Viruses, Corals and the Origin of Metazoans

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Biology

by

Steven Douglass Quistad

Committee in charge:

University of California, San Diego

Professor Stephen Hedrick
Professor Martin Tresguerres

San Diego State University

Professor Forest Rohwer, Chair
Professor Robert Edwards
Professor Roland Wolkowicz

2015
The Dissertation of Steven Douglass Quistad is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego
San Diego State University
2015
DEDICATION

This dissertation is dedicated to the giants who’s shoulders I stand on today.
EPIGRAPH

Roll on, deep and dark blue ocean, roll. Ten thousand fleets sweep over thee in vain. Man marks the earth with ruin, but his control stops with the shore.

-Lord Byron
### TABLE OF CONTENTS

Signature Page ........................................................................................................ iii

Dedication................................................................................................................ iv

Epigraph .................................................................................................................. v

Table of Contents ................................................................................................... vi

List of Figures ......................................................................................................... x

List of Tables ........................................................................................................... xiii

Acknowledgements ............................................................................................... xiv

Vita ........................................................................................................................... xvii

Abstract of the Dissertation .................................................................................... xviii

#### Chapter 1 Introduction....................................................................................... 1
  The Cambrian Explosion and the birth of metazoans ........................................ 2
  The unexpected complexity of cnidarians ............................................................. 2
  Coral viruses .......................................................................................................... 3
  References .............................................................................................................. 4

#### Chapter 2 Using viral populations to predict immune system structure
  in uncharacterized systems ..................................................................................... 7
  Abstract................................................................................................................. 7
  Introduction .......................................................................................................... 9
  Results .................................................................................................................. 12
    Extraction of viral DNA and bioinformatic analysis ........................................ 12
    Construction of the predicted viral interactome ................................................. 13
    Bioinformatic analysis of select viral sequences matching to host proteins ... 16
    Expression of predicted immune proteins from virome analysis .................... 18
  Discussion .......................................................................................................... 23
Chapter 3 Evidence of an ancient retroviral infection of a cnidarian ancestor

Abstract............................................................................................................. 41
Introduction ......................................................................................................... 42
Results .................................................................................................................. 45
Identification of a novel ERV family in Acropora digitifera............................. 45
Distribution of the ERV family across phyla and viromes.............................. 49
LTR as a promoter element .............................................................................. 52
Discussion .......................................................................................................... 57
Summary ............................................................................................................ 57
Fibroblast Growth Factor Receptors (FGFRs) ................................................. 57
PBX1 transcription factors and the Cambrian Explosion............................... 58
Precambrian origin of ERV family ................................................................. 59
ERVs, GRNs, and metazoan evolution ............................................................. 60
Predictions of ERV-mediated metazoan evolution ........................................ 60
Viral addiction and metazoan evolution ......................................................... 61
Conclusions ....................................................................................................... 63
References .......................................................................................................... 64
Acknowledgments ............................................................................................. 68
Appendix ............................................................................................................ 69

Chapter 4 Evolution of TNF-induced apoptosis reveals 550 million years of functional conservation................................................................. 76
Abstract............................................................................................................. 76
Introduction ......................................................................................................... 77
Results .................................................................................................................. 81
Bioinformatic analysis of the Acropora digitifera apoptotic repertoire reveals high-predicted conservation .............................................................. 81
HuTNFα causes apoptosis in Acropora yongei ............................................... 84
HuTNFα causes myosin fragmentation and results in an acidic shift in the coral proteome ................................................................. 86
HuTNFα causes bleaching of Acropora yongei ........................................ 88
Crude AdTNF1 causes apoptosis in coral and is involved with bleaching .................................................................................. 88
Purified Coral AdTNF1 causes apoptosis in human T-Lymphocytes .............................................................. 89
Discussion ........................................................................................................... 92
Materials and Methods ....................................................................................... 97
Bioinformatic analysis of the TNFRSF and TNFSF in Acropora digitifera ................................................................. 97
Human antibody arrays ..................................................................................... 97
2D-Gel Electrophoresis of A. yongei stimulated with HuTNFα ................................................................. 98
Production of crude AdTNF1 utilizing a Prolactin secretion scaffold ................................................................................... 99
Direct stimulation of coral cells with HuTNFα and crude AdTNF1 ................................................................................ 100
Immunohistochemistry of coral cells exposed to HuTNFα and crude AdTNF1-FLAG ................................................................. 101
Caspase activity of coral cells exposed to HuTNFα ........................................ 102
Flow cytometric analysis of coral bleaching .................................................. 102
Flow cytometric analysis of Jurkat cell viability ............................................. 103
References ........................................................................................................ 105
Acknowledgments ........................................................................................... 110
Appendix ........................................................................................................... 111

