12-10-2020

A Comparative Genomics Exploration of Inter-partner Metabolic Signaling in the Coral-algal Symbiosis

Katherine E. Dougan
Florida International University, kdoug023@fiu.edu

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

A COMPARATIVE GENOMICS EXPLORATION OF INTER-PARTNER METABOLIC SIGNALING IN THE CORAL-ALGAL SYMBIOSIS

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOLOGY

by

Katherine E. Dougan

2021

To: Dean Michael R. Heithaus
College of Arts, Sciences and Education

This dissertation, written by Katherine E. Dougan, and entitled A Comparative Genomics Exploration of Inter-Partner Metabolic Signaling in the Coral-Algal Symbiosis, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

_____
Joana Figueiredo

_____
Fernando Noriega

_____
Laurie Richardson

_____
Wensong Wu

_____
Mauricio Rodriguez-Lanetty, Major Professor

Date of Defense: December 10, 2020

The dissertation of Katherine E. Dougan is approved.

_____
Dean Michael R. Heithaus
College of Arts, Sciences and Education

_____
Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2021

DEDICATION

I dedicate this dissertation to my parents, Milagros and Kevin Dougan, who have always encouraged me to pursue my dreams in science and told me I could achieve anything I put my mind to. I would not have accomplished this without them and am forever indebted to them for all of their unconditional support, guidance, and love.

ACKNOWLEDGMENTS

First and foremost, thank you to my advisor Dr. Rodriguez-Lanetty for all of your support and guidance during my Ph.D. Thank you for giving me the freedom to take a chance on my PhD research project and for helping me focus and develop it into what it has now become. With your supervision, I have experienced incredible academic and personal growth and I am extremely grateful for all that I have learned from you. Thank you as well to my Ph.D. committee members - Dr. Joana Figueiredo, Dr. Fernando Noriega, Dr. Laurie Richardson, and Dr. Wensong Wu - for all of your guidance over the years. Thank you to Florida International University for a teaching assistantship, Dissertation Year Fellowship and Tropics Summer Fellowship. Thank you to the members of the Rodriguez-Lanetty Lab, past and present, for all of your help and support. I would especially like to thank Dr. Anthony Bellantuono. You were not only irreplaceable to my research and Ph.D., but you have also become a trusted friend and mentor. Thank you to all of my amazing friends–there are no words to describe what your friendship through the good and hard times has meant to me. Thank you to my wonderful family for all of your love, support, and encouragement throughout this journey. Thank you, Mom and Dad, for everything you have done and continue to do–there is just simply too much to put it into words here. Thank you to my brother, Thomas, for making sure my head never got too big. And last but certainly not least, thank you, Aunt Laurie for my first introduction to a lab all that time ago and the steady supply of advice that followed.

ABSTRACT OF THE DISSERTATION

A COMPARATIVE GENOMICS EXPLORATION OF INTER-PARTNER METABOLIC SIGNALING IN THE CORAL-ALGAL SYMBIOSIS

by

Katherine E. Dougan

Florida International University, 2021

Miami, Florida

Professor Mauricio Rodriguez-Lanetty, Major Professor

At the foundation of coral reef ecosystems is the symbiosis between the coral host and its microbial community, particularly its photoautotrophic algae from the family Symbiodiniaceae. As a symbiosis centered around nutritional exchange, determining the mechanisms involved in the maintenance of this cooperative exchange is central to understanding how it breaks down. As the nutritional transfer primarily consists of sugars, this work first focuses on the cnidarian insulin signaling pathway, an evolutionarily important metazoan pathway involved in diverse functions, most notably metabolism. This dissertation unveiled 360 putative cnidarian insulin-like peptides (cnILPs) from existing transcriptomic datasets, where they were previously missed due to the bioinformatic methods employed. Significantly, symbiotic corals and anemones possessed the greatest diversity in insulin-like peptides compared to other cnidarian taxa. Conserved transcriptional responses of the cnILPs were also detected, particularly cnILP-B down-regulation in response to symbiosis along with a non-specific cnILP up-regulation in response to thermal stress. These trends coincide well with known transcriptional responses of ILPs in diverse organisms ranging from the nematode C.

elegans to humans, implicating for the first time that insulin signaling similarly functions in symbiosis and stress response in non-bilaterians. This dissertation also focused on the genome of the thermotolerant Durusdinium trenchii, which is well-known to confer thermotolerance on diverse coral species. We identified considerable duplication of gene blocks, more than 10-100x that of other Symbiodinaceae species, in support of previous hypotheses regarding a near or whole genome duplication event. Importantly, within these duplicated gene regions we detected extensive positive selection on genes central to the maintenance and repair of chloroplast structures like thylakoid membranes and photosystem II, a primary site of damage during photoinhibition. Widespread genome duplication and adaptive selection on photosynthetic functions is significant as it aligns with previous physiological studies identifying this as a factor in the thermotolerance of *D. trenchii*. This dissertation for the first time not only substantiates previous hypotheses of genome duplication in *D. trenchii*, but connects this duplication to the acquisition of thermotolerance in *D. trenchii*. Altogether, this dissertation highlights the importance for further investigations into the functions of the insulin signaling pathway in coral-algal symbioses and stress response, as well as confirms genomic duplication and selection as contributing to the evolutionary acquisition of thermotolerance in the symbiont *D. trenchii.*

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

CHAPTER-TABLE PAGE

LIST OF FIGURES

CHAPTER-FIGURE PAGE

Chapter 1: Introduction

1.1 The insulin-like signaling pathway

1.1.1 The invertebrate insulin-like signaling pathway

The insulin-like signaling pathway (ISP) is the central pathway responsible for responding to changing glucose levels in metazoans (Chang, Chiang, & Saltiel, 2004). The ISP is an evolutionarily important metabolic pathway that is integral to the regulation of development, metabolism, growth, and longevity in most, if not all, metazoans (Barbieri, Bonafè, Franceschi, & Paolisso, 2003; Claeys et al., 2002; De Meyts, 2004). Genetic work employing mutational and knockout studies of the insulin receptor have suggested that the downstream signaling pathways and functions mediated by the insulin receptor are conserved among nematodes, molluscs, insects, and mammals (González, Farnés, Vasconcelos, & Pérez, 2009). Studies on the hydrozoan *Hydra vulgaris* built upon these findings, showing that the activity of this pathway is likely conserved in non-bilaterians from the phylum Cnidaria as well (Steele, Lieu, Mai, Shenk, & Sarras Jr, 1996).

Outside of vertebrates, insulin-like proteins are generally referred to as insulin-related peptides (IRPs) or insulin-like peptides (ILPs). In invertebrates, these ILPs have primarily been studied in model organisms such as the nematode *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster*, the mosquito *Anopheles stephensi*, or the silkmoth *Bombyx mori (Jin Chan & Steiner, 2000)*. First discovered in *B. mori* in 1984 and referred to as bombyxin (Nagasawa, Kataoka, Hori, et al., 1984; Nagasawa, Kataoka, Isogai, et al., 1984), the ILPs were shown to possess the same A- and B- chain structure

with six conserved cysteine residues that bridge the two chains via disulfide bonds that are found in canonical insulin. Although the ILP structure is conserved in invertebrates, the overall number can far exceed that of invertebrates as evidenced by the over 30 ILPs in *B. mori* (Aslam, Kiya, Mita, & Iwami, 2011), 40 ILPs in *C. elegans (Zheng et al., 2018),* and 8 ILPs in *D. melanogaster* (Grönke, Clarke, Broughton, Andrews, & Partridge, 2010). On the contrary, the insulin family in vertebrates is made up of 10 known peptides including insulin, insulin-like growth factors, and relaxins/ILPs (Shabanpoor, Separovic, & Wade, 2009).

1.1.2 Insulin-like signaling in symbiosis and stress

The ISP has been implicated in the inter-partner communication for several organisms, although it has not been extensively studied and the mechanisms are not fully understood. In the mosquito *A. stephensi*, the translocation of soluble products from the human malaria parasite *Plasmodium falciparum* have been shown to induce the synthesis of two *A. stephensi* ILPs (Marquez et al., 2011). It is believed that by activating the mosquito's ISP, *P. falciparum* can suppress its host's immune system via alterations in metabolism (Marquez et al., 2011; Pietri et al., 2016; Pietri, Pietri, Potts, Riehle, & Luckhart, 2015). In the fruit fly *D. melanogaster*, however, the increased ISP activity resulting from *Wolbachia* infection results in extended lifespan and rescues DILP mutants (Grönke et al., 2010; Ikeya, Broughton, Alic, Grandison, & Partridge, 2009). While examples of inter-partner signaling via ISP have not been extensively studied elsewhere, transcriptomic studies have implicated them in the endosymbiosis between the salamander *Ambystoma maculatum* and the green algal *Oophila amblystomatis* (Burns,

Zhang, Hill, Kim, & Kerney, 2017) and even in the coral-algal symbiosis (Yuyama, Ishikawa, Nozawa, Yoshida, & Ikeo, 2018).

Compared to its role in symbiotic communication, the response and function of the ISP under stress has been well documented, particularly within model organisms. The ISP is perhaps most well-known in this area for its role in regulating the onset of the dauer life stage in *C. elegans* (Riddle, Blumenthal, Meyer, & Priess, 1997). The ISP normally initiates the dauer stage in response to unfavorable conditions such as limited food availability or temperatures stress, but it can also be triggered via knockout of the ISP regardless of environmental conditions (Riddle et al., 1997). Dauer formation results in an extended lifespan as well as increased resistance to stress and age-related diseases via signaling with downstream ISP effectors such as FoxO and PTEN. Reductions in ISP activity, outside of the dauer life stage, in general confer increased fitness and resilience under stress including oxidative stress (Larsen, 1993), ultraviolet stress (Wolff et al., 2006), and heat stress (Gems et al., 1998). Similarly to *C. elegans*, gene knockouts in the ISP of the fruit fly *D. melanogaster* have also been connected to the diapause life stage and increased longevity and stress resistance (Piper, Selman, McElwee, & Partridge, 2008).

Outside of increased longevity from dietary restriction, the ISP in invertebrates has been shown to modulate growth and reproduction in response to temperature changes as well as pathogens. For example, temperature changes and their effects on ISP signaling have been linked to alterations in reproduction in snakes (Sparkman, Byars, Ford, & Bronikowski, 2010) and growth in trout (Gabillard et al., 2003). Similarly, the ISP has

also been shown to mediate energetic trade-offs between growth and reproduction in *D. melanogaster* in response to stressors such as temperature (Burks et al., 2000). Reductions in growth and reproduction via ISP signaling is also considered to be a trade-off for increased immunity, an essential response during period of pathogenic challenge (Schwenke, Lazzaro, & Wolfner, 2016). Indeed, flies in the diapause stage or under dietary restriction are more resistant to infections and have higher expression levels of important immune regulatory genes (Kubrak, Kučerová, Theopold, & Nässel, 2014; Libert, Chao, Zwiener, & Pletcher, 2008; Schwenke et al., 2016).

1.2 The coral-algal symbiosis

1.2.1 Nutritional foundation for the symbiosis

Corals are ecosystem engineers for the "rainforest of the sea," or coral reefs. Valued at $9.9 trillion USD per year in goods and services (Costanza et al., 1997), coral reef ecosystems are important to biodiversity, recreation, tourism, fishing, coastal protection, and jobs despite comprising only 1% of the Earth's surface . At the foundation of coral reef ecosystems are the symbiotic interactions between the coral host and its microbial community, which includes viruses, bacteria, archaea, algae, and fungi (Rosenberg, Koren, Reshef, Efrony, & Zilber-Rosenberg, 2007).

Of particular importance within this holobiont is the photoautotrophic algae from the family Symbiodiniaceae (Order: Suessiales), which persist within a host-derived vacuole termed the symbiosome and provide the coral animal with sufficient energy to survive (Wakefiel & Kempf, 2001). This nutritional recycling is believed to be responsible for the evolutionary success of coral-algal symbioses in the oligotrophic environments where

coral reefs are usually found (Stat, Carter, & Hoegh-Guldberg, 2006). They also engage in symbioses with diverse organism, such as their ectosymbiotic relationships with Cardiid clams where they are located in an extension of the stomach as extracellular symbionts (Leggat et al., 2002). Originally classified as the genus *Symbiodinium* that contained nine clades from A-I, they were recently revised to recognize this group as the family Symbiodiniaceae (LaJeunesse et al., 2018). Currently, this family contains nine distinct genera with the aforementioned clades now described as the genera *Symbiodinium*, *Breviolum*, *Cladocopium*, *Durusdinium*, *Effrenium*, *Fugacium*, and *Gerakladium*.

Many studies and reviews have targeted the coral-algal symbiosis on a cellular level to gain a mechanistic awareness of how and why the coral-algal symbiosis is disrupted in response to stress (Weis, 2008). Research has targeted a diverse array of mechanisms involved in maintaining this symbiosis including immunity, cell cycle regulation, calcification, and nutritional transfer (Antonelli, Rutz, Sammarco, & Strychar, 2016; Davy, Allemand, & Weis, 2012; Weis, 2008). As nutritional transfer between the two partners is central to the evolutionary success of this symbiosis, it is no surprise that has been the most widely studied facet of the coral-algal symbiosis (Davy et al., 2012; Kirk & Weis, 2016; Weis & Allemand, 2009). The metabolic regulation of nutritional transfer between the two symbiotic partners becomes even more important under environmental stress that requires an energetically-costly stress response.

The translocation of photosynthetically-fixed carbon by the coral symbionts to the coral animal is central to the ecological success of the coral-algal symbiosis. First

demonstrated in the anemone *A. elegantissima* (Muscatine & Hand, 1958), metabolic transfer has become the most well-studied aspect of the coral-algal symbiosis that provides up to 99% of the coral's daily carbon requirements (Muscatine, Falkowski, Porter, & Dubinsky, 1984). Several hypotheses have been developed to explain the regulatory controls of the photosynthetic transfer and how it might break down. One common explanation is through the limitation of nutrients, particularly nitrogen and phosphorus. Many studies have shown that increased environmental loads of nitrogen and phosphorus can cause increased concentrations of the algal symbiont in coral tissues along with decreased transfer of photosynthates (Ferrier-Pages, Gattuso, Dallot, & Jaubert, 2000; Szmant, 2002). While some studies have instead documented increases in growth following elevated nutrient levels (Muller-Parker, McCloskey, Hoegh-Guldberg, & McAuley, 1994), decreased photosynthetic transfer and subsequent coral growth is more common (Ferrier-Pages et al., 2000; Stambler, Popper, Dubinsky, & Stimson, 1991; Steven & Broadbent, 1997).

When stress exceeds a critical threshold, corals will exhibit a general stress response called coral bleaching where the coral host and Symbiodiniaceae dissociate from each other, thereby losing this input of carbon (Glynn, 1991). This thermal-induced symbiotic breakdown, or dysbiosis, is primarily believed to occur through light-induced photodamage and photoinhibition on the photosynthetic apparatus, particularly in photosystem II (Weis, 2008). As the coral animal is dependent upon the organic carbon produced by their photoautotrophic algal symbionts for their daily energy requirements, the loss of Symbiodiniaceae and their photosynthetic products causes them to starve. If

the stress is too severe or persists for too long, bleaching will result in coral mortality (Glynn, 1991).

1.2.2 The genus *Durusdinium*

The genus *Durusdinium* contains four described species including *Durusdinium trenchii* (D1-4), *Durusdinium boreum* (type D15), *Durusdinium glynni*, and *Durusdinium eurythalpos* (D8-13) (LaJeunesse et al., 2014). Phylogenetic analysis of marker genes shows that *D. eurythalpos* and *D. boreum* have diverged extensively, while it is not possible to differentiate *D. glynni* and *D. trenchii* from each other using the cp23S and ribosomal large subunit marker genes (Drew C Wham, Ning, & LaJeunesse, 2017). The diversification of *D. eurythalpos and D. boreum* from the other *Durusdinium* species is likely a result of their specialization with the zebra coral *Oulastrea crispata*, a brooder coral that often transmits its endosymbionts vertically (Thornhill, Lewis, Wham, & LaJeunesse, 2014). Phylogenetic analysis further suggests ecological diversification between the two species, which is reflected in their overlapping, but distinct, latitudinal distributions where *D. boreum* is found in more northern latitudes *(LaJeunesse et al., 2014)*.

While both *D. glynni* and *D. trenchii* are found in tropical regions, *D. glynni* forms symbioses in the Eastern Pacific with brooding corals from the genera *Montipora* and *Pocillopora*, while *D. trenchii* is the most host-generalist species within the *Durusdinium* genus and associates with diverse brooding and broadcast spawning coral species across the Caribbean and Indo-Pacific (LaJeunesse et al., 2014). Similarly to *D. trenchii*, *D. glynnii* is also noted for conferring thermotolerance on the coral holobiont and is often

found in corals that experience greater and/or more dynamic temperature regimes (Cunning, Ritson-Williams, & Gates, 2016; McGinley et al., 2012; Rowan, 2004). Microsatellite studies have revealed that *D. trenchii* is diallelic at nearly every loci and has four alleles at one loci that is diallelic in *D. glynnii*, suggesting genome duplication has been a factor in the evolution of these two species and perhaps their acquired thermotolerance (LaJeunesse et al., 2014; Drew C Wham et al., 2017; Drew C. Wham, Pettay, & LaJeunesse, 2011).

Believed to be an invasive species of Symbiodiniaceae from the Indo-Pacific, *Durusdinium trenchii* is an ecologically-important species of Symbiodiniaceae as a result of its resilience to thermal stress and broad host range (Pettay, Wham, Smith, Iglesias-Prieto, & LaJeunesse, 2015). Many studies have documented the increased thermal resilience conferred on corals by hosting *D. trenchii* (Berkelmans & Van Oppen, 2006; Boulotte et al., 2016; Jones, Berkelmans, van Oppen, Mieog, & Sinclair, 2008; Keshavmurthy et al., 2012; Manzello et al., 2019; Mieog, van Oppen, Cantin, Stam, & Olsen, 2007). Even when free-living in cultures, *D. trenchii* exhibits greater thermotolerance than other Symbiodiniaceae species. Out of six different species of Symbiodiniaceae across five different genera, *D. trenchii* was the only species to exhibit no changes in net ROS production or gene expression of the antioxidants catalase and superoxide dismutase under heat stress (McGinty, Pieczonka, & Mydlarz, 2012).

Although the mechanisms underlying the increased thermotolerance in *D. trenchii* are not fully understood, it seems that one component in this is increased structural integrity in thylakoid membranes (Takahashi, Whitney, & Badger, 2009). Membrane lipids have

been shown to be indicative of sensitivity to thermal-induced bleaching (Tchernov et al., 2004), particularly certain lipids that are found in higher abundance within *D. trenchii* such as sulfoquinovosyldiacylglycerols as well as higher ratios of digalactosyldiacylglycerol to monogalactosyldiacylglycerol lipids. What other factors contribute to *D. trenchii*'s increased thermotolerance and how, such as higher turnover and repair rates of photosystem II during photoinhibition which is one of the main underlying causes of heat-induced bleaching , remains to be elucidated (Warner, Fitt, & Schmidt, 1999).

1.3 Dissertation objectives, hypotheses, and organization

The purpose of my dissertation was to first investigate the sequence and structural diversity of the insulin-like peptides across the phylum Cnidaria, focusing on the major Classes including Anthozoa, Hydrozoa, Cubozoa, Scyphozoa, and Staurozoa. After elucidating the diversity of ILPs across the phylum and the level of conservation, the phylogenetic analysis was used as a framework to systematically investigate the transcriptional responses of cnidarian ILPs to symbiosis and stress, particularly elevated temperatures and ocean acidification. Finally, my dissertation investigates the genomic underpinnings of thermotolerance in coral symbionts by producing the first genome assembly for the thermotolerant coral symbiont *D. trenchii,* a known coral symbiont that confers greater thermotolerance on its hosts.

Chapter two addressed the following questions:

Do symbiotic coral hosts possess insulin-like peptides? What is the diversity of insulin-like peptides across different taxonomic groups within the phylum Cnidaria?

Chapter three addressed the following questions and hypotheses:

Are there conserved patterns in the gene expression of ILPs across cnidarian species in response to symbiosis and stress?

H_{3a}: The expression of ILPs in different Cnidarian species will change in response to the coral-algal symbiosis as it is centered around the exchange of nutritional compounds of which the insulin signaling pathway is a master regulator of and known to be stimulated by.

H_{3b}: Considering the well-documented functional role of the insulin signaling pathway in stress response, I also postulate that they will exhibit similar expression changes considering the high level of structural conservation in this pathway as evidenced by the activation of cnidarian insulin receptors with mammalian insulin.

Chapter four addressed the following questions:

What genomic signatures within the genome of *Durusdinium trenchii* might underly its higher thermotolerance compared to other Symbiodiniaceae species?

Does the genome sequence of *Durusdinium trenchii* support previous predictions regarding a near- or whole-genome duplication event in its evolutionary history? Further, if extensive duplication is identified will this be connected to any genomic signatures of thermotolerance as has been previously hypothesized?

1.4 References

Antonelli, P. L., Rutz, S. F., Sammarco, P. W., & Strychar, K. B. (2016). Evolution of symbiosis in hermatypic corals: A model of the past, present, and future.

Nonlinear Analysis: Real World Applications, 32, 389-402. doi:10.1016/j.nonrwa.2016.05.004

Aslam, A. F., Kiya, T., Mita, K., & Iwami, M. (2011). Identification of novel bombyxin genes from the genome of the silkmoth *Bombyx mori* and analysis of their expression. *Zoological science, 28*(8), 609-616. doi:10.2108/zsj.28.609

Barbieri, M., Bonafè, M., Franceschi, C., & Paolisso, G. (2003). Insulin/IGF-I-signaling pathway: an evolutionarily conserved mechanism of longevity from yeast to humans. *American Journal of Physiology, Endocrinology, and Metabolism, 285*(5), E1064-1071. doi:10.1152/ajpendo.00296.2003

Berkelmans, R., & Van Oppen, M. J. H. (2006). The role of zooxanthellae in the thermal tolerance of corals: a 'nugget of hope' for coral reefs in an era of climate change. *Proceedings of the Royal Society B: Biological Sciences, 273*, 2305-2312. doi:10.1098/rspb.2006.3567

Boulotte, N. M., Dalton, S. J., Carroll, A. G., Harrison, P. L., Putnam, H. M., Peplow, L. M., & van Oppen, M. J. H. (2016). Exploring the *Symbiodinium* rare biosphere provides evidence for symbiont switching in reef-building corals. *Nature Publishing Group, 54*, 2693-2701. doi:10.1038/ismej.2016.54

Burks, D. J., Mora, J. F. D., Schubert, M., Withers, D. J., Myers, M. G., Towery, H. H., . . . White, M. F. (2000). IRS-2 pathways integrate female reproduction and energy homeostasis. *Nature, 407*(6802), 377-382. doi:10.1038/35030105

Burns, J. A., Zhang, H., Hill, E., Kim, E., & Kerney, R. (2017). Transcriptome analysis illuminates the nature of the intracellular interaction in a vertebrate-algal symbiosis. *eLife, 6*. doi:10.7554/eLife.22054.001

Chang, L., Chiang, S.-H., & Saltiel, A. R. (2004). Insulin signaling and the regulation of glucose transport. *Molecular Medicine, 10*(7-12), 65-71. doi:10.2119/2005-00029.Saltiel

Claeys, I., Simonet, G., Poels, J., Van Loy, T., Vercammen, L., De Loof, A., & Vanden Broeck, J. (2002). Insulin-related peptides and their conserved signal transduction pathway. *Peptides, 23*(4), 807-816. doi:10.1016/S0196-9781(01)00666-0

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., . . . van den Belt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature, 387*(6630), 253-260. doi:10.1038/387253a0

Cunning, R., Ritson-Williams, R., & Gates, R. D. (2016). Patterns of bleaching and recovery of *Montipora capitata* in Kāne ʻohe Bay, Hawai ʻi, USA. *Marine Ecology Progress Series, 551*, 131-139. doi:10.3354/meps11733

Davy, S. K., Allemand, D., & Weis, V. M. (2012). Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiology and Molecular Biology Reviews, 76*(2), 229-261. doi:10.1128/MMBR.05014-11

De Meyts, P. (2004). Insulin and its receptor: Structure, function and evolution. *BioEssays, 26*(12), 1351-1362. doi:10.1002/bies.20151

Ferrier-Pages, C., Gattuso, J. P., Dallot, S., & Jaubert, J. (2000). Effect of nutrient enrichment on growth and photosynthesis of the zooxanthellate coral *Stylophora pistillata*. *Coral Reefs, 19*(2), 103-113. doi:10.1007/s003380000078

Gabillard, J.-C., Weil, C., Rescan, P.-Y., Navarro, I., Gutiérrez, J., & Le Bail, P.-Y. (2003). Effects of environmental temperature on IGF1, IGF2, and IGF type I receptor expression in rainbow trout (*Oncorhynchus mykiss*). *General and Comparative Endocrinology, 133*(2), 233-242. doi:10.1016/s0016-6480(03)00167-9

Gems, D., Sutton, A. J., Sundermeyer, M. L., Albert, P. S., King, K. V., Edgley, M. L., . . . Riddle, D. L. (1998). Two pleiotropic classes of daf-2 mutation affect larval arrest, adult behavior, reproduction and longevity in *Caenorhabditis elegans*. *Genetics, 150*(1), 129-155.

Glynn, P. W. (1991). Coral reef bleaching in the 1980s and possible connections with global warming. *Trends in Ecology and Evolution, 6*(6), 175-179. doi:10.1016/0169-5347(91)90208-F

González, J. P., Farnés, O. C., Vasconcelos, A. T. R., & Pérez, A. G. (2009). Conservation of key members in the course of the evolution of the insulin signaling pathway. *Biosystems, 95*(1), 7-16. doi:10.1016/j.biosystems.2008.06.003

Grönke, S., Clarke, D.-F., Broughton, S., Andrews, T. D., & Partridge, L. (2010). Molecular Evolution and Functional Characterization of *Drosophila* Insulin-Like Peptides. *PLoS Genetics, 6*(2), e1000857. doi:10.1371/journal.pgen.1000857

Ikeya, T., Broughton, S., Alic, N., Grandison, R., & Partridge, L. (2009). The endosymbiont *Wolbachia* increases insulin/IGF-like signalling in *Drosophila*. *Proceedings of the Royal Society of London B: Biological Sciences, 276*(1674), 3799-3807. doi:10.1098/rspb.2009.0778

Jin Chan, S., & Steiner, D. F. (2000). Insulin through the ages: phylogeny of a growth promoting and metabolic regulatory hormone. *American Zoology, 40*(2), 213-222. doi:10.1093/icb/40.2.213

Jones, A. M., Berkelmans, R., van Oppen, M. J., Mieog, J. C., & Sinclair, W. (2008). A community change in the algal endosymbionts of a scleractinian coral following a natural bleaching event: field evidence of acclimatization. *Proceedings of the Royal Society B: Biological Sciences, 275*, 1359-1365. doi:doi.org/10.1098/rspb.2008.0069

Keshavmurthy, S., Hsu, C. M., Kuo, C. Y., Meng, P. J., Wang, J. T., & Chen, C. A. (2012). Symbiont communities and host genetic structure of the brain coral *Platygyra verweyi*, at the outlet of a nuclear power plant and adjacent areas. *Molecular Ecology, 21*(17), 4393-4407. doi:10.1111/j.1365-294X.2012.05704.x

Kirk, N. L., & Weis, V. M. (2016). Animal–*Symbiodinium* Symbioses: Foundations of Coral Reef Ecosystems. In *The Mechanistic Benefits of Microbial Symbionts. Advances in Environmental Microbiology, vol. 2* (pp. 269-294): Springer, Cham.

