to detect with genome-wide sequencing, use of such techniques throughout the geographic range of these corals may help settle their phylogeny.

2.2 INTRODUCTION

Inconsistencies between conventional- and genetics-based taxonomy are common across taxa (Patterson et al. 1993). Morphological characters have long been the basis of systematics, yet they can be misleading in cases such as convergent evolution or extensive phenotypic plasticity (Potter et al. 1997; Fukami et al. 2004; Fritz et al. 2007). While molecular phylogenetics has illuminated species relationships by identifying and resolving some of these issues, uncertainties often persist, and disagreement among molecular studies is not uncommon. The inferences made by molecular studies are influenced by numerous factors such as the number and type of markers and the models used to analyze them (Brocchieri 2001; Yang & Rannala 2012). Moreover, the field of epigenetics has uncovered novel mechanisms of phenotypic variation and inheritance that has led to reappraisal of the traditional theory of molecular evolution (Noble 2015; Skinner 2015).

Reef-building corals exhibit substantial inter- and intraspecific variation in growth forms that has proved particularly problematic for the taxonomic delineation of species (Fukami et al. 2004; Forsman et al. 2009; Flot et al. 2011). Conventional taxonomy based on morphological features has obscured coral phylogenies and in some cases overestimated species diversity revealed by way of genetic analyses (Fukami et al. 2004; Forsman et al. 2010; Prada et al. 2014). In other cases, supposed ecomorphs of the same species have turned out to be sibling species (Knowlton et al. 1992), and genetic data have also uncovered cryptic species among similar

growth forms of what was thought to be a single species (Keshavmurthy et al. 2013; Schmidt-Roach et al. 2013). The confusion surrounding coral taxonomy has consequences for coral conservation, because a species cannot be conserved and managed if it cannot be defined. For example, 25 coral species are currently listed under the U.S. Endangered Species Act (NOAA 2015), and the ambiguous taxonomic status of some of these species has necessitated revision using molecular data (Forsman et al. 2010).

Several studies have concluded that morphological plasticity is among the most plausible hypotheses for instances of both over- and underestimation of species diversity based on conventional taxonomy (Forsman et al. 2010; Flot et al. 2011; Keshavmurthy et al. 2013; Prada et al. 2014). Indeed, a recent reciprocal transplant study confirmed the role of phenotypic plasticity in alternative morphotypes of *Pocillopora* spp. (Paz-García et al. 2015). Morphological plasticity is likely to be particularly prevalent and adaptive in corals due to their sessility, longevity, modularity, and indeterminate growth (Jackson & Coates 1986; Sebens 1987; Todd 2008). In a review of 26 studies that tested for morphological plasticity in approximately 20 different species of reef corals, 92% of studies documented evidence for plasticity (Todd 2008).

The expanding field of epigenetics may hold promise for understanding phenotypic plasticity in corals. Epigenetic processes are increasingly recognized as molecular signatures of phenotypic variation (Duncan et al. 2014). DNA methylation, the most widely studied epigenetic mark, involves the addition of a methyl group to a cytosine, most commonly in the context of a CpG dinucleotide pair. Along with other epigenetic features such as histone modifications and small RNA molecules, methylation patterns can influence gene expression, though the mechanisms appear to be diverse, complex, and context-dependent (Gavery & Roberts 2014; Duncan et al. 2014). Unlike the DNA sequences of the genome itself, which change relatively

little during an individual's lifetime, genome DNA methylation patterns are not fixed, and can be influenced by environmental stimuli (Duncan et al. 2014). An expanding number of studies have documented differential patterns of DNA methylation associated with alternative phenotypes in a range of taxa (Kucharski et al. 2008; Fonseca Lira-Medeiros et al. 2010; Smith et al. 2015, 2016; Schield et al. 2016). Among different morphotypes of mangroves living in distinct habitats, for example, (Fonseca Lira-Medeiros et al. 2010) found little genetic variation but high levels of epigenetic variation. In the genomes of two alternative morphotypes of threespine sticklebacks, (Smith et al. 2015) identified 77 differentially methylated regions whose functions were associated with known adaptive phenotypes. Interestingly, even among different species of Darwin's finches, epigenetic differences were better correlated with traditional phylogenetic relationships than were genetic differences (Skinner et al. 2014). Thus, an increased understanding of epigenetics may revise definitions of biological diversity and evolution (Skinner 2015), with implications for future conservation efforts.

The taxonomic status of branching corals of the genus *Porites* in the tropical Western Atlantic has been uncertain. There are currently three recognized species*:* *P. porites* (Pallas, 1766), *P. furcata* (Lamarck, 1816), and *P. divaricata* (Lesueur, 1820). Branch diameter and corallite features have traditionally been used to define these species, and while one study found morphological variation to be nearly continuous (Brakel 1977), other studies have identified morphological breaks, with some supporting traditional taxonomic delineations (Weil 1992; Budd et al. 1994; Jameson 1997; Jameson & Cairns 2012). Molecular studies have also failed to reach consensus (Weil 1992; Budd et al. 1994; Forsman et al. 2009; Prada et al. 2014), but earlier studies used allozymes, which could be influenced by epigenetic effects. The most recent and thorough population genetic analysis found no support for upholding the three named

species, finding no significant variation between supposed species across 11 genetic markers and multiple geographic sub-regions (Prada et al. 2014). The authors acknowledged that it is possible that their study overlooked diversity somewhere in the genome, but identified phenotypic plasticity as a plausible explanation for their results. To further explore the potential source of phenotypic variation in branching *Porites* spp., we compared genetic and epigenetic diversity in these corals using restriction site associated DNA (RAD) sequencing.

2.3 METHODS

Specimen collection and DNA extraction

Corals were collected in May 2016 within approximately 4 km of the Smithsonian Institution's Carrie Bow Cay Field Station in Belize (16° 48' 9.39" N, 88° 4' 54.99" W). In the field, the three Caribbean branching *Porites* spp. are distinguished primarily by branch diameter and habitat type, so collections targeted a broad range of branch diameters (6 - 26 mm), depths (0.5 - 17 m) and habitats (mangrove to forereef) to sample as much variation as possible (Fig. 2.1). A total of thirty specimens were collected, but we obtained sufficient sequence data from only 27 of these (Table 2.1).

Figure 2.1. Examples of branching *Porites* spp. collected in this study, showing variation in branch diameter. Clockwise from top left, specimen branch diameters were 6, 10, 17.5 and 21 mm.

Table 2.1. *Porites* spp. specimen data. For descriptions of habitat types on the Belize Barrier Reef, refer to Rützler & Macintyre (1982). Under symbiont clade, "NoAmp" indicates samples for which PCR amplification of the cp23S marker failed despite repeated attempts.

Sample #	Coordinates	Habitat	Depth (m)	Symbiont clade	Branch diameter (mm)	Genetic group
101	N 16° 48.126', W 88° 4.729'	Lower spur & groove	7.9	C	19.5	2
103	N 16° 48.294', W 88° 4.666'	Inner reef slope	16.2	C	10.5	3
104	N 16° 48.294', W 88° 4.666'	Inner reef slope	16.5	C	9	1
105	N 16° 46.792', W 88° 4.599'	Lower spur & groove	9.8	C	11.5	2
106	N 16° 46.792', W 88° 4.599'	Lower spur & groove	7.0	C	17	2
107	N 16° 46.821', W 88° 4.521'	Inner reef slope	12.5	C	11.5	3
108	N 16° 48.120', W 88° 4.751'	Upper spur & groove	7.0	NoAmp	10.5	3
109	N 16° 46.821', W 88° 4.521'	Inner reef slope	14.6	C	10	1
110	N 16° 48.146', W 88° 4.658'	Inner reef slope	14.0	C	14	2
111	N 16° 48.120', W 88° 4.751'	Upper spur & groove	7.9	C	14	2
112	N 16° 46.571', W 88° 4.524'	Lower spur & groove	8.5	C	13	3
114	N 16° 46.798', W 88° 4.511'	Outer ridge	15.5	C	9	1
115	N 16° 48.354', W 88° 4.998'	Backreef	1.2	C	20	2
116	N 16° 48.354', W 88° 4.998'	Backreef	1.2	NoAmp	10	3
117	N 16° 48.354', W 88° 4.998'	Backreef	1.2	A	13	3
118	N 16° 49.645', W 88° 6.364'	Mangrove	0.6	C	7	3
121	N 16° 49.645', W 88° 6.364'	Mangrove	0.6	C	7.5	3
122	N 16° 49.645', W 88° 6.364'	Mangrove	0.6	C	6	1
123	N 16° 48.181', W 88° 4.885'	Backreef	0.8	C	21	2
124	N 16° 48.115', W 88° 4.983'	Lagoon/cut	4.6	C	16.5	3
125	N 16° 46.798', W 88° 4.511'	Outer ridge	16.5	C	10	1
126	N 16° 46.798', W 88° 4.511'	Outer ridge	17.4	C	26	2
127	N 16° 48.115', W 88° 4.983'	Lagoon/cut	2.7	A	15	2
128	N 16° 48.181', W 88° 4.885'	Backreef	0.6	NoAmp	10	3
129	N 16° 48.115', W 88° 4.983'	Turtlegrass	0.9	A	10.5	3
130	N 16° 48.181', W 88° 4.885'	Backreef	0.8	C	17.5	2
131	N 16° 48.181', W 88° 4.885'	Turtlegrass	0.5	A	10	3

Branch tips were cut with shears and placed in a conical tube. Within one hour of collection, specimens were transferred to tubes containing salt-saturated DMSO (SS-DMSO) for DNA preservation (Gaither et al. 2011). Coral tissue preserved in SS-DMSO began to slough off

the skeleton after several days and was easily removed using forceps. Small pieces of tissue (~0.5 μl volume) were washed three times via centrifugation with phosphate buffered saline prior to DNA extraction using Qiagen DNeasy Blood and Tissue kits according to the manufacturer's protocol, with an overnight lysis with proteinase K. Samples were further purified via overnight ethanol precipitation, followed by resuspension in Qiagen AE buffer. Genomic DNA was checked for yield and quality via fluorescence (Qubit BR assay) and gel electrophoresis, respectively.

The remaining coral skeletons were soaked in a 10% bleach solution for 24 hrs, then dried at 40 °C for 48 hrs. The diameter of branch tips at their widest cross-section was measured to the nearest 0.5 mm with vernier calipers.

Symbiont genotyping

Reef corals commonly engage in species-specific associations with symbiotic dinoflagellates (*Symbiodinium* spp.), and these associations can also be related to genetic structure within a given host (Bongaerts et al. 2010; Finney et al. 2010). To identify the dominant *Symbiodinium* type associated with each coral, an approximately 700 base pair region of domain V of the cp23S-rDNA region was PCR-amplified using primer pair 23S1 (5'-CACGACGTTGTAAAACGACGGC TGTAACTATAACGGTCC-3') and 23S2 (5'-GGATAACAATTTCACACAGGCCATCGTATTGAACCCAGC-3') (Santos et al. 2002). PCR was performed in 25 μl volumes containing 1X green buffer (Promega), 2.5 mM MgCl2, 240 μM dNTP, 5 pmol of each primer, 1U Taq (GoTaq, Promega), and 1-20 ng of template DNA. Reactions were carried out in an Applied Biosystems Veriti thermocycler under the following conditions: initial denaturing period of 1 min at 95 °C, 35 cycles of 95 °C for 45 s, 55 °C for 45

s, and 72 °C for 1 min, and a final extension period of 7 min. PCR products were cleaned (NEB Monarch kit) and checked on a 1% agarose gel. Cleaned PCR products (10 ng/µL) were sent to Sequetech Corporation (Mountain View, CA) for Sanger sequencing using the forward primer. Chromatograms were edited with Geneious v9.1.5, converting bases with a Phred score < 30 to Ns. Sequences were then aligned using ClustalW and queried against the GenBank nucleotide database in Geneious.

RAD library preparation and sequencing

Double digest RADseq (ddRADseq) libraries were prepared following the methods of (Peterson et al. 2012). Only samples with high molecular weight DNA were used for library preparation. For each sample, a minor variation of ddRADseq called EpiRADseq was also employed to evaluate methylated loci (Schield et al. 2016). Both methods use two restriction enzymes—a rare cutter and a common cutter—to perform a double digest of the DNA at specific restriction sites throughout the genome, with the only difference that EpiRADseq uses a methylation-sensitive common cutter that will not cut methylated loci. Both methods used the rare cutter PstI (5'-CTGCAG-3' recognition site), while the common cutter MspI (5'-CCGG-3' recognition site) was used for ddRADseq and the methylation-sensitive isoschizomer HpaII (also 5'-CCGG-3' recognition site) was used for EpiRADseq preparations. Double digests of 300-500 ng gDNA per sample were carried out using 20 units of each enzyme in the manufacturer's supplied buffer (New England Biolabs, NEB) for 5 hours at 37 °C. Samples were cleaned using magnetic beads (Sera-Mag SpeedBeads) prior to ligation of barcoded Illumina adapters onto the fragments (Peterson et al. 2012). After two rounds of bead cleanup, samples were pooled into 12 libraries (along with 18 other samples not analyzed in this study), which were then subjected to

automated size-selection of fragments between 415 and 515 bp using a Pippin Prep (Sage Science). Libraries were then PCR amplified using Phusion Taq (NEB) and Illumina-indexed primers (Peterson et al. 2012), followed by a single round of bead cleanup. Final library fragment sizes and concentrations were evaluated with D1000 ScreenTape on an Agilent 2200 TapeStation. Libraries were sent to the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley, where their concentrations were verified via qPCR prior to 100 bp, paired-end sequencing in equimolar ratios on the Illumina HiSeq 4000.

RAD sequence assembly

Sequences were assembled using ipyrad v0.3.41 (Eaton 2014). We used the 'denovo - reference' assembly method with the *Symbiodinium minutum* (clade B; GenBank accession GCA_000507305.1 (Shoguchi et al. 2013)) and *Symbiodinium kawagutii* (clade F; http://web.malab.cn/symka_new/data/Symbiodinium_kawagutii.assembly.935Mb.fa.gzb.fa.gz (Lin et al. 2015)) genomes used as reference to subtract symbiont reads from the de novo assembly. We concatenated these genomes into a single reference file. Step one of the ipyrad workflow demultiplexed the data in each pool by identifying restriction overhangs and barcode sequences associated with each sample; zero barcode mismatches were tolerated. Demultiplexed samples were then combined in a single directory for further steps. In step two, reads were trimmed of barcodes and adapters and quality filtered using a q-score threshold of 20, with bases below this score converted to Ns and any reads with more than 5 Ns removed. Step three mapped reads to the concatenated symbiont reference genomes with BWA using the default bwa mem setting and removed any mapped reads. With the remaining reads, similar clusters of reads were identified using a threshold of 85% similarity and aligning them. We chose 85% as a moderately

conservative clustering threshold to avoid over-splitting of loci (Harvey et al. 2015). Next, step four performed joint estimation of heterozygosity and error rate (Lynch 2008) based on a diploid model assuming a maximum of 2 consensus alleles per individual. Step five used the parameters from step four to determine consensus bases calls for each allele, and removed consensus sequences with greater than 5 Ns per end of paired-end reads. With consensus sequences identified, step six clustered and aligned reads for each sample to consensus sequences. Finally, step seven filtered the dataset according to maximum number of indels allowed per read end (8), maximum number of SNPs per locus (20), maximum proportion of shared heterozygous sites per locus (0.5), and minimum number of samples per locus (15).

Henceforth, we will use the term locus to refer to a consensus paired-end read. The term SNP refers specifically to a single nucleotide polymorphism on a locus, while the term CpG refers to a cytosine-guanine dinucleotide pair that can be either methylated or non-methylated at the 5'-CCGG-3' restriction site of each locus.

SNP analysis

We analyzed unlinked SNPs that were sampled by ipyrad at 1 SNP per locus with the least amount of missing data; SNPs were sampled randomly if they had equal amounts of missing data. We were able to estimate the SNP error rate, defined as the proportion of SNP mismatches between pairs of the same individuals (Mastretta-Yanes et al. 2015), by treating ddRADseq and EpiRADseq samples as technical replicates and computing pairwise differences between individuals using the dist.gene function in the R package ape (Paradis et al. 2004). We then analyzed a SNP dataset from ddRADseq libraries that had no missing data across samples.

While this reduced the dataset considerably, it allowed us to test variation across a similar set of loci for both the genetic and epigenetic analysis, as explained further below.

Genetic differentiation among samples was examined using multidimensional scaling (k = 2) using the cmdscale R function. Discriminant analysis of principal components (DAPC) in the R package adegenet (Jombart 2008) was used to further examine these patterns (but note that DAPC was performed on raw data and was separate from the MDS analysis). Rather than applying prior assumptions about the identity of each coral specimen to assign population groups, we used the find.clusters function in adegenet to identify groups. This k-means clustering function reduces the data with principal components analysis (PCA) before estimating the number of clusters with the lowest Bayesian information criterion (BIC). All PCs were retained for the analysis and a maximum of 10 clusters was specified. Once groups were identified, cross-validation (xvalDapc) was used to estimate the number of PCs to retain for the subsequent DAPC analysis. In the groups identified by the DAPC analysis, genetic differentiation between groups was evaluated by computing pairwise Weir and Cockerham's FST using the R package hierfstat (Goudet 2005).

DNA methylation analysis

Methylation detection with EpiRADseq data involves analysis of read counts (Schield et al. 2016). Presence of a given read in the EpiRADseq dataset indicates that the locus is not methylated at the restriction cut site. Conversely, if a locus is methylated at the 5'-CCGG-3' cut site, HpaII is blocked from cutting, and the locus will be absent from the dataset. Hence, read counts of zero are informative, but they could also be the result of missing data through processes such as allele dropout or variation in library size or fragment size selection. To control

for this, we used ddRADseq data to normalize the EpiRADseq data. Any loci with zeros in the ddRADseq library were treated as absent and removed, thereby leaving zeros in the EpiRADseq library only where the locus was counted in the ddRADseq library. This had the added benefit that it resulted in analysis of a similar set of loci as the SNP analysis described above. EpiRADseq and ddRADseq read counts per locus were highly correlated, with the exception of methylated loci, which, as expected, had low or zero abundances in EpiRADseq libraries (Fig. 2.2A). The dataset comprising read counts for both EpiRADseq and ddRADseq was subject to normalization of read counts using TMM normalization in the R package edgeR (Robinson et al. 2010). Residuals of linear regressions of EpiRADseq and ddRADseq libraries provided optimal differentiation of methylated loci from non-methylated loci (Fig. 2.2B&C). Finally, a binary dataset was created from these data by setting a residual threshold of -1 for methylated loci such that residuals less than -1 were coded as methylated (1) and residuals greater than -1 were coded as non-methylated (0) (Fig. 2.2C). These data were evaluated using multidimensional scaling as described above for the SNP analysis.