Chapter 5 Viruses and the origin of metazoan immunity ........................................... 120
Abstract ......................................................................................................... 120
Summary ........................................................................................................ 121
Introduction .................................................................................................... 123
Viruses and the origin of TNF-induced apoptosis ......................................... 124
Apoptosis and multicellularity ....................................................................... 124
TNFRs and viral addiction .............................................................................. 126
Phage: An ancient metazoan selective-microbial ....................................... 129
The ubiquitous mucosal environment ........................................................... 129
Niche exclusion and immunology ................................................................. 130
BAM: An ancient metazoan selective microbial ........................................... 131
Conclusions .................................................................................................... 134
Future perspectives ......................................................................................... 134
References ...................................................................................................... 136
Acknowledgments .......................................................................................... 141
LIST OF FIGURES

Figure 2.1: Bioinformatic pipeline used to identity viral homologs to predicted coral immune proteins ................................................................. 13

Figure 2.2: The predicted coral-viral interactome ............................................. 14

Figure 2.3: Viral sequences matching to coral immune proteins .................... 15

Figure 2.4: Primary sequence alignments and GC-content of translated viral gene segments matching to predicted coral immune proteins ........................................ 17

Figure 2.5: Differentially expressed genes predicted from viral populations ........ 19

Figure 2.6: Localization of viral hits to unknown coral proteins ...................... 21

Figure 2.7: General pipeline to characterize unknown immune systems ........... 22

Supplementary Figure 2.1: Porites rus tissue used for virome extraction ........... 37

Supplementary Figure 2.2: 18S and 16S PCR of viromes before library preparation 37

Supplementary Figure 2.3: Coral Nod-Like Receptors (NLRs) with the total number of viral hits indicated to the left of each protein ........................................... 38

Supplementary Figure 2.4: Herpes proteins with human homologs categorized into cellular function .............................................................................. 38

Figure 3.1: Bioinformatic pipeline used to identify novel ERV family .............. 46

Figure 3.2: ERV family characterization in Acropora digitifera .......................... 48

Figure 3.3: Phylogenetic tree of the LTR element in genomes and viral metagenomes ......................................................................................... 50

Figure 3.4: A proposed model for an ERV infection of the Proto-cnidarian ancestor 51

Figure 3.5: Predicted PBX1 binding site within LTR region and primary sequence alignment of the PBX1 protein across metazoans ........................................... 53

Figure 3.6: The location of the LTR element within genomic scaffolds containing Fibroblast Growth Factor Receptors (FGFRs) in three cnidarians .......... 55
Figure 3.7: Proposed model of ERV-mediated metazoan evolution

Supplementary Figure 3.1: Representative primary sequence alignments of 54 amino acid motif extracted from the LTR element

Supplementary Figure 3.2: Abundance of the LTR element among host genomes and viromes

Figure 4.1: Bioinformatic analysis of coral and human proteins involved with Death Receptor Signaling

Figure 4.2: Effect of HuTNFα on Acropora yongei cell populations

Figure 4.3: Proteomic analysis of coral exposed to HuTNFα using Two-Dimensional Gel Electrophoresis

Figure 4.4: Effect of AdTNF1 on Human T-Lymphocyte cellular viability

Supplementary Figure 4.1: Primary sequence alignments of Acropora digitifera and Homo sapiens Death Receptor Pathway proteins

Supplementary Figure 4.2: Phylogenetic trees of Acropora digitifera and Homo sapiens Death Receptor Pathway proteins

Supplementary Figure 4.3: HuTNFα affects the expression of unknown coral proteins bound to human antibodies involved with Death Receptor Signaling

Supplementary Figure 4.4: Proteomic analysis of coral exposed to HuTNFα using human antibody arrays

Supplementary Figure 4.5: The effect of HuTNFα on coral bleaching

Supplementary Figure 4.6: The physiological and cellular responses of coral to crude AdTNF1

