Comparative Profiling of coral symbiont communities from the Caribbean, Indo-Pacific, and Arabian Seas

Dissertation by

Chatchanit Arif

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

December 2014
In memory of Sugra Arif,
You are always in my heart
EXAMINATION COMMITTEE APPROVALS FORM

Committee Chairperson: Christian Voolstra
Committee Member: Uli Stingl
Committee Member: Timothy Ravasi
Committee Member: John Burt
ABSTRACT

Coral reef ecosystems are in rapid decline due to global and local anthropogenic factors. Being among the most diverse ecosystems on Earth, a loss will decrease species diversity, and remove food source for people along the coast.

The coral together with its symbionts (i.e. Symbiodinium, bacteria, and other microorganisms) is called the ‘coral holobiont’. The coral host offers its associated symbionts suitable habitats and nutrients, while Symbiodinium and coral-associated bacteria provide the host with photosynthates and vital nutrients. Association of corals with certain types of Symbiodinium and bacteria confer coral stress tolerance, and lack or loss of these symbionts coincides with diseased or bleached corals. However, a detailed understanding of the coral holobiont diversity and structure in regard to diseases and health states or across global scales is missing.

This dissertation addressed coral-associated symbiont diversity, specifically of Symbiodinium and bacteria, in various coral species from different geographic locations and different health states. The main aims were (1) to expand the scope of existing technologies, (2) to establish a standardized framework to facilitate comparison of symbiont assemblages over coral species and sites, (3) to assess Symbiodinium diversity in the Arabian Seas, and (4) to elucidate whether coral health states have conserved bacterial footprints.

In summary, a next generation sequencing pipeline for Symbiodinium diversity typing of the ITS2 marker is developed and applied to describe Symbiodinium diversity in corals around the Arabian Peninsula. The data show that corals in the Arabian Seas are dominated by a single Symbiodinium type, but harbor a rich variety of types in low abundant. Further, association with different Symbiodinium types is structured according to geographic locations. In addition, the application of 16S
rRNA gene microarrays to investigate how differences in microbiome structure relate to differences in health and disease demonstrate that coral species share common microbial footprints in phenotypically similar diseases that are conserved between regional seas. Moreover, corals harbor bacteria that are species-specific and distinct from the diseased microbial footprints. The existence of conserved coral disease microbiomes allows for cataloging diseases based on bacterial assemblage over coral species boundaries and will greatly facilitate future comparative analyses.
ACKNOWLEDGEMENTS

I would like to use this opportunity to express my gratitude to everyone who supported me through my PhD dissertation. I am especially indebted to King Abdullah University for the Graduate Fellowship I received, which granted me to pursue my education.

I would like to express my deepest gratitude towards my supervisor, Dr. Christian Voolstra, who believed in my potential and provided constant supervision, guidance, and constructive criticisms from beginning to end without fail. I would also like to extend my appreciation to my PhD committee members, Dr. Uli Stingl, Dr. Timothy Ravasi, and Dr. John Burt, for their invaluable advice and comments.

I am also grateful to Dr. Cornelia Roder for her reassurance and encouragement when I was vulnerable, Dr. Camille Daniels for her continual assistance, particularly, to ensure a smooth transition between my projects, Dr. Manuel Aranda, who were patient and who taught me many tricks in the lab, and finally, Dr. Till Bayer, who will always be my hero when it comes to bioinformatics.

This dissertation involved extensive trips to various countries for sample collections. I am very thankful to Dr. Voolstra, Dr. Roder, Dr. Ernesto Weil, Dr. Matthew Neave, and Chat Scott, who helped me with sample collections and were part of the adventures in different countries. I am also grateful to Dr. Suchana Chavanich and Dr. Weil who provided lab spaces during my sampling trips to Thailand and Puerto Rico.

Lastly, I will not have made it through without the supports from my family. I would like to thank my parents, Choomchet Arif and Sugra Arif, who always had faith in me. And most importantly, I am very grateful to my husband, Hussain Moiz, who was my bolster and cushion until I achieved the goal.
# TABLE OF CONTENTS

EXAMINATION COMMITTEE APPROVALS FORM ......................................................... 3
ABSTRACT .................................................................................................................. 5
ACKNOWLEDGEMENTS ......................................................................................... 7
TABLE OF CONTENTS ............................................................................................ 7
LIST OF ABBREVIATIONS ..................................................................................... 8
LIST OF FIGURES .................................................................................................... 13
LIST OF TABLES ...................................................................................................... 15

## INTRODUCTION ......................................................................................................... 16
  Coral Reef Ecosystems ............................................................................................... 16
  The Coral Holobiont .................................................................................................. 16
  Coral-associated photosynthetic dinoflagellates in the genus Symbiodinium .......... 17
  Coral-Associated Bacteria ....................................................................................... 23
  Coral Diseases ......................................................................................................... 23

## REFERENCES ............................................................................................................ 30

## OBJECTIVES ............................................................................................................. 36

## CHAPTER 1: ASSESSING SYMBIODINIUM DIVERSITY IN SCLERACTINIAN CORALS VIA NEXT-GENERATION SEQUENCING-BASED GENOTYING OF THE ITS2 rRNA REGION .......................................................... 38

### 1.1 ABSTRACT ......................................................................................................... 39

### 1.2 INTRODUCTION ............................................................................................... 40

### 1.3 MATERIALS AND METHODS ......................................................................... 43
  1.3.1 Sample collection and processing ................................................................. 43
  1.3.2 DGGE analysis of the ITS2 rDNA ................................................................. 44
  1.3.3 454 pyrosequencing of Symbiodinium ITS2 rDNA .................................... 46
  1.3.4 454 pyrosequencing data analysis .................................................. 47
  1.3.5 Single cell PCRs on Symbiodinium cells ............................................. 48

### 1.4 RESULTS ............................................................................................................ 49
  1.4.1 Sequence variation among ITS2 copies in the genomes of Symbiodinium ... 49
  1.4.2 Taxon-based analysis of Symbiodinium ITS2 diversity in isoclonal cultures 54
  1.4.3 Taxon-based analysis of ITS2 diversity in environmental samples .......... 55

### 1.5 DISCUSSION AND CONCLUSION ................................................................ 57
  1.5.1 Sequence variation among ITS2 copies in the genomes of Symbiodinium .... 57
  1.5.2 Taxon-based analysis of ITS2 diversity .................................................. 60
  1.5.3 Considerations for the application of an OTU-based pipeline to the analysis of Symbiodinium ITS2 diversity ................................................................. 63

### 1.6 TABLES AND FIGURES ................................................................................... 65

### 1.7 SUPPLEMENTARY INFORMATION ............................................................... 72

### 1.8 REFERENCES .................................................................................................. 82

## CHAPTER 2: NEXT GENERATION SEQUENCING-BASED ANALYSIS OF CORAL-ASSOCIATED SYMBIODINIUM DIVERSITY AROUND THE ARABIAN PENINSULA ......................................................... 86

### 2.1 ABSTRACT ......................................................................................................... 87

### 2.2 INTRODUCTION ............................................................................................... 88

### 2.3 MATERIALS AND METHODS ........................................................................ 91
  2.3.1 Sample Collection and environmental conditions ...................................... 91
  2.3.2 DNA extraction and 454 pyrosequencing ............................................. 92
CHAPTER 3: BACTERIAL PROFILING OF WHITE PLAGUE DISEASE (WPD) IN A COMPARATIVE CORAL SPECIES FRAMEWORK

2.3.3 Pyrosequencing data processing ................................................. 93
2.3.4 Statistical analysis ...................................................................... 94
2.4 RESULTS ...................................................................................... 95
  2.4.1 Pyrosequencing data ................................................................. 95
  2.4.2 Symbiodinium diversity around the Arabian Peninsula .......... 96
  2.4.3 PhyloChip-based analysis of coral-associated Symbiodinium in the Arabian seas 96
  2.4.4 OTU-based analysis of Symbiodinium diversity around the Arabian Peninsula ... 98
  2.4.5 OTU-based Symbiodinium diversity within a host species .......... 99
  2.4.6 OTU-based Symbiodinium diversity comparisons between Acropora, Porites, and Favia around the Arabian Peninsula .................................................. 100
2.5 DISCUSSION ............................................................................... 101
  2.5.1 PhyloChip-based vs. OTU-based analysis ................................... 102
  2.5.2 Symbiodinium distribution around the Arabian Peninsula ....... 103
  2.5.3 High diversity of Symbiodinium association within coral hosts .......... 104
  2.5.4 Thermally tolerant Symbiodinium types .................................. 105
  2.5.5 Effects of salinity on coral-Symbiodinium association .............. 108
2.6 CONCLUSION ............................................................................ 109
2.7 TABLES AND FIGURES ................................................................ 110
2.8 SUPPLEMENTARY INFORMATION .............................................. 120
2.9 REFERENCES ............................................................................. 120

CHAPTER 4: BACTERIAL PROFILING OF WHITE PLAGUE DISEASE ACROSS CORALS AND OCEANS INDICATES A CONSERVED AND DISTINCT DISEASE MICROBIOME

4.1 ABSTRACT .................................................................................. 150
4.2 INTRODUCTION .......................................................................... 151
4.3 MATERIAL AND METHODS ......................................................... 154
  4.3.1 Study site and sample collection .............................................. 154
  4.3.2 Sample processing and data generation .................................... 154
  4.3.3 Data analysis ........................................................................ 155
4.4 RESULTS .................................................................................... 157
  4.4.1 Bacterial richness in healthy and diseased corals ...................... 157
  4.4.2 Differentially abundant OTUs in healthy and diseased corals from the Caribbean ...... 158
4.4.3 Comparison of bacterial patterns in WPD from the Caribbean and the Indo-Pacific .......................................................... 159
4.4.4 Phylogenetic position of the coral host and bacterial community structure ...... 161
4.5 DISCUSSION ........................................................................................................................................ 162
  4.5.1 Bacterial richness in healthy and diseased corals ................................................................. 162
  4.5.2 Towards elucidation of bacterial disease footprints and microbiomes ........... 162
4.6 CONCLUSION ......................................................................................................................................... 165
4.7 FIGURES AND TABLES ............................................................................................................................ 166
4.8 SUPPLEMENTARY INFORMATION ........................................................................................................ 171
4.9 REFERENCES ....................................................................................................................................... 171
OVERALL CONCLUSION .......................................................................................................................... 174
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>KAUST</td>
<td>King Abdullah University of Science and Technology</td>
</tr>
<tr>
<td>SCUBA</td>
<td>Self-Contained underwater breathing apparatus</td>
</tr>
<tr>
<td>NGS</td>
<td>Next generation sequencing</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal ribonucleic acid</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction fragment length polymorphisms</td>
</tr>
<tr>
<td>DGGE</td>
<td>Denaturing gradient gel electrophoresis</td>
</tr>
<tr>
<td>ITS</td>
<td>Internal transcribed spacer</td>
</tr>
<tr>
<td>SML</td>
<td>Surface mucus layer</td>
</tr>
<tr>
<td>OTU</td>
<td>Operation taxonomic unit</td>
</tr>
<tr>
<td>BBD</td>
<td>Black band disease</td>
</tr>
<tr>
<td>WPD</td>
<td>White plague disease</td>
</tr>
<tr>
<td>WS</td>
<td>White syndrome</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>FSW</td>
<td>Filtered seawater</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>MUSCLE</td>
<td>Multiple sequence comparison by log-expectation</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic local alignment search tool</td>
</tr>
<tr>
<td>NAST</td>
<td>Nearest Alignment Space Termination</td>
</tr>
<tr>
<td>PERMANOVA</td>
<td>Permutational analysis of variance</td>
</tr>
<tr>
<td>ANOSIM</td>
<td>Analysis of similiarity</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>SIMPER</td>
<td>Similarity percentage analysis</td>
</tr>
<tr>
<td>MDS</td>
<td>Multi dimensional scaling</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>CAP</td>
<td>Canonical analysis of PCA</td>
</tr>
<tr>
<td>PCA</td>
<td>Principle component analysis</td>
</tr>
<tr>
<td>FDR</td>
<td>False discovery rate</td>
</tr>
<tr>
<td>SST</td>
<td>Sea surface temperature</td>
</tr>
<tr>
<td>Chl-a</td>
<td>Chlorophyll a</td>
</tr>
<tr>
<td>POC</td>
<td>Particulate organic carbon</td>
</tr>
<tr>
<td>PIC</td>
<td>Particulate inorganic carbon</td>
</tr>
<tr>
<td>PSU</td>
<td>Practical salinity units</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

FIGURE 1 THE EUKARYOTIC NUCLEAR RDNA ARRAYS. ETS= EXTERNAL TRANSCRIBED SPACER, ITS =INTERNAL TRANSCRIBED SPACER. .................................................. 19

FIGURE 1.1 (A) FREQUENCY DISTRIBUTION OF SEQUENCE READS FOR THE 10 MOST COMMON ITS2 COPIES FROM ISOCLOINAL CULTURE SAMPLES. (B) RAREFACT CURVE ILLUSTRATING ITS2 SEQUENCE DIVERSITY DISTRIBUTION IN ISOCLOINAL CULTURES. ........................................... 68

FIGURE 1.2 DENATURING GRADIENT GEL ELECTROPHORESIS (DGGE) FINGERPRINT OF SAMPLES USED IN THIS STUDY WITH CLADES/SUBCLADES INDICATED .................................. 69

FIGURE 1.3 REPRODUCIBILITY OF PYROSEQUENCING rDNA VARIANTS FOR REPLICATED PCRS (USING THE SAME DNA EXTRACT) OF ISOCLOINAL CULTURES CCMP 2467 AND RT-147 TAKING THE TOP 10 MOST ABUNDANT SEQUENCE VARIANTS INTO ACCOUNT. THE SIZES OF THE CIRCLES REPRESENT RELATIVE ABUNDANCE; CIRCLES WITH AN ASTERISK REPRESENT DISTINCT ITS2 VARIANTS. ........................................... 70

FIGURE 1.4 COMPARISON OF TOP 10 MOST COMMON VARIANTS FOUND FROM PYROSEQUENCING ITS2 FROM (A) CULTURE RT-152 (SYMBIODINIUM GOREAE OR TYPE C1) AND SEVERAL FIELD-COLLECTED SAMPLES REPRESENTING CLADE C SYMBIODINIUM; (B) C1C (SAMPLE P.WAJ.D7), (C) C1H (SAMPLE P.MAQ.R2.19), AND (D) C41 (SAMPLE A.AF.B6). THE SIZE OF THE CIRCLES REPRESENT RELATIVE ABUNDANCE, NUMBERS IN CIRCLES REPRESENT ORDER OF DISTINCT ITS2 VARIANTS FOR A GIVEN SAMPLE. SHARED ITS2 VARIANTS ARE DEPICTED IN A BLACK OUTER CIRCLE REPRESENTING ITS2 TYPE C1, AND A GREY OUTER CIRCLE REPRESENTING ITS2 SUBTYPE C. ........................................... 71

FIGURE 2.1 SEA SURFACE TEMPERATURE (SST) OBTAINED FROM NASA Aqua Moderate Resolution Images (MODIS) IN (A) WINTER (JAN-MAR) AND (B) SUMMER (JUL-SEP) OF 2013. ................................. 110

FIGURE 2.2 MAPS SHOWING SAMPLING SITES AND NUMBER OF CORAL SPECIMENS COLLECTED. SAMPLES FROM THE RED SEA WERE COLLECTED AT JANABU 23 (1), JANABU AYONA (2), KAUST INNA FASR (3), KAUST AL FAHAI (4), KAUST SHIB NAZAR (5), AL LITH SHALLOW REEF (6), AND AL LITH SOUTH REEF (7). COLLECTION SITES AT THE GULF INCLUDE AL YASSAT (8), RAS GHANADA (9), AND SAADHAYAT (10). SAMPLES FROM GULF OF OMAN WERE COLLECTED AT AL AQAH (11), DIBBA ROCK (12), DAMANYAT ISLANDS (13-15), FAHAI ISLAND (16), SAIFAT ASH SHEIKH (17), AND BANDAR KHAYRAN ISLET (18). .. 111

FIGURE 2.3 CUMULATIVE NUMBER OF CORAL GENERA COLLECTED FROM GULF OF OMAN (BLUE), THE GULF (GREEN) AND THE RED SEA (RED). TOTAL N=796. .................................................................................. 112

FIGURE 2.4 PIE CHARTS SHOW PERCENT DISTRIBUTION OF SYMBIODINIUM spp. IN EACH REGION VIA PHYLOTYPED-BASED ANALYSIS (A), AND OTU-BASED ANALYSIS (B). RARE VARIANTS AND RARE OTUS INDICATE CUMULATIVE PERCENTAGE OF SYMBIODINIUM TYPES/OTUS THAT ARE ≤3% ABUNDANT IN EACH REGION. C38* AND C3* ARE VARIANTS THAT ARE 96.91% AND 99.63% MATCH TO C38 AND C3 VIA BLASTN, RESPECTIVELY. ........................................................................... 113

FIGURE 2.5 PHYLOTYPED-BASED SYMBIODINIUM TYPES ASSOCIATED WITH CORAL GENERA FROM A) GULF OF OMAN, B) RED SEA AND C) THE GULF THAT REPRESENT ON AVERAGE AT LEAST 5% IN ABUNDANCE (OTHERS REPRESENT SYMBIODINIUM TYPES WITH AVERAGE PERCENTAGE BELOW 5%). ITS2 VARIANTS THAT DO NOT HAVE A 100% MATCH TO ANY SYMBIODINIUM spp. IN THE DATABASE ARE REPORTED AS THE CLOSEST SYMBIODINIUM TYPES IN TERMS OF PERCENT IDENTITY OBTAINED FROM BLASTN ANNOTATION. .......................................................... 114


FIGURE 2.7 CANONICAL ANALYSIS OF PCA CONDUCTED FROM BRAY-CURTIS SIMILARITIES OF OTU ABUNDANCE PATTERNS OF THE THREE CORAL GENERA i.e. ACROPORA spp. (Cross), PORITES spp. (Circle) AND Favia spp. (Triangle), COLLECTED FROM GULF OF OMAN (Blue), THE GULF (Green) AND THE RED SEA (Red), AND THEIR DEGREES OF CORRELATION WITH FIVE ENVIRONMENTAL VARIABLES, INCLUDING 5-YEAR ANNUAL AVERAGE OF CHLOROPHYLL A (CHL-A), PARTICULATE ORGANIC CARBON (POC), PARTICULATE INORGANIC CARBON (PIC), SEA SURFACE TEMPERATURE (SST), AND AN ANNUAL AVERAGE OF SALINITY. THE LENGTH AND DIRECTION OF VECTORS ILLUSTRATE THE STRENGTH AND DIRECTION OF CORRELATION WITH EACH AXIS. .................................................................................. 117
**Figure 3.1** Multidimensional (MD) scaling plot based on Bray-Curtis distances of normalized PhyloChip™ HybScores of healthy (circles) and diseased (triangles) specimens of the corals *P. duerdeni* (white) and *P. lutea* (black) illustrating the similarity of associated bacterial communities. Stress represents the goodness of fit of the data onto the MD ordination.

**Figure 4.1** Multidimensional scaling (MDS) plot derived from Bray-Curtis distances of normalized PhyloChip™ HybScores (*n* = 7,200 OTUs). Healthy (open) and WPD-affected (filled) specimens of the corals *P. duerdeni* (square), *P. lutea* (circle), *O. faveolata* (triangle), and *O. franksi* (diamond) are shown. Stress represents the goodness of fit of the data onto the MD ordination.

**Figure 4.2** Displayed are bacterial families and OTUs that showed a > 2-fold abundance difference between HH and DD over samples from *P. duerdeni* (white), *P. lutea* (light gray), *O. faveolata* (black), and *O. franksi* (dark gray).
LIST OF TABLES

| TABLE 1 | OVERVIEW OF COMMON CORAL DISEASES AND THEIR PATHOGENS. ..........................................................27 |
| TABLE 1.1 | SYMBIODINIUM ITS2 PYROSEQUENCING OVERVIEW. DEPICTED ARE CLADE TYPES (WHEN CULTURES WERE USED), THE RESPECTIVE ANIMAL HOST AND COLLECTION SITE, AND DGGE-TYPING RESULTS (BASED ON THE BRIGHTEST BAND). FOR ITS2 PYROSEQUENCING DATA, THE TOTAL NUMBER OF SEQUENCE READS (AFTER FILTERING) AND THE NUMBER OF DISTINCT ITS2 COPIES FOR EACH SAMPLE IS PROVIDED. SEQUENCE READS FOR EACH SAMPLE ARE FURTHER ASSIGNED TO CLADE TYPES DETAILING NUMBER OF SEQUENCE READS, NUMBER OF DISTINCT ITS2 COPIES WITHIN CLADES, AND CLADE TYPE OF MOST ABUNDANT ITS2 COPY. IN CASES WHERE THE CLADE TYPE OF THE MOST ABUNDANT ITS2 COPY IS DIFFERENT FROM THE DATABASE SEQUENCE, THE PERCENT IDENTITY IS INDICATED. IN CASES WHERE THERE ARE 2 OR MORE EQUALLY ABUNDANT ITS2 COPIES, ALL CLADE TYPES ARE INDICATED. ..........................................................65 |
| TABLE 1.2 | INTRAGENOMIC DIVERSITY OF ITS2 GENES IN ISOCLOINAL CULTURES (ONLY ITS2 COPIES REPRESENTED BY AT LEAST 100 READS WERE CONSIDERED). DIFFERENCES ARE CALCULATED AS UNCORRECTED PAIRWISE DISTANCES BETWEEN ALIGNED DNA SEQUENCES. THE DIFFERENCE BETWEEN THE MINIMUM AND MAXIMUM GENETIC DIFFERENCE OF ITS2 VARIANTS WITHIN CULTURES CAN BE USED TO DERIVE A SIMILARITY CUTOFF FOR AN OPERATIONAL TAXONOMIC UNIT (OTU)-BASED FRAMEWORK, UPON WHICH INTRAGENOMIC DIVERSITY IS CONTAINED WITHIN A GIVEN OTU. .............................................................. 66 |
| TABLE 1.3 | SYMBIODINIUM ITS2 OTU-BASED ANALYSIS OVERVIEW. DEPICTED ARE SAMPLES USED IN THIS STUDY AND THE DESIGNATED CLADE TYPE BASED ON DGGE-TYPING. FOR ITS2 PYROSEQUENCING DATA, THE TOTAL NUMBER OF SEQUENCE READS (AFTER FILTERING AND CLADE-BASED ALIGNMENT) AND THE NUMBER OF OTUS AT A 97% SIMILARITY CUTOFF FOR EACH SAMPLE IS PROVIDED. DISTRIBUTION OF SEQUENCE READS OVER OTUS ON A CLADE BASIS AND THE ITS2 CLADE TYPES MOST SIMILAR TO AN OTU ARE DETAILED. IN CASES WHERE THE OTU IS DIFFERENT FROM THE ITS2 CLADE TYPE DATABASE SEQUENCE, THE PERCENT IDENTITY IS INDICATED. .................................................................67 |
| TABLE 2.1 | THE TABLE SHOWS SAMPLING SITES, NUMBER OF SAMPLES, AND NUMBER OF CORAL GENERA COLLECTED FROM GULF OF OMAN, THE GULF AND THE RED SEA. PYROSEQUENCING READS ARE INDICATED AS WELL AS NUMBERS OF ITS2 VARIANTS AND NUMBER OF OTUS PER SITE PER CLADE. ..........................................................118 |
| TABLE 2.2 | PERMANOVA SUMMARY STATISTICS OF THE DIFFERENCES IN OTU ABUNDANCE BETWEEN REGIONS (GULF OF OMAN, RED SEA, AND THE GULF), CORAL GENUS (PORITES SPP., ACROPORA SPP., AND FAVIA SPP.), REEFS (NESTED WITHIN REGIONS), AND REGIONS CROSSED WITH CORAL GENUS, USING 999 PERMUTATIONS. .................................................................................................................................119 |
| TABLE 3.1 | NUMBER OF DISTINCT TAXONOMIC RANKS IDENTIFIED BY PHYLOCIPHTM IN COMPARISON TO CLONE LIBRARY SEQUENCING OF A POOL OF 96 CLONES FROM EACH SAMPLE (N = 477). .................................................................140 |
| TABLE 3.2 | NUMBER OF DETECTED OTUS OVER ALL SAMPLES WITH PHYLOCIPHTM MICROARRAYS. HH: HEALTHY, DD: DISEASED. ........................................................................................................................................140 |
| TABLE 3.3 | SUMMARY STATISTICS OF TWO-WAY CROSSED ANOSIM AND TWO-WAY ANOVA. .................................140 |
| TABLE 3.4 | OVER- OR UNDER-REPRESENTATION OF BACTERIAL FAMILIES OF OTUS DIFFERENTIALLY ABUNDANT BETWEEN CORAL SPECIES, AND CONGREGATED FOLD-CHANGE DIFFERENCES BETWEEN HEALTHY AND DISEASED SPECIMENS OF P. DUERDENI AND P. LUTEA (ONLY FAMILIES THAT WERE REPRESENTED BY AT LEAST 5 BACTERIAL TAXA WERE CONSIDERED). ..................................................................................................................142 |
| TABLE 3.5 | OVER- OR UNDER-REPRESENTATION OF BACTERIAL FAMILIES OF OTUS DIFFERENTIALLY ABUNDANT BETWEEN HEALTH STATES OF P. DUERDENI AND P. LUTEA, AND CONGREGATED FOLD CHANGE DIFFERENCES BETWEEN HEALTHY (HH) AND DISEASED (DD) SPECIMENS (ONLY FAMILIES THAT WERE REPRESENTED BY AT LEAST 5 BACTERIAL TAXA WERE CONSIDERED). .................................................................................................................................143 |
| TABLE 4.1 | OTU RICHNESS IS SHOWN FOR ALL FOUR SPECIES/CONDITION COMBINATIONS FROM THE CARIBBEAN (O. FAVEOLATA HH, O. FAVEOLATA DD, O. FRANKSI HH, O. FRANKSI DD) AND FROM THE INDO-PACIFIC (P. DUERDENI HH, P. DUERDENI DD, P. LUTEA HH, P. LUTEA DD). BACTERIAL TAXA WERE COUNTED PRESENT WHEN DETECTED IN 2 OUT OF 3 REPLICATES OF ANY SPECIES/CONDITION COMBINATION. ..................................................................................................................143 |
| TABLE 4.2 | SUMMARY STATISTICS OF TWO-WAY ANOVA SEPARATING SPECIES AND CONDITION EFFECTS BETWEEN O. FAVEOLATA AND O. FRANKSI. DATA BASED ON NORMALIZED HYBRIDSCORES OF 11,256 OTUS PRESENT OVER ALL CORAL SAMPLES. ..................................................................................................................146 |
| TABLE 4.3 | PERMANOVA SUMMARY STATISTICS FOR FACTORS AFFECTING BACTERIAL COMMUNITIES IN HEALTHY AND WPD-AFFECTED CORALS FROM THE CARIBBEAN (I.E. PUERTO RICO) AND THE INDO-PACIFIC (I.E. GULF OF THAILAND). DF = DEGREES OF FREEDOM. SS = SUM OF SQUARES. MS = MEAN SQUARES. N.S. = NOT SIGNIFICANT ..................................................................................................................146 |
INTRODUCTION

Coral Reef Ecosystems

Coral reefs are one of the most diverse ecosystems of the world (1, 2). They are habitat to a wide range of marine life (3), including nurseries for one third of the world’s marine fish species (4). Rich and diverse marine fish species make coral reef ecosystems a main source of income for over tens of millions of people that live along the coasts in more than hundred countries around the globe (5). In addition, marine tourism industries, which include activities such as SCUBA, snorkeling, and other coastal activities, attract numerous tourists each year. The estimated economic value of global coral reef ecosystems is on average about USD 353,000/hectare/year (6).

The structure of coral reefs is formed by the deposition of calcium carbonate during the growth of corals that is cemented by crustose coralline algae (CCA). This hard substrate does not only act as a wave breaker and a barrier to natural disasters such as tsunami and hurricanes, but is also important in the settlement process of new coral recruits (2).

A coral is a meta-organism in the sense that it is composed of various organisms in one entity, i.e. the coral animal host, single-celled photosynthetic algae in the genus *Symbiodinium*, and microorganisms such as bacteria, archaea, and viruses (7, 8). Together they are called the coral holobiont (7, 8). The photosynthesis performed by *Symbiodinium* helps corals survive in oligotrophic conditions.

With an increase in human population and frequent environmental disturbances, coral reefs are in decline. Overpopulation has led to urbanization, pollution, and over-use of resources, such as overfishing (9), which together threaten over 57% of the world’s reef ecosystems (10). Furthermore, changes in environmental conditions such as elevated seawater temperature (1, 11, 12) and ocean acidification (13), have led to
more incidences of coral bleaching and coral diseases worldwide (14, 15). As a consequence, one third of reef-building coral species are at risk of becoming endangered species (16). In fact, it is estimated that 30% of the world’s coral cover has been lost and that 60% may be lost by 2030 (9). Therefore, urgent conservation actions are needed to reduce further decline of coral reef ecosystems. Some of the actions include managing runoff systems (17), artificial reef construction (18), creating marine protected areas (MPAs), and promoting sustainable fishing (19).

**The Coral Holobiont**

Corals belong to the phylum Cnidaria. Cnidarians are diploblastic, simple animals that consist of endo- and ectoderm. Corals provide microhabitats, both internally (gastroderm, and skeleton) and externally (coral mucus), for microbiota from all three domains of life, i.e. eukaryotes, bacteria and archaea (7, 20). The coral as a host provides carbon dioxide, nitrogenous waste, and shelter to all of its symbionts (7). In exchange, the coral’s endosymbiotic dinoflagellates in the genus *Symbiodinium*, supply up to 95% of photosynthetically fixed carbon to satisfy the host’s respiratory requirements (21). As for coral-associated bacteria, various studies have elucidated their potential roles, including fixing nitrogen (22-25), decomposing organic materials (26), producing metabolites such as antibiotic compounds (27-29), and occupying the space in the coral to prevent colonization of pathogens (7, 30).

The coral is a complex holobiont, whose interactions between the different members are crucial. Disruption in the interactive role of even one of these parties often leads to the collective damage of the host and its symbionts (7). For instance, bleaching coral results in a dissociation between the coral host and its associated *Symbiodinium*, which are crucial to the host survival (31, 32). Another example is a shift in coral-
associated microbial communities when a coral is under stress, which may lead to
disease (33-36).

**Coral-associated photosynthetic dinoflagellates in the genus Symbiodinium**

The unicellular photosynthetic dinoflagellates in the genus *Symbiodinium* are
intracellular symbionts of many marine invertebrate including cnidarians, molluscs,
and some protists (37, 38). They are also found as free-living organisms in seawater
column. Corals have an obligatory symbiotic relationship to these symbionts since
corals depend on symbiont photosynthetic products in the form of sugar, amino acids,
and glycerol (39), (40). *Symbiodinium* also enhances calcification rates of
scleractinian corals (41). In turn, the coral host provides *Symbiodinium* with inorganic
materials such as ammonium and phosphate and also a suitable environment for them
to survive within the host tissue. *Symbiodinium* occur in high density within a coral
host and can reach up to $10^6$ cells/cm$^2$ of host tissue (42, 43).