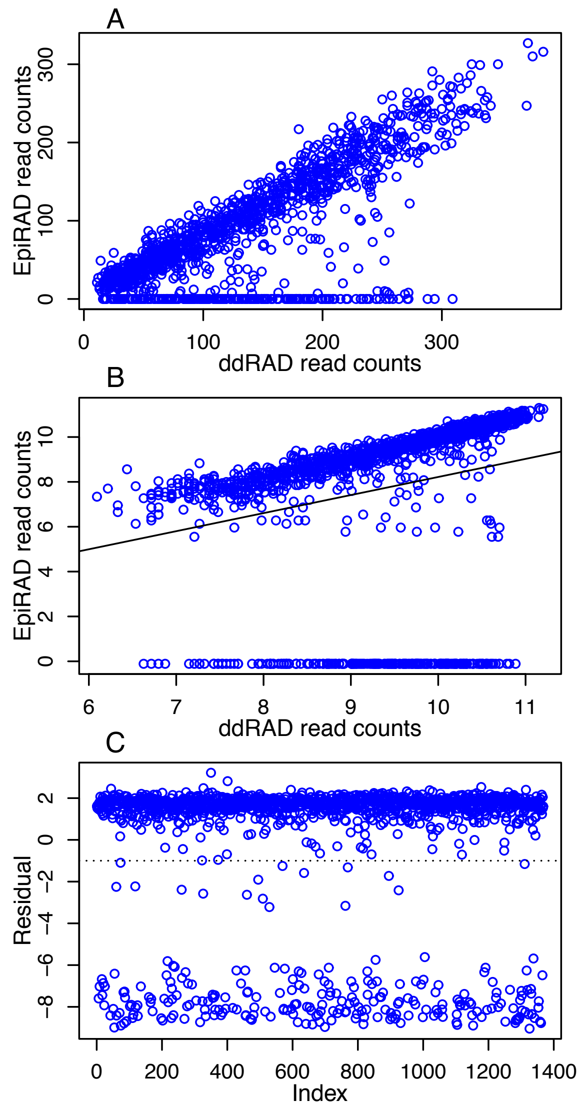

Figure 2.2. Determining methylated loci in a representative sample. (A) ddRADseq and EpiRADseq raw read counts for each locus (each point is a locus) were highly correlated except for methylated reads, which were absent or at low abundance in the EpiRAD library and cluster at the bottom of the y-axis. (B) ddRADseq and EpiRADseq read counts after TMM normalization, showing regression line from linear model used for derivation of residuals. (C) Residuals from the linear regression of the normalized data. Non-methylated reads have positive residuals and cluster at the top while methylated reads have negative residuals and cluster at the bottom. The binary dataset derived from these data was based on a residual threshold of ≤ -1 (dotted line) for designating loci as methylated.

EpiRADseq read counts could theoretically represent variable levels of methylation within a sample, such as through variation in methylation across pooled replicates, or among different tissue or cell types (Schield et al. 2016). However, our analysis of ddRADseq and EpiRADseq read counts in tandem indicated that much of the variability in read counts occurs in both libraries and is thus likely related to library preparation effects such as fragment size-selection or PCR bias (Davey et al. 2013). In the original EpiRADseq method developed by Schield et al. (2016), PCR effects were minimized by using unique molecular identifier sequences attached to fragments. We suspect that fragment size-selection effects could be equally important in driving read count variability. For example, abundant reads could simply represent the mean or mode fragment size within the library after size-selection, while less abundant reads could represent the tails of the distribution. In either case, our analysis suggests that much of the read count variation above zero does not reflect methylation levels, and by normalizing EpiRADseq data to the ddRADseq data and creating a binary dataset, we removed much of the potential bias described above.

A repository with the complete bioinformatic workflow described above can be accessed at https://github.com/jldimond/Branching-*Porites*.

2.4 RESULTS

Data yield

An average of 3.24 million paired-end reads per sample were obtained based on restriction overhangs and barcodes, with an average of 2.63 million reads per sample remaining after q-score and adapter filtering. After filtering for symbionts and minimum read depth, an average of 26,330 consensus loci per sample were obtained, with a total of 135,980 unique

consensus loci across samples. Obtaining datasets with no missing data across samples substantially reduced the number of loci; the final SNP dataset consisted of 1,113 unlinked SNPs (1,113 consensus paired-end loci with 1 SNP sampled per locus) across 27 samples, while the methylation dataset consisted of 1,712 CpGs (1,712 consensus paired-end loci with one CpG cut site per locus) across 25 samples. Differences in the number of SNPs/CpGs between datasets were due to the different numbers of samples; two samples, 101 and 112, were removed from the methylation dataset due to low coverage. This reduced the number of samples in the methylation dataset but resulted in a greater number of shared CpG sites.

Symbiont identity

BLAST searches of *Symbiodinium* cp23S sequences indicated that the dominant symbiont in the majority of corals was a clade C *Symbiodinium* with high similarity to *Symbiodinium* subclade C3 (Table 2.1). Four corals occurring in shallow water habitats (≦ 2.7 m) hosted clade A *Symbiodinium* with high similarity to subclade A3. We were unable to amplify cp23S in three of the specimens despite repeated attempts.

Genetic patterns

Based on SNP mismatches between technical replicates, the SNP error rate was estimated to be 3.6% (standard deviation, SD, 3.1%), which is on the lower end of the range that has been reported previously (Mastretta-Yanes et al. 2015) (Fig. 2.3). In other words, multilocus genotypes of technical replicates were 96.4% similar on average. In contrast, nearly all non-replicate pairwise comparisons exhibited greater differences, with the exception of two pairs of individuals (109 & 114, 127 & 115) that exhibited similarity within the range of the SNP error

(Fig. 2.3). The high similarity of these individuals suggests they are possible clones. Specimens 109 and 114 were separated by approximately 45 m, while specimens 127 and 115 were separated by approximately 450 m.

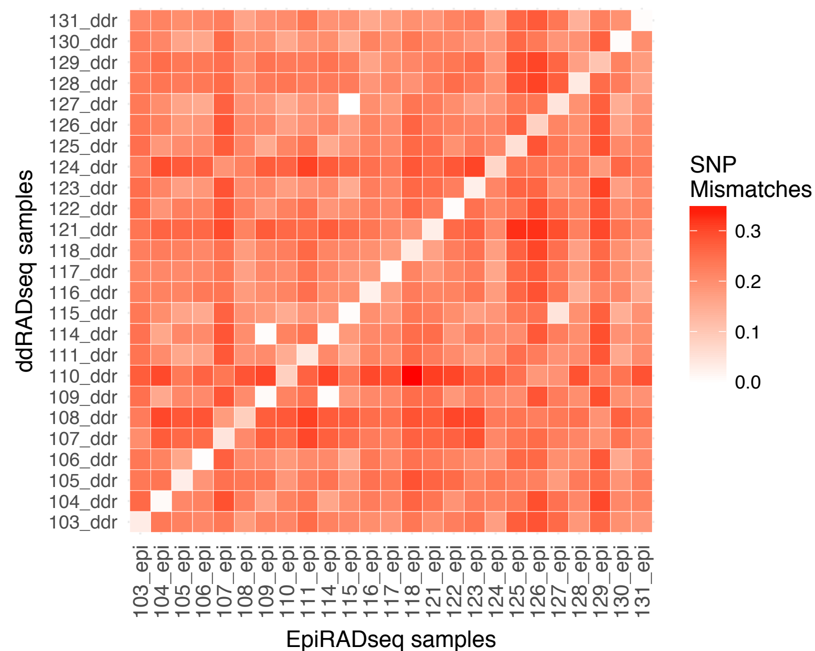

Figure 2.3. Pairwise comparisons of SNP mismatches between all 25 individuals with both ddRADseq and EpiRADseq data. The color ramp represents the proportion of SNP mismatches, also known as genetic distance, between pairs. ddRADseq and EpiRADseq data for the same individual were used as technical replicates to estimate the SNP error rate and are expressed along the diagonal. Two pairs of outliers with a low proportion of mismatches (109/114, 115/127) are possible clones.

Multidimensional scaling of samples based on SNPs suggested genetic structure of *Porites* spp., with samples clustering into three generally well-separated groups (Fig. 2.4A). K-means clustering of SNPs showed the strongest support for three clusters according to BIC values (Fig. 2.4B). Based on results of cross-validation, nine PCs (the maximum suggested given

the sample size) were retained for DAPC analysis of the three groups, explaining 64% of the variance. These were further collapsed into two discriminant functions accounting for 76% and 24% of the remaining variance, respectively. As with the MDS analysis, DAPC indicated clear separation of the three SNP groups (Fig. 2.4C). The two pairs of potential clones were in two separate groups (109 & 114 in group 1, 127 & 115 in group 2). Similar values of FST were observed between groups (FST for each pairwise comparison: 1&2 = 0.194; 1&3 = 0.207; 2&3 = 0.191).

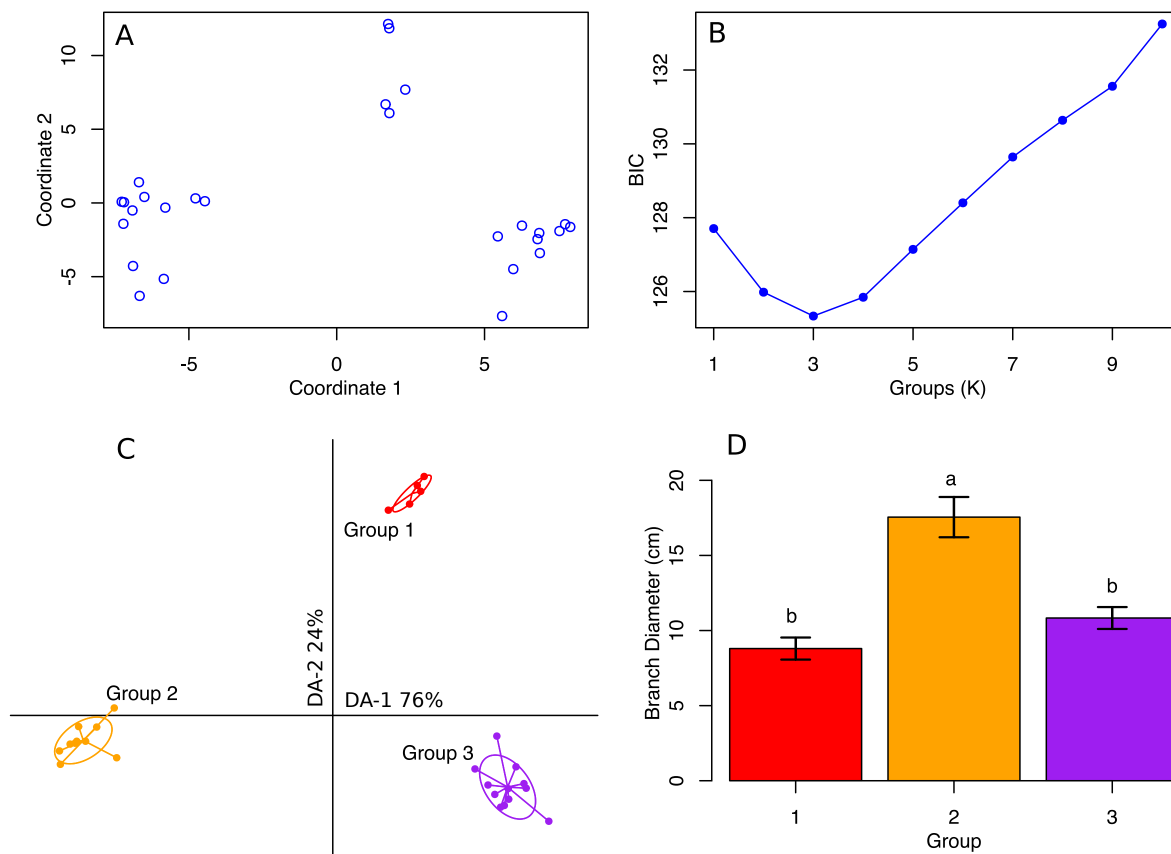

Figure 2.4. Genetic and phenotypic variation in *Porites* spp. (A) Multidimensional scaling plot of multilocus genotypes derived from SNPs in the *Porites* spp. specimens. Each point represents a specimen. Distances between points represent Euclidean distances projected in two dimensions. (B) Results of *find.clusters* analysis on SNP data identifying optimal number of

groups. K = 3 groups was determined to be optimal based on the lowest value of the Bayesian information criterion (BIC). (C) Results of DAPC analysis on SNP data using the three optimal groups identified by *find.clusters.* The analysis was based on 9 retained principal components explaining 64% of the variance, further collapsed to two discriminant functions accounting for all of this variance (76% and 24% for each axis, respectively). (D) Comparison of branch diameter in the three *Porites* groups identified by the DAPC analysis (mean ± 1 SE). Different letters denote groups determined to be significantly different in pairwise t-tests.

We evaluated potential factors associated with genetic structure using multiple regression. For the genetic variable, we used the first discriminant axis from the DAPC analysis of SNPs. This was regressed against collection depth, symbiont type, habitat, and branch diameter. The model explained 61% of the variation in the SNP variable. The function calc.relimp in the R package relaimpo was used to estimate the relative importance of each model component. The relative importance of depth, symbiont type, habitat, and branch diameter in the model was 0.5%, 2%, 31%, and 66.5%, respectively. We further evaluated branch diameter with a one-way ANOVA, followed by pairwise t-tests with Bonferroni adjustment. Group 2 branch diameter was significantly different from both groups 1 and 3 ($p < 0.001$), while groups 1 and 3 were not significantly different from each other ($p = 0.724$) (Fig. 2.4D).

Epigenetic patterns

Among the 1,712 CpGs, a range of 314-360 were methylated per sample, yielding a mean methylation level of 19.5% (SD 0.8%). However, there were two outliers with high levels of methylation. This was due largely to low read count loci being categorized as methylated. A more stringent minimum read count threshold of 10 reads per locus was therefore applied to the ddRADseq dataset and this resulted in 1,368 CpGs with a range of 238-263 methylated CpGs per

sample, corresponding to a methylation level of 18.3% (SD 0.5%). Of the 1,368 CpGs, 208 were differentially methylated, 131 were constitutively methylated across all samples, and 1,029 were constitutively non-methylated across all samples.

Methylation patterns in the MDS analysis were less clear than SNP patterns, with no clear grouping and most of the variation spread across much of the horizontal axis (Fig. 2.5A). K-means clustering of CpG methylation indicated the strongest support for a single group (Fig. 2.5B), so DAPC was not appropriate.

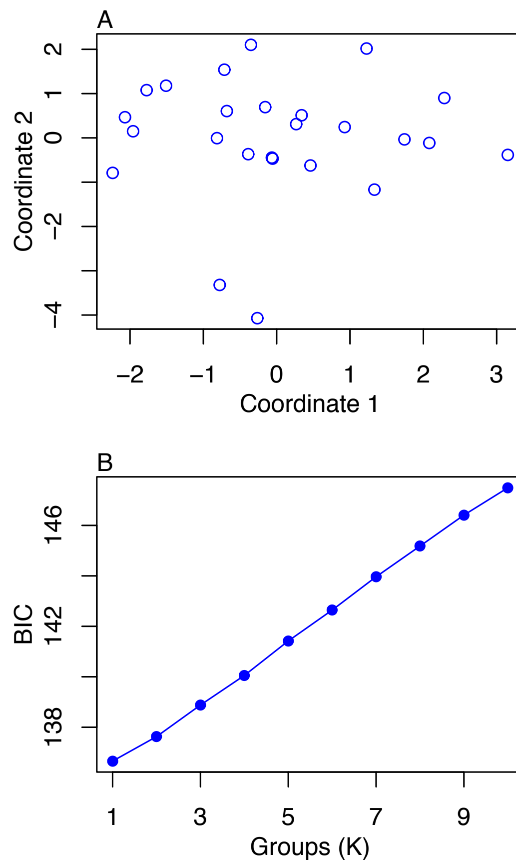

Figure 2.5. Examining patterns in *Porites* spp. CpG methylation. (A) Multidimensional scaling plot of multilocus epi-genotypes in the specimens. Each point represents a specimen. Distances between points represent Euclidean distances projected in two dimensions. (B) Results of find.clusters analysis on multilocus epi-genotypes identifying optimal number of groups. K =

1 groups was determined to be optimal based on the lowest value of the Bayesian information criterion (BIC). In other words, this analysis did not find any epigenetic structuring.

Instead, we applied hierarchical clustering to the 208 differentially methylated CpGs and 25 samples to visualize patterns (Fig. 2.6). No clear clustering was observed according to depth, branch diameter, habitat type, symbiont type, or genetic group.

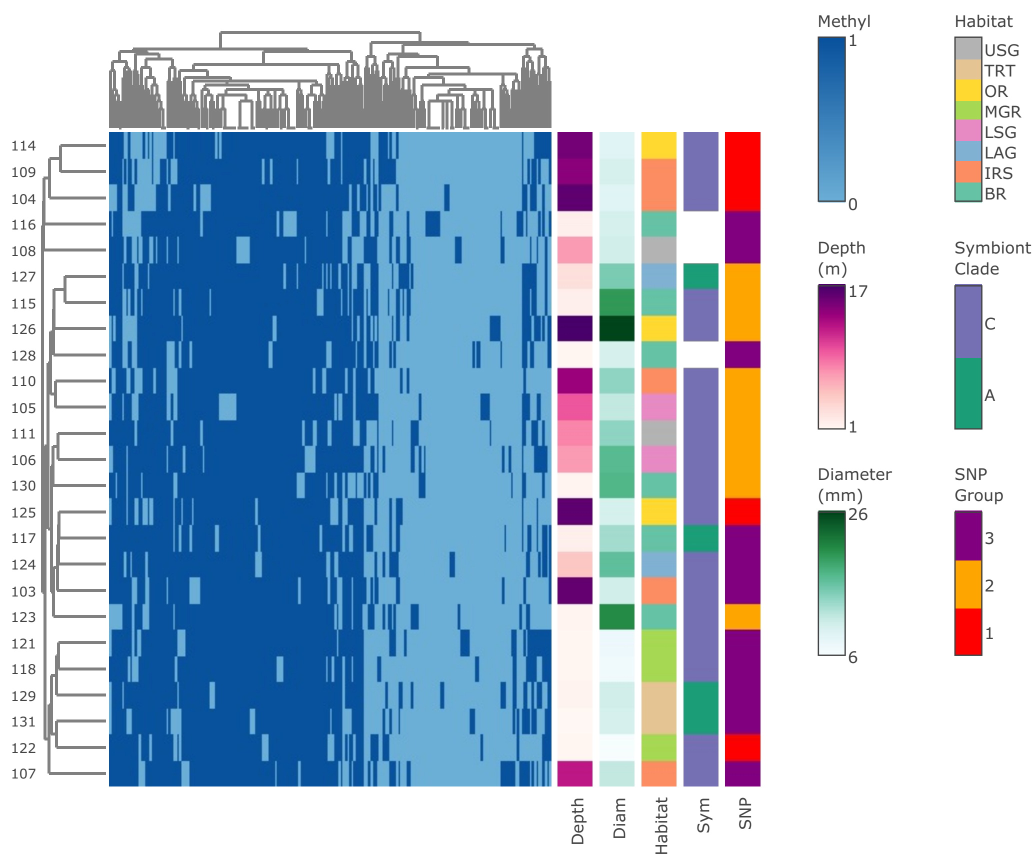

Figure 2.6. Differentially methylated CpGs among *Porites* spp. individuals. Methylated CpGs are shown in dark blue. Hierarchical clustering was used for ordering of both rows (samples) and columns (CpGs) and is shown on the right and top sides of the plot. Sample ID numbers are also shown on the left-hand side. Color bars on the right-hand side of the plot depict sample phenotypic and environmental data. From left, these include water depth, branch

diameter, habitat type (see Table 2.1 for full habitat names), symbiont clade, and the SNP (genetic) groups identified in Fig. 2.4.

Linkages between genetic and epigenetic variation

We compared genetic and epigenetic variation by calculating pairwise genetic and epigenetic distances, again using the dist.gene function in the R package ape. This was the same statistic generated in Fig. 2.3 for the SNP data, measuring the proportion of pairwise mismatches in multilocus SNPs and CpG methylation status. Excluding the two putative clones (109 & 114, 127 & 115) as outliers, the 298 unique pairwise comparisons yielded a weak positive correlation between genetic and epigenetic distance (Spearman rank correlation, 0.17; Fig. 2.7). Notably, however, the two putative clones had both the lowest genetic distances as well as the lowest epigenetic distances, though the epigenetic distances were much larger (13-16%) than the genetic distances (0%) in these individuals.