Kubrak, O. I., Kučerová, L., Theopold, U., & Nässel, D. R. (2014). The sleeping beauty: how reproductive diapause affects hormone signaling, metabolism, immune response and somatic maintenance in *Drosophila melanogaster*. *PLoS ONE, 9*(11), e113051. doi:10.1371/journal.pone.0113051

LaJeunesse, T. C., Parkinson, J. E., Gabrielson, P. W., Jeong, H. J., Reimer, J. D., Voolstra, C. R., & Santos, S. R. (2018). Systematic Revision of Symbiodiniaceae Highlights the Antiquity and Diversity of Coral Endosymbionts. *Current Biology, 28*(16), 2570-2580.e2576. doi:10.1016/J.CUB.2018.07.008

LaJeunesse, T. C., Wham, D. C., Pettay, D. T., Parkinson, J. E., Keshavmurthy, S., & Chen, C. A. (2014). Ecologically differentiated stress-tolerant endosymbionts in the dinoflagellate genus *Symbiodinium* (Dinophyceae) Clade D are different species. *Phycologia, 53*(4), 305-319. doi:10.2216/13-186.1

Larsen, P. L. (1993). Aging and resistance to oxidative damage in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences, 90*(19), 8905-8909. doi:10.1073/pnas.90.19.8905

Leggat, W., Marendy, E. M., Baillie, B., Whitney, S. M., Ludwig, M., Badger, M. R., & Yellowlees, D. (2002). Dinoflagellate symbioses: strategies and adaptations for the acquisition and fixation of inorganic carbon. *Functional Plant Biology, 29*(3), 309-309. doi:10.1071/PP01202

Libert, S., Chao, Y., Zwiener, J., & Pletcher, S. D. (2008). Realized immune response is enhanced in long-lived puc and chico mutants but is unaffected by dietary restriction. *Molecular immunology, 45*(3), 810-817. doi:10.1016/j.molimm.2007.06.353

Manzello, D. P., Matz, M. V., Enochs, I. C., Valentino, L., Carlton, R. D., Kolodziej, G., . . . Jankulak, M. (2019). Role of host genetics and heat-tolerant algal symbionts in sustaining populations of the endangered coral *Orbicella faveolata* in the Florida Keys with ocean warming. *Global Change Biology, 25*(3), 1016-1031. doi:10.1111/gcb.14545

Marquez, A. G., Pietri, J. E., Smithers, H. M., Nuss, A., Antonova, Y., Drexler, A. L., . . . Luckhart, S. (2011). Insulin-like peptides in the mosquito *Anopheles stephensi*: Identification and expression in response to diet and infection with *Plasmodium falciparum*. *General and Comparative Endocrinology, 173*(2), 303-312. doi:10.1016/j.ygcen.2011.06.005

McGinley, M. P., Aschaffenburg, M. D., Pettay, D. T., Smith, R. T., LaJeunesse, T. C., & Warner, M. E. (2012). Transcriptional response of two core photosystem genes in *Symbiodinium* spp. exposed to thermal stress. *PLoS ONE, 7*(12), e50439. doi:10.1371/journal.pone.0050439

McGinty, E. S., Pieczonka, J., & Mydlarz, L. D. (2012). Variations in Reactive Oxygen Release and Antioxidant Activity in Multiple *Symbiodinium* Types in Response to Elevated Temperature. *Microbial Ecology, 64*(4), 1000-1007. doi:10.1007/s00248-012-0085-z

Mieog, J. C., van Oppen, M. J., Cantin, N. E., Stam, W. T., & Olsen, J. L. (2007). Real-time PCR reveals a high incidence of *Symbiodinium* clade D at low levels in four scleractinian corals across the Great Barrier Reef: implications for symbiont shuffling. *Coral Reefs, 26*, 449-457. doi:doi.org/10.1007/s00338-007-0244-8

Muller-Parker, G., McCloskey, L. R., Hoegh-Guldberg, v., & McAuley, P. J. (1994). Effect of Ammonium Enrichment on Animal and Algal Biomass of the Coral *Pocillopora damicornis*. *Pacific Science, 48*(3), 273-283.

Muscatine, L., Falkowski, P. G., Porter, J. W., & Dubinsky, Z. (1984). Fate of Photosynthetic Fixed Carbon in Light- and Shade-Adapted Colonies of the Symbiotic Coral *Stylophora pistillata*. *Proceedings of the Royal Society of Biological Sciences, 222*, 181-202. doi:10.1098/rspb.1984.0058

Muscatine, L., & Hand, C. (1958). Direct evidence for the transfer of materials from symbiotic algae to the tissues of a coelenterate. *Proceedings of the National Academy of Sciences, 44*(12), 1259-1263. doi:10.1073/pnas.44.12.1259

Nagasawa, H., Kataoka, H., Hori, Y., Isogai, A., Tamura, S., Suzuki, A., . . . Fujishita, M. (1984). Isolation and some characterization of the prothoracicotropic hormone from *Bombyx mori*. *General and Comparative Endocrinology, 53*(1), 143-152. doi:doi: 10.1016/0016-6480(84)90234-x

Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, A., Ishizaki, H., . . . Suzuki, A. (1984). Amino-terminal amino acid sequence of the silkworm prothoracicotropic hormone: homology with insulin. *science, 226*(4680), 1344-1345. doi:10.1126/science.226.4680.1344

Pettay, D. T., Wham, D. C., Smith, R. T., Iglesias-Prieto, R., & LaJeunesse, T. C. (2015). Microbial invasion of the Caribbean by an Indo-Pacific coral zooxanthella. *Proceedings of the National Academy of Sciences, 112*(24), 7513-7518. doi:10.1073/pnas.1502283112

Pietri, J. E., Pakpour, N., Napoli, E., Song, G., Pietri, E., Potts, R., . . . Luckhart, S. (2016). Two insulin-like peptides differentially regulate malaria parasite infection in the mosquito through effects on intermediary metabolism. *Biochemical Journal, 473*(20), 3487-3503. doi:10.1042/BCJ20160271

Pietri, J. E., Pietri, E. J., Potts, R., Riehle, M. A., & Luckhart, S. (2015). *Plasmodium falciparum* suppresses the host immune response by inducing the synthesis of insulin-like peptides (ILPs) in the mosquito *Anopheles stephensi*. *Developmental & Comparative Immunology, 53*(1), 134-144. doi:10.1016/J.DCI.2015.06.012

Piper, M. D. W., Selman, C., McElwee, J. J., & Partridge, L. (2008). Separating cause from effect: How does insulin/IGF signalling control lifespan in worms, flies and mice? *Journal of Internal Medicine, 263*(2), 179-191. doi:10.1111/j.1365-2796.2007.01906.x

Riddle, D. L., Blumenthal, T., Meyer, B., & Priess, J. (1997). Genetic and Environmental Regulation of Dauer Larva Development. In *C. elegans II. 2nd edition* (ed.). Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Rosenberg, E., Koren, O., Reshef, L., Efrony, R., & Zilber-Rosenberg, I. (2007). The role of microorganisms in coral health, disease and evolution. *Nature Reviews Microbiology, 5*(5), 355-362. doi:10.1038/nrmicro1635

Rowan, R. (2004). Thermal adaptation in reef coral symbionts. *Nature, 430*, 742. doi:10.1038/430742a

Schwenke, R. A., Lazzaro, B. P., & Wolfner, M. F. (2016). Reproduction–immunity trade-offs in insects. *Annual review of entomology, 61*, 239-256. doi:doi.org/10.1146/annurev-ento-010715-023924

Shabanpoor, F., Separovic, F., & Wade, J. D. (2009). The human insulin superfamily of polypeptide hormones. *Vitamins & Hormones, 80*, 1-31. doi:10.1016/S0083-6729(08)00601-8

Sparkman, A., Byars, D., Ford, N., & Bronikowski, A. (2010). The role of insulin-like growth factor-1 (IGF-1) in growth and reproduction in female brown house snakes (*Lamprophis fuliginosus*). *General and Comparative Endocrinology, 168*(3), 408-414. doi:10.1016/j.ygcen.2010.05.006

Stambler, N., Popper, N., Dubinsky, Z., & Stimson, J. (1991). Effects of Nutrient Enrichment and Water Motion on the Coral *Pocillopora damicornis*. *Pacific Science, 45*(3), 299-307.

Stat, M., Carter, D., & Hoegh-Guldberg, O. (2006). The evolutionary history of *Symbiodinium* and scleractinian hosts-Symbiosis, diversity, and the effect of climate change. *Perspectives in Plant Ecology, Evolution and Systematics, 8*(1), 23-43. doi:10.1016/j.ppees.2006.04.001

Steele, R., Lieu, P., Mai, N. H., Shenk, M. A., & Sarras Jr, M. P. (1996). Response to insulin and the expression pattern of a gene encoding an insulin receptor homologue suggest a role for an insulin-like molecule in regulating growth and patterning in *Hydra*. *Development genes and evolution, 206*(4), 247-259. doi:10.1007/s004270050050

Steven, A. D. L., & Broadbent, A. D. (1997). Growth and metabolic responses of *Acropora palifera* to long term nutrient enrichment. *Proc. 8th Int. Coral Reef Symp*.

Szmant, A. M. (2002). Nutrient enrichment on coral reefs : Is it a major cause of coral reef decline. *Estuaries, 25*, 743-766. doi:10.1007/BF02804903

Takahashi, S., Whitney, S. M., & Badger, M. R. (2009). Different thermal sensitivity of the repair of photodamaged photosynthetic machinery in cultured *Symbiodinium* species. *Proceedings of the National Academy of Sciences, 106*(9), 3237-3242. doi:10.1073/pnas.0808363106

Tchernov, D., Gorbunov, M. Y., De Vargas, C., Yadav, S. N., Milligan, A. J., Häggblom, M., & Falkowski, P. G. (2004). Membrane lipids of symbiotic algae are diagnostic of sensitivity to thermal bleaching in corals. *Proceedings of the National Academy of Sciences, 101*(37), 13531-13535.

Thornhill, D. J., Lewis, A. M., Wham, D. C., & LaJeunesse, T. C. (2014). Host-specialist lineages dominate the adaptive radiation of reef coral endosymbionts. *Evolution, 68*(2), 352-367. doi:10.1111/evo.12270

Wakefiel, T. S., & Kempf, S. C. (2001). Development of host- and symbiont-specific monoclonal antibodies and confirmation of the origin of the symbiosome membrane in a cnidarian-dinoflagellate symbiosis. *The Biological bulletin, 200*(2), 127-143. doi:10.2307/1543306

Warner, M. E., Fitt, W. K., & Schmidt, G. W. (1999). Damage to photosystem II in symbiotic dinoflagellates: a determinant of coral bleaching. *Proceedings of the National Academy of Sciences, 96*(14), 8007-8012. doi:10.1073/pnas.96.14.8007

Weis, V. M. (2008). Cellular mechanisms of Cnidarian bleaching: stress causes the collapse of symbiosis. *The Journal of Experimental Biology, 211*, 3059-3066. doi:10.1242/jeb.009597

Weis, V. M., & Allemand, D. (2009). What determines coral health? *Science (New York, N.Y.), 324*(5931), 1153-1155. doi:10.1126/science.1172540

Wham, D. C., Ning, G., & LaJeunesse, T. C. (2017). *Symbiodinium glynnii* sp. nov., a species of stress-tolerant symbiotic dinoflagellates from pocilloporid and montiporid corals in the Pacific Ocean. *Phycologia, 56*(4), 396-409. doi:10.2216/16-86.1

Wham, D. C., Pettay, D. T., & LaJeunesse, T. C. (2011). Microsatellite loci for the host-generalist "zooxanthella" *Symbiodinium trenchi* and other Clade D *Symbiodinium*. *Conservation Genetics Resources, 3*, 541-544. doi:10.1007/s12686-011-9399-2

Wolff, S., Ma, H., Burch, D., Maciel, G. A., Hunter, T., & Dillin, A. (2006). SMK-1, an essential regulator of DAF-16-mediated longevity. *Cell, 124*(5), 1039-1053. doi:10.1016/j.cell.2005.12.042

Yuyama, I., Ishikawa, M., Nozawa, M., Yoshida, M.-a., & Ikeo, K. (2018). Transcriptomic changes with increasing algal symbiont reveal the detailed process underlying establishment of coral-algal symbiosis. *Scientific Reports, 8*, 1-11. doi:doi.org/10.1038/s41598-018-34575-5

Zheng, S., Chiu, H., Boudreau, J., Papanicolaou, T., Bendena, W., & Chin-Sang, I. (2018). A functional study of all 40 *Caenorhabditis elegans* insulin-like peptides. *Journal of Biological Chemistry, 293*(43), 16912-16922. doi:10.1074/jbc.RA118.004542

Chapter 2: Phylogenetic analysis reveals expansion of insulin-like peptide protein family in Anthozoa

2.1 Abstract

The insulin superfamily of hormones is an evolutionarily important group of hormones responsible for regulating a diverse array of functions including metabolism, growth, development, and longevity in animals. Insulin-like peptides have been found in almost every metazoan phyla and have been hypothesized to pre-date the emergence of Metazoa. In spite of their widespread nature, investigations of insulin-like peptide diversity in non-bilaterian phyla have been few, and resulted in a lack of appreciation for the potential diversity of the family of hormones within the nonbilaterian phyla. We data mined the NCBI Short Read Archive for transcriptomic sequencing data from 93 cnidarian species spread across 10 orders from all the cnidarian classes including Anthozoa, Cubozoa, Hydrozoa, Scyphozoa, and Staurozoa. Identification of putative cnidarian insulin-like peptides was accomplished with the combination of NCBI-BLAST and HHBlits, which allowed the identification of remote homologs using structure. In great contrast to the previously appreciated diversity of seven insulin-like peptides within the phylum cnidaria, we recovered 360 putative cnILPs across the 93 cnidarian species and classified them into four structural types (A-C) according to the patterns of their conserved cysteines that form the disulfide bonds essential to the final insulin-like structure. Type A had the canonical insulin motifs of three disulfide bonds and six conserved cysteines while types B and C had two additional B chain cysteines resulting in an additional intra-B chain disulfide bond. The taxonomic distribution of the cnILPs suggests that there was

an expansion and diversification of cnILPs within the class Anthozoa. Specifically, this diversification in cnILPs likely occurred after the split between the sub-classes Hexacorallia and Octocorallia as the Octocorallian order Alcyonacea only possessed two cnILPs in contrast to the eight and seven cnILPs found in the Hexacorallian orders Scleractinia and Actiniaria, respectively. Genomic mapping of these eight cnILPs confirms that they are distinct genes rather than isoforms and the result of alternative splicing. This dissertation is the first record of such extensive ILP diversity in a non-bilaterian as besides the three cnILPs previously identified in *H. vulgaris*, no other non-bilaterian species has been shown to possess more than one ILP. Future studies should investigate whether the diverse repertoire of the Scleractinian cnILPs results in a similar level of functional diversity as the group contains the ecologically important, yet threatened, corals.

2.2 Introduction

The insulin signaling pathway (ISP) is an evolutionarily important metabolic pathway that is integral to linking fluctuating nutrient levels and environmental cues with the regulation of development, metabolism, growth, and longevity in most, if not all, metazoans (Barbieri, Bonafè, Franceschi, & Paolisso, 2003; Jin Chan & Steiner, 2000; Pertseva & Shpakov, 2002; Saltiel & Kahn, 2001; Taniguchi, Emanuelli, & Kahn, 2006). Activated by the binding of insulin, insulin-like growth factors (IGFs) or insulin-like peptides (ILPs) to their transmembrane receptor, this pathway controls a multitude of downstream signaling cascades including the phosphoinositide 3-kinase (PI3K), Ras/MAPK (mitogen-activated protein kinase), and mTOR pathways. In fact,

bioinformatic investigations have revealed that all ISP downstream effector pathways are conserved between vertebrates and invertebrates (González, Farnés, Vasconcelos, & Pérez, 2009).

Insulin-like peptides have been found in almost all animal phyla, suggesting that they predate the emergence of Metazoans around 600 million years ago. Surprisingly, the identification of similar proteins to insulin in protozoans, fungi, and even bacteria potentially push back the evolutionary origins of insulin much farther than previously thought to about 1-2 billion years ago (Le Roith, Shiloach, Roth, & Lesniak, 1980), long before the appearance of Metazoans. Recent years have led to a large expansion in the investigation and discovery of ILPs in many invertebrate taxa including species from the groups Tunicata, Nematoda, Crustacea, Mollusca, and Annelida (see Cherif et al. (2019)). In invertebrates, ILPs have even been identified in the non-bilaterian species *Hydra vulgaris* (Steele, Lieu, Mai, Shenk, & Sarras Jr, 1996) and *Nematostella vectensis* (Anctil, 2009) from the phylum Cnidaria along with *Geodia cydonium* from the phylum Porifera (Robitzki et al., 1989)

Despite the low amino acid sequence similarity between ILPs, there is a stunningly high level of conservation in the final ILP protein structure as mammalian insulin is able to activate the insulin-like receptor in the cnidarian hydrozoan *H. vulgaris* (Steele et al., 1996). Synthesized as pre-prohormones, ILPs most commonly consist of four domains including a signal peptide, B chain, C peptide, and A chain. The signal and C peptides are cleaved by proprotein convertase, followed by the bridging of the B and A chains via several disulfide bonds at conserved cysteine residues and yields the final insulin

molecule (Rutter, Pullen, Hodson, & Martinez-Sanchez, 2015). Recent work has classified the diverse insulin-like family into α, β, and γ types according to the position of these disulfide bonds (Pierce et al., 2001). In the γ type ILPs that include the canonical insulin hormone structure, there is one conserved intra-chain disulfide bond on the A chain along with four cysteines that are responsible for bridging the A and B chains through two inter-chain disulfide bonds. ILP β type structures, found in multiple invertebrate phyla, possess an additional inter-chain disulfide bond resulting in three total inter-chain disulfide bonds along with the single intra-chain disulfide bond. Similarly, γ ILP types, which have only been found in *C. elegans*, possess three inter-chain disulfide bonds but lack the intra-chain disulfide bond usually found on the A chain.

A focal point in the research surrounding invertebrate ILPs is how the diversity in the types and number of ILPs present in a species connects to the functional diversity of their ISP. While vertebrates have three ISP receptors including the insulin receptor, IGF1, and IGF2 receptors, invertebrates only possess a single insulin-like receptor (González et al., 2009; Pertseva & Shpakov, 2002). This reduced diversity in the receptors implies that counter to vertebrates where diversity in both the peptide and receptor can result in functional diversity, functional diversity in invertebrate ISPs may be more dependent on ILP diversity. Interestingly, invertebrates exhibit a wide range in the number of ILPs anywhere from three in the cnidarian *H. vulgaris* to 40 in *C. elegans* and 38 in the silkworm *Bombyx mori* (Aslam, Kiya, Mita, & Iwami, 2011; Iwami, 2000). Functional work in *C. elegans*, *D. melanogaster*, *D, melanogaster*, and the oyster *Crassostrea gigas* has demonstrated that ILPs are expressed in different tissues at different developmental

stages implying there is functional diversity (Baumeister, Schaffitzel, & Hertweck, 2006; Cherif et al., 2019; Jia, Chen, & Riddle, 2004; Mizoguchi & Okamoto, 2013; Wu & Brown, 2006). Furthermore, ILPs in *C. elegans* can be both agonists and antagonists of the ISP and its downstream functions (Zheng et al., 2018). Multiple invertebrate insulin-like peptides possess metabolism-controlling functions including bombyxin from silk moths (Satake et al., 1997) and locust molluscan insulin-related proteins (Lagueux, Lwoff, Meister, Goltzene, & Hoffmann, 1990; Smit et al., 1996).

While cnidarian ILPs (cnILPs) have only been identified in the hydrozoan *Hydra vulgaris* (Steele et al., 1996) and the anthozoan *Nematostella vectensis* (Putnam et al., 2007), no studies have identified members of the protein family in other members of the diverse cnidarian phylum that includes the ecologically important corals. Considering the ancient evolutionary origins of the ILP family, it seems unlikely that corals do not possess ILPs despite the lack of discovery of coral ILPS in the >150 genomic and transcriptomics studies that have been conducted and are publicly available. As the final protein structures of insulin peptides are highly conserved but the amino acid sequence is not between invertebrates and vertebrates, this failure of identifying coral ILPs might be partially due to the dependence of the field on sequence similarity-based approaches that are unable to identify structurally conserved proteins, despite the divergent amino acid sequences. However, gene ontology analyses from transcriptomic studies have implicated the coral ISP in the response to thermal and nutrient stress (Rosic et al., 2014).

Herein, we employ data mining and transcriptomics to perform an in-depth search for ILPs across all classes and 10 orders within the phylum Cnidaria, one of the earliest

diverging metazoan phyla (for review see Collins (2009)). We also use phylogenetics to characterize the diversity of ILPs within and among each of the classes. Our work will serve as a foundation for future work on the functional role of ILPs in diverse cnidarians including the ecologically important and endangered corals.

2.3 Methods

2.3.1 Data mining and transcriptome assembly

In order to investigate the diversity of insulin-like peptides in Cnidarians, we data mined the NCBI Short Read Archive to conduct a comparative transcriptomics study on insulin-like peptides and their receptor by analyzing transcriptomic data from 93 species across 10 orders in the cnidarian classes Anthozoa (n=50; Orders: Actiniaria, Alcyonacea, Scleractinia), Cubozoa (n=5; Orders: Carybdeida, Chirodropida), Hydrozoa (n=25; Orders: Anthoathecata, Siphonophorae), Scyphozoa (n=8; Orders: Semaeostomeae, Rhizostomeae), and Staurozoa (n=5; Orders: Stauromedusae). Transcriptomic data quality was evaluated and trimmed using FastQC (Andrews, 2010) and Trimmomatic (Bolger, Lohse, & Usadel, 2014), respectively, and then assembled using Trinity v2.6.5 with a 400bp minimum transcript cut-off (Haas et al., 2013).

2.3.2 Assessment of *de novo* transcriptome assembly quality

The resulting transcriptome assemblies were first filtered to remove contaminants using Kraken, an ultra-fast metagenomic sequence classifier, against the standard Kraken database of all bacterial, archaeal, and viral genomes present in RefSeq as of January 10, 2019 (Wood & Salzberg, 2014). Transcriptome assemblies were then additionally filtered

using the spliced aligner Hisat2 (Kim, Paggi, Park, Bennett, & Salzberg, 2019) against a custom database built from Cnidarian the symbiont genomes *Symbiodinium microadriaticum*, *Breviolum minutum*, *Cladocopium goreaui*, *Fugacium kawagutii*, *Chlorella variabilis*, *Chromera velia*, *Chlorella vulgaris*, and *Chlorella sorokiniana*. Transcriptome assembly completeness was assessed using BUSCO (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) in combination with the core metazoan single-copy ortholog dataset and quality was assessed using Transrate (Smith-Unna, Boursnell, Patro, Hibberd, & Kelly, 2016). The final dataset for subsequent analysis only included more than 80% completeness according to BUSCO.

2.3.3 Identification of putative insulin-like peptides

As a result of the low sequence similarity, NCBI-BLAST+ was used in combination with HHblits and the reference PDB70 database to identify remote homologs below 20% amino acid identity (Remmert, Biegert, Hauser, & Söding, 2012). To determine if biases inherent to the different assemblies contributed to the number of transcripts identified for a particular gene, a regression analysis of the identified putative cnidarian ILPs in a species was conducted in R (R Core Team 2014) against various metrics of the assembly process. These metrics included the number of input bases, number of transcripts in transcriptome assembly, complete BUSCOs, single BUSCOs, duplicated BUSCOs, fragmented BUSCOs, missing BUSCOs, and the species' taxonomic group (i.e., Scleractinia, Hexacorallia (minus Scleractinia), Octocorallia, Hydrozoa, Scyphozoa, Staurozoa, and Cubozoa).

2.3.4 Phylogenetic inference

Protein sequences for vertebrate insulin as well as the *Drosophila* insulin-like peptides (DILPs) from 19 species in the genus *Drosophila* were used as outgroups for the phylogenetic analysis. This included *Drosophila* insulin-like peptides from the species *D. arizonae, D. biarmipes, D. bipectinata, D. busckii, D. elegans, D. eugracilis, D. ficusphila, D. hydei, D. kikkawai, D. melanogaster, D. miranda, D. navojoa, D. obscura, D. rhoipaloa, D. serrata, D. simulans, D. suzukii, D. takahasii,* and *D. willistoni*. A multiple sequence alignment of the putative cnILPs, DILP 1-7, and vertebrate insulin was conducted with MAFFT using the E-INS-I iterative refinment algorithm that is intended for genes with multiple conserved domains and long gaps (Katoh, Kuma, Toh, & Miyata, 2005). Phylogenetic inference of the multiple sequence alignment was performed using a maximum likelihood approach with the program IQ-TREE (Nguyen, Schmidt, Von Haeseler, & Minh, 2015) which performs comparably to more well-known phylogenetic programs (i.e., RaxML, PhyML, FastTree) while being time-efficient on large phylogenomic data sets (Zhou, Shen, Hittinger, & Rokas, 2018). The IQTree program performs model testing and selection on over 545 evolutionary models using ModelFinder (Kalyaanamoorthy, Minh, Wong, von Haeseler, & Jermiin, 2017) before subsequent tree construction. We assessed branch support using the ultrafast bootstrap (Hoang, Chernomor, Von Haeseler, Minh, & Vinh, 2018), SH-like approximate ratio test (Guindon et al., 2010), and approximate Bayes test (Anisimova, Gil, Dufayard, Dessimoz, & Gascuel, 2011). To prevent local optima from influencing the final tree structure the phylogenetic analysis was conducted 20 consecutive times and selected the tree with the highest log likelihood as the final tree.

2.3.5 Structural analysis

Signal peptides were identified using the SignalP-5.0 Server (Petersen, Brunak, Von Heijne, & Nielsen, 2011) and propeptide cleavage sites were detected using the ProP 1.0 Server that is designed to identify eukaryotic furin-specific and general proprotein convertase (PC) cleavage sites using neural networks (Duckert, Brunak, & Blom, 2004). For cnILPs that did not have an identified PC cleavage site between the B chain and C peptide and/or C peptide and A chain, where PC sites are to be expected in ILPs, the cnILPs were visually examined to determine if a putative PC site was present. As PC enzymes recognize single or paired basic motifs following a (R/K)-2nX-R↓ (n = 0–3 aa) pattern, any instances that met these qualifications were noted as a putative PC cleavage site. To determine the genomic architecture and location for putative cnILPs the protein sequences were searched against the genomes for the Scleractinian species *A. palmata, A. millepora,* and *A. digitifera*.

2.4 Results

2.4.1 Classification of the cnidarian insulin-like peptides

We used the remote homology protein detection program HHPred to identify ILPs from 10 orders in the phylum Cnidaria (Table 2-1), resulting in 360 putative cnILPs across the 93 species investigated in this work (Figure 2-1) compared to the 7 cnILPs previously identified in *Hydra vulgaris* and *Nematostella vectensis*. As the overall sequence in ILPs is not conserved, they are generally classified according to the number and distribution of cysteine residues in the A and B chains, which are the locations for the disulfide bonds that bridge the two chains and are integral to the insulin-like structure. We classified the

putative cnILPs into three groups (cnILPs A-C) according to the number and distribution of the conserved cysteine. Group A exhibited the characteristic γ type structure found in vertebrate insulin of six conserved cysteine residues and three disulfide bonds (Table 2-2). The remaining cnILPs were classified into groups B and C because of the two additional conserved cysteines in their B chains, resulting in a total of eight conserved cysteines and four disulfide bonds (Table 2-2). These cnILPs were classified into either group B or C according to the specific B chain motif they possessed as there were two distinctive motifs created by the additional cysteine residues. Compared to Group C, Group B had a more conserved B chain motif ($CX_{10}CCX_{3-6}C$) where the two additional cysteines were downstream of the canonical B chain cysteine residues while Group C had a more variable B chain motif with the additional cysteines instead found upstream of the two canonical B chain cysteine residues (Table 2-2). While some of the four patterns had internal dibasic PC cleavage sites in the B chain (Figure 2-2) that would remove either one (cnILP-C3) or both (cnILP-C4) of the additional B chain cysteines, these dibasic sites had a very low probability according to the ProP server of being a PC cleavage site and were therefore kept as C-type cnILPs.

2.4.2 Diversity of the cnidarian insulin-like peptides

Through the employment of phylogenetics, we were able to further identify ten distinct cnILPs based upon conserved sequence elements across the three cnILP groups, some of which were found across multiple classes within the cnidarian phylum (Table 2-2). These phylogenetic patterns yielded three group A cnILPs (cnILP-A1 through cnILP-A3), one group B cnILP (cnILP-B), and six group C cnILPs (cnILP-C1 through cnILP-C6). These

three cnILPs exhibited varying levels of conservation within the cnidarian phylum. While CnILP-A2 was the most prevalent among the orders as it was found in nine out of the ten orders in this study, it was followed by cnILP-B with eight orders, cnILP-A1 with six orders, and then then finally cnILP-A3 and the group C cnILPs with two orders. Hexacorallians, a subclass of the class Anthozoa, possessed the greatest diversity of cnILPs as Scleractinia had 7 distinct cnILPs similarly to Actiniaria which had 6 distinct ILPs. Out of the two other Anthozoan subclasses (Ceriantharia and Octocorallia), the order Alcyonacea from the subclass Ceriantharia was also included in our study but only two distinct cnILPs (cnILP-A2 and cnILP-B) were identified in the five species examined. The next highest diversity was found in the classes Cubozoa and Scyphozoa who possessed the exact same cnILP repertoire with a total of three distinct cnILPs (cnILP-A1, cnILP-A2, cnILP-B) while Hydrozoans possessed only two of these cnILPs (cnILP-A1, cnILP-A2), missing CnILP-B. The stalked jellyfish from the order Staurozoa had the fewest number of cnILPs, only possessing the common CnILP-B structure, which was common to every other cnidarian class in our study including Anthozoa, Cubozoa, Scyphozoa, and Staurozoa except for Hydrozoa. Only three putative *H. vulgaris* cnILPs were recovered from the transcriptome assembly generated as a part of our work of which all three matched to the previously identified *H. vulgaris* cnILPs (Steele et al., 1996). Of these, the previously identified *H. vulgaris* ILP1 (Acc: ADA67985.1) was classified in the cnILP-A2 group with 100% sequence identity and ILP2 and ILP3 to the cnILP-A1 group with 97.9% and 100% sequence identity between the database entries and the cnILP sequences generated in our study.