*Symbiodinium microadriaticum* was once assumed to be the only species in the genus
*Symbiodinium* (44), because all observed *Symbiodinium* cells showed highly similar
morphologies as non-mobile yellow-brown coccoid cysts under a light microscope.
However, accumulative evidence, such as organisms’ behavior, biochemical
component, physiological characteristics and host infectivity, suggest that
*Symbiodinium* were more diverse than previously thought (reviewed by (45)). The
notion was confirmed with the advent of molecular technologies in the early 1990s
when Rowan and Powers (46) first examined *Symbiodinium* phylogenetic
relationships using restriction fragment length polymorphisms (RFLPs) and
sequencing of the small ribosomal subunit RNA (SSU) gene from various marine
invertebrate hosts. Their phylogeny identified divergent lineages within the genus
*Symbiodinium* that exhibited sequence differences comparable to that observed among
different families and orders of other dinoflagellate groups (47). However, the conserved SSU gene is slow evolving and hence did not provide a sufficient resolution to distinguish closely related *Symbiodinium* species (46, 48, 49).

Researchers then turned to less conserved, faster evolving DNA regions including the nuclear large subunit (LSU) (50, 51), the Internal Transcribed Spacer regions (ITS) (i.e. ITS1 and ITS2) (48, 52, 53), the chloroplast ribosomal large subunit (cp23S) (54, 55), and cytochrome oxidase-b (56). These markers differ in their ability to resolve *Symbiodinium* into ecologically distinct units depending on the level of conservation of each gene. Other markers, including the chloroplast Photosystem II protein D1 (psbA) non-coding region (57, 58), and microsatellites were also recently employed—often in combination with each other— to further improve genetic resolution and delimit species boundaries (59). To date, the genus *Symbiodinium* are classified into nine functionally distinct evolutionary entities or phylogenetic clades (A to I), with each clade containing multiple genetic varieties denoted as subclades or types. Among the markers, the ITS2 region has been the most commonly used DNA marker to assess *Symbiodinium* diversity from a diverse array of hosts and over large geographic distances (60-64).

![Diagram](image)

**Figure 1.** The eukaryotic nuclear rDNA arrays. ETS= external transcribed spacer, ITS = internal transcribed spacer.
*Symbiodinium* eukaryotic nuclear rDNA consists of the 18S, 5.8S and 28S, which are separated by ITS1 and ITS2 (Figure 1). The ITS2 region is a relatively less conserved and a faster evolving region in comparison to the 18S, 5.8S and 28S rDNA genes; hence, it enables the classification of *Symbiodinium* diversity beyond phylogenetic clade levels. This makes ITS2 region suitable for *Symbiodinium* diversity typing in various hosts in different locations. However, because eukaryotic ribosomal genes are arranged in tandem repeat arrays, there are many variants of ITS2 within each *Symbiodinium* cell or genome. In essence, concerted evolution acts to homogenize these ITS2 variants into a single dominant sequence. However, the process is imperfect and leads to the maintenance of heterogeneity in the *Symbiodinium* rDNA, resulting in the presence of a high number of rare ITS2 variants within an individual *Symbiodinium* cell (45, 48). The boundary between intra- versus inter-genomic variants of ITS2 is still unclear and this inhibits the use of ITS2 as marker to classify *Symbiodinium* into distinct taxa. The term ITS2 types are adopted to refer to the variations in the *Symbiodinium* diversity via ITS2 marker (42). Two molecular methods are commonly utilized when studying *Symbiodinium* ITS2 diversity i.e. bacterial cloning, and denaturing gradient gel electrophoresis (DGGE) and bands sequencing (review by (65)). Each method has its advantages and disadvantages. Bacterial cloning of the ITS2 DNA amplification products has been shown to generate inflated number of ITS2 variants with high intra-genomic variations, which may or may not represent true ITS2 sequences (66). However, *Symbiodinium* ITS2 diversity studies often focus on the dominant ITS2 types to denote the physiologically and ecologically important symbionts, since it is yet unclear what biological role low abundant *Symbiodinium* species play and how they influence the coral (67). DGGE fingerprinting on the other hand is able to separate ITS2 variants based on their
sequence composition, and the dominant ITS2 type (the most prominent band in the profile) is commonly sequenced to assess the dominant *Symbiodinium* species. However, DGGE might underestimate the true diversity of *Symbiodinium* because the resolution of DGGE can only detect symbionts above 5-10% abundance (66). Nevertheless, both methods are time consuming, laborious, and relatively expensive, hence, only minimal number of samples can be studied per time.

Corals associate with different *Symbiodinium* clades or types in different locations and different environmental conditions. *Symbiodinium* clades A, B, C, D, F, G are reported to associate with corals, other cnidaria, giant clams, and sponges (68). Some coral species show a specific relationship with the symbionts they harbor (specialists) (42, 69, 70), while others are more flexible (generalists) (71, 72). A single coral colony can harbor heterogeneous types of *Symbiodinium* (73, 74), with one dominant type accompanied by lower abundant symbiont types (75). Single coral colonies are also shown to associate with different genotypes of the same *Symbiodinium* subtypes. For example, Baum (76) reported that despite specific association between *Acropora palmata* and *Symbiodinium* ‘fitti’ or symbiont type A3 across the Caribbean, a single colony of *A. palmata* can harbor single or even mixed strains of A3 over time, and further, that *A. palmata* across the Caribbean are associated with different genotypes of A3. Moreover, corals of the same species can harbor different types of symbionts in presence of different environmental parameters such as irradiance, temperature, and depth (63, 71, 77, 78). Studies showed that presence of a particular *Symbiodinium* type (such as clade D) might influence a coral’s tolerance to thermal stress and its resistance to bleaching (79-82). In the Caribbean, corals predominantly harbor *Symbiodinium* clade A, B, and C (73), (42). In the Pacific, clade C are commonly associated with many coral species (64, 68, 83),
and in the Red Sea, many coral species are commonly associated with *Symbiodinium* clade A and C (84, 85). *Symbiodinium* clade D was proposed to be a stress-resistant symbiont capable to proliferate in harsh conditions such as high sea temperatures and high turbidity (86). However, the data are still inconclusive since symbionts in clade D are not always found to associate with corals in elevated seawater temperature, but are often found associated with corals that are starting to bleach or have recently recovered from bleaching. For instance, studies by Baker (80) reported a shift of *Symbiodinium* from the commensal clade C to clade D during bleaching event, and the symbiont reverse to original clade C once the water temperature returned to normal. Therefore, *Symbiodinium* clade D might be an opportunistic symbiont that colonizes corals during bleaching (bleached corals in the process of recovering their steady-state symbiont communities (53, 87, 88). Baker (89) also reported the difference in carbon fixation and nitrogen assimilation between symbionts from clade C1 and D at different temperatures and showed that clade C1 symbionts are more efficient in acquiring nitrogen at normal temperatures, while clade D is more efficient in carbon fixation at higher than normal temperatures. These results provide partial explanation to the global omnipresence of clade C, while clade D only has a competitive advantage in corals undergoing thermal stress. Nevertheless, it is important to note that *Symbiodinium* clade D consists of many ecological entities, some of which are yet to be identified (90), and most likely not all of them represent the thermally tolerant traits. Therefore, merely referring clade D as thermo tolerant clade may be an over generalization. Instead, the specific subtypes within clade D that exhibit thermally tolerant properties should be experimentally verified and subsequently reported.
**Coral-Associated Bacteria**

While tropical shallow water corals have long been recognized to exist in close and obligate symbioses with endosymbiotic unicellular algae (dinoflagellates) of the genus *Symbiodinium*, the importance of the diverse community of bacteria became only recently established (91). The initial studies of coral-associated bacteria were dependent on cultured bacteria that are associated with the coral Surface Mucus Layer (SML) (92, 93). However, because the majority of microbes are uncultivable, the true diversity of coral-associated microbes was unknown. Only after Pace and Olsen (94) introduced culture-independent methods for microbial studies, scientists have discovered the highly diverse microbial communities in nearly every ecosystem studied, including corals.

The culture-independent methods introduced by Pace and Olsen (94) involved the use of 16S rDNA sequencing, a gene coding for the prokaryotic small subunit (SSU) of ribosomes as a phylotypic marker. The feasibility of this sequence as a marker is highlighted by the fact that it is evolutionary conserved among all prokaryotes, yet includes 9 hyper-variable regions, V1-V9. These hyper-variable regions are used to differentiate between bacterial ribotypes whose relationships can be viewed by phylogenetic analysis. Weisburg (95) developed a universal primer pair called 27F and 1492R that probes for almost the entire region of the 16S rDNA gene and successfully works in many bacterial species. To compare bacterial diversity, it is common practice to apply a provisional cutoff of >97% similarities to 16S rRNA gene sequences that defines bacterial taxa as Operation Taxonomic Units (OTUs) (96). Applying an OTU cutoff in order to provisionally assign diversity in ITS2 sequences to species diversity has proven more difficult in *Symbiodinium* given the deep divergence within the genus.
Rohwer et al. (97) were the first to apply 16S sequencing-based methods to study bacteria associated with the coral *Montastraea franksi*. The samples were collected from five separate reefs targeting the V3 hyper-variable region. Their findings led to the discovery of various groups of novel bacterial taxa, and showed that a group of alphaproteobacteria closely related to *Silicibacter lacuscaerulensis* is present in all *M. franski*, suggesting specific microbe-coral associations (97). In the following year, Rohwer et al. (7) sequenced up to 1,000 bacterial 16S rDNAs from 14 coral samples of three different species, namely *M. franksi, Doploria strigosa,* and *Porites astreoides*. The study yielded 430 distinct bacterial ribotypes, half of which were less than 93% identical to those found in any previous study (7). The finding suggested that different coral species harbor distinct communities of bacteria (7). In addition, a study by Littman (98) that looked at bacterial community profiles of three *Acropora* species (i.e. *Acropora millepora, Acropora tenuis,* and *Acropora valida*) from two different locations in the Great Barrier Reef demonstrated that although the bacterial profiles of these closely related coral species were similar, the dominant genera of the bacterial communities differed between locations.

Corals appear to host distinct communities of bacteria in different parts of their bodies (99), including their SMLs (30, 35, 93, 100), their tissues (101, 102), and their underlying calcium carbonate skeletons (91). Sweet et al. (99) show that the choice of sampling method used to study coral-microbe association in each compartment of corals is important in order to minimize carryover contaminations. For example, they showed that “milked” mucus represents microbial communities in the coral gastrovascular cavity rather than coral SML as previously assumed (99). Coral SMLs harbor the most diverse community of bacteria (up to $10^5$-$10^6$ colony forming units (cfu) per ml) (91, 93, 103). Bourne and Munn (100) showed that the majority of
bacteria associated with mucus and tissue slurry of *Pocillopora damicornis* from the Great Barrier Reefs are *alphaproteobacteria*, and *gammaproteobacteria* respectively. Their findings support the results of Rohwer et al. (97) and Frias-Lopez et al. (101) who showed the dominance of *gammaproteobacteria* associated with tissue slurries of *M. franksi* and *D. strigosa*, respectively. Studies have also shown that bacterial communities in corals are significantly different from the surrounding seawater (30, 100, 101), further emphasizing that they may have evolved and adapted to the coral-specific environment.

The knowledge of coral-associated microbes is still in its infancy. Since the advent of sequencing-based methods, many researchers have omitted cultivating bacteria for reasons of feasibility, ease, and throughput. Interestingly, studies have shown that bacteria derived from 16S rDNA genes of uncultured and cultured communities are distinct (97, 104). Suzuki (104) claimed that there is very little overlap between partial SSU rDNA sequences from the bacterial isolates and the clone libraries of DNA extracted directly from the samples. This led them to conclude that while the majority of bacteria are not readily cultivable, many of the cultured varieties are not well represented in sequence databases (104, 105). One reason might be because majority of sequence data come from uncultured bacteria. For this reason, sequencing of 16S rDNA genes from bacterial isolates along with the clone library construction (integrative approach) is recommended to get a comprehensive picture of marine microbial studies (106-108). For instance, Rohwer (97) used SSU rDNA of both cultured and non-cultured bacteria to study communities associated with *M. franksi* and reported the dramatic difference in species composition of the cultured microbial population as compared to the uncultured ones.
One of the challenges in studying coral-associated microbes is producing a reference of bacterial species present in relatively healthy corals in order to understand true changes in coral-associated microbes with changing environmental conditions. This includes defining the type of relationship that microbes have with corals, which can be one or more of the following (28, 99, 109, 110):

- **Mutualistic** (beneficial symbionts): microbes that benefit from properties of the coral while providing benefits to the coral in return.
- **Commensalistic** (transient symbionts): microbes that get trapped in the coral SML from the surrounding environment (e.g. sediments and seawater). They do no harm or no good to the coral.
- **Parasitic** (pathogens): microbes that cause diseases in corals (satisfy Koch’s postulates) or have been reported to cause diseases in other marine organisms.
- **Opportunistic**: any coral associated microbes that take advantage of the system by proliferating or becoming virulent in certain conditions such as elevated seawater temperature, i.e. *Vibrio* spp., and may lead to diseases.

In addition, changes in the abundance of certain coral-associated bacterial community members may denote changes in condition of the coral host. Thus both qualitative (bacterial taxa) and quantitative (abundance) information regarding coral-associated bacteria are important in informing the studies of coral bleaching and diseases.

**Coral Diseases**

The perception of coral disease has been established about 30 years ago. Antonius (114) was the first to report a coral disease that showed a dark band which grew across apparently healthy tissue of scleractinian corals in reefs of the Caribbean. He referred to it as Black Band Disease (BBD) (114). Subsequently, so-called White Band and White Plague diseases were also reported (126). To date, more than 18
coral diseases have been identified (34). Out of the 18 diseases, six causative pathogens of five coral diseases were reported to satisfy Koch’s postulates. They include *Vibrio carchariae* for White Band type II, *Aurantimonas coralicida* and *Thalassomonas loyana* for White Plague type II, *Aspergillus sydowii* for Aspergillosis, *Serratia marcescens* for White Pox, and *Vibrio shiloi* for Bacterial Bleaching of *Oculina patagonica* (127). Table 1 summarizes the 13 most common coral diseases and their putative causative pathogens, if known.

Table 1 Overview of common coral diseases and their pathogens

<table>
<thead>
<tr>
<th>Disease</th>
<th>Acronym</th>
<th>Species Infected</th>
<th>Pathogen</th>
<th>Literature References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterial Bleaching</td>
<td>BBL</td>
<td><em>Oculina patagonica</em></td>
<td><em>Vibrio shiloi</em></td>
<td>(97)</td>
</tr>
<tr>
<td>Bacterial Bleaching</td>
<td>BBL</td>
<td><em>Pocillopora damicornis</em></td>
<td><em>Vibrio coralliilyticus</em></td>
<td>(106)</td>
</tr>
<tr>
<td>Aspergillosis</td>
<td>ASP</td>
<td>Octocorals (Gorgonians)</td>
<td><em>Aspergillus sydowii</em></td>
<td>(107)</td>
</tr>
<tr>
<td>Black Band</td>
<td>BBD</td>
<td>Many</td>
<td>Consortium</td>
<td>(108, 109)</td>
</tr>
<tr>
<td>White Band I</td>
<td>WBD-I</td>
<td>Acroporids</td>
<td>Unknown</td>
<td>(110)</td>
</tr>
<tr>
<td>White Band II</td>
<td>WBD-II</td>
<td>Many</td>
<td><em>Vibrio carchariae</em></td>
<td>(111)</td>
</tr>
<tr>
<td>White Plague (Caribbean)</td>
<td>WPD</td>
<td>Many</td>
<td><em>Aurantimonas coralicida</em></td>
<td>(110, 112, 113)</td>
</tr>
<tr>
<td>White Plague (Red Sea)</td>
<td>WPD</td>
<td>Mainly <em>Favia, Goniastrea</em></td>
<td><em>Thalassomonas loyana</em></td>
<td>(114, 115)</td>
</tr>
<tr>
<td>White Pox</td>
<td>WPX</td>
<td><em>Acropora palmata</em></td>
<td><em>Serratia marcescens</em></td>
<td>(116)</td>
</tr>
<tr>
<td>Yellow Blotch</td>
<td>YBS</td>
<td>Many</td>
<td><em>Vibrio sp.</em></td>
<td>(117)</td>
</tr>
<tr>
<td>Brown Band</td>
<td>BrB</td>
<td>Many</td>
<td>Ciliate</td>
<td>(118)</td>
</tr>
<tr>
<td>Porites</td>
<td>PTR</td>
<td>Porites</td>
<td><em>Podocotyloides stenometra</em></td>
<td>(119)</td>
</tr>
<tr>
<td>Skeleton Eroding Band</td>
<td>SEB</td>
<td>Many</td>
<td><em>Halofoliculina corallasia</em></td>
<td>(120)</td>
</tr>
</tbody>
</table>

Despite the proposed identification of pathogens for many diseases, some pathogens could not be verified in the subsequent studies. For example, although the isolation of *Aurantimonas coralicida* satisfied Koch’s postulates as a causative agent for WPD-
type II in the Caribbean, studies by Pantos et al. (33) and Sunagawa (36), both failed to identify Aurantimonas coralicida when applying 16S rDNA genes to study the bacterial community associated with White Plague-like disease of Montastrea annularis and M. faveolata from the Caribbean, respectively. Instead, they found certain commensal bacteria that increased in abundance in the diseased specimens and suggested these changes to be opportunistic proliferation (36). It is likely that Aurantimonas coralicida and Thalassomonas loyana are not the only causative agents of WPD-type II. In fact, it is likely that a single pathogen is not responsible for coral diseases in different species. For example, a consortium of pathogens has been shown to be involved in causing BBD (128), and this might be the case for many other coral diseases including WPD. However, whether changes in bacterial community associated to coral diseases are secondary reaction towards diseased corals (110), or whether they are responsible for the onset of the diseases still remain to be investigated.

Another challenge is the difficulty to discern some phenotypically similar diseases. Diseases such as White Plague, White Pox, and White Band appear to have similar characteristics. This is because coral diseases are identified according to their morphological and phenotypic framework during field observations rather than underlying bacterial assemblages. (34). Hence, it creates ambiguity and causes misidentification when researchers encounter diseased corals. It also raises the question as to whether the same characteristics of coral diseases reported worldwide are indicative of the same disease or if they simply look very similar. Taken together, we are faced with a situation, in which phenotypically similar diseases in different species are sometimes grouped together (as in the case of WPD) and sometimes are not (as in the case of White Pox, White Band, etc.). In addition, classification is
irrespective of any common identified pathogen. Hence, global and comprehensive framework is needed to verify if phenotypically similar coral diseases harbor the same structure of bacterial communities and whether it is suffice to regard them as the same disease.
REFERENCES


102. FEMS microbiology ecology 68(5):2214.


OBJECTIVES

This dissertation aimed to assess the diversity of coral-associated symbionts, specifically _Symbiodinium_ and bacteria, in various coral species from various geographic locations and health conditions. The main focus was to expand the scope of technologies, and establish a standard framework to facilitate studies of coral’s symbionts assemblage in multi coral species from different locations.

Studies of _Symbiodinium_ diversity are still majorly dependent on traditional methods like DGGE, and bacterial cloning, which are time consuming and not feasible for large scale diversity typing. Therefore, one of the aims of this dissertation was to develop a pipeline for NGS to assess _Symbiodinium_ ITS2 diversity, which was elucidated in Chapter 1. The chapter also provided a comparison of data from NGS and DGGE to verify the accountability of the pipeline. The pipeline was further utilized in Chapter 2 to survey _Symbiodinium_ diversity around the Arabian Peninsula. Seas around the Arabian Peninsula represent some of the most extreme conditions for corals to persist. However, studies of coral holobiont diversity here are still very limited. Hence, the second chapter aimed to provide a comprehensive data on _Symbiodinium_ diversity that are associated with corals around the Arabian Peninsula via NGS platform using the pipeline developed in Chapter 1.

Two challenges in coral disease studies include 1) to correctly characterize stressed corals to a specific disease, and 2) to identify the causative pathogen for the disease. However, the characterization of some diseases itself remain ambiguous due to similarity in appearance of the symptoms produced by the stressed corals. Whether phenotypically similar diseases undergo the same changes in microbial community is still unknown. To address this issue, Chapters 3 and 4 were devoted to conduct a global comparative study of bacterial community associated with phenotypically
similar coral diseases in order to assess if common bacterial footprint exists that are conserved between coral species and locations. 3<sup>rd</sup> Generation 16S rDNA Microarrays (Phylochip) were utilized for a standardized comparison between coral species and oceans for these two chapters.
Chapter 1: Assessing *Symbiodinium* diversity in scleractinian corals via next-generation sequencing-based genotyping of the ITS2 rDNA region

Chatchanit Arif\(^1\), Camille Daniels\(^1\), Till Bayer\(^2\), Eulalia Banguera-Hinestroza\(^1\), Adrian Barbrook\(^3\), Christopher J Howe\(^3\), Todd C LaJeunesse\(^4\), Christian R Voolstra\(^1\).

\(^1\)King Abdullah University of Science and Technology, Saudi Arabia

\(^2\)GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany

\(^3\)Department of Biochemistry, University of Cambridge, United Kingdom
1.1 Abstract

The persistence of coral reef ecosystems relies on the symbiotic relationship between scleractinian corals and intracellular, photosynthetic dinoflagellates in the genus *Symbiodinium*. Genetic evidence indicates that these symbionts are biologically diverse and exhibit discrete patterns of environmental and host distribution. This makes the assessment of *Symbiodinium* diversity critical to understanding the symbiosis ecology of corals. Here we applied pyrosequencing to the elucidation of *Symbiodinium* diversity via analysis of the Internal Transcribed Spacer 2 (ITS2) region, a multicopy genetic marker commonly used to analyze *Symbiodinium* diversity. Replicated data generated from isoclonal *Symbiodinium* cultures showed that all genomes contained numerous, yet mostly rare, ITS2 sequence variants. Pyrosequencing data was consistent with more traditional denaturing gradient gel electrophoresis (DGGE) approaches to the screening of ITS2 PCR amplifications, where the most common sequences appeared as the most intense bands. Further, we developed an Operational Taxonomic Unit (OTU)-based pipeline for *Symbiodinium* ITS2 diversity typing to provisionally resolve ecologically discrete entities from intragenomic variation. A genetic distance cutoff of 0.03 collapsed intragenomic ITS2 variants of isoclonal cultures into single OTUs. When applied to the analysis of field-collected coral samples, our analyses confirm that much of the commonly observed *Symbiodinium* ITS2 diversity can be attributed to intragenomic variation. We conclude that by analyzing *Symbiodinium* populations in an OTU-based framework, we can improve objectivity, comparability, and simplicity when assessing ITS2 diversity in field-based studies.
1.2 Introduction

Coral reef communities depend critically on the relationship between scleractinian corals and photosynthetic endosymbionts in the genus *Symbiodinium*. While the coral host provides a light-rich, sheltered environment and inorganic nutrients, the dinoflagellate algae provide photosynthetically fixed carbon (1). Some corals seem to rely strictly on a distinct symbiont type, whereas others harbor different types of *Symbiodinium* (2). Although host identity is the primary predictor of symbiont identity, coral-associated *Symbiodinium* diversity also co-varies with geographic location, water depth, and health state (3-10). The current decline in coral reef cover resulting from global (e.g. ocean warming) and local (e.g. pollution, overfishing) anthropogenic factors has intensified the need to examine the spatial, geographical, and ecological distribution of *Symbiodinium*-coral associations and how these relate to environmental extremes.

Various DNA markers have been used to describe genetic diversity within *Symbiodinium* (11). Rowan and Powers (12) were the first to examine *Symbiodinium* evolutionary relationships using restriction fragment length polymorphisms (RFLPs) by sequencing of the small ribosomal subunit RNA (SSU) gene from dinoflagellates from various marine invertebrate hosts. Their phylogeny identified divergent lineages within this genus that exhibited sequence differences comparable to those observed among dinoflagellates from different taxonomic families and orders (13). These lineages are commonly referred to as clades. So far, 9 clades are identified (clades A – I), of which representatives in 6 clades (i.e. A, B, C, D, F, G) have been shown to persist in association with corals, other cnidarians, giant clams, and sponges (14, 15). It was soon recognized that the conserved SSU gene is not able to resolve species (12, 16, 17). Researchers turned to more variable DNA regions, including the nuclear large
subunit (LSU) (18, 19), the Internal Transcribed Spacer regions (17, 20, 21), the chloroplast large subunit (cp23S) (15, 22), and cytochrome oxidase b (11). A systematic survey by Sampayo, Dove and Lajeunesse (11) targeting ribosomal, mitochondrial, and chloroplast genes employing 13 distinct genetic analyses found that different markers showed remarkable concordance, but differed in their relative ability to resolve ecologically distinct units. Other rapidly evolving markers, including the chloroplast psbA non-coding region (23, 24) and microsatellites were recently employed, often in combination with each other, to improve genetic resolution further and delimit species boundaries (25). Only recently, Barbrook, Voolstra and Howe (26) sequenced the entire chloroplast genome of a *Symbiodinium* sp. type C3, where individual genes reside on minicircles that will provide additional genetic markers. Currently, however, the ITS2 region is still the most commonly employed DNA marker used to assess *Symbiodinium* diversity from a diverse array of hosts and over large geographic distances (6, 27-30).

LaJeunesse (31) determined that ITS2, when analyzed through the targeting of numerically dominant intragenomic variants, provides sufficient resolution to resolve many ecologically distinct *Symbiodinium* spp. In order to achieve this, Denaturing Gradient Gel Electrophoresis (DGGE) was used to screen PCR amplifications for common ITS2 sequence variants diagnostic of a particular *Symbiodinium* “type” (6, 7, 30, 31). More recently, researchers have also employed bacterial cloning and sequencing and found more sequence diversity within the populations of *Symbiodinium* residing in a host than identified with DGGE (32-35). Because rDNA represents an extreme example of a multicopy gene arrayed in tandem, intragenomic variation in the form of pseudogenes or numerous low-abundant functional variants affect how ITS2 data are interpreted (11, 24, 36). For instance, it has been shown that
cloning-and-sequencing based approaches are potentially prone to inflated diversity estimates through the technique's tendency to recover intragenomic variants that are of low abundance (i.e. limited diagnostic value), as well as by introducing additional sequence artifacts generated during the PCR and cloning steps (36). However, sequencing a substantial number of clones may resolve how rDNA data are best analyzed and interpreted by allowing differentiation between intragenomic and interspecific ITS2 sequence variants, albeit at a high cost (11). Assessment has also shown that individuals from different species possess widely differing amounts of intragenomic variation (36), yet the full extent of this variation is at present unknown. In both regards, the application of pyrosequencing-based methods is projected to overcome the limited resolution and high cost associated with Sanger-based ITS2 diversity assessments, as a high number of sequences are produced at comparatively low cost per sequence.

While the study of prokaryotic diversity is now routinely conducted via pyrosequencing-based amplicon-typing of the 16S region (37, 38), only a limited number of studies have utilized high throughput sequencing to assess eukaryotic diversity. For instance, Stoeck et al. (39, 40) and Amaral-Zettler, McCliment, Ducklow and Huse (41) sequenced the variable regions of the small subunit (SSU) and large subunit (LSU) of the 18S region to estimate eukaryotic diversity.

Pyrosequencing is now being applied to assess Symbiodinium diversity (42-44), but there is a need to ground truth this new approach. In this study, we applied pyrosequencing of ITS2 rDNA amplicons (320-360bp) to genotype Symbiodinium diversity in several isoclonal and replicated cultures, two cultured isolates mixed at different ratios, and field-collected coral specimens. The application of a high-throughput pyrosequencing approach holds the promise to improve assessment of the
relative degree of sequence variation found in these eukaryotic genomes. Further, a pyrosequencing approach may provide a more accurate detection of low abundance background *Symbiodinium*. Our aim was to understand diversity of the ITS2 gene at the genome level and to compare pyrosequencing results to data obtained from DGGE-typing. Furthermore, we sought to analyze *Symbiodinium* diversity in an OTU-based framework in order to improve objectivity, comparability, and simplicity when assessing ITS2 to study *Symbiodinium* composition in environmental samples of corals.

1.3. Materials and Methods

1.3.1 Sample collection and processing

Isoclonal cultures of *Symbiodinium* sp. (CCMP2467: *S. microadriaticum*, KB8 *S. microadriaticum*, rt-147: undescribed clade B type, and rt-064: undescribed B1 type) were cultured at 23°C in f/2 medium (45) on a 12h:12h light-dark cycle (daytime: 6 am to 6 pm; night-time: 6 pm to 6 am, light intensity 80 µmolm⁻²s⁻¹). The salt content in the medium was set to 40 g/l, matching the average salinity characteristic of the Red Sea. Coral samples were collected from various reefs in the Red Sea between 2011 and 2012 with SCUBA at depths between 4 and 12 m. More specifically, five specimens of *Pocillopora verrucosa* were collected from reefs at Maqna (n=2), Al Wajh (n=2), and Doga (n=1) as well as one specimen of *Acropora hemprichii* from Al Fahal reef at Thuwal. Specimens of about 1-3 cm² tissue were collected with Hammer and Chisel and stored in Whirl-Paks during diving, subsequently washed with 0.22 µm filtered seawater (FSW), preserved in DMSO/NaCl buffer (46), and stored at 4°C until further processing. For *Symbiodinium* culture DNA extraction, cells from each
culture were counted using a hemocytometer (Hausser Scientific, Horsham, PA) under a light microscope (Leica DM2500, Wetzlar, Germany), and approximately 8 x 10^6 cells per culture were used for DNA extraction. Mixed samples were generated by combining cells of two *Symbiodinium* species (a strain of *S. microadriaticum* CCMP2467 from clade A and strain rt-147 from clade B) in a ratio of 1:1 (4 x 10^6 cells : 4 x 10^6 cells) and in a ratio of 1:3 (2 x 10^6 cells : 6 x 10^6 cells), respectively. Cells were spun at 1,934 g for 10 min, and subsequently washed with DNase-free water. 500 µl of 0.5 mm sterile glass beads (BioSpec, Bartlesville, OK) were added to the pelleted cells together with 400 µl of buffer AP1 and 4 µl RNAse (Qiagen, Hilden, Germany). Samples were bead-beaten for 90 s with a Tissue Lyser II (Qiagen, Hilden, Germany). DNA was isolated with the Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions. For the analysis of technical variation, DNA from isoclonal cultures of CCMP2467 and rt-147 were isolated once but amplified in distinct PCR reactions with a different barcoded primer. For DNA extraction of environmental samples, approximately 50 mg of coral tissue was transferred to 1.5 ml tubes (Eppendorf, Hamburg, Germany). 500 µl of 0.5 mm sterile glass beads (BioSpec, Bartlesville, OK) were added together with 400 µl of buffer AP1 and 4 µl RNAse (Qiagen, Hilden, Germany). Samples were bead-beaten for 90 s with a Tissue Lyser II (Qiagen, Hilden, Germany). DNA was isolated with the Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer’s protocol.