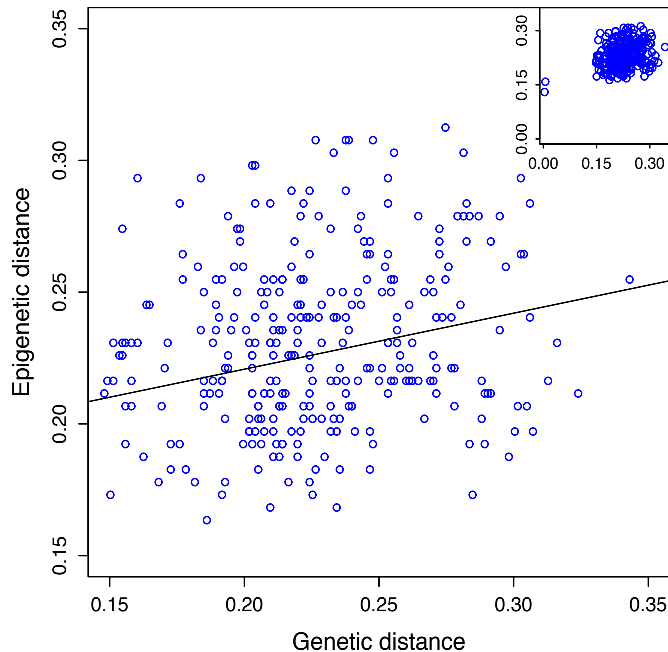

Figure 2.7. Association between pairwise genetic distance and pairwise epigenetic distance. Each point represents a unique pairwise comparison between two samples among the 25 samples that had both genetic and epigenetic data (298 unique pairwise comparisons total, exclusive of two outliers). For a given pair, genetic and epigenetic distances represent the proportion of mismatches between multilocus genotypes (SNPs) and multilocus epi-genotypes (CpG methylation status), respectively. A linear trend line is shown (Spearman rank correlation = 0.17). **Inset:** All data inclusive of the outliers (samples 109/114 and 127/115); these putative clones showed the lowest levels of both genetic and epigenetic differentiation.

2.5 Discussion

Our sample of branching *Porites* spp. from Belize exhibited clear genetic differentiation, supporting the idea that these corals comprise three separate and fairly well-defined groups. While this result is in general agreement with older studies based on allozymes (Weil 1992; Budd et al. 1994), it contrasts with the most recent study based on 11 genetic markers (Prada et al. 2014). The inferences we were able to make likely reflect the large number of markers we

sampled. Reduced representation genome sampling methods like RADseq have shown great promise in resolving phylogenetic relationships, especially in recalcitrant taxa like cnidarians for which traditional mitochondrial and nuclear markers have shown limited success (Pante et al. 2015; Combosch & Vollmer 2015; Herrera & Shank 2016; Rosser et al. 2017). Despite the number of markers we sampled, however, our study lacks sufficient geographic breadth to draw strong phylogenetic conclusions. Indeed, all evaluations of branching *Porites* spp. to date, including ours, have been limited by geographic scope, insufficient colony sample size, or low genome coverage. A thorough reexamination of branching *Porites* spp. phylogeny across the Caribbean / Western Atlantic region may be warranted, assessing both genome-wide variation and multivariate morphological traits. Ideally, such a study would incorporate randomized collections in different habitats to permit assessment of potential habitat selection.

Differences in branch diameter, particularly between group 2 and the other two groups, suggest that the genetic differentiation we observed is associated with colony-level morphological variation. Variation in branch thickness is the primary diagnostic feature used for *Porites* spp. identification in the field, and is likely what historically prompted delimitation of species groups. Corallite-level variation in morphology is also well documented, and most studies using these methods have identified distinct groups based on multivariate corallite characters (Weil 1992; Budd et al. 1994; Jameson 1997). Two of these studies found evidence supporting the three recognized species (Weil 1992; Jameson 1997), while one found up to five morphospecies among specimens collected from three regions (Budd et al. 1994), and another identified continuous variation without any clear breaks (Brakel 1977). However, despite the continuous variation observed by Brakel (1977), he judged the variation to be largely genetic (fixed) based on the observation that 1) corals from very different environments can have very

similar morphologies, and 2) corals from very similar environments can have very different morphologies. This concurs with our analysis suggesting that morphotype variation in *Porites* spp. has a genetic basis. On the other hand, we also observed evidence for phenotypic plasticity, such as in the case of three specimens collected from nearby mangrove prop roots exhibiting similar branch morphologies; one of these colonies (122) was assigned to group 1 while the others (118, 121) were assigned to group 3.

Based on the lack of consensus among previous studies concerning the taxonomic status of branching *Porites* spp., and hypotheses that phenotypic plasticity might underlie this incongruence, we did not anticipate finding strong genetic differentiation and instead we hypothesized that epigenetic patterns might prove more informative. Instead, patterns of DNA methylation were less conclusive than genetic patterns. Levels of methylation did not vary greatly among samples, and even among differentially methylated CpGs, most were either methylated or unmethylated in a majority of samples. Similarly, little variation in methylation was observed in a recent study of threespine sticklebacks, with only 737 differentially methylated CpG sites identified out of 1,445,567 sites examined across eight individuals (Smith et al. 2015). However, these differentially methylated loci were associated with alternative phenotypes (Smith et al. 2015). The only phenotypic pattern to emerge from our data was a suggestion that individuals in group 2, which also tended to be more thickly branched, were generally grouped close together epigenetically. Thus, there is a possibility that epigenetic variation is coupled with genetic variation in these individuals. Genetic-epigenetic coupling is further suggested by the weak positive correlation between pairwise genetic and epigenetic distance. This relationship implies a heritable component, albeit small, to the methylation patterns.

The two pairs of putative clones deserve special consideration. These individuals are likely clones because even though some branching *Porites* spp. colonies are hermaphrodites capable of self-fertilization (Schlöder & Guzman 2008), Mendelian segregation leads to an extremely low probability of identical parental and offspring genotypes (Stoddart 1983). The physical distances separating these pairs suggests that they are unlikely to be clones resulting from fragmentation, but possibly the result of asexually produced larvae (Harrison 2011) or dispersal via parrotfish corallivory (Boulay et al. 2014). If they are indeed clones, the similar methylation patterns in these individuals would reflect inheritance via mitosis rather than sexual recombination through the germline. However, regardless of whether methylation profiles were transmitted sexually or asexually in these individuals, it is notable that while genetic distance was near zero, epigenetic distance was relatively high. This could reflect divergence of the methylome due to environmental effects. Methylation levels in corals have recently been shown to be at least partially under environmental influence over short time scales (Putnam et al. 2016). Clearly, the methylation patterns we observed in *Porites* spp. exhibit some unexplained variance, and this may reflect the diverse habitat conditions and environmental histories experienced by the corals analyzed here; the factors depicted in Fig. 5 illustrate this complexity, and yet these represent only a fraction of the factors that could influence the methylome. For example, we did not account for coral age, and methylation is known to change with age in numerous taxa, including invertebrates (e.g. Lian et al. 2015).

Reduced representation genome sequencing techniques are currently very popular and are enhancing our ability to probe molecular processes, but they are not without error. Our analysis of ddRADseq and EpiRADseq libraries in tandem provided robust control for error in both the SNP and methylation datasets. Technical replicates have been advocated as a means to assess

genotyping error, and by assessing genotypes of the same individuals from the two libraries we confirmed that this error was within the range documented elsewhere (Mastretta-Yanes et al. 2015; Recknagel et al. 2015). For the methylation analysis, comparing EpiRADseq libaries to ddRADseq libraries was a key factor in controlling for library composition effects, reducing the likelihood of false positives. This paired library approach appears to be a strong alternative to the unique molecular identifier approach used by Schield et al. (2016).

In addition to genetic data, prevailing symbiont populations could provide an additional means to evaluate whether branching *Porites* spp. exhibit species-level differentiation (e.g. Pinzón and LaJeunesse 2011). Species-specific associations between host corals and *Symbiodinium* are common in the Caribbean, particularly among brooding species such as *Porites* (Finney et al. 2010; Bongaerts et al. 2015). Furthermore, branching *Porites* spp. appear to associate with host-specialist symbionts that are not common in other hosts (Finney et al. 2010; Bongaerts et al. 2015). Branching *Porites* spp. also tend to host distinct symbionts from their congener *P. astreoides* (Finney et al. 2010; Bongaerts et al. 2015). Although cp23S is not a fine-scale genetic marker, the majority of corals we analyzed hosted the same *Symbiodinium* clade C phylotype, while three individuals in group 3 and one in group 2 hosted the clade A phylotype (Table 2.1). A comprehensive reexamination of branching *Porites* spp. would be wise to include an assessment of *Symbiodinium* communities.

Conclusion

Contrary to our expectations, branching *Porites* spp. morphotype variation was better explained by genetic patterns than epigenetic patterns. This analysis benefited from the resolution afforded by genome-wide sequencing, and may justify a more thorough analysis of

branching *Porites* spp. phylogeny throughout the tropical Western Atlantic. Although patterns of DNA methylation were not as conclusive as genetic patterns, there was some evidence of covariation between genetic and epigenetic variation. This possibility, as well as potential environmental influence on methylation in corals, will require further study. Given the increasingly powerful molecular biology tools available for work in environmental epigenomics, stronger inferences about the extent, variability, and potential functions of epigenetic processes in corals are only a matter of time.

2.6 ACKNOWLEDGEMENTS

We thank the Smithsonian Institution's Caribbean Coral Reef Ecosystems Program for field support, and the Belize Fisheries Department for specimen export permitting. RADseq training and materials were generously provided by Adam Leaché and Kevin Epperly. Daniel Thornhill and Terra Hiebert provided advice on symbiont genotyping. Sam White, Hollie Putnam, Katherine Silliman, and Megan Hintz provided helpful comments on the manuscript. The manuscript was further improved by thoughtful critique from Zac H. Forsman and an anonymous reviewer. This study was supported by the Hall Conservation Genetics Research Award (UW-CoEnv), the ARCS Foundation Seattle Chapter, the John E. Halver Fellowship (UW-SAFS), and National Science Foundation Award OCE-1559940. This is contribution 997 to the Smithsonian's Caribbean Coral Reef Ecosystems Program.

2.7 DATA ACCESSIBILITY

Symbiont cp23S DNA sequences were deposited in GenBank (accession numbers

KY649212-KY649238). Demultiplexed RADseq reads were deposited in the NCBI Short Read Archive (SRA accession numbers SAMN06566335-SAMN06566364). A repository detailing assembly and analysis methods can accessed at https://github.com/jldimond/Branching-*Porites*.

2.8 Literature Cited

Bongaerts P, Carmichael M, Hay KB *et al.* (2015) Prevalent endosymbiont zonation shapes the depth distributions of scleractinian coral species. *Royal Society open science*, **2**, 140297.

Bongaerts P, Riginos C, Ridgway T *et al.* (2010) Genetic divergence across habitats in the widespread coral Seriatopora hystrix and its associated *Symbiodinium*. *Plos One*, **5**, e10871.

Boulay JN, Hellberg ME, Cortés J, Baums IB (2014) Unrecognized coral species diversity masks differences in functional ecology. *Proceedings of the Royal Society of London B: Biological Sciences*, **281**, 20131580.

Brakel W (1977) Corallite variation in *Porites* and the species problem in corals. *Proceedings of the Third Internatioeedinal Coral Reef Symposium*, **1**, 457–462.

Brocchieri L (2001) Phylogenetic inferences from molecular sequences: review and critique. *Theoretical Population Biology*, **59**, 27–40.

Budd AF, Johnson KG, Potts DC (1994) Recognizing morphospecies in colonial reef corals: I. Landmark-based methods. *Paleobiology*, **20**, 484–505.

Combosch DJ, Vollmer SV (2015) Trans-Pacific RAD-Seq population genomics confirms introgressive hybridization in Eastern Pacific *Pocillopora* corals. *Molecular Phylogenetics and Evolution*, **88**, 154–162.

Davey JW, Cezard T, Fuentes-Utrilla P *et al.* (2013) Special features of RAD Sequencing data: implications for genotyping. *Molecular Ecology*, **22**, 3151–3164.

Duncan EJ, Gluckman PD, Dearden PK (2014) Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype? *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution*, **322**, 208–220.

Eaton DAR (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.

Finney JC, Pettay DT, Sampayo EM *et al.* (2010) The relative significance of host-habitat, depth, and geography on the ecology, endemism, and speciation of coral endosymbionts in the genus *Symbiodinium*. *Microbial Ecology*, **60**, 250–263.

Flot J, Blanchot J, Charpy L *et al.* (2011) Incongruence between morphotypes and genetically delimited species in the coral genus *Stylophora*: phenotypic plasticity, morphological convergence, morphological stasis or interspecific hybridization? *BMC Ecology*, **11**, 22.

Fonseca Lira-Medeiros C, Parisod C, Fernandes R *et al.* (2010) Epigenetic variation in mangrove plants occurring in contrasting natural environment. *Plos One*, **5**, e10326.

Forsman ZH, Barshis DJ, Hunter CL, Toonen RJ (2009) Shape-shifting corals: molecular markers show morphology is evolutionarily plastic in *Porites*. *BMC Evolutionary Biology*, **9**, 45.

Forsman ZH, Concepcion GT, Haverkort RD *et al.* (2010) Ecomorph or endangered coral? DNA and microstructure reveal hawaiian species complexes: Montipora dilatata/flabellata/turgescens & M. patula/verrilli. *Plos One*, **5**, e15021.

Fritz U, Hundsdörfer A, Široký P *et al.* (2007) Phenotypic plasticity leads to incongruence between morphology-based taxonomy and genetic differentiation in western Palaearctic tortoises (Testudo graeca complex; Testudines, Testudinidae). *Amphibia-Reptilia*, **28**, 97–121.

Fukami H, Budd AF, Paulay G *et al.* (2004) Conventional taxonomy obscures deep divergence between Pacific and Atlantic corals. *Nature*, **427**, 832–835.

Gaither M, Szabó Z, Crepeau M, Bird C, et al. (2011) Preservation of corals in salt-saturated DMSO buffer is superior to ethanol for PCR experiments. *Coral Reefs*, **30**, 329–333.

Gavery MR, Roberts SB (2014) A context dependent role for DNA methylation in bivalves. *Briefings in functional genomics*, **13**, 217–222.

Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular ecology resources*, **5**, 184–186.

Harrison PL (2011) Sexual Reproduction of Scleractinian Corals. In: *Coral Reefs: An Ecosystem in Transition* (eds Dubinsky Z, Stambler N), pp. 59–85. Springer Netherlands, Dordrecht.

Harvey MG, Judy CD, Seeholzer GF *et al.* (2015) Similarity thresholds used in DNA sequence assembly from short reads can reduce the comparability of population histories across species. *PeerJ*, **3**, e895.

Herrera S, Shank TM (2016) RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. *Molecular Phylogenetics and Evolution*, **100**, 70–79.

Jackson JBC, Coates AG (1986) Life cycles and evolution of clonal (modular) animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **313**, 7–22.

Jameson S (1997) Morphometric analysis of the Poritidae (Anthozoa: Scleractinia) off Belize. In: *Proceedings of the 8th International Coral Reef Symposium* (eds Lessios H, McIntyre I), pp. 1591–1596. International Society for Reef Studies, Smithsonian Tropical Research Institute, Panama.

Jameson S, Cairns S (2012) Neotypes for *Porites Porites* (Pallas, 1766) and *Porites divaricata*

Le Sueur, 1820 and remarks on other western Atlantic species of *Porites* (Anthozoa: Scleractinia). *Proceedings of the Biological Society of Washington*, **125**, 189–207.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Keshavmurthy S, Yang S-Y, Alamaru A *et al.* (2013) DNA barcoding reveals the coral "laboratory-rat", *Stylophora pistillata* encompasses multiple identities. *Scientific reports*, **3**, 1520.

Knowlton N, Weil E, Weigt LA, Guzmán HM (1992) Sibling Species in Montastraea annularis, Coral Bleaching, and the Coral Climate Record. *Science*, **255**, 330–333.

Kucharski R, Maleszka J, Foret S, Maleszka R (2008) Nutritional control of reproductive status in honeybees via DNA methylation. *Science*, **319**, 1827–1830.

LaJeunesse T (2002) Diversity and community structure of symbiotic dinoflagellates from Caribbean coral reefs. *Marine biology*, **141**, 387–400.

Lian S, He Y, Li X, Zhao B, Hou R, Hu X, Zhang L, Bao Z (2015) Changes in global DNA methylation intensity and DNMT1 transcription during the aging process of scallop *Chlamys farreri*. *Journal of Ocean University of China*, **14**, 685-690.

Lin S, Cheng S, Song B *et al.* (2015) The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. *Science*, **350**, 691–694.

Lynch M (2008) Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. *Molecular Biology and Evolution*, **25**, 2409–2419.

Mastretta-Yanes A, Arrigo N, Alvarez N *et al.* (2015) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population

genetic inference. *Molecular ecology resources*, **15**, 28–41.

NOAA (2015, October 15). Corals. Retrieved March 11, 2017, from http://www.nmfs.noaa.gov/pr/species/invertebrates/corals.htm

Noble D (2015) Evolution beyond neo-Darwinism: a new conceptual framework. *The Journal of Experimental Biology*, **218**, 7–13.

Pante E, Abdelkrim J, Viricel A *et al.* (2015) Use of RAD sequencing for delimiting species. *Heredity*, **114**, 450–459.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**, 289–290.

Patterson C, Williams DM, Humphries CJ (1993) Congruence between molecular and morphological phylogenies. *Annual review of ecology and systematics*, **24**, 153–188.

Paz-García DA, Hellberg ME, García-de-León FJ, Balart EF (2015) Switch between Morphospecies of *Pocillopora* Corals. *The American Naturalist*, **186**, 434–440.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *Plos One*, **7**, e37135.

Pinzón JH, LaJeunesse TC (2011) Species delimitation of common reef corals in the genus *Pocillopora* using nucleotide sequence phylogenies, population genetics and symbiosis ecology. *Molecular Ecology*, **20**, 311-325.

Potter D, Lajeunesse T, Saunders G, et al. (1997) Convergent evolution masks extensive biodiversity among marine coccoid picoplankton. *Biodiversity and conservation*, **6**, 99–107.

Prada C, DeBiasse MB, Neigel JE *et al.* (2014) Genetic species delineation among branching Caribbean *Porites* corals. *Coral reefs (Online)*, **33**, 1019–1030.

Putnam HM, Davidson JM, Gates RD (2016) Ocean acidification influences host DNA methylation and phenotypic plasticity in environmentally susceptible corals. *Evolutionary applications*, **9**, 1165–1178.

Recknagel H, Jacobs A, Herzyk P, Elmer KR (2015) Double-digest RAD sequencing using Ion Proton semiconductor platform (ddRADseq-ion) with nonmodel organisms. *Molecular ecology resources*, **15**, 1316–1329.

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Rosser NL, Thomas L, Stankowski S *et al.* (2017) Phylogenomics provides new insight into evolutionary relationships and genealogical discordance in the reef-building coral genus *Acropora*. *Proceedings. Biological Sciences / the Royal Society*, **284**.

Rützler K, Macintyre I (1982) The habitat distribution and community structure of the barrier reef complex at Carrie Bow Cay, Belize. In: *The Atlantic barrier reef ecosystem at Carrie Bow Cay, Belize, I. Structure and communities.* , pp. 9–45.

Santos S, Taylor D, Kinzie R *et al.* (2002) Molecular phylogeny of symbiotic dinoflagellates inferred from partial chloroplast large subunit (23S)-rDNA sequences. *Molecular Phylogenetics and Evolution*, **23**, 97–111.

Schield DR, Walsh MR, Card DC *et al.* (2016) EpiRADseq: scalable analysis of genomewide patterns of methylation using next-generation sequencing. *Methods in ecology and evolution / British Ecological Society*, **7**, 60–69.

Schlöder C, Guzman H (2008) Reproductive patterns of the Caribbean coral *Porites furcata* (Anthozoa, Scleractinia, Poritidae) in Panama. *Bulletin of marine science*, **82**, 107–117.

Schmidt-Roach S, Lundgren P, Miller KJ *et al.* (2013) Assessing hidden species diversity in the

coral *Pocillopora damicornis* from Eastern Australia. *Coral reefs (Online)*, **32**, 161–172.