2.4.3 Conservation of cnILP sequences

At the generic level, conservation of cnILP amino acid sequences is high within the Scleractinian genus *Acropora*, for example, the eight identified cnILPs have an overall percent identity ranging from 80.4% to 91.9% (Table 2-3) in contrast to the 40% to 83% found in *Drosophila* spp (Grönke, Clarke, Broughton, Andrews, & Partridge, 2010). While the level of conservation for a particular ILP across *Acropora* spp. is high, the low level of sequence similarity between different ILPs identified in our study suggests that the different phylogenetic clades represent different ILP genes rather than isoforms. In support of the cnILPs being distinct genes, mapping of the different ILPs to the genomes reveals similar patterns among *Acropora* spp. and confirms that none of the identified groups map to the same genomic locus. Four (cnILP-A2a, cnILP-A2b, cnILP-A3, cnILP-B) of the seven *Acropora* cnILPs were not only located on the same genomic scaffold but also in the exact same order in the Pacific species *A. digitifera* and *A. millepora* and the Caribbean species *A. palmata*. This conserved syntenic order was identified in the cnILP-C3 and cnILP-C4 peptides that also mapped together to a separate scaffold in these species. Interestingly, in both cases the three species had introns located in the same positions for the cnILPs, comparable lengths in the introns, and comparable distances between the different cnILPs on the scaffold (Figure 2-3).

2.5 Discussion

Herein, we greatly expand our awareness of the impressive diversity of cnILPs through our identification of 360 putative cnILPs across 93 cnidarian species. This increased awareness of cnILP diversity expands our previous knowledge of three cnILPs in the

hydrozoan *H. vulgaris* (Steele et al., 1996) and 4 cnILPs in the anthozoan *N. vectensis* (Putnam et al., 2007). We recovered all three of the previously identified *H. vulgaris* ILPs with 100% sequence identity in two and 97.9% sequence identity in the third, demonstrating the accuracy of the cnILP sequences generated from the *de novo* trancriptome assemblies in our work. While a handful of functional studies have been conducted on the ISP in *H. vulgaris* (Mortzfeld et al., 2019; Sebestyén et al., 2017; Steele et al., 1996), the species only contains a small fraction of the ILP diversity present in cnidarians meaning there are several structural types of cnILPs that have yet to be functionally examined. In particular, no functional studies have yet to directly assess cnILPs in the ecologically important Scleractinian corals, which possess an impressive eight distinct cnILPs.

The low sequence homology of the identified cnILPs is a common characteristic in taxonomic studies of ILPs in other systems (Grönke et al., 2010; Jin Chan & Steiner, 2000). Because of their divergent sequences, classification of ILPs in invertebrates generally reflects the structural position of the conserved cysteine residues that form disulfide bridges both between and within the two chains. We were able to identity conserved A and B chain cysteine motifs and adapted this phylogenetic framework to propose a classification system for the cnILPs. All of the identified cnILPs from group A appear to be of the γ type ILP structure as α and β type ILP structures possess an additional third disulfide bond between the A and B chains and no additional cysteines were conserved in the A chain of the cnILPs to allow a third inter-chain disulfide bond (Cherif et al., 2019; Zheng et al., 2018). We adapted a classification framework to group the cnILPs in groups from A-C according to the A and B chain motifs where group A had

the three canonical ILP disulfide bonds and groups B and C had an additional intra-B chain disulfide bond. The cnILPs within these four groups were then classified further according to their sequence similarity to facilitate comparison in downstream analyses and recognize the sequence diversity within these four groups. This multi-level classification system ultimately resulted in a total of ten putative, distinct cnILP proteins across the phylum.

The taxonomic distribution of the cnILPs suggests that there was an expansion and diversification of cnILPs within the class Anthozoa. Specifically, this diversification likely occurred after the split between the sub-classes Hexacorallia and Octocorallia as the Octocorallian order Alcyonacea only possessed two cnILPs in contrast to the eight and seven cnILPs found in the Hexacorallian orders Scleractinia and Actiniaria, respectively. Genomic mapping of each of these eight cnILPs to individual loci confirms that they are distinct genes rather than isoforms and the result of alternative splicing (Figure 2-3). Further, the four Scleractinian cnILPs from groups A and B with the three canonical disulfide bonds are found on the same scaffold in the exact same orientation within three separate *Acropora* spp. genomes, similarly to *C. elegans* where ILPs are found in sets of 2-7 ILPs from the same type. The level of ILP diversity found in corals specifically is not unusually high when compared to other invertebrates as *Drosophila* spp. possess eight ILPs and *C. elegans* has 40 ILPs (Grönke et al., 2010; Zheng et al., 2018). However, these patterns in corals are the first record of such extensive ILP diversity in a non-bilaterian as besides the three cnILPs previously identified in *H. vulgaris*, no other non-bilaterian species has been shown to possess more than one ILP. Significantly, the stark contrast between the level of cnILP diversity found in

Hexacorallia compared to the rest of the phylum begs the question as to whether this similarly reflects their level of functional diversification in Hexacorallian species.

The unusual addition of an intra-B chain disulfide bond in the group B and C cnILPs is not the first instance of this kind of ILP structure in animals. The insulin-like androgrenic gland hormone of crustaceans (AGH) that controls sexual differentiation in males has a B chain cysteine structure ($CX_8CX_1CX_{14}C$) similar to that of group C and D cnILPs (Manor et al., 2007; Ventura et al., 2009). The group B and C cnILP B chain motifs are also similar to *Conus floridulus* con-Ins F2c (Safavi-Hemami et al., 2015) that was identified in cone snail venom. When injected into their prey, the venom ILP overwhelms the prey's insulin receptors leading to hypoglycemic shock. The increasing awareness of diverse B and A chain motifs is even expanding in *Homo sapiens* through the discovery of the human insulin-like growth factor-like (ILGFL) proteins (Emtage et al., 2006). These secreted proteins have a very unusual motif ($CX_5CX_{10}CCX_{13}CX_{2–3}CX_4CX_3CCX_{12–22}CX_{8–10}C$) but their predicted structure, sequence, and gene expression patterns are similar to the *H. sapiens* IGF family leading to their classification within the insulin hormone superfamily.

Within the diverse cnILP repertoire found in Scleractinian corals, a seemingly high level of sequence conservation can be found on a generic level. Within the fourteen species from the genus *Acropora* that was included in our study, for example, there is an amazing overall percent identity ranging from 80.4% to 91.9% for the different cnILPs (Table 2-3), which stands in stark contrast to the 40% to 76% found in the genus *Drosophila* (Grönke et al., 2010). One potential explanation for this variability in sequence

divergence revolves around the hypothesis that coral genera have experienced reticulate evolution over time (Veron, 1995). Reticulate evolution suggests that species within a genus are not the result of the diversification from a single ancestral species, but rather through consecutive hybridization and separation events of species that ultimately resulted in many interconnected lineages within a genus. This hybridization is feasible within the genus *Acropora* through the synchronous "mass spawning" events that are common to the genus of corals. Particularly prevalent in and important to Indo-Pacfic *Acropora* spp., mass spawning events or the near simultaneous release of eggs and sperm from closely related coral species can facilitate the fertilization between their eggs and sperm, resulting in hybrids (Baird, Guest, & Willis, 2009; Willis, van Oppen, Miller, Vollmer, & Ayre, 2006). Hybridization might facilitate a higher level of sequence conservation within the *Acropora* genus compared to other genera that contain species reproductively isolated from their counterparts because of the constant mixing of genetic material over evolutionary time (Veron, 1995). Genetic mixing is not feasible between Caribbean and Indo-Pacific species within the *Acropora* genus, implying there must be an alternate reason for the level of sequence conservation between species from these two regions. As the two Caribbean species *A. palmata* and *A. cervicornis* are nearly identical to their Pacific counterparts, it is unlikely that reticulate evolution is the main causative factor behind the high level of cnILP sequence conservation within this genus. The last potential connection between the Caribbean and Pacific species from this genus were lost either during the closing of the Isthmus of Panama 3.0–3.5 MYA (Coates & Obando, 1996; Duque-Caro, 1990; Keigwin, 1982) or through the Tethys Sea via the Mediterranean during the mid-Miocene ~12 MYA (Steininger & Rögl, 1984). However,

studies in other taxa suggest that genetic divergence between Caribbean and Pacific lineages in marine species may have occurred before the closing of the Isthmus of Panama because of changing oceanographic conditions between 12.9 and 7.0 Mya (Knowlton, Weigt, Solorzano, Mills, & Bermingham, 1993). As phylogenetic analysis and divergence estimates place the splitting of the subgenera *Sophophora* and *Drosophila* within the *Drosophila* genus around 40 MYA (Russo, Takezaki, & Nei, 1995), much earlier than the potential genetic separation within the *Acropora* genus, perhaps the conservation level might be due to . Overall, this contradictory level of ILP divergence in different genera suggests that the more recent reproductive isolation of the *Acropora* genus combined with their reticulate evolution might be a significant factor in their high level of sequence level conservation.

2.6 Conclusion

While invertebrates are known for having large numbers of ILPs, only seven ILPs have been characterized in non-bilaterian species including the hydroid *H. vulgaris* and the anemone *N. vectensis*. We discovered a large repertoire of ILPs as evidenced by the 360 putative cnILPs found across 93 cnidarian species. Notably, Hexacorallians including the reef-building Scleractinian corals followed by anemones from the order Actiniaria displayed the most extensive cnILP diversity and show signs for duplication and diversification in their cnILP sequences. As functional studies on ILPs have only been conducted in *H. vulgaris*, functional studies within Hexacorallians are needed to assess the functional implications of the extensive cnILP diversity found Hexacorallians. These findings implicate corals as having the most extensive cnILP diversity found outside of

bilaterians and begs the question of how they might be involved in metabolism, immunity, stress response, longevity, and the endosymbioses of reef-building corals.

Class	Orders	# species
Anthozoa	Actiniaria	15
	Alcyonacea	30
	Scleractinia	5
Cubozoa	Carybdeida	4
	Chirodropida	1
Hydrozoa	Anthoathecata	5
	Siphonophorae	20
Scyphozoa	Rhizostomeae	4
	Semaeostomeae	4
Staurozoa	Stauromedusae	5

Table 2-1. Taxonomic breakdown of the cnidarian species included in this study to investigate the diversity in cnidarian insulin-like peptide sequences and structures.

Class	Order		B chain motif	A chain motif
Structure Type A consensus:			CX_nC	CCX_nCX_nC
A1	Cubozoa	Carybdeida	$CX_{11}C$	CCX_3CX_9C
		Chirodropida		
	Scyphozoa	Rhizostomeae		
		Semaeostomeae		
	Hydrozoa	Anthoathecata		
		Siphonophorae		
		Scleractinia		
A2a	Anthozoa	Actiniaria	$CX_{13/14}C$	CCX_3CX_8C
		Scleractinia		$CCX_{3/4}CX_9C$
A2b	Anthozoa	Actiniaria	$CX_{13}C$	$CCX_{3/5}CX_9C$
		Scleractinia	$CX_{13/14}C$	CCX_3CX_8C
A2	Anthozoa	Alcyonacea	$CX_{13}C$	CCX_3CX_8C
	Cubozoa	Carybdeida	$CX_{12}C$	
		Chirodropida		CCX_4CX_8C
	Scyphozoa	Rhizostomeae		
		Semaeostomeae		
	Hydrozoa	Anthoathecata		CCX_6CX_8C
		Siphonophorae		$CCX_{5-8}CX_8C$
A3	Anthozoa	Actiniaria	$CX_{11}C$	CCX_3CX_8C
		Scleractinia		

Structure Type B consensus:			Cx_nCCx_nC	CCx_nCx_nC
B	Anthozoa	Actiniaria	$CX_{10}CCX_5C$	CCX_3CX_9C
		Alcyonacea	$CX_{10}CCX_{3/6}C$	
		Scleractinia	$CX_{10}CCX_3C$	
	Cubozoa	Carybdeida	$CX_{10}CCX_{3/5}C$	$CCX_3CX_{8\text{-}9}C$
		Chirodropida	$CX_{10}CCX_{3/5}C$	
	Scyphozoa	Rhizostomeae	$CX_{10}CCX_{3/5}C$	CCX_3CX_9C
		Semaeostomeae	$CX_{10}CCX_3C$	
	Staurozoa	Stauromedusae	$CX_{11}CCX_4C$	
Structure Type C consensus:			$Cx_nCx_nCx_nC$	CCx_nCx_nC
C1	Anthozoa	Scleractinia	$CX_{12}CX_3CX_{11}C$	CCX_3CX_8C
C2			$CX_{11/14/15}CX_{3/6}CX_{11}C$	
C3			$CX_{2/5}CX_{14/15/16}CX_{11}C$	
C4			$CX_{2/5}CX_{14/15/16}CX_{11}C$	
C5		Actiniaria	$CX_5CX_{15}CX_{11}C$	
C6			$CX_{4\text{-}6}CX_{15/16}CX_{11}C$	

Table 2-2. Breakdown of the classification of the putative cnILPs identified in the phylogenetic analysis, their B and A chain cysteine motifs, and their taxonomic distribution in the phylum

	Signal peptide	B chain	C peptide	A chain	Overall
cnILP-A2a (n=11)	80.6%	85.0%	86.4%	93.1%	88.2%
cnILP-A2b (n=9)	83.3%	82.1%	68.8%	86.8%	80.4%
cnILP-A3 (n=10)	66.7%	78.9%	88.9%	92.9%	82.8%
cnILP-B (n=12)	76.7%	90.8%	80.0%	96.2%	85.8%
cnILP-C1 (n=12)	96.2%	86.9%	96.6%	86.4%	91.9%
cnILP-C2 (n=7)	78.3%	92.1%	80.0%	91.7%	86.5%
cnILP-C3 (n=9)	82.6%	77.8%	85.0%	92.3%	84.1%
cnILP-C4 (n=11)	78.6%	85.4%	90.9%	93.1%	86.4%

Table 2-3. Percent sequence identity for the signal peptide, B chain, C peptide, and A chain regions as well as the overall cnILP in *Acropora spp.*

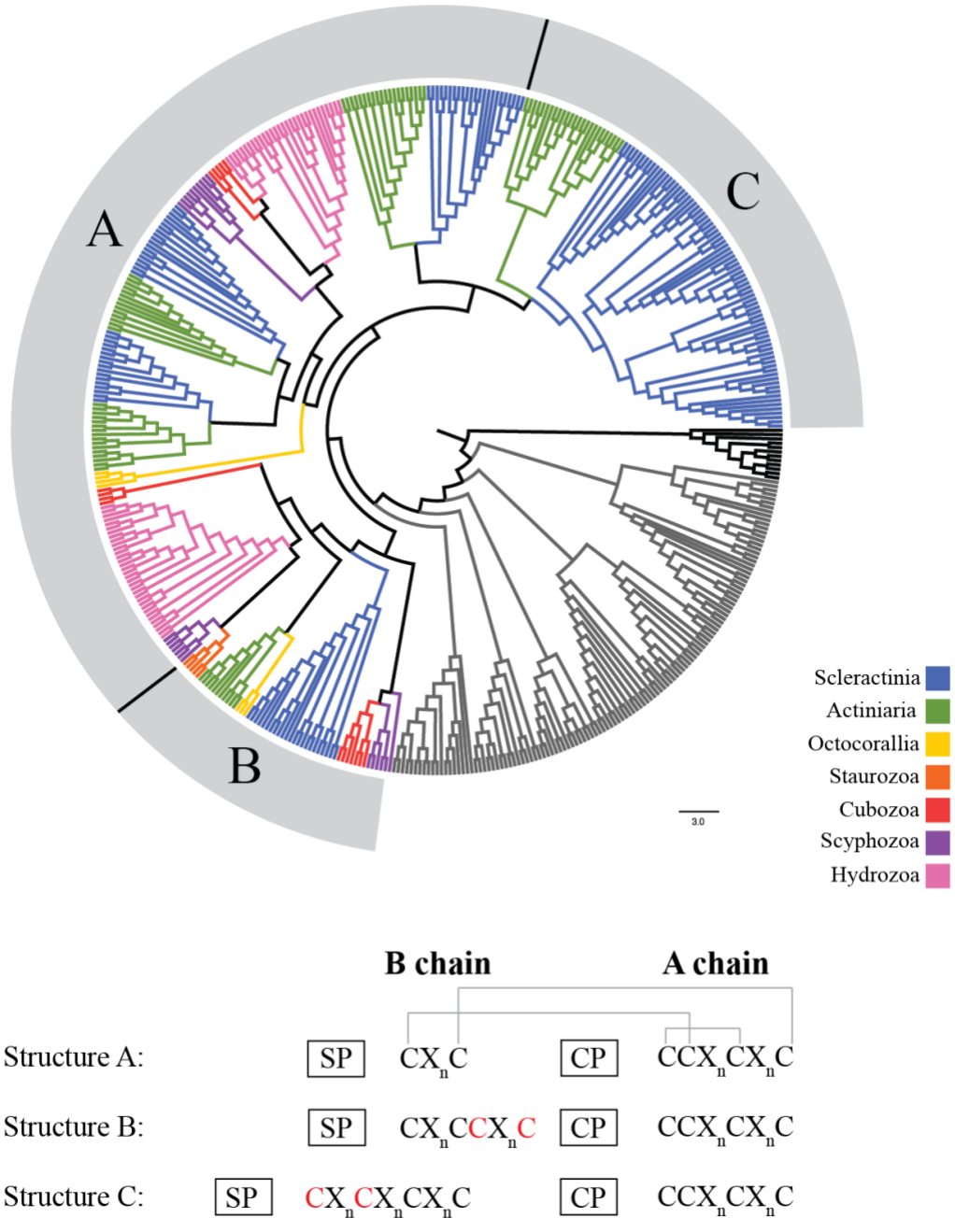


Figure 2-1. Phylogeny of the cnidarian insulin-like peptides showing the distribution of the three types of ILP structures throughout different taxonomic groups within Cnidaria. The disulfide bonds linking the conserved cysteine residues between the two chains are indicated by the black lines and the additional two cysteines found in the type B and C cnILP structures are colored red.

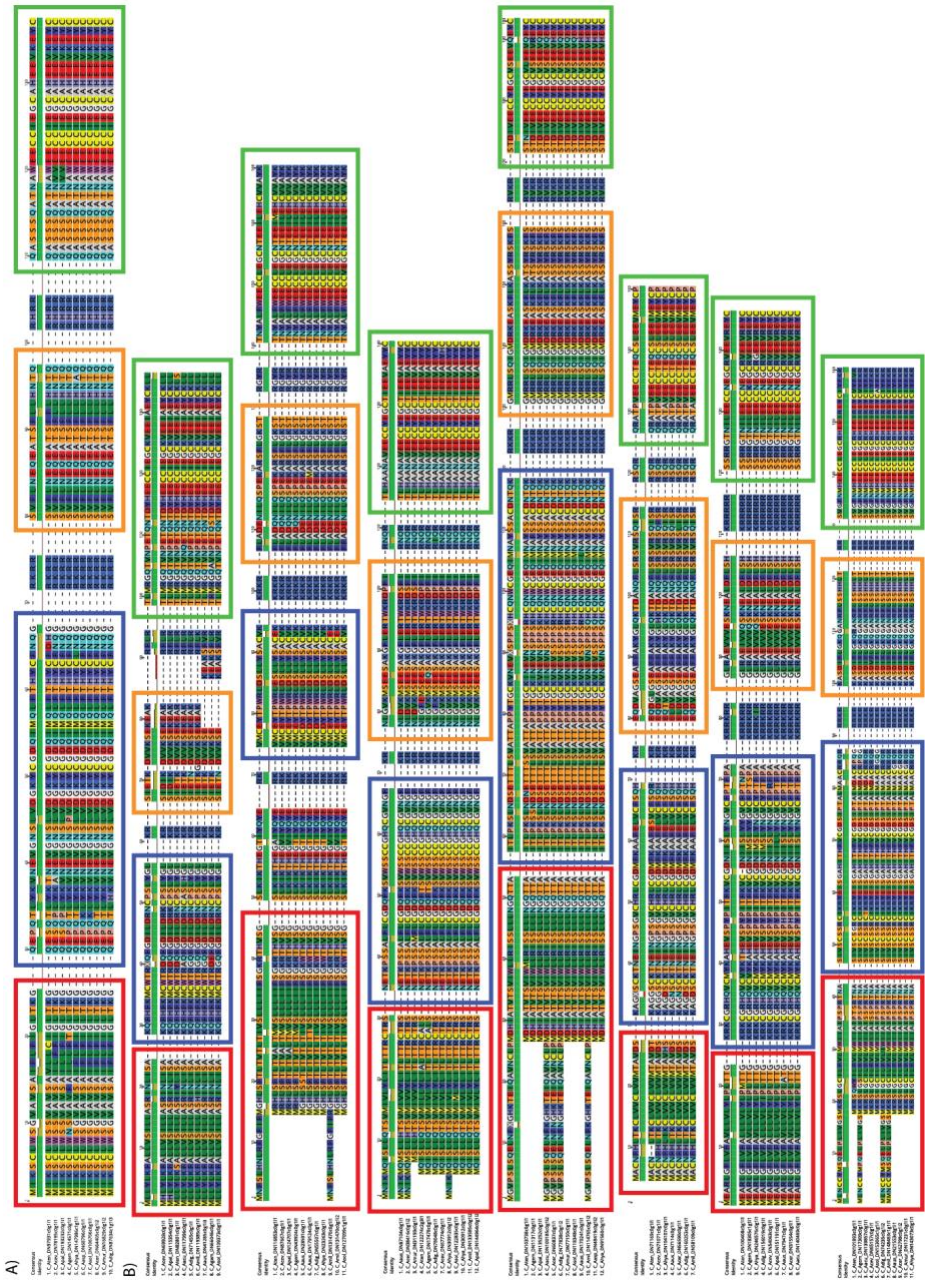


Figure 2-2. Multiple sequence alignment of the *Acropora spp.* cnILPs (top to bottom) cnILP-A3, cnILP-A2b, cnILP-A2a, cnILP-B, cnILP-C1, cnILP-C2, cnILP-C3, cnILP-C4 from the genus Acropora with boxes denoting the different domains of the signal peptide (red), B chain (blue), C peptide (orange), and A chain (green).

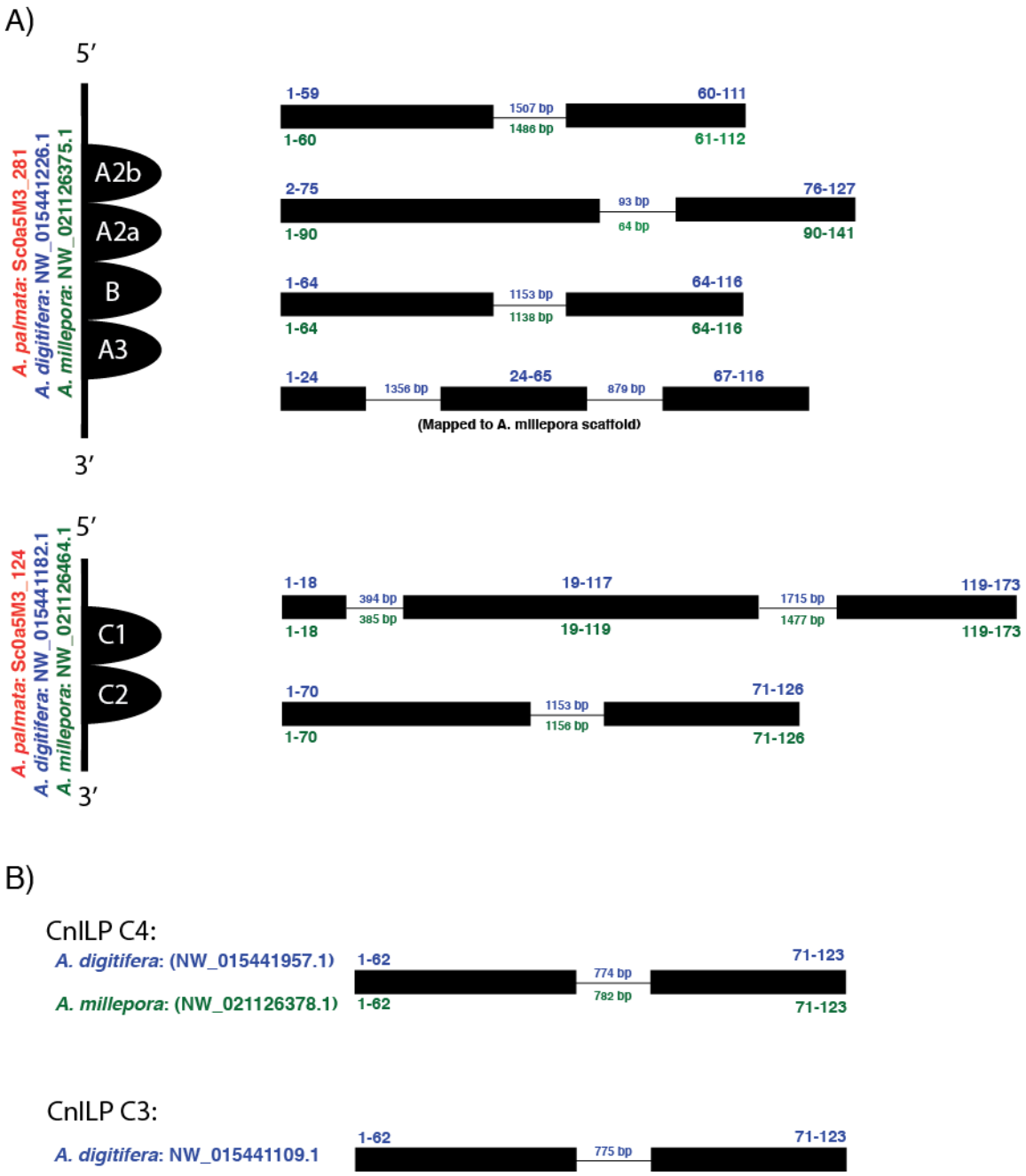


Figure 2-3. Mapping of the cnILPs from the species A. digitifera, A. millepora, and A. palmata to their respective genomes.

2.7 References

Anctil, M. (2009). Chemical transmission in the sea anemone *Nematostella vectensis*: A genomic perspective. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 4*(4), 268-289. doi:10.1016/j.cbd.2009.07.001

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. In: Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Anisimova, M., Gil, M., Dufayard, J.-F., Dessimoz, C., & Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic biology, 60*(5), 685-699. doi:10.1093/sysbio/syr041

Aslam, A. F., Kiya, T., Mita, K., & Iwami, M. (2011). Identification of novel bombyxin genes from the genome of the silkmoth *Bombyx mori* and analysis of their expression. *Zoological science, 28*(8), 609-616. doi:10.2108/zsj.28.609

Baird, A. H., Guest, J. R., & Willis, B. L. (2009). Systematic and biogeographical patterns in the reproductive biology of scleractinian corals. *Annual Review of Ecology, Evolution, and Systematics, 40*, 551-571. doi:10.1146/annurev.ecolsys.110308.120220

Barbieri, M., Bonafè, M., Franceschi, C., & Paolisso, G. (2003). Insulin/IGF-I-signaling pathway: an evolutionarily conserved mechanism of longevity from yeast to humans. *American Journal of Physiology-Endocrinology And Metabolism, 285*(5), E1064-E1071.

Baumeister, R., Schaffitzel, E., & Hertweck, M. (2006). Endocrine signaling in *Caenorhabditis elegans* controls stress response and longevity. *Journal of endocrinology, 190*(2), 191-202. doi:10.1677/joe.1.06856

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Cherif, M., Berthelin, C. H., Adeline, B., Rivière, G., Favrel, P., & Kellner, K. (2019). Molecular evolution and functional characterisation of insulin related peptides in

molluscs: Contributions of *Crassostrea gigas* genomic and transcriptomic-wide screening. *General and Comparative Endocrinology, 271*, 15-29.