Supplementary Figure 4.7: Live/dead cell assay of Jurkat cells exposed to FasL and purified AdTNF1
Figure 5.1: A proposed model for the development of metazoan immunity............. 122

Figure 5.2: The TNFR-FADD-c-FLIP addiction module. ............................................. 127

Figure 5.3: Genomic scaffolds taken from *Acropora digitifera* expressing a c-FLIP homolog with viral-domains indicated. ................................................................. 128

Figure 5.4: The Bacterial Adherence to Mucus (BAM) model of immunity .......... 133
LIST OF TABLES

Supplementary Table 2.1: Pre-processing viromes .................................................. 39
Supplementary Table 2.2: Viral sequences matching to coral immune proteins .... 39
Supplementary Table 2.3: Transmembrane region prediction in uncharacterized coral proteins .................................................................................................................................................. 40
Supplementary Table 3.1: Transcription features of novel LTR located within the Acropora digitifera genome .......................................................................................................................................................... 72
Supplementary Table 3.2: Distribution of LTR across viromes and genomes .......... 72
Supplementary Table 3.3: Representative ERV genome located within scaffold 1125 of the Acropora digitifera genome .......................................................................................................................................................... 73
Supplementary Table 3.4: MEME analysis of the 54 amino acid motif located within LTR region .................................................................................................................................................................. 74
Supplementary Table 3.5: Cnidarian genome scaffolds that contain both the novel LTR and genes encoding for Fibroblast Growth Factor Receptors (FGFRs) .......... 74
Supplementary Table 3.6: Additional genes that are associated with the novel LTR identified here in all three cnidarians investigated ............................................................................................................. 75
Supplementary Table 4.1: Previously identified TNF receptor/ligand family members across evolution .................................................................................................................................................. 118
Supplementary Table 4.2: Proteins identified through mass spectrometry with the best hit to either the Acropora digitifera or Symbiodinium proteome ......................... 119
ACKNOWLEDGEMENTS

None of this work would have been possible without the intellectual guidance and once-in-lifetime research opportunities provided to me by my mentor Dr. Forest Rohwer. Through his support I was given the opportunity to experience some of the most remote regions on this planet from tropical coral atolls to the frigid Russian Arctic where both of us decided that iceberg diving was not for us. These experiences further developed my intense desire to understand how the world works from both a scientific and cultural perspective. The most important skill Forest has taught me is the process of searching for interesting scientific questions to investigate while remaining humble in any success that may be achieved.

During my dissertation work I have had the honor to interact with a broad group of extremely talented individuals. In particular I would like to thank Gina Spidel and Leslie Rodelander for their friendship and administrative support during my time in the Rohwer Lab. I can say with confidence that we would not have made it to Russia (and returned in one piece) without both of you. I would also like to thank Drs. Linda Wegley Kelly and Katie Barott for their scientific mentorship and friendship. I thank Drs. Roland Wolkowicz and Robert Edwards for providing me with the opportunity to rotate in your labs. Both experiences provided me with unique skill sets that directly shaped my dissertation work. To all of the Rohwer Lab members over the years I thank you for supporting a collaborative and dynamic working environment that always kept work interesting. For financial support I acknowledge the National Science Foundation as well as the ARCS foundation.
While in the Rohwer Lab I have had the unique opportunity to participate in multiple research expeditions for which I am extremely grateful. For the 2011 RAMPS cruise to Wake Island I thank Dr. Jamison Gove, Chip Young, and NOAA; despite a tsunami we were able to still complete all of our diving. For work performed in Moorea, French Polynesia I thank Drs. Craig Nelson, Andi Haas, and Linda Wegley Kelly as well as the Richard B. Gump Field Station; the chow mein baguette sandwich will always hold a special place in my heart. For the opportunity to explore the most northern archipelago in the world, Franz Josef Land, I thank the expedition leaders Drs. Enric Sala and Maria Gavrilo, Daria Martynova for her copepod enthusiasm, the entire PristineSeas team, and the National Geographic Society.

Lastly I thank my friends and family. To Billy I thank you for your support while writing this dissertation and look forward to our future adventures. To my parents Virginia and Gary thank you for supporting the curiosities of an overly energetic child and to my siblings Kelly, Kerry, and David thanks for putting up with my science lab and planting machine.