Sebens KP (1987) The ecology of indeterminate growth in animals. *Annual review of ecology and systematics*, **18**, 371–407.

Shoguchi E, Shinzato C, Kawashima T *et al.* (2013) Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Current Biology*, **23**, 1399–1408.

Skinner MK (2015) Environmental Epigenetics and a Unified Theory of the Molecular Aspects of Evolution: A Neo-Lamarckian Concept that Facilitates Neo-Darwinian Evolution. *Genome Biology and Evolution*, **7**, 1296–1302.

Skinner MK, Gurerrero-Bosagna C, Haque MM *et al.* (2014) Epigenetics and the evolution of Darwin's Finches. *Genome Biology and Evolution*, **6**, 1972–1989.

Smith TA, Martin MD, Nguyen M, Mendelson TC (2016) Epigenetic divergence as a potential first step in darter speciation. *Molecular Ecology*, **25**, 1883–1894.

Smith G, Smith C, Kenny JG, Chaudhuri RR, Ritchie MG (2015) Genome-wide DNA methylation patterns in wild samples of two morphotypes of threespine stickleback (*Gasterosteus aculeatus*). *Molecular Biology and Evolution*, **32**, 888–895.

Stoddart J (1983) Asexual production of planulae in the coral *Pocillopora damicornis*. *Marine Biology*.

Todd PA (2008) Morphological plasticity in scleractinian corals. *Biological Reviews of the Cambridge Philosophical Society*, **83**, 315–337.

Weil E (1992) Genetic and morphological variation in Caribbean and eastern Pacific *Porites* (Anthozoa, Scleractinia). Preliminary results. *Proc 7th int coral Reef Symp*.

Yang Z, Rannala B (2012) Molecular phylogenetics: principles and practice. *Nature Reviews.*

Genetics, **13**, 303–314.

Chapter 3. RESPONSE OF DNA METHYLATION TO ENVIRONMENTAL CHANGE IN THE REEF CORAL *PORITES ASTREOIDES*

A version of this chapter is being submitted to a special issue of Frontiers in Marine Science as: Dimond JL, Roberts SB. Response of DNA methylation to environmental change in the reef coral *Porites astreoides*

3.1 ABSTRACT

Phenotypic acclimatization is an essential organismal response to environmental change that may be rooted in epigenetic mechanisms. In reef building corals, organisms that are severely threatened by environmental change, limited evidence suggests that DNA methylation is an environmentally responsive mediator of acclimatization. We investigated the degree of change in DNA methylation of the reef coral *Porites astreoides* in response to simulated environmental change. Coral colonies were sampled from a variety of habitats on the Belize Barrier Reef and transplanted to a common garden for one year. We used restriction site associated DNA sequencing, including a methylation-sensitive variant, to subsample the genome and assess changes in DNA methylation levels after a year in the common garden. Methylation changes among the 629 CpG loci we recovered were subtle, with an average of 2% of loci per individual changing in methylation status over the one-year period. Despite these limited changes, coral methylomes were more similar to each other after a year in the common garden together, indicating convergence of methylation profiles in a common environment. Differentially methylated loci showed matches with both coding and non-coding RNA sequences with putative roles in intracellular signaling, apoptosis, gene regulation, and epigenetic crosstalk. The subtle yet significant convergence of methylation profiles in *P. astreoides* concur with the emerging

studies to date suggesting that DNA methylation in reef corals is an environmentally responsive process reflective of acclimatization.

3.2 Introduction

Epigenetic processes, which contribute to gene regulation without affecting underlying DNA sequences, are increasingly recognized as molecular mechanisms that shape phenotypic responses (Duncan et al. 2014). Moreover, epigenetic signatures of organisms can change over their lifetimes, acting as potential records of, and responses to, environmental changes (Duncan et al. 2014; Hofmann 2017; Eirin-Lopez and Putnam 2019). The expanding evidence for both consequential functions of epigenetic processes and their plasticity is therefore driving interest among environmental and evolutionary biologists in search of the molecular basis of phenotypic plasticity, local adaptation, and responses to climate change (Duncan et al. 2014; Hofmann 2017; Eirin-Lopez and Putnam 2019).

Although epigenetics encompasses a suite of molecular processes that appear to interact together, DNA methylation is the best understood and most widely studied of these processes (Eirin-Lopez and Putnam 2019). In the animal kingdom, 5-methylcytosine is the most common form of DNA methylation and is almost exclusively associated with CpG motifs. Invertebrate genomes are generally more sparsely methylated than vertebrate genomes, and methylation tends to be concentrated within gene bodies (i.e., introns and exons) of housekeeping genes (Sarda et al. 2012). While our understanding of the function of gene body methylation is only in its infancy, current evidence suggests that it helps ensure transcriptional fidelity, consistency, and efficiency, and may also be involved in alternative mRNA splicing (Neri et al. 2017; Flores et al. 2012).

Stable yet labile, epigenetic marks like DNA methylation can persist over generations, but they can also be primed and altered by environmental changes. In this way, epigenetic processes are thought to impart environmental "memories" in organisms (Iwasaki and Paszkowski 2014). Environmental memories may be particularly relevant in organisms such as plants and sessile invertebrates because they must weather any changes in their environment. Indeed, there are numerous examples of environmentally inducible epigenetic modifications in plants (Kinoshita and Seki 2014; Iwasaki and Paszkowski 2014). For sessile invertebrates, however, far less is known, and studies are just beginning to emerge.

Tropical reef corals are long-lived, sessile invertebrates that are thought to be particularly reliant on physiological acclimatization and phenotypic plasticity to cope with environmental variation (Gates and Edmunds 1999; Todd 2008). The underlying basis of this plasticity could lie, at least in part, in epigenetic mechanisms like DNA methylation (Roberts and Gavery 2012). However, in order to mediate phenotypic plasticity, DNA methylation itself must be plastic. To date, only a few studies have evaluated the response of DNA methylation to environmental change in corals. In a comparative study of two species of corals, Putnam et al. (2016) found that global methylation levels of *Montipora capitata* did not change in response to reduced pH conditions, while those of *Pocillopora damicornis* were responsive. In another study of simulated ocean acidification conditions with *Stylophora pistillata*, Liew et al. (2018) observed modifications in methylation levels of genes involved in cell cycle and body size pathways that were reflected by phenotypic changes. Finally, in a reciprocal transplant study, Dixon et al. (2018) reported genome-wide changes in methylation in *Acropora millepora* that were correlated with physiological and transcriptional plasticity. In the latter two studies, there was evidence that

changes in methylation were associated with acclimatization. However, there is still limited information on the extent to which DNA methylation responds to environmental change.

In this study, we used a common garden approach to test the response of DNA methylation to environmental change in the Western Atlantic reef coral *Porites astreoides*. Corals were resampled after a year in a common garden and methylation levels were assessed using restriction site associated DNA sequencing (RADseq) techniques. We hypothesized that DNA methylation would be responsive to this manipulation and that methylation profiles among colonies would be more similar to each other after a year in a common environment.

3.3 METHODS

Common garden experiment

In November 2015, 19 colonies of *P. astreoides* (approximately 20 cm diameter) were transplanted from their home site to a common garden in the shallow (~1 m depth) backreef in front of Carrie Bow Cay (CBC; 16° 48’ 9″N, 88° 4’ 55″W), Belize. Coral colonies were haphazardly selected from shallow habitats (1-3 m depth) within a 20 km radius of CBC. Some colonies were collected from windward backreef habitats similar to those of CBC, while others were collected from inshore habitats. Upon collection, colonies were first sampled for DNA by chipping off a small piece of the colony with hammer and chisel, then preserving the fragment in salt-saturated DMSO solution at room temperature until extraction. Colonies were then halved, and one half was brought to the common garden, an approximately 9 m^2 area, where they were reattached to the substratum using A-788 splash zone compound. Fourteen of the original 19 colonies also had their remaining half reattached to the substratum from which they were collected, serving as controls. Numbered aluminum tags were attached adjacent to each coral to permit later identification, and subsurface floats were moored near all control colonies left at

their site of origin. Colonies were resampled one year later in November 2016, again removing and preserving a small fragment for DNA extraction.

Environmental analysis

To characterize the surrounding physical environment, remotely sensed data from the AQUA MODIS satellite sensor were used (NASA 2016). Previous studies have reported strong correlations between remotely sensed sea surface temperatures (SST) and those recorded from data loggers moored in shallow benthic habitats, suggesting that remotely sensed data is appropriate for studies of shallow benthic environments (Smale and Wernberg 2009; Pearce et al. 2006). Monthly AQUA MODIS climatology datasets for the period November 2015 to October 2016 were used. For SST, we used the 11µ band nighttime dataset (NASA 2016). For chlorophyll concentration, the OCx algorithm was used (NASA 2016). Datasets were imported into R as raster images for analysis. Additionally, Belize basemap (Meerman and Clabaugh 2017) and coral reef basemap (UNEP-WCMC 2018) shapefiles were used to provide spatial context.

DNA extraction and sequencing

DNA extraction and library preparation and sequencing followed methods described in detail by Dimond et al. (2017). Briefly, libraries were prepared according to the double-digest RADseq (ddRADseq) and EpiRADseq methods of Peterson et al. (2012) and Schield et al. (2016), respectively. This created tandem libraries for each sample; the ddRADseq library used a methylation insensitive common cutter (MspI) targeting 5'-CCGG-3' motifs, while the EpiRADseq library used a methylation sensitive common cutter (HpaII, also targeting 5'-CCGG-

3'). ddRADseq and EpiRADseq rely on a size-selection step to ensure targeted sequencing of a small subset of fragments within a narrow size range, with the ultimate goal of sequencing genomic intervals that will be present across many samples (Peterson et al. 2012). Paired-end, 100 bp libraries were sequenced in equimolar ratios on the Illumina HiSeq 4000. A total of 96 samples, half ddRADseq and half EpiRADseq, were sequenced on a single lane; some of these samples were used in a separate study.

Symbiont genotyping

Most reef corals engage in obligate symbiotic associations with dinoflagellates of the family Symbiodiniaceae (LaJeunesse et al. 2018). To inform the sequence assembly workflow described below regarding the choice of symbiont genome used to subtract symbiont reads, symbiont genotypes for each sample were determined at the genus (formerly clade) level via NCBI BLAST (v. 2.6.0) queries of ddRADseq libraries. A custom BLAST database was generated by searching for all *Symbiodinium* (formerly *Symbiodinium* clade A), *Breviolum* (*Symbiodinium* clade B), *Cladocopium* (*Symbiodinium* clade C), and *Durusdinium* (*Symbiodinium* clade D) records in the NCBI nucleotide database. Search terms were as follows: ((("*Symbiodinium*"[Organism] AND "*Symbiodinium* sp. clade A"[Organism]) OR "*Symbiodinium* sp. clade B"[Organism]) OR "*Symbiodinium* sp. clade C"[Organism]) OR "*Symbiodinium* sp. clade D"[Organism]. Due to wide variation in the number of records for each taxon, the full dataset was standardized by random sampling to retain 20,000 records per taxon (80,000 total records retained out of the original 447,000). The resulting sequences were used as a BLAST database against which each ddRADseq library was queried with BLASTN using max_target_seqs = 1 and an e-value of 1e-20. Reads that aligned to more than one taxon were

removed, and any reads mapping to 18S rDNA were also removed since this gave many ambiguous matches.

The efficacy of this approach was tested using a prior dataset for which symbiont genotyping had also been performed using PCR amplification and Sanger sequencing of cp23S rDNA amplicons (methods described and data reported in Dimond et al. (2017)). Ten samples of branching *Porites* spp. were tested for correspondence between symbiont identities determined via BLAST searches of ddRADseq reads to those determined via cp23S amplicon sequencing.

Coral sequence assembly

Coral sequences were assembled with ipyrad v.0.5.15 (Eaton 2014), using the "denovo – reference" assembly method to exclude symbiont reads from the assembly. Given the results of the symbiont genotyping analysis identifying the dominant symbiont taxon as *Symbiodinium* (formerly *Symbiodinium* clade A; see Results) in all corals, the *Symbiodinium microadriaticum* (GenBank Accession no. GCA_001939145.1 (Aranda et al. 2016)) genome was used as a reference to exclude symbiont reads. Step one of ipyrad demultiplexed the reads by identifying restriction overhangs and barcode sequences associated with each sample; zero barcode mismatches were tolerated. Demultiplexed samples were then combined in a single directory for further steps. In step two, reads were trimmed of barcodes and adapters and quality filtered using a q-score threshold of 20, with bases below this score converted to Ns and any reads with more than 5 Ns excluded. Next, reads were mapped to the symbiont reference genome with bwa mem using default settings and any mapped reads were excluded from further analysis. Similar clusters of the remaining reads were then aligned using a threshold of 90% similarity. Step four performed joint estimation of heterozygosity and error rate (Lynch 2008) based on a diploid

model assuming a maximum of 2 consensus alleles per individual. Step five used the parameters from step four to determine consensus bases calls for each allele and removed consensus sequences with > 5 Ns per end of paired-end reads. With consensus sequences identified, step six clustered and aligned reads for each sample to consensus sequences. Lastly, the data were filtered according to maximum number of indels allowed per read end (8 indels), maximum number of SNPs per locus (20), maximum proportion of shared heterozygous sites per locus (0.3) and minimum number of samples for a locus to be reported (20). From this final dataset, which included monomorphic loci, a subset of the data was chosen that maximized the number of resampled individuals (sampled in both 2015 and 2016) with a robust set of shared loci (i.e., loci with missing data were excluded). Several samples were excluded from analysis if either one or both of the years in which they were sampled had low data yield (e.g., due to low starting DNA quality). As a definition of terms used here, the term locus to refers to a consensus paired-end read. The term SNP refers to a single nucleotide polymorphism on a locus, while the term CpG refers to a cytosine-guanine dinucleotide pair that can be either methylated or non-methylated at the 5'-CCGG-3' restriction site of each locus.

Genetic analysis

Unlinked SNPs were scored by ipyrad at 1 SNP per locus with the least amount of missing data; SNPs were sampled randomly if they had equal amounts of missing data. The SNP error rate, defined as the proportion of SNP mismatches between pairs of datasets from identical individuals (Mastretta-Yanes et al. 2015), was estimated by treating ddRADseq and EpiRADseq libraries as technical replicates and calculating pairwise differences between individuals using the *dist.gene* function in the R package ape (Paradis et al. 2004).

Epigenetic analysis

Analysis of DNA methylation using EpiRADseq data relies on read count information, as read counts of loci using this technique are inversely related to their methylation frequency (Schield et al. 2016). The analysis followed methods described in detail by Dimond et al. (2017) using a tandem ddRAD/EpiRAD approach in which reads that were present in the ddRAD library but absent in the EpiRAD library were considered methylated. As in Dimond et al. (2017), the residuals from linear regression of ddRAD vs. EpiRAD read counts were used to ascertain methylation status, however, instead of manually setting the methylated/unmethylated threshold for a locus, k-means clustering was used to differentiate methylated from unmethylated loci using k=2. The superheat R package was used for heatmap visualization of methylation patterns (Barter and Yu 2018). The potential functions of differentially methylated loci were determined via a web-based BLASTN search using the default 'nr' nucleotide database and an e-value threshold of e^{-5}.

Data accessibility

Demultiplexed sequence reads can be accessed in the NCBI sequence read archive (SRA) under accession number SRP132538 (https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP132538), or under BioProject PRJNA433592 (https://www.ncbi.nlm.nih.gov//bioproject/PRJNA433592). Analysis workflows can be accessed at https://github.com/jldimond/P.ast-transplant.

3.4 RESULTS

Sample recovery and data yield

Of the 19 colony halves transplanted to the common garden, only one was not recovered a year later. Of the 14 control halves left at their site of origin, three were not relocated, while a fourth had been overtaken and killed by a damselfish garden. Although many samples were submitted for sequencing, low sequencing yield from several samples effectively excluded them from the final analysis due to insufficient data. Sequencing yield was highly variable, producing an average of 3.4 (s.d. = 3.0) million reads per sample. An average of 7539 (s.d. = 4799) consensus loci per sample were included in the final assembly. If data were insufficient (based on summary statistics given by ipyrad) for one or both sampling years for a given coral colony, the colony was excluded from analysis. The optimal balance between colony sample size and number of shared loci was achieved with N = 8 colonies sampled in both 2015 and 2016, including two controls, sharing 629 loci. Thus, 629 SNPs and CpGs were analyzed.

Environmental analysis

Of the eight colonies included in the final analysis, four (colonies 5, 6, 10, and 11) originated from inshore environments approximately 6-8 km west of the barrier reef, while the other four (colonies 2, 3, 8, and 9) were collected from offshore environments just shoreward (within 500 m) of the barrier reef (Fig. 3.1). Inshore reefs were characterized by slightly cooler annual temperatures, a larger seasonal temperature range, and higher chlorophyll a concentration (Fig. 3.1A-C). These effects likely reflect the relatively shallow depth and longer residence time of water in the lagoon and its greater susceptibility to seasonal variation in solar heating and wind patterns, in addition to freshwater runoff and terrestrial nutrient sources. Most corals

brought to the common garden originated in habitats with a greater seasonal temperature range and high chlorophyll concentration; the slightly lower mean temperature of these habitats was probably not biologically meaningful (Fig. 3.1D).

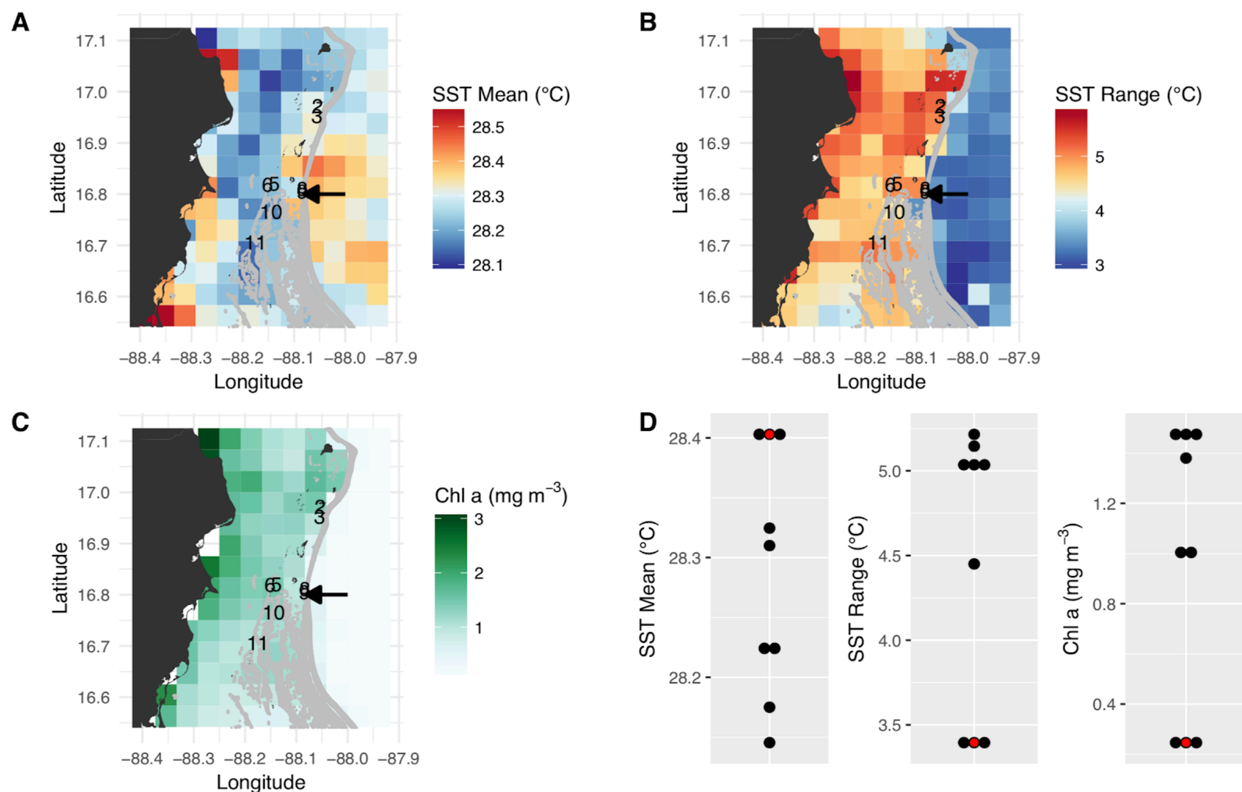

Figure 3.1. Satellite-derived environmental conditions associated with coral collection sites and the common garden along the Belize Barrier Reef. (A-C) Mean annual sea surface temperature (SST), annual SST range, and annual mean chlorophyll a concentration. The collection sites of the eight analyzed coral colonies are depicted by colony numbers, and the arrows show the location of the common garden (colonies 8 and 9 originated from just north of the common garden site). (D) Summary of native habitat and common garden environmental conditions. The common garden is depicted in red.