Coates, A., & Obando, J. (1996). The geological evolution of the Central American Isthmus. *Science Advances, 2*(8), e1600883. doi:10.1126/sciadv.1600883

Collins, A. G. (2009). Recent insights into cnidarian phylogeny. *Smithsonian contributions to the marine sciences*(38).

Duckert, P., Brunak, S., & Blom, N. (2004). Prediction of proprotein convertase cleavage sites. *Protein Engineering Design and Selection, 17*(1), 107-112. doi:10.1093/protein/gzh013

Duque-Caro, H. (1990). Neogene stratigraphy, paleoceanography and paleobiogeography in northwest South America and the evolution of the Panama Seaway. *Palaeogeography, Palaeoclimatology, Palaeoecology, 77*(3-4), 203-234. doi:10.1016/0031-0182(90)90178-A

Emtage, P., Vatta, P., Arterburn, M., Muller, M. W., Park, E., Boyle, B., . . . Tang, Y. T. (2006). IGFL: A secreted family with conserved cysteine residues and similarities to the IGF superfamily. *Genomics, 88*(4), 513-520. doi:10.1016/j.ygeno.2006.05.012

González, J. P., Farnés, O. C., Vasconcelos, A. T. R., & Pérez, A. G. (2009). Conservation of key members in the course of the evolution of the insulin signaling pathway. *Biosystems, 95*(1), 7-16. doi:10.1016/j.biosystems.2008.06.003

Grönke, S., Clarke, D.-F., Broughton, S., Andrews, T. D., & Partridge, L. (2010). Molecular evolution and functional characterization of *Drosophila* insulin-like peptides. *PLoS Genet, 6*(2), e1000857.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology, 59*(3), 307-321. doi:10.1093/sysbio/syq010

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., . . . Lieber, M. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols, 8*(8), 1494-1512. doi:10.1038/nprot.2013.084

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Molecular biology and evolution, 35*(2), 518-522. doi:10.1093/molbev/msx281

Iwami, M. (2000). Bombyxin: an insect brain peptide that belongs to the insulin family. *Zoological science, 17*(8), 1035-1044.

Jia, K., Chen, D., & Riddle, D. L. (2004). The TOR pathway interacts with the insulin signaling pathway to regulate *C. elegans* larval development, metabolism and life span. *Development, 131*(16), 3897-3906. doi:10.1242/dev.01255

Jin Chan, S., & Steiner, D. F. (2000). Insulin through the ages: phylogeny of a growth promoting and metabolic regulatory hormone. *American Zoologist, 40*(2), 213-222.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods, 14*(6), 587-589. doi:10.1038/nmeth.4285

Katoh, K., Kuma, K.-i., Toh, H., & Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic acids research, 33*(2), 511-518. doi:10.1093/nar/gki198

Keigwin, L. (1982). Isotopic paleoceanography of the Caribbean and East Pacific: role of Panama uplift in late Neogene time. *science, 217*(4557), 350-353. doi:10.1126/science.217.4557.350

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology, 37*(8), 907-915. doi:10.1038/s41587-019-0201-4

Knowlton, N., Weigt, L. A., Solorzano, L. A., Mills, D. K., & Bermingham, E. (1993). Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the Isthmus of Panama. *science, 260*(5114), 1629-1632. doi:10.1126/science.8503007

Lagueux, M., Lwoff, L., Meister, M., Goltzene, F., & Hoffmann, J. A. (1990). cDNAs from neurosecretory cells of brains of *Locusta migratoria* (Insecta, Orthoptera) encoding a novel member of the superfamily of insulins. *European journal of biochemistry, 187*(1), 249-254. doi:10.1111/j.1432-1033.1990.tb15302.x

Le Roith, D., Shiloach, J., Roth, J., & Lesniak, M. A. (1980). Evolutionary origins of vertebrate hormones: substances similar to mammalian insulins are native to unicellular eukaryotes. *Proceedings of the National Academy of Sciences, 77*(10), 6184-6188.

Manor, R., Weil, S., Oren, S., Glazer, L., Aflalo, E. D., Ventura, T., . . . Sagi, A. (2007). Insulin and gender: an insulin-like gene expressed exclusively in the androgenic gland of the male crayfish. *General and Comparative Endocrinology, 150*(2), 326-336. doi:10.1016/j.ygcen.2006.09.006

Mizoguchi, A., & Okamoto, N. (2013). Insulin-like and IGF-like peptides in the silkmoth Bombyx mori: discovery, structure, secretion, and function. *Frontiers in physiology, 4*, 217.

Mortzfeld, B. M., Taubenheim, J., Klimovich, A. V., Fraune, S., Rosenstiel, P., & Bosch, T. C. (2019). Temperature and insulin signaling regulate body size in *Hydra* by the Wnt and TGF-beta pathways. *Nature communications, 10*(1), 1-13. doi:10.1038/s41467-019-11136-6

Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution, 32*(1), 268-274. doi:10.1093/molbev/msu300

Pertseva, M., & Shpakov, A. (2002). Conservatism of the insulin signaling system in evolution of invertebrate and vertebrate animals. *Journal of Evolutionary Biochemistry and Physiology, 38*(5), 547-561.

Petersen, T. N., Brunak, S., Von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods, 8*(10), 785-786. doi:10.1038/nmeth.1701

Pierce, S. B., Costa, M., Wisotzkey, R., Devadhar, S., Homburger, S. A., Buchman, A. R., . . . Pasquinelli, A. A. (2001). Regulation of DAF-2 receptor signaling by human insulin and ins-1, a member of the unusually large and diverse *C. elegans* insulin gene family. *Genes & development, 15*(6), 672-686. doi:10.1101/gad.867301

Putnam, N. H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., . . . Kapitonov, V. V. (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *science, 317*(5834), 86-94. doi:10.1126/science.1139158

Remmert, M., Biegert, A., Hauser, A., & Söding, J. (2012). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature methods, 9*(2), 173-175. doi:10.1038/nmeth.1818

Robitzki, A., Schröder, H., Ugarkovic, D., Pfeifer, K., Uhlenbruck, G., & Müller, W. (1989). Demonstration of an endocrine signaling circuit for insulin in the sponge *Geodia cydonium*. *The EMBO journal, 8*(10), 2905-2909.

Rosic, N., Kaniewska, P., Chan, C.-K. K., Ling, E. Y. S., Edwards, D., Dove, S., & Hoegh-Guldberg, O. (2014). Early transcriptional changes in the reef-building coral *Acropora aspera* in response to thermal and nutrient stress. *BMC Genomics, 15*(1), 1052.

Russo, C., Takezaki, N., & Nei, M. (1995). Molecular phylogeny and divergence times of drosophilid species. *Molecular biology and evolution, 12*(3), 391-404. doi:10.1093/oxfordjournals.molbev.a040214

Rutter, G. A., Pullen, T. J., Hodson, D. J., & Martinez-Sanchez, A. (2015). Pancreatic β-cell identity, glucose sensing and the control of insulin secretion. *Biochemical Journal, 466*(2), 203-218.

Safavi-Hemami, H., Gajewiak, J., Karanth, S., Robinson, S. D., Ueberheide, B., Douglass, A. D., . . . Bandyopadhyay, P. K. (2015). Specialized insulin is used for chemical warfare by fish-hunting cone snails. *Proceedings of the National Academy of Sciences, 112*(6), 1743-1748.

Saltiel, A. R., & Kahn, C. R. (2001). Insulin signalling and the regulation of glucose and lipid metabolism. *Nature, 414*(6865), 799-806.

Satake, S. I., Masumura, M., Ishizaki, H., Nagata, K., Kataoka, H., Suzuki, A., & Mizoguchi, A. (1997). Bombyxin, an insulin-related peptide of insects, reduces the major storage carbohydrates in the silkworm *Bombyx mori*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 118*(2), 349-357.

Sebestyén, F., Póliska, S., Rácz, R., Bereczki, J., Lénárt, K., Barta, Z., . . . Tökölyi, J. (2017). Insulin/IGF Signaling and life history traits in response to food availability and perceived density in the cnidarian *Hydra vulgaris*. *Zoological science, 34*(4), 318-325.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics, 31*(19), 3210-3212.

Smit, A., Spijker, S., Van Minnen, J., Burke, J., De Winter, F., Van Elk, R., & Geraerts, W. (1996). Expression and characterization of molluscan insulin-related peptide VII from the mollusc *Lymnaea stagnalis*. *Neuroscience, 70*(2), 589-596.

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M., & Kelly, S. (2016). TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome research, 26*(8), 1134-1144.

Steele, R., Lieu, P., Mai, N. H., Shenk, M. A., & Sarras Jr, M. P. (1996). Response to insulin and the expression pattern of a gene encoding an insulin receptor homologue suggest a role for an insulin-like molecule in regulating growth and patterning in *Hydra*. *Development genes and evolution, 206*(4), 247-259. doi:10.1007/s004270050050

Steininger, F. F., & Rögl, F. (1984). Paleogeography and palinspastic reconstruction of the Neogene of the Mediterranean and Paratethys. *Geological Society, London, Special Publications, 17*(1), 659-668.

Taniguchi, C. M., Emanuelli, B., & Kahn, C. R. (2006). Critical nodes in signalling pathways: insights into insulin action. *Nature reviews Molecular cell biology, 7*(2), 85-96.

Ventura, T., Manor, R., Aflalo, E. D., Weil, S., Raviv, S., Glazer, L., & Sagi, A. (2009). Temporal silencing of an androgenic gland-specific insulin-like gene affecting phenotypical gender differences and spermatogenesis. *Endocrinology, 150*(3), 1278-1286.

Veron, J. E. N. (1995). *Corals in space and time: the biogeography and evolution of the Scleractinia*: Cornell University Press.

Willis, B. L., van Oppen, M. J., Miller, D. J., Vollmer, S. V., & Ayre, D. J. (2006). The role of hybridization in the evolution of reef corals. *Annu. Rev. Ecol. Evol. Syst., 37*, 489-517.

Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome biology, 15*(3), 1-12.

Wu, Q., & Brown, M. R. (2006). Signaling and function of insulin-like peptides in insects. *Annu. Rev. Entomol., 51*, 1-24.

Zheng, S., Chiu, H., Boudreau, J., Papanicolaou, T., Bendena, W., & Chin-Sang, I. (2018). A functional study of all 40 *Caenorhabditis elegans* insulin-like peptides. *Journal of Biological Chemistry, 293*(43), 16912-16922. doi:10.1074/jbc.RA118.004542

Zhou, X., Shen, X.-X., Hittinger, C. T., & Rokas, A. (2018). Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic data sets. *Molecular biology and evolution, 35*(2), 486-503.

Chapter 3: Transcriptomic analysis of the cnidarian insulin-like peptides implicates them in symbiosis and stress response

3.1 Abstract

The pleiotropic roles of the insulin-like signaling pathway are conserved throughout metazoans where its maintains a central role in controlling processes such as growth, reproduction, development, immunity, stress response, and longevity. Studies in model organisms have implicated the insulin-like signaling pathway as being susceptible to manipulation by microorganisms, both mutualists and parasites, to evade the immune system through changes in metabolic pathways. We characterized the response of the cnidarian insulin-like peptides (cnILPs) in response to symbiosis, heat stress, and ocean acidification stress to understand factors that influence cnILP gene expression and hypothesize on its function in cnidarians. We identified conserved patterns of cnILP expression that include the down-regulation of cnILPs, particularly from the cnILP-B group, in response to stable symbiosis suggesting that insulin signaling might function in symbiotic communication between corals and algae. We also identified an up-regulation of cnILP expression in response to thermal stress, but not ocean acidification stress, suggesting that insulin pathway activity is central to thermal stress responses but not necessarily all environmental changes. In particular, patterns in cnILP gene expression suggest it may underlie the variable physiological responses between thermally susceptible and resilient populations. Our work begins to highlight the potential functional importance of insulin signaling within the coral animal and hypothesize potential functions of the insulin pathway that should be investigated in future studies.

3.2 Introduction

Insulin/insulin-like peptides (ILPs) and the insulin/insulin-like peptide signaling pathway (IIS) is an evolutionarily conserved pathway that regulates metabolism, growth, stress resistance, development, reproduction, aging, and life span in both vertebrates and invertebrates (Barbieri, Bonafè, Franceschi, & Paolisso, 2003; Colwell, 2012; Saltiel & Kahn, 2001; Taniguchi, Emanuelli, & Kahn, 2006). The pleiotropic functions of the IIS equip it with the necessary capabilities to integrate environmental signals and respond to stresses such as starvation, thermal stress, oxidative stress, and infection, thereby contributing to phenotypic plasticity (Taguchi & White, 2008). From *Caenorhabditis elegans* to humans, decreased insulin signaling and increased insulin sensitivity have been tied to increased stress resistance, disease resistance, and longevity (Giovannucci, 2002; Kimura, Tissenbaum, Liu, & Ruvkun, 1997; Piper, Selman, McElwee, & Partridge, 2008; Taguchi & White, 2008).

The signals governing the transcription and release of insulin-like peptides are as diverse as the physiological functions they influence and include nutritional inputs, hormones, neurotransmitters, and transcription factors. Most famously, glucose (Kréneisz, Chen, Fridell, & Mulkey, 2010) and other sugars (Miyamoto, Slone, Song, & Amrein, 2012) can initiate ILP transcription along with certain amino acids (Colombani et al., 2003; Géminard, Rulifson, & Léopold, 2009). Glutamate may also be a second messenger of sorts that can enhance the release of insulin, although its importance to insulin secretion is still under debate. Other nutritional inputs for ILP transcription and secretion include incretin hormones (Holst, 2007; Scheen, Castillo, & Lefèbvre, 1996) and

neurotransmitters like acetycholine (Gilon & Henquin, 2001; Rodriguez-Diaz et al., 2011) and GABA/γ-aminobutyric acid (Braun et al., 2010). Reactive oxygen species at low concentrations seem to be required for glucose-stimulated ILP production (Leloup et al., 2009; Pi et al., 2007) although they can become repressive on ILP transcription at high levels during oxidative stress. Similarly, elevated levels of other nutritional signals like glucose (Dai et al., 2016; Gleason, Gonzalez, Harmon, & Paul Robertson, 2000) and fatty acids, particularly palmitate and palmitoleate (Hagman, Hays, Parazzoli, & Poitout, 2005; Kelpe et al., 2003; Ritz-Laser et al., 1999), can suppress ILP transcription.

A growing body of literature indicates that these various signals can be an avenue through which other organisms can manipulate the reactions of host immune systems to infection via IIS-dependent signaling. Foreign organisms' control of the immune response by IIS could be accomplished through alterating the host's carbohydrate and lipid metabolism, as many studies have demonstrated the interplay between metabolism and immunity. In *Drosophila*, where the IIS has been extensively studied, changes to metabolism can alter the inflammatory response to infection by foreign pathogens (Chambers, Song, & Schneider, 2012; Hang et al., 2014; Liu et al., 2012). On one hand, insulin signaling has been directly implicated in the reallocation of energy sources to fuel the immune response to infection (DiAngelo, Bland, Bambina, Cherry, & Birnbaum, 2009). However, invading organisms might also directly manipulate this link between metabolism and immunity to facilitate their persistence and survival within the host (Hang et al., 2014).

Manipulation of a host's IIS by a foreign organism to facilitate its persistence is not only seen in pathogenesis, but also in symbiosis. While the mechanisms and evolutionary significance of this immune manipulation via metabolism yet to be completely elucidated, growing evidence suggests that the influencing the signaling connecting IIS and immunity can be advantageous to symbionts (Demas, Adamo, & French, 2011; DiAngelo et al., 2009). Endosymbionts from the bacterial genus *Wolbachia* manipulate various arthropods in this manner including the fruit fly *Drosophila* (Grönke, Clarke, Broughton, Andrews, & Partridge, 2010; Ikeya, Broughton, Alic, Grandison, & Partridge, 2009) and crustaceans (Huang, Ye, Huang, Yang, & Gong, 2014; Rosen et al., 2010). This type of manipulation of host immunity is also seen in the protozoan *Plasmodium falciparum*, the human malaria parasite, as it represses mosquito immune responses by directly inducing the synthesis of two insect ILPs (Pietri, Pietri, Potts, Riehle, & Luckhart, 2015), altering intermediary metabolism (Marquez et al., 2011; Pietri et al., 2015), and suppressing NF-κB-dependent mosquito immune defenses (Pietri et al., 2015). Parallel to this *in vitro* work in model organisms, transcriptomic investigation into the onset of symbiosis between the salamander *Ambystoma maculatum* and the green alga *Oophila amblystomatis* have also implicated altered IIS sensitivity combined with repression of NF-κB-dependent immunity (Burns, Zhang, Hill, Kim, & Kerney, 2017).

Manipulation of host immunity via metabolism might also be present in corals where they participate in an ecologically important endosymbiosis with the algal family Symbiodiniaceae that depends on the nutritional transfer of glucose and other sugars, lipids, amino acids, and reactive oxygen species between the two partners. The parallels between the coral-algal symbiosis and insect symbioses where metabolism is manipulated

to repress host immunity in favor of symbiont persistence are striking, especially in context of the coral's dependence on receiving photosynthetically produced glucose from the algae. A recent transcriptomic investigation on the establishment of symbiosis between the coral *Acropora tenuis* and its algal endosymbionts revealed gene ontology enrichment for the categories insulin-like growth factor signaling pathway among down-regulated transcripts (Yuyama & Higuchi, 2014). Studies directly investigating the functional responses of the cnidarian IIS or ILPs are sparse, with only one study that investigated the effect of food level of population density simply finding increased expression of the insulin-like receptor under increased population density but no effect of either factor on ILP expression. Taking into account scleractinian corals possess the most diverse cnILP, their dependence on the nutritional transfer between the two partners, and the ecological importance of this symbiosis, this advocates corals as a prime target for investigating the function of IIS in endosymbiosis. In this work we aim to begin understanding the potential functional responses of cnILPs by examining and summarizing their transcriptomic responses to the onset and/or presence of symbiosis, taxonomic identity of algal endosymbionts, temperature stress, and ocean acidification across many species.

3.3 Methods

3.3.1 Identification of SRA Bioprojects for analysis

We searched the NCBI Short Read Archive (SRA) for SRA Bioprojects that examined transcriptomic changes in response to symbiosis, temperature stress, and ocean acidification stress in cnidarians. We identified six bioprojects to investigate symbiotic

cnILP gene expression including PRJNA309168 (A. R. Mohamed et al., 2016), PRJNA398338 (Amin R. Mohamed et al., 2018), PRJDB4715 (Yuyama & Higuchi, 2014), PRJEB21012 (unpublished), PRJNA385711 (Meron et al., 2019), and PRJDB7145 (Ishii et al., 2019). Both PRJNA385711 and PRJDB7145 were also used to examine the effects of heat stress on cnILP expression along with the additional bioprojects PRJNA308355 (Zhou et al., 2017), PRJNA177515 (Barshis et al., 2013), and PRJNA340085 (Hou et al., 2018). Bioprojects related to ocean acidification included PRJNA380146 (Yuan et al., 2018), PRJNA149513 (Moya et al., 2012), PRJNA260269 (Moya et al., 2015), and PRJEB21531 (González-Pech, Vargas, Francis, & Wörheide, 2017).

3.3.2 Preparation of transcriptomes for gene expression analysis

Gene-coding transcripts for each of the species' previously assembled reference transcriptomes were identified using TransDecoder (Haas & Papanicolaou, 2017) under default settings. Transcripts were then clustered at 99% identity using CD-HIT-EST (Fu, Niu, Zhu, Wu, & Li, 2012) to remove duplicate sequences. The FASTQ files for each of the bioprojects were assessed and then trimmed for quality using FastQC (Andrews, 2010) and Trimmomatic (Bolger, Lohse, & Usadel, 2014), respectively. Quality trimmed reads were then mapped back to the clustered transcriptome using Salmon (Roberts & Pachter, 2013) with either 19 bp, 23 bp, 27bp, or 31 bp kmer sizes for the salmon index depending on the bioproject's read length. Mapping output was then used for downstream gene expression analyses.

3.3.3 Differential gene expression analysis

Differential gene expression analysis was conducted in R using the edgeR package (Robinson, McCarthy, & Smyth, 2010). Custom scripts were constructed for each individual bioproject according its experimental set-up. Low read filtering was standardized across the datasets using a threshold that required 0.2 counts per million in at least two samples followed by the standard edgeR normalization. Dispersion was modeled using the estimateGLMRobustDisp function within edgeR, which uses estimated observation weights for reads to perform a robust regression to the model. A likelihood ratio test was used to determine differentially expressed genes using a threshold of FDR < 0.1 and a log2FC of at least 0.5 in magnitude.

3.4 Results

3.4.1 Effect of symbiont presence and type on cnILP gene expression

To investigate the effect of symbiosis on coral cnILP gene expression, we re-examined two transcriptomic studies from the species *Acropora tenuis* (Yuyama et al. 2018) and *Euphyllia paradivisa* (Meron et al., 2019). Yuyama and Higuchi (2014) infected aposymbiotic *A. tenuis* with one of two Symbiodiniaceae species, *Cladocopium goreaui* or *Durusdinium trenchii*, and assessed their gene expression in comparison to aposymbiotic controls after 10 and 20 days. While there was no differential expression of cnILPs in individuals infected with *C. goreaui* compared to controls, individuals exposed to *D. trenchii* displayed up-regulation of cnILP-A2b and cnILP-C3 after 10 days exposure and down-regulation of cnILP-B at 20 days. In Meron et al. (2019), symbiotic and aposymbiotic *E. paradivisa*, acquired via long-term maintenance under no light, were

exposed to one of three temperature regimes that included no heat stress at 25 °C, 28 °C heat stress, and 31 °C heat stress. Under no heat stress at 25 °C, symbiotic *E. paradivisa* had reduced gene expression of cnILP-B and cnILP-C4. We also analyzed two symbiotic datasets from the anemone *Exaiptasia pallida* (Ishii et al., 2019; Lehnert, Burriesci, & Pringle, 2012), which is used as a model system for coral research. Within both of these datasets we identified the down-regulation of cnILP-B in symbiotic individuals, along with the sole up-regulation of cnILP-A3 in Ishii et al. (2019).

In addition to the presence of symbiosis, we also detected changes in cnILP gene expression among corals hosting different Symbiodiniaceae species, specifically the thermally tolerant *D. trenchii* and *C. goreauii*. In Yuyama and Higuchi (2014), *A. tenuis* infected with *D. trenchii* displayed higher expression of cnILP-A3, cnILP-A2b, cnILP-B, cnILP-C2, and cnILP-C4 after 10 days but decreased expression of cnILP-A3 and cnILP-C1 after 20 days. We also examined an additional dataset that contrasted gene expression under heat and light stress in *Acropora solitaryensis* when it was hosting either the thermotolerant *D. trenchii* or *C. goreauii* (ACC: PRJDB6866). In this study, corals hosting *D. trenchii* that were not exposed to stress displayed higher expression of cnILP-A2a and cnILP-C1 in combination with decreased cnILP-C4 expression.

We further examined the unique contrast between two studies examining transcriptomic responses of the coral *Acropora digitifera* after 4, 12, and 48 hours of infection with an algal species that was either a mutualistic endosymbiont, *Breviolum minutum* (A. R. Mohamed et al., 2016), or a non-symbiotic coral-associated alga, *Chromera velia* (Amin R. Mohamed et al., 2018), respectively, compared to aposymbiotic control treatments.

When exposed to its endosymbiont *B. minutum* in A. R. Mohamed et al. (2016), there was no differential expression of cnILPs at any of the timepoints between corals infected with a mutualist and control treatments. As time progressed, however, individuals infected with a mutualist alga developed reduced cnILP-A3 expression at 12 hours post-infection that persisted through the 48 hour timepoint, at which time they also developed increased cnILP-B expression.

When exposed to the non-mutualist *C. velia* in Amin R. Mohamed et al. (2018), there was differential expression between infected corals and aposymbiotic at 48 hrs when there was a large up-regulation of cnILP expression including cnILP-A2a, cnILP-A2b, cnILP-A3, cnILP-B and cnILP-C1 in infected corals. Within the *C. velia*-infected treatment, there was no change in cnILP expression between the 4 and 12 hour timepoints. However, at 48 hours cnILP-B and cnILP-C4 displayed increased expression compared to both the 4 and 12 hour timepoints and cnILP-A2b and cnILP-C1 were increased when compared to the 12 hour timepoint. The aposymbiotic *A. digitifera* from both of the above studies also experienced a drastic reduction after 12 hours in cnILP expression, specfically cnILP-A2a, cnILP-A2b, cnILP-A3, cnILP-B, cnILP-C1, cnILP-C3, and cnILP-C4. The decreased expression of cnILP-A3 and cnILP-C1 was maintained through the 48 hour timepoint. There was no differential expression of cnILPs between 12 and 48 hours in the aposymbiotic controls from either study or the *B. minutum*-infected treatments.

3.4.2 Effect of thermal stress on symbiotic cnILP gene expression

To understand the response of cnILPs to hyperthermal stress we examined six separate studies that address different facets of the thermal stress response of corals. In Barshis et al. (2013) (Acc: PRJNA177515), they utilized thermally resilient *Acropora hyacinthus* corals from highly variable (HV) pools and thermally sensitive *A. hyacinthus* from moderately variable (MV) pools located on the same reef. These two groups of varying thermal tolerance were then exposed to either control or thermal stress conditions for 72 hours and assessed for gene expression. Our analysis identified no change in cnILP gene expression in HV corals exposed to heat stress while MV corals displayed increased expression of cnILP-A3, cnILP-A2a, cnILP-C1, and cnILP-C3. Contrasting the baseline cnILP expression levels between HV and MV corals under no heat stress revealed that HV corals possessed lower expression of cnILP-A3, cnILP-A2a, and cnILP-B. Under heat stress, HV corals maintained lower expression of these three cnILPs along with the addition of decreased expression in cnILP-C1 and cnILP-C3. Heat stress also caused up-regulation of cnILP-A3 in the coral *Galaxea fascicularis* from Hou et al. (2018) after 18 hours of heat stress although there was no differential expression of other cnILPs and no change in cnILP expression at 10 hours.

Another study (Acc: PRJDB6866) examined the effect of both elevated heat and light conditions after one or two days on *Acropora solitaryensis* corals infected with either *C. goreaui* or *D. trenchii* symbionts. After one day of stress, individuals containing *C. goreaui* had up-regulated cnILP-A3 and cnILP-C4 that was joined by the addition of down-regulated cnILP-C1, cnILP-C2, and cnILP-C3 after two days of stress. In

individuals containing *D. trenchii,* however, after one day of stress there was up-regulated cnILP-A3 and down-regulated cnILP-C1. This contrasting regulation was joined by the addition of down-regulated cnILP-B, cnILP-C2, and cnILP-C3 and up-regulated cnILP-C4 after two days of stress. There were also differences in baseline cnILP expression levels between *A. solitaryensis* that were infected with the different symbionts with those containing *D. trenchii* exhibiting higher expression of cnILP-A2a and cnILP-C1 and lower expression of cnILP-C4. After one day of stress, *D. trenchii*-infected individuals had lower expression of cnILP-A3, cnILP-C1, and cnILP-C4 compared to those infected with *C. goreaui*. After two days of stress, individuals with *D. trenchii* had lower expression of cnILP-A3 and cnILP-C4 but higher expression of cnILP-C2 and cnILP-3.

3.4.3 cnILP expression under ocean acidification

We examined four transcriptomic datasets on the response of corals to the effects of ocean acidification including Yuan et al. (2018) (PRJNA380146) in *Acropora gemmifera,* Moya et al. (2012) in *Acropora millepora*, Moya et al. (2015) in *A. millepora,* and González-Pech et al. (2017) in *Montipora digitata.* Differential gene expression of cnILPs was only present in (Yuan et al., 2018) where cnILP-B and cnILP-C4 were up-regulated under high acidic conditions compared to controls. There was no change in cnILP gene expression in the medium acidification exposure compared to the control treatment.

3.5 Discussion

In our work we identified conserved changes in cnILP gene expression, most significantly that decreased cnILP expression might be a hallmark of symbiotic gene expression in anemones and corals. Suppression of cnILP expression was supported by all five published transcriptomic studies investigating mutualistic symbiosis where we identified cnILP down-regulation, particular of cnILP-B, between aposymbiotic controls and individuals infected with mutualist symbionts (Ishii et al., 2019; Lehnert et al., 2012; Meron et al., 2019) or in the temporal progression of symbiosis establishment (A. R. Mohamed et al., 2016; Yuyama & Higuchi, 2014). The consistently decreased cnILP expression during symbiosis is significant as decreased insulin signaling is generally associated with increased resistance to stress and longevity (Barbieri et al., 2003; Kimura et al., 1997; Taguchi & White, 2008).