**1.3.2 DGGE analysis of the ITS2 rDNA**

DGGE ITS2 diversity typing was performed following the protocol detailed in LaJeunesse (31). Briefly, the *Symbiodinium* ITS2 region was amplified with the primer pair ITSintfor2 and ITS2CLAMP using PCR conditions described in
LaJeunesse, et al. (47) with the following modifications: the annealing temperature was maintained at 52°C for 27 cycles after 20 cycles of touchdown amplification. PCR products were mixed with 10 µl Ficoll-based loading buffer and concentrated by a speed vacuum before loading on a 8% poly-acrylamide gel using a Cipher DGGE kit (CBS Scientific Company, Del Mar, CA). Gels were run at 150V for 15 hours, stained for 30 minutes with 1x SYBR Green (Invitrogen, Carlsbad, CA), and visualized on a Dark Reader Transilluminator (Clare Chemical Research, Dolores, CO). Prominent band(s) were excised from the DGGE gel with a sterile scalpel. Each band was transferred into an Eppendorf tube that contained 500 µl DNAse-free water and incubated at 4°C for 24 hrs. 2 µl of this were used for re-amplification as described in LaJeunesse (31) and purified with Illustra ExoStar (SelectScience, Bath, UK) enzyme mix following the manufacturer’s instructions. Successful amplification was verified by running products on a 1% agarose gel stained with 1x SYBR Safe (Invitrogen, Carlsbad, CA). Samples were sent for bi-directional Sanger sequencing at the KAUST BioScience Core Laboratory (Thuwal, Saudi Arabia). Sequences were processed in CodonCode Aligner (CodonCode Corporation, Centerville, MA). After quality trimming, forward and reverse sequences were assembled into contigs. For phylogenetic assignment of ITS2 sequences, we built a custom BLAST database of ITS2 types collected from 409 ITS2 sequences taken from GeoSymbio (Franklin et al. 2012) (denoted as GS), 7 ITS2 sequences from Scott Santos’ database (www.auburn.edu/~santosr/sequencedatasets.htm) (denoted as ST), and 17 DGGE ITS2 sequences from Todd LaJeunesse’s SD2-GED database (https://131.204.120.103/srsantos/symbiodinium/sd2_ged/database/views.php) denoted as LJ. ITS2 sequences were assigned to the ITS2 types that represented maximum percent identity in the BLASTn hits.
1.3.3 454 pyrosequencing of *Symbiodinium* ITS2 rDNA

PCR amplification of the ITS2 gene for pyrosequencing was performed using primers ITSintfor2 and ITS2-reverse that generated an amplicon of around 320 bp. The primer sequences were 5’- CCATCTCATCCCCGCTGCTCTCCGACTCAG(N)8GAATTG CAGAACTCCGTG-3’ (454-ITSintfor2) and 5’- CCTATCCCCCTGTGTCGCTTTGGC AGTCTCAGGGGATCCATATGCTAAGCTAGGAGGTT-3’ (454-ITS2-reverse).

Primers included 454 Lib-L library adapters (underlined) and a barcode (shown as N) (48). PCRs were run in triplicate per sample with 12.5 µl of Qiagen Multiplex PCR Kit (Qiagen, Hilden, Germany), 0.1 µM primers, 20 – 50 ng DNA, and DNAse-free water to make a total volume of 25 µl. The following PCR conditions were used: initial denaturation for 15 min at 94°C, followed by 35 cycles of 94°C for 30 s, 51°C for 30 s, 72°C for 30 s, and a final extension step of 10 min at 72°C. PCR products were run on a 1% agarose gel stained with 1x SYBR Safe (Invitrogen, Carlsbad, CA) to visualize successful amplification. For each sample, triplicate PCR products were pooled and their DNA concentrations were measured using a Qubit 2.0 (Invitrogen, Carlsbad, USA). 20 ng of the triplicated PCR samples were combined and ran on a 1% agarose gel to remove excess primers. The gel band was excised, purified with the Qiagen MinElute Gel Extraction Kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions, quantified with Qubit 2.0 (Invitrogen, Carlsbad, USA), and quality checked via Bioanalyzer (Agilent, Santa Clara, CA). 40 ng of a pooled library (i.e. pooled triplicated PCRs from all samples) were submitted to KAUST BioScience Core Laboratory (Thuwal, Saudi Arabia) for sequencing using Titanium FLX chemistry on a quarter of a picotiter plate. Raw sequencing data were retrieved with Roche 454 amplicon processing pipeline.
1.3.4 454 pyrosequencing data analysis

A total of 218,475 reads with a median length of 314 bp were obtained from sequencing and processed using the software mothur v.1.31.2 (49). Sequences were de-noised using PyroNoise (50). Forward primer and barcode sequence were removed from reads by the trim.seqs function in mothur. All sequences that met the following criteria were discarded: barcodes (>0 mismatches), forward primer (>2 mismatches), ambiguities (>0 bp), homopolymers (>4 bp), and short sequence length (<250 bp). Cutadapt version 1.1 (51) was applied to remove the reverse primer (overall error rate set to 0.15). All identical sequences were subsequently collapsed and representative sequences were retained via unique.seqs command in mothur. After chimera removal with UCHIME as implemented in mothur (52), singletons (i.e. sequences detected only once across the entire dataset) were also discarded. 197,181 sequences were retained for the remainder of the analyses. From these data, frequency distributions of ITS2 variants for all samples were obtained via count.seqs command in mothur. Genetic distances for within-culture ITS2 diversity were calculated with the dist.seqs command in mothur based on MUSCLE-aligned ITS2 copies that were represented by at least 100 reads in any given isoclonal culture.

For the OTU-based framework analysis, ITS2 sequences were assorted into their respective clades based on pairwise distances via pairwise.seqs command, and subsequent clustering with average neighbor algorithm using a cutoff value of 0.15 (empirically determined to effectively cluster sequences into clades). Sequences assigned to distinct clades were aligned with MUSCLE (53) and trimmed to equal length using the screen.seqs and filter.seqs commands in mothur. Sequences that were shorter than 90% of sequences in each clade were discarded resulting in 197,128 sequences. A distance matrix was calculated using the aligned sequences within each
clade. Sequences were clustered with the average neighbor algorithm (54) at a 97% similarity cutoff, as this cutoff clustered ITS2 variants from isoclonal cultures into a single OTU (i.e. species). The most abundant sequence in each OTU was chosen as the representative ITS2 copy and was annotated via BLASTn against the custom ITS2 database to determine a specific *Symbiodinium* type. In addition, the most abundant ITS2 copy from each pyrosequenced sample was aligned to the sequence derived from the most prominent DGGE band of the same sample to verify whether they were identical. A list of succession of commands and an unattended script are available as supplementary information (Supplemental File S1.1, Supplemental File S1.2).

### 1.3.5 Single cell PCRs on *Symbiodinium* cells

In order to conduct single cell-based PCRs on *Symbiodinium*, we developed an approach modified from Frommlet and Iglesias-Rodríguez (55). Cells from cultures CCMP2467, KB8, rt-064, and rt-147 were harvested with the following changes: 1 ml of each culture was filtered through a 40 µm cell strainer (BD Biosciences, San Jose, CA), spun down at 1934 g for 10 min, and washed 3 times with 1 ml of 1x TE buffer. Cell pellets were subsequently re-suspended with 500 µl of 1x TE, counted using a hemocytometer (Hausser Scientific), and diluted to a concentration of 1 cell/µl (‘dilution-to-extinction’ approach). 1 µl was subsequently transferred into 96 well plates and the presence of single cells was confirmed using an inverted microscope (Leica DMI3000B, Wetzlar, Germany). Cells were disrupted by triplicate freezing of cells for 2 min in Liquid Nitrogen and immediate thawing for 2 min at 95°C. Next, a single 0.5 mm glass bead (BioSpec, Bartlesville, OK) was added to each well and the cell was homogenized with a Tissue-Lyzer II (Qiagen, Hilden, Germany) for 1 min. PCR master mix containing 12.5 µl Qiagen Multiplex PCR Kit (Hilden, Germany), 0.1 µM of ITSintfor2 and ITS2-Reverse (31), and DNase-free water was added to
make a total volume of 25µl. PCR conditions were identical to the conditions used for 454 sequencing. Subsequently, 1 µl of the PCR product was used as template in a 25µl clade-specific PCR reaction containing 12.5 µl Qiagen Multiplex PCR Kit (Hilden, Germany), DNase-free water, 0.1 µM of degenerated reverse primer (5’-TCWCYTGTCTGACTTCATGC-3’), and 0.1 µM of either clade A- (5’-TGGCACTGGCATGC-3’), B- (5’-ATTGCTGCTTCGCTTTCC-3’), or C- (5’-TGCTTAACTTGCCCAAC-3’) specific forward primers generating 100bp, 180bp, and 210bp amplicons, respectively. PCR conditions were 94°C for 15 min followed by 35 cycles of 30 sec at 94°C, 30 sec at 53°C, and 30 sec at 72°C and a final step of 10 min at 72°C. Amplicons were checked on a 1% agarose gel, cloned with Invitrogen TOPO-TA cloning kit (Carlsbad, CA), and sent for Sanger sequencing to the KAUST BioScience Core Laboratory (Thuwal, Saudi Arabia). CodonCode Aligner (CodonCode Corporation, Centerville, MA) was used to process sequences and annotated via BLASTn using the local Symbiodinium ITS2 database.

1.4 Results

1.4.1 Sequence variation among ITS2 copies in the genomes of Symbiodinium

We applied 454-based amplicon sequencing of the ITS2 rDNA region from Symbiodinium to a collection of replicated isoclonal cultures (n = 7), pooled cultures (n = 2), and environmental samples (n = 6) (Table 1.1). A total of 218,475 reads were obtained from pyrosequencing with a median length of 314 bp (mean length = 323.11 ± 32.61 SD). After filtering of sequences, 197,181 reads were retained with an average of 13,145.40 sequences per sample, representing 1,487 unique ITS2 sequences (Table 1.1, Supplemental File S1.3).
Depending on the taxon investigated, we identified between 102 and 331 distinct ITS2 sequence variants, including cultured strains and field-collected specimens (mean = 219.67). Taking only isoclonal culture samples into account, we identified on average 230.86 ITS2 sequence variants per culture indicating that there is a broad range in the number of distinct ITS2 sequence variants found within *Symbiodinium* genomes (Table 1.1). Despite the high number of distinct ITS2 copies, read counts for the different ITS2 copies showed a highly uneven distribution (Figure 1.1A). When sorting ITS2 variants of isoclonal cultures by sequence read abundance, the most abundant ITS2 copies were on average ~20 times more prevalent than the second most common ITS2 copy (all clades 21.84-fold, clade A’s 8.25-fold, clade B’s 39.88-fold, clade C’s 8.49-fold). Further, the 5 most abundant ITS2 copies from any culture made up >80% of associated reads, indicating that only few distinct ITS2 genes make up the majority of genomic gene copies. This was substantiated by a rarefaction analysis, which indicated that most of the numerous ITS2 copies were captured at very low abundance in each genome (Figure 1.1B). For instance, subsampling of isoclonal cultures to 2,000 reads yielded on average less than half of the distinct ITS2 copies we were able to recover taking all sequence reads into account. Additionally, the rarefaction curves at this sampling depth did not approach saturation. This effectively illustrates that intragenomic diversity lies in low abundant genomic ITS2 copies, which also seems to far exceed what could be captured by 'traditional' sequencing methods. Comparison of pyrosequencing data of culture CCMP2467 (type A1) to 25 sequences generated by cloning-and-sequencing showed that both techniques identified the same most dominant sequence, but numerical ranking of the next most common sequences did not particularly match up well between cloning and 454 data (not shown). Comparing pyrosequencing data to DGGE fingerprinting of
isoclonal *Symbiodinium* cultures, DGGE yielded a single dominant band accompanied by few, faint, background bands. The dominant ITS2 variants produced by 454 pyrosequencing (numerically abundant) and DGGE (brightest band) were identical in sequence, and were representative sequences of the *Symbiodinium* type analyzed (Table 1.1, Figure 1.2). Accordingly, 454 pyrosequencing and DGGE produced uniform results in regard to identifying the ITS2 variants most representative of the genome.

To further assess levels of intragenomic ITS2 diversity, we calculated uncorrected genetic distances between all ITS2 variants of a given isoclonal culture (Table 1.2). For each sample, we only considered ITS2 types represented by at least 100 reads. This was done in order to avoid diversity inflation by potential contamination, methodological artifacts, or ultralow abundant ITS2 copies. The resulting median genetic distance between ITS2 copies from isoclonal cultures was in all cases below 0.02, and median intragenomic variation ranged from 0.003 to 0.017 (Table 1.2). Therefore, most sequence variants within any culture differed by only one or two nucleotide substitutions from each other.

The reproducibility of pyrosequencing rDNA varied depending on the scale of comparison. In some cases, different tallies of distinct ITS2 copies were calculated between the sequencing of different PCR amplifications conducted from the same DNA extract. Our replicated analysis of culture CCMP2467 recovered 331 and 241 distinct ITS2 sequences from 16,681 and 8,565 sequence reads, respectively. Of these ITS2 sequences 184 were identical (55.59% and 76.35%, respectively). Repeated sequencing of culture rt-147 was less variable, and retrieved 225 and 241 distinct ITS2 variants at a comparable sequencing depth, of which 168 were identical (74.67% and 69.71%, respectively) (Table 1.1). The comparison between the top 10 most
abundant sequence variants showed consistency, but there were some discrepancies between replicated samples (Figure 1.3). Sequence variants that represented 2-3% of the total variation were typically recovered between replicates, but not always in the same relative abundance. The top 3 most common sequences, however, were always recovered, and in the same order of abundance between replicates (Figure 1.3). To further elucidate this pattern, the top 10 most common variants from pyrosequencing ITS2 of culture rt-152 (Symbiodinium goreaui or type C1) were compared to the composition of top 10 variants identified from several field-collected samples representing clade C Symbiodinium, including types C1c, C1h, and C41, (Figure 1.4). For each of these samples, there were 1 to 5 common sequence variants representing >5% of sequence reads for the respective specimen. Each sequence set contained the C1 sequence, sensu LaJeunesse (17), but in very different proportions. The C1 sequence comprised 68.36% of the genome of S. goreaui (rt-152 or type C1), 34.00% for type C1c (P.Waj.D7), 34.08% for type C1h (P.Maq.R2.19), and 11.82% in type C41 (A.Af.B6). These proportions were similar to the relative band intensities observed from their respective DGGE fingerprint profile (Figure 1.2). The “c” sequence found in co-dominance with C1 (19.95% vs. 34.00%) in the ribosomal array of Symbiodinium type C1c also was detected in two other types (C1h and C1) at significantly lower concentrations and explains why this sequence variant was not resolved by DGGE fingerprinting (Figure 1.2).

Among sequences recovered from PCRs of DNA extracted from the artificially mixed cultures, CCMP2467 (S. microadriaticum or type A1) and rt-147 (type B1), in a 1:1 or 1:3 cell ratio, more than 99% of ITS2 reads were identified only as clade B. This result was consistent with DGGE profiling conducted on the same samples, which showed only a single prominent band matching that of the fingerprint for culture rt-
The band diagnostic of type A1, *S. microadriaticum*, was absent in both mixed samples (Figure 1.2). Sequences derived from 454 (most abundant) and DGGE (brightest band) of both pooled samples were identical and identified as type B1 via BLASTn. These results suggest that either clade B has a far greater copy number of rDNA than clade A, or PCR bias of rt-147 over CCMP2467. This bias might arise from differences in DNA isolation efficiency or preferential primer amplification as discussed in Thornhill, Xiang, Pettay, Zhong and Santos (56). In conclusion, we could not detect both entities in their initially mixed proportions.

Looking further at ITS2 type composition from pyrosequencing reads of isoclonal cultures, we found that besides the majority of sequences being representatives of the underlying ITS2 type, very low abundant ITS2 types were present that were representatives of other clades (Table 1.1). These variants matched with the dominant sequence variants (or nearly so) of isolates representing other *Symbiodinium* clades also under analysis, and occurred between 0.01% and 0.40% of all reads (Table 1.1). To understand the cause of this observation further, we conducted a series of single cell PCRs on cultures CCMP2467, KB8, rt-064, and rt-147 with primers specific for Clades A, B, and C. Clade-specific PCRs performed on single *Symbiodinium* cells from these cultures yielded PCR products with the corresponding clade-specific primers, and failed to yield amplicons using primers specific to a clade other than the template, with the exception of culture rt-064 (type B1). For this culture, we retrieved PCR products in 2 out of 7 cases upon amplification with clade A-specific primers. From both PCR products, a total of 8 clones were sent for sequencing, of which 5 were identical to the identified clade B1 in 454 sequences and 3 were discarded due to reduced quality.
1.4.2 Taxon-based analysis of Symbiodinium ITS2 diversity in isoclonal cultures

Studies of microbial diversity in an OTU framework based on pyrosequencing of the 16S gene have revolutionized our understanding of bacterial diversity and distribution (37). Similarly, the ITS2 gene is amenable to analysis in a taxon-based framework under the premise of derivation of appropriate cutoffs to denoting clades and species/types. Given the deep divergence between *Symbiodinium* species from different clades (comparable to differences between orders in other dinoflagellates), we had to devise a strategy where taxonomic delineation was conducted on the clade level first and subsequently at the type level.

Clade-separation was empirically determined by pairwise similarity calculation of all sequences and subsequent clustering using a similarity cutoff of >0.10. This approach consistently clustered reads into different clades, which was confirmed by BLASTing representatives of each cluster against our custom ITS2 database. As mothur calculates exact distance cutoffs, the next higher distant cutoff from 0.10 was chosen for data analysis (here: 0.15). After aligning, trimming, and discarding sequences that were shorter than 90% of the reads in each clade, 197,128 sequences remained, of which 101,239 sequences belonged to clade A, 82,273 belonged to clade B, and 13,616 sequences belonged to clade C (Table 1.3).

For the determination of species- or type-level cutoffs we used data from isoclonal cultures. Our aim was to determine a cutoff that effectively clustered all reads from culture samples assorted to a clade (step above) into 1 corresponding OTU at the ‘species’ level. Average neighbor clustering of reads based on uncorrected pairwise distances at 0.03 provided the cutoff where all cultures collapsed into 1 OTU for a given clade (Table 1.3). Therefore, 97% sequence similarity was implemented as a cutoff value for species-/type-level OTU-based analyses within clades. Comparing
results from sequence- (Table 1.1) and OTU-based (Table 1.3) analyses effectively illustrates that the high number of distinct ITS2 copies that we identified in isoclonal culture samples could be collapsed to within-species (i.e. within OTU$_{0.03}$) ITS2 variation in a taxon-based analysis. Three distinct OTU$_{0.03}$ (1 of each Clade A, B, C) represented data from all cultures. For the pooled culture samples, 2 OTU$_{0.03}$ were correctly identified, but the relative abundances of reads did not reflect the initial ratio of cells used. Accordingly, while diversity was correctly recovered for cells from pooled isoclonal cultures, the relative abundances were not.

1.4.3 Taxon-based analysis of ITS2 diversity in environmental samples

We applied our OTU-based framework to analyze ITS2 diversity in environmental samples based on the cutoffs we derived from the isoclonal cultures (Supplemental File S1.4). Similar to the results for the isoclonal culture samples, the amount of type-level OTU$_{0.03}$ was dramatically lower than the number of distinct ITS2 copies. Diversity in environmental samples was comprised of only 6 OTU$_{0.03}$ that represented a total of 841 distinct ITS2 sequences, demonstrating that taxon-based framework analyses drastically and effectively reduce complexity and diversity of primary sequence data. Of the 6 OTU$_{0.03}$ that we identified across all environmental samples, 3 OTUs could be assigned to clade A (referred to as OTU1$_{A0.03}$, OTU2$_{A0.03}$, OTU3$_{A0.03}$), 2 OTUs to clade B (referred to as OTU1$_{B0.03}$, OTU2$_{B0.03}$), and 1 OTU to clade C (referred to as OTU1$_{C0.03}$) (Table 1.3). OTU1$_{A0.03}$, OTU1$_{B0.03}$, and OTU1$_{C0.03}$ were the most common ITS2 types representing >99% of all sequences for a given clade. These OTUs were classified belonging to *Symbiodinium* types *A1*, *B1*, and *C1*, each with 100% sequence identity (Table 1.3). Interestingly, all environmental samples were dominated by a single OTU$_{0.03}$, although all environmental samples comprised more than one OTU, but the fraction of sequence reads representing
additional OTUs was low. This diversity was not captured in the DGGE fingerprint. For instance, DGGE fingerprinting of samples P.Dog.R3.2 and P.Maq.R2.7 only showed association with *Symbiodinium A1*. In our pyrosequencing data, although both samples were dominated by OTU1<sub>A0.03</sub> representing clade *A1* (~99% of all reads), 0.27% and 0.38% of reads represented OTU2<sub>A0.03</sub>, and 0.02% of all reads from sample P.Maq.R2.7 represented OTU3<sub>A0.03</sub> (Table 1.3). Similarly, while P.Maq.R2.19 was dominated by OTU1<sub>C0.03</sub> in pyrosequencing data (98.06% of reads), it was also associated with clade *A1* (OTU1<sub>A0.03</sub>: 1.12%) and clade *B1* (OTU1<sub>B0.03</sub>: 0.82%). Finally, even though P.Maq.R2.7 was the most diverse sample, it still could be represented by only 6 OTUs, effectively demonstrating how OTU-based analyses collapse the majority of genetic diversity into intragenomic (i.e. within OTU) variation.

Comparing the OTU-based analysis to DGGE fingerprinting, we found that for the majority of samples analyzed, the representative OTUs were identical to the *Symbiodinium* ITS2 type derived from DGGE rDNA fingerprinting when the most prominent band was sequenced. For sample A.Af.B6, however, the most prominent DGGE band was represented by ITS2 type *C41*. While this type was also the most abundant pyrosequencing read in that sample, *C41* collapsed into a single OTU (i.e. OTU1<sub>C0.03</sub>) with a *C1* ITS2 type. Similarly, in some cases DGGE fingerprinting identified ITS2 types that were not readily identified as distinct OTUs from pyrosequencing data. For instance, while P.Maq.R2.19 was primarily associated with a clade *C1* ITS2 type, it was also associated with clade *C1h* based on the DGGE profile. Further, sample P.Waj.D.7 showed an association with *C1c* in addition to clade *C1* in the DGGE profile. As was the case for ITS type *C41*, *C1h* and *C1c* were readily detected in pyrosequencing data, but represented by/collapsed to a single OTU
(i.e. OTU1_{c0.03}). This can be attributed to a single base pair difference between C1, C41, Clh, and Clc. From a taxon-based analysis point-of-view, inter-genomic divergence did not exceed intra-genomic divergence, and accordingly could not be resolved into distinct OTUs.

1.5 Discussion and Conclusion

1.5.1 Sequence variation among ITS2 copies in the genomes of *Symbiodinium*

We applied 454 pyrosequencing to analyze intra- and intergenomic *Symbiodinium* rDNA diversity in order to gain an in-depth perspective on the relative homogeneity of this high multi-copy gene (36). By utilizing isoclonal cultures, we further resolved the degree of genomic homogeneity relative to abundance of ITS2 sequence variants across the ribosomal array, we extended sequencing resolution in comparison to classical cloning-and-sequencing-based approaches, and we were able to empirically derive cutoffs for application of these types of data in a taxon-based framework. Our analysis protocol was then applied to environmental samples to assess overall feasibility of the approach.

A concern regarding utilization of ITS2 genes for estimating *Symbiodinium* diversity is the multi-copy nature of ribosomal DNA (31). The genomes of eukaryotes contain one to several numerically dominant sequence variants (57-59). Species lineages of *Symbiodinium* exist that have two or more co-dominant ITS2 copies that together are diagnostic of the species (11). The rank abundance plot of ITS2 variants from isoclonal samples (Figure 1.1A) revealed that while there seems to be extensive intragenomic variation in ITS2 sequences, each genome was characterized (in these particular examples) by the presence of one numerically dominant ITS2 sequence.
Provided that the process of concerted evolution is correctly understood (57), common intragenomic variants persist in the genomes among the individuals of a genetically recombining population over long evolutionary time-scales (11, 25). The number of ITS2 sequence variants follows a long tail abundance distribution with dominant variants (often one or two) being present at high frequency and a high number of rare variants at much lower frequencies (Figure 1.1). This disparity among intragenomic variants indicates that gene conversion drives the relative homogenization of this gene array, but is not rapid enough to dampen out the appearance and partial spread of new mutations. As rare variants are removed or replaced through concerted evolution, variants derived from the dominant sequence are continually being generated. This process, however, leads to evolutionary stability of the dominant sequence variant(s). For this reason the numerically dominant intragenomic variant is used to conservatively characterize *Symbiodinium* (31, 36). Accordingly, DGGE-based analyses focus only on the dominant *Symbiodinium* ITS2 sequence variants in a sample, to resolve the most abundant *Symbiodinium* species present in the host. However, DGGE-based analyses are relatively labor and time intensive, which might prohibit studying a large number of samples. In comparison, pyrosequencing requires little hands-on time and has become comparatively cheap. In addition, the resolution of additional *Symbiodinium* that might occur at low abundances (<5%) may be detected. DGGE can usually detect the co-existence of a second or third *Symbiodinium* if present at >5-10% of the sample (60), with some species (e.g., type *D1* in *Pocillopora*) requiring 10-30% abundance to be detected (61). In contrast, pyrosequencing detected sequences below 1% in our data. Accordingly, DGGE-based ITS2 typing can miss the presence of background species that may be ecologically important (62).
The application of deep sequencing can be used to better understand intra-genomic rDNA diversity and its evolution within a genome. Pyrosequencing repeatedly identified a large number of ITS2 sequence variants diagnostic of a particular genome (Figure 1.3) beyond the resolution of DGGE- or cloning-based approaches. Still, the median intragenomic sequence divergence was <0.02 (Table 1.2) similar to values calculated when only the most abundant sequence variants were considered (11, 35); and this value was subsequently used to inform cutoff estimates for the resolution of separate taxa.

The analysis of mixed cultured isolates at different ratios showed that one must be conservative when interpreting abundance in samples with two or more Symbiodinium species. This is mainly due to the differences in rDNA copy number between Symbiodinium species (2, 61, 63, 64), or differences in DNA extraction efficiency among different Symbiodinium spp. Pooled cultures of Symbiodinium A1 (CCMP2467) and B1 (rt-147) in a 1:1 ratio, generated far more sequence reads for Symbiodinium B1, despite pooling the same number of Symbiodinium cells from both cultures (Table 1.1). While genomic complexity in terms of distinct copies was similar for both cultures, clade B has several times more ITS copies than clade A, explaining the unequal read distribution we recovered (LaJeunesse unpubl. data). Accordingly, this has to be taken into account when using rDNA in estimating the relative abundances of samples containing members of two or more clades.

Our discovery of ITS2 sequence variants diagnostic of a different clade in pyrosequencing data from isoclonal cultures prompted further investigations via single cell PCRs. Given the absence of consistent foreign clade amplification in single cell PCRs as well as the vast evolutionary timescales between Symbiodinium clades (estimated at tens of millions of years), the most parsimonious explanation is DNA
contamination (e.g., in the form of aerosols). In a recent review, van Oppen and Gates (65) report on the retrieval of highly divergent sequences (diagnostic of other clade entities) at ultralow abundance indicating that contamination is potentially a significant issue in laboratories that extensively employ PCRs of the ITS2 marker. The finding of unexpected ‘background’ sequences would need to be confirmed by next-generation sequencing these samples in laboratories that have not previously worked on Symbiodinium ITS2.

1.5.2 Taxon-based analysis of ITS2 diversity

While many studies have applied cluster-based approaches to analyze Symbiodinium ITS2 diversity within environmental samples (35, 66), few studies yet have targeted analyzing Symbiodinium diversity with high throughput sequencing data in an OTU-based framework (43, 44). We think that this is the most suitable approach in dealing with data produced from next generation sequencing of rDNA. It should be noted that the application of high throughput sequencing of the ITS2 region is not limited to the Roche 454 platform. Given current read-length improvements of the Illumina MiSeq platform (2*300bp Paired-End libraries) in addition to its higher throughput, data from this platform can readily be applied to the analysis of ITS2 in a taxon-based framework. In particular, our OTU-based pipeline developed here for mothur is compatible and portable to data generated on the Illumina MiSeq platform with minor changes to the quality trimming steps (Phase I, Supplemental File 1.1) based on differences in the .sff (Roche 454) and .fastq (Illumina MiSeq) file format. The challenge lies in the derivation of appropriate cutoffs for the separation of sequences into clades and species or types. Even for a well-studied marker such as 16S, taxonomic cutoffs are not necessarily accurate (67, 68). Here we used (replicated) sequencing of isoclonal cultures representing clades A, B, and C in order to better
understand inter- and intragenomic diversity and empirically derive taxonomic cutoffs. Future studies should sequence rDNA from additional cultures representing a diverse set of species.

On the clade level, due to the large evolutionary distance in the genus *Symbiodinium*, ITS2 sequences from different clades are difficult if not impossible to align. Accordingly, taxon-based framework analyses will fail to separate ITS2 sequences into ecological and evolutionarily discrete entities, when sequences from all clades are considered at the same time. Rather, it will lead to exaggerated genetic distance estimates retrieved from the alignment of non-homologous DNA characters. To resolve ITS2 types correctly, we found that sequences need to be separated into distinct clades first, and a subsequent OTU-based analysis has to be carried out for each group separately. For our data, ITS2 sequences separated into clades upon applying a cutoff value of 0.15. However, application of this cutoff to ITS2 sequences from our database considering all clades (i.e. A-I) did not correctly cluster ITS2 sequences into all distinct clades (data not shown). This might be due to difficulties in calculating pairwise similarities between sequences when all subtypes of *Symbiodinium*, e.g. A1, A2, A3, or B1, B2, B3 are present in one dataset. To separate ITS2 sequences into *Symbiodinium* clade levels, we suggest manual binning upon determination of an appropriate similarity cutoff before conducting OTU-based ITS2 type analyses (Supplemental File S1.4).

Similarly, as for the separation of ITS2 reads into clades, there is no *a priori* cutoff value that consistently sorts ITS2 sequences into distinct species (25). In fact, *Symbiodinium* species classification should not be conducted using ITS2 sequence divergence alone, as ITS2 may not consistently resolve species. For instance, two distinct *Symbiodinium* lineages that are not undergoing genetic exchange can have the
same ITS2 sequence (25). Accordingly, a taxonomic cutoff value that clusters existent ITS2 variants into OTUs should be regarded as provisional, pending subsequent analysis with independent and more rapidly evolving markers. For this reason, grouping of ITS2 variants into provisional OTUs for the (initial) characterization and comparison of *Symbiodinium* diversity seems a more feasible approach than trying to definitely resolve ITS2 sequence differences into species.

Since rDNA is not a single copy gene, a cutoff value that clusters intra-genomic ITS2 variants into a single OTU might be indicated as a valid cutoff value. In our data, clustering sequences from isoclonal cultures assorted to clades with the average neighbor algorithm (54) at a pairwise genetic distance cutoff of 0.03 collapsed intragenomic ITS2 variation into a single OTU. For this reason, we adopted 0.03 as a cutoff to determine *Symbiodinium* community composition from pyrosequencing data retrieved from environmental coral specimens. The application of this cutoff to samples from coral host tissues resulted in only 6 distinct OTUs describing *Symbiodinium* diversity, despite numerous sequenced ITS2 variants. This provides an indication of how much of the variation we sequenced might be attributable to actual intra- rather than inter-genomic diversity. Applying an OTU-based approach to the analysis of coral symbiont diversity suggests that most coral colonies appear to be dominated by a single genetic entity (11, 25, 69, 70). However, ITS2 sequences of types *C1, C1h, C1c*, and *C41* were represented by the same OTU at a 0.03 cutoff underscoring the difficulties that are associated with matching ITS2 sequence variance to closely related *Symbiodinium* with separate ecological distributions. Because these entities share a recent common ancestor as members of the *C1*-radiation (sensu 25), the genomes of these *Symbiodinium* shared the ancestral sequence, *C1*, at different relative abundances (Fig. 4). An exception was a
Pocillopora verrucosa specimen (P.Maq.R2.7, Table 1.3): in this sample we detected all 6 OTUs representing the entire OTU diversity from our data. The significance of colonies possibly harboring highly mixed Symbiodinium assemblages is not fully understood and should be the effort of future studies.