Symbiont genotyping

For the test dataset evaluating the efficacy of symbiont genotyping via BLAST searches using ddRADseq reads, the dominant symbiont taxon detected via BLAST search was identical to the dominant symbiont taxon detected via cp23S Sanger sequencing in all cases (Fig. 3.2). Among the *P. astreoides* samples, all individuals hosted ≥ 98% *Symbiodinium* (formerly *Symbiodinium* clade A) across both years. This justified the use of the *S. microadriaticum* genome to subtract symbiont sequences during de novo assembly of the RADseq data.

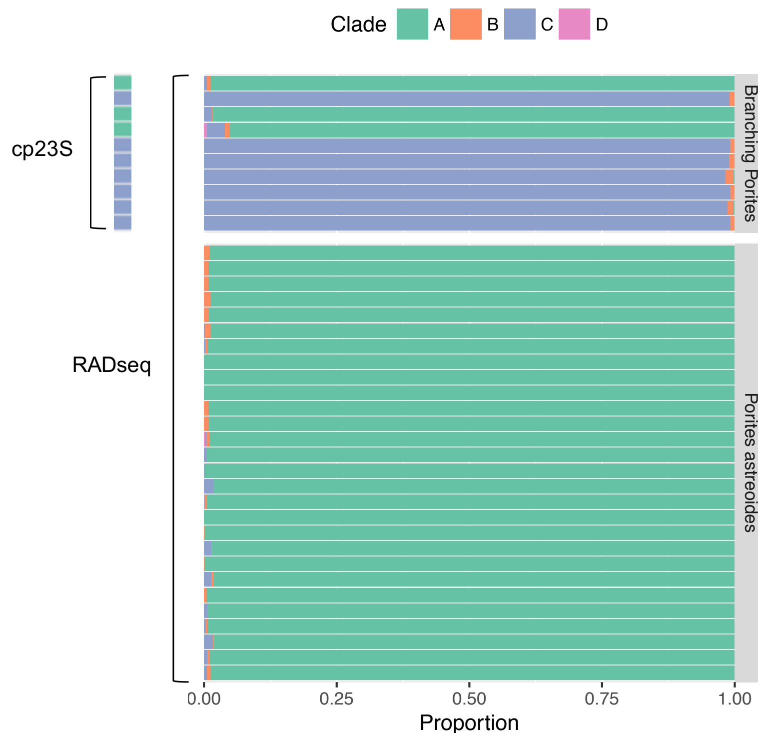

Figure 3.2. Symbiont identification using BLAST searches of ddRADseq data against custom databases of *Symbiodinium* (formerly *Symbiodinium* clade A), *Breviolum* (*Symbiodinium* clade B), *Cladocopium* (*Symbiodinium* clade C), and *Durusdinium* (*Symbiodinium* clade D). Top panel shows symbionts of branching *Porites* spp. collected for a prior study (Dimond et al. 2017); the dominant symbiont found in BLAST searches was identical to the dominant symbiont identified via cp23S Sanger sequencing. Bottom panel shows results from analysis of the *P. astreoides* colonies that are the primary subject of this study.

Genetic analysis

Based on genotyping and analysis of genetic distance among technical replicates (ddRAD and EpiRAD libraries combined), the SNP error rate was estimated to be 1.2% (s.d. = 0.7%), well below previously published estimates of reduced representation sequencing data (Fig. 3.3) (Mastretta-Yanes et al. 2015; Recknagel et al. 2015; Dimond et al. 2017). In other words, genotyping was 98.8% accurate. Genetic distance analysis also indicated that resampling of corals from 2015 to 2016 was accurate and no errors (e.g., sampling, misidentification) were made, showing SNP calling error comparable to the SNP error rate reported above for technical replicates (1.7%, s.d. = 0.9%). Chimerism (within-colony genetic variation resulting from fusion of juvenile colonies) and mosaicism (within-colony genetic variation arising from somatic mutations) is also not uncommon among scleractinian corals (Schweinsberg et al. 2015), and this analysis shows no evidence for these phenomena among the colonies sampled here. By contrast, genetic distance among all non-replicate and non-resampling pairwise comparisons between individuals was much greater, averaging 16.3% (s.d. = 1.4%). No clones were identified.

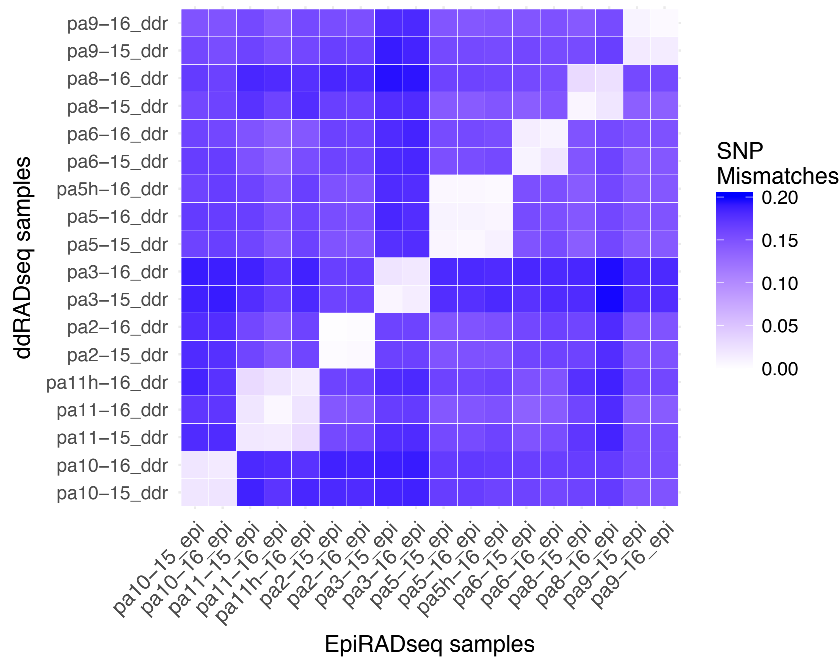

Figure 3.3. Pairwise SNP mismatches between the *P. astreoides* samples. EpiRADseq samples are shown along the x-axis, while ddRADseq samples are shown along the y-axis. Values along the diagonal illustrate the SNP error rate using the two libraries as technical replicates, while off-diagonal values show the error associated with resampling corals from 2015 to 2016. Two controls that were left at their site of origin (pa11h and pa5h) are included.

Epigenetic analysis

Across all samples and years, an average of 18.6% (s.d. = 0.9%) of CpGs were methylated (Fig. 3.4). As indicated by the low variance, most loci were either methylated or unmethylated across all samples and years; 73% of loci were constitutively unmethylated across samples and years, 12% were constitutively methylated, and 15% were differentially methylated (Fig. 3.4). The low SNP error rate reported above also adds confidence to the epigenetic analysis as it indicates that reads were correctly assigned to consensus loci at the level of each sample.

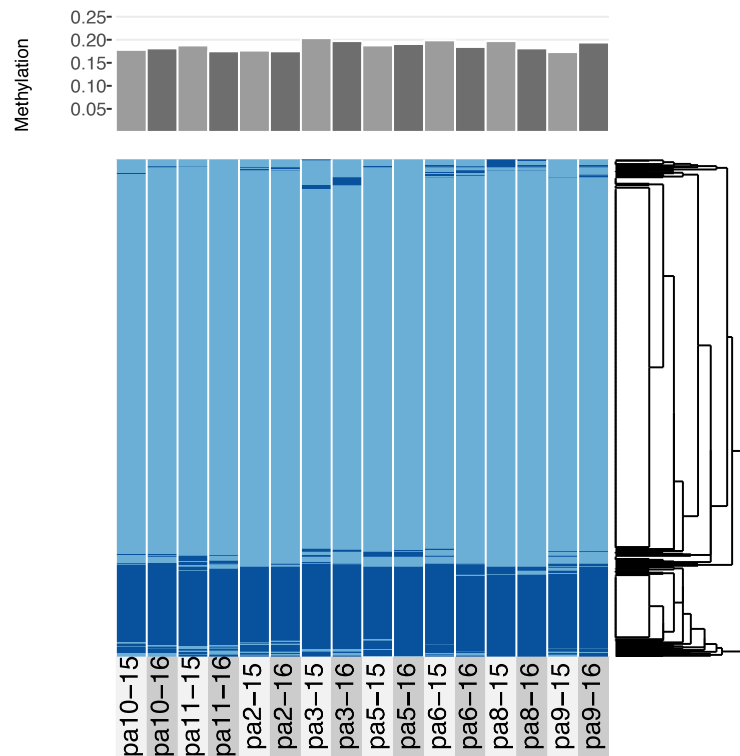

Figure 3.4. Comparative methylation profiles of the eight *P. astreoides* samples from 2015 to 2016. Every two vertical bars is a single individual, with the first bar showing 2015 methylation and the second showing 2016 methylation. Top panel shows proportion of methylated loci, while bottom panel shows a heatmap with methylated loci in dark blue and unmethylated loci in light blue, ordered via hierarchical clustering.

All corals underwent some degree of change in their methylation status, including the two controls (Fig. 3.5A). The mean percentage of loci changing methylation state per colony was 2.0% (s.d. = 0.9%). Among the two colonies with controls, the percentage of loci that changed methylation state was equivalent for colony pa11 and its control (2.9%), while for colony pa5, percent change was 1.0% and 1.7% for the common garden and control, respectively. Although there was no significant overall change in percent CpG methylation from 2015 to 2016 (paired t-

test, df = 7, p = 0.511; Fig. 3.5B), there was evidence for convergence of the methylation status among colonies toward a more similar methylome after a year in the common garden together, as pairwise differences between colonies were significantly lower in 2016 (paired t-test, df = 27, p < 0.001; Fig. 3.5C).

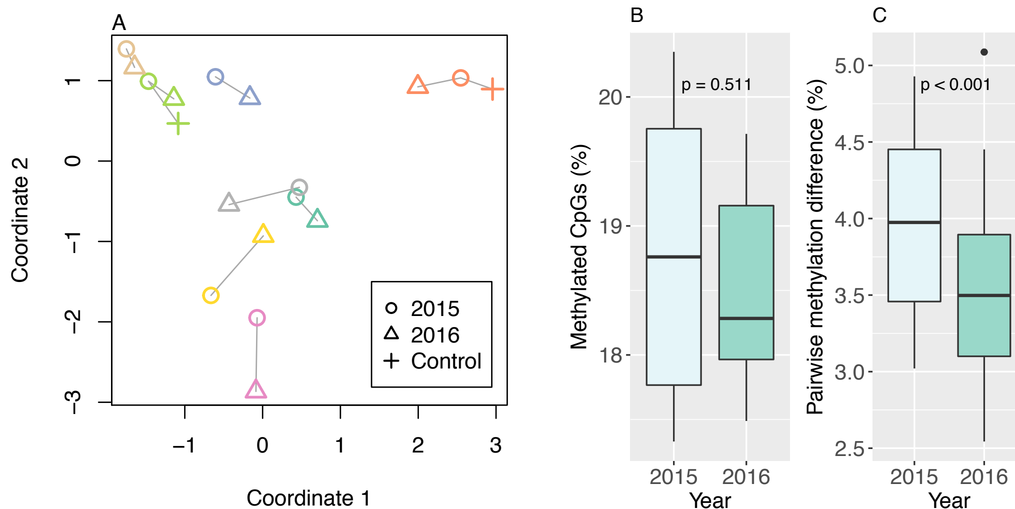

Figure 3.5. Methylation changes from 2015 to 2016. (A) Multidimensional scaling plot of methylation. Individual colonies are shown with different colors, and vectors show relative direction and magnitude of change from 2015 to 2016, including two controls that were left at their site of origin. (B) Percent methylated CpGs by year. (A) Pairwise methylation difference by year, including all pairwise comparisons between colonies.

Of the differentially methylated loci, most (76 of 95 = 80%) were differentially methylated only in a single individual colony. To further evaluate differentially methylated loci with some degree of confidence, only loci differentially methylated in more than one individual were considered further. Nineteen loci (19 of 95 = 20%) met this criterion and eight returned BLASTN hits. All matches were coral RNA sequences; six were mRNA sequences, while two were non-coding (nc) RNA sequences (Table 3.1). Four of the sequences had predicted products,

including a protein kinase, a helicase-like ribonucleoprotein, a FAM98A-like protein, and a tax-1 binding protein homolog.

Table 3.1. BLAST hits for select differentially methylated loci. See text for details.

Locus ID	Description	E value	GenBank Accession
20278	PREDICTED: Orbicella faveolata uncharacterized LOC110068412, mRNA	1.0E-12	XM_020775789.1
36065	PREDICTED: Orbicella faveolata protein FAM98A-like , mRNA	1.0E-06	XM_020762555.1
50692	PREDICTED: Acropora millepora tax1-binding protein 1 homolog (LOC114946951), mRNA	2.0E-17	XM_029323598.1
51711	PREDICTED: Acropora digitifera uncharacterized LOC107337846 (LOC107337846), mRNA	5.0E-12	XM_015903081.1
66685	PREDICTED: Orbicella faveolata U5 small nuclear ribonucleoprotein 200 kDa helicase-like, partial mRNA	1.0E-13	XM_020757069.1
69826	PREDICTED: Orbicella faveolata uncharacterized LOC110057897, transcript variant X5, ncRNA	2.0E-16	XR_002297590.1
76792	PREDICTED: Pocillopora damicornis uncharacterized LOC113678647, ncRNA	1.0E-05	XR_003446911.1
82175	PREDICTED: Orbicella faveolata calcium-independent protein kinase C-like, mRNA	1.0E-18	XM_020748568.1

3.5 DISCUSSION

Our analysis indicates that transplantation to a new environment for a one-year period did not elicit widespread changes in the methylome of *P. astreoides*. Yet, there was also evidence that corals converged toward similar methylomes in the common garden. This study indicates that DNA methylation in corals is responsive to environmental change, while also showing that the magnitude of methylation change may be subtle. Similarly, in a 3-month reciprocal transplant study of *Acropora millepora* on the Great Barrier Reef, Dixon et al. (2018) found that DNA methylation was considerably less responsive than gene expression to transplantation. These relatively slight methylation changes, however, were correlated with measures of physiological competency in novel environments, with transplanted individuals whose methylomes became more similar to those of native individuals showing more robust physiological profiles (Dixon et al. 2018). The small yet significant convergence of methylation patterns observed after one year in the common garden in our study is in agreement with the results of Dixon et al. (2018).

Responsiveness of DNA methylation to changing environmental conditions is not a foregone conclusion, as shown by Putnam et al. (2016) in their study of global methylation responses of two coral species to low pH. The relatively environmentally sensitive coral *Pocillipora damicornis* exhibited significant changes in methylation, while the more robust coral *Montipora capitata* did not. Interestingly, whereas *P. damicornis* performed poorly and showed limited evidence of acclimatization to the 6-week exposure to experimental conditions, *M. capitata* performed well. This suggests that changes in methylation are not necessarily associated with acclimatization. However, only bulk changes in methylation were quantified by Putnam et al. (2016), so it is possible that individual genes in *M. capitata* underwent both increases and decreases in methylation to contribute to homeostasis. Indeed, this is the scenario reported by Dixon et al. (2018) and in our study. Here, approximately 18.5% CpG methylation was maintained before and after transplantation despite an average of 2% of loci changing their methylation state, reflecting a combination of increases and decreases in methylation among individual colonies and loci. Likewise, Dixon et al. (2018) reported no genome-wide increases or decreases in methylation in response to transplantation, but instead that methylation changed in a so-called "seesaw" pattern whereby changes in methylation among hypomethylated genes were mirrored by changes in the opposite direction among hypermethylated genes. Their study, along with a study by Liew et al. (2018) on pCO_2-mediated changes in methylation and phenotype in *Stylophora pistillata*, have concluded that environmentally induced changes in DNA methylation are associated with homeostatic regulation. The precise mechanisms of this regulation, however, will require further study to fully understand.

The relatively small changes in methylation we detected here could be at least partially attributed to the limited difference in native habitat environmental conditions relative to the

common garden. Mean temperature among collections sites and the common garden varied only slightly, while temperature range and chlorophyll concentration were somewhat more variable. It is noteworthy, however, that Dixon et al. (2018) measured relatively small changes in methylation despite much greater environmental differences between transplantation habitats, albeit with only a three-month experimental duration. In addition, it is worth noting that colony pa9 exhibited some of the largest changes in methylation despite being collected within 500m of the common garden. This could suggest that some of the methylation change was unrelated to environmental change, or that habitat microhabitat was a factor. For example, small-scale variation in water flow conditions or shading may be biologically significant (e.g. Anthony and Hoegh-Guldberg 2003). On the other hand, several studies have noted considerable variance in methylation patterns that is incompletely explained by genetics and known environmental conditions (Dimond et al. 2017; Durante et al. 2019; Rondon et al. 2017).

A further consideration with respect to whether methylation changes were truly responding to the common garden environment is that changes in methylation among the two control colonies that remained at their site of origin were similar in magnitude to the experimental colonies. This could suggest that methylation simply changes as corals age, or that the experimental act of transplantation (halving colonies and reattaching them) caused methylation changes. Age-related changes in methylation, for example, are well-documented among humans and other vertebrates (Horvath 2013), and there is also some evidence among invertebrates (Lian et al. 2015). However, the strongest evidence that methylation was in fact responding to the common garden environment is the significant reduction in pairwise differences in colony methylation profiles after a year in the common garden together.

Although lack of an annotated *P. astreoides* genome limited the scope of our analysis and the inferences we were able to draw regarding potential functional implications of changes in methylation, a handful of loci exhibiting consistent differences in methylation provide some clues. Most of these loci were apparently associated with coding sequences, which is consistent with gene body methylation as the primary form of methylation among invertebrates (Sarda et al. 2012). One locus was associated with a sequence coding for a calcium-independent protein kinase C-like. Calcium-independent protein kinase C is involved in intracellular signaling, a biological process that is typically associated with the hypomethylated fraction of the genome (Dimond and Roberts 2016). By contrast with hypermethylated housekeeping genes that tend to exhibit consistent expression across conditions and tissues, hypomethylated genes such as those involved in cell-cell signaling are characterized by their inducibility in response to environmental change (Dimond and Roberts 2016).

A gene encoding a putative tax1-binding protein homolog was another differentially methylated locus. These proteins are involved in negative regulation of apoptotic processes via negative regulation of NF-κB transcription factor activity. Interestingly, deregulation of host NF-κB is associated with dinoflagellate symbiosis in cnidarians (Mansfield et al. 2017); loss of symbionts (bleaching) is associated with elevated levels of NF-κB. Perhaps differential methylation of a tax1-binding protein gene was associated with symbiotic homeostatic maintenance in the new environment.

Another locus resembled a gene encoding a U5 small nuclear ribonucleoprotein 200 kDa helicase, which is involved in mRNA splicing via its role in the spliceosome. Epigenetic factors are widely implicated in alternative mRNA splicing, which is a major source of protein diversity,

and hence, phenotypic variation (Luco et al. 2011). DNA methylation itself has been identified as a modulator of alternative splicing (Lev Maor et al. 2015).