Surprisingly, in symbiotic *A. digitifera* larvae there was decreased expression of cnILP-A3 in place of cnILP-B, which instead displayed expression increased at 48 hours. However, our finding is supported by a separate dataset from *A. tenuis* where a rapid increase in algal density of *D. trenchii*-infected corals yielded increased cnILP-B expression after 10 days when compared to *C. goreauii*-infected corals, which contained almost no endosymbiotic algae. Nevertheless, cnILP-B expression decreased within *D. trenchii*-infected corals between 10 and 20 days as algal density increased, suggesting that there might be an initial up-regulation of cnILP-B during the onset of symbiosis that is later followed by decreased expression during stable symbiosis after it is established. Time-dependent cnILP expression patterns and/or stages during symbiosis would

resemble the symbiotic expression pattern of the mosquito *Anopheles stephensi*'s ILP3 (AsILP3), which is up-regulated 24-48 hours after infection with the malaria-causing protozoan *Plasmodium falciparum* (Marquez et al., 2011; Pietri et al., 2015). However after the initial development of the parasitism and suppression of the host's immune system, additional AsILP3 decreases the prevalence of infection as it has been hypothesized to not be important to increased development (Pietri et al., 2016).

Previous transcriptomic studies in corals have already implicated the insulin-like signaing pathway in symbiosis. Gene ontology enrichment analysis from Yuyama and Higuchi (2014) identified the term *insulin-like growth factor signaling pathway* as down-regulated in *D. trenchii*-infected corals after 20 days, but not 10 days, after their populations had increased significantly within the coral. Considering that down-regulated cnILP-B was the only cnILP we identified as differentially expressed between 10 and 20 days in these corals, its disappearance suggests that the down-regulation of cnILP-B might be connected to the altered insulin-like growth factor signaling pathway activity identified in their study under high symbiont density. Perhaps there is a similar pattern within symbiotic cnidarians where induction of cnILP-B synthesis, and its downstream signaling components, is essential for the establishment of symbiosis, but continued expression after the initial symbiotic establishment becomes detrimental. Regardless, this down-regulation of IIS activity is surprising considering glucose, one of the main photosynthates translocated to the coral host from its symbiont on a daily basis, should theoretically induce IIS activity and ILP synthesis.

A recent study using cell-specific transcriptomics in *E. pallida* has proposed a model of symbiotic establishment in corals that directly implicates mTORC1 (mechanistic target of rapamycin complex 1) in sensing symbiont-derived nutrients at the surface of the symbiosome, similarly to mammalian lysosomes, and triggering cellular proliferation while preventing autophagic degradation of the symbiont-containing symbiosomes (Voss et al., 2019). Interestingly, in contrast to fed aposymbiotic anemones in our study, mTORC1 activation in symbiotic anemones was continuously elevated and never decreased. Many studies in model organisms have connected the overactivation of TOR signaling because of high nutrient inputs to the development of a negative feedback mechanism that represses insulin signaling, and which ultimately causes the development of insulin resistance, a risk factor of type 2 diabetes in humans (Shah, Wang, & Hunter, 2004). While insulin resistance can be detrimental, it is believed to hold evolutionary importance by facilitating survival via diverting the majority of glucose to biosynthetic processes to sustain energetically costly processes such as growth in organisms who must persist through acute stress events or periods of prolonged nutrient deprivation (Soeters & Soeters, 2012).

The hypothesis that alterations in IIS activity are either a byproduct of, or perhaps even necessary for, the establishment of mutualistic symbiosis in corals is further supported by contrasting cnILP expression between corals exposed to mutualist and non-mutualist algae in *A. digitifera*. While infection with its mutualist *B. minutum* resulted in no difference in cnILP expression between the symbiotic treatments and aposymbiotic controls, infection with the non-mutualistic *C. velia* led to a large induction in cnILP expression after 48 hours when compared to aposymbiotic corals, particularly cnILP-A3,

cnILP-A2a, cnILP-A2b, cnILP-B, and cnILP-C1. This up-regulation was also reflected in the progression of cnILP expression over time in the symbiotic treatment, as non-mutualistic infection led to an accumulation of cnILP up-regulation over time, in stark contrast to the more muted response to *B. minutum* characterized by an initial down-regulation of cnILP-A3 followed by increased expression of only cnILP-B. This larger cnILP response to the non-mutualist is not surprising, considering other studies on microarrays have also demonstrated that competent Symbiodiniaceae species, in comparison to non-competent strains, seem to be able to either avoid detection by or influence the coral's immune response, leading to an almost non-existent change in the gene expression (Schnitzler & Weis, 2010; Voolstra et al., 2009). In light of this, the marked elevation in cnILP expression in corals infected with non-mutualist algae suggests that down-regulation of IIS may be an important aspect of the establishment of stable symbiosis with their endosymbiotic algae.

Up-regulation of cnILPs is not only a stress response to non-mutualist symbionts, but also seems to potentially be a more generalized stress response in cnidarians, such as in the response to thermal stress in particular. While we looked for differential expression of cnILPs in both thermal stress and ocean acidification stress, out of the four studies investigating ocean acidification there was only a minor up-regulation in cnILP gene expression in the work by (Yuan et al., 2018), suggesting the IIS might play a minor role in responding to pH changes. However, in Meron et al. (2019) where we identified decreased expression of cnILP-B and cnILP-C4 in symbiotic *E. paradivisa* corals, heat stress reversed the pattern of symbiotic cnILP down-regulation as there was increasing expression of these two cnILPs under elevated temperatures in symbiotic corals, resulting

in no differential expression of cnILPs between symbiotic and aposymbiotic corals under 31 °C heat stress. We additionally identified cnILP-A3 up-regulation in response to heat stress within the coral *G. fascicularis*. Surprisingly, the response of cnILP expression under heat stress in corals was in direct contrast to its response in the anemone *E. pallida*, where we identified a down-regulation in cnILP-C5 gene expression under heat stress in symbiotic anemones Ishii et al. (2019). This contradiction suggests that perhaps that functional role of IIS in the heat stress response of symbiotic cnidarians might be complex, or that there might have been divergence in IIS signaling patterns and outcomes between corals and anemones over evolutionary time.

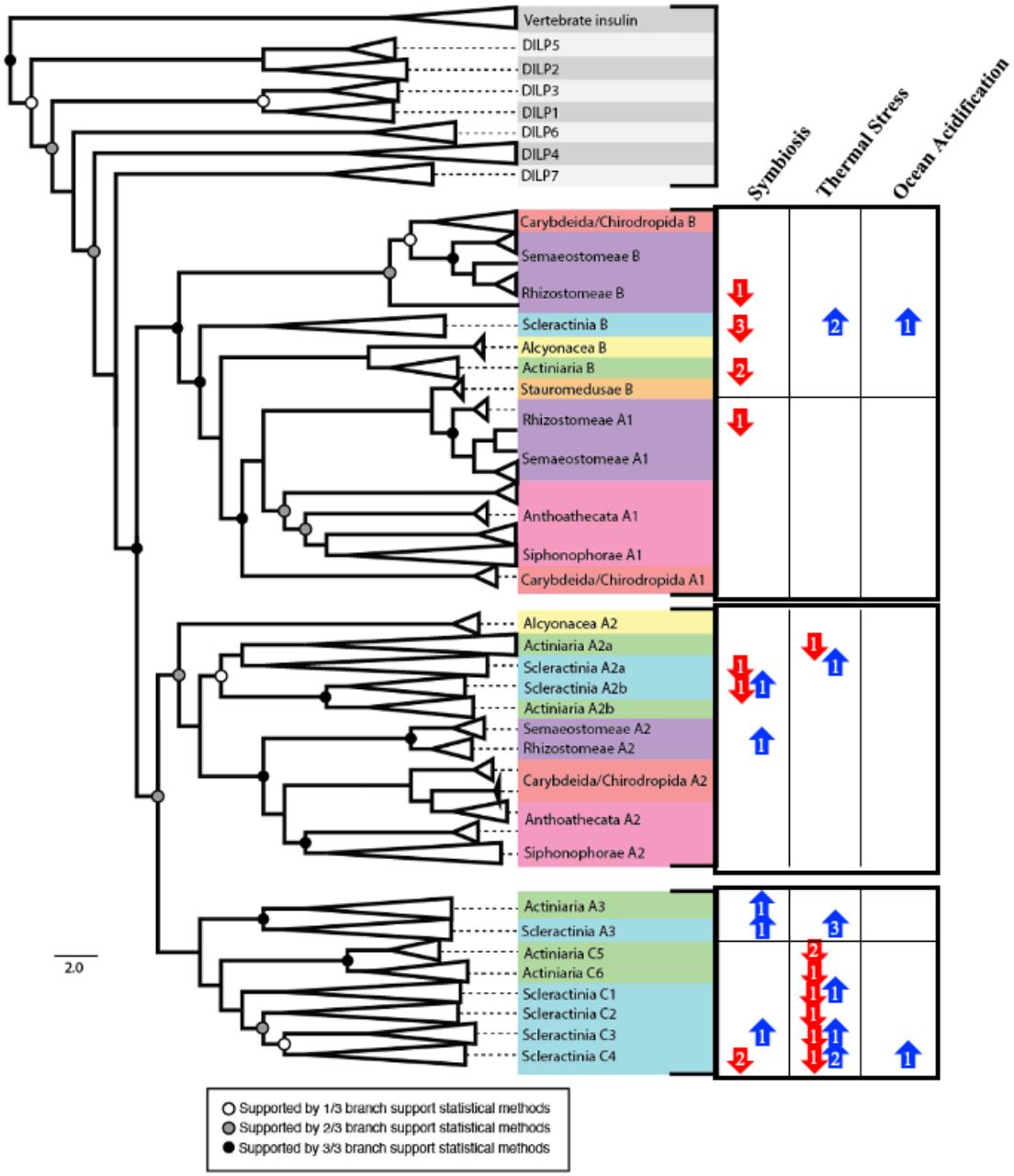
Variability in cnILP gene expression responses to heat stress might also be influenced by differing levels of physiological resilience to stress between individuals, or species. In Barshis et al. (2013), they used reef corals that are naturally exposed to medium variability (MV) temperature changes and are less resistant to thermal stress and slow growing compared to corals that experience high variability (HV) temperature changes to examine their thermal stress responses. They found that HV corals exhibited a markedly reduced transcriptomic response to heat stress as they seemed to be more predisposed to resist thermal stress due to their frontloading of stress response genes. In agreement with this finding, we identified no cnILP differential expression under heat stress in HV corals, but up-regulation of cnILP-A2a, cnILP-A3, cnILP-C1, and cnILP-C3 in MV corals. Further, under heat stress HV corals had markedly lower expression of cnILPs than MV corals, including decrease cnILP-A2a, cnILP-A3, cnILP-B, cnILP-C1, and cnILP-C3. The HV corals even exhibited lower baseline expression of cnILP-A2a, cnILP-A3, and cnILP-B under no heat stress compared to MV corals by more than 3 fold,

although this was not statistically significant. Regardless, the overall lower expression of cnILPs in the thermally resilient HV corals under heat stress, and potentially baseline expression, parallels IIS patterns seen in many other organisms where decreased insulin signaling is associated with increased stress resistance and longevity (Barbieri et al., 2003; Kimura et al., 1997; Piper et al., 2008; Salih & Brunet, 2008; Taguchi & White, 2008). Altogether, the commonalities between the expression patterns in resilient coral populations and model systems suggests that insulin signaling and its evolutionary role in stress resistance and longevity might be an important mechanism in the stress responses of corals.

3.6 Conclusion

The insulin-like signaling pathway is an evolutionarily important signaling pathway with pleiotropic functions that are essential to many aspects of metazoan life. Within cnidarians, where studies into this pathway are remarkably absent, we identified various patterns in cnidarian IIS activity that are either conserved among cnidarians or comparable to that seen in other metazoans, from *Drosophila* to humans. These gene expression patterns implicate the cnidarian IIS in responding to, and perhaps regulating, the onset of symbiosis with their endosymbiotic algae. It further implicates insulin signaling in the response to stress, particularly thermal stress, and as a signaling component during the breakdown of symbiosis. Understanding how this pathway functions in the coral-algal symbiosis will help give us a better understanding of the potential signaling regulation that forms the coral animal's response to symbiosis and stress.

Figure 3-1. Diagram showing the numbers and types of cnILPs that exhibited gene expression changes in response to symbiosis, temperature stress, or ocean acidification stress. Red arrows indicated decreased expression and blue arrows increased expression while the numbers in the arrows indicate the number of gene expression studies it was detected.



3.7 References

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. In: Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Barbieri, M., Bonafè, M., Franceschi, C., & Paolisso, G. (2003). Insulin/IGF-I-signaling pathway: an evolutionarily conserved mechanism of longevity from yeast to humans. *American Journal of Physiology, Endocrinology, and Metabolism, 285*(5), E1064-1071. doi:10.1152/ajpendo.00296.2003

Barshis, D. J., Ladner, J. T., Oliver, T. A., Seneca, F. O., Traylor-Knowles, N., & Palumbi, S. R. (2013). Genomic basis for coral resilience to climate change. *Proceedings of the National Academy of Sciences, 110*(4), 1387-1392. doi:10.1073/pnas.1210224110

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Braun, M., Ramracheya, R., Bengtsson, M., Clark, A., Walker, J. N., Johnson, P. R., & Rorsman, P. (2010). γ-aminobutyric acid (GABA) is an autocrine excitatory transmitter in human pancreatic β-cells. *Diabetes, 59*(7), 1694-1701. doi:10.2337/db09-0797

Burns, J. A., Zhang, H., Hill, E., Kim, E., & Kerney, R. (2017). Transcriptome analysis illuminates the nature of the intracellular interaction in a vertebrate-algal symbiosis. *eLife, 6*. doi:10.7554/eLife.22054.001

Chambers, M. C., Song, K. H., & Schneider, D. S. (2012). Listeria monocytogenes Infection Causes Metabolic Shifts in *Drosophila melanogaster*. *PLoS ONE, 7*(12), e50679-e50679. doi:10.1371/journal.pone.0050679

Colombani, J., Raisin, S., Pantalacci, S., Radimerski, T., Montagne, J., & Léopold, P. (2003). A nutrient sensor mechanism controls *Drosophila* growth. *Cell, 114*(6), 739-749. doi:10.1016/S0092-8674(03)00713-X

Colwell, J. A. (2012). Mechanisms of insulin action. *JAMA: The Journal of the American Medical Association, 308*(10), 1042-1042. doi:10.1001/jama.308.10.1042-a

Dai, C., Kayton, N. S., Shostak, A., Poffenberger, G., Cyphert, H. A., Aramandla, R., . . . Powers, A. C. (2016). Stress-impaired transcription factor expression and insulin secretion in transplanted human islets. *Journal of Clinical Investigation, 126*(5), 1857-1870. doi:10.1172/JCI83657

Demas, G. E., Adamo, S. A., & French, S. S. (2011). Neuroendocrine-immune crosstalk in vertebrates and invertebrates: implications for host defence. *Functional Ecology, 25*(1), 29-39. doi:10.1111/j.1365-2435.2010.01738.x

DiAngelo, J. R., Bland, M. L., Bambina, S., Cherry, S., & Birnbaum, M. J. (2009). The immune response attenuates growth and nutrient storage in *Drosophila* by reducing insulin signaling. *Proceedings of the National Academy of Sciences of the United States of America, 106*(49), 20853-20858. doi:10.1073/pnas.0906749106

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics, 28*(23), 3150-3152. doi:10.1093/bioinformatics/bts565

Géminard, C., Rulifson, E. J., & Léopold, P. (2009). Remote control of insulin secretion by fat cells in *Drosophila*. *Cell Metabolism, 10*(3), 199-207. doi:10.1016/j.cmet.2009.08.002

Gilon, P., & Henquin, J. C. (2001). Mechanisms and physiological significance of the cholinergic control of pancreatic β-cell function. *Endocrine Reviews, 22*(5), 565-604. doi:10.1210/er.22.5.565

Giovannucci, E. (2002). Nutrition, insulin, insulin-like growth factors and cancer. *Hormone and Metabolic Research, 35*(11-12), 694-704. doi:10.1055/s-2004-814147

Gleason, C. E., Gonzalez, M., Harmon, J. S., & Paul Robertson, R. (2000). Determinants of glucose toxicity and its reversibility in the pancreatic islet β-cell line, HIT-T15. *American Journal of Physiology - Endocrinology and Metabolism, 279*(5 42-5). doi:10.1152/ajpendo.2000.279.5.e997

González-Pech, R. A., Vargas, S., Francis, W. R., & Wörheide, G. (2017). Transcriptomic resilience of the *Montipora digitata* holobiont to low pH. *Frontiers in Marine Science, 4*. doi:10.3389/fmars.2017.00403

Grönke, S., Clarke, D.-F., Broughton, S., Andrews, T. D., & Partridge, L. (2010). Molecular evolution and functional characterization of *Drosophila* insulin-like peptides. *PLoS Genetics, 6*(2), e1000857. doi:10.1371/journal.pgen.1000857

Haas, B., & Papanicolaou, A. (2017). TransDecoder. Retrieved from https://github.com/TransDecoder/TransDecoder/

Hagman, D. K., Hays, L. B., Parazzoli, S. D., & Poitout, V. (2005). Palmitate inhibits insulin gene expression by altering PDX-1 nuclear localization and reducing MafA expression in isolated rat islets of Langerhans. *Journal of Biological Chemistry, 280*(37), 32413-32418. doi:10.1074/jbc.M506000200

Hang, S., Purdy, A. E., Robins, W. P., Wang, Z., Mandal, M., Chang, S., . . . Watnick, P. I. (2014). The acetate switch of an intestinal pathogen disrupts host insulin signaling and lipid metabolism. *Cell Host and Microbe, 16*(5), 592-604. doi:10.1016/j.chom.2014.10.006

Holst, J. J. (2007). The physiology of glucagon-like peptide 1. *Physiological Reviews, 87*(4), 1409-1439. doi:10.1152/physrev.00034.2006

Hou, J., Xu, T., Su, D., Wu, Y., Cheng, L., Wang, J., . . . Wang, Y. (2018). RNA-Seq reveals extensive transcriptional response to heat stress in the stony coral *Galaxea fascicularis*. *Frontiers in Genetics, 9*. doi:10.3389/fgene.2018.00037

Huang, X., Ye, H., Huang, H., Yang, Y., & Gong, J. (2014). An insulin-like androgenic gland hormone gene in the mud crab, *Scylla paramamosain*, extensively expressed and involved in the processes of growth and female reproduction. *General and Comparative Endocrinology, 204*, 229-238. doi:10.1016/j.ygcen.2014.06.002

Ikeya, T., Broughton, S., Alic, N., Grandison, R., & Partridge, L. (2009). The endosymbiont *Wolbachia* increases insulin/IGF-like signalling in *Drosophila*. *Proceedings of the Royal Society of London B: Biological Sciences, 276*(1674), 3799-3807. doi:10.1098/rspb.2009.0778

Ishii, Y., Maruyama, S., Takahashi, H., Aihara, Y., Yamaguchi, T., Yamaguchi, K., . . . Minagawa, J. (2019). Global shifts in gene expression profiles accompanied with environmental changes in cnidarian-dinoflagellate endosymbiosis. *G3: Genes, Genomes, Genetics, 9*(7), 2337-2347. doi:10.1534/g3.118.201012

Kelpe, C. L., Moore, P. C., Parazzoli, S. D., Wicksteed, B., Rhodes, C. J., & Poitout, V. (2003). Palmitate inhibition of insulin gene expression is mediated at the transcriptional level via ceramide synthesis. *Journal of Biological Chemistry, 278*(32), 30015-30021. doi:10.1074/jbc.M302548200

Kimura, K. D., Tissenbaum, H. A., Liu, Y., & Ruvkun, G. (1997). daf-2, an Insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *science, 942*. doi:10.1126/science.277.5328.942

Kréneisz, O., Chen, X., Fridell, Y. W. C., & Mulkey, D. K. (2010). Glucose increases activity and Ca2+ in insulin-producing cells of adult *Drosophila*. *NeuroReport, 21*(17), 1116-1120. doi:10.1097/WNR.0b013e3283409200

Lehnert, E. M., Burriesci, M. S., & Pringle, J. R. (2012). Developing the anemone *Aiptasia* as a tractable model for cnidarian-dinoflagellate symbiosis: the transcriptome of aposymbiotic *A. pallida*. *BMC genomics, 13*(1), 271-271. doi:10.1186/1471-2164-13-271

Leloup, C., Tourrel-Cuzin, C., Magnan, C., Karaca, M., Castel, J., Carneiro, L., . . . Pénicaud, L. (2009). Mitochondrial reactive oxygen species are obligatory signals for glucose-induced insulin secretion. *Diabetes, 58*(3), 673-681. doi:10.2337/db07-1056

Liu, T. F., Brown, C. M., El Gazzar, M., McPhail, L., Millet, P., Rao, A., . . . McCall, C. E. (2012). Fueling the flame: bioenergy couples metabolism and inflammation. *Journal of Leukocyte Biology, 92*(3), 499-507. doi:10.1189/jlb.0212078

Marquez, A. G., Pietri, J. E., Smithers, H. M., Nuss, A., Antonova, Y., Drexler, A. L., . . . Luckhart, S. (2011). Insulin-like peptides in the mosquito *Anopheles stephensi*: Identification and expression in response to diet and infection with *Plasmodium falciparum*. *General and Comparative Endocrinology, 173*(2), 303-312. doi:10.1016/j.ygcen.2011.06.005

Meron, D., Maor-Landaw, K., Weizman, E., Ben-Asher, H. W., Eyal, G., Banin, E., . . . Levi, O. (2019). The algal symbiont modifies the transcriptome of the scleractinian coral *Euphyllia paradivisa* during heat stress. *Microorganisms, 7*(8), 256-256. doi:10.3390/microorganisms7080256

Miyamoto, T., Slone, J., Song, X., & Amrein, H. (2012). A fructose receptor functions as a nutrient sensor in the *Drosophila* brain. *Cell, 151*(5), 1113-1125. doi:10.1016/j.cell.2012.10.024

Mohamed, A. R., Cumbo, V., Harii, S., Shinzato, C., Chan, C. X., Ragan, M. A., . . . Miller, D. J. (2016). The transcriptomic response of the coral *Acropora digitifera* to a competent *Symbiodinium* strain: the symbiosome as an arrested early phagosome. *Molecular Ecology, 25*(13), 3127-3141. doi:10.1111/mec.13659

Mohamed, A. R., Cumbo, V. R., Harii, S., Shinzato, C., Chan, C. X., Ragan, M. A., . . . Miller, D. J. (2018). Deciphering the nature of the coral–*Chromera* association. *The ISME Journal, 12*(3), 776-790. doi:10.1038/s41396-017-0005-9

Moya, A., Huisman, L., Ball, E. E., Hayward, D. C., Grasso, L. C., Chua, C. M., . . . Miller, D. J. (2012). Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to CO2-driven acidification during the initiation of calcification. *Molecular Ecology, 21*(10), 2440-2454. doi:10.1111/j.1365-294X.2012.05554.x

Moya, A., Huisman, L., Forêt, S., Gattuso, J. P., Hayward, D. C., Ball, E. E., & Miller, D. J. (2015). Rapid acclimation of juvenile corals to CO2-mediated acidification by upregulation of heat shock protein and Bcl-2 genes. *Molecular Ecology, 24*(2), 438-452. doi:10.1111/mec.13021

Pi, J., Bai, Y., Zhang, Q., Wong, V., Floering, L. M., Daniel, K., . . . Collins, S. (2007). Reactive oxygen species as a signal in glucose-stimulated insulin secretion. *Diabetes, 56*(7), 1783-1791. doi:10.2337/db06-1601

Pietri, J. E., Pakpour, N., Napoli, E., Song, G., Pietri, E., Potts, R., . . . Luckhart, S. (2016). Two insulin-like peptides differentially regulate malaria parasite infection in the mosquito through effects on intermediary metabolism. *Biochemical Journal, 473*(20), 3487-3503. doi:10.1042/BCJ20160271

Pietri, J. E., Pietri, E. J., Potts, R., Riehle, M. A., & Luckhart, S. (2015). *Plasmodium falciparum* suppresses the host immune response by inducing the synthesis of insulin-like peptides (ILPs) in the mosquito *Anopheles stephensi*. *Developmental & Comparative Immunology, 53*(1), 134-144. doi:10.1016/J.DCI.2015.06.012

Piper, M. D. W., Selman, C., McElwee, J. J., & Partridge, L. (2008). Separating cause from effect: How does insulin/IGF signalling control lifespan in worms, flies and mice? *Journal of Internal Medicine, 263*(2), 179-191. doi:10.1111/j.1365-2796.2007.01906.x

Ritz-Laser, B., Meda, P., Constant, I., Klages, N., Charollais, A., Morales, A., . . . Philippe, J. (1999). Glucose-induced preproinsulin gene expression is inhibited by the free fatty acid palmitate 1. *Endocrinology, 140*(9), 4005-4014. doi:10.1210/endo.140.9.6953

Roberts, A., & Pachter, L. (2013). Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature methods, 10*(1), 71-73. doi:10.1038/nmeth.2251

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140. doi:10.1093/bioinformatics/btp616

Rodriguez-Diaz, R., Dando, R., Jacques-Silva, M. C., Fachado, A., Molina, J., Abdulreda, M. H., . . . Caicedo, A. (2011). Alpha cells secrete acetylcholine as a non-neuronal paracrine signal priming beta cell function in humans. *Nature Medicine, 17*(7), 888-892. doi:10.1038/nm.2371

Rosen, O., Manor, R., Weil, S., Gafni, O., Linial, A., Aflalo, E. D., . . . Sagi, A. (2010). A sexual shift induced by silencing of a single insulin-like gene in crayfish: ovarian upregulation and testicular degeneration. *PLoS ONE, 5*(12), e15281-e15281. doi:10.1371/journal.pone.0015281

Salih, D. A., & Brunet, A. (2008). FoxO transcription factors in the maintenance of cellular homeostasis during aging. *Current Opinion in Cell Biology, 20*(2), 126-136. doi:10.1016/j.ceb.2008.02.005

Saltiel, A. R., & Kahn, C. R. (2001). Insulin signalling and the regulation of glucose and lipid metabolism. *Nature, 414*(6865), 799-806. doi:10.1038/414799a

Scheen, A. J., Castillo, M. J., & Lefèbvre, P. J. (1996). Assessment of residual insulin secretion in diabetic patients using the intravenous glucagon stimulatory test: Methodological aspects and clinical applications. In *Diabetes Metabolism* (pp. 397-406).