1.5.3 Considerations for the application of an OTU-based pipeline to the analysis of Symbiodinium ITS2 diversity

There is probably no single genetic marker that can correctly classify all Symbiodinium diversity into distinct species (71, 72). Typically, a distinct ITS2 variant that is numerically dominant in the genome is tentatively regarded as a new type, or “species”, of Symbiodinium (6, 31); and only when diagnostic of an ecologically distinct population. However, for (initial) characterization of Symbiodinium diversity in a large number of samples, elucidation of Symbiodinium species using multimarker molecular evidence in combination with morphological, ecological, and physiological data (as detailed in, e.g. 71) is not (yet) feasible. Here we suggest adoption of an OTU-based framework to analyze high-throughput Symbiodinium ITS2 sequence data to provisionally assess species diversity (similar to the use of DGGE fingerprinting of ITS2 rDNA) until ‘true’ species are formally described using the convergence of independent lines of genetic, ecological, and in some cases morphological evidence (73). Although the application of a taxon-based framework will miss some ecologically and evolutionarily distinct entities, it provides a conservative and comparable estimate of Symbiodinium diversity based on rDNA. The development of sensible cutoffs to delineate Symbiodinium diversity in an OTU-based framework allows harmonization and re-analysis of existing and new data under a common set of rules. We specifically acknowledge that some species will be contained in a single OTU and not resolve at a cutoff of >97% similarity. For
instance, clade C, which is by far the most diverse clade with a broad radiation, would require higher cut-off values to resolve species. But since ITS2 data alone are unable to unequivocally diagnose species, our intentions were to provide a reasonable cutoff that investigators can apply to compare diversity within and between their samples, and that deflates rather than inflates *Symbiodinium* species diversity estimates. Further, continuing efforts in the (genetic description) of *Symbiodinium* species (71, 73, 74) and the ease of typing a large number of samples might encourage the use of multiple markers in the analysis of *Symbiodinium* diversity, which will facilitate analysis of high throughput sequencing data in a phylogeny-based context. This perspective hopefully stimulates further comparative analyses of ocean basins and species in a standardized framework that might provide greater insight in symbiont diversity of marine invertebrates and the acclimation of corals to environmental change.
### 1.6 Tables and Figures

#### Table 1.1 *Symbiodinium* ITS2 pyrosequencing overview.
Depicted are clade types (when cultures were used), the respective animal host and collection site, and DGGE-typing results (based on the brightest band). For ITS2 pyrosequencing data, the total number of sequence reads (after filtering) and the number of distinct ITS2 copies for each sample is provided. Sequence reads for each sample are further assigned to clade types detailing number of sequence reads, number of distinct ITS2 copies within clades, and clade type of most abundant ITS2 copy. In cases where the clade type of the most abundant ITS2 copy is different from the database sequence, the percent identity is indicated. In cases where there are 2 or more equally abundant ITS2 copies, all clade types are indicated.

<table>
<thead>
<tr>
<th>Sample name</th>
<th>Symbiodinium culture type</th>
<th>Host and Collection Site</th>
<th>DGGE/typing</th>
<th>No. of ITS2 reads after filtering</th>
<th>No. of distinct ITS2 copies</th>
<th>Most abundant ITS2 copy</th>
<th>No. of ITS2 reads</th>
<th>No. of distinct ITS2 copies</th>
<th>Most abundant ITS2 copy</th>
<th>Most abundant ITS2 copy%</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCMP2467 (1)</td>
<td>A1</td>
<td>Stylophora pistillata, Gulf of Aqaba</td>
<td>B1</td>
<td>16,681</td>
<td>331</td>
<td>GS_A1 (16,589)</td>
<td>26</td>
<td>5</td>
<td>GS_A1 (16,589)</td>
<td>17 G5_C1</td>
</tr>
<tr>
<td>CCMP2467 (2)</td>
<td>A1</td>
<td>Stylophora pistillata, Gulf of Aqaba</td>
<td>B1</td>
<td>8,565</td>
<td>241</td>
<td>GS_A1 (8,537)</td>
<td>14</td>
<td>4</td>
<td>GS_A1 (8,537)</td>
<td>8 G5_C1 (100%)</td>
</tr>
<tr>
<td>KB8</td>
<td>A1</td>
<td>Cassiopea xamachana, Hawaii</td>
<td>A1</td>
<td>14,894</td>
<td>292</td>
<td>GS_A1 (14,882)</td>
<td>9</td>
<td>4</td>
<td>GS_A1 (14,882)</td>
<td>106 G5_C1</td>
</tr>
<tr>
<td>rt/147 (1)</td>
<td>B1</td>
<td>Pseudoterogorgia bipinnata, Jamaica</td>
<td>B1</td>
<td>16,843</td>
<td>225</td>
<td>GS_A1 (16,837)</td>
<td>0</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>rt/147 (2)</td>
<td>B1</td>
<td>Pseudoterogorgia bipinnata, Jamaica</td>
<td>B1</td>
<td>18,989</td>
<td>241</td>
<td>GS_B1 (18,989)</td>
<td>0</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>rt/064</td>
<td>B1</td>
<td>Cassiopea xamachana, Jamaica</td>
<td>B1</td>
<td>16,042</td>
<td>175</td>
<td>GS_A1 (16,040)</td>
<td>0</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>rt/152</td>
<td>C1</td>
<td>Discosoma sp., Jamaica</td>
<td>C1</td>
<td>3,116</td>
<td>111</td>
<td>GS_A1 (3,104)</td>
<td>11</td>
<td>4</td>
<td>GS_A1 (3,104)</td>
<td>/</td>
</tr>
<tr>
<td>CCMP2467:rt/147 (1:3)</td>
<td>A1/B1</td>
<td>Pooled cultures CCMP2467:rt/147 (1:3)</td>
<td>B1</td>
<td>14,915</td>
<td>223</td>
<td>GS_A1 (14,897)</td>
<td>18</td>
<td>7</td>
<td>GS_A1 (14,897)</td>
<td>/</td>
</tr>
<tr>
<td>P.Dog.R3.2</td>
<td>/</td>
<td>Pocillopora sp., Doga, Red Sea</td>
<td>A1</td>
<td>22,732</td>
<td>289</td>
<td>GS_A1 (22,724)</td>
<td>7</td>
<td>1</td>
<td>GS_A1 (22,724)</td>
<td>/</td>
</tr>
<tr>
<td>P.Maq.R2.19</td>
<td>/</td>
<td>Pocillopora sp., Maqna, Red Sea</td>
<td>C41</td>
<td>3,460</td>
<td>182</td>
<td>GS_A1 (3,400)</td>
<td>3</td>
<td>1</td>
<td>GS_A1 (3,400)</td>
<td>/</td>
</tr>
<tr>
<td>P.Waj.R1.5</td>
<td>/</td>
<td>Pocillopora sp., Al Wahj, Red Sea</td>
<td>C41</td>
<td>20,384</td>
<td>287</td>
<td>GS_A1 (20,080)</td>
<td>108</td>
<td>16</td>
<td>GS_A1 (20,080)</td>
<td>/</td>
</tr>
</tbody>
</table>

*Percent identity indicated if not 100% match.*
Table 1.2 Intragenomic diversity of ITS2 genes in isoclonal cultures (only ITS2 copies represented by at least 100 reads were considered). Differences are calculated as uncorrected pairwise distances between aligned DNA sequences. The difference between the minimum and maximum genetic difference of ITS2 variants within cultures can be used to derive a similarity cutoff for an Operational Taxonomic Unit (OTU)-based framework, upon which intragenomic diversity is contained within a given OTU.

<table>
<thead>
<tr>
<th>Symbiodinium culture type</th>
<th>Symbiodinium culture</th>
<th>Mean uncorrected genetic distance</th>
<th>Median uncorrected genetic distance</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Difference (Max-Min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCMP2467 (1)</td>
<td>A1</td>
<td>0.006</td>
<td>0.008</td>
<td>0.004</td>
<td>0.008</td>
<td>0.004</td>
</tr>
<tr>
<td>CCMP2467 (2)</td>
<td>A1</td>
<td>0.006</td>
<td>0.008</td>
<td>0.004</td>
<td>0.008</td>
<td>0.004</td>
</tr>
<tr>
<td>KB8</td>
<td>A1</td>
<td>0.013</td>
<td>0.008</td>
<td>0.004</td>
<td>0.008</td>
<td>0.004</td>
</tr>
<tr>
<td>rt-147 (1)</td>
<td>B1</td>
<td>0.005</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>rt-147 (2)</td>
<td>B1</td>
<td>0.017</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>rt-264</td>
<td>B1</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>rt-152</td>
<td>C1</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
</tbody>
</table>
Table 1.3 Symbiodinium ITS2 OTU-based analysis overview. Depicted are samples used in this study and the designated clade type based on DGGE-typing. For ITS2 pyrosequencing data, the total number of sequence reads (after filtering and clade-based alignment) and the number of OTUs at a 97% similarity cutoff for each sample is provided. Distribution of sequence reads over OTUs on a clade basis and the ITS2 clade type most similar to an OTU is detailed. In cases where the OTU is different from the ITS2 clade type database sequence, the percent identity is indicated.

<table>
<thead>
<tr>
<th>Culture or Sample name</th>
<th>DGGE typing</th>
<th>No. of ITS2 sequences</th>
<th>No. of OTUs</th>
<th>No. of sequences OTU1adj</th>
<th>ITS2 clade types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isoclonal cultures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCMP2467 (1) A1</td>
<td></td>
<td>16,681</td>
<td>3</td>
<td>16,589</td>
<td>A1</td>
</tr>
<tr>
<td>CCMP2467 (2) A1</td>
<td></td>
<td>8,565</td>
<td>3</td>
<td>8,537</td>
<td>A1</td>
</tr>
<tr>
<td>KB8 A1</td>
<td></td>
<td>14,891</td>
<td>3</td>
<td>14,879</td>
<td>A1</td>
</tr>
<tr>
<td>rt-147 (1) BI</td>
<td></td>
<td>16,836</td>
<td>2</td>
<td>6</td>
<td>B1</td>
</tr>
<tr>
<td>rt-147 (2) BI</td>
<td></td>
<td>18,987</td>
<td>1</td>
<td>0</td>
<td>B1</td>
</tr>
<tr>
<td>rt-064 BI</td>
<td></td>
<td>16,042</td>
<td>2</td>
<td>2</td>
<td>B1</td>
</tr>
<tr>
<td>rt-152 Cl</td>
<td></td>
<td>3,116</td>
<td>3</td>
<td>1</td>
<td>C1</td>
</tr>
<tr>
<td>Pooled cultures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCMP2467:rt-147 (1:1) B1</td>
<td></td>
<td>15,231</td>
<td>2</td>
<td>100</td>
<td>B1</td>
</tr>
<tr>
<td>CCMP2467: rt-147 (1:3) B1</td>
<td></td>
<td>14,911</td>
<td>2</td>
<td>18</td>
<td>B1</td>
</tr>
<tr>
<td>Field-collected coral specimens</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P.Dog.R3.2 A1</td>
<td></td>
<td>22,733</td>
<td>4</td>
<td>22,663</td>
<td>A1</td>
</tr>
<tr>
<td>P.Maq.R2.19 C1h</td>
<td></td>
<td>3,653</td>
<td>3</td>
<td>41</td>
<td>C1</td>
</tr>
<tr>
<td>P.Maq R2.7 A1</td>
<td></td>
<td>18,327</td>
<td>6</td>
<td>18,146</td>
<td>A1</td>
</tr>
<tr>
<td>P.Waj.R1.5 A1</td>
<td></td>
<td>20,284</td>
<td>4</td>
<td>20,080</td>
<td>A1</td>
</tr>
<tr>
<td>P.Waj.D7 C1c</td>
<td></td>
<td>3,579</td>
<td>3</td>
<td>38</td>
<td>C1</td>
</tr>
<tr>
<td>A.Af.B6 C4l</td>
<td></td>
<td>3,292</td>
<td>3</td>
<td>4</td>
<td>C4l</td>
</tr>
</tbody>
</table>

Some reads discarded due to filter and clade alignment
Figure 1.1 (A) Frequency distribution of sequence reads for the 10 most common ITS2 copies from isoclonal culture samples. (B) Rarefaction curve illustrating ITS2 sequence diversity distribution in isoclonal cultures.
**Figure 1.2** Denaturing Gradient Gel Electrophoresis (DGGE) fingerprint of samples used in this study with clades/subclades indicated.
**Figure 1.3** Reproducibility of pyrosequencing rDNA variants for replicated PCRs (using the same DNA extract) of isoclonal cultures CCMP2467 and rt-147 taking the top 10 most abundant sequence variants into account. The sizes of the circles represent relative abundance; circles with an asterisk represent distinct ITS2 variants.
**Figure 1.4** Comparison of top 10 most common variants found from pyrosequencing ITS2 from (A) culture rt-152 (Symbiodinium goreaui or type C1) and several field-collected samples representing clade C Symbiodinium: (B) C1c (sample P.Waj.D7), (C) C1h (sample P.Maq.R2.19), and (D) C41 (sample A.Af.B6). The size of the circles represent relative abundance, numbers in circles represent order of distinct ITS2 variants for a given sample. Shared ITS2 variants are depicted in a black outer circle representing ITS2 type C1, and a grey outer circle representing ITS2 subtype c.
1.7 Supplementary Information

Supplemental File S1.1 Symbiodinium ITS2 analysis pipeline with mothur (version 1.31.2).

_Symbiodinium_ ITS2 Analysis Pipeline with mothur (version 1.31.2), Voolstra lab, April 2014

*Note: the main difference in comparison to OTU-based analyses with 16S is that ITS2 sequences from different clades cannot be aligned properly; for this reason sequences must be separated by clade after general sequence trimming and filtering*

Phase I: Quality trimming

1. Extract Fasta, Qual, Flow files from .sff file

   ```
   mothur > sffinfo(sff=H2SYBJJ03.sff, flow=T)
   mothur > summary.seqs(fasta=H2SYBJJ03.fasta)
   ```

2. Sort out bad seqs from the flowgram (need oligos file), split seqs by barcodes (flow file)

   ```
   mothur > trim.flows(flow=H2SYBJJ03.flow, oligos=oligos.oligos, bdiffs=0, pdiffs=2, minflows=250, maxflows=800, fasta=T, processors=4)
   ```

3. Denoise the flowgrams with the PyroNoise Algorithm

   # LookUp_Titanium.pat file needed; download from mothur wiki

   ```
   mothur > shhh.flows(file=H2SYBJJ03.flow.files, processors=4)
   ```

   ð H2SYBJJ03.shhh.fasta, H2SYBJJ03.shhh.names

4. Trim sequences according to parameters

   # 2 errors in primer, 0 errors in barcode; discard any seqs ≥ 1 ambiguous bp; discard seqs < 250bp
   # oligo file holds barcode and forward sequencing primer sequence

   ```
   mothur > trim.seqs(fasta=H2SYBJJ03.shhh.fasta, name=H2SYBJJ03.shhh.names, oligos=oligos.oligos, pdiffs=2, bdiffs=0, maxambig=0, maxhomop=4, minlength=250, processors=4, allfiles=T)
   ```

   ð H2SYBJJ03.shhh.scrap.fasta

   # Check why seqs are being discarded from trim.seqs in the scrap.fasta file

   ```
   bash$ cut -d ' ' -f 1 H2SYBJJ03.shhh.scrap.fasta | perl -F'|' -ane 'Sh\{ $F[1] }++; END {print "$_\t$h\{_\}" for keys %h}'> H2SYBJJ03.shhh.scrap.fasta.count
   ```

   # Check summary.seqs to determine number of seqs kept

   ```
   mothur > summary.seqs(fasta=H2SYBJJ03.shhh.trim.fasta, name=H2SYBJJ03.shhh.trim.names)
   ```

   # If many sequences are lost, it might be sensible to increase maxhomop

5. Trim reverse primer using cutadapt version 1.1

   # Allow error rate = 0.15 to allow 2-3 indel/mismatches; reverse primer seq should be in revcomp direction

   ```
   bash$ cutadapt -a ACCCGCTGAACCTAAGCATATGGATCCC -e 0.15 H2SYBJJ03.shhh.trim.fasta > H2SYBJJ03.shhh.revtrim.fasta
   ```
6. Collapse identical seqs
# Only the representative seqs are kept to reduce computation time
# counts of actual number of seqs are retained

```bash
mothur > unique.seqs(fasta=H2SYBJJ03.shhh.trim.revtrim.fasta, name=H2SYBJJ03.shhh.trim.names)
mothur > summary.seqs(fasta=H2SYBJJ03.shhh.trim.revtrim.unique.fasta, name=H2SYBJJ03.shhh.trim.revtrim.names)
mothur > count.seqs(name=H2SYBJJ03.shhh.trim.revtrim.names, group=H2SYBJJ03.shhh.groups)
```

7. Check for Chimeras using UCHIME and remove them
# (http://drive5.com/usearch/manual/uchime_algo.html)

```bash
mothur > chimera.uchime(fasta=H2SYBJJ03.shhh.trim.revtrim.unique.fasta, name=H2SYBJJ03.shhh.trim.revtrim.names, group=H2SYBJJ03.shhh.groups, processors=4)
mothur > remove.seqs(accnos=H2SYBJJ03.shhh.trim.revtrim.unique.uchime.accnos, fasta=H2SYBJJ03.shhh.trim.revtrim.unique.fasta, name=H2SYBJJ03.shhh.trim.revtrim.names, group=H2SYBJJ03.shhh.groups)
mothur > summary.seqs(name=H2SYBJJ03.shhh.trim.revtrim.pick.names)
```

8. Remove singletons in mothur
# Seqs detected only once across the entire dataset are not included in further analyses

```bash
mothur > split.abund(cutoff=1, fasta=H2SYBJJ03.shhh.trim.revtrim.unique.pick.fasta, name=H2SYBJJ03.shhh.trim.revtrim.pick.names, group=H2SYBJJ03.shhh.pick.groups)
```

# Keep record of sequences kept/lost
```bash
mothur > summary.seqs(fasta=H2SYBJJ03.shhh.trim.revtrim.unique.pick.abund.fasta, name=H2SYBJJ03.shhh.trim.revtrim.pick.abund.names)
```

# count.seqs for assessing frequency distribution of ITS2 copies
```bash
mothur > count.seqs(name=H2SYBJJ03.shhh.trim.revtrim.pick.abund.names, group=H2SYBJJ03.shhh.pick.abund.groups)
```

# Sanity check: align the most abundant seqs in all samples to their respective DGGE seqs to check if they are identical
# Copy to files with short names, its2.fasta, its2.names and its2.groups

```bash
bash$ cp H2SYBJJ03.shhh.trim.revtrim.unique.pick.abund.fasta its2.fasta
bash$ cp H2SYBJJ03.shhh.trim.revtrim.pick.abund.names its2.names
bash$ cp H2SYBJJ03.shhh.pick.abund.groups its2.groups
```

Phase II: Clades separation

9. Cluster Sequences according to clades and assess divergence cutoffs for clades
# First, calculate pairwise similarity between seqs using pairwise.seqs command
# Then cluster based on similarity scores (which pairwise.seqs command gives out) using average neighbor option

```bash
mothur > pairwise.seqs(fasta=its2.fasta, calc=onegap, countends=F, processors=4)
mothur > cluster(column=its2.dist, name=its2.names, method=average)
mothur > make.shared(list=its2.an.list, group=its2.groups)
```
its2.an.shared (provides size and number of OTUs over samples for given cutoff)

10. Assign OTUs to clades

# mothur calculates exact distant cutoffs and provides only results from those cutoffs. Therefore, we have to check seqs in each cutoff and choose a cutoff that separate clusters into different clades (usually it should be > 0.10); caution: use next higher cutoff value as mothur per default selects previous lower one

# Get representative sequence for each OTU (i.e. clade level at > 0.10)

```r
mothur > get.oturep(column=its2.dist, name=its2.names, fasta=its2.fasta, list=its2.an.list, group=its2.groups, label=0.15)
```

# Find out which OTU represents which clade via BLASTn (in this data set: 3 sequences)

```r
mothur > system(blastn -db ~/Databases/ITS2KAUST -query its2.an.0.15.rep.fasta -out 0.15.blast -outfmt 6 -max_target_seqs 1)
```

# Display BLAST outputs

```r
mothur > system(cat 0.15.blast)
```

# In this example: at 0.15 cutoff, there are 3 clusters and they belong to clade A, B and C therefore, this is the cutoff to choose. CHECK via BLASTn that the OTUs generated belong to different clades

11. Generate a fasta file of all sequences associated to OTUs at the clade level

```r
mothur > bin.seqs(name=its2.names, fasta=its2.fasta, list=its2.an.list, label=0.15)
```

⇒ its2.an.0.15.fasta (this is the complete 454 data sorted according to clade level)

12. Separate sequences by clade-level OTUs to different fasta files

# Save files as its2.A.fasta, its2.B.fasta and its2.C.fasta

# For example, using grep

```bash
bash$ grep -A 1 -P '>.+\t1$' its2.an.0.15.fasta > its2.A.fasta
bash$ grep -A 1 -P '>.+\t2$' its2.an.0.15.fasta > its2.B.fasta
bash$ grep -A 1 -P '>.+\t3$' its2.an.0.15.fasta > its2.C.fasta
```

13. Sanity check: BLAST fasta file from each clade to check that all sequences match the same clade.

```bash
bash$ blastn -db ~/Databases/ITS2KAUST -query its2.A.fasta -out its2.A.blast -outfmt 6 -max_target_seqs 1
bash$ blastn -db ~/Databases/ITS2KAUST -query its2.B.fasta -out its2.B.blast -outfmt 6 -max_target_seqs 1
bash$ blastn -db ~/Databases/ITS2KAUST -query its2.C.fasta -out its2.C.blast -outfmt 6 -max_target_seqs 1
```

14. Collapse to unique seqs per clade to reduce computation time

# Do for each clade identified

```r
mother > unique.seqs(fasta=its2.A.fasta)
mother > summary.seqs(fasta=its2.A.unique.fasta, name=its2.A.names)
```

15. Align unique seqs of each clade with MUSCLE

# Do for each clade identified

```r
mothur > system(muscle -in its2.A.unique.fasta -out its2.A.unique.align)
```
16. Trim seqs from each clade to equal length

# The format of aligned fasta files need to be modified since MUSCLE does not put "." at the beginning or end of alignments, and outputs block fasta format. The following command changes that

# Do for each clade identified

bash$ perl -ne 'if ($_. == 1) {print; next}; chomp; if (/^$/) { print "\n", $_[0], "\n" } else { print }'
its2.A.unique.align | perl -pe 's/^(-+)/." x length($1)/eg; s/(.-)"/"." x length($1)/eg;' > its2.A.unique.dots.align

mothur > summary.seqs(fasta=its2.A.unique.dots.align, name=its2.A.names)

# The following needs to be adjusted for the actual alignment; choose start and end so that a majority of the sequences can be retained (here: optimize option = 90 was used in the screen.seqs, which automatically trim to keep 90% of the sequences)

mothur > screen.seqs(fasta=its2.A.unique.dots.align, optimize=start-end, criteria=90, name=its2.A.names)

mothur > summary.seqs(fasta=its2.A.unique.dots.good.align, name=its2.A.names)

# The filter.seqs command will cut off ‘overhangs’ at both ends of the alignment, and remove any columns that are all gaps

mothur > filter.seqs(fasta=its2.A.unique.dots.good.align, trump=., vertical=T)
mothur > summary.seqs(fasta=its2.A.unique.dots.good.filter.fasta, name=its2.A.good.names)

17. De-collapse unique seqs for each clade to get all seqs, save to fasta file

# Do for each clade identified

mothur > deunique.seqs(fasta=its2.A.unique.dots.good.filter.fasta, name=its2.A.good.names)

⇒ its2.A.redundant.fasta

18. Extract seq IDs and sample names from each clade's fasta file for use as a seq ID file and a group file in mothur

# to continue analysis in mothur group files are needed for each clade
# Save as its2.all.A.seqIDs and its2.all.A.groups, accordingly for all clades identified
# Do for each clade identified

bash$ grep \> its2.A.redundant.fasta | perl -pe '\s/>//>' its2.all.A.seqIDs

# Get the group information from main group file, write to new clade group file.

bash$ perl -ne 'BEGIN{open $IN, ">", "its2.groups"; while (<$IN>) {chomp; Sdb{[\{ split \s/ / \}[0]}}=\$_; } chomp; if ( exists $db{\$_} ) { print "$db{\$_}\n" }' its2.all.A.seqIDs > its2.all.A.groups

Phase III: ITS2 type separation

# It is recommended to do the following steps in separate directories per clade
# Start the community analysis for each clade separately but over all samples

19. Use the outputs from step 18. Collapse identical sequences to reduce computation time

# Do for each clade identified

mothur > unique.seqs(fasta=its2.A.redundant.fasta)

20. Calculate uncorrected pairwise distance and cluster seqs into OTUs at a 0.03 cutoff

# Do for each clade identified
21. Get a representative sequence (the most abundant sequence) for each OTU at a 0.03 cutoff and annotate with local database (BLASTn)
   # Do for each clade identified
   
   bash$ blastn -db ~/Databases/ITS2KAUST -query its2.A.redundant.unique.an.0.03.rep.fasta -out its2.A.0.03rep.blast -outfmt 6 -max_target_seqs 3

22. Get ITS2 sequence distribution over all samples for any given clade (optional)
   # Do for each clade identified
   
   Supplemental File S1.2 Script file for unattended mothur ITS2 analysis.
   
   System editing and chimera check
   system(mkdir edit)
   system(mv H2SYBJJ03.shhh.fasta H2SYBJJ03.shhh.names oligos.oligos ./edit)
   set.dir(output=./edit, input=./edit)
   
   # Trimming, make group file
   trim.seqs(fasta=H2SYBJJ03.shhh.fasta, name=H2SYBJJ03.shhh.names, oligos=../oligos.oligos, 
pdiffs=2, bdiffs=0, maxambig=0, maxhomop=4, minlength=250, processors=4, allfiles=T)
   system(cut -d ' ' -f 1 ./edit/H2SYBJJ03.shhh.scrap.fasta | perl -F'|' -ane 'sh{ $F[1] }++'; END{print "$ \sh{$_}" for keys $\sh;}' > ./edit/H2SYBJJ03.shhh.scrap.fasta.count)
   summary.seqs(fasta=H2SYBJJ03.shhh.trim.fasta, name=H2SYBJJ03.shhh.trim.names)
   system(cutadapt -a ACCCGCTGAACTTAAGCATATGGATCCC -e 0.15 ./edit/H2SYBJJ03.shhh.trim.fasta > ./edit/H2SYBJJ03.shhh.trim.revtrim.fasta)
   unique.seqs(fasta=H2SYBJJ03.shhh.trim.revtrim.fasta, name=H2SYBJJ03.shhh.trim.revtrim.names)
   system(mkdir revtrim)
   system(cutadapt -a ACCCGCTGAACTTAAGCATATGGATCCC -e 0.15 ./edit/H2SYBJJ03.shhh.trim.revtrim.fasta > ./edit/H2SYBJJ03.shhh.trim.revtrim.revtrim.fasta)
   unique.seqs(fasta=H2SYBJJ03.shhh.trim.revtrim.revtrim.fasta, name=H2SYBJJ03.shhh.trim.revtrim.revtrim.names)
   summary.seqs(fasta=H2SYBJJ03.shhh.trim.revtrim.unique.fasta, name=H2SYBJJ03.shhh.trim.revtrim.unique.names)
   
   # Chimera check
chimera.uchime(fasta=H2SYBJJ03.shhh.trim.revtrim.unique.fasta, 
name=H2SYBJJ03.shhh.trim.revtrim.names, group=H2SYBJJ03.shhh.groups, processors=4)

remove.seq(accnos=H2SYBJJ03.shhh.trim.revtrim.unique.accnos, 
fasta=H2SYBJJ03.shhh.trim.revtrim.unique.fasta, name=H2SYBJJ03.shhh.trim.revtrim.names, 
group=H2SYBJJ03.shhh.groups)

summary.seqs(name=H2SYBJJ03.shhh.trim.revtrim.pick.names)

# Remove singletons

split.abund(cutoff=1, fasta=H2SYBJJ03.shhh.trim.revtrim.unique.pick.fasta, 
name=H2SYBJJ03.shhh.trim.revtrim.pick.names, group=H2SYBJJ03.shhh.pick.groups)

summary.seqs(fasta=H2SYBJJ03.shhh.trim.revtrim.unique.pick.abund.fasta, 
name=H2SYBJJ03.shhh.trim.revtrim.pick.abund.names)

count.seqs(name=H2SYBJJ03.shhh.trim.revtrim.pick.abund.names, 
group=H2SYBJJ03.shhh.pick.abund.groups)

# Counting on clade level

system(mkdir clustering)

set.dir(output=./clustering, input=./clustering)

system(cp ./edit/H2SYBJJ03.shhh.trim.revtrim.unique.pick.abund.fasta ./clustering/its2.fasta)

system(cp ./edit/H2SYBJJ03.shhh.trim.revtrim.pick.abund.names ./clustering/its2.names)

system(cp ./edit/H2SYBJJ03.shhh.pick.abund.groups ./clustering/its2.groups)

# Calculate distance matrix

pairwise.seqs(fasta=its2.fasta, calc=onegap, countends=F, processors=4)

# Cluster

cluster(column=its2.dist, name=its2.names, method=average)

make.shared(list=its2.an.list, group=its2.groups)

# Find OTU identities (which clade)

get.oturep(column=its2.dist, name=its2.names, fasta=its2.fasta, list=its2.an.list, group=its2.groups, label=0.15)

system(blastn -db ./its2_ref_blast_db/ITS2KAUST -query ./clustering/its2.an.0.15.rep.fasta -out ./clustering/0.15.blast -outfmt 6 -max_target_seqs 1)

system(cat ./clustering/0.15.blast)

# Sort sequences to OTU to enable per-clade processing and alignment

bin.seqs(name=its2.names, fasta=its2.fasta, list=its2.an.list, label=0.15)

# Analysis clade A

system(mkdir ./cladeA)

set.dir(output=./cladeA, input=./cladeA)

# Get sequences for this OTU/clade. The important part is the number 1 in '>'.+t1S which tells grep which OTU to pick out

system(grep -A 1 -P '>'.+t1S' ./clustering/its2.an.0.15.fasta > ./cladeA/its2.A.fasta)

system(blastn -db ./its2_ref_blast_db/ITS2KAUST -query ./cladeA/its2.A.fasta -out ./cladeA/its2.A.blast -outfmt 6 -max_target_seqs 1)

unique.seqs(fasta=its2.A.fasta)

summary.seqs(fasta=its2.A.unique.fasta, name=its2.A.names)