The last mRNA-associated locus coded for a putative FAM98A protein. These proteins have numerous associated biological processes, including positive regulation of cell proliferation and gene expression. However, the most intriguing function involves protein methylation. A study of human colorectal cancer found that FAM98A was required for expression of an arginine methyltransferase (Akter et al. 2017). Protein methylation has been widely studied in histones, and there is ample evidence for epigenetic crosstalk between methylated DNA and methylated histones (Du et al. 2015).

Further evidence of epigenetic crosstalk in the loci responsive to transplantation is suggested by two loci associated with putative non-coding RNAs (ncRNAs). Along with DNA methylation and histone modifications, ncRNAs are considered part of the epigenetic machinery. ncRNAs are considerably more abundant than mRNAs and play numerous roles, particularly gene regulation (Kornienko et al. 2013). In some cases, they are involved in directing DNA and histone methylation patterns (Miska and Ferguson-Smith 2016).

Conclusion

This work shows that DNA methylation is an environmentally responsive epigenetic process, and is consistent with its putative role in acclimatization. We were able to detect subtle changes in *P. astreoides* methylation associated with experimental transplantation. Loci responding to transplantation were associated with signaling, apoptosis, gene regulation and epigenetic crosstalk, yet much remains to be learned about the functional implications of methylation changes in these differentially methylated genes. This study helps set the stage for

future work on potential roles of epigenetic mechanisms in the molecular acclimatization processes of reef corals.

3.6 LITERATURE CITED

Akter, K.A., Mansour, M.A., Hyodo, T. and Senga, T. 2017. FAM98A associates with DDX1-C14orf166-FAM98B in a novel complex involved in colorectal cancer progression. The International Journal of Biochemistry & Cell Biology 84, 1–13.

Anthony, K.R.N. and Hoegh-Guldberg, O. 2003. Variation in coral photosynthesis, respiration and growth characteristics in contrasting light microhabitats: an analogue to plants in forest gaps and understoreys? Functional ecology 17(2), 246–259.

Aranda, M., Li, Y., Liew, Y.J., et al. 2016. Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. Scientific reports 6, 39734.

Barter, R.L. and Yu, B. 2018. Superheat: An R package for creating beautiful and extendable heatmaps for visualizing complex data. Journal of computational and graphical statistics : a joint publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America 27(4), 910–922.

Dimond, J.L., Gamblewood, S.K. and Roberts, S.B. 2017. Genetic and epigenetic insight into morphospecies in a reef coral. Molecular Ecology 26(19), 5031–5042.

Dimond, J.L. and Roberts, S.B. 2016. Germline DNA methylation in reef corals: patterns and potential roles in response to environmental change. Molecular Ecology 25(8), 1895–1904.

Dixon, G., Liao, Y., Bay, L.K. and Matz, M.V. 2018. Role of gene body methylation in acclimatization and adaptation in a basal metazoan. Proceedings of the National Academy of Sciences of the United States of America 115(52), 13342–13346.

Duncan, E.J., Gluckman, P.D. and Dearden, P.K. 2014. Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype? Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution 322(4), 208–220.

Durante, M.K., Baums, I.B., Williams, D.E., Kemp, D.W. and Vohsen, S. 2019. What drives phenotypic divergence among coral clonemates of *Acropora palmata* ? Molecular Ecology.

Du, J., Johnson, L.M., Jacobsen, S.E. and Patel, D.J. 2015. DNA methylation pathways and their crosstalk with histone methylation. Nature Reviews. Molecular Cell Biology 16(9), 519–532.

Eaton, D.A.R. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30(13), 1844–1849.

Eirin-Lopez, J.M. and Putnam, H.M. 2019. Marine Environmental Epigenetics. Annual review of marine science 11(1), 335–368.

Flores, K., Wolschin, F., Corneveaux, J.J., Allen, A.N., Huentelman, M.J. and Amdam, G.V. 2012. Genome-wide association between DNA methylation and alternative splicing in an invertebrate. BMC Genomics 13, 480.

Gates, R.D. and Edmunds, P.J. 1999. The physiological mechanisms of acclimatization in tropical reef corals. American Zoologist 43, 30–43.

Hofmann, G.E. 2017. Ecological epigenetics in marine metazoans. Frontiers in Marine Science 4.

Horvath, S. 2013. DNA methylation age of human tissues and cell types. Genome biology, 14(10), 3156.

Iwasaki, M. and Paszkowski, J. 2014. Epigenetic memory in plants. The EMBO Journal 33(18), 1987–1998.

Kinoshita, T. and Seki, M. 2014. Epigenetic memory for stress response and adaptation in plants. Plant & Cell Physiology 55(11), 1859–1863.

Kornienko, A.E., Guenzl, *P*.M., Barlow, D.P. and Pauler, F.M. 2013. Gene regulation by the act of long non-coding RNA transcription. BMC Biology 11, 59.

LaJeunesse, T.C., Parkinson, J.E., Gabrielson, P.W., et al. 2018. Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. Current Biology 28(16), 2570-2580.e6.

Lev Maor, G., Yearim, A. and Ast, G. 2015. The alternative role of DNA methylation in splicing regulation. Trends in Genetics 31(5), 274–280.

Lian, S., He, Y., Li, X., Zhao, B., Hou, R., Hu, X., Zhang, L., and Bao, Z. 2015. Changes in global DNA methylation intensity and DNMT1 transcription during the aging process of scallop *Chlamys farreri*. Journal of Ocean University of China, 14(4), 685-690.

Liew, Y.J., Zoccola, D., Li, Y., et al. 2018. Epigenome-associated phenotypic acclimatization to ocean acidification in a reef-building coral. Science advances 4(6), 8028.

Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R. and Misteli, T. 2011. Epigenetics in alternative pre-mRNA splicing. Cell 144(1), 16–26.

Lynch, M. 2008. Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. Molecular Biology and Evolution 25(11), 2409–2419.

Mansfield, K.M., Carter, N.M., Nguyen, L., Cleves, P.A., Alshanbayeva, A., Williams, L.M., Crowder, C., Penvose, A.R., Finnerty, J.R., Weis, V.M. and Siggers, T.W. 2017. Transcription factor NF-κB is modulated by symbiotic status in a sea anemone model of cnidarian bleaching. Scientific reports, 7(1), 16025.

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T.H., Piñero, D. and Emerson, B.C. 2015. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Molecular ecology resources 15(1), 28–41.

Meerman, J. and Clabaugh, J. 2017. Biodiversity and Environmental Resource Data System of Belize. Online. http://www.biodiversity.bz Accessed on 8/24/2018.

Miska, E.A. and Ferguson-Smith, A.C. 2016. Transgenerational inheritance: Models and mechanisms of non-DNA sequence-based inheritance. Science 354(6308), 59–63.

NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group (2016) Moderate Resolution Imaging Spectroradiometer (MODIS) Ocean Color Data, NASA OB.DAAC. https://oceancolor.gsfc.nasa.gov/cgi/l3. Accessed on 8/24/2018.

Neri, F., Rapelli, S., Krepelova, A., et al. 2017. Intragenic DNA methylation prevents spurious transcription initiation. Nature 543(7643), 72–77.

Paradis, E., Claude, J. and Strimmer, K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20(2), 289–290.

Pearce, A., Faskel, F. and Hyndes, G. 2006. Nearshore sea temperature variability off Rottnest Island (Western Australia) derived from satellite data. International journal of remote sensing 27(12), 2503–2518.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. and Hoekstra, H.E. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. Plos One 7(5), e37135.

Putnam, H.M., Davidson, J.M. and Gates, R.D. 2016. Ocean acidification influences host DNA methylation and phenotypic plasticity in environmentally susceptible corals. Evolutionary applications 9(9), 1165–1178.

Recknagel, H., Jacobs, A., Herzyk, P. and Elmer, K.R. 2015. Double-digest RAD sequencing using Ion Proton semiconductor platform (ddRADseq-ion) with nonmodel organisms. Molecular ecology resources 15(6), 1316–1329.

Roberts, S.B. and Gavery, M.R. 2012. Is There a Relationship between DNA Methylation and Phenotypic Plasticity in Invertebrates? Frontiers in physiology 2, 116.

Rondon, R., Grunau, C., Fallet, M., et al. 2017. Effects of a parental exposure to diuron on Pacific oyster spat methylome. Environmental epigenetics 3(1), dvx004.

Sarda, S., Zeng, J., Hunt, B.G. and Yi, S.V. 2012. The evolution of invertebrate gene body methylation. Molecular Biology and Evolution 29(8), 1907–1916.

Schield, D.R., Walsh, M.R., Card, D.C., Andrew, A.L., Adams, R.H. and Castoe, T.A. 2016. EpiRADseq: scalable analysis of genomewide patterns of methylation using next-generation sequencing. Methods in ecology and evolution 7(1), 60–69.

Schweinsberg, M., Weiss, L.C., Striewski, S., Tollrian, R. and Lampert, K.P. 2015. More than one genotype: how common is intracolonial genetic variability in scleractinian corals? Molecular Ecology, 24(11), 2673-2685.

Smale, D.A. and Wernberg, T. 2009. Satellite-derived SST data as a proxy for water temperature in nearshore benthic ecology. Marine Ecology Progress Series 387, 27–37.

Todd, P.A. 2008. Morphological plasticity in scleractinian corals. Biological Reviews of the Cambridge Philosophical Society 83(3), 315–337.

UNEP-WCMC, WorldFish Centre, WRI, TNC. 2018. Global distribution of warm-water coral reefs, compiled from multiple sources including the Millennium Coral Reef Mapping Project. Version 4.0. Includes contributions from IMaRS-USF and IRD (2005), IMaRS-USF (2005) and Spalding et al. (2001). Cambridge (UK): UN Environment World Conservation Monitoring Centre. URL: http://data.unep-wcmc.org/datasets/1 Accessed on 8/24/2018.

Chapter 4. DNA METHYLATION PROFILING OF A CNIDARIAN-ALGAL SYMBIOSIS USING NANOPORE SEQUENCING

4.1 Abstract

Symbiosis with algae is common among cnidarians such as corals and sea anemones, and is associated with physiological changes that are in turn driven by changes in gene expression and, potentially, epigenetic processes such as DNA methylation. While it is still unclear exactly how methylation may interact with gene expression, there is evidence that methylation can influence transcript quality through regulation of exon splicing and alternative promoters. I leveraged the sensitivity to base modifications and full-length cDNA reads possible using nanopore sequencing with the Oxford Nanopore MinION to probe the effect of symbiosis on methylation and transcription in the sea anemone *Anthopleura elegantissima*. A draft genome comprising 243 Mb was first assembled using reads generated from aposymbiotic anemones. Symbiosis with the chlorophyte *Elliptochloris marina* was not associated with significant remodeling of the host methylome. However, I did identify three genomic regions with consistently different methylation according to symbiotic state. The region exhibiting the strongest difference was associated with a DNA polymerase zeta that is noted for its role in translesion synthesis, which opens interesting questions about the biology of this symbiosis. Targeted studies of the regions identified here may lead to new insights about this symbiosis and the role of DNA methylation.

4.2 Introduction

Endosymbiosis with algae is a widespread phenomenon among cnidarians inhabiting shallow tropical and temperate waters. The cnidarian-algal symbiosis is most notable as the trophic foundation of coral reef ecosystems, and plays a key role in the biology of scleractinian reef-building corals. For the host cnidarian, symbiosis can afford considerable nutritional benefits in the form of photosynthetically-fixed carbon that is translocated from the symbionts to the host cells in which they reside (Davy et al. 2012). Not surprisingly, symbiosis has physiological consequences for the host, many of which have been recently elucidated by studies of symbiosis-induced changes in host gene and protein expression (Rodriguez-Lanetty et al. 2006; Lehnert et al. 2014; Ganot et al. 2011; Oakley et al. 2016). These studies, involving a handful of facultatively symbiotic cnidarians capable of living naturally with and without symbionts, have revealed numerous affected pathways, including metabolism, membrane transport, cell adhesion, cell proliferation, cell-cell recognition, apoptosis, and oxidative stress (Rodriguez-Lanetty et al. 2006; Lehnert et al. 2014; Ganot et al. 2011; Oakley et al. 2016). Altogether, some 2-5% of the host transcriptome appears to be altered by symbiosis (Rodriguez-Lanetty et al. 2006; Lehnert et al. 2014), yet the processes ultimately governing these changes at the molecular level have been largely unexplored.

Epigenetic processes are often implicated as regulators of gene expression and phenotype, and include a suite of processes that modify the genome without changing base sequences. DNA methylation, which most commonly involves conversion of cytosine to 5-methylcytosine, has traditionally been the most heavily researched and relatively well-understood epigenetic modification (Duncan et al. 2014). DNA methylation in invertebrate genomes is relatively sparse, yet it occurs most prominently in actively transcribed genes, where

it is termed gene body methylation (Zemach et al. 2010; Sarda et al. 2012). While the function of gene body methylation remains a subject of study and debate, perhaps the most compelling evidence to date suggests that it plays a role in transcriptional fidelity by reducing spurious transcription (Neri et al. 2017). Indeed, other research suggesting that gene body methylation influences alternative splicing (Maunakea et al. 2013), alternative promoters (Maunakea et al. 2010), and the rate of transcriptional elongation (Lorincz et al. 2004) is not inconsistent with a role in transcriptional fidelity. Collectively, these studies suggest that DNA methylation influences when, where, and at what rate RNA polymerase should proceed. Because the majority of gene body methylation tends to occur in highly conserved housekeeping genes essential for cellular function (Sarda et al. 2012; Dimond and Roberts 2016), this supports the hypothesis that methylation is particularly important where transcriptional fidelity is essential.

To date, only one study has investigated DNA methylation in relation to the cnidarian-algal symbiosis. In their study of symbiotic and aposymbiotic *Exaiptasia pallida* sea anemones, Li et al. (2018) found support for a role of DNA methylation in maintaining transcriptional homeostasis under different symbiotic states. They found numerous symbiosis genes that were differentially methylated according to symbiotic state, and their data supported the hypothesis that DNA methylation reduces spurious transcription. In essence, Li et al. (2018) found that heavily methylated genes produced transcripts with less variable exon composition than more sparsely methylated genes.

Continued advances in genomic sequencing technologies are allowing ever-greater resolution of methylomes and transcriptomes. With the advent of nanopore sequencing, in which the base composition of full-length nucleic acid molecules passing through a nanopore is quantified using electrical currents, base modifications such as DNA methylation can be readily

detected without additional library preparation steps (Jain et al. 2016; Simpson et al. 2017). Moreover, the ability to sequence full-length molecules permits much greater resolution than short-read technologies for the detection of RNA splicing patterns. Together, these properties make nanopore sequencing a promising tool for functional genomics research on the potential interplay of the methylome and the transcriptome.

In this study, our goal was to identify differentially methylated regions associated with symbiosis and determine if methylation is associated with alternative RNA splicing. I used nanopore technology to investigate methylation and transcription in symbiotic and aposymbiotic *Anthopleura elegantissima* sea anemones. A common member of intertidal communities along the Pacific coast of North America, *A. elegantissima* is unique in its ability to exist either aposymbiotic or symbiotic with one (and occasionally both) of two distinct algal symbionts: the chlorophyte *Elliptochloris marina* or the dinoflagellate *Brevolium* (= *Symbiodinium*) *muscatinei* (Dimond et al. 2011). This facultative symbiosis makes *A. elegantissima* an excellent model system for research on cnidarian-algal symbiosis.

4.3 Methods

Specimen collection

At Lawrence Point, Orcas Island, Washington, USA, eight *A. elegantissima* were collected within an approximately 15 cm x 15 cm area from the underside of an intertidal boulder in June 2018. The individuals were chosen from within this small area because they likely represented a single clone, and yet individuals in two different symbiotic states could be seen. Four aposymbiotic individuals and four symbiotic individuals hosting the chlorophyte *Elliptochloris marina* were chosen; the symbiotic status of these sea anemones can be easily

discerned in the field based on color alone, since aposymbiotic individuals appear white and symbiotic animals are green (Figure 1). This sampling scheme was chosen to select two clearly different, naturally occurring phenotypes while minimizing environmental and genetic effects. Specimens were flash frozen in liquid nitrogen within one hour of collection. Just prior to freezing, tentacles excised from anemones were used to confirm the absence of *E. marina* in aposymbiotic animals using epifluorescence microscopy; red fluorescence from the chlorophyll-containing *E. marina* makes their presence readily apparent.

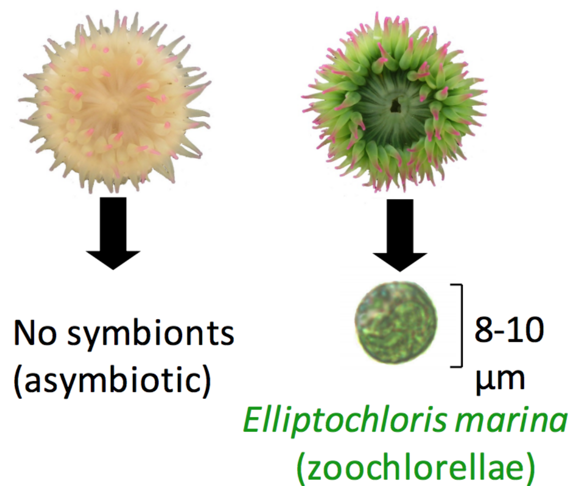

Figure 4.1. Symbiotic and aposymbiotic *A. elegantissima* are easily discernable based on the green color imparted by the symbionts.

DNA sequencing

Frozen sea anemone tissue was ground to a fine powder with a mortar and pestle in liquid nitrogen prior to genomic DNA extraction with the Qiagen DNeasy Blood and Tissue kit. The Qubit BR assay was used to asses DNA quantity, followed by 1% agarose gel electrophoresis to confirm the presence of high molecular weight DNA; DNA was estimated to be ≥ 10 kb. Four

DNA libraries were prepared for sequencing with the Oxford Nanopore Technologies (ONT) MinION platform. The MinION was connected to an Apple iMac with 1TB SSD, 32 GB RAM, 3.5 GHz Intel Core i7, and running MacOS Sierra 10.12.5 and MinKNOW 1.15.4. The first library was prepared with gDNA from a single aposymbiotic individual using the PCR-free, transposase-based ONT Rapid Sequencing Kit (SQK-RAD004) following manufacturer guidelines. Sequencing was performed on two FLO-MIN 106D R9.4 flow cells. Next, three libraries with one aposymbiotic individual and one symbiotic individual each (both barcoded) were prepared using the PCR-free ONT Ligation Sequencing Kit with the Native Barcoding Expansion Kit (SQK-LSK109 and EXP-NBD103), following the manufacturer's 1D native barcoding gDNA protocol. Each library was run on a separate FLO-MIN 106D R9.4 flow cell. Basecalling and demultiplexing was performed with ONT Albacore Sequencing Pipeline Software v. 2.3.3, and sequencing adapters were removed with Porechop v. 0.2.4.

RNA sequencing

The same six symbiotic and aposymbiotic individuals sequenced in the latter three DNA sequencing runs were also used for RNA sequencing. RNA was extracted from the remaining frozen sea anemone tissue powder used for DNA extraction using the Qiagen RNeasy Mini Kit with QIAshredder homogenization columns. RNA was quantified using the Qubit RNA BR assay prior to mRNA enrichment with the NEBNext Poly(A) mRNA Magnetic Isolation Module. Three poly(A) mRNA libraries were then prepared using the ONT cDNA-PCR Sequencing Kit with the PCR Barcoding Kit (SQK-PCS108 and SQK-PBK004), following the manufacturer's cDNA-PCR Barcoding protocol. As with the DNA libraries, each library included one aposymbiotic and one symbiotic individual, and libraries were each sequenced on a separate

FLO-MIN 106D R9.4 flow cell. Basecalling, demultiplexing, and adapter removal were performed as described above.