Schnitzler, C. E., & Weis, V. M. (2010). Coral larvae exhibit few measurable transcriptional changes during the onset of coral-dinoflagellate endosymbiosis. *Marine Genomics, 3*(2), 107-116. doi:10.1016/j.margen.2010.08.002

Shah, O. J., Wang, Z., & Hunter, T. (2004). Inappropriate activation of the TSC/Rheb/mTOR/S6K cassette induces IRS1/2 depletion, insulin resistance, and cell survival deficiencies. *Current Biology, 14*(18), 1650-1656. doi:10.1016/j.cub.2004.08.026

Soeters, M. R., & Soeters, P. B. (2012). The evolutionary benefit of insulin resistance. *Clinical Nutrition, 31*(6), 1002-1007. doi:10.1016/j.clnu.2012.05.011

Taguchi, A., & White, M. F. (2008). Insulin-like signaling, nutrient homeostasis, and life span. *Annual Review of Physiology, 70*(1), 191-212. doi:10.1146/annurev.physiol.70.113006.100533

Taniguchi, C. M., Emanuelli, B., & Kahn, C. R. (2006). Critical nodes in signalling pathways: insights into insulin action. *Nature Reviews: Molecular Cell Biology, 7*(2), 85-96. doi:10.1038/nrm1837

Voolstra, C. R., Schwarz, J. A., Schnetzer, J., Sunagawa, S., Desalvo, M. K., Szmant, A. M., . . . Medina, M. (2009). The host transcriptome remains unaltered during the establishment of coral-algal symbioses. *Molecular Ecology, 18*(9), 1823-1833. doi:10.1111/j.1365-294X.2009.04167.x

Voss, P. A., Gornik, S. G., Jacobovitz, M. R., Rupp, S., Dörr, M. S., Maegele, I., & Guse, A. (2019). Nutrient-dependent mTORC1 signaling in coral-algal symbiosis. *bioRxiv*, 1-24. doi:10.1101/723312

Yuan, X., Yuan, T., Huang, H., Jiang, L., Zhou, W., & Liu, S. (2018). Elevated CO2 delays the early development of scleractinian coral *Acropora gemmifera*. *Scientific Reports, 8*(1), 2787-2787. doi:10.1038/s41598-018-21267-3

Yuyama, I., & Higuchi, T. (2014). Comparing the effects of symbiotic algae (*Symbiodinium*) clades C1 and D on early growth stages of *Acropora tenuis*. *PLoS ONE, 9*(6), e98999-e98999. doi:10.1371/journal.pone.0098999

Zhou, Z., Wu, Y., Zhang, C., Li, C., Chen, G., Yu, X., . . . Huang, B. (2017). Suppression of NF-κB signal pathway by NLRC3-like protein in stony coral *Acropora aculeus*

under heat stress. *Fish and Shellfish Immunology, 67*, 322-330. doi:10.1016/j.fsi.2017.06.027

Chapter 4: The Durusdinium trenchii genome: evolution of thermotolerance in a coral symbiont

4.1 Abstract

Dinoflagellates from the family Symbiodinaceae engage in diverse symbiotic associations with cnidarian hosts, particularly the ecologically significant corals. The increase in thermal-induced bleaching outbreaks in the past couple of decades has led to increased levels of thermal stress and bleaching, resulting in increasing levels of coral mortality. Coral symbionts from the genus *Durusdinium*, particularly the species *Durusdinium trenchii,* are notable for their contributions to the thermal resilience of the coral holobiont. In this work we investigate the genomic underpinnings of the thermotolerance of *D. trenchii,* which is believed to be the byproduct of a whole or near-whole genome duplication event. In this genome we identified an extensive level of gene duplication that was 10-100x greater than any other species from this family currently available. Further, we examined these duplicated gene blocks for evidence of positive selection and identified pervasive duplication and subsequent positive selection on genes related to thylakoid membrane integrity, the photosynthetic apparatus, and the mitigation of oxidative stress damage. Our study suggests that large-scale gene duplication followed by positive selection on genes related to photosynthesis might have likely contributed to the acquisition of thermal tolerance in *D. trenchii*, and thus contributed to the ecological significance of this species.

4.2 Introduction

The family Symbiodiniaceae encompasses a diverse assortment of dinoflagellates (Order: Suessiales) that hold particular importance due to their symbiotic relationships with marine invertebrates, especially corals (Yellowlees, Rees, & Leggat, 2008). Considering the large declines in coral populations due to thermal-induced bleaching, understanding not only the genomic basis for symbiosis but also symbiosis breakdown is of great interest to the coral research community (Weis, 2008). In this aspect, comparative genomics can also be valuable to inform about the mechanisms underlying dysbiosis, the disruption of the host and its microalgal symbionts (Bhattacharya et al., 2016; González-Pech, Bhattacharya, Ragan, & Chan, 2019).

Different genera and species of Symbiodiniaceae have diverse physiological responses to their environments and stressors such as elevated temperatures (Grégoire, Schmacka, Coffroth, & Karsten, 2017; Mansour, Pollock, Díaz-Almeyda, Iglesias-Prieto, & Medina, 2018). Although different genera and species of Symbiodiniaceae have variable tolerances to environmental stress, the *Durusdinium* genus contains the notable species *D. trenchii,* a thermotolerant species that is a relatively recent invader to the Caribbean (Pettay, Wham, Smith, Iglesias-Prieto, & LaJeunesse, 2015). This species has been shown to increase in abundance in corals in response to thermal stress and increase the thermal tolerance of its host while also causing negative long-term effects on the growth of its coral host (Silverstein, Cunning, & Baker, 2017).

Previous comparative genomics studies within the Symbidoiniaceae family have revealed extensive differences in genomic architecture including gene content, orientation, and synteny along with genomic enrichment in the functions of transmembrane transport, mechanisms of response to reactive oxygen species, and protection against UV radiation (Aranda et al., 2016; González-Pech, Bhattacharya, et al., 2019; González-Pech, Chen, et al., 2019; González-Pech, Stephens, et al., 2019; Lin et al., 2015; Liu et al., 2018). Previous genomic studies have focused primarily on species from the genus *Symbiodinium* (previously refered as Clade A; LaJeunesse et al. 2028). Genomic patterns present in the ecologically important *D. trenchii* might reveal underlying genomic characteristics of thermal tolerance.

Considering the increasing prevalence and severity of thermal-induced bleaching, understanding what increases the thermal tolerance of the coral holobiont is a critical aim within the coral field. Uncovering the genomic underpinnings to the thermotolerance associated with *D. trenchii* would be an immense resource for the generation of new testable hypotheses for ongoing functional work aimed at improving the thermotolerance of the coral holobiont to conserve coral reef ecosystems.

4.3 Methods

4.3.1 Symbiodiniaceae cultures

The *D. trenchii* strain CCMP2556 was originally isolated by Dr. Mary Alice Coffroth (Buffalo University, New York, USA) from an *Orbicella faveolata* coral sampled at a

depth of 10 m on Tennessee Reef, Florida Keys (24.7335° N, 20.7669° W). The cultures used in this study were acquired from the Provasoli-Guillard National Center for Marine Algae and Microbiota (Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine) and maintained for one year in PROV-50 medium (NCMA) with an antimicrobial cocktail (kanamycin [50 μg/mL], ampicillin [100 μg/mL], streptomycin [50 μg/mL], amphotericin B [2.5 μg/mL]) to minimize bacterial and fungal contamination. During this time, the culture was passaged in the antimicrobial cocktail and inspected via light microscopy to confirm culture quality and the absence of microbial contamination on a monthly basis. PCR amplification and sequencing of the CP23S marker sequence was also regularly conducted for comparison to the known *D. trenchii* sequence to verify the genetic identity of the cultures. No other *Durusdinium* spp. were maintained in the laboratory prior to or during this time.

Just prior to DNA isolation, the culture was cleaned using methods adapted from Su et al. (2007) to further minimize bacterial contamination. Briefly, $5x10^7$ log phase cells were pelleted via centrifugation (500 *g*, 3 min) and resuspended in 40 mL sterile PROV-50 medium using the aforementioned antimicrobial cocktail. Cells were once again pelleted and resuspended in PROV-50 medium with 0.005% Tween-80 and EDTA (0.1 M, pH 8.0). Resuspended cells were incubated at 25°C for 1 hour on a platform rocker followed by the addition of lysozyme (0.5 mg/mL) and SDS (0.25%) and another incubation at room temperature (RT) with rocking for 10 minutes. Finally, cells were washed three times with PROV-50 medium containing the antimicrobial cocktail.

4.3.2 Chromium Library DNA extraction and sequencing

Genomic DNA for the 10X Chromium library was extracted following the recommended MagAttract HMW DNA Kit (Qiagen), with slight modifications to the extraction protocol. Once *D. trenchii* cells were in exponential log phase growth, 50 mL of $5x10^7$ cells were collected, pelleted, washed with SDS (0.25%), and then washed an additional three times in sterile culture medium. The resulting cell pellet was flash frozen in liquid nitrogen and briefly ground in a liquid nitrogen-chilled mortar and pestle with acid-washed 425-600 μm glass beads (Sigma). Approximately 50 mg of the ground cells were then used as input for the extraction protocol using the MagAttract HMW DNA Kit (Qiagen). In addition to using a larger input sample, wide-bore pipette tips were used for all steps involving the handling of lysate or DNA, and mixing was performed by gentle inversion instead of the manufacture-indicated pipetting. The final gDNA was eluted in Buffer AE (Qiagen; 10 mM Tris-Cl 0.5 mM EDTA, pH 9.0) and stored at -20° C.

To confirm DNA integrity, the DNA sample was analyzed by PFGE which indicated the presence of an intact sample with a mean fragment size of approximately 100 kilobases. Briefly, the 10X Genomics Chromium Library preparation approach captures long-range distance information by using microfluidics to partition ~5-10 fragments of gDNA into each of over a million micelles, with each micelle containing a unique barcode. This method allows the reconstruction of long-range linkage information between read pairs that dramatically improves the quality of the genome assembly and is able to resolve whole genome duplications, which has been hypothesized to occur in *D. trenchii*. The *D.*

trenchii chromium library was constructed using 1 ng of gDNA as input and then sequenced across two HiSeq X lanes using 150 bp PE chemistry (performed by HudsonAlpha).

4.3.3 De novo genome assembly

Using the 10X Chromium sequencing, two preliminary genomes were assembled using the 10X Genomics Supernova assembler. As the estimated *D. trenchii* coverage of ~100x exceeded the optimal coverage range of the supernova assembler (38-56x), we used two complementary, sub-sampling approaches that either disregard or account for barcode information to achieve optimal read coverage. The first approach sub-samples by read number, reducing it to a specified max number of reads to be used as input for the regardless of their corresponding barcodes. The second approach sub-samples by barcode and only selects a certain fraction of the barcodes to be used as assembly input, thereby reducing the genome coverage while maintaining coverage within the individual molecules (i.e., barcodes). This resulted in two high-quality, preliminary assemblies generated through sub-sampling to 600 million reads (~50x coverage) and barcode fraction of 0.65. As a whole, or at least partial, genome duplication has been predicted for the *D. trenchii* genome, we exported the assemblies in the megabubbles format so that only unique sequence regions would be exported and prevent artificially inflating the genome length. These two preliminary assemblies were then merged using Metassembler into one high-quality assembly and subjected to two consecutive rounds of Pilon to

correct single base differences, small indels, larger indels or block substitution events, and fill gaps.

4.3.4 Identification and removal of putative contaminants

The presence of putative contaminant scaffolds in the genome assembly were investigated by creating a taxon-annotated GC-coverage plot (Figure 1) with the Blobtools suite (Laetsch & Blaxter, 2017) to identify scaffolds that deviated by read coverage, taxonomic sequence similarity, and/or GC content. Read coverage was obtained via bowtie mapping of Trimmomatic (Bolger, Lohse, & Usadel, 2014) quality trimmed reads to the genome assembly. The taxonomic identity of scaffolds was assigned to the group containing the species with the highest summed bit score for all unique BLASTN hits across the scaffold. Bit scores were obtained through BLASTN searches to a custom database containing all complete viral, archaea, bacteria, fungi, invertebrate, protozoa, and human genomes in NCBI (downloaded 7 September, 2018). Scaffolds were classified as putative contaminants and removed from the assembly if they were both labelled as non-Symbiodiniaceae from the summed bit scores and exhibited low read coverage of more than 1.5*IQR below the median coverage of scaffolds assigned to the Symbiodiniaceae group from the bit score values.

4.3.5 Gene prediction and comparative analysis

Prediction of gene models was carried out using the same workflow outlined in (Chen, González-Pech, Stephens, Bhattacharya, & Chan, 2020) in order to prevent biases

resulting from different bioinformatic approaches. This workflow incorporates multiple evidence types including transcriptomic sequencing, machine learning, sequence homology, and *ab initio* prediction before integrating this with an *a priori* knowledge of the differing quality levels of different evidence types to generate high-quality gene model predictions. The resulting gene model predictions were assessed by PCA comparison to other Suessiales gene predictions using the same method. Genome completeness of the *D. trenchii* genome was assessed using BLASTP searches of predicted proteins against the set of 458 core eukaryotic orthologs from CEGMA (Parra, Bradnam, & Korf, 2007).

Gene models were functionally annotated using BLASTP ($E \leq 10^{-5}$; query/target cover ≥ 50%) searches against the SWISS-PROT database followed by the TrEMBL database for predicted proteins with no significant hits. Pfam domains within predicted proteins were annotatued using PfamScan ($E \leq 0.001$) in conjunction with the Pfam-A database. Orthofinder v2.3.10 (Emms & Kelly, 2015) was used to infer homologous gene sets among Suessiales species (Supplementary Table S2) using NCBI-BLAST. PfamScan was used to predict protein domains and test for enrichment of protein domains in *D. trenchii* compared to other Symbidoiniaceae species. The protein domain enrichment test was conducted using a hypergeometric test with phyper from the R stats package. The presence of duplicated genes in collinear gene blocks was investigated using MCScanX (Wang et al., 2012). The identified genes from duplicated gene blocks were then used as input to identify enriched gene functions in these duplicated genomic regions using the R package topGO and the elimination method combined with a Fisher's exact test.

4.4 Results

4.4.1 Genome assembly

We used 10X Chromium Linked-Read sequencing to generate a *de novo* genome assembly for *D. trenchii* CCMP2556. The 1.676 Gb genome assembly consisted of 20,106 and had a scaffold N50 of 745 Mb (Table 4-1). According to k-mer distribution estimates, the estimated genome size for *D. trenchii* was 1.72 Gb, which is comparable to other Symbiodiniaceae genomes (Figure 4-2).

4.4.2 Repeat content

The percentage of total repeat content in the *D. trenchii* genome is comparable to that of other Symbiodiniaceae species at 25.9% (Figure 4-3). LTR repeats were much more common in the *D. trenchii* genome at 7.7% compared to other Symbiodiniaceae species where they make up less than 0.7% of their genomes. LTR repeats are quite common in the non-symbiotic dinoflagellate *Polarella glacialis*, where these repeats make up around 12% of their genomes. Similarly to *Breviolum minutum*, *Cladocopium goreaui*, and *Fugacium kawagutii*, LINE repeats were not particularly abundant in *D. trenchii* although they are common in species from the genus *Symbiodinium*.

4.4.3 Gene models

There were 39,826 total genes predicted from the *D. trenchii* genome, with metrics of the predicted gene models comparable to other Symbiodiniaceae (Table 4-2). PCA analysis

reveals that the *D. trenchii* gene model metrics are similar to those from the genera Symbiodinium and Polarella (Figure 4-4). Investigating these gene model predictions on an orthogroup level revealed that the patterns of total shared orthogroups between different Suessiales species reflect their phylogenetic relationships (Fig. 4-5). While Polarella and Symbiodinium species cluster according by genus, there is a similar level of shared orthogroups between the Breviolum, Cladocopium, Fugacium, and Durusdinium genera. *D. trenchii* exhibited the highest level of similarity in orthogroups with the species *B. minutum* followed by *C. goreaui* SCF055-01.

4.4.4 Extensive gene duplication in *D. trenchii*

We used MCScanX to identify areas of gene duplication that exist in collinear blocks of five or more genes, resulting in a total of 13,901 duplicated genes. Except for *S. tridacnidorum* CCMP2592 (297 genes in 56 duplicated gene blocks), the number of duplicated genes in duplicated gene blocks in *D. trenchii* is anywhere from 10-100x greater than the rest of Symbiodiniaceae. (Gonzalez-Pech et al. 2019). Of these 13,901 duplicated genes, 2,762 genes displayed evidence of purifying selection and 4,518 genes of positive selection (Figure 4-7). Gene ontology enrichment of these duplicated genes under positive selection revealed enrichment for a variety of functions, particularly those related to the chloroplasts including *thylakoid membrane organization*, *photoreactive repair*, *PSII associated light-harvesting complex II catabolic process*, *photoinhibition*, and *chloroplast rRNA processing* (Table 4-3). Duplicated genes experiencing positive selection also included many chaperone proteins including HSP70, DNAJ, DNAJ-1,

DNAJ-2, DNAJ-16, Hsp90 co-chaperone Cdc37, DNAJ-A2, DNAJ-B2, DNAJ-B6, and ClpB. There was also positive selection on many proteins from the photosystems including psaC, psbN, FCP, FCPE, FCPF, petC, petH.

4.4.5 Protein domain enrichment

We identified sixteen Pfam domains as being over-represented in the *D. trenchii* genome (Table 4-7). The majority of these domains consisted of eight repeat domains with one Kelch domain (PF13854), one FNIP domain (PF05725), one leucine-rich repeat (LRR) domain (PF13306), two tetratricopeptide (TPR) domains (PF00515, PF13181), and three PPR domains (PF01535, PF13041, PF13812). Two zinc finger domains were also over-represented including the Zinc finger C3HC4-type domain (PF13920) and the MYND zinc finger domain (PF01753) that functions in protein-protein interactions to repress transcription. Two enriched domains involved in cellular signaling include the MIB_HERC2 domain (PF06701) and the Ephrin_rec_like domain (PF07699). Other over-represented domains included the Tenascin EGF domain found in membrane and secreted proteins as well as a carbonic anhydrase domain (PF00484).

4.5 Discussion

We identified the most extensive amount of duplicated genes in syntenic, collinear gene blocks in *D. trenchii* compared to any previous Suessiales genome study thus far. This finding aligns with those from earlier microsatellite studies that have long implicated whole genome duplication as a factor in the evolution of *D. trenchii* (LaJeunesse et al.,

2014; Drew C Wham, Ning, & LaJeunesse, 2017; Drew C. Wham, Pettay, & LaJeunesse, 2011). Although k-mer estimate of the total genome length is smaller than our actual genome assembly, k-mer estimates of genome size are prone to underestimation especially when there is a high level of repetitiveness or duplication in a genome. Further, the identification of LTR repeats as a dominant repeat type making up 7.2% of the *D. trenchii* stands in stark contrast to the rest of Symbiodiniaceae, where LTR repeats fail to exceed 1% of those genomes. *D. trenchii* is not alone in this, however, as another Suessiales species, *Polarella glacialis*, also has higher amounts of LTR repeats at 12% of their genome (Stephens et al. 2020). LTR repeats have long been implicated in diverse plant species as being a main transposable element that signifies an evolutionary history of genomic restructuring in regards to genome length, structure, and polyploidy (Huang et al., 2020). Analysis of orthologous gene sets further supports this hypothesis as *D. trenchii* has more species-specific orthogroups than other Symbiodiniaceae species, which might be the result of gene duplication and then divergence leading to novel functions or isoforms.

Large-scale gene duplication followed by positive selection on genes related to photosynthesis likely contributed to the acquisition of thermal tolerance in *D. trenchii*. Positive selection on duplicated genes related to chloroplast membranes, particularly *chloroplast outer membrane* and *thylakoid membrane organization*, is significant as increased amounts of unsaturated polyunsaturated fatty acids increase the stability and strength of thylakoid membranes when exposed to ROS (Tchernov et al., 2004). In particular, all three annotated digalactosyldiacylglycerol synthases (one DGDG2, two

DGDG1 proteins) whose lipid products have previously been implicated in the increased stability and heat tolerance in *D. trenchii* compared to other Symbiodiniacean species (Rosset et al., 2019) exhibited positive selection in the duplicated gene blocks. In addition, there was also duplication and positive selection on genes in the photosynthetic apparatus within the thylakoid membranes including photosystem I (psaC), photosystem II (psbN, FCP, FCPB, FCPE, FCPF), and the cytochrome b_6f complex (petC, petH, CCB1). A previous transcriptomic analysis of adaptive selection in *D. trenchii* agrees with our finding of enrichment for positive selection in genes related to the chloroplast structure, particularly the thylakoid membrane (Ladner, Barshis, & Palumbi, 2012). Importantly, in this work we now connect this extensive positive selection on genes involved in chloroplast's structure and the photosynthetic apparatus to pervasive gene block duplication.

Another emergent theme within these positively-selected, duplicated gene blocks was that of genes involved in repairing damage from oxidative stress. Repair of photosynthetic damage, represented by *negative regulation of nitric oxide biosynthetic process, PSII associated light-harvesting complex II catabolic process*, and *photoinhibition* was largely due to the presence of Deg proteases and FtsH metalloproteases (FtsH, FtsH3, FtsH6), which function together in the turnover of photodamaged D1 protein, along with photo-lyases (cryD, phrA). A recent study on the transcriptional response to symbiosis and heat stress in *D. trenchii* further detected up-regulation of these proteins in response to both symbiosis and heat stress (Bellantuono, Dougan, Granados-Cifuentes, & Rodriguez-Lanetty, 2019). Various antioxidants also

contributed to this enrichment including peroxiredoxin BAS1, superoxide dismutase (sodA), glutathione peroxidase 1 (GPX), and ascorbate peroxidase (APX1).

Heat shock and chaperone proteins are another important tool for Symbiodiniaceae species in mitigating oxidative stress (Leggat, Hoegh-Guldberg, Dove, & Yellowlees, 2007), and were a dominant member within the positively-selected, duplicated genes of *D. trenchii.* Enrichment for *regulation of chaperone-mediated protein folding* encompassed and wide array of chaperone proteins (HSP70, DNAJ, DNAJ1, DNAJ2, DNAJ-16, Hsp90 co-chaperone Cdc37, DNAJ-A2, DNAJ-B2, DNAJ-B6, DNAJ-B14, DnaJ-C1, DnaJ-C10, DnaJ-C21, DnaJ-C27, LDJ2, clpB, p23-1). This was supplemented by chaperone receptors in the outer envelope of the chloroplast (AKR2B, OEP61), which are involved in the HSP70-dependent targeting of proteins to the chloroplasts in stress response (Kim, Na, Park, Baek, & Kim, 2019; Von Loeffelholz et al., 2011). Supported by the findings from a previous transcriptome investigation on positive selection in *D. trenchii* (Ladner et al., 2012), the expansion in the repertoire of chaperone and heat shock proteins in *D. trenchii* is significant in context of its thermotolerance. Previous work has shown that corals engaging in symbiosis with *D. trenchii* compared to their native symbionts display a greater oxidative stress response that resembles that of minor heat stress itself in some cases (Cunning & Baker, 2020; Matthews et al., 2017), presumably due to the increased ROS generation and transfer from *D. trenchii.* As *D. trenchii* is known to persist through higher temperatures not only in symbiosis but when free-living as well, perhaps this expansion in chaperone and heat shock proteins has contributed to the ability of *D. trenchii* to withstand higher levels of ROS generation and stress.

Of the various enrichment analyses and duplication investigations conducted in this study, the one common finding among them was the presence of enriched levels carbonic anhydrase genes and/or domains in *D. trenchii*. Not only were they identified as positively enriched on the protein domain level, but carbonic anhydrase genes were also the single driver of the most enriched GO category *carbon utilization* (Supplementary Table 7). Carbonic anhydrase proteins are known to be enriched in Symbiodiniaceae compared to other dinoflagellate species and represent one adaptation that enables them to engage in symbioses with corals, anemones, and clams (Aranda et al., 2016). As studies have shown that an increased availability of dissolved inorganic carbon to the symbionts can increase photosynthesis (Goiran, Al-Moghrabi, Allemand, & Jaubert, 1996; Tansik, Fitt, & Hopkinson, 2017), it is significant that we not only identified enriched levels of carbonic anhydrase genes but also duplication of and signatures of positive selection on this protein family in *D. trenchii.* Although prior phylogenetic analyses on carbonic anhydrases already show that they tend to be specific to and show genus-specific expansion and diversification in Symbiodiniaceae, it might be particularly important to the unique biology of *D. trenchii*.

Although repeat-containing proteins have long been known to be important to Symbiodiniaceae species and their symbiosis with metazoans, we identified a significant level of protein domain enrichment for these proteins. Ten out of the sixteen enriched protein domains were in fact repeat-containing proteins (Figure 6B), many of which were also commonly found as one of the top ten most abundant protein domains in *D. trenchii* as well as other Symbiodiniaceae species. Repeat-containing proteins are known to be

integral to host-symbiont interactions. In particulary, we detected enrichment for the LRR_5 protein domain (PF13306), which contains TvBspA-like-625 proteins. The leucine-rich repeat protein TvBspA-like-625 was first identified in the parasite *Trichomonax vaginalis* where it is believed to function as an extracellular protein that binds to host membrane proteins when establishing infection (Hirt, Harriman, Kajava, & Embley, 2002). We identified enrichment of five TPR and PPR repeat collectively, that are commonly implicated in protein-protein interactions in symbiosis (Shinzato et al. 2014, Frank et al. 2019)

4.6 Conclusion

In this work we investigate the genome of the thermotolerant *D. trenchii*, an ecologically important symbiont of corals that confers increased resilience under thermal stress. We identified extensive duplication of genes as evidenced by the duplication of collinear gene blocks of greater than give genes. We further examined these duplicated gene blocks for patterns related to the thermotolerance. Through this, we discovered extensive duplication and positive selection on gene functions related to the chloroplasts and thylakoid membrane integrity, which has been previously implicated in increased resilience to thermal stress in Symbiodiniaceae. We further detected duplication and positive selection on a diverse array of chaperone and heat shock proteins that are integral to the mitigation of damage resulting from oxidative stress. This aligns with previous studies showing that cnidarians hosting *D. trenchii* exhibit higher levels of oxidative stress response but are more resilient to bleaching. Altogether, this suggests that although

levels of oxidative stress are higher in *D. trenchii*, the extensive duplication in its genome might have contributed to novel functions related to thermoresilience and contributed to the ecological significance of this species.

Libraries	
Paired-end-lane-1-read1	388,811,062
Paired-end-lane-1-read2	388,811,062
Paired-end-lane-2-read1	399,364,351
Paired-end-lane-2-read2	399,364,351
Total num. of reads	1,576,350,826
Total read length (bp)	236,452,623,900
Assembly	
Overall G+C (%)	49.82
Number of scaffolds	20,106
Assembly length (bp)	1,667,027,443
N50 scaffold length (bp)	745,532
Max. scaffold length (Mbp)	4.429

Number of contigs	86,653
N50 contig length (bp)	118,893
Max. contig length (Mbp)	282.969
Gap (%)	7.66%
Estimated genome size (bp)	1,728,967,727
Assembled fraction of genome (%)	96.4

Table 4-1. *Durusdinium trenchii* sequencing library and final genome assembly metrics.

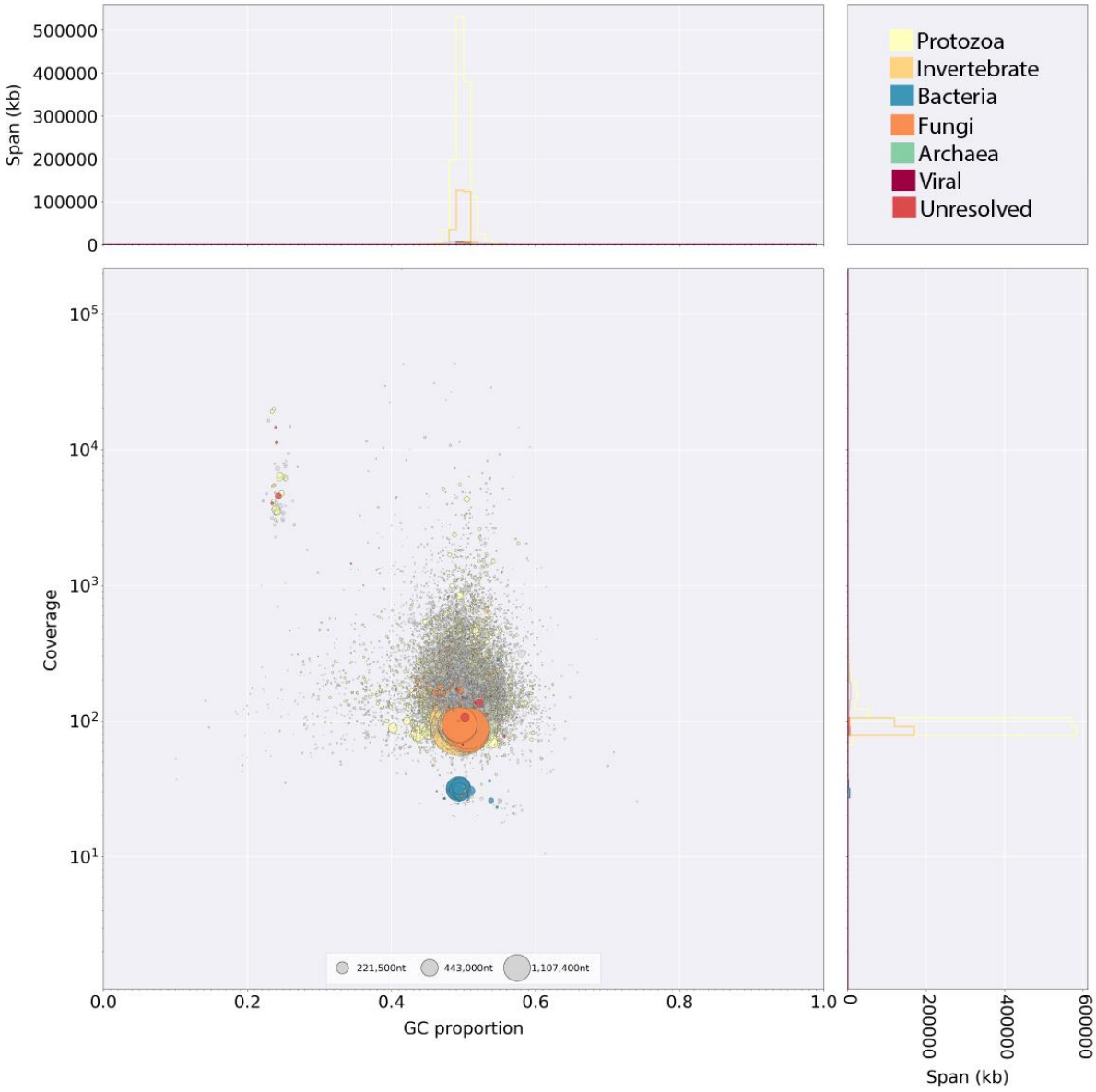


Figure 4-1. Blobtools taxon-annotated GC-coverage plot for D. trenchii genome assembly to investigate putative contaminant scaffolds.