# Align with MUSCLE (may take long)

system(muscle -i ./cladeA/its2.A.unique.fasta -o ./cladeA/its2.A.unique.align)

system/perl -ne if ($_. eq 1) { print; next}; chomp; if (/^>/) { print "$\n\n\n" } else { print ]

./cladeA/its2.A.unique.align | perl -pe 's/(\+/)/\n\n\n\n/x length($1)eg; s/(\-)\n\n\n\n/x length($1)eg;' > ./cladeA/its2.A.unique.dotalign

summary.seqs(fasta=its2.A.unique.dotalign, name=its2.A.names)

#!!! the values need to be adjusted in the following
screen.seqs(fasta=its2.A.unique.dots.align, optimize=start-end, criteria=90, name=its2.A.names)
summary.seqs(fasta=its2.A.unique.dots.good.align, name=its2.A.names)
filter.seqs(fasta=its2.A.unique.dots.good.align, trump=., vertical=T)
summary.seqs(fasta=its2.A.unique.dots.good.filter.fasta, name=its2.A.good.names)
# make groups file
deunique.seqs(fasta=its2.A.unique.dots.good.filter.fasta, name=its2.A.good.names)
system(grep > ./cladeA/its2.A.redundant.fasta | perl -pe 's/>//'> ./cladeA/its2.all.A.seqIDs)
# the following uses => instead of commas in the Perl because of a bug in mothur
system perl -ne 'BEGIN{open $IN => "./clustering/its2.groups"; while (<$IN>){|chomp; $db{\{split \'/\'.{\[0\]}=>\$_\}}} chomp; if ( exists $db{\$_} ) { print "$db{\$_}\n" } }./cladeA/its2.all.A.seqIDs > ./cladeA/its2.all.A.groups)

## Clustering on type level (i.e. OTU0.03)
system(mkdir ./cladeB)
set.dir(output=./cladeB, input=./cladeB)
# get sequences for this OTU/clade. The important part is the number 2 in '>.+\t2$' which tells grep which OTU to pick out
system(grep -A 1 -P '>.+=\t2$' ./clustering/its2.an.0.15.fasta > ./cladeB/its2.B.fasta)
system(blastn -db ./its2_ref_blast_db/ITS2KAUST -query ./cladeB/its2.B.fasta -out ./cladeB/its2.B.blast -outfmt 6 -max_target_seqs 1)
unique.seqs(fasta=its2.B.fasta)
summary.seqs(fasta=its2.B.unique.fasta, name=its2.B.names)
# align with MUSCLE (may take long)
system(muscle -in ./cladeB/its2.B.unique.fasta -out ./cladeB/its2.B.unique.align)
system perl -ne 'if ($._ == 1) {print next;}; chomp; if (/^>/) { print "\n\n\$_\n" x length($1)/eg; s/(-$\$_)$\/_ x length($1)/eg;} > ./cladeB/its2.B.unique.dot
summary.seqs(fasta=its2.B.unique.dots.align, name=its2.B.names)
# !!! the values need to be adjusted in the following
screen.seqs(fasta=its2.B.unique.dots.align, optimize=start-end, criteria=90, name=its2.B.names)
summary.seqs(fasta=its2.B.unique.dots.good.align, name=its2.B.names)
filter.seqs(fasta=its2.B.unique.dots.good.align, trump=., vertical=T)
summary.seqs(fasta=its2.B.unique.dots.good.filter.fasta, name=its2.B.good.names)
# make groups file
deunique.seqs(fasta=its2.B.unique.dots.good.filter.fasta, name=its2.B.good.names)
system(grep > ./cladeB/its2.B.redundant.fasta | perl -pe 's/>//'> ./cladeB/its2.all.B.seqIDs)
# the following uses => instead of commas in the Perl because of a bug in mothur
system(perl -ne 'BEGIN {open $IN => "./clustering/its2.groups"; while (<$IN>) {chomp; $db{split /\s/ \[0\]} = $$_;} chomp; if ( exists $db{$_} ) { print "$db{$_}\n" } } ./cladeB/its2.all.B.seqIDs > ./cladeB/its2.all.B.groups')

## Clustering on type level (i.e. OTU0.03)

## Clustering on type level (i.e. OTU0.03)

system(mkdir ./cladeB/opt2)
system(cp ./cladeB/its2.all.B.seqIDs > ./cladeB/its2.all.B.groups)

system(mkdir ./cladeC)
system(grep -A 1 -P '>.+\t3' ./clustering/its2.an.0.15.fasta > ./cladeC/its2.C.fasta)

system(blastn -db ./its2_ref_blast_db/ITS2KAUST-query .cladeC/its2.C.fasta -out ./cladeC/its2.C.blast -outfmt 6 -max_target_seqs 3)

system(muscle -in ./cladeC/its2.C.unique.fasta -out ./cladeC/its2.C.unique.align)

# !!! the values need to be adjusted in the following

system(mkdir ./cladeC/opt2)
system(cp ./cladeC/its2.all.C.seqIDs > ./cladeC/its2.all.C.groups)

## Clustering on type level (i.e. OTU0.03)
unique.seqs(fasta=its2.C.redundant.fasta)
dist.seqs(fasta=its2.C.redundant.unique.fasta)
cluster(column=its2.C.redundant.unique.dist, name=its2.C.redundant.names, method=average)
make.shared(list=its2.C.redundant.unique.an.list, group=its2.all.C.groups)
get.oturep(fasta=its2.C.redundant.unique.fasta, list=its2.C.redundant.unique.an.list,
name=its2.C.redundant.names, label=0.03, method=abundance)
system(blastn -db ./its2_ref_blast_db/ITS2KAUST -query ./cladeC/opt2/its2.C.redundant.unique.an.0.03.rep.fasta -out ./cladeC/opt2/its2.C.0.03rep.blast -outfmt 6 -max_target_seqs 3)
count.seqs(name=its2.C.redundant.names, group=its2.all.C.groups)
rarefaction.single(shared=its2.C.redundant.unique.an.shared)
summary.single(shared=its2.C.redundant.unique.an.shared, calc=chao-sobs-simpson-invsimpson, label=0.03)

Supplemental File S1.3 Overview over ITS2 classification, read counts, and
reference sequence for all ITS2 variants identified.

Data archived online KAUST library.
Supplemental File S1.4 *Symbiodinium* ITS2 454 data analysis pipeline

**ITS2 Symbiodinium 454 Data Analysis Pipeline**

**Phase I: Quality trimming**

Raw 454 reads → Discard the following: seq length < 250bp, ambiguous seq >0bp, barcode mismatch >0bp, primer mismatch >2bp, homopolymer >4 bp) → Trim reverse primer via cutadapt (error rate = 0.15) → Remove singletons (seqs represented only once across all dataset) → Discard chimera seqs (UCHIME)

Create a file denoting ITS2 variants frequency distribution over all samples

**Phase II: Clades separation**

BLAST all distinct ITS2 copies to local ITS2 database → Calculate pairwise distances between all sequences → Cluster all sequences (average neighbor) and select a cutoff value that separates sequences into clades* → Create a make.shared file to view the break down of sequences in all samples into OTUs at different cutoffs

Trim sequences in each clade to start and end at the same position (discard sequences shorter than 90% of all sequences) → Create a fasta file for each clade and align sequences in each clade via MUSCLE

**Phase III: ITS2 type separation**

Calculate distance matrix for sequences in each clade → Cluster seqs into OTUs (average neighbor, 0.03 cutoff) → Create a make.shared file for each clade to view the break down of sequences in all samples into OTUs

Create a file denoting ITS2 variants frequency distribution over all samples in each clade → Annotate the representative sequences to the local database (BLASTn) → Choose the most abundant sequences as a representative sequence for each OTU

*The cutoff value to separate seqs into clades depends on clade compositions in the data and might be different depending on the 454 dataset. In this study, the cutoff value was 0.15. For some datasets, it might be necessary to bin sequences into their corresponding clades manually.
1.8 References


LaJeunesse TC & Pinzon JH (2007) Screening intragenomic rDNA for dominant variants can provide a consistent retrieval of evolutionarily persistent ITS (rDNA) sequences. Molecular Phylogenetics and Evolution 45(1):417-422.


CHAPTER 2: Next Generation Sequencing-based analysis of coral-associated *Symbiodinium* diversity around the Arabian Peninsula

Chatchanit Arif\(^1\), John Burt\(^2\), Cornelia M. Roder\(^1\), Adrian Barbrook\(^3\), Christopher J. Howe\(^3\), Sergey Dobretsov\(^4\), Todd LaJeunesse\(^5\), Christian R. Voolstra\(^1\)

\(^1\)King Abdullah University of Science and Technology, Saudi Arabia,  
\(^2\)New York University, Abu Dhabi  
\(^3\)University of Cambridge, United Kingdom,  
\(^4\)Sultan Qaboos University, Oman,  
\(^5\)Penn State University, United States
2.1 Abstract

Coral reef ecosystems rely on the symbiosis between scleractinian corals and intracellular, photosynthetic dinoflagellates of the genus *Symbiodinium*. These symbionts are biologically diverse and exhibit discrete patterns of depth, geographical, and coral host distribution. This makes the assessment of *Symbiodinium* diversity critical to our understanding of ecological resilience of corals. Corals in the Red Sea and Arabian/Persian Gulf persist in high temperature environments, making them good candidates to study how the symbiosis between corals and their symbionts is affected by environmental change. In contrast, the Gulf of Oman is characterized by upwelling of cold deep-sea water. Despite the significance of the Arabian Seas as ‘Future Ocean’ environments, research on these regions is still very limited.

To understand the potential impact of climate change on coral-dinoflagellate symbioses, we applied next generation sequencing of 800 coral specimens from the Arabian Seas covering over 40 genera via analysis of the Internal Transcribed Spacer 2 (ITS2) rRNA gene from *Symbiodinium*. Besides phylotype-based analyses, we analyzed *Symbiodinium* diversity in an OTU-based framework. This allowed us to compare ocean basins and coral genera in a standardized manner.

Corals in the Red Sea, Arabian/Persian Gulf, and Gulf of Oman display distinct associations with *Symbiodinium* that are different from the Caribbean and Indo-Pacific. Our data show that corals in the Arabian seas are dominated by *Symbiodinium* clade C. Further, the application of an OTU-based framework to the analysis of coral symbiont diversity suggests that most coral colonies appear to be dominated by a single *Symbiodinium* genetic entity, despite substantial intragenomic diversity.
The ease of typing a large number of samples at a comparatively small cost with high throughput sequencing methods in an OTU-based framework facilitates comparative analyses of ocean basins and coral hosts that will hopefully provide further insight into the nature of coral resilience in regard to the underlying symbiont type.

2.2 Introduction

Some of the most extreme and the most understudied marine diversity in the world are the seas surrounding the Arabian Peninsula, including the Red Sea, Gulf of Oman, and Persian/Arabian Gulf (henceforth referred to as ‘the Gulf’). The Red Sea is oligotrophic and has high temperature and salinity due to low precipitation rates, low freshwater inflow, and limited exchange with the Indian Ocean (1). Moreover, these parameters show a gradient from north to south i.e. temperature in the Red Sea is lower in the north (24°C) and increases towards to south (30°C) (Figure 2.1), while salinity is highest in the north (41PSU) and gradually decreases towards the south (36PSU) (2, 3). The environmental conditions in the Gulf are arguably among the most extreme in the world where coral reefs exist. Corals in the Gulf are exposed to high fluctuation in temperature (ranging from 11-36°C, Figure 2.1), and salinity maxima (>39PSU throughout the Gulf, and >70PSU at the Gulf of Salwahi) (4). The conditions experienced by corals in the Red Sea and the Gulf are generally beyond the limits that corals elsewhere can survive, and extend the threshold previously believed to be tolerable by corals (5). Hence, studying coral reef diversity in these regions might be the key to discover evolutionary adaptations that corals acquired in order to survive under extreme environmental conditions.
On the contrary, the conditions in Gulf of Oman are less extreme than the Gulf and the Red Sea despite being physically connected via the Strait of Hormuz and the Gulf of Aden. This is due to monsoonal upwelling that occurs in the summer, bringing cool, nutrient rich deeper waters to the surface (6). And because of this, summer sea temperatures in Gulf of Oman are relatively benign compared with the Gulf and the Red Sea, and fluctuate between 23-31°C, while in winter the temperatures are between 22-23°C (Figure 2.1) (7). Salinity is also quite constant in the Gulf of Oman and is around 36.5 PSU throughout the year (7). The conditions in Gulf of Oman in terms of temperature and salinity profiles are therefore, comparable to other major oceans around the world (i.e. the Great Barrier Reefs (8)). The contrasting profiles between the Red Sea, the Gulf, and Gulf of Oman serve as interesting localities for comparison of *Symbiodinium* diversity associated to marine invertebrates host, especially in corals.

Reef building corals are the key organisms that structure reef ecosystems and provide habitat to a diverse set of marine species, some of which are important economically and environmentally. The ability of sclerectinian corals to form reef structures critically depends on their photosynthetic endosymbiont in the genus *Symbiodinium* since the symbionts are the main energy provider to the host (9). There are six clades of *Symbiodinium* (clade A-D, F and G) (10-13) that are found to be in association with corals and other invertebrates. These clades are further diversified into hundreds of subclades, some of which varies in their nutritional benefits to the hosts (14-16), and in their responses to thermal stress and light intensity (17-23). For example, *Symbiodinium* types within clade D are presumed to be thermally tolerant and help prevent corals from bleaching, or quicken the recovery of bleached corals (20-22, 24).
To date, *Symbiodinium* diversity has been studied in many coral species from various locations including the Caribbean, the Indo-Pacific and the Great Barrier Reefs e.g. (25-31). Corals can harbor multiple genotypes of symbionts within individual colonies (11, 32-37), with one or two types being dominant in the symbiont population (38-42). Moreover, some species of coral host specific *Symbiodinium* clades or subclades across space and time (specialists) (25, 26, 28, 43), while other hosts are more flexible and associate with various *Symbiodinium* types (generalists) (44, 45). Corals are also able to shift community of their associated symbionts to the ones that would enhance their resilience towards environmental stress such as increased temperatures (35), but whether the *Symbiodinium* community is acquired *de novo* or was selected from background symbiont community remains unresolved.

Internal transcribed spacer 2 (ITS2), a spacer region within rRNA genes, is a commonly used marker for *Symbiodinium* diversity typing. Because of the tandem repeat rRNA genes, ITS2 is a multi-copy marker, and hence, discerning inter- from intra-genomic variations has proven to be a challenge. Traditional methods used for symbiont ITS2 typing, including Denaturing Gradient Gel Electrophoresis (DGGE) and bacterial cloning, are considered to either under- or over-estimate the symbionts community structure (46, 47). DGGE, despite being very robust in typing the dominant symbiont and separating different ITS2 variants, lacks the sensitivity to correctly identify background communities especially those that are <5-10% abundant (21, 46). On the other hand, bacterial cloning tends to produce pseudo-ITS2 types, which complicate the interpretation of true symbionts community within a host (47). Recently, Next Generation Sequencing (NGS) has been utilized for typing *Symbiodinium* ITS2 diversity,
which enable the identification of symbionts that are below 1% in abundance (41, 42, 48). The high throughput technology creates an opportunity for scientists to study the unexplored background symbionts community structure. An NGS pipeline recently developed by Arif et al. (42) successfully collapsed inter-genomic ITS2 variants of isoclonal cultures into a single Operational Taxonomic Units (OTUs) using a 97% similarities cutoff. Although application of OTU-based analysis for Symbiodinium ITS2 diversity typing is still provisional, and could be too conservative in OTU estimations, the approach efficiently reduces the complexity of the data produced by high throughput sequencing and allows for economic and efficient analysis of a large number of coral colonies in a common framework.

This chapter aims to provide a comprehensive assessment of coral-Symbiodinium association in the seas around the Arabian Peninsula with NGS technologies using the pipeline developed recently (42), and also to assess whether the diversity of Symbiodinium are structured according to geographic locations or physiological conditions.

2.3 Materials And Methods

2.3.1 Sample Collection and environmental conditions

A total of 796 specimens of corals representing 43 genera were collected according to their prevalence at several reefs in the Red Sea (January and March 2014), the Gulf (September 2012), and Gulf of Oman (September 2011-12, and May 2012), by SCUBA at depths between 3-15 meters (Figure 2.2 and Figure 2.3). Corals were classified to their respective genus level via photographs. All
samples were preserved in DMSO/NaCl buffer (49) at 4°C until DNA extraction. For all sites, 4km resolutions of five year annual average (2009-2013) of sea surface temperature (SST), Chlorophyll a (Chl-a), Particulate Organic Carbon (POC) and Particulate Inorganic Carbon (PIC) data were obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard the NASA Aqua satellite (http://gdata1.sci.gsfc.nasa.gov). Average annual salinity (September 2013 to September 2014) in the Red Sea was obtained from CTD measurements at the sampling sites. As for The Gulf and Gulf of Oman, annual average salinity were obtained from Bauman (50).

2.3.2 DNA extraction and 454 pyrosequencing

DNA from specimens collected was extracted with Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following manufacturer’s protocol with minor modification. Briefly, 100mg of specimen were added to 1.5ml eppendorf tubes containing 500µl sterile glass beads (BioSpec, Bartlesville, OK), 400µl of buffer AP1 and 4µl RNAse (Qiagen, Hilden, Germany). Samples were homogenized with Qiagen tissue Lyzer II (Qiagen, Hilden, Germany) for 1min and extraction were continued according to protocol. All DNA samples were quantified with Qubit broad range DNA assay using SpectraMax Paradigm Multi-Mode Microplate Detection Platform (Molecular Devices, CA, USA) and normalized to 30ng/ml for PCR reaction. PCR amplification of the ITS2 gene for pyrosequencing was performed using primers ITSintfor2 and ITS2-reverse. The primer sequences were 5’- CCAATTGCAGAAC TCCGTG-3’ (454-ITSintfor2) and 5’- CCTATCCCTG TGTGCCCTGGCAGTCTCAGGG GATCCCATATGCTTAAGTTCAGCGGGT-3’ (454-ITS2-reverse). Primers included 454 Lib-L library adapters (underlined) and
a barcode (shown as N (51)). PCRs were run in triplicate per sample with 12.5µl of Qiagen Multiplex PCR Kit (Qiagen, Hilden, Germany), 0.1µM primers, 30ng DNA, and volume adjusted to 25µl with DNAse-free water. The following PCR conditions were used: 15min at 94°C, followed by 35 cycles of 94°C for 30s, 51°C for 30s, 72°C for 3s, and a final extension step of 10min at 72°C. PCR products were run on a 1% agarose gel stained with 1x SYBR Safe (Invitrogen, Carlsbad, CA) to visualize successful amplification. For each sample, triplicate PCR products were pooled and their DNA concentrations were measured with Qubit broad range DNA assay using SpectraMax Paradigm Multi-Mode Microplate Detection Platform (Molecular Devices, CA, USA). 8ng of each pooled sample were combined and ran on a 1% agarose gel to remove excess primers. The gel band was excised, purified with the Qiagen MinElute Gel Extraction Kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions, quantified with Qubit 2.0 (Invitrogen, Carlsbad, USA), and quality checked via Bioanalyzer (Agilent, Santa Clara, CA). 100ng of each pooled library were submitted to Macrogen (Korea) for sequencing using Titanium FLX chemistry. Two libraries of PCR products from Gulf of Oman samples and one library of PCR product from the Gulf samples were sequenced on a quarter of a picotiter plate per library. Two PCR products from the Red Sea samples were sequenced on half a picotiter plate per library. Raw sequencing data were retrieved with Roche 454 amplicon processing pipeline.

2.3.3 Pyrosequencing data processing

Libraries from Gulf of Oman produced 322,527 and 310,441 reads with average read length of 314.30bp and 310.05bp respectively. Gulf samples produced 325,252 total reads with average read length of 318.20bp. Two libraries from Red
Sea samples produced 934,080 and 948,659 reads with average read length of 313.31 and 309.66 respectively. The sequences of each library were primer trimmed, barcode removed, and chimera checked according to the pipeline developed by Arif et al. (42). The reads from all regions were pooled together, singletons were removed, and redundant reads were collapsed to 30,968 distinct ITS2 types, with an average of 165.98 distinct ITS2 copies per samples. For the OTU-based framework analysis, ITS2 variants were assigned to their corresponding ITS2 types via BLASTn to the custom database before assorted into their respective clades. The sequences within each clade were aligned with MUSCLE (52), trimmed to equal length, and discarded for being shorter than 90% of the reads, which resulted in further reduction of reads to 2,245,665. Finally, the sequences were clustered at 0.03 cutoff based on the pipeline reported in Arif et al. (42). OTUs from each clade were designated as OtuA00001, OtuA00002, OtuB00001, OtuB00002, and continued through OtuG00001. The OTUs were subsequently assigned to their corresponding ITS2 types via BLASTn to the custom database as reported in Arif et al. (42).

2.3.4 Statistical analysis

Multivariate statistical analysis was performed with PRIMER v6 (PRIMER-E Ltd, Ivybridge, UK) software (53). For each OTU, reads were first converted to percent total, then log(x+1) transformed to reduce the scale differences between the most and the least abundant reads within a sample. Next, similarities between samples were calculated with Bray-Curtis resemblance matrix. Similarities and differences in OTU composition between regions (i.e. the Gulf, Gulf of Oman and the Red Sea), and between coral genera were tested using permutational multivariate analysis of variance (PERMANOVA) with three fixed factors (i.e.
‘region’, ‘reef’ (nested within region), and ‘coral genus’) using 999 permutations. To predict the potential environmental variables (including salinity, SST, POC, PIC and Chl-a) that drive the differences between OTU abundance in our dataset, canonical analysis of PCA (CAP) was performed using 999 permutations. PERMANOVA and CAP analyses were performed using the PERMANOVA+ add-on for PRIMER statistical software package (PRIMER-E Ltd, Plymouth, United Kingdom (54)).

2.4 Results

2.4.1 Pyrosequencing data

In total, specimens collected from three regions around the Arabian Peninsula represent corals from 43 genera: 35 of which were collected from the Red Sea, 33 from the Gulf of Oman, and 13 from the Gulf (Table 2.1, Figure 2.3, and Supplemental File 2.1). Porites, Acropora, Pocillopora, Favia, Montipora, and Platygyra were among the six most common coral genera collected across all three regions, together representing about half (51.50%) of the total number of specimens collected. Our pyrosequencing data produced a total of 2,688,970 raw reads, 15.46% of which were discarded in quality filtering steps resulting in 2,273,302 reads of high quality, which represented 30,968 distinct ITS2 variants from clades A (4,943), B (2), C (19,978), D (5,268), F (9) and G (2) (Table 2.1). The remaining 766 variants could not be assigned to any Symbiodinium types in the database and were excluded from further analysis. As represented in Table 2.1, the Red Sea had the highest number of ITS2 variants of 22,676 (14,807 variants in clade C (65.3%), 4,439 variants in clade A (19.6%), and 3,210 variants in clade D
(14.2%)), while the Gulf had the least number of variants of 2,991, most of which also belonged to clade C (2,275; 76.1%), and some in clade A (400; 13.4%). As for Gulf of Oman, out of the total 10,753 variants, 5,982 (55.6%) and 3,325 (30.9%) belonged to clade C and D, respectively, and 1,148 variants belonged to clade A (10.7%, Table 2.1).

2.4.2 Symbiodinium diversity around the Arabian Peninsula

Although there were over 30,000 Symbiodinium ITS2 variants from the pyrosequencing data, only a few variants represented more than 5% in each region as illustrated in Figure 2.4A. The Red Sea, for example, were dominated with only three types of ITS2 variants (type A1 (17.07%), C41 (15.68%) and C1 (7.47%)) with nearly two third of the variants (59.79%) below 5% abundant. Gulf of Oman, on the other hand, were mainly represented with six Symbiodinium types (C39 (12.49%), D1 (11.12%), C1 (10.11%), C15 (7.66%), A1 (5.99%), and D5 (5.65%)), with almost half the variants (46.98%) below 5% abundant. As for the Gulf, in addition to Symbiodinium type A1 that represented 6.79% in abundance, nearly half of the variants belong to Symbiodinium type C3 (48.68%), while 11.29% and 5.94% of the variants were almost identical to Symbiodinium types C38 and C3 (96.91% and 99.65% identity respectively via BLASTN) (Figure 2.4A).

2.4.3 Phylotype-based analysis of coral-associated Symbiodinium in the Arabian seas

Corals were typically associated with one to two dominant Symbiodinium types, which represented about 70-100% of total diversity within samples, with high number of rare ITS2 variants following a long tail distribution. The data showed
that dominant *Symbiodinium* communities were structured according to regions rather than being specific to any coral genera across geographic locations (Figure 2.5).

Scleractinian corals from each region differed in their association with dominant symbiont types. On average, *Symbiodinium* types C41 (~29-72% distribution) and C1 (~28-58% distribution) dominated hard corals from the Red Sea, with some coral genera being commonly associated with other symbionts types, for example, associations of *Stylophora* and *Pocillopora* with A1 (48% and 63% respectively), *Porites* with C15 (51%), and *Leptoria* with D1 (61%) (Figure 2.5B). Likewise, while most of the stony corals from the Gulf were associated with *Symbiodinium* type C3 (~58-70% distribution on average), some coral genera were also associated with A1 such as *Acanthastrea* (36%), *Cyphastrea* (10%), *Montipora* (21%) and *Platygyra* (9%). Interestingly, a *Goniastrea* specimen sampled from the Gulf was dominated with unclassified symbiont types (88%), rendering future investigation on *Goniastrea* from the Gulf and the classification of these unknown symbiont types. The distribution of symbionts in hard corals from Gulf of Oman on the other hand, were more variable. For instance, some corals were dominated with symbiont type C39 such as *Acanthastrea* (32%), *Echinophyllia* (41%), *Echinopora* (30%), *Favia* (32%), *Favites* (32%), *Hydnophora* (65%), *Leptoria* (69%), *Lobophyllia* (72%), and *Symphyllia* (53%), while other corals were commonly associated with type C1 i.e. *Astreopora* (74%), *Goniastrea* (28%), and *Psammocora* (47%). Also, symbiont type D1 dominated a few coral genera in Gulf of Oman including *Acanthastrea* (32%), *Cyphastrea* (39%), *Galaxea* (40%), and *Leptastrea* (38%). Finally, *Montipora* and *Acropora* from Gulf of Oman
were mainly associated with symbiont type A1 (40% and 64% respectively) (Figure 2.5A).

Similarly, soft corals were associated with different dominant symbiont types at different locations. For example, in the Red Sea, *Xenia* were dominated with symbiont types C3 (23.5%) and D3 (30.0%), while *Sarcophyton* and *Sinularia* were dominated with C65 (70.3% and 52.6% respectively), except for one *Sinularia* specimen (ALb_019) that was dominated with C1 (35%) and C1b (16%) (Figure 2.5B, and Supplemental File 2.1). *Xenia* from Gulf of Oman on the other hand, were only dominated with symbiont types D5 (52.7%) and D1a (33.0%), while *Sarcophyton* and *Sinularia* were commonly associated with symbiont types C3 (28.1% and 30.7% respectively), C3w (16.5% and 20.7% respectively), and C65 (7.2% and 9.5% respectively) (Figure 2.5A).

### 2.4.4 OTU-based analysis of *Symbiodinium* diversity around the Arabian Peninsula

The 30,968 distinct ITS2 copies clustered into 243 OTUs (A=7, B=1, C=220, D=11, F=3 and G=1) at a 0.03 distance cutoff (Table 2.1). Among 243 OTUs, 16 OTUs were shared in all regions, 53 OTUs were shared between Gulf of Oman and the Red Sea, 23 OTUs were shared between the Gulf and Red Sea, and only 17 OTUs were shared between Gulf of Oman and the Gulf, (Figure 2.6A). Some OTUs were unique to each region (Gulf of Oman=33 OTUs, the Gulf=14 OTUs, Red Sea=135 OTUs) (Figure 2.6A). As for the OTU compositions, samples from the Red Sea had the highest diversity of *Symbiodinium* (195 OTUs), most of which belonged to clade C (180 OTUs), while there were 38 *Symbiodinium* OTUs in the Gulf, and 87 OTUs in Gulf of Oman.
Of the 243 *Symbiodinium* OTUs, three prevalent OTUs i.e. OtuC00001, OtuA00001 and OtuD00001 cumulatively represent 96.58% of total OTUs, while the other 240 OTUs were rare, each representing <1% of total data. The most predominant OTU in our data was OtuC00001 (59.45%), representing 58.77% of the OTUs from the Red Sea, 52.06% of the OTUs from Gulf of Oman and 89.88% of the OTUs from the Gulf (Figure 2.4B). The Red Sea and Gulf of Oman were also dominated with OtuA00001 (26.89% and 10.51%, respectively), and OtuD00001 (10.31% and 32.20%, respectively) (Figure 2.4B). Apart from OtuC00001, corals in The Gulf were also associated with OtuA00001 (8.43%) (Figure 2.4B). The Gulf had the least diverse OTUs among the three regions, while the Red Sea, despite being the most diverse in symbiont diversity (195 OTUs), most OTUs (192 OTUs) were rare (<5% in abundance) (Figure 2.4B). Additionally, the most diverse clade in the Red Sea was clade C (180 OTUs), however, only 1 OTU (OtuC00001) were mainly represented, while the rest were <5% abundant. It is important to note that different clade C symbiont types that were associated with each region i.e. C41, C3, C1, and C15, were all represented as OtuC00001 because ITS2 sequences of these symbiont types differed only by one to three nucleotide substitutions from one another and could not be separated into distinct OTUs with a 0.03 cutoff (data not shown). Hence, although OTU-based analysis does simplify the complexity of the pyrosequencing data, it does so at a cost of being too conservative to separate certain ITS2 types into a separate OTU.

### 2.4.5 OTU-based *Symbiodinium* diversity within a host species

The majority (93.21%) of the coral hosts from all three regions harbored multiple symbiont OTUs per specimen, all of which had a long tail distribution where only
one or two symbiont OTUs dominated in the symbiont profiles. Intriguingly, 6.78% of the overall samples harbored one OTU exclusively (yet, each had their congeneric harboring >1 OTUs) (Supplemental File 2.2), however, none of the corals from the Gulf were exclusively associated to only a single OTU. Apart from having the most diverse types of *Symbiodinium*, coral genera from the Red Sea also hosted the highest diversity of symbionts within specimens compared to their counterparts in the Gulf and Gulf of Oman. On average, each coral from the Red Sea harbored 10.03 OTUs (Figure 2.6B). For example, *Pavona* sampled from Ayona reef in Yanbu, Red Sea (YBb_011) was associated with 20 OTUs, the maximum number of OTUs harbored by a single sample in our dataset (Supplemental File 2.2). On the other hand, corals in the Gulf of Oman and the Gulf harbored, on average, almost half the number of OTUs within a colony compared to the Red Sea (4.98 OTUs in Gulf of Oman, and 6.33 OTUs in the Gulf) (Figure 2.6B).