Chapter 2, in full, is in preparation for submission. Steven D. Quistad, Yanwei Lim, Genivaldo Gueiros Z. Silva, Craig E. Nelson, Andreas F. Haas, Linda Wegley Kelly, Robert A. Edwards and Forest Rohwer; 2015. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is under review in the Proceedings of the National Academy of Sciences. Steven D. Quistad, Aaron C. Hartmann, Linda Wegley Kelly, and Forest
Rohwer; 2015. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, is a reprint of the material as it appears in the Proceedings of the National Academy of Sciences. Steven D. Quistad, Aleksandr Stotland; Katie L. Barott, Cameron Smurthweites, Brett Hilton, Juris Grasis, Roland Wolkowicz, and Forest Rohwer; 2014. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, in preparation for submission. Steven D. Quistad, and Forest Rohwer; 2015. The dissertation author was the primary investigator and author of this paper.
VITA

2015  Doctor of Philosophy in Biology, University of California at San Diego and San Diego State University (Advisor: Dr. Forest Rohwer)

2010  Bachelor of Arts in Biology, University of California Santa Barbara

2009 - 2010  Research Assistant Marine Science Institute, University of California Santa Barbara (Advisor: Dr. David Valentine)

2008  Research Assistant, Marine Science Institute, University of California Santa Barbara (Advisors: Dr. Michael O’Donnell (post doc), Dr. Gretchen Hofmann (PI))

2007  Research Assistant, Neuroscience Research Institute, University of California Santa Barbara, (Advisor: Dr. Stuart Feinstein)


PUBLICATIONS


Quistad SD, AC Hartmann, Wegley Kelley L, and Rohwer F. Viral expansion of Gene Regulatory Networks: A proposed model of animal diversification (in revision).


Quistad SD, and Rohwer F. Viruses and the origin of metazoan immunity (in preparation)
Viruses, Corals and the Origin of Metazoans

by

Steven Douglass Quistad

Doctor of Philosophy in Biology

University of California, San Diego, 2015
San Diego State University, 2015

Professor Forest Rohwer, Chair

Viruses drive the evolution of animal life through direct manipulation of host biology. The objective of this dissertation was to understand how the viruses associated with reef-building corals interact with host cellular processes in order to provide insight into the evolution of Metazoans. One mechanism viruses use to manipulate their host involves the expression of viral homologs to host immune proteins. To determine whether viral populations could be used to predict immune system structure a bioinformatic approach was utilized. Viral sequences were
compared to the host proteome using BLASTx revealing viral homologs to host proteins involved with cell cycle control, transcription, and immunity. Viruses can also impact the evolution of host genomes through the donation of genetic material. Here I provide evidence of an ancient endogenous retroviral infection of the proto-cnidarian ancestor. This novel ERV family is associated with Fibroblast Growth Factor Receptors (FGFRs) in all three cnidarians investigated suggesting that it is involved with cnidarian-specific developmental traits. Viruses affect host immunity and genome structure however hosts can resist viral invasion through programmed cell death or apoptosis. The Tumor Necrosis Factor Receptor ligand Superfamily (TNFRSF) is a central mediator of apoptosis in vertebrates, though previously uncharacterized in cnidarians. Bioinformatic analysis revealed that corals possess more TNFR’s than any organism described thus far, including humans. Furthermore, corals harbor all of the central components of the canonical death receptor pathway. Empirical studies demonstrated that Human TNFα could bind to coral cells, induce apoptotic blebbing, and cause coral bleaching. The reciprocal experiment where human cells were exposed to a coral TNF also resulted in apoptosis. These results demonstrate remarkable conservation of the TNF-apoptotic response that has been functionally maintained for over 550 million years. In summation this dissertation work provides novel insight into cnidarian biology with broad application to the origin of metazoan diversity.
CHAPTER 1

Introduction: Corals, viruses and explosion of animal life

Abstract

Viruses have driven the evolution of metazoan life through persistent interactions between viral evasion tactics and the host immune response. As obligate intracellular parasites they are unable to replicate independently of a host cell. Mechanisms of viral-driven evolution are well established in vertebrates however viruses have yet to be placed within the context of the Cambrian Explosion and the origin of metazoans. This work utilizes reef-building corals as a representative organism of the Precambrian phyla Cnidaria to explore how viruses have driven the evolution of animal life. Chapter