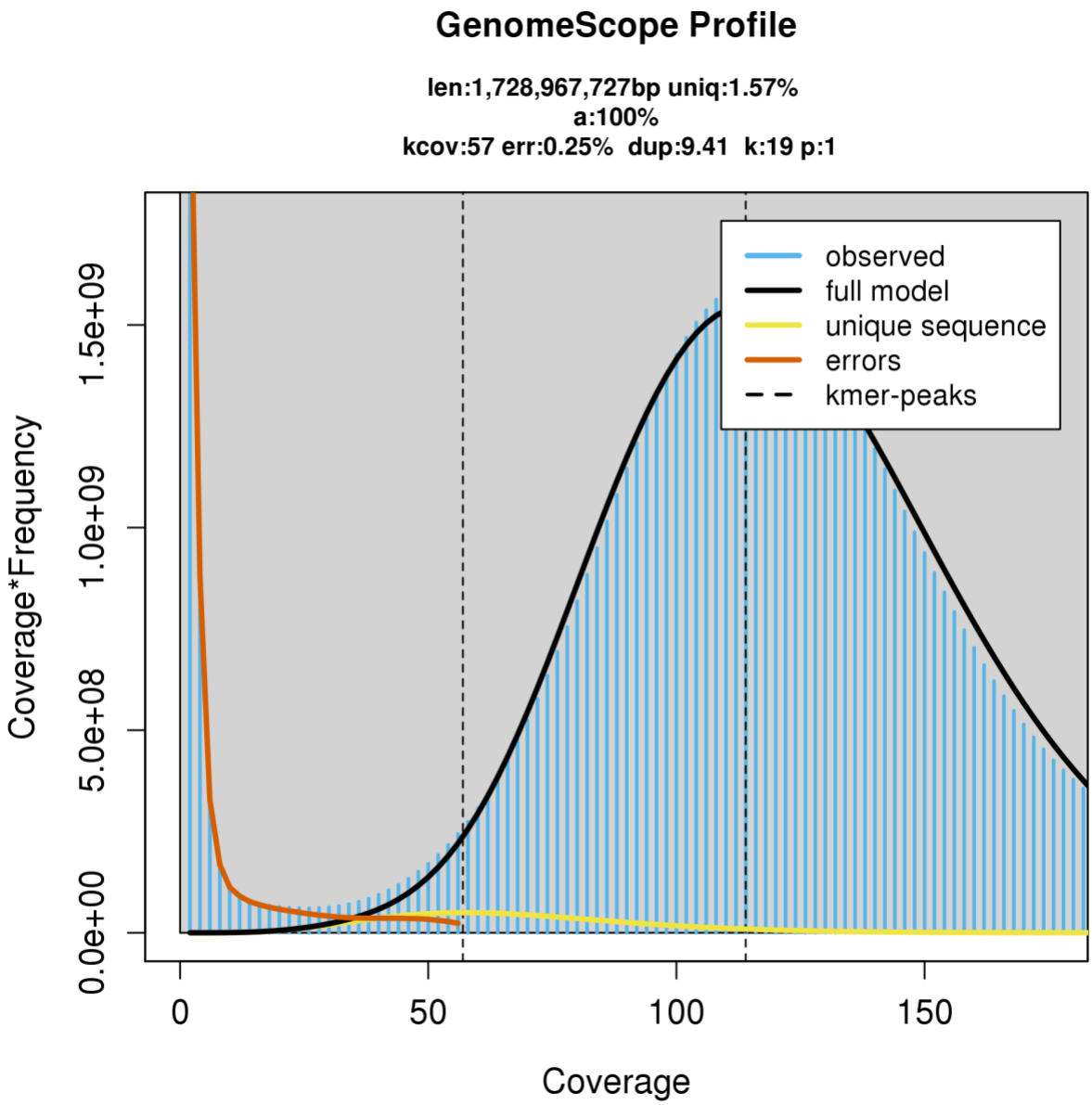


Figure 4-2. Genomescope2 plot depicting the k-mer plot and predicted D. trenchii genome size at k-mer 19 bp.

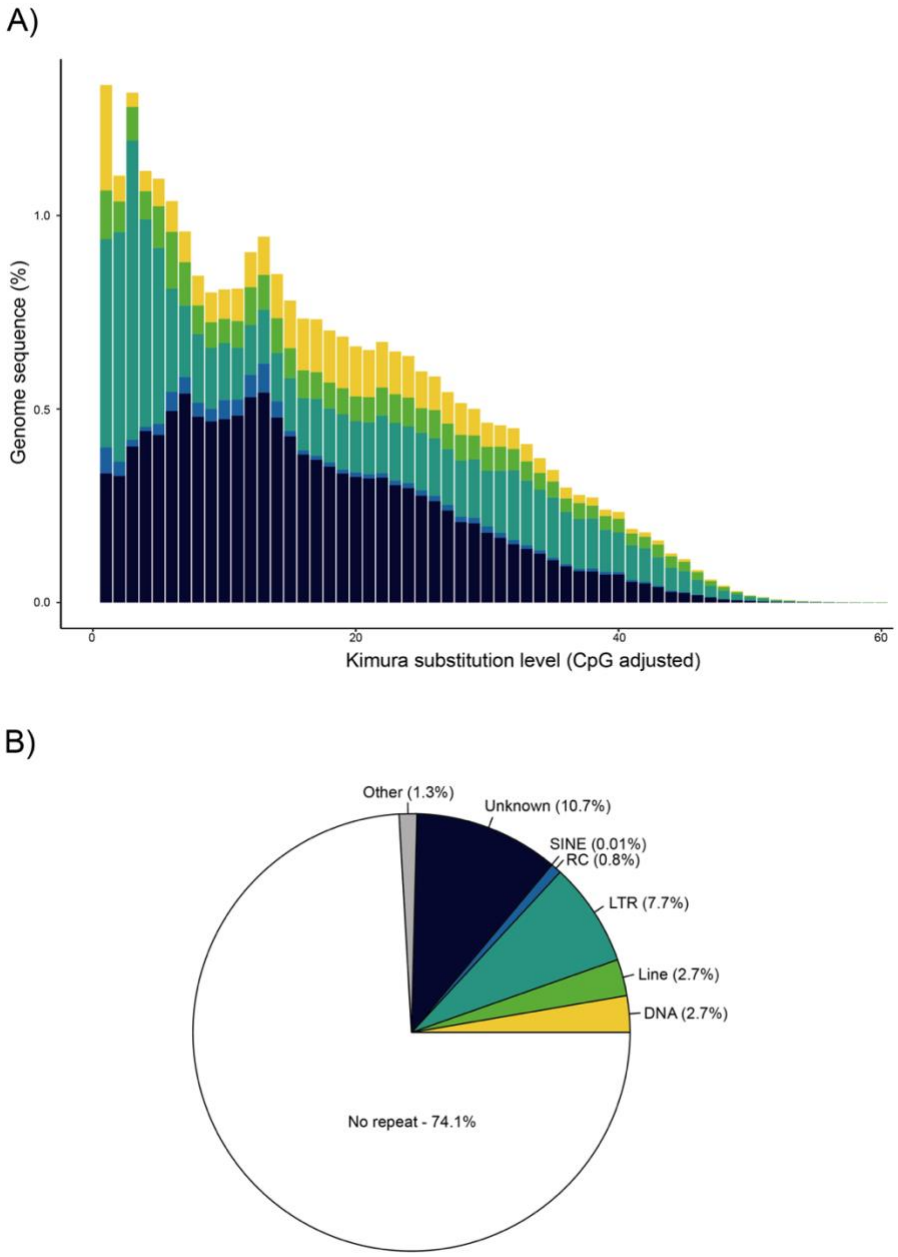


Figure 4-3. Diagrams of the (A) repeat landscape in D. trenchii according to the Kimura substitution level and the (B) size breakdown of different repeat types according to the percentage of the genome length.

Genes	**D. trenchii CCMP2556**
Number of genes	39,826
Average gene (exons + introns) length (bp)	13,184.34
Average CDS length (bp)	1,355.48
Gene content % (total gene length/total assembly length)	31.50
CDS G+C (%)	56.12
Supported by transcript data (%)	52.87
Exons	
Total number	618,252
Average number per gene	15.52
Average length (bp)	87.32
Total length (bp)	53,983,438
Introns	

Total number		578,426
Number of genes with introns		35,462
Average length (bp)		814.45
Total intron length (bp)		471,096,168
G+C (%)		
Intron-exon boundaries		
5'donor splice sites %	GC (canonical)	51.64
	GT (non-canonical)	19.64
	GA (non-canonical	28.81
Nucleotide after the AG 3'-acceptor splice sites (%)	G	97.06
	A	1.84
	T	0.78
	C	0.32

Intergenic regions	
Total number	34,239
Average length (bp)	854,483,645.00
G+C (%)	24,956.44

Table 4-2. Metrics of the D. trenchii gene model predictions used in the principal component analysis.

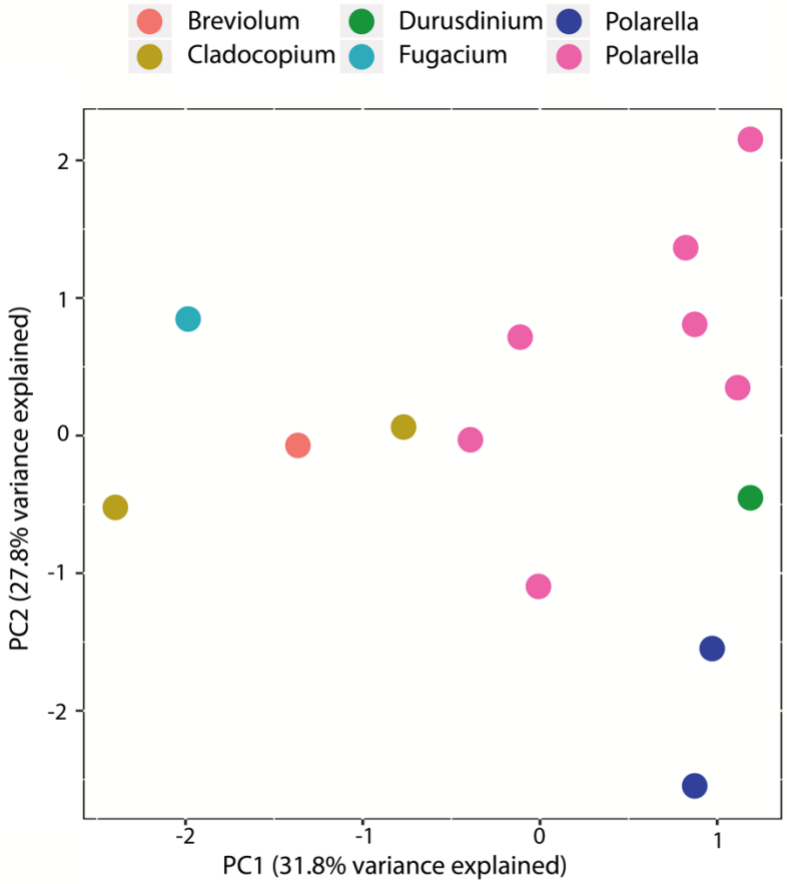


Figure 4-4. Principle component analysis of various metrics of the gene model predictions within Suessiales.

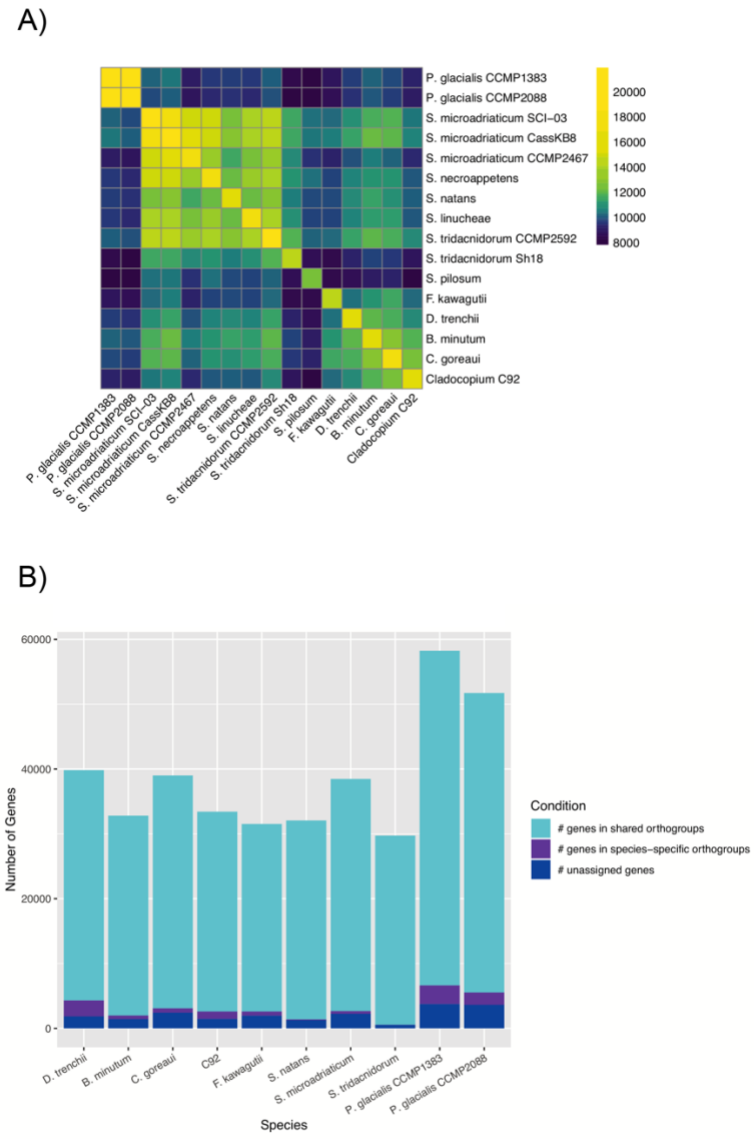


Figure 4-5. Representation of (A) shared orthologous gene sets across Suessiales identified using OrthoFinder2 as well as the (B) number of genes in orthologous gene sets shared by other Suessiales species, species-specific gene sets, and genes that were not assigned to an orthologous gene set.

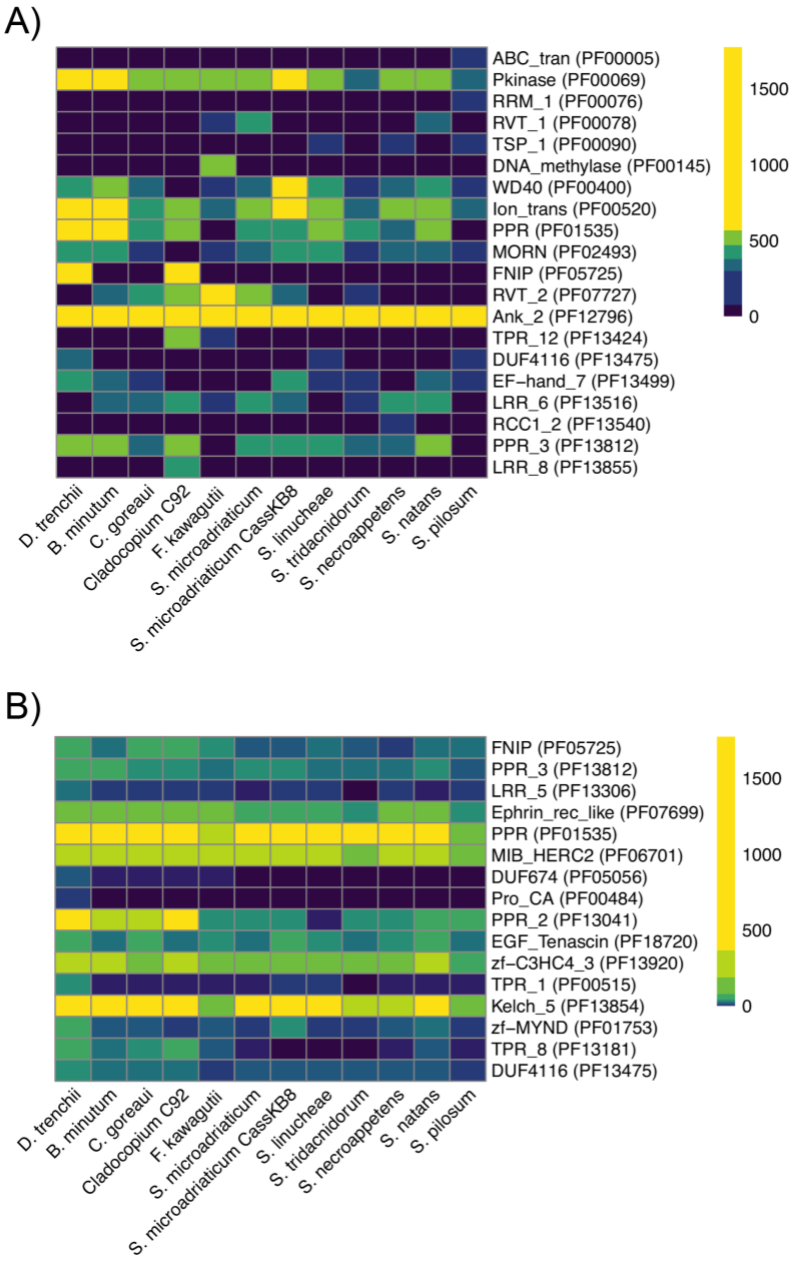


Figure 4-6. Heatmaps of Pfam protein domains that are (A) one of the top ten most abundant protein domains in one of the Symbiodiniaceae species or were (B) significantly enriched in the D. trenchii genome.

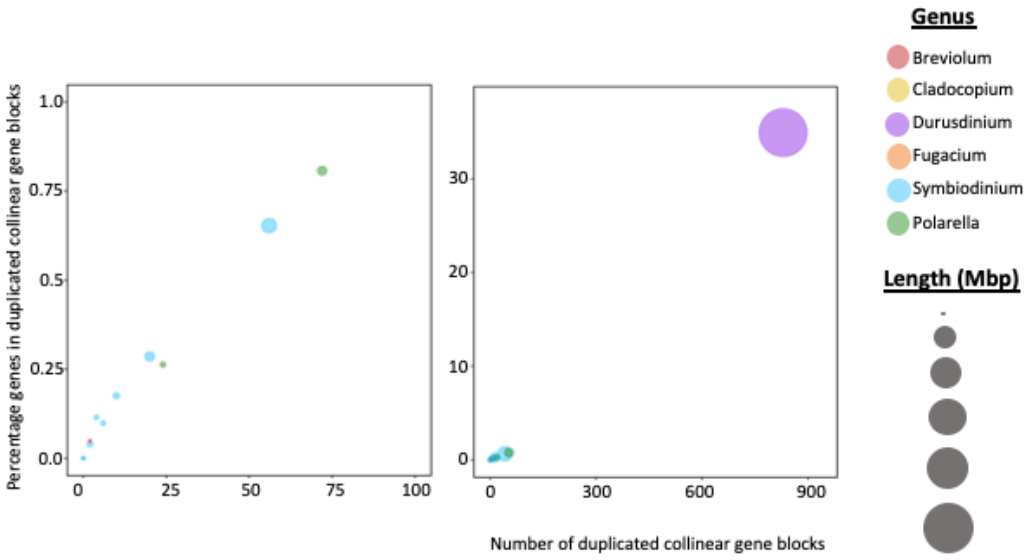


Figure 4-7. Scatterplot showing the percentage of genes in duplicated collinear gene blocks on the y-axis versus the total number of duplicated gene blocks on the x-axos in Suessiales species excluding D. trenchii on the left and including D. trenchii on the right.

GO ID	GO Term	p-value	Selective Pressure
Biological Process			
GO:0010027	thylakoid membrane organization	7.00E-05	Positive
GO:0031122	cytoplasmic microtubule organization	0.00011	Positive
GO:0090333	regulation of stomatal closure	0.00015	Positive
GO:0071372	cellular response to follicle-stimulating hormone stimulus	0.00019	Positive
GO:0046689	response to mercury ion	0.00027	Positive
GO:0007631	feeding behavior	0.00033	Positive
GO:0070555	response to interleukin-1	0.00036	Positive
GO:0010364	regulation of ethylene biosynthetic process	0.00037	Positive
GO:0044752	response to human chorionic gonadotropin	0.00037	Positive

GO:0097069	cellular response to thyroxine stimulus	0.00037	Positive
GO:0110025	DNA strand resection involved in replication fork processing	0.00037	Positive
GO:1905090	negative regulation of parkin-mediated stimulation of mitophagy in response to mitochondrial depolarization	0.00037	Positive
GO:2000490	negative regulation of hepatic stellate cell activation	0.00037	Positive
GO:0042127	regulation of cell proliferation	0.00039	Positive
GO:0009624	response to nematode	0.00051	Positive
GO:0010103	stomatal complex morphogenesis	0.00083	Positive
GO:0007052	mitotic spindle organization	0.00096	Positive
GO:0035304	regulation of protein dephosphorylation	0.001	Positive
GO:1905421	regulation of plant organ morphogenesis	0.001	Positive
GO:1902117	positive regulation of organelle assembly	0.00109	Positive

GO:1902476	chloride transmembrane transport	0.00111	Positive
GO:0055064	chloride ion homeostasis	0.00135	Positive
GO:0000719	photoreactive repair	0.00163	Positive
GO:0002176	male germ cell proliferation	0.00163	Positive
GO:0009958	positive gravitropism	0.00163	Positive
GO:0010304	PSII associated light-harvesting complex II catabolic process	0.00163	Positive
GO:0019227	neuronal action potential propagation	0.00163	Positive
GO:0072382	minus-end-directed vesicle transport along microtubule	0.00163	Positive
GO:0090671	telomerase RNA localization to Cajal body	0.00163	Positive
GO:0046685	response to arsenic-containing substance	0.00166	Positive
GO:0010923	negative regulation of phosphatase activity	0.00196	Positive
GO:0071300	cellular response to retinoic acid	0.00196	Positive

GO:0034097	response to cytokine	0.00204	Positive
GO:0090501	RNA phosphodiester bond hydrolysis	0.00219	Positive
GO:0060296	regulation of cilium beat frequency involved in ciliary motility	0.00237	Positive
GO:0120162	positive regulation of cold-induced thermogenesis	0.00239	Positive
GO:0045453	bone resorption	0.00242	Positive
GO:0097530	granulocyte migration	0.00242	Positive
GO:0010483	pollen tube reception	0.00265	Positive
GO:1901133	kanamycin biosynthetic process	0.00265	Positive
GO:1901315	negative regulation of histone H2A K63-linked ubiquitination	0.00265	Positive
GO:1904851	positive regulation of establishment of protein localization to telomere	0.00265	Positive
GO:0009611	response to wounding	0.00291	Positive

GO:0046958	nonassociative learning	0.00304	Positive
GO:0060339	negative regulation of type I interferon-mediated signaling pathway	0.00304	Positive
GO:0030433	ubiquitin-dependent ERAD pathway	0.00311	Positive
GO:0051298	centrosome duplication	0.0035	Positive
GO:0060972	left/right pattern formation	0.0035	Positive
GO:0071260	cellular response to mechanical stimulus	0.0035	Positive
GO:0007586	digestion	0.00358	Positive
GO:0010608	posttranscriptional regulation of gene expression	0.00388	Positive
GO:0035729	cellular response to hepatocyte growth factor stimulus	0.00391	Positive
GO:0042066	perineurial glial growth	0.00391	Positive
GO:0071716	leukotriene transport	0.00391	Positive

GO:1903818	positive regulation of voltage-gated potassium channel activity	0.00391	Positive
GO:0000056	ribosomal small subunit export from nucleus	0.00434	Positive
GO:0000733	DNA strand renaturation	0.00434	Positive
GO:0000738	DNA catabolic process, exonucleolytic	0.00434	Positive
GO:0006382	adenosine to inosine editing	0.00434	Positive
GO:0006729	tetrahydrobiopterin biosynthetic process	0.00434	Positive
GO:0007194	negative regulation of adenylate cyclase activity	0.00434	Positive
GO:0009944	polarity specification of adaxial/abaxial axis	0.00434	Positive
GO:0010148	transpiration	0.00434	Positive
GO:0015722	canalicular bile acid transport	0.00434	Positive
GO:0015949	nucleobase-containing small molecule interconversion	0.00434	Positive

GO:0016121	carotene catabolic process	0.00434	Positive
GO:0016446	somatic hypermutation of immunoglobulin genes	0.00434	Positive
GO:0022617	extracellular matrix disassembly	0.00434	Positive
GO:0031571	mitotic G1 DNA damage checkpoint	0.00434	Positive
GO:0048281	inflorescence morphogenesis	0.00434	Positive
GO:0051900	regulation of mitochondrial depolarization	0.00434	Positive
GO:0009074	aromatic amino acid family catabolic process	0.00461	Positive
GO:0071347	cellular response to interleukin-1	0.00465	Positive
GO:0045019	negative regulation of nitric oxide biosynthetic process	0.00504	Positive
GO:1901259	chloroplast rRNA processing	0.00504	Positive
GO:0032496	response to lipopolysaccharide	0.00523	Positive

GO:0019852	L-ascorbic acid metabolic process	0.00524	Positive
GO:0140014	mitotic nuclear division	0.00556	Positive
GO:0019886	antigen processing and presentation of exogenous peptide antigen via MHC class II	0.00595	Positive
GO:0070507	regulation of microtubule cytoskeleton organization	0.0061	Positive
GO:0009960	endosperm development	0.0062	Positive
GO:0046822	regulation of nucleocytoplasmic transport	0.00628	Positive
GO:0090276	regulation of peptide hormone secretion	0.00628	Positive
GO:0032469	endoplasmic reticulum calcium ion homeostasis	0.00694	Positive
GO:0048150	behavioral response to ether	0.00694	Positive
GO:1990074	polyuridylation-dependent mRNA catabolic process	0.00694	Positive
GO:0006536	glutamate metabolic process	0.00701	Positive

GO:0030317	flagellated sperm motility	0.00722	Positive
GO:0071383	cellular response to steroid hormone stimulus	0.00722	Positive
GO:0071230	cellular response to amino acid stimulus	0.0075	Positive
GO:0007342	fusion of sperm to egg plasma membrane involved in single fertilization	0.00751	Positive
GO:0016043	cellular component organization	0.00754	Positive
GO:0106134	positive regulation of cardiac muscle cell contraction	0.00757	Positive
GO:0043393	regulation of protein binding	0.00767	Positive
GO:0045053	protein retention in Golgi apparatus	0.00773	Positive
GO:0009247	glycolipid biosynthetic process	0.00778	Positive
GO:0032880	regulation of protein localization	0.00814	Positive
GO:0019614	catechol-containing compound catabolic process	0.00818	Positive

GO:0070374	positive regulation of ERK1 and ERK2 cascade	0.00818	Positive
GO:0000819	sister chromatid segregation	0.00871	Positive
GO:0061014	positive regulation of mRNA catabolic process	0.00879	Positive
GO:0043434	response to peptide hormone	0.00888	Positive
GO:0006203	dGTP catabolic process	0.00902	Positive
GO:0010205	photoinhibition	0.00902	Positive
GO:0033194	response to hydroperoxide	0.00902	Positive
GO:0046061	dATP catabolic process	0.00902	Positive
GO:1901029	negative regulation of mitochondrial outer membrane permeabilization involved in apoptotic signaling pathway	0.00902	Positive
GO:0034249	negative regulation of cellular amide metabolic process	0.00921	Positive