### 2.4.6 OTU-based *Symbiodinium* diversity comparisons between *Acropora*, *Porites*, and *Favia* around the Arabian Peninsula

To determine if coral samples in our data were significantly different according to their associations with symbionts, we applied multivariate statistical analysis on the three most common coral genera sampled across all three regions (*Acropora* spp. (n=97), *Porites* spp. (n=88), and *Favia* spp. (n=58), Figure 2.1). PERMANOVA test showed that OTU compositions in samples were significantly different between all factors tested (i.e. region, coral genus, reef (nested within region), and region cross with coral genus). Region acted as the most discriminating factor separating samples in our data (Pseudo-\(F = 33.045, P <0.001\)) (Table 2.2).
To visualize the dissimilarities of *Symbiodinium* abundance pattern between *Acropora* spp., *Porites* spp., and *Favia* spp. from Gulf of Oman, the Gulf and the Red Sea, and to predict the potential environmental variables (SST, Chl a, POC, PIC and salinity) that drive these differences, canonical analysis of PCA (CAP) was conducted (Figure 2.7). Supporting PERMANOVA test, samples were clearly clustered according to their respective regions. Interestingly, a couple of *Porites* specimens from Gulf of Oman (OMa_025 and OMd_014) were dominated with *Symbiodinium* OtuD00001 rather than OtuC00001, which were opposite to other *Porites* samples from this region that were dominated with OtuC00001, and hence, represented as outliers in the CAP ordination. As for environmental variables, salinity was positively correlated to the dissimilarities of *Symbiodinium* profiles in the Gulf (CAP1; \(r = 0.86\), CAP2; \(r = 0.12\)), while Chl-a and POC, were negatively correlated to the dissimilarities between *Symbiodinium* community in the Red Sea and Gulf of Oman respectively (Chl-a: CAP1; \(r = -0.415\), CAP2; \(r = -0.618\), and POC: CAP1; \(r = -0.251\), CAP2; \(r = -0.742\)). SST and PIC did not have high correlation to the differences of *Symbiodinium* diversity in the data (\(|r| < 0.25\)).

### 2.5 Discussion

This chapter provides a comprehensive survey on coral-*Symbiodinium* association in one of the most extreme but understudied geographical regions in the world.. The data represent *Symbiodinium* diversity that are associated to almost 800 corals specimens of 43 genera from the Red Sea, the Gulf, and Gulf of Oman. *Symbiodinium* diversity typing of numerous samples has proven useful with NGS, which is not only more convenient, but also more robust for identifying diversity
of symbiont within a host, from the dominant to the rare types. This chapter includes both phylotype- and OTU-based analysis to facilitate direct comparison to studies that applied traditional molecular methods as well as to provide thorough analyses of Symbiodinium-host association, and hence, represents one of the most comprehensive Symbiodinium diversity surveys to date.

2.5.1 Phylotype-based vs. OTU-based analysis

For the application of NGS to Symbiodinium ITS2 diversity studies, it is imperative that both phylotype- and OTU-based methods are used simultaneously. A phylotype-based analysis allows direct association of corals to the symbionts they host (usually the most dominant types), but at the same time relies on previously described symbiont types that are already present in the database. OTU-based framework on the other hand, is able to estimate symbionts diversity generated from NGS (dominant and rare types) into ecological entities or OTUs (without previous description of ITS2 types), which dramatically reduces the complexity of the data, but sometimes at a cost of not correctly separating different ITS2 types into distinct OTUs (42). Therefore, the best approach for NGS data processing is to use both analyses to overcome the limitation of each method, and to explore symbiont types that are the most common and the most rare.

As prices for sequencing technology are gradually decreasing, scientists are increasingly applying NGS for Symbiodinium ITS2 analysis. Nevertheless, the NGS-based Symbiodinium diversity analysis is still in its infancy, and needs to be standardized to facilitate data comparison between studies. If the discovery of single copy genes that are relatively fast evolving and that provide resolution to classify Symbiodinium into species level was identified, perhaps applying OTU-
based analysis would correctly resolve this complex issue and provide a more rapid development of *Symbiodinium* diversity studies in the future.

### 2.5.2 *Symbiodinium* distribution around the Arabian Peninsula

The Arabian Peninsula as a whole shares only a few types of *Symbiodinium* with the Caribbean and the Pacific. Marine invertebrates in the Caribbean are most commonly associated with *Symbiodinium* types B1, C1 and C3, while in the Pacific; clade C symbionts are the most dominant, specifically types C3, C1 and C21 in Southern Great Barrier Reefs (30, 38, 55). *Symbiodinium* diversity around the Arabian Peninsula is dominated with clade A1, various types of clade C (i.e. C41, C3, C1, and C39), and clade D. Our results support Hume et al. (56) study in showing a prevalence of *Symbiodinium* type C3 in the Gulf. Overall, Red Sea has the most diverse types of *Symbiodinium* OTUs present, which is also reflected in the association of coral specimens with the most diverse type of symbionts OTUs compared to other two regions. However, most of the *Symbiodinium* diversity in the Red Sea is present in low density with only three types (A1, C41 and D1) being dominant.

The analyses show that coral-symbiont association is strongly defined by geographic location rather than coral genera. The association of corals to distinct types of symbionts in different regions, despite being biogeographically connected via Strait of Hormuz and Gulf of Aden is very intriguing. This finding is in contrast to some studies that show stability of host-symbiont association over spatial scales. For instance, *Acropora palmata* is specifically dominated by *Symbiodinium* type A3 or *Symbiodinium ‘fitti’* throughout the Caribbean (37), the specific association of *Porites* spp. with *Symbiodinium* type C15 across the Pacific (56-59), and the prevalence of *Symbiodinium* type C1 and C3 in various corals
species across Indo-Pacific and Western Atlantic-Caribbean (60). The prevalence of certain types of *Symbiodinium* in particular sites i.e. C3 in the Gulf, and C41 in the Red Sea might indicate speciation events that separate *Symbiodinium* types in the Red Sea from Gulf of Oman and the Gulf, and allow co-evolution between the host and *Symbiodinium* over time (25).

Corals from the Gulf host the least diverse *Symbiodinium* types compared to Gulf of Oman and the Red Sea. This might be due to the extreme environmental conditions in the Gulf in terms of high temperature variations as well as high salinity that enhance only certain types of *Symbiodinium* to be successful. Nevertheless, it should be emphasized that *Symbiodinium* diversity shown in this chapter represents a snapshot of *Symbiodinium* prevalence at a specific time point, and that the diversity presented here may change if samples were collected in different seasons or at a different time point. A long term *Symbiodinium* diversity assessment i.e. spanning 5-10 years, flanking a coral bleaching event would unequivocally provide evidence if coral-*Symbiodinium* association are stable or if there are any symbionts shuffling within hosts.

### 2.5.3 High diversity of *Symbiodinium* association within coral hosts

By applying NGS for *Symbiodinium* genotyping, this chapter shows that corals are associated with quite a diverse type of *Symbiodinium* spp. at any point in time. However, *Symbiodinium* diversity within the corals are not equally distributed, but are dominated with only one to two types with the rest being rare in abundance (42). Therefore, previous reports that applied traditional molecular techniques on coral-*Symbiodinium* association might underestimate the actual symbiont population within a coral due to limited sensitivity. Hence, the notion that certain coral species harbor specific *Symbiodinium* types (25, 26, 28, 43) might hold true
for the dominant types of symbionts they harbor, and should be revisited with NGS-based analysis to explore their potential association with rare symbiont community.

The following are possible explanations to the differences in distribution of *Symbiodinium* within a host. 1.) Corals chose their underlying symbionts from the surrounding seawater, and depending on cumulative benefits the symbionts provide the hosts, and/or the environmental conditions the hosts are facing, corals retain some symbionts as dominant and some in the background. This might explain why a dominant symbiont in one sample can be rare in another (25). 2.) *Symbiodinium* choose to infect hosts with suitable niches for their optimal proliferation, and 3.) there is a competition for space between *Symbiodinium* within a host that results in a successful colonization of a certain symbiont to become dominant and others being in the background. To understand different effects between hosting a rare symbiont type and a dominant type of symbiont, an experiment where a coral is infected with its commensal symbiont type compared to its conspecific that is infected with rare symbiont type exclusively overtime would reveal the difference in benefits a coral receives from hosting different symbionts, and shed lights into why might one type be more successful in a particular host than another.

### 2.5.4 Thermally tolerant *Symbiodinium* types

Some of the *Symbiodinium* types around the Peninsula including A1, C3, and C1 are suggested to be thermally resilient, for example, *Symbiodinium* clade A is more thermally tolerant than clade B and C supposedly due to its capability to produce UV-protective mycosporine-like amino acids (MAAs) (61), and are mostly associated with corals in shallow reefs (25). Similarly, *Porites* sp. from
the Gulf harboring *Symbiodinium* type C3 deals better with heat stress than *Porites* sp. from Pacific Ocean (Fiji) that harbor C15, according to a study by Hume et al. (56). Also, different strains of *Symbiodinium* type C1 can exhibit different responses to thermal stress, with one strain being more thermally tolerant and prevent coral from bleaching, while other strains do not (personal communication with Rachel Levin). Hence, the prevalence of these thermally tolerant *Symbiodinium* types in this region is not unusual. However, there are many other *Symbiodinium* spp. that are also prevalent in the Arabian Peninsula, but their potential resilience at high temperature is at present unknown. For example, *Symbiodinium* type C41, C39, and other clade C types in the Red Sea, and Gulf of Oman, could potentially be endosymbiont types that can tolerate heat stress. Thermal stress coupled with physiological studies of these subclades might reveal their unique properties that allow them to be successful in this extreme marine condition.

Symbionts within clade D are also commonly considered to be resilient at high temperatures, and are regarded by some to be opportunistic types that infect compromised coral hosts during or post-bleaching incidents (20-24). Accordingly, clade D symbionts were expected to dominate the Red Sea and the Gulf due to their high temperature profiles and recent bleaching event in 2010 (62). Interestingly, symbionts in clade D are not the most dominant types in corals collected in this study. A few possible explanations for the relatively low occurrence of clade D compared to clade A and C in our data are: 1.) Symbionts within clade D might not always be thermally resistant. The resilience of corals to bleaching depends on the underlying symbionts (63), however, the resilience of corals may as well depend on the compatibility between symbionts and the host in
order to achieve optimal benefits to the holobiont. In fact, not all corals harboring symbiont clade D are thermally resistant. For example, while juveniles *Acropora millepora* harboring symbiont clade D are more thermally tolerant than when harboring symbiont type C1, juveniles *Acropora tenuis* harboring symbiont clade C1 are more tolerant towards high temperature than when harboring symbiont D type (64); 2.) It is possible that not all symbionts within clade D exhibit thermally tolerant properties and merely referring to clade D as thermally tolerant type may be over generalization, hence, more research is needed to identify specific types or species within clade D that exhibit tolerance towards high temperature; 3.) Symbionts in clade D may not only be an opportunistic symbiont and can co-evolve with a suitable host. For example, clade D symbionts are dominant in all *Pocillopora* spp. from Gulf of Oman despite dominance of other symbiont types in the region. This specificity might be unique only in Gulf of Oman since *Pocillopora* in the Red Sea harbor various types of symbionts in clade A and C. It is also possible that *Pocillopora* spp. from these two regions belonged to different species, which would further explain the specificity of host-*Symbiodinium* relationship; 4.) More expansive sampling throughout the Arabian Peninsula might result in a better diversity assemblage of *Symbiodinium* in this region, especially in the Gulf, where the data in this chapter are only based on corals along the coast off Abu Dhabi. Studies have shown association of corals in the Gulf to clade D symbionts, for instance, Baker (20) reported on diversity of *Symbiodinium* clade D at Jana and Karan islands off eastern Saudi Arabian coast, and Mostafavi, P. G., et al. (65) reported on the dominance of *Symbiodinium* clade D in Kish and Larak Islands on the Iranian side of the Gulf; and finally, 5.) The inaccuracy of copy numbers derived from PCR-based *Symbiodinium* typing
cannot be neglected. As reported in Arif, et al. (42), even though the diversity of symbionts is accurate via pyrosequencing, the relative abundance is not, due to bias in PCR amplification. To attest the *Symbiodinium* distribution within a sample, one would have to count *Symbiodinium* cells from the specimens and confirm their taxa using molecular markers. FISH-FLOW, a recently introduced technique by McIlroy, S. E., et al. (66) that combines fluorescent in situ hybridization (FISH) and flow cytometry to do genotype-specific labeling of *Symbiodinium* and to quantify the abundance of endosymbiont cells *in situ* could be applied to confirm the PCR-based *Symbiodinium* distribution within a sample.

### 2.5.5 Effects of salinity on coral-*Symbiodinium* association

The analysis in this chapter indicates strong correlation of salinity to the differences in OTU abundance patterns between geographic regions. While environmental variables such as temperature and light intensity on corals and *Symbiodinium* have been vastly studied, only a few studies focused on the effect of salinity to the coral holobiont (67-69). Chartrand, K. M., et al. (70) studied the effect of *Siderastrea radians* and its symbiont to hyposaline water conditions from river inflow into Florida Bay, and discovered that corals are tolerant to some extent to the reduced in salinity, but when salinity is below 10, *Symbiodinium* decreases in photosynthetic activities. Moreover, Wicks, L. C (71) suggests that coral distribution depends on the host’s and not only their symbionts’ tolerance towards environmental variables, including salinity.

The Gulf represents one of the most extreme environments for coral reefs to reside in the world. Exposure to variable temperatures and high salinity might be a limiting factor for diversity and distribution of both corals and *Symbiodinium* in this region (72). Although corals in the Gulf may have adapted to high
temperature and high salinity, further investigation on the effects of these environmental variables acting together on coral holobiont would shed some light on the upper limit of coral resilience. Moreover, comparison of physiological properties between the most abundant symbiont type in the Gulf (C3) and the most abundant symbiont type in the Red Sea (A1 and C41), might give insights into the differences in their response to the hypersaline environment and leads to an explanation to why C3 are more successful in this harsh environmental conditions than A1 or C41.

2.6 Conclusion

This chapter utilizes NGS to analyze *Symbiodinium* ITS2 diversity associated to over forty coral genera around the Arabian Peninsula, where coral reef ecosystems thrive in one of the most extreme marine environments on the planet. NGS allows deep sequencing of the symbiont community within a sample and is ideal for studying a large number of samples. Moreover, NGS enables detection of very low abundant ITS2 variants that would normally be missed by traditional methods. This chapter emphasizes the use of both phylotype- and OTU-based analyses for understanding *Symbiodinium* diversity and to avoid inaccurate estimations of *Symbiodinium* assemblages. This chapter provides evidence that corals around the Arabian Peninsula are associated with a diversity of *Symbiodinium* types that are strongly structured according to geographic location. The analysis also shows that most of the diversity of *Symbiodinium* communities within a coral are low in abundance and can be considered background level. Exploring environmental properties, e.g. the effects of salinity to symbionts association around the Peninsula might lead to the discovery of new heat and salinity tolerant symbiont types.
2.7 Tables And Figures

Figure 2.1 Sea surface temperature (SST) obtained from NASA Aqua Moderate Resolution Images (MODIS) in (A) winter (Jan-Mar) and (B) summer (Jul-Sept) of 2013 (http://gdata1.sci.gsfc.nasa.gov).
Figure 2.2 Maps showing sampling sites and number of coral specimens collected. Samples from the Red Sea were collected at Yanbu 23 (1), Yanbu Ayona (2), KAUST Inna Fasr (3), KAUST Al Fahal (4), KAUST Shib Nazar (5), Al Lith Shallow Reef (6), and Al Lith South Reef (7). Collection sites at the Gulf include Al Yassat (8), Ras Ghanada (9), and Saadiyat (10). Samples from Gulf of Oman were collected at Al Aqah (11), Dibba Rock (12), Damanyat islands (13-15), Fahal Island (16), Saifat Ash Shiekh (17), and Bandar Khayran Islet (18).
Figure 2.3 Cumulative Number of coral genera collected from the Gulf of Oman (blue), The Gulf (green) and the Red Sea (red). Total n=796.
Figure 2.4 Pie charts show percent distribution of *Symbiodinium* spp. in each region via phylotype-based analysis (A), and OTU-based analysis (B). Rare variants and rare OTUs indicate cumulative percentage of *Symbiodinium* types/OTUs that are <5% abundant in each region. C38* and C3* are variants that are 96.91% and 99.65% match to C38 and C3 via BLASTN, respectively.
**Figure 2.5** Phylotype-based *Symbiodinium* types associated with coral genera from A) Gulf of Oman, B) Red Sea and C) the Gulf that represent on average at least 5% in abundance (others represent *Symbiodinium* types with average percentage below 5%). ITS2 variants that do not have a 100% match to any *Symbiodinium* spp. in the database are reported as the closest *Symbiodinium* types in terms of percent identity obtained from BLASTN annotation.

**A) Gulf of Oman**

**B) Red Sea**
C) The Gulf

The Symbiodinium distribution on average per coral genera for the Gulf region is shown in the bar chart. The distribution varies across different coral genera, with some genera having a higher percentage of Symbiodinium than others. The chart includes a legend that identifies different Symbiodinium types with specific colors, allowing for a clear visual representation of the distribution patterns.
**Figure 2.6** A) A Venn diagram showing number of *Symbiodinium* OTUs found in Gulf of Oman, the Gulf, and the Red Sea as well as number of OTUs that are shared between them. B) A Box plot displaying average number of OTUs per sample at Gulf of Oman (*n*=439), The Gulf (*n*=101), and the Red Sea (*n*=256). Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend to minimum and maximum values.
**Figure 2.7** Canonical Analysis of PCA conducted from Bray-Curtis similarities of OTU abundance patterns of the three coral genera i.e. *Acropora* spp. (cross), *Porites* spp. (circle) and *Favia* spp. (triangle), collected from Gulf of Oman (blue), The Gulf (green) and The Red Sea (Red), and their degrees of correlation with five environmental variables, including 5-year annual average of chlorophyll a (Chl-a), particulate organic carbon (POC), particulate inorganic carbon (PIC), sea surface temperature (SST), and an annual average of salinity. The length and direction of vectors illustrate the strength and direction of correlation with each axis.
The table shows sampling sites, number of samples, and number of OTUs per site per clade. Pyrosequencing reads are indicated as well as numbers of ITS2 variants and coral genera collected from Gulf of Oman, the Gulf and the Red Sea.

Table 2.1
Table 2.2 PERMANOVA summary statistics of the differences in OTU abundance between regions (Gulf of Oman, Red Sea, and the Gulf), coral genus (*Porites* spp., *Acropora* spp., and *Favia* spp.), reefs (nested within regions), and regions crossed with coral genus, using 999 permutations.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>Pseudo-F</th>
<th>P-value</th>
<th>Permutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regions</td>
<td>2</td>
<td>52395</td>
<td>26198</td>
<td>33.045</td>
<td>0.001</td>
<td>999</td>
</tr>
<tr>
<td>Coral Genus</td>
<td>2</td>
<td>11857</td>
<td>5928.4</td>
<td>7.4779</td>
<td>0.001</td>
<td>999</td>
</tr>
<tr>
<td>Reefs (Regions)</td>
<td>15</td>
<td>37506</td>
<td>2500.4</td>
<td>3.1539</td>
<td>0.002</td>
<td>999</td>
</tr>
<tr>
<td>Regions x Coral Genus</td>
<td>4</td>
<td>28409</td>
<td>7102.2</td>
<td>8.9586</td>
<td>0.001</td>
<td>997</td>
</tr>
</tbody>
</table>

*Df = degrees of freedom; SS = sum of squares; MS = mean squares*
2.8 Supplementary Information

Supplemental File S2.1 The table lists all coral samples collected, their collection sites, depth, their associated Symbiodinium phylotypes, and environmental data. Data are archived online KAUST library.

Supplemental File S2.2 Overview over Symbiodinium ITS2 OTUs classification, read counts, and reference sequence for all OTUs identified Data are archived online KAUST library.

2.9 References


CHAPTER 3: Bacterial profiling of White Plague Disease (WPD) in a comparative coral species framework

Cornelia Roder¹, Chatchanit Arif¹, Till Bayer¹, Manuel Aranda¹, Camille Daniels¹, Ahmed Shibli¹, Suchana Chavanich², Christian R Voolstra¹

¹King Abdullah University of Science and Technology, Saudi Arabia
²Chulalongkorn University, Thailand
3.1 Abstract

Coral reefs are threatened throughout the world. A major factor contributing to their decline is outbreaks and propagation of coral diseases. Due to the complexity of coral-associated microbe communities, little is understood in terms of disease agents, hosts, and vectors. It is known that compromised health in corals is correlated with shifts in bacterial assemblages colonizing coral mucus and tissue. However, general disease patterns remain to a large extent ambiguous as comparative studies over species, regions, or diseases are scarce. Here, we compare bacterial assemblages of samples from healthy colonies and such displaying symptoms of White Plague Disease (WPD) of two different coral species (*Pavona duerdeni* and *Porites lutea*) from the same reef in Koh Tao, Thailand, using 16S rRNA gene microarrays. In line with other studies we found an increase of bacterial diversity in diseased corals, and a higher abundance of taxa from the families that include known coral pathogens (Alteromonadaceae, Rhodobacteraceae, Vibrionaceae). In our comparative framework analysis, we found differences in microbial assemblages between coral species and coral health states. Notably, patterns of bacterial community structures from healthy and diseased corals were maintained over species boundaries. Moreover, microbes that differentiated the two coral species did not overlap with microbes that were indicative of healthy and diseased corals. This suggests that while corals harbor distinct species-specific microbial assemblages, disease-specific bacterial abundance patterns exist that are maintained over coral species boundaries.
3.2 Introduction

One of the most recognized features of tropical, shallow-water corals is their symbiosis with photosynthetic unicellular algae (zooxanthellae) that provide photosynthetically fixed carbon to satisfy their host’s respiratory requirements (1) and facilitate calcification (2). Corals also live in association with numerous other microorganisms such as bacteria, archaea, protists, endolithic algae, fungi, and viruses (3), the significance of which is only partially understood (4, 5). The sum of all organisms is referred to as the coral holobiont (3).

It is now being recognized that bacteria contribute significantly to the biology of higher order organisms (6), and accordingly, bacteria associated with corals are considered a vital component of the coral holobiont. Their potential roles include nitrogen fixation (7), decomposition of organic materials (8), production of antibiotic compounds (9, 10), and occupation of space to prevent colonization by pathogens (11). Coral-associated bacteria have been shown to be host species-specific, diverse, and complex (12-14), and this assemblage comprises a unique signature that differs from bacterial communities in the surrounding water column (13, 15, 16).

Several studies have been conducted that highlight the role of bacteria in coral diseases (3-5, 17-22). Coral diseases appear as changes in tissue color in the form of patches or bands on the coral surface, associated with subsequent tissue damage, necrosis, and tissue loss (23). In many areas, disease outbreaks have led to massive die-offs of reef-building corals that resulted in habitat loss for reef-associated organisms, with propensity for irreversible ecosystem change (23-26). To date, the exact number of coral diseases remains unknown (27). Their characterization is mainly based on field observations of altered phenotypes. As a
result, the same disease might have been defined several times and in different ways depending on the species or the region affected (23). For most diseases, our knowledge on causative agents, modes of transmission, or disease reservoirs is missing (25). It is unknown whether the same pathogens cause similar/same disease characteristics in different coral hosts, or whether the same shifts in microbial assemblages result in the same disease-phenotype in different coral species (28, 29). Furthermore, it is not known whether diseases with a similar phenotype are caused by similar underlying mechanisms, i.e. if they are associated with comparable bacterial changes or species (28). Answers to these questions might not only enable a clearer disease nomenclature, but will result in a better understanding of the mechanisms driving coral disease outbreak and progression and will eventually lead to a better understanding of coral holobiont pathology (27, 30).

White Plague Disease (WPD) is one of the first described coral diseases (31). Records show that WPD was responsible for several virulent outbreaks, and it is held responsible for major reef declines worldwide, especially in the Caribbean (26, 27, 32-34). Corals affected by a WPD phenotype show a pronounced line of bright, white tissue that separates the colored (living) part of the coral from bare, rapidly algal-colonized skeleton (26). Three types of WPD, I (31), II (33), and III (26), have been described that differ in the rate of progression across a coral’s surface and affect different species (26, 35). Richardson, et al. (36) initially suggested a species of *Sphingomonas* as the causative pathogen, but Denner (17) proposed *Aurantimonas coralicida* as the final WPD-causing pathogen in corals from the Caribbean. Similarly, *Thalassomonas loyana* (37) has been proposed to be the causative agent of White Plague-like disease in the Red Sea. However,
neither of these bacteria could be unequivocally verified as the responsible pathogen in subsequent studies (18, 20, 21, 38). Consequently, it is debatable whether a definitive pathogen for WPD exists, or whether different pathogens or bacterial consortia produce a similar disease phenotype in different coral species. Given the inherent difficulties of assigning a pathogen to WPD, and thereby proofing a causal relationship, Willis et al (39) suggested that coral diseases from the Great Barrier Reef (and by extension the Indo-Pacific) that produce a phenotype of white bands of tissue and/or skeleton should be referenced collectively as White Syndrome (WS), unless the underlying disease etiology is known. Here we employed an alternative approach and tested whether healthy (HH) and diseased (DD) coral colonies displaying a WPD-characteristic phenotype (26, 31) from the Indo-Pacific share similarities in underlying microbial community patterns and are comparable to WPD-affected corals and studies from the Caribbean.

Sunagawa et al (20) was the first study that used 16S rRNA gene microarrays (PhyloChips™, Second Genome) to assess bacterial community changes in WPD in Montastraea faveolata and demonstrated the overall feasibility of the method. In this study, we used PhyloChips™ to profile microbial communities of healthy (HH) and diseased (DD) colonies of two coral species (Porites lutea and Pavona duerdeni) displaying signs of WPD collected from the same reef in Koh Tao, Thailand. Our aim was to examine microbial community differences within and between species, and between coral health states (HH vs. DD). Additionally, 16S rRNA gene clone library sequencing was conducted for all samples to compare the two different methods for assaying coral associated bacterial community structure.
3.3 Material And Methods

3.3.1 Sample collection
Sampling took place offshore of Sairee Beach (10.097908’ N, 99.825163’ E), Koh Tao Island, in the Gulf of Thailand during non-monsoon season in January 2011. Tissue was sampled from three healthy colonies and three colonies displaying signs of WPD between 4-7 m depth by SCUBA using hammer and chisel from two coral species, *Pavona duerdeni* and *Porites lutea*. Diseased colonies displayed an abrupt band of white, exposed coral skeleton that separated living tissue from algal-colonized dead coral skeleton. Samples from healthy colonies were chiseled off the uppermost part of the colonies, while samples displaying WPD symptoms were taken from the interface of healthy and diseased tissue. All samples were handled wearing gloves and directly transferred into sterile Whirl-Pak sampling bags. On board, corals were rinsed with filtered seawater (0.22 µm) and wrapped in aluminum foil. One liter of seawater was sampled from the water column above the reef and filtered (20 mm Hg) onto 0.22 µm Durapore PVDF filters (Millipore, Billerica, Massachusetts). All samples were immediately flash frozen in liquid nitrogen on board and stored at -80°C until subsequent DNA extraction.

3.3.2 DNA extraction
Coral samples were crushed to powder in liquid nitrogen using autoclaved mortars and pestles. Aliquots of 50-100 mg of coral powder and the disrupted filter holding the microbial community of the water column were utilized for DNA extraction using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). DNA concentrations were quantified on a NanoDrop 2000C spectrophotometer.
130

(Thermo Fisher Scientific, Waltham, MA) and with a Qubit fluorometer using the Quant-IT dsDNA Broad Range Assay Kit (Invitrogen, Carlsbad, CA).

3.3.3 PhyloChip™ PCR and hybridization

DNAs were shipped on dry ice to Second Genome Inc. (San Bruno, CA) for assaying on the PhyloChip™ G3 platform. Up to 500 ng of PCR product was applied to each PhyloChip™ G3 following previously described procedures (40). Briefly, the 16S rRNA amplicons and a mixture of amplicons at known concentrations (spike-mix) were combined, fragmented using DNAse1 (Invitrogen, Carlsbad, CA), and biotin-labeled using the recommended protocol for Affymetrix Prokaryotic Arrays. Labeled products were hybridized overnight at 48°C and 60 rpm. The arrays were washed, stained, and scanned as previously described (40).

3.3.4 PhyloChip™ data transformation and normalization

Details on probe selection, probe scoring, data acquisition, and preliminary data analysis are according to Hazen et al. (2010). Array fluorescence intensity was collected as integer values ranging from 1 to 65,536 (2⁰ to 2¹⁶). Subsequent log2 transformation yielded decimal numbers ranging from 0 to 16 that were multiplied by 1,000 yielding a range of 0 to 16,000 (HybScore). To correct for uneven hybridization, differences in hybridization intensities, and scale, intensity HybScores were loess-normalized using the normalize.loess function in the affy package (41) in the R statistical environment (R Core 42). A microbial taxon was regarded present if it was identified in two of three replicates of any species/condition combination (P. duerdeni HH, P. duerdeni DD, P. lutea HH, P. lutea DD), or determined present in the water sample based on the method in
Hazen, et al. (40). Of 59,222 bacterial operational taxonomic units (OTUs) assayed on the PhyloChip™, 29,103 were called present over all samples (Supplemental File S3.2, Supplemental File S3.3).

3.3.5 PhyloChip™ data analysis

To visualize similarities within and between species-condition combinations, a multidimensional scaling plot based on Bray-Curtis distances of OTU abundance data was generated using the libraries MASS and vegan in the R statistical environment (R Core 42). A corresponding 2-way crossed (species and condition) Analysis of Similarity (ANOSIM) on the basis of the same resemblance matrix (Bray Curtis distances of OTU abundances between samples) and using 9,999 permutations was conducted in the PRIMER v6 software (43). The degree of correspondence between the distances among points implied by multidimensional scaling (MDS) was measured by a stress function of the form \[ \sqrt{\sum \sum (f(x_{ij}) - d_{ij})^2 / \text{scale}} \]. In the equation, \( d_{ij} \) refers to the Bray Curtis distance between samples, \( f(x_{ij}) \) is some function of the input data, and \( \text{scale} \) refers to a constant scaling factor used to keep stress values between 0 and 1. The smaller the stress, the better the representation. Normalized HybScores were analyzed using the TM4 software (44). A two-way factorial ANOVA was conducted based on the 14,213 OTUs present in coral samples to determine differentially abundant OTUs between HH and DD samples and between species, as well as combinations thereof. Corresponding \( p \)-values were False Discovery Rate (FDR)-adjusted via R software package QVALUE (45) with a FDR cutoff rate of 10%. Hierarchical clustering was performed using Euclidean distance using the heatmap.2 function in the gplots package in the statistical environment R (R Core 42) on HybScores averaged over triplicates. Bacterial family overrepresentation was analyzed via
Chi-Square test with Yates’ Correction by comparing number of differentially abundant OTUs in a family in relation to all OTUs assayed for that family on the PhyloChip™. Only families that were represented by at least five taxa were considered.