A. *elegantissima draft genome generation*

Using reads from aposymbiotic individuals only, a draft *A. elegantissima* genome was generated with wtdbg2, a *de novo* sequence assembly program designed for long noisy reads such as those produced by ONT sequencing (Ruan and Li 2019). Wtdbg2 first assembles raw reads without error correction, then builds a consensus from intermediate assembly output. ONT settings with default values were used.

Methylation analysis

DNA methylation patterns for each sequenced *A. elegantissima* individual were analyzed with Nanopolish v. 0.10.1 (Simpson et al. 2017). Nanopolish analyzes CpG motifs in short (11-34 bp) k-mer sequences and distinguishes 5-methylcytosine from unmethylated cytosine based on signal disruptions in raw ONT FAST5 sequence data. The draft *A. elegantissima* genome was used by Nanopolish for sequence mapping. Nanopolish output was summarized using the helper script calculate_methylation_frequency.py (Simpson et al. 2017), which summarizes methylation frequency by genomic coordinates. A threshold of at least 3 called sites per k-mer per sample was set to exclude low-coverage data. To investigate methylation patterns at individual genomic regions, pairwise t-tests, Cohen's d effect size tests, and absolute mean differences between aposymbiotic and symbiotic methylation were computed at each genomic position. The Benjamini-Hochberg procedure was applied to t-test p-values for false discovery rate (FDR)

correction at $\alpha = 0.05$. The Python tool methplotlib v. 0.1.1 was used to visualize regions of interest graphically and to further summarize data.

Transcriptome mapping

Nanopore RNAseq data were mapped to the genome subset using minimap2 v. 2.14 with ONT settings. Minimap2 specializes in spliced alignments of long, noisy reads. Additionally, an Illumina-based *A. elegantissima* transcriptome assembled by Kitchen et al. (2015) was mapped with BWA using the *mem* algorithm. Transcriptome data were also used to identify putative gene functions where differential methylation was observed, via web-based blastn searches against the nr database.

4.4 RESULTS

Sequencing

Sequencing statistics are shown in Table 1. The Rapid DNA sequencing protocol, in which a single aposymbiotic individual was sequenced, produced a total of 3.8 Gb from two flow cell runs. Mean read length from Rapid sequencing libraries (1,260 bp) was relatively short by comparison with read length obtained from the Ligation DNA sequencing protocol (overall mean 4,177 bp). Ligation sequencing libraries produced an average of 688 Mb per sample, with a maximum read length of 72,055 bp. As expected, and by contrast with all DNA libraries, cDNA libraries produced much shorter reads, averaging 623 bp. cDNA libraries produced an average of 1.4 million reads, or 927 Mb, per sample. Interestingly, there was a trend of more cDNA reads obtained from aposymbiotic versus symbiotic *A. elegantissima.*

Table 4.1. Sequencing statistics by sample. Most samples were sequenced in groups of two, and this is reflected in the library number. All libraries except the Rapid library were run on one flow cell each. The statistics reflect sequences that passed quality filtering only, and only barcoded sequences are reported for Ligation and cDNA-PCR libraries. The prefix 'A' under the sample category denotes an aposymbiotic sea anemone, whereas 'G' denotes a symbiotic anemone.

Sample	Library	Total reads	Total base pairs	Mean bp	Max bp	N50 bp
A2 DNA	Rapid (2 flow cells)	3,044,962	3,835,969,017	1,260	48,696	1,753
A4 DNA	Ligation #1	105,499	494,824,366	4,711	54,995	7,668
G2 DNA	Ligation #1	139,157	669,243,165	4,765	70,128	8,465
A3 DNA	Ligation #2	150,068	647,455,529	4,314	50,946	7,252
G4 DNA	Ligation #2	347,615	1,208,181,769	3,476	72,055	6,373
A1 DNA	Ligation #3	153,161	632,263,257	4,128	54,707	6,581
G3 DNA	Ligation #3	130,673	479,395,377	3,669	65,555	7,048
A3 RNA	cDNA-PCR #1	3,391,203	2,746,777,670	810	7,792	857
G2 RNA	cDNA-PCR #1	652,984	474,537,710	727	7,219	776
A4 RNA	cDNA-PCR #2	1,221,031	640,624,926	525	3,492	617
G3 RNA	cDNA-PCR #2	540,170	234,974,717	435	3,356	482
A1 RNA	cDNA-PCR #3	1,384,544	819,520,978	592	5,514	650
G4 RNA	cDNA-PCR #3	999,287	648,871,716	649	5,044	698

Draft genome

The wtdbg2-generated draft genome comprised 243 Mb, including 5359 contigs with an N50 of 87 kb and N90 of 19.2 kb. All aposymbiotic sequences were used to generate the draft genome, providing a total of 5.6 Gb and an estimated coverage of 23x.

Methylation analysis

After filtering for minimum read coverage, a total of 452,841 CpGs in 171,545 k-mers were analyzed for methylation, accounting for 7.8% of the 5,818,825 CpGs in the *A. elegantissima* draft genome. Because Nanopolish cannot resolve methylation of individual CpGs within ~6 bases each other, k-mers often included more than one CpG (mean = 2.6 CpGs per k-mer). Overall CpG methylation frequency was very low; mean methylation frequency was 0.055 (S.D. = 0.002) for aposymbiotic anemones and 0.059 (S.D. = 0.001) for symbiotic anemones. While there was a trend of higher methylation among symbiotic individuals, the difference was not significant (t-test, $p = 0.074$; Figure 2A). Principal components analysis of methylation patterns also did not reveal any clear differences between symbiotic states, although there was some separation of aposymbiotic and symbiotic individuals along PC2 (Figure 2B).

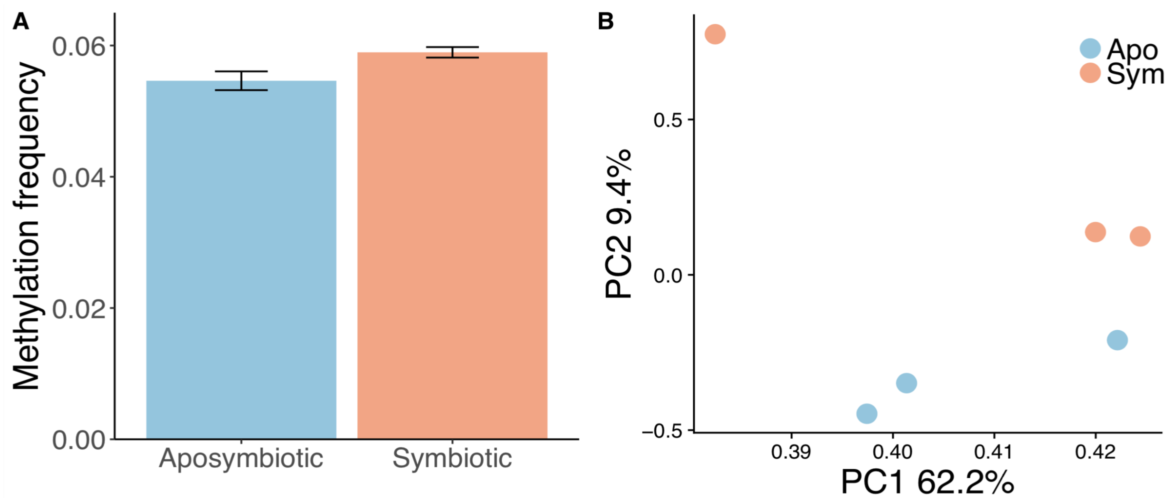

Figure 4.2. Comparative methylation frequency among aposymbiotic and symbiotic *A. elegantissima.* (A) Mean (± SE) methylation frequency. (B) Principal components analysis of methylation frequency.

At each genomic position, none of the t-tests passed FDR correction, so instead a Cohen's d value of 4 and an absolute methylation frequency difference of 0.4 were used to identify regions of interest. These criteria resolved 46 regions of interest, which were each viewed graphically for further evaluation. Of these 46 regions, three regions were identified that showed particularly strong and consistent differences in methylation between aposymbiotic and symbiotic *A. elegantissima.* On contig 4024, an approximately 5 kb region showed consistently higher methylation among aposymbiotic individuals (Figure 3). No ONT RNAseq data mapped to this region, however, transcripts from Kitchen et al. (2015) indicated that most of this region comprised several uncharacterized cnidarian genes as well as two transcripts showing hits to the peptidase A17 family of proteins.

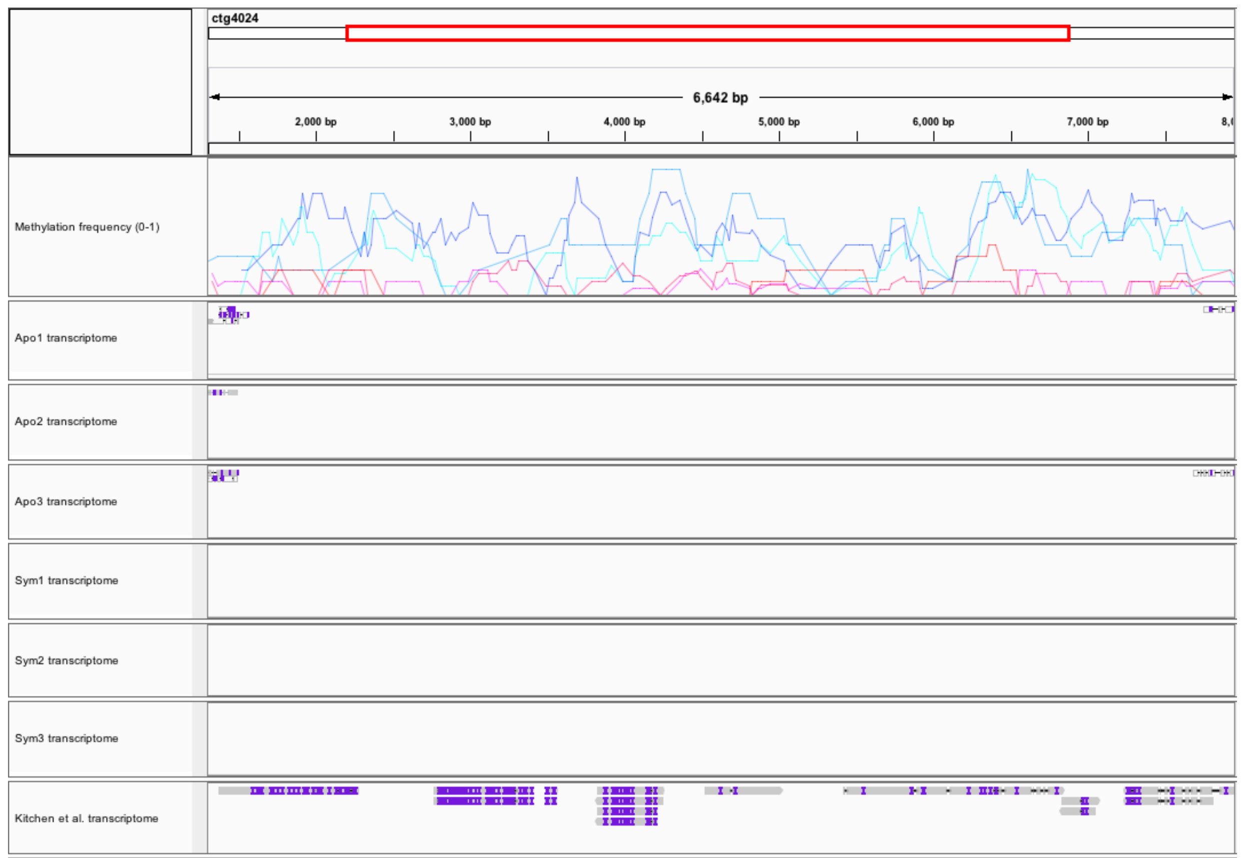

Figure 4.3. Methylation and transcription patterns associated with contig 4024. The top panel shows methylation frequency on a scale of 0-1; traces in blue hues represent aposymbiotic individuals, while red hues represent symbiotic individuals. The middle six panels depict ONT RNAseq reads (Apo = aposymbiotic, Sym = symbiotic). The last panel shows the transcriptome from (Kitchen et al. 2015).

Next, on contig 107, an approximately 2 kb region showed consistently higher methylation among aposymbiotic individuals (Figure 4). In this case, transcripts from both (Kitchen et al. 2015) and a limited number of ONT RNAseq transcripts indicated that this is a region of active transcription. The entire region, including the region several kb upstream of the differentially methylated region, was associated with a protein SFI1 homolog. The resolution of splicing patterns afforded by ONT RNAseq data were well illustrated in this region, yet the data

were too sparse to draw inferences about differences in transcript abundance or isoform composition between symbiotic states.

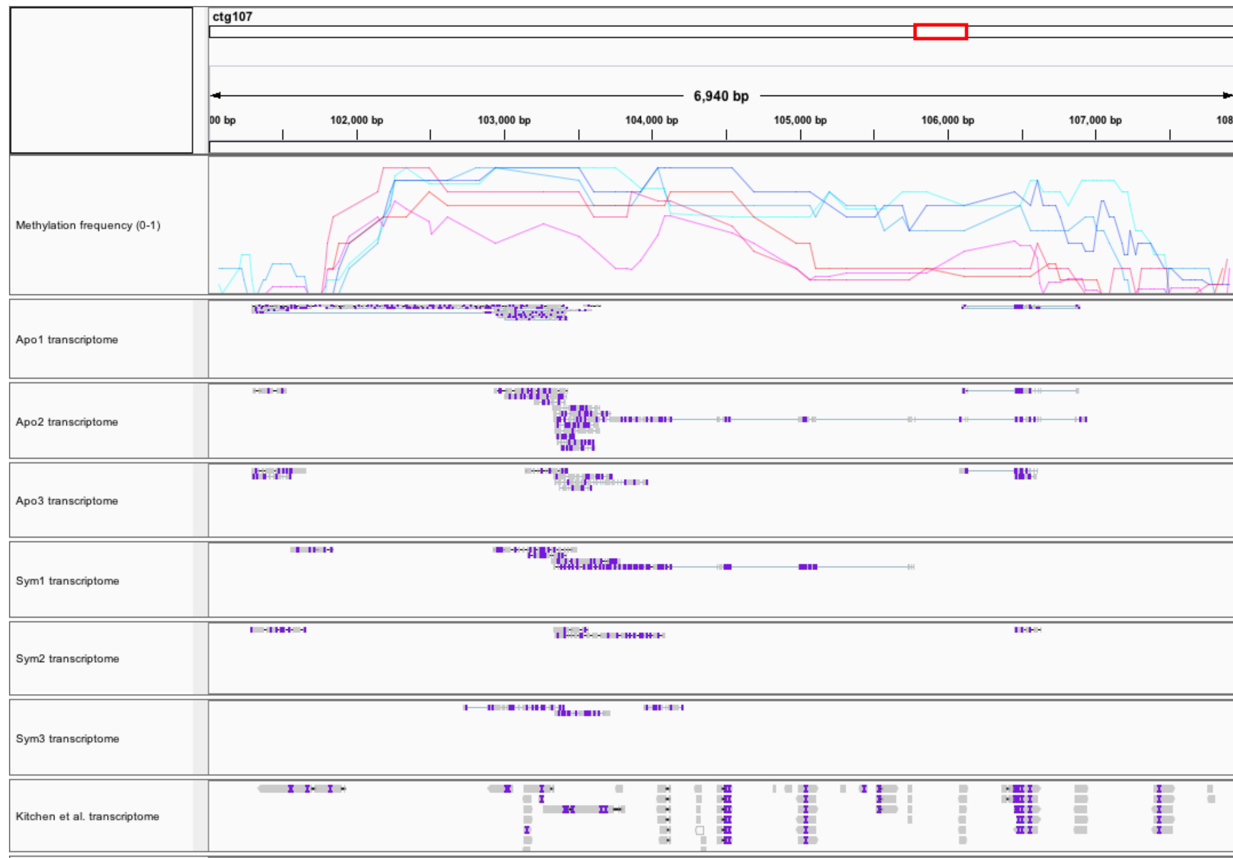

Figure 4.4. Methylation and transcription patterns associated with contig 107. The top panel shows methylation frequency on a scale of 0-1; traces in blue hues represent aposymbiotic individuals, while red hues represent symbiotic individuals. The middle six panels depict ONT RNAseq reads (Apo = aposymbiotic, Sym = symbiotic). The last panel shows the transcriptome from Kitchen et al. (2015).

Lastly, the most compelling differentially methylated region was found on contig 1212, in which an approximately 10 kb region showed consistently high methylation frequency among

symbiotic individuals, while methylation frequency among aposymbiotic individuals was near zero (Figure 5). This region had very few mapped ONT RNAseq reads, yet 21 distinct transcripts from Kitchen et al. (2015) were mapped. While eight of these transcripts were either uncharacterized or without hits, the remaining 13 were interspersed along the entire length of the differentially methylated region and were all associated with a DNA polymerase zeta catalytic subunit. This region was flanked on either side by regions of both high methylation and high expression among aposymbiotic and symbiotic *A. elegantissima.*

Figure 4.5. Methylation and transcription patterns associated with contig 1212. The top panel shows methylation frequency on a scale of 0-1; traces in blue hues represent aposymbiotic individuals, while red hues represent symbiotic individuals. The middle six panels depict ONT

RNAseq reads (Apo = aposymbiotic, Sym = symbiotic). The last panel shows the transcriptome from (Kitchen et al. 2015).

4.5 Discussion

Nanopore sequencing on the ONT MinION platform combined with the Nanopolish methylation detection algorithm (Simpson et al. 2017) provided high-quality DNA methylation data at near-base resolution. In addition, cDNA sequencing on the same platform resulted in many full-length or near full-length transcripts in which splicing patterns could be clearly discerned. Perhaps most importantly, we were able to generate a largely complete *A. elegantissima* genome with which to map gDNA and cDNA reads, and this genome is being used as part of a larger effort to develop a fully annotated, published genome. Based on the 23x coverage estimate as well as similar genome size estimates from another concurrent *A. elegantissima* genome effort, the draft genome generated here is likely to be largely complete. The concurrent *A. elegantissima* genome project, using Illumina short-read data, has reported a contig size of 240 Mb and a scaffold size of 316 Mb (Holland Elder, personal communication). A fully annotated, hybrid genome using both data sources is forthcoming (Holland Elder, personal communication).

Surprisingly, methylation levels and patterns did not differ greatly between aposymbiotic and symbiotic anemones hosting *E. marina.* This was unexpected, since Li et al. (2018) found somewhat more pronounced differences between symbiotic and aposymbiotic *E. pallida.* In their study, albeit with much greater genomic coverage (53x per individual) than our study, Li et al. (2018) found over two thousand genes (approximately 7.3% of *E. pallida* genes) exhibiting significant methylation differences according to symbiotic state. Aside from lower genomic

coverage in our study, two other factors may have contributed to the relatively subtle effects of symbiosis observed here. First, differences in experimental design may have been a factor. Whereas our study was a mensurative study of anemones in their natural state, Li et al. (2018) performed experimental bleaching and reinfection of anemones with symbionts. Secondly, in addition to obvious differences in host taxa between studies, the dinoflagellate symbiont of *E. pallida* in Li et al. (2018) is very different from the chlorophyte symbiont of *A. elegantissima* in our study. It is possible that symbiosis with *E. marina* does not elicit significant changes in host methylation because of its relatively limited contributions to host nutrition (Bergschneider and Muller-Parker 2008), or because it is more effectively able to evade detection by the host's immune system than dinoflagellate symbionts. For example, there is evidence that successful establishment of cnidarian symbiosis by highly compatible symbionts is associated with limited changes in gene expression relative to unsuccessful establishment by less compatible symbionts (Voolstra et al. 2009).