GO:0032655	regulation of interleukin-12 production	0.00941	Positive
GO:0044344	cellular response to fibroblast growth factor stimulus	0.00941	Positive
GO:0010940	positive regulation of necrotic cell death	0.00949	Positive
GO:0015694	mercury ion transport	0.00949	Positive
GO:0070327	thyroid hormone transport	0.00949	Positive
GO:1901074	regulation of engulfment of apoptotic cell	0.00949	Positive
GO:1901086	benzylpenicillin metabolic process	0.00949	Positive
GO:1903644	regulation of chaperone-mediated protein folding	0.00949	Positive
GO:2000623	negative regulation of nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	0.00949	Positive
Cellular Component			
GO:0019898	extrinsic component of membrane	0.00012	Positive

GO:0009707	chloroplast outer membrane	0.00041	Positive
GO:0043235	receptor complex	0.00043	Positive
GO:0031514	motile cilium	0.00056	Positive
GO:1902710	GABA receptor complex	0.00064	Positive
GO:0008278	cohesin complex	0.00065	Positive
GO:0017109	glutamate-cysteine ligase complex	0.00131	Positive
GO:0043661	peribacteroid membrane	0.00131	Positive
GO:0036128	CatSper complex	0.00159	Positive
GO:0005813	centrosome	0.00191	Positive
GO:0042644	chloroplast nucleoid	0.0025	Positive
GO:0005887	integral component of plasma membrane	0.00255	Positive
GO:0016939	kinesin II complex	0.00352	Positive
GO:0032127	dense core granule membrane	0.00352	Positive

GO:0016272	prefoldin complex	0.00375	Positive
GO:0005783	endoplasmic reticulum	0.00436	Positive
GO:0000323	lytic vacuole	0.00483	Positive
GO:0030658	transport vesicle membrane	0.0056	Positive
GO:0031965	nuclear membrane	0.00609	Positive
GO:0030425	dendrite	0.00626	Positive
GO:0038039	G protein-coupled receptor heterodimeric complex	0.00736	Positive
GO:0019897	extrinsic component of plasma membrane	0.00752	Positive
GO:0005777	peroxisome	0.0076	Positive
GO:0044426	cell wall part	0.00779	Positive
GO:0005901	caveola	0.0082	Positive
GO:1904115	axon cytoplasm	0.00868	Positive

Molecular Function

GO:0004616	phosphogluconate dehydrogenase (decarboxylating) activity	0.00031	Positive
GO:0004838	L-tyrosine:2-oxoglutarate aminotransferase activity	0.00031	Positive
GO:0005149	interleukin-1 receptor binding	0.00031	Positive
GO:0008832	dGTPase activity	0.00031	Positive
GO:0019855	calcium channel inhibitor activity	0.00031	Positive
GO:0032567	dGTP binding	0.00031	Positive
GO:0080079	cellobiose glucosidase activity	0.00031	Positive
GO:0016595	glutamate binding	0.00032	Positive
GO:0004843	thiol-dependent ubiquitin-specific protease activity	0.00046	Positive
GO:0004724	magnesium-dependent protein serine/threonine phosphatase activity	0.00056	Positive

GO:0004345	glucose-6-phosphate dehydrogenase activity	0.00068	Positive
GO:0004357	glutamate-cysteine ligase activity	0.0007	Positive
GO:0005315	inorganic phosphate transmembrane transporter activity	0.00081	Positive
GO:0004888	transmembrane signaling receptor activity	0.00112	Positive
GO:0008559	xenobiotic transmembrane transporting ATPase activity	0.00116	Positive
GO:0004108	citrate (Si)-synthase activity	0.00138	Positive
GO:0008124	4-alpha-hydroxytetrahydrobiopterin dehydratase activity	0.00138	Positive
GO:0102130	malonyl-CoA methyltransferase activity	0.00138	Positive
GO:0005227	calcium activated cation channel activity	0.00157	Positive
GO:0046481	digalactosyldiacylglycerol synthase activity	0.00233	Positive
GO:0070016	armadillo repeat domain binding	0.00233	Positive

GO:0005262	calcium channel activity	0.00236	Positive
GO:0004460	L-lactate dehydrogenase (cytochrome) activity	0.00244	Positive
GO:0008483	transaminase activity	0.00284	Positive
GO:0010436	carotenoid dioxygenase activity	0.0037	Positive
GO:0015379	potassium:chloride symporter activity	0.0037	Positive
GO:0019784	NEDD8-specific protease activity	0.0037	Positive
GO:0051920	peroxiredoxin activity	0.0037	Positive
GO:0102162	all-trans-8'-apo-beta-carotenal 15,15'-oxygenase	0.0037	Positive
GO:0005384	manganese ion transmembrane transporter activity	0.00402	Positive
GO:0008201	heparin binding	0.00414	Positive
GO:0019894	kinesin binding	0.00466	Positive

GO:0000175	3'-5'-exoribonuclease activity	0.00518	Positive
GO:0090729	toxin activity	0.00537	Positive
GO:0004471	malate dehydrogenase (decarboxylating) (NAD+) activity	0.00576	Positive
GO:0008017	microtubule binding	0.00663	Positive
GO:0008157	protein phosphatase 1 binding	0.00773	Positive
GO:0008251	tRNA-specific adenosine deaminase activity	0.00773	Positive
GO:0008144	drug binding	0.00816	Positive
GO:0005237	inhibitory extracellular ligand-gated ion channel activity	0.00839	Positive
GO:0042937	tripeptide transmembrane transporter activity	0.00928	Positive
GO:0043015	gamma-tubulin binding	0.00942	Positive
GO:0004683	calmodulin-dependent protein kinase activity	0.00972	Positive

Table 4-3. Significantly enriched gene ontology categories in genes located in duplicated collinear gene blocks experiencing adaptive selection, either positive or purifying selection.

Pfam Domain	adj.p	# in *D. trenchii*	# in Symbioiniaceae
Repeat domain-containing genes			
FNIP (PF05725)	2.00E-242	1062	3714
LRR_5 (PF13306)	7.00E-18	69	208
PPR (PF01535)	6.00E-09	634	5191
PPR_2 (PF13041)	3.00E-12	293	1955
PPR_3 (PF13812)	8.00E-04	537	4703
TPR_1 (PF00515)	2.00E-11	58	205
TPR_8 (PF13181)	2.00E-02	71	467
Kelch_5 (PF13854)	2.00E-03	24	93
Zinc finger domains			
zf-C3HC4_3 (PF13920)	3.00E-03	68	420

zf-MYND (PF01753)	2.00E-02	39	217
Miscellaneous			
Ephrin_rec_like (PF07699)	6.00E-09	169	1057
MIB_HERC2 (PF06701)	5.00E-11	33	80
Pro_CA (PF00484)	4.00E-04	65	369
EGF_Tenascin (PF18720)	3.00E-05	14	28
Unknown domains			
DUF674 (PF05056)	4.00E-04	6	6
DUF4116 (PF13475)	1.00E-02	335	2878

Table 4-4. Protein domains that are enriched in D. trenchii compared to other Symbiodiniaceae species

4.7 References

Aranda, M., Li, Y., Liew, Y. J., Baumgarten, S., Simakov, O., Wilson, M. C., . . . Bajic, V. B. (2016). Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Scientific reports, 6*, 39734.

Bellantuono, A. J., Dougan, K. E., Granados-Cifuentes, C., & Rodriguez-Lanetty, M. (2019). Free-living and symbiotic lifestyles of a thermotolerant coral endosymbiont display profoundly distinct transcriptomes under both stable and heat stress conditions. *Molecular Ecology, 28*(24), 5265-5281.

Bhattacharya, D., Agrawal, S., Aranda, M., Baumgarten, S., Belcaid, M., Drake, J. L., . . . Falkowski, P. G. (2016). Comparative genomics explains the evolutionary success of reef-forming corals. *Elife, 5*, 5741-5741. doi:10.7554/eLife.13288

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Chen, Y., González-Pech, R. A., Stephens, T. G., Bhattacharya, D., & Chan, C. X. (2020). Evidence That Inconsistent Gene Prediction Can Mislead Analysis of Dinoflagellate Genomes. *Journal of Phycology, 56*(1), 6-10. doi:10.1111/jpy.12947

Cunning, R., & Baker, A. C. (2020). Thermotolerant coral symbionts modulate heat stress-responsive genes in their hosts. *Molecular Ecology, 29*(15), 2940-2950.

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology, 16*(1), 157-157. doi:10.1186/s13059-015-0721-2

Goiran, C., Al-Moghrabi, S., Allemand, D., & Jaubert, J. (1996). Inorganic carbon uptake for photosynthesis by the symbiotic coral/dinoflagellate association I. Photosynthetic performances of symbionts and dependence on sea water bicarbonate. *Journal of Experimental Marine Biology and Ecology, 199*(2), 207-225.

González-Pech, R. A., Bhattacharya, D., Ragan, M. A., & Chan, C. X. (2019). Genome evolution of coral reef symbionts as intracellular residents. *Trends in ecology & evolution*, *34*(9), 799-806.

González-Pech, R. A., Chen, Y., Stephens, T. G., Shah, S., Mohamed, A. R., Lagorce, R., . . . Chan, C. X. (2019). Genomes of Symbiodiniaceae reveal extensive sequence divergence but conserved functions at family and genus levels. *bioRxiv*, 800482.

González-Pech, R. A., Stephens, T. G., Chen, Y., Mohamed, A. R., Cheng, Y., Burt, D. W., . . . Chan, C. X. (2019). Structural rearrangements drive extensive genome divergence between symbiotic and free-living Symbiodinium. *bioRxiv*, 783902.

Grégoire, V., Schmacka, F., Coffroth, M. A., & Karsten, U. (2017). Photophysiological and thermal tolerance of various genotypes of the coral endosymbiont Symbiodinium sp. (Dinophyceae). *Journal of Applied Phycology*, *29*(4), 1893-1905. doi:10.1007/s10811-017-1127-1

Hirt, R. P., Harriman, N., Kajava, A. V., & Embley, T. M. (2002). A novel potential surface protein in Trichomonas vaginalis contains a leucine-rich repeat shared by micro-organisms from all three domains of life. *Molecular and biochemical parasitology*, *125*(1-2), 195-199.

Huang, L., Feng, G., Yan, H., Zhang, Z., Bushman, B. S., Wang, J., . . . Nie, G. (2020). Genome assembly provides insights into the genome evolution and flowering regulation of orchardgrass. *Plant biotechnology journal*, *18*(2), 373-388.

Kim, J., Na, Y. J., Park, S. J., Baek, S.-H., & Kim, D. H. (2019). Biogenesis of chloroplast outer envelope membrane proteins. *Plant cell reports*, 1-10.

Ladner, J. T., Barshis, D. J., & Palumbi, S. R. (2012). Protein evolution in two co-occurring types of Symbiodinium: an exploration into the genetic basis of thermal tolerance in Symbiodiniumclade D. *BMC evolutionary biology*, *12*(1), 217.

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, *6*, 1287-1287. doi:10.12688/f1000research.12232.1

LaJeunesse, T. C., Wham, D. C., Pettay, D. T., Parkinson, J. E., Keshavmurthy, S., & Chen, C. A. (2014). Ecologically differentiated stress-tolerant endosymbionts in

the dinoflagellate genus Symbiodinium (Dinophyceae) Clade D are different species. *Phycologia, 53*(4), 305-319. doi:10.2216/13-186.1

Leggat, W., Hoegh-Guldberg, O., Dove, S., & Yellowlees, D. (2007). Analysis of an EST library from the dinoflagellate (Symbiodinium sp.) symbiont of reef-building corals 1. *Journal of Phycology, 43*(5), 1010-1021.

Lin, S., Cheng, S., Song, B., Zhong, X., Lin, X., Li, W., . . . Ji, Z. (2015). The Symbiodinium kawagutii genome illuminates dinoflagellate gene expression and coral symbiosis. *Science, 350*(6261), 691-694.

Liu, H., Stephens, T. G., González-Pech, R. A., Beltran, V. H., Lapeyre, B., Bongaerts, P., . . . Forêt, S. (2018). Symbiodinium genomes reveal adaptive evolution of functions related to coral-dinoflagellate symbiosis. *Communications biology, 1*(1), 1-11.

Mansour, J. S., Pollock, F. J., Díaz-Almeyda, E., Iglesias-Prieto, R., & Medina, M. (2018). Intra- and interspecific variation and phenotypic plasticity in thylakoid membrane properties across two Symbiodinium clades. *Coral Reefs, 37*(3), 841-850. doi:10.1007/s00338-018-1710-1

Matthews, J. L., Crowder, C. M., Oakley, C. A., Lutz, A., Roessner, U., Meyer, E., . . . Davy, S. K. (2017). Optimal nutrient exchange and immune responses operate in partner specificity in the cnidarian-dinoflagellate symbiosis. doi:10.1073/pnas.1710733114

Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics, 23*(9), 1061-1067. doi:10.1093/bioinformatics/btm071

Pettay, D. T., Wham, D. C., Smith, R. T., Iglesias-Prieto, R., & LaJeunesse, T. C. (2015). Microbial invasion of the Caribbean by an Indo-Pacific coral zooxanthella. *Proceedings of the National Academy of Sciences of the United States of America, 112*(24), 7513-7518. doi:10.1073/pnas.1502283112

Rosset, S., Koster, G., Brandsma, J., Hunt, A., Postle, A., & D'Angelo, C. (2019). Lipidome analysis of Symbiodiniaceae reveals possible mechanisms of heat stress tolerance in reef coral symbionts. *Coral Reefs, 38*(6), 1241-1253.

Silverstein, R. N., Cunning, R., & Baker, A. C. (2017). Tenacious D: Symbiodinium in clade D remain in reef corals at both high and low temperature extremes despite impairment. *Journal of Experimental Biology, 220*(7), 1192-1196.

Tansik, A. L., Fitt, W. K., & Hopkinson, B. M. (2017). Inorganic carbon is scarce for symbionts in scleractinian corals. *Limnology and Oceanography, 62*(5), 2045-2055.

Tchernov, D., Gorbunov, M. Y., De Vargas, C., Yadav, S. N., Milligan, A. J., Häggblom, M., & Falkowski, P. G. (2004). Membrane lipids of symbiotic algae are diagnostic of sensitivity to thermal bleaching in corals. *Proceedings of the National Academy of Sciences, 101*(37), 13531-13535.

Von Loeffelholz, O., Kriechbaumer, V., Ewan, R. A., Jonczyk, R., Lehmann, S., Young, J. C., & Abell, B. M. (2011). OEP61 is a chaperone receptor at the plastid outer envelope. *Biochemical Journal, 438*(1), 143-153.

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., . . . Guo, H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research, 40*(7), e49-e49.

Weis, V. M. (2008). Cellular mechanisms of Cnidarian bleaching: stress causes the collapse of symbiosis. *The Journal of Experimental Biology, 211*(Pt 19), 3059-3066. doi:10.1242/jeb.009597

Wham, D. C., Ning, G., & LaJeunesse, T. C. (2017). Symbiodinium glynnii sp. nov., a species of stress-tolerant symbiotic dinoflagellates from pocilloporid and montiporid corals in the Pacific Ocean. *Phycologia, 56*(4), 396-409.

Wham, D. C., Pettay, D. T., & LaJeunesse, T. C. (2011). Microsatellite loci for the host-generalist "zooxanthella" Symbiodinium trenchi and other Clade D Symbiodinium. *Conservation Genetics Resources, 3*(3), 541-544. doi:10.1007/s12686-011-9399-2

Yellowlees, D., Rees, T. A. V., & Leggat, W. (2008). Metabolic interactions between algal symbionts and invertebrate hosts. *Plant, Cell & Environment, 31*(5), 679-694. doi:10.1111/j.1365-3040.2008.01802.x

Chapter 5: Final conclusions and synthesis

5.1 Conclusions

The purpose of this dissertation was to explore methods of interpartner metabolic signaling in the coral-algal symbiosis. First, I characterized the diversity of the cnidarian insulin-like peptides (ILPs) and explored their transcriptional response to symbiosis, thermal stress, and ocean acidification. Second, I investigated the genome of the thermotolerant coral symbiont *Durusdinium trenchii* to characterize the genomic basis for its contributions to the thermotolerance of the coral holobiont.

My first two data chapters explore the insulin signaling pathway (ISP) in cnidarians, an evolutionarily important signaling pathway that mediates diverse functions including metabolism, growth, development, reproduction, and immunity (Barbieri, Bonafè, Franceschi, & Paolisso, 2003; Pertseva & Shpakov, 2002). In Chapter 2, I first target building a framework for studies into the cnidarian ILPs by elucidating the diversity in ILP structures and sequences as well as their conservation across different taxonomic groups within Cnidaria. Targeting 93 species from 10 Orders across the major Cnidarian Classes of Anthozoa, Hydrozoa, Cubozoa, Scyphozoa, and Staurozoa, we surprisingly recovered 360 putative cnidarian ILPs that were previously un-annotated in their respective studies due to the methods employed. Normally grouped according to the number and location of the conserved cysteine residues where disulfide bonds form bridging the two chains, we identified three distinct structural types of cnILPs. Type A had the canonical insulin motifs of three disulfide bonds and six conserved cysteines while types B and C had two additional B chain cysteines resulting in an additional intra-

B chain disulfide bond. The taxonomic distribution of the cnILPs suggests that there was an expansion and diversification of cnILPs within the class Anthozoa. Specifically, this diversification likely occurred after the split between the sub-classes Hexacorallia and Octocorallia as the Octocorallian order Alcyonacea only possessed two cnILPs in contrast to the eight and seven cnILPs found in the Hexacorallian orders. Specifically, Anthozoans have a greater diversity of the canonical type A insulin structure in addition to possessing four distinct type C ILPs, which was unique to Anthozoa. This greatly expands our understanding of the diversity in non-bilaterian ILPs, which was previously limited to only three in *H. vulgaris* (Steele, Lieu, Mai, Shenk, & Sarras Jr, 1996) and four in *N. vectensis* (Putnam et al., 2007), neither of which are symbiotic. This previously unrecognized diversity of ILPs in non-bilaterians, particularly that of symbiotic corals and anemones, is a prime target for future functional work aimed at further elucidating the symbiotic communication between the two partners.

Following the establishment of this phylogenetic framework, we systematically investigated the transcriptional responses of the previously un-annotated, but conserved cnILPs across cnidarian taxa in response to symbiosis, temperature stress, and ocean acidification. Although the ISP is well-known for its functional importance in stress response, particularly the onset of the dauer larva or diapause-like state in *C. elegans* (Riddle, Blumenthal, Meyer, & Priess, 1997) and *D. melanogaster* (DiAngelo, Bland, Bambina, Cherry, & Birnbaum, 2009), respectively, it has also been implicated in symbiotic communication. Research in the fruit fly and mosquito have suggested that translocated products from an endosymbiont can manipulate host metabolic pathways via the ISP and, in turn, suppress the host's immune system enabling its persistence

(Marquez et al., 2011; Negri, 2012; Pietri et al., 2016; Pietri, Pietri, Potts, Riehle, & Luckhart, 2015). Significantly, we identified a strong trend for cnILP down-regulation in response to symbiosis, particularly that of cnILP-B of which a single peptide is found in every cnidarian Class except Hydrozoa. Under thermal stress we identified baseline patterns of a non-specific down-regulation of cnILPs in more thermally resilient coral populations with increases in cnILP expression when putting forth a response to thermal stress. The connection between decreased ISP activity and increased longevity and stress resistance is not a new topic and has been previously described all the way from the nematode *C. elegans* to humans (Barbieri et al., 2003). This work for the first time highlights not only an extensive cnILP repertoire in cnidarians, but also demonstrates the presence of comparable transcriptional responses in cnidarian ILPs as seen in other invertebrates as well as vertebrates. Future work is essential for gaining functional and mechanistic understandings of how this evolutionarily important pathway functions in cnidarians.

Chapter 4 investigates for the first time the genome of a thermotolerant symbiont of corals, specifically the ecologically important *Durusdinium trenchii*. Unique compared to its counterparts within the *Durusdinium* genus in regards to its broad geographic distribution and host generalist nature, it has been shown in numerous studies to confer increased resistance to thermal stress on its coral hosts (Silverstein, Cunning, & Baker, 2017). Previous microsatellite studies have implicated a whole- or near-whole genome duplication event over the course of this species' evolution, hypothesizing that extensive duplication followed by diversification likely contributed to its thermotolerance (Wham, Ning, & LaJeunesse, 2017). In this work we used 10X Genomics linked-read sequencing

to capture this whole genome duplication and investigate whether there is genomic evidence of duplication as well as selection on functions that have contributed to the greater thermotolerance of *D. trenchii*. We identified extensive duplication of collinear gene blocks of five or more genes conserved in the exact same order in more than one genomic location. Significantly, we show that level of duplication of these collinear gene blocks was 10-100x greater than any other available genomes from species in the order Suessiales. Further, upon examination of these duplicated gene blocks for evidence of positive selection, we identified pervasive duplication and subsequent positive selection on genes related to thylakoid membrane integrity, the photosynthetic apparatus, and the mitigation of oxidative stress damage, all of which have been implicated as major factors in varying thermotolerance among Symbiodiniaceae species (Tchernov et al., 2004; Weis, 2008). Our study suggests that large-scale gene duplication followed by positive selection and the fixation of advantageous variants of these duplicated genes, largely which are related to photosynthesis, likely contributed to the acquisition of thermal tolerance in *D. trenchii*, and thus contributed to the ecological significance of this species.

Overall, my dissertation work has shown that there is an extensive diversity heretofore unknown in the non-bilaterian phylum Cnidaria, particularly symbiotic anemones and reef-building corals. It characterizes the varying diversity in cnidarian ILPs and their conserved transcriptional responses across different taxonomic groups within Cnidaria. Specifically, it identifies a strong trend for down-regulation, particularly in the type B cnILP structure, in response to symbiosis as well as non-specific down-regulation of cnILP expression in more resilience groups of corals. Most significantly, the identified

transcriptional response of the cnILPs align with previous studies spanning invertebrates and vertebrates, including humans. Finally, my dissertation work also yielded the first genome of a thermotolerant coral symbiont, *Durusdinium trenchii*. Its genome confirms previous reports of extensive genomic duplication and reveals this large scale duplication was followed by extensive positive selection on genes important to photosynthesis, and is likely a central factor that resulted in the thermotolerance of *D. trenchii*.

5.2 References

Barbieri, M., Bonafè, M., Franceschi, C., & Paolisso, G. (2003). Insulin/IGF-I-signaling pathway: an evolutionarily conserved mechanism of longevity from yeast to humans. *American Journal of Physiology, Endocrinology, and Metabolism, 285*(5), E1064-1071. doi:10.1152/ajpendo.00296.2003

DiAngelo, J. R., Bland, M. L., Bambina, S., Cherry, S., & Birnbaum, M. J. (2009). The immune response attenuates growth and nutrient storage in *Drosophila* by reducing insulin signaling. *Proceedings of the National Academy of Sciences of the United States of America, 106*(49), 20853-20858. doi:10.1073/pnas.0906749106

Marquez, A. G., Pietri, J. E., Smithers, H. M., Nuss, A., Antonova, Y., Drexler, A. L., . . . Luckhart, S. (2011). Insulin-like peptides in the mosquito *Anopheles stephensi*: Identification and expression in response to diet and infection with *Plasmodium falciparum*. *General and Comparative Endocrinology, 173*(2), 303-312. doi:10.1016/j.ygcen.2011.06.005

Negri, I. (2012). *Wolbachia* as an “infectious” extrinsic factor manipulating host signaling pathways. *Frontiers in Endocrinology, 2*, 115-115. doi:10.3389/fendo.2011.00115

Pertseva, M., & Shpakov, A. (2002). Conservatism of the insulin signaling system in evolution of invertebrate and vertebrate animals. *Journal of Evolutionary Biochemistry and Physiology, 38*(5), 547-561.

Pietri, J. E., Pakpour, N., Napoli, E., Song, G., Pietri, E., Potts, R., . . . Luckhart, S. (2016). Two insulin-like peptides differentially regulate malaria parasite infection in the mosquito through effects on intermediary metabolism. *Biochemical Journal, 473*(20), 3487-3503. doi:10.1042/BCJ20160271

Pietri, J. E., Pietri, E. J., Potts, R., Riehle, M. A., & Luckhart, S. (2015). *Plasmodium falciparum* suppresses the host immune response by inducing the synthesis of insulin-like peptides (ILPs) in the mosquito *Anopheles stephensi*. *Developmental & Comparative Immunology, 53*(1), 134-144. doi:10.1016/J.DCI.2015.06.012

Putnam, N. H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., . . . Kapitonov, V. V. (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *science, 317*(5834), 86-94. doi:10.1126/science.1139158

Riddle, D. L., Blumenthal, T., Meyer, B., & Priess, J. (1997). Genetic and environmental regulation of dauer larva development. In *C. elegans II. 2nd edition* (ed.). Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Silverstein, R. N., Cunning, R., & Baker, A. C. (2017). Tenacious D: *Symbiodinium* in clade D remain in reef corals at both high and low temperature extremes despite impairment. *Journal of Experimental Biology, 220*(7), 1192-1196. doi:10.1242/jeb.148239

Steele, R., Lieu, P., Mai, N. H., Shenk, M. A., & Sarras Jr, M. P. (1996). Response to insulin and the expression pattern of a gene encoding an insulin receptor homologue suggest a role for an insulin-like molecule in regulating growth and patterning in *Hydra*. *Development genes and evolution, 206*(4), 247-259. doi:10.1007/s004270050050

Tchernov, D., Gorbunov, M. Y., De Vargas, C., Yadav, S. N., Milligan, A. J., Häggblom, M., & Falkowski, P. G. (2004). Membrane lipids of symbiotic algae are diagnostic of sensitivity to thermal bleaching in corals. *Proceedings of the National Academy of Sciences, 101*(37), 13531-13535. doi:10.1073/pnas.0402907101

Weis, V. M. (2008). Cellular mechanisms of Cnidarian bleaching: stress causes the collapse of symbiosis. *The Journal of Experimental Biology, 211*, 3059-3066. doi:10.1242/jeb.009597

Wham, D. C., Ning, G., & LaJeunesse, T. C. (2017). *Symbiodinium glynnii* sp. nov., a species of stress-tolerant symbiotic dinoflagellates from pocilloporid and montiporid corals in the Pacific Ocean. *Phycologia, 56*(4), 396-409. doi:10.2216/16-86.1

VITA

KATHERINE E. DOUGAN

2014	B.S., Microbiology University of Michigan Ann Arbor, Michigan
2011-2014	Research Assistant Supervisor: Dr. Alison Gould The University of Michigan Ann Arbor, Michigan
2015-2019	Teaching Assistant Florida International University Miami, Florida
2018	Tropics Program Research Fellow Florida International University Miami, Florida
2019	M.S., Biology Florida International University Miami, Florida
2019	Dissertation Year Fellow Florida International University Miami, Florida

PUBLICATIONS

C Iha, KE Dougan, JA Varela, V Avila, CJ Jackson, KA Bogaert, Y Chen, LM Judd, R Wick, KE Holt, MM Pasella, F Ricci, SI Repetti, M Medina, VR Marcelino, CX Chan, H Verbruggen (2021). Genomic adaptations to an endolithic lifestyle in the coral-associated alga *Ostreobium*. Current Biology 31, 1-10. DOI: 10.1016/j.cub.2021.01.018

KE Dougan, MC Ladd, C Fuchs, R Vega Thurber, DE Burkepile, M Rodriguez-Lanetty (2020). Nutrient pollution and predation differentially affect innate immune pathways in the coral *Porites porites*. *Frontiers in Marine Science* 7:563865. DOI: 10.3389/fmars.2020.563865

AJ Bellantuono*, KE Dougan*, C Granados-Cifuentes, M Rodriguez-Lanetty (2019). Free-living and symbiotic lifestyles of a thermotolerant coral endosymbiont display

profoundly distinct transcriptomes under both stable and heat stress conditions. *Molecular Ecology*, 28(24):5265-5281. DOI: 10.1111/mec.15300

G Bonthond, DG Merselis, KE Dougan, T Graff, W Todd, JW Fourqurean, M Rodriguez-Lanetty (2018). Inter-domain microbial diversity within the coral holobiont *Siderastrea siderea* from two depth habitats. PeerJ 6:e4323. DOI: 10.7717/peerj.4323

TA Hendry, J.R. de Wet, KE Dougan, and PV Dunlap (2016). Genome evolution in the obligate but environmentally active luminous symbionts of flashlight fish. *Genome Biology and Evolution*, 8:2203-2213. DOI: 10.1093/gbe/evw161

AL Gould, KE Dougan, S Koel, PV Dunlap (2016). Life history of the symbiotically luminous cardinal fish *Siphamia tubifer* (Perciformes: Apogonidae) in Okinawa, Japan. *Journal of Fish Biology,* 89:1359-1377. DOI: 10.1111/jfb.13063

*authors contributed equally