3.3.6 Cloning and Sequencing

16S rRNA genes PCRs were run using coral DNAs according to the PhyloChip™ PCR protocol and primers (Hazen et al. 2010) to generate clone libraries. PCR products were cleaned with the MinElute PCR Purification Kit (Qiagen, Hilden, Germany). Clones for each sample were produced with a PCR Cloning Kit (Qiagen, Hilden, Germany) and picked into a 96 well plate, which contained a 25 µL mastermix that consisted of 1 x Multiplex Mix (Qiagen, Hilden, Germany), 0.2 µM of each M13F (-43: 5’-AGGGTTTTCCAGTCACGACGTT-3’) and M13R (-49: 5’-GAGCGGGATAACAATTTCACACAGG-3’) primer, and DNAse-RNase free water. M13 PCR conditions were 94°C for 15 min, 30 cycles of (94°C for 30 s, 55°C for 90 s, 72°C for 90 s), and a final extension of 72°C for 10 min. The 16S rRNA clones were sequenced bi-directionally with M13F (-21: 5’-TGTAAAACGACGGCCAGT-3’) and M13R (-29: 5’-CAGGAAACAGCTATGACC-3’) on an ABI 3730xl (Applied Biosystems by Life Technologies, CA) at the KAUST BioScience Core Facility. Sequence data have been submitted to the GenBank database under accession numbers KC527063 - KC527539.

3.3.7 Clone library analysis

16S rRNA gene sequences represented on the PhyloChip™ microarray were extracted from the Green Genes 2011 sequence dataset (46) resulting in 59,112 sequences, which were used to create a BLAST database. Clone sequences were
quality-trimmed, assembled, aligned, and checked for orientation in Codon Code Aligner (Codon Code Corporation, Centerville, MA) to obtain full-length 16S rRNA genes. 16S rRNA genes were queried with BLAST 2.2.26+ (BLASTN) (47) to assign a taxonomic level of classification to the clone sequence as described in DeSantis, et al. (48). Briefly, clone sequence and BLASTN hit were aligned to the Green Genes 16S rRNA genes alignment using NAST (49), and Lane mask (50) was applied using mothur (51). DNADIST (52) was used to calculate the sequence similarity between sequence pairs using the F84 model assuming a transition/transversion ratio of 2.0, and an A, C, G, T base frequency of 0.2509, 0.2276, 0.3156, 0.2057 respectively. The obtained similarity values were split into taxonomic groups according to the DNADIST percent similarity (Phylum (≥80%), Class (≥85%), Order (≥90%), Family (≥92%), Subfamily (≥94%), OTU (≥97%). In addition to the PhyloChip™ subset of 16S rRNA gene sequences, cloned 16S rRNA genes were also compared to the full Green Genes 2011 database.

3.4 RESULTS

3.4.1 PhyloChip™ and clone library comparison

To determine the amount of bacterial taxa that were not assayed on the PhyloChip™, we conducted a comparison of PhyloChip™ to clone library sequencing (Table 3.1). On the phylum level, all sequences identified by clone libraries were also detected by the PhyloChip™. Similarly, for all lower taxonomic ranks, the percentage of assigned 16S rRNA clones via Green Genes database and the PhyloChip™ was highly similar. It is worth noting that only
about 50% of all 16S rRNA genes could be annotated on the family level, and only about 40% of 16S rRNA genes on the OTU level, irrespective of the technique used. At the OTU level (i.e. ≥ 97% similarity), the PhyloChip™ missed only 15 clones that were successfully assigned to a 16S rRNA sequence via Green Genes database.

3.4.2 Patterns of bacterial richness and diversity in healthy and diseased corals

Of the 59,222 microbial OTUs assayed on the PhyloChip™ G3 microarray, 29,103 were deemed present in our samples (Table 3.2). Of these, 14,213 were present in corals and 18,418 OTUs were found in reef water. Diseased fragments had about one third more bacterial OTUs than their healthy counterparts, and *P. lutea* contained more than double the amount of bacterial OTUs than *P. duerdeni* irrespective of the health state (i.e. HH or DD).

To elucidate patterns of species and health state differences, we compared species-condition differences using ANOSIM (Table 3.3) and plotted the results in a MDS ordination (Figure 3.1). Samples significantly (p < 0.01) clustered according to coral species and condition (Table 3.3) as visualized by a partitioning of the samples along the two axes in the MDS ordination (Figure 3.1) indicating that microbial communities in corals separate according to species and disease. However, we found varying distances between replicates of species and conditions that emphasize that a high degree of natural variation between coral colonies seems to exist. The strength of difference (R) between microbial communities of the two coral species (*P. duerdeni* vs. *P. lutea*) and between the two health conditions (HH vs. DD) was equally significant and displayed a similar and high R value (R = 0.65 for species, R = 0.54 for condition) (Table 3.3). It is
important to note that the difference between health states is irrespective of the
coral species, and hence, a strong pattern of microbial community stratification in
healthy and diseased coral tissue exists that is consistent over coral species
boundaries.

3.4.3 Differentially abundant OTUs between species and disease states
A two-way ANOVA on differentially abundant OTUs between all four species-
condition combinations (P. duerdeni HH, P. duerdeni DD, P. lutea HH, P. lutea
DD) identified a total of 1,003 OTUs that were differentially abundant between
coral species, and 629 OTUs that were differentially abundant between healthy
and diseased samples (Table 3.3). The difference between coral species and
conditions was similar, although species differences were more pronounced. This
result corroborates the ANOSIM analysis where strong differences with species
differentiation were more pronounced than conditions. Notably, none of the OTUs
identified was significant in both comparisons (i.e. showed a species x condition
interaction) (Table 3.3, Supplemental File S3.4) indicating that OTUs that are
different between species are distinct from OTUs that are different between health
states.

The majority of OTUs that showed significant differences in abundance between
the two coral species were more abundant in P. duerdeni than in P. lutea
(Supplemental File S3.1A). This was true for healthy samples (655 vs. 348
OTUs), as well as for diseased samples (651 vs. 352 OTUs). We aggregated the
1,003 bacterial OTUs to the level of family. A Chi-Square analysis showed an
overrepresentation of bacteria belonging to the families Bacillaceae,
Comamonadaceae, Enterobacteriaceae, Lachnospiraceae, and Streptococcaceae
among the differentially abundant OTUs that separate coral species (df = 2, all p < 0.01, Table 3.4).

About two thirds of OTUs significantly different between HH and DD were more abundant in diseased specimens (P. duerdeni: 428 OTUs DD vs. 201 OTUs HH; P. lutea: 429 OTUs DD vs. 200 OTUs HH) (Supplemental File S3.1B). Comparison of HH and DD samples via congregated family fold-change differences of the 629 OTUs showed a higher abundance of bacteria belonging to the families Comamonadaceae, Enterobacteriaceae, and Streptococcaceae in HH samples (among others). In contrast, bacteria belonging to the families Colwelliaceae, Pseudomonadaceae, Rhizobiaceae, and Rhodobacteraceae were overrepresented and more abundant in diseased samples (Chi-Square, df = 2, all p < 0.01, Table 3.5). Changes in abundance were highest for bacteria belonging to the families Oceanospirillaceae, Rhodobacteraceae, and Vibrionaceae (all > 4-fold more abundant in diseased tissues for both coral species).

3.5 Discussion

Coral-associated microbes constitute an essential component in coral holobiont functioning (3). In particular, bacteria seem to play important roles in coral health and disease that still need to be further defined. One approach to identify common bacterial species is to conduct microbial studies in a comparative coral species framework. By choosing two species from the same coral reef, we limited variation in environmental variables in order to focus on the difference between coral species and coral health states. Here we characterized the abundance patterns of bacterial OTUs associated with healthy and diseased samples of P. duerdeni and P lutea in a standardized comparison via 16S rRNA gene
microarrays. The general feasibility of the PhyloChip™ platform to assess microbial community patterns in coral disease has been established by Sunagawa et al (20). In regard to taxonomic diversity and identification of OTUs from corals collected at Sairee Beach in Thailand, PhyloChip™ microarrays yielded comparable results to clone library-sequencing efforts. Both methods identified all OTUs to the phylum level and half of the OTUs to the family level, while about 60% of all the sequences failed to be annotated to the level of OTUs with either method. We found a higher number of OTUs in our study (between 2,756 OTUs in *P. duerdeni* HH and 10,848 OTUs in *P. lutea* DD) in comparison to sequence-based studies that looked at bacterial diversity in corals (e.g. Barott, *et al.* (53): between 163 and 461 OTUs per sample; Cardenas, Rodriguez, Pizarro, Cadavid and Arevalo-Ferro (21): between 256 and 378 OTUs per sample; Koren and Rosenberg (54): 400 OTUs; Lins-de-Barros, *et al.* (55): 354 OTUs). However, our estimates are well in line with estimates from Kellogg *et al.* (56) that identified between 1,112 and 9,240 OTUs with PhyloChips™ in a comparison of sampling methods for coral microbial community analysis.

Our data suggest that a lower bacterial diversity and abundance is associated with healthy corals, which has also been reported by Pantos, *et al.* (38), Sunagawa, *et al.* (20), and Cróquer, Bastidas, Elliott and Sweet (22). We identified *Pseudomonadaceae* and *Rhodobacteraceae* as prominent families promoted in colonies displaying WPD symptoms. *Rhodobacteraceae* have been proposed to be opportunistic due to uncontrolled propagation in disease by Sunagawa *et al.* (20). Furthermore, bacterial taxa of the family *Vibrionaceae* were more abundant in diseased samples as has been shown previously (20, 27, 57). Cardenas, Rodriguez, Pizarro, Cadavid and Arevalo-Ferro (21) conducted a study with a similar
experimental design and compared the microbiome of healthy and WPD-affected corals from two species (*Diploria strigosa* and *Siderastrea siderea*) in the Caribbean via 16S rRNA gene amplicon sequencing but the authors did not find consistent bacterial shifts over coral species. The use of pooled replicates by Cardenas et al (21) for the different conditions and species might have influenced the ability to statistically test for coral species or condition specificity. Alternatively, WPD-affected corals in the Caribbean might display a different pattern. We did not find *A. coralicida* (GenBank ID EF512716.1), the putative WPD pathogen from the Caribbean in any of the coral samples using clone libraries or the PhyloChip™ microarray. Also, *T. layana* (GenBank ID AY643537.2), a proposed causative agent of White Plague-like disease from the Red Sea was not identified during our cloning efforts and neither did we detect it on the microarray. This is consistent with results of other WPD-investigating studies that failed to discover either of these bacteria (20, 21, 38), which might be a result of investigating phenotypically similar, but not identical diseases (28, 39). It could also be argued that pathogens are subject to evolutionary change, which has been shown in other coral diseases (58). In this regard, the loss of pathogenicity due to changes in environmental conditions (19), repression through a newly, more favorably structured holobiont microbial assemblage (59), or control through bacteriophages (60) could be possible explanations.

When comparing healthy and diseased samples, there is a clear trend from bacterial communities low in diversity and abundance to mixed and variable assemblages with high numbers of unclassified bacteria, many of which were also identified in the surrounding water (data not shown). Most notably, we found no
overlap between OTUs differentially abundant between coral species and their health states.

### 3.6 Conclusion

Our data indicate that phenotypically similar coral diseases are accompanied by a common shift in bacterial community patterns in two different coral species collected from the same reef. At the same time, corals display species-specific bacterial communities that are different from disease-associated bacteria. Health and disease were as strong a discriminator between samples as species. One important consequence is that microbial community patterns (‘bacterial footprints’) might exist, which classify healthy and diseased coral specimens over species boundaries. In this regard, our study represents an approach to compare and analyze microbial assemblages of coral disease in a standardized framework (i.e. via PhyloChip™ profiles) that might aid in the classification and categorization of coral diseases. Future studies should incorporate measures over geographical distances in the same and different species in order to understand whether these patterns are only regionally or globally conserved.
3.7 Tables And Figures

Table 3.1 Number of distinct taxonomic ranks identified by PhyloChip™ in comparison to clone library sequencing of a pool of 96 clones from each sample (n = 477).

<table>
<thead>
<tr>
<th>Taxonomic rank (% cutoff)</th>
<th>Clones classified in Greengenes database (2011)</th>
<th>Clones detected by PhyloChip</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Phylum (≥80%)</td>
<td>474</td>
<td>99.37</td>
</tr>
<tr>
<td>Class (≥85%)</td>
<td>266</td>
<td>55.77</td>
</tr>
<tr>
<td>Order (≥90%)</td>
<td>254</td>
<td>53.25</td>
</tr>
<tr>
<td>Family (≥92%)</td>
<td>249</td>
<td>52.20</td>
</tr>
<tr>
<td>Subfamily (≥94%)</td>
<td>230</td>
<td>48.22</td>
</tr>
<tr>
<td>OTU (≥97%)</td>
<td>200</td>
<td>41.93</td>
</tr>
<tr>
<td>unclassified</td>
<td>3</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Abbreviation: OTU, operational taxonomic unit.

Table 3.2 Number of detected OTUs over all samples with PhyloChip™ microarrays. HH: healthy, DD: diseased.

<table>
<thead>
<tr>
<th>PhyloChip</th>
<th>No. of OTUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detected in coral and water</td>
<td>29103</td>
</tr>
<tr>
<td>Detected in coral</td>
<td>14213</td>
</tr>
<tr>
<td>in <em>Pavona duerdeni</em> HH</td>
<td>2756</td>
</tr>
<tr>
<td>in <em>Pavona duerdeni</em> DD</td>
<td>4434</td>
</tr>
<tr>
<td>in <em>Porites lutea</em> HH</td>
<td>7580</td>
</tr>
<tr>
<td>in <em>Porites lutea</em> DD</td>
<td>10848</td>
</tr>
<tr>
<td>Detected in water</td>
<td>18418</td>
</tr>
</tbody>
</table>

Abbreviations: DD, diseased; HH, healthy; OTU, operational taxonomic unit.
Table 3.3 Summary statistics of two-way crossed ANOSIM and two-way ANOVA.

<table>
<thead>
<tr>
<th>ANOSIM (based on Bray-Curtis distances)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differences between species (P. duerdeni vs P. lutea)</td>
</tr>
<tr>
<td>Strength of difference R: 0.65</td>
</tr>
<tr>
<td>Significance P&lt;0.01</td>
</tr>
<tr>
<td>Differences between conditions (HH vs DD)</td>
</tr>
<tr>
<td>Strength of difference R: 0.54</td>
</tr>
<tr>
<td>Significance P&lt;0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANOVA (14 213 OTUs, FDR &lt;0.1)</th>
<th>No. of OTUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species significant (P. duerdeni vs P. lutea)</td>
<td>1003</td>
</tr>
<tr>
<td>Condition significant (HH vs DD)</td>
<td>629</td>
</tr>
<tr>
<td>Interaction significant (species × condition)</td>
<td>0</td>
</tr>
</tbody>
</table>

Abbreviations: ANOSIM, analysis of similarity; ANOVA, analysis of variance; DD, diseased; FDR, false discovery rate; HH, healthy.
Table 3.4 Over- or under-representation of bacterial families of OTUs differentially abundant between coral species, and congregated fold-change differences between healthy and diseased specimens of *P. duerdeni* and *P. lutea* (only families that were represented by at least 5 bacterial taxa were considered).

<table>
<thead>
<tr>
<th>Bacterial family</th>
<th>OTU count</th>
<th>OTU count</th>
<th>Chi-square</th>
<th>P-value</th>
<th>Mean fold-change difference between healthy corals</th>
<th>Mean fold-change difference between diseased corals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>More abundant in <em>P. duerdeni</em></strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aquabacteriaceae</td>
<td>12</td>
<td>310</td>
<td>3.9220</td>
<td>&lt;0.05</td>
<td>1.92</td>
<td>1.87</td>
</tr>
<tr>
<td>Bacillaceae</td>
<td>38</td>
<td>264</td>
<td>16.9856</td>
<td>&lt;0.0001</td>
<td>2.04</td>
<td>1.69</td>
</tr>
<tr>
<td>Bacteroidaceae</td>
<td>5</td>
<td>37</td>
<td>1.1628</td>
<td>ns</td>
<td>1.23</td>
<td>1.13</td>
</tr>
<tr>
<td>Burkholderiaceae</td>
<td>7</td>
<td>146</td>
<td>0.6845</td>
<td>ns</td>
<td>1.45</td>
<td>1.82</td>
</tr>
<tr>
<td>Clostridaceae</td>
<td>13</td>
<td>176</td>
<td>0.0002</td>
<td>ns</td>
<td>2.08</td>
<td>1.52</td>
</tr>
<tr>
<td>Clostridales Family XI Incertae Sedis</td>
<td>6</td>
<td>101</td>
<td>0.0467</td>
<td>ns</td>
<td>2.25</td>
<td>1.99</td>
</tr>
<tr>
<td>Corynebacteriaceae</td>
<td>57</td>
<td>632</td>
<td>3.0345</td>
<td>ns</td>
<td>1.58</td>
<td>1.24</td>
</tr>
<tr>
<td>Flavobacteriaceae</td>
<td>27</td>
<td>629</td>
<td>6.4096</td>
<td>&lt;0.05</td>
<td>2.64</td>
<td>1.93</td>
</tr>
<tr>
<td>Lachnospiraceae</td>
<td>136</td>
<td>1508</td>
<td>9.3007</td>
<td>&lt;0.01</td>
<td>2.46</td>
<td>1.68</td>
</tr>
<tr>
<td>Lactobacillaceae</td>
<td>11</td>
<td>148</td>
<td>0.0000</td>
<td>ns</td>
<td>2.59</td>
<td>1.68</td>
</tr>
<tr>
<td>Pelagiibacteraceae</td>
<td>8</td>
<td>238</td>
<td>5.0707</td>
<td>&lt;0.05</td>
<td>1.58</td>
<td>1.83</td>
</tr>
<tr>
<td>Planococcaceae</td>
<td>5</td>
<td>32</td>
<td>1.8656</td>
<td>ns</td>
<td>2.62</td>
<td>2.31</td>
</tr>
<tr>
<td>Porphyromonadaceae</td>
<td>7</td>
<td>40</td>
<td>4.0121</td>
<td>&lt;0.05</td>
<td>1.3</td>
<td>1.68</td>
</tr>
<tr>
<td>Prevotellaceae</td>
<td>13</td>
<td>99</td>
<td>0.0010</td>
<td>ns</td>
<td>1.56</td>
<td>1.49</td>
</tr>
<tr>
<td>Pseudomonadaceae</td>
<td>24</td>
<td>797</td>
<td>18.3400</td>
<td>&lt;0.0001</td>
<td>1.09</td>
<td>1.24</td>
</tr>
<tr>
<td>Rhodobacteraceae</td>
<td>10</td>
<td>355</td>
<td>8.3811</td>
<td>&lt;0.01</td>
<td>2.64</td>
<td>1.59</td>
</tr>
<tr>
<td>Rhodospirillaceae</td>
<td>10</td>
<td>211</td>
<td>1.2334</td>
<td>ns</td>
<td>1.23</td>
<td>1.21</td>
</tr>
<tr>
<td>Rikenellaceae</td>
<td>7</td>
<td>46</td>
<td>2.7799</td>
<td>ns</td>
<td>1.91</td>
<td>1.71</td>
</tr>
<tr>
<td>Ruminococcaceae</td>
<td>57</td>
<td>616</td>
<td>3.7217</td>
<td>ns</td>
<td>2.79</td>
<td>2.13</td>
</tr>
<tr>
<td>Staphylococcaceae</td>
<td>14</td>
<td>323</td>
<td>2.9319</td>
<td>ns</td>
<td>1.67</td>
<td>1.41</td>
</tr>
<tr>
<td>Streptococcaceae</td>
<td>76</td>
<td>299</td>
<td>186.8956</td>
<td>&lt;0.0001</td>
<td>3.18</td>
<td>2.28</td>
</tr>
<tr>
<td>unclassified</td>
<td>38</td>
<td>618</td>
<td>0.5811</td>
<td>ns</td>
<td>2.13</td>
<td>1.89</td>
</tr>
<tr>
<td>Veillonellaceae</td>
<td>10</td>
<td>112</td>
<td>0.2855</td>
<td>ns</td>
<td>2.57</td>
<td>2.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bacterial family</th>
<th>OTU count</th>
<th>OTU count</th>
<th>Chi-square</th>
<th>P-value</th>
<th>Mean fold-change difference between healthy corals</th>
<th>Mean fold-change difference between diseased corals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>More abundant in <em>P. lutea</em></strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commonbacteriaceae</td>
<td>101</td>
<td>903</td>
<td>20.4645</td>
<td>&lt;0.0001</td>
<td>4.59</td>
<td>3.39</td>
</tr>
<tr>
<td>Denitrofibracae</td>
<td>6</td>
<td>88</td>
<td>0.0161</td>
<td>ns</td>
<td>1.28</td>
<td>1.09</td>
</tr>
<tr>
<td>Enterobacteriaceae</td>
<td>104</td>
<td>891</td>
<td>36.6887</td>
<td>&lt;0.0001</td>
<td>2.66</td>
<td>2.08</td>
</tr>
<tr>
<td>Moraxellaceae</td>
<td>5</td>
<td>163</td>
<td>3.0356</td>
<td>ns</td>
<td>2.16</td>
<td>1.97</td>
</tr>
<tr>
<td>Propionibacteriaceae</td>
<td>6</td>
<td>65</td>
<td>0.1546</td>
<td>ns</td>
<td>1.57</td>
<td>1.09</td>
</tr>
<tr>
<td>Rikenellaceae II</td>
<td>32</td>
<td>332</td>
<td>2.5767</td>
<td>ns</td>
<td>1.53</td>
<td>1.27</td>
</tr>
</tbody>
</table>

Abbreviations: ANOVA, analysis of variance; NS, not significant; OTU, operational taxonomic unit.
Table 3.5 Over- or under-representation of bacterial families of OTUs differentially abundant between health states of *P. duerdeni* and *P. lutea*, and congregated fold-change differences between healthy (HH) and diseased (DD) specimens (only families that were represented by at least 5 bacterial taxa were considered).

<table>
<thead>
<tr>
<th>Bacterial family</th>
<th>OTU count ANOVA (total 629)</th>
<th>OTU count PhyloChip (total 14 213)</th>
<th>Chi-square</th>
<th>P-value</th>
<th>Mean fold-change difference between HH vs DD <em>P. duerdeni</em></th>
<th>Mean fold-change difference between HH vs DD <em>P. lutea</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Aquabacteriaceae</em></td>
<td>12</td>
<td>134</td>
<td>0.1030</td>
<td>ns</td>
<td>1.86</td>
<td>2.60</td>
</tr>
<tr>
<td><em>Bacillaceae</em></td>
<td>9</td>
<td>211</td>
<td>0.3936</td>
<td>ns</td>
<td>1.15</td>
<td>1.21</td>
</tr>
<tr>
<td><em>Burkholderiaceae</em></td>
<td>7</td>
<td>332</td>
<td>0.0000</td>
<td>ns</td>
<td>1.16</td>
<td>2.19</td>
</tr>
<tr>
<td><em>Comamonadaceae</em></td>
<td>6</td>
<td>903</td>
<td>18.3690</td>
<td>&lt;0.0001</td>
<td>1.69</td>
<td>2.69</td>
</tr>
<tr>
<td><em>Enterobacteriaceae</em></td>
<td>20</td>
<td>801</td>
<td>6.4893</td>
<td>&lt;0.01</td>
<td>1.41</td>
<td>2.07</td>
</tr>
<tr>
<td><em>Mesorhizobiales</em></td>
<td>10</td>
<td>163</td>
<td>0.6779</td>
<td>ns</td>
<td>2.35</td>
<td>2.66</td>
</tr>
<tr>
<td><em>Streptococcaceae</em></td>
<td>23</td>
<td>209</td>
<td>17.3175</td>
<td>&lt;0.0001</td>
<td>2.36</td>
<td>1.91</td>
</tr>
<tr>
<td><em>Xanthomonadaceae</em></td>
<td>7</td>
<td>120</td>
<td>0.2089</td>
<td>ns</td>
<td>2.48</td>
<td>3.29</td>
</tr>
</tbody>
</table>

**More abundant in HH**

| *Alteromonadaceae*        | 5                            | 95                                | 0.0171     | ns      | 1.35                                                     | 1.61                                                 |
| *Clostridiaceae*          | 5                            | 176                               | 0.6491     | ns      | 1.78                                                     | 6.41                                                 |
| *Colwelliaceae*           | 6                            | 29                                | 18.3690    | <0.0001 | 3.94                                                     | 4.08                                                 |
| *Coronobacteriaceae*      | 17                           | 632                               | 3.9720     | <0.05   | 1.13                                                     | 2.17                                                 |
| *Flavobacteriaceae*       | 27                           | 629                               | 0.0033     | ns      | 3.18                                                     | 2.97                                                 |
| *Lachnospiraceae*         | 35                           | 1508                              | 15.9206    | <0.0001 | 1.22                                                     | 1.73                                                 |
| *Oceanospirillaceae*      | 9                            | 264                               | 23.8756    | <0.0001 | 5.35                                                     | 7.84                                                 |
| *Pulmonibacteriaceae*     | 8                            | 258                               | 0.7250     | ns      | 2.55                                                     | 2.33                                                 |
| *Pseudomonadaceae*        | 57                           | 797                               | 12.6293    | <0.0001 | 3.12                                                     | 4.66                                                 |
| *Rhizobiaceae*            | 11                           | 97                                | 8.0640     | <0.01   | 3.58                                                     | 1.93                                                 |
| *Rhodobacteraceae*        | 178                          | 355                               | 1150.8208  | <0.0001 | 5.28                                                     | 7.11                                                 |
| *Rhodospirillaceae*       | 9                            | 36                                | 0.0054     | ns      | 2.36                                                     | 2.79                                                 |
| *Rikenellaceae II*        | 7                            | 146                               | 3.3470     | ns      | 1.83                                                     | 1.55                                                 |
| *Ruminococcaceae*         | 11                           | 616                               | 0.3196     | <0.01   | 1.15                                                     | 1.01                                                 |
| *Sphingomonadaceae*       | 5                            | 142                               | 0.0093     | ns      | 1.75                                                     | 1.30                                                 |
| *unclassified*            | 25                           | 618                               | 0.1224     | ns      | 1.73                                                     | 1.48                                                 |
| *Vibroniaceae*            | 12                           | 310                               | 4.8116     | <0.05   | 5.46                                                     | 4.38                                                 |

Abbreviations: ANOVA, analysis of variance; NS, not significant; OTU, operational taxonomic unit.
Figure 3.1 Multidimensional (MD) scaling plot based on Bray-Curtis distances of normalized PhyloChip™ HybScores of healthy (circles) and diseased (triangles) specimens of the corals *P. duerdeni* (white) and *P. lutea* (black) illustrating the similarity of associated bacterial communities. Stress represents the goodness of fit of the data onto the MD ordination.
3.8 Supplementary Information

Supplemental File S3.1 Heatmap diagram displaying clustered normalized HybScores of significantly differentially abundant OTUs (A) between species, and (B) between conditions. Color key and histogram are displayed on top left corner of heatmaps. HH: healthy, DD: diseased, Pav: *Pavona duerdeni*, Por: *Porites lutea*.

Supplemental File S3.2 HybScores for all present OTUs (n = 29,103) of PhyloChip™ microarrays.

Data are archived online KAUST library.

Supplemental File S3.3 Presence/Absence of 29,103 OTUs over all samples on PhyloChip™ microarrays (0 = absent, 1 = present).

Data are archived online KAUST library.

Supplemental File S3.4 Two-way ANOVA on 14,213 OTUs that were detected present in healthy and diseased specimens of the corals *P. duerdeni* and *P. lutea*.

Data are archived online KAUST library.
3.9 References


42. R Development Core Team (2010) A Language and Environment for Statistical Computing.


Chapter 4: Bacterial Profiling of White Plague Disease across corals and oceans indicates a conserved and distinct disease microbiome

Cornelia Roder¹, Chatchanit Arif¹, Camille Daniels¹, Ernesto Weil²,
Christian R Voolstra¹

¹King Abdullah University of Science and Technology, Saudi Arabia
²University of Puerto Rico, Puerto Rico, United States
4.1 Abstract

Coral diseases are characterized by microbial community shifts in coral mucus and tissue, but causes and consequences of these changes are vaguely understood due to the complexity and dynamics of coral-associated bacteria. We used 16S rRNA gene microarrays to assay differences in bacterial assemblages of healthy and diseased colonies displaying White Plague Disease (WPD) signs from two closely related Caribbean coral species, *Orbicella faveolata* and *Orbicella franksi*. Analysis of differentially abundant operational taxonomic units (OTUs) revealed strong differences between healthy and diseased specimens, but not between coral species. A subsequent comparison to data from two Indo-Pacific coral species (*Pavona duerdeni* and *Porites lutea*) revealed distinct microbial community patterns associated with ocean basin, coral species, and health state. Coral species were clearly separated by site, but also the relatedness of the underlying bacterial community structures resembled the phylogenetic relationship of the coral hosts. In diseased samples, bacterial richness increased and putatively opportunistic bacteria were consistently more abundant highlighting the role of opportunistic conditions in structuring microbial community patterns during disease. Our comparative analysis shows that it is possible to derive conserved bacterial footprints of diseased coral holobionts that might help in identifying key bacterial species related to the underlying etiopathology. Furthermore, our data demonstrate that similar-appearing disease phenotypes produce microbial community patterns that are consistent over coral species and oceans, irrespective of the putative underlying pathogen. Consequently, profiling coral diseases by microbial community structure over multiple coral species might allow development of a
comparative disease framework that can inform on cause and relatedness of coral diseases.

4.2 Introduction

Corals are animals that live in a symbiotic relationship with photosynthetic dinoflagellates of the genus *Symbiodinium* as well as a rich bacterial community among other microorganisms that are collectively referred to as the coral holobiont (1). A coral’s associated microbial community contributes fundamentally to the holobiont’s functioning due to its role in coral nutrition (2) and host defense (3-5). Coral diseases are considered one of the most destructive local and geographic forces that impact corals, and are responsible for major reef ecosystem declines over the past decades (6-11).

Coral disease is defined as any abnormal condition affecting the coral holobiont (12), often described as a progressive loss of coral tissue due to either viral, fungal, protozoan, or bacterial infections (7, 13), and facilitated by environmental factors (e.g. high sea surface temperatures). It usually manifests through tissue discoloration and eventually tissue loss (necrosis). While the causative agents remain unknown for most diseases (14), it has been shown that compromised health in corals is accompanied by shifts in the microbial community associated with the coral holobiont (15-19). However, it is unclear whether infection of a single pathogen or opportunistic infections secondary to exposure to physiological stress triggers the re-structuring of microbial communities in coral disease (20). While this is mainly due to the complexity and dynamics of the host microbial assemblages (1, 21, 22), difficulty in conducting experiments underwater, an
overall lack of information on the structure and composition of the “natural” bacterial community of corals, and differences in applied methodologies further complicate the comparison of data. Sanger cloning-and-sequencing approaches are now being complemented by high throughput methodologies, and comparative analyses have shown that data from different platforms produce similar results (18, 23). However, unequal sample read representation and the use of different 16S amplicon sites hinder a direct comparison between studies. PhyloChip™ 16S rRNA gene microarrays provide a standardized platform, and have been successfully used to uncover microbial community patterns in coral disease (18, 19, 24).