Although there was no evidence for substantial remodeling of the host methylome associated with symbiosis three regions were identified with consistently different methylation between aposymbiotic and symbiotic *A. elegantissima*. The region on contig 4024 was more strongly methylated among aposymbiotic anemones and was associated with a peptidase from the A17 family of proteins, which are homologous to aspartic proteinases encoded by retrotransposons. While gene bodies are the primary targets of methylation in invertebrate genomes, methylation of transposable elements does occur (Suzuki and Bird 2008). Methylation of transposable elements is well understood in plants, where it prevents transposition, but less is known about its function in animals (Suzuki and Bird 2008).

The region identified on contig 107 was associated with a protein SFI1 homolog and exhibited reduced methylation among symbiotic anemones. Protein SFI1 plays a role in cytoskeleton structure via centrosome-associated contractile fibers, and studies on yeast indicate that it is required for the transition from G_2 to M phase (Ma et al. 1999). In this context, it is noteworthy that the cnidarian-algal symbiosis is associated with a general deregulation of the host cell cycle (Rodriguez-Lanetty et al. 2006). Perhaps reduced methylation of this region in symbiotic anemones contributes to cell cycle deregulation.

By far the most compelling differentially methylated region was located on contig 1212, where much higher methylation among symbiotic individuals was consistent for a stretch of approximately 10 kb. This region was noted for multiple transcripts associated with a DNA polymerase zeta catalytic subunit. DNA polymerase zeta is a translesion synthesis polymerase that is able to bypass DNA lesions and complete transcription of damaged DNA, albeit at the cost of potential mutations (Gan et al. 2008). In yeast, for example, DNA polymerase zeta is critical for error-free replication past thymine glycol, a common DNA lesion created by free radicals (Johnson et al. 2003). Given that oxidative stress is a prominent physiological consequence of hosting algal endosymbionts (Weis 2008), this raises intriguing possibilities about how symbiosis might influence this process. Since methylation is generally associated with active transcription, which is highlighted by flanking regions, one hypothesis is that higher methylation of this DNA polymerase zeta among symbiotic anemones would be associated with higher transcription. That being said, it cannot be assumed that higher methylation is associated with higher transcription or exon inclusion, as the relationship between methylation and transcription is nuanced (Jones 2012) and the inclusion of exons can be either enhanced or suppressed by DNA methylation in a context-specific manner (Yearim et al. 2015). Future work

targeting methylation and transcription in this region more in-depth would be worthwhile and may lead to further insights.

Two recent studies with symbiotic cnidarians have found limited correlation between differentially expressed and differentially methylated genes. In their study of symbiosis in *E. pallida,* Li et al. (2018) found numerous differentially methylated and differentially expressed genes, yet very little overlap between the two sets of genes. Similarly, Dixon et al. (2018) found only weak correlation between changes in methylation and transcription at the individual gene level. Interestingly, however, both of these studies found associations between changes in methylation and transcription when considering gene functional groups or pathways rather than individual genes. Both studies concluded that there is some complementarity between methylation and transcription at a broad functional level, yet precise mechanisms remain a subject of speculation.

Conclusion

Nanopore sequencing using the ONT MinION platform has clear potential to be a powerful tool for investigations of DNA methylation and gene expression. Subtle differences in methylation associated with symbiosis in *A. elegantissima* were observed. While *E. marina* does not appear to strongly influence the host methylome, three genomic regions with clearly differential methylation were identified. In particular, the potential involvement of DNA polymerase zeta and translesion synthesis in this symbiosis deserves further attention, and may yield new insights into the cnidarian-algal symbiosis as well as the potential functions of DNA methylation.

4.6 LITERATURE CITED

Bergschneider, H. and Muller-Parker, G. 2008. Nutritional role of two algal symbionts in the temperate sea anemone *Anthopleura elegantissima* brandt. *The Biological bulletin* 215(1), 73–88.

Davy, S.K., Allemand, D. and Weis, V.M. 2012. Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiology and Molecular Biology Reviews* 76(2), 229–261.

Dimond, J.L., Bingham, B.L., Muller-Parker, G., Wuesthoff, K. and Francis, L. 2011. Seasonal stability of a flexible algal-cnidarian symbiosis in a highly variable temperate environment. *Limnology and oceanography* 56(6), 2233–2242.

Dimond, J.L. and Roberts, S.B. 2016. Germline DNA methylation in reef corals: patterns and potential roles in response to environmental change. *Molecular Ecology* 25(8), 1895–1904.

Dixon, G., Liao, Y., Bay, L.K. and Matz, M.V. 2018. Role of gene body methylation in acclimatization and adaptation in a basal metazoan. *Proceedings of the National Academy of Sciences of the United States of America* 115(52), 13342–13346.

Duncan, E.J., Gluckman, P.D. and Dearden, P.K. 2014. Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype? *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution* 322(4), 208–220.

Gan, G.N., Wittschieben, J.*P.*, Wittschieben, B.Ø. and Wood, R.D. 2008. DNA polymerase zeta (pol zeta) in higher eukaryotes. *Cell Research* 18(1), 174–183.

Ganot, *P.*, Moya, A., Magnone, V., Allemand, D., Furla, *P.* and Sabourault, C. 2011. Adaptations to endosymbiosis in a cnidarian-dinoflagellate association: differential gene expression and specific gene duplications. *PLoS Genetics* 7(7), e1002187.

Jain, M., Olsen, H.E., Paten, B. and Akeson, M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* 17(1), 239.

Johnson, R.E., Yu, S.-L., Prakash, S. and Prakash, L. 2003. Yeast DNA polymerase zeta (zeta) is essential for error-free replication past thymine glycol. *Genes & Development* 17(1), 77–87.

Jones, *P.*A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews. Genetics* 13(7), 484–492.

Kitchen, S.A., Crowder, C.M., Poole, A.Z., Weis, V.M. and Meyer, E. 2015. De novo assembly and characterization of four anthozoan (phylum cnidaria) transcriptomes. *G3 (Bethesda, Md.)* 5(11), 2441–2452.

Lehnert, E.M., Mouchka, M.E., Burriesci, M.S., Gallo, N.D., Schwarz, J.A. and Pringle, J.R. 2014. Extensive differences in gene expression between symbiotic and aposymbiotic cnidarians. *G3 (Bethesda, Md.)* 4(2), 277–295.

Li, Y., Liew, Y.J., Cui, G., et al. 2018. DNA methylation regulates transcriptional homeostasis of algal endosymbiosis in the coral model *Aiptasia*. *Science advances* 4(8), 2142.

Lorincz, M.C., Dickerson, D.R., Schmitt, M. and Groudine, M. 2004. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nature Structural & Molecular Biology* 11(11), 1068–1075.

Maunakea, A.K., Chepelev, I., Cui, K. and Zhao, K. 2013. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Research* 23(11), 1256–1269.

Maunakea, A.K., Nagarajan, R.*P.*, Bilenky, M., et al. 2010. Conserved role of intragenic DNA

methylation in regulating alternative promoters. *Nature* 466(7303), 253–257.

Ma, *P.*, Winderickx, J., Nauwelaers, D., et al. 1999. Deletion ofSFI1, a novel suppressor of partial Ras-cAMP pathway deficiency in the yeastSaccharomyces cerevisiae, causes G2 arrest. *Yeast* 15(11), 1097–1109.

Neri, F., Rapelli, S., Krepelova, A., et al. 2017. Intragenic DNA methylation prevents spurious transcription initiation. *Nature* 543(7643), 72–77.

Oakley, C.A., Ameismeier, M.F., Peng, L., Weis, V.M., Grossman, A.R. and Davy, S.K. 2016. Symbiosis induces widespread changes in the proteome of the model cnidarian *Aiptasia*. *Cellular Microbiology* 18(7), 1009–1023.

Rodriguez-Lanetty, M., Phillips, W.S. and Weis, V.M. 2006. Transcriptome analysis of a cnidarian-dinoflagellate mutualism reveals complex modulation of host gene expression. *BMC Genomics* 7, 23.

Ruan, J. and Li, H. 2019. Fast and accurate long-read assembly with wtdbg2. *BioRxiv*.

Sarda, S., Zeng, J., Hunt, B.G. and Yi, S.V. 2012. The evolution of invertebrate gene body methylation. *Molecular Biology and Evolution* 29(8), 1907–1916.

Simpson, J.T., Workman, R.E., Zuzarte, *P.*C., David, M., Dursi, L.J. and Timp, W. 2017. Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods* 14(4), 407–410.

Suzuki, M.M. and Bird, A. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews. Genetics* 9(6), 465–476.

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

1823–1833.

Weis, V.M. 2008. Cellular mechanisms of Cnidarian bleaching: stress causes the collapse of symbiosis. *The Journal of Experimental Biology* 211(Pt 19), 3059–3066.

Yearim, A., Gelfman, S., Shayevitch, R., et al. 2015. HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. *Cell reports* 10(7), 1122–1134.

Zemach, A., McDaniel, I.E., Silva, *P.* and Zilberman, D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328(5980), 916–919.

APPENDIX A

Preliminary results of the effects of a DNA methylation inhibitor on coral growth and bleaching during thermal stress

Introduction

Epigenetic processes influence gene regulation without affecting underlying DNA sequences, and are considered potential molecular mechanisms involved in phenotypic plasticity (Duncan et al. 2014). DNA methylation is a relatively well-known epigenetic process, and it most commonly involves the addition of a methyl group to a cytosine adjacent to a guanine. In plants and invertebrates, methylation is most prevalent in gene bodies, the transcribed regions of the genome (Zemach et al. 2010). Gene body methylation appears to have a range of functions, including regulating alternative splicing, repressing intragenic promoter activity, and reducing the efficiency of transcriptional elongation (Duncan et al. 2014). Given that these effects may influence phenotypic plasticity, there is growing interest in how DNA methylation may be involved in the response of corals to environmental change.

To date, only a handful of studies have investigated epigenetic processes in corals (Dixon et al., 2014, 2016, Dimond and Roberts 2016, Marsh et al. 2016, Putnam et al. 2016). DNA methylation patterns in coral transcriptomes have generally corroborated findings reported in studies of other invertebrate species (Gavery and Roberts, 2010, Sarda et al., 2012), showing that hypomethylated genes tend to be those with inducible functions (Dixon et al., 2014, Dimond and Roberts 2016, Marsh et al. 2016). Moreover, genes found to be differentially expressed in response to transplantation, elevated temperature, and reduced pH tend to have lower levels of methylation (Dixon et al., 2014, Dimond and Roberts 2016). It has been hypothesized that these

inducible, hypomethylated genes may be the most open to transient or conditional methylation, such as through environmental effects (Roberts and Gavery 2012). Recently, Putnam et al. (2016) demonstrated that differential genome-wide methylation of *Pocillopora damicornis* was associated with increased tolerance to reduced pH conditions.

Here I report a preliminary experiment testing inhibition of DNA methylation in corals using 5-aza-2′-deoxycytidine, also known as Decitabine. Once incorporated into DNA, this cytosine analog inhibits *de novo* methylation. I examined the effect of Decitabine on coral growth and bleaching under thermal stress, testing the hypothesis that inhibition of DNA methylation affects coral thermal tolerance plasticity.

Methods

The experiment was based at the Smithsonian Institution's Carrie Bow Cay Field Station (CBC), approximately 20 km off the Belize coast, situated adjacent to the Mesoamerican Barrier Reef. A 12-day factorial experiment was performed in the flow-through seawater aquaria system at CBC. For the experiment, I chose the branching coral *Porites porites*. I cut four replicate nubbins from each of four different colonies in the backreef (1-2 m depth) at Carrie Bow Cay. Nubbins were glued to small petri dish lids with A-788 compound, then buoyant-weighed and photographed. Coral nubbins were placed in independent glass beakers with 80 mL seawater. There were two water baths: one maintained at ambient seawater temperature for the first eight days of the experiment ("No Ramping"), and another with a heater that was used to raise the temperature approximately 0.3 ºC per day ("Ramping") (Fig. 1). For the final four days of the experiment, corals from the No Ramping treatment were placed in the Ramping treatment. The water levels of the water baths was maintained just below the tops of the coral beakers.

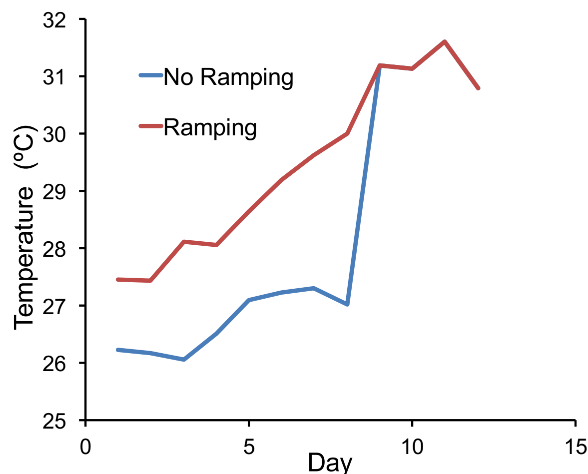

Figure 1. Daily mean temperature in the experimental treatments.

A stock solution of Decitabine at a concentration of 219 uM was made by dissolving 10 mg of the compound in 200 ml filtered seawater. This stock solution was kept refrigerated. For a final concentration of 10 uM in the experimental beakers, a dose which has been previously shown to disrupt DNA methylation in honeybees [20], I added 3.6 ml of the stock into beakers filled with 80 ml seawater. Water changes were made twice daily. The beakers were covered with a glass plate to reduce evaporation and rain water entry.

At the conclusion of the experiment, corals were again weighed and photographed. Percent weight change was computed, as well as percent bleaching, which was estimated by changes in coral color brightness with coral images measured in Image J software.

Results and discussion

All corals exhibited positive growth by the end of the experiment, but there were no significant differences in growth according to temperature or Decitabine treatment (ANOVA, p >

0.05; Fig. 2). In contrast, while corals in all treatments lost pigment during the experiment, there was a clear effect of Decitabine addition on coral bleaching, with Decitabine treatment associated with bleaching regardless of temperature treatment (ANOVA, p <0.001; Fig. 3).

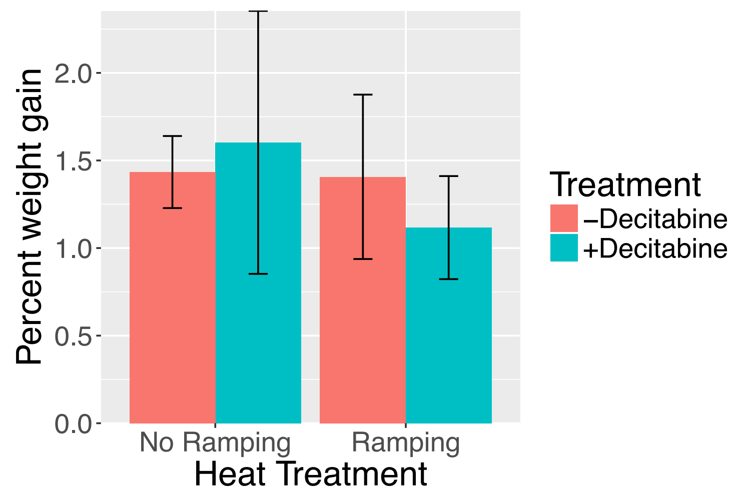

Figure 2. Percent weight gain (+/- std. error) of *Porites porites* nubbins in the experimental treatments.

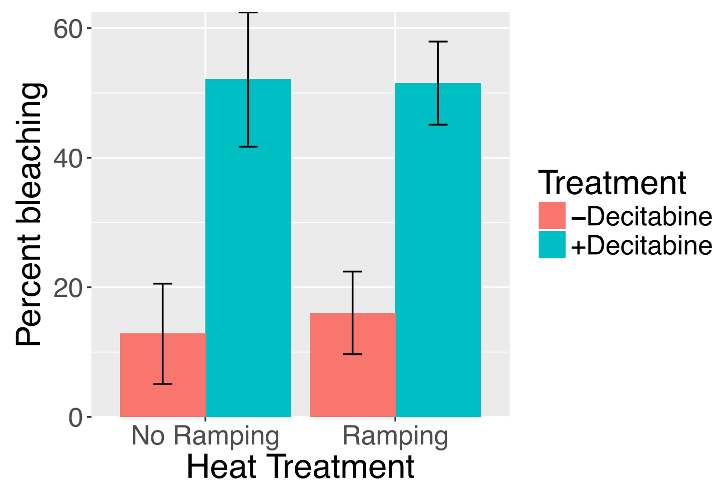

Figure 2. Percent bleaching (+/- std. error) of *Porites porites* nubbins in the experimental treatments.

This preliminary experiment demonstrates that addition of a DNA methylation inhibitor has a clear effect on *Porites porites.* Although it is possible that this effect reflects toxicity, the concentration used was the same used by Amarasinghe et al. (2014) in a study of bumble bee methylation. The authors argued that their results did not reflect toxicity and were well below levels found to be toxic in *Drosophila melanogaster.* In the present study, lack of a significant negative effect on coral growth argues against toxicity.

An alternative explanation for the bleaching effects is inhibition of *de novo* methylation in either the host, the symbionts, or both. The implication is that methylation inhibition disrupts symbiosis, perhaps particularly during thermal stress. Future work will be needed to evaluate this hypothesis, including determining actual changes in methylation. Ultimately, the goal of such experimental studies is to determine if DNA methylation plays a causative role in phenotypic plasticity, and especially in acclimative responses to environmental change that will be critical to corals' survival in coming decades.

Acknowledgements

This work was supported by a 2016 ISRS Graduate Fellowship.

Literature Cited

Amarasinghe HE, Clayton CI, Mallon EB (2014) Methylation and worker reproduction in the bumble-bee (*Bombus terrestris*). Proc R Soc B 281:20132502

Dimond JL, Roberts SB (2015) Germline DNA methylation in reef corals: patterns and potential roles in response to environmental change. Mol Ecol

Dixon GB, Bay LK, Matz MV (2014) Bimodal signatures of germline methylation are linked with gene expression plasticity in the coral *Acropora millepora*. BMC Genomics 15:1109

Dixon, G., Bay, L. K., & Matz, M. V. (2016). Genic DNA methylation drives codon bias in stony corals. Molecular Biology and Evolution, msw100.

Duncan EJ, Gluckman PD, Dearden PK (2014) Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype? J Exp Zool B Mol Dev Evol 322:208–220

Gavery MR, Roberts SB (2014) A context dependent role for DNA methylation in bivalves. Brief Funct Genomics elt054.

Marsh, A. G., Hoadley, K. D., & Warner, M. E. (2016). Distribution of CpG Motifs in Upstream Gene Domains in a Reef Coral and Sea Anemone: Implications for Epigenetics in Cnidarians. PloS one *11*, e0150840.

Putnam, H. M., Davidson, J. M., & Gates, R. D. (2016). Ocean acidification influences host DNA methylation and phenotypic plasticity in environmentally susceptible corals. Evolutionary Applications *9*, 1165-1178.

Roberts SB, Gavery MR (2011) Is there a relationship between DNA methylation and phenotypic plasticity in invertebrates? Front Physiol 2:116

Sarda S, Zeng J, Hunt BG, Yi SV (2012) The evolution of invertebrate gene body methylation. Mol Biol Evol 29:1907–1916

Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328:916–919

VITA

James Dimond obtained his B.S. in Natural Resources from the University of Vermont in 2002. He worked for the following year in the lab of Dr. Emily Carrington at the University of Rhode Island as a research assistant studying ecomechanics of byssal attachment in mussels. He continued to work in Dr. Carrington's lab for his M.S. on the ecology of symbiosis in the temperature coral *Astrangia poculata*, which he completed in 2006. From 2005-2007 he worked as a biological collector in the Marine Resources Division of the Marine Biological Laboratory in Woods Hole, MA. Following a year-long trip backpacking around the world in 2007-2008, he worked as a research assistant to Drs. Brian Bingham and Gisèle Muller-Parker from 2008-2013 on symbiosis in the sea anemone *Anthopleura elegantissima* at Western Washington University's Shannon Point Marine Center. He began his PhD work in the lab of Dr. Steven Roberts at the UW School of Aquatic and Fishery Sciences in 2014.