White Plague Disease (WPD) is one of the most destructive and widespread coral diseases in Caribbean reefs (10, 25-28). It presents as a bright white band (i.e. clean skeletal structure resulting from disappearing tissue) that initiates at the base or sides of a colony, and separates the living tissue from recently settled turf algae on the exposed skeleton that quickly advances across the colony surface. Depending on the type of WPD (I, II, or III), progression rates vary and different coral species are affected (7, 29). *Aurantimonas coralicida* (30) and *Thalassomonas loyana* (31) were proposed as causative agents of WPD or WPD-like in corals from the Caribbean and the Red Sea, respectively. However, subsequent studies were unable to detect either of these two putative pathogens (15, 18, 19, 32, 33) suggesting that different pathogens must be able to produce highly similar disease phenotypes (6, 29, 34). In the Great Barrier Reef and Indo-Pacific region, phenotypes of WPD-like etiopathology have been denominated White Syndrome (WS) (11) and strains of the coral bleaching pathogen *Vibrio coralliilyticus* have been identified as potential infectious agents in a number of
coral species (35). Accordingly, indistinguishable disease phenotypes are produced by different pathogens (29), and disease nomenclature can be mistaken. For this reason, we refer to coral colonies displaying visual characteristics of White Syndrome, White Plague, or White Plague-like Disease as WPD, acknowledging that this neither includes nor excludes the presence of the pathogens *A. coralicida*, *T. loyana*, or *V. coralliilyticus*. On the other hand, how this convergent phenotypic resemblance relates to similarities in shifts in the underlying microbial community structure is at present unknown.

In this study, we analyzed microbial communities of healthy and WPD-affected coral tissues of *Orbicella faveolata* and *Orbicella franksi* (former genus *Montastraea*, Budd, Fukami, Smith and Knowlton (36)) from the Caribbean (Puerto Rico). Subsequently, we compared these data to microbial communities of two coral species, *Pavona duerdeni* and *Porites lutea*, displaying WPD characteristics (25, 37) from the Indo-Pacific (Gulf of Thailand) (19). We employed 16S rRNA gene microarrays (PhyloChips™) assaying 59,222 operational taxonomic units (OTUs) to profile microbial communities of healthy (HH) and diseased (DD) coral colonies in a standardized framework. We aimed to determine whether microbial community patterns of healthy and diseased colonies are not only consistent between species from the same site (19), but also over larger geographical distances, and how these patterns change between closely related and more distantly related coral species.
4.3 Material And Methods

4.3.1 Study site and sample collection
Sampling took place on September 5 and 6, 2011, at Weimberg reef (between N 17°53’17.40 / W 66°59’52.90 and N 17°53’25.40 / W 66°59’19.00) off the southwest coast of Puerto Rico. Two coral species (*O. faveolata* and *O. franksi*) were sampled via SCUBA between 16 to 22 m depth. From both species, tissue samples from three healthy colonies (displaying no visible signs of stress) and three colonies with WPD phenotype were collected. All corals were of similar size and all healthy samples were collected from the upper, non-shaded surface of the coral colony using hammer and chisel and were immediately transferred to sterile Whirl-pak bags. Samples of corals displaying signs of WPD were taken directly from the interface of healthy and diseased tissue. All samples were kept on ice during transportation.

4.3.2 Sample processing and data generation
Upon return to the laboratory, samples were rinsed with filtered seawater (0.22 µm) to remove loosely associated microbes. Rinsed samples were subsequently flash frozen in liquid nitrogen and ground to powder using mortar and pestle. Samples were then processed as described in Roder, *et al.* (19). Briefly, DNA was extracted from the coral powder using the DNeasy Plant kit (Qiagen, Hilden, Germany). After quantification of DNA using a NanoDrop 2000C spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and a Qubit fluorometer (Quant-IT dsDNA Broad Range Assay Kit, Invitrogen, Carlsbad, CA), DNA was shipped on dry ice to Second Genome Inc. (San Bruno, CA) for hybridization to the PhyloChip™ G3 platform as described in Hazen, *et al.* (38).
Up to 500 ng of PCR product was applied to each PhyloChip™ G3 following previously described procedures (38). Hybridized arrays were washed, stained, and scanned as previously described (38). Array fluorescence intensities (HybScores) (Supplemental File S4.1) were Loess-normalized using the normalize.loess function in the affy package (39) in the R statistical environment (40) to obtain abundance data for OTUs present in *O. faveolata* and *O. franksi* (n = 11,256 OTUs). A microbial taxon was regarded present if it was identified in two of three replicates of any species/condition combination (*O. faveolata* HH, *O. faveolata* DD, *O. franksi* HH, *O. franksi* DD) (Supplemental File S4.2). Comparisons of these data to PhyloChip™ results from healthy and WPD-affected samples of *P. lutea* and *P. duerdeni* collected in the Indo-Pacific (19) were conducted on Loess-normalized HybScores of shared OTUs (n = 7,200 OTUs). An OTU was considered present if it was detected in two of three replicates for any species/condition combination (i.e. *O. faveolata* HH, *O. faveolata* DD, *O. franksi* HH, *O. franksi* DD, *P. lutea* HH, *P. lutea* DD, *P. duerdeni* HH, *P. duerdeni* DD).

### 4.3.3 Data analysis

We calculated OTU richness for the total dataset as well as for any species/condition combination. To derive total richness of both datasets, i.e. from the Caribbean (this study) and the Indo-Pacific (19), the amount of all OTUs present were determined, and OTUs shared between both, unique to either dataset, or shared between conditions were compared.

Statistical evaluation of the Caribbean dataset included all bacterial OTUs present (n = 11,256 OTUs). First, differentially abundant OTUs between species (*O. faveolata* vs. *O. franksi*), condition (HH vs. DD), and their interactions were
determined based on normalized HybScores applying a two-way factorial analysis of variance (ANOVA) (Supplemental File S4.3) using the TM4 software (41). P-values were adjusted applying a 10% False Discovery Rate (FDR) using the QVALUE software package in R (42).

Relationships between samples from the two combined datasets (Caribbean and Indo-Pacific) were illustrated using multidimensional scaling (MDS) based on Bray-Curtis distances between samples using MASS and vegan library in R (40), in which the stress value depicts the accuracy of the ordination. Effects of the three factors ‘site’ (Caribbean vs. Indo-Pacific), ‘species’ (P. duerdeni vs. P. lutea vs. O. faveolata vs. O. franksi), and ‘condition’ (HH vs. DD) were calculated using the permutational multivariate analysis of variance (PERMANOVA) add-on in PRIMER-E (43). In the PERMANOVA design, we nested the factor ‘species’ within the factor ‘site’, as different species pairs were sampled at the two study sites. 999 permutations (only 998 unique for ‘condition’, only 997 unique for interaction term ‘site’ × ‘condition’, and only 997 unique for interaction term ‘condition’ × ‘species(site)’) of residuals under a reduced model were conducted (44, 45). The resulting Pseudo-F and associated p-values can be interpreted as equal to results of univariate ANOVAs, however, are based on the multivariate Bray-Curtis distance measures (43).

Differentially abundant OTUs for the combined data from the Caribbean and Indo-Pacific were determined via two-way ANOVA, as described for the Caribbean samples above. We classified those OTUs that were more than 2-fold differentially abundant in the same direction (HH or DD) in all four species (P. duerdeni, P. lutea, O. faveolata, O. franksi) as footprint bacterial species in WPD. These respective microbial key players were additionally sorted according to their
respective families. Finally, we compared phylogenetic relationships of the coral species to differences in bacterial assemblages (n = 7,200 OTUs) in HH samples. Phylogenetic relationships between the four coral species (*O. faveolata*, *O. franksi*, *P. lutea*, and *Pavona duerdeni*) (36, 46, 47) were compared to dendrograms based on similarities in bacterial assemblages. Dendrograms were constructed by averaging normalized OTU HybScores over samples for all HH species and subsequent application of Euclidean distance clustering (average linkage) with 1,000 bootstraps using the TM4 software (41). Trees were visualized using TreeView (48).

4.4 Results

4.4.1 Bacterial richness in healthy and diseased corals

Of the 59,222 microbial OTUs assayed on the PhyloChip™ G3 microarray, 11,256 OTUs were present in the coral samples collected from the Caribbean. OTU numbers were similar for both species (*O. faveolata* and *O. franksi*) with more than a 50% increase in OTU richness in diseased corals compared to healthy specimens (Table 4.1). The numbers of detected OTUs were similar to results from a previous study in the Indo-Pacific, where we assayed healthy and WPD-affected colonies in *P. duerdeni* and *P. lutea* (19). In both surveys, we observed increased richness in DD samples compared to HH. Combining OTU richness from both studies, we found a total of 18,269 distinct OTUs over the four different coral species. Of these, around 40% (7,200 OTUs) were shared between coral species from both regions, while the remaining OTUs were distributed unevenly between corals from the Caribbean (4,056 non-shared OTUs) and the Indo-Pacific.
Of all OTUs present in either study, more than three times as many OTUs (7,122) were found in diseased samples in comparison to OTUs from healthy samples (2,335). *A. coralicida* or *T. loyana*, the two proposed causative agents of WPD, were not detected in any sample from both datasets analyzed with the PhyloChip™ platform. Further, strains of *V. coralliilyticus*, identified as proposed WS pathogens by Sussman, Willis, Victor and Bourne (35), were not represented on the PhyloChip™ platform, and accordingly, not assayed.

### 4.4.2 Differentially abundant OTUs in healthy and diseased corals from the Caribbean

A two-way ANOVA was conducted to test for significant differences in OTU abundances between conditions (HH vs. DD), species (*O. faveolata* vs. *O. franksi*), and combinations thereof (Table 4.2). While no bacterial taxon differed significantly in abundance between the two coral species, 2,411 OTUs were significantly different between HH and DD samples. Log2 fold-changes in bacterial abundance ranged from -5.03 (OTU 72172, genus Fusobacterium: ~32-fold higher in HH) to +3.10 (OTU 61563, family Rhodobacteraceae: ~8-fold higher in DD) in *O. faveolata* (Supplemental File S4.3). In *O. franksi*, log2 bacterial abundance changes ranged from -6.29 (OTU 76854, genus Spirochaeta: >70-fold higher in HH) to +5.08 (OTU 51567, phylum Acidobacteria, class PAUC37f: ~30-fold higher in DD) (Supplemental File S4.3). Average log2 fold changes were higher in *O. franksi* (log2FC = 2.42 ± 0.41) in comparison to *O. faveolata* (log2FC = 1.21 ± 0.70).

Approximately half of the OTUs significantly different between HH and DD samples were more abundant in HH compared to DD for both *Orcibella* species, about 40% were more abundant in DD, and less than 10% of the bacterial OTUs
either increased or decreased when assessing condition-specific OTUs per coral species (Supplemental File S4.3). Only a single bacterial taxon, *Jannaschia* sp., was found to differ significantly in abundance in a species-and-condition-type interaction confirming that bacterial abundance patterns in healthy and diseased corals seem to be conserved and distinct from coral species-specific bacteria for the majority of OTUs analyzed, as found by Roder, *et al.* (19).

### 4.4.3 Comparison of bacterial patterns in WPD from the Caribbean and the Indo-Pacific

We combined data from this study with data from a previous study in the Indo-Pacific (19) to compare bacterial abundance patterns in healthy and diseased corals across oceans (i.e. the Caribbean and the Indo-Pacific) and coral species (i.e. *O. franksi, O. faveolata, P. duerdeni, P. lutea*). For this analysis, we only considered OTUs present in both datasets (n = 7,200). Relationships based on OTU diversity of the samples were displayed in an MDS ordination (Figure 4.1) and the relative contributions of ‘site’, ‘species’, and ‘condition’ were tested via PERMANOVA (Table 4.3). While all factors contributed highly significant to the structure of the data (*P* < 0.001), the denominator ‘site’ was the most pronounced factor as indicated by the highest Pseudo-F value (Pseudo-F = 10.40). The contribution of the factor ‘site’ was more than twice as high as that of ‘condition’ (Pseudo-F = 4.90) or ‘species’ (Pseudo-F = 2.82). Since the combined datasets did not comprise the same coral species in both regions, the factor ‘species’ was nested within the factor ‘site’ and therefore yielded a lower Pseudo-F, even though the sum of squares was higher than for the factor ‘condition’ (20.72 vs. 18.02). Furthermore, interaction terms of ‘site’ and ‘condition’ were significant, while there was no significant interaction between the factors ‘condition’ and
species’. This pattern recaptures what we saw when analyzing diseased corals from the Indo-Pacific: a strong separation by condition and a slightly stronger separation by species, but no interaction between species and condition (19). PERMANOVA results were corroborated by the MDS plot where the largest separation was between the two sites (i.e. HH samples of either ocean were distributed towards the edges of the plane), while all DD samples were arranged amid the HH samples. This indicates that bacterial communities associated with different corals, irrespective of species or site, are more similar when affected by WPD than when corals are healthy (Figure 4.1).

In order to identify key bacterial species and community shifts (i.e. bacterial footprints) within the microbiomes of healthy and diseased corals, we conducted a two-way ANOVA and determined all those OTUs that were more than two-fold different between HH and DD in all four species and assorted them to bacterial families (Figure 4.2). For those taxa > 2-fold more abundant in HH corals, only 2 bacterial families (Lachnospiraceae, Prevotellaceae) were represented by more than 1 OTU, whereas 5 bacterial families (Desulfomicrobiaceae, Lactobacillaceae, Rikenellaceae, Streptococcaceae, Xanthomonadaceae) were each represented by 1 OTU. Except for Desulfomicrobium orale (OTU 57563), none of the OTUs could be assigned to a described bacterial species. Those bacterial taxa with more > 2-fold higher abundances in DD corals were comprised of 4 OTUs from 4 families (Alteromonadaceae, Flavobacteriaceae, Phyllobacteriaceae, Rhodospirillaceae) and 12 OTUs belonging to the family Rhodobacteraceae. Accordingly, the family Rhodobacteraceae contributed 75% of the key OTUs in DD samples, and also represented the highest fold changes. The only OTU assigned to the species level from this bacterial family was
Rhodobacter sphaeroides (OTU 61632), while all other bacterial taxa were taxonomically unclassified at the species level.

4.4.4 Phylogenetic position of the coral host and bacterial community structure

While the main factor structuring all eight species-condition combinations appeared to be region, bacterial community profiles of DD samples were closely related to each other irrespective of species or site (Figure 4.1). To further understand the contribution of phylogenetic positioning of a coral species to its microbial abundance pattern, HH samples were clustered based on microbial diversity and compared to the phylogeny between the four coral species (Figure 4.3). Samples from healthy colonies recaptured phylogenetic positioning of the coral species. Furthermore, similarity between corals decreased with increasing phylogenetic distance, i.e. healthy specimens of the two closely related Orbicella species harbored a more similar microbial abundance pattern than the more distantly related P. duerdeni and P. lutea. The greatest difference between HH bacterial abundance patterns of all four species was between complex and robust corals that are evolutionarily separated by > 200 million years (46). However, species pairs from a given site in our study (i.e. O. franksi and O. faveolata from the Caribbean, P. duerdeni and P. lutea from the Indo-Pacific) coincided with affiliation to the robust (O. franksi and O. faveolata) and complex (P. duerdeni and P. lutea) clades, so we were not able to disentangle whether the differences arose from phylogenetic or site affiliation.
4.5 Discussion

4.5.1 Bacterial richness in healthy and diseased corals

The numbers of OTUs detected in samples from the Caribbean were comparable to those detected in other studies applying PhyloChips™ in corals (18, 19, 24), but higher than those in sequencing-based surveys (e.g. 15). Richness and diversity of bacterial taxa associated with two *Orbicella* species were similar within the same condition (HH and DD) as has been shown for microbial communities of closely related sponge species (49). We also found higher numbers of bacterial taxa associated with diseased samples, which concurs with previous studies on coral disease (17-19, 32). This is suggested to be a consequence of bacterial colonization from the surrounding environment on compromised and vulnerable coral tissues (18-20), and supported in this work by a larger number of shared OTUs, and an increase in richness of unclassified bacteria in DD samples. It is worthwhile noting that bacterial diversity has been reported to be similar and overlapping across different oceans (50). Accordingly, colonization of bacteria from the surrounding water column would result in similar microbial profiles of diseased tissue even between geographically distinct regions.

4.5.2 Towards elucidation of bacterial disease footprints and microbiomes

A great advantage of the PhyloChip™ microarray is that it represents a standardized platform in which data over different studies can be easily integrated, comprehensively analyzed, and compared in a common framework. While our recent analysis of two coral species from the same reef in the Indo-Pacific indicated that microbial community patterns of health and disease are conserved
over coral species boundaries (19), we were not able to test whether these patterns hold true over geographical distances (i.e. regionally and globally) in coral species. The experimental design of this study matches and complements that of Roder, et al. (19), enabling us to integrate and compare healthy and diseased states of four coral species (P. duerdeni, P. lutea, O. faveolata, O. franksi) comprising three genera (Pavona, Porites, Orbicella) from two distinct regions (Caribbean and Indo-Pacific). Most interestingly, almost half of all the bacterial taxa found in either dataset were shared (n = 7,200 OTUs), even though we analyzed different coral species and different regions at different depths. We found that region is the strongest separating factor between the two datasets. However, this measure took inadvertently the closer phylogenetic relationship between O. faveolata and O. franksi into account, so that we were not able to differentiate between similarity due to common reef of origin, or similarity due to phylogenetic relatedness. Profiling bacterial diversity in healthy and diseased tissues from the same coral species over geographical distances will unequivocally resolve the relative contribution of phylogenetic similarity vs. health condition (e.g. data from this study and Kellogg, et al. (51)).

Between closely related coral species (i.e. O. faveolata and O. franksi) microbial abundance patterns were undistinguishable (as indicated by no significant different OTUs between species in the two-way ANOVA) and support the notion that there is a close relationship between a coral host and its bacterial assemblage (52). Indeed, grouping of corals based on bacterial community structure of HH specimens (Figure 4.3) indicates that microbial assemblage resembles phylogenetic position of the coral host. Besides, diseased corals were more similar to each other than to healthy ones as indicated by the MDS plot (Figure 4.1). Last,
we detected no bacteria in the dataset from the Indo-Pacific (19) and only one OTU in the dataset from the Caribbean (this study) that were significantly different between species and at the same time significantly different between conditions. These data substantiate the previous notion that shared disease-specific microbial community patterns exist that is distinct from non-shared species-specific bacterial assemblages (19).

Bacterial taxa that were consistently found in WPD-affected coral species analyzed here comprised only few families and support the hypothesis of a bacterial abundance pattern in WPD that is structured by opportunists. For instance, Pelagibacteraceae are abundant in ocean surface bacterioplankton communities (53), and Rhodospirillaceae have been reported to become abundant in heterotrophic environments (54). Most notable, OTUs from the family Rhodobacteraceae made up 75% of all OTUs more than 2-fold enriched in DD, and fold-changes were more drastic than for any other bacterial family. Since none of the members of the Rhodobacteracea family are known to be potentially pathogenic, we hypothesize that the reason for their colonization success must either lie in the specific characteristics of this family (e.g. competence for rapid proliferation) or be due to prevailing environmental conditions (e.g. high abundance in environment). Many members of the Rhodobacteraceae are phototrophs, and their ability to accumulate in response to availability of organic substrate (54) might present optimal conditions for effective growth in or on compromised coral tissues. Accumulation of opportunistic bacteria has been suggested to result in phenotypically similar mortality patterns (diseases) that might not necessarily have the same cause (20). WS, WPD, or WPD-like infections have been reported for a variety of coral species and regions (7, 11, 13,
37, 51), but only three bacterial species have been proposed as causative agents (30, 31, 35). Also, rates of the disease’s progression have been shown to be different (25, 26, 55). Conversely, while it is likely that distinct pathogens might cause WPD in different coral species, similarity in phenotypes and etiopathologies might be driven by opportunistic bacteria that are conserved over species boundaries and that structure microbial abundance patterns.

4.6 Conclusion

Taken together, our data indicate that 1) disease-specific microbial abundance patterns exist, which are 2) conserved across coral species and oceans, and that are 3) largely different from species-specific abundance patterns. While the increase in bacterial diversity from WPD specimens over ocean boundaries render the hypothesis of colonization of opportunistic bacteria a likely scenario, the trigger of opportunistic colonization remains unknown and might well be of varying origin and importance. Conserved microbial community patterns provide the opportunity to derive bacterial families and species whose relative abundance can serve as indicators for health, stress, and disease. The elucidation of commonalities in coral diseases beyond species and sites would allow the establishment of ‘bacterial footprints’, i.e. shared community profiles for a given disease that can inform on the state of the coral holobiont irrespective of the coral species and/or region under study. Furthermore, such bacterial footprints would allow placing different coral diseases into a comparative framework and specifically test evolutionary hypotheses in regard to cause, origin and relatedness of different coral diseases.
4.7 Figures and Tables

Table 4.1 OTU richness is shown for all four species/condition combinations from the Caribbean (\textit{O. faveolata} HH, \textit{O. faveolata} DD, \textit{O. franksi} HH, \textit{O. franksi} DD) and from the Indo-Pacific (\textit{P. duerdeni} HH, \textit{P. duerdeni} DD, \textit{P. lutea} HH, \textit{P. lutea} DD). Bacterial taxa were counted present when detected in 2 out of 3 replicates of any species/condition combination.

<table>
<thead>
<tr>
<th>PhyloChip</th>
<th># Bacterial</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>This study - Puerto Rico, Caribbean</strong></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>11,256</td>
</tr>
<tr>
<td>in \textit{Orbicella faveolata} HH</td>
<td>4,336</td>
</tr>
<tr>
<td>in \textit{Orbicella faveolata} DD</td>
<td>6,791</td>
</tr>
<tr>
<td>in \textit{Orbicella franksi} HH</td>
<td>4,538</td>
</tr>
<tr>
<td>in \textit{Orbicella franksi} DD</td>
<td>7,448</td>
</tr>
<tr>
<td><strong>Roder et al. (2013) - Gulf of Thailand, Indo-Pacific</strong></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>14,213</td>
</tr>
<tr>
<td>in \textit{Pavona duerdeni} HH</td>
<td>2,756</td>
</tr>
<tr>
<td>in \textit{Pavona duerdeni} DD</td>
<td>4,434</td>
</tr>
<tr>
<td>in \textit{Porites lutea} HH</td>
<td>7,580</td>
</tr>
<tr>
<td>in \textit{Porites lutea} DD</td>
<td>10,848</td>
</tr>
<tr>
<td><strong>Both studies</strong></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>18,269</td>
</tr>
<tr>
<td>Unique Puerto Rico (Caribbean)</td>
<td>4,056</td>
</tr>
<tr>
<td>Unique Gulf of Thailand (Indo-Pacific)</td>
<td>7,013</td>
</tr>
<tr>
<td>Shared (Puerto Rico and Gulf of Thailand)</td>
<td>7,200</td>
</tr>
<tr>
<td>Associated with HH (Caribbean and Indo-Pacific)</td>
<td>2,335</td>
</tr>
<tr>
<td>Associated with DD (Caribbean and Indo-Pacific)</td>
<td>7,122</td>
</tr>
</tbody>
</table>
Table 4.2 Summary statistics of two-way ANOVA separating species and condition effects between *O. faveolata* and *O. franksi*. Data based on normalized HybScores of 11,256 OTUs present over all coral samples.

<table>
<thead>
<tr>
<th>Two-Way ANOVA (11,256 OTUs, FDR &lt; 0.1)</th>
<th># Bacterial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species significant (<em>Orbicella franksi</em> vs. <em>Orbicella faveolata</em>)</td>
<td>0</td>
</tr>
<tr>
<td>Condition significant (HH vs. DD)</td>
<td>2,411</td>
</tr>
<tr>
<td>Interaction significant: species x condition</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.3 PERMANOVA summary statistics for factors affecting bacterial communities in healthy and WPD-affected corals from the Caribbean (i.e. Puerto Rico) and the Indo-Pacific (i.e. Gulf of Thailand). df = degrees of freedom. SS = Sum of Squares. MS = Mean Squares. n.s. = not significant.

<table>
<thead>
<tr>
<th>df</th>
<th>site</th>
<th>condition</th>
<th>species (site)</th>
<th>site × condition</th>
<th>condition × species(site)</th>
<th>residuals</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>SS</td>
<td>38.28</td>
<td>18.02</td>
<td>20.72</td>
<td>11.33</td>
<td>10.28</td>
<td>58.87</td>
<td>157.50</td>
</tr>
<tr>
<td>MS</td>
<td>38.28</td>
<td>18.02</td>
<td>10.36</td>
<td>11.33</td>
<td>5.14</td>
<td>3.68</td>
<td></td>
</tr>
<tr>
<td>Pseudo-F</td>
<td>10.40</td>
<td>4.90</td>
<td>2.82</td>
<td>3.08</td>
<td>1.40</td>
<td>1.40</td>
<td></td>
</tr>
<tr>
<td>Pperm</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>n.s.</td>
<td>n.s.</td>
<td></td>
</tr>
<tr>
<td>permutations</td>
<td>999</td>
<td>998</td>
<td>999</td>
<td>997</td>
<td>997</td>
<td>997</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.1 Multidimensional scaling (MDS) plot derived from Bray-Curtis distances of normalized PhyloChip™ HybScores (n = 7,200 OTUs). Healthy (open) and WPD-affected (filled) specimens of the corals *P. duerderni* (square), *P. lutea* (circle), *O. faveolata* (triangle), and *O. franksi* (diamond) are shown. Stress represents the goodness of fit of the data onto the MD ordination.
Figure 4.2 Displayed are bacterial families and OTUs that showed a > 2-fold abundance difference between HH and DD over samples from *P. duerdeni* (white), *P. lutea* (light gray), *O. faveolata* (black), and *O. franksi* (dark gray).
Figure 4.3 Relationship between healthy coral species in regard to differences in their bacterial community based on Euclidean distance of bacterial abundance (i.e. HybScores) from 7,200 shared OTUs (1,000 bootstraps). The dendrogram recaptures the phylogenetic relationship of the corals *O. franksi*, *O. faveolata*, *P. duerdeni*, and *P. lutea*. 
4.8 Supplementary Information

Supplemental File S4.1 HybScores for all present OTUs \( n = 11,256 \) of PhyloChip™ microarrays from *Orbicella faveolata* HH, DD and *Orbicella franksi* HH, DD.

Data are archived online KAUST library.

Supplemental File S4.2 Presence/Absence calls of 11,256 OTUs over all samples on PhyloChip™ microarrays \( (0 = \text{absent}, 1 = \text{present}) \).

Data are archived online KAUST library.

Supplemental File S4.3 Two-way ANOVA on 11,256 OTUs that were detected present in healthy and diseased specimens of the corals *Orbicella faveolata* and *Orbicella franksi*.

Data are archived online KAUST library.

4.9 References


47. Fukami H, et al. (2008) Mitochondrial and nuclear genes suggest that stony corals are monophyletic but most families of stony corals are not (Order Scleractinia, Class Anthozoa, Phylum Cnidaria). *PLoS One* 3(9):e3222.


OVERALL CONCLUSION

The coral is critically dependent on its associated symbiont community. Disruption of symbiotic associations leads to deterioration of coral health and the integrity of the coral holobiont as a whole. This dissertation focused on the importance of two symbiont members of the coral holobiont, namely *Symbiodinium* and bacteria. The thesis provides detailed insights into symbiont diversity and distribution in over widespread geographic locations around the globe.

This dissertation implemented four different techniques for diversity analyses, and emphasized the importance of developing standardized techniques in order to facilitate comparisons between studies. Chapters one and three utilized two techniques in parallel i.e. traditional techniques versus more recently developed techniques to show advantages and disadvantages of both. While traditional techniques such as bacterial clone libraries and DGGE could correctly identify the diversity of abundant bacteria, more advanced techniques such as 16S rRNA gene microarrays (Phylochips), and Next-Generation Sequencing (NGS) were proven to deliver greater resolution, and hence, greater insight into the underlying ecology (e.g. in regard to previously underexplored low(er) abundant community members). On the other hand, the utilization of ever-new techniques prior to standardization would impede the advancement of research as a common framework for direct comparison between studies. For example, even though many studies are considering NGS-based analysis for *Symbiodinium* diversity, the differences in data processing and derivation of OTUs further complicate the data interpretation and lead to ambiguity in defining *Symbiodinium* spp. or OTUs. Because technologies will keep advancing, and the application of new techniques
in ecology is inevitable, it is important to ground-truth new techniques and lay down a standardized framework that should be agreed upon among the science community before being widely implemented. This would not only enhance direct comparisons between studies, but will also accelerate the advancement in ecological research.

Although there are comprehensive surveys on diversity of *Symbidinium* associated with marine invertebrates around the world, most studies only focused on the dominating *Symbiodonim* spp., while considerably fewer studies attended to the diversity of rare *Symbiodinium* types. The main reason is the limit in resolution of molecular methods being applied. Moreover, utilizing ITS2 region as a marker for *Symbiodinium* diversity typing further complicates data interpretation due to the difficulty in discerning intragenomic from intergenomic variations. Hence, defining *Symbiodinium* species is not straightforward as defining bacterial species through 16S rRNA genes, and needs multiple markers in complementation. This dissertation contributes greatly to the field by introducing an NGS-based pipeline for *Symbiodinium* ITS2 diversity typing. Through NGS, we confirmed that: (1) ITS2 variants within a genome of a *Symbiodinium* species are arranged in disequilibrium, whereby a single most dominant variant exists that most likely represents the ITS2 type of the *Symbiodinium* species; (2) Genetic distances between ITS2 variants within a genome of a *Symbiodinium* species are small and can be grouped into an OTU with a distinct distance cutoff (≥97%), and (3), Applying this similarity cutoff to *Symbiodinium* spp. associated to coral samples indicates that corals harbor a diverse community of *Symbiodinium* spp., whereby only one or two symbiont types are (co-)dominant, but accompanied by a substantially high number of background symbionts. The survey on *Symbiodinium*
spp. diversity in the Arabian Peninsula via NGS confirmed the complex diversity of *Symbiodinium* types within a coral, and warrants further investigation of the rare symbionts and their potential roles to the holobiont. Nevertheless, we have to acknowledge the urgent need for a better marker to define *Symbiodinium* into a distinct species, as it is not possible to unambiguously disentangle ITS2 diversity into ecological discrete species. Future research should invest into discovering new species markers besides the sole use of ITS2 as a marker for *Symbiodinium* diversity typing.

One of the most important factors that limit the knowledge of microbial diversity associated with corals is the lack of baseline information regarding microbial assemblage in coral holobionts. With comparisons between corals that appeared healthy as opposed to those that appeared stressed, scientists have developed indicators for various ‘diseases’. Some pathogens for the diseases have also been suggested, though many could not be subsequently verified. This dissertation focused on bacteria associated with multiple coral species over multiple geographic locations, and reinforced the notion that microbial diversity in corals is (1) distinct from seawater, (2) distinct between coral species, and (3) distinct in corals from different geographic locations, although certain bacterial species in healthy corals are conserved over geographical boundaries. Further, the dissertation provides evidence that these attributes do not hold true for corals that display symptoms of WPD. In fact, the analyses reported here indicate that common assemblages of bacteria in phenotypically similar coral diseases exist, and that these are conserved over coral species and geographic boundaries. For example, the increase in abundance of *Rhodobacteraceae* family in all coral species showing signs of WPD argues for conserved ‘bacterial footprints’ that can
be used to put diseases into a phylogenetic and evolutionary context to each other. If ‘footprints’ are conserved over coral species displaying similar disease symptoms, it would strongly indicate that the shift in microbial community in diseased corals is due to secondary reactions of the stressed holobiont, and that a single pathogen that causes an infection might or might not exist. Further investigation that considers the association of other members in the holobiont to the disease formation, such as viruses, archaea, and fungi, should be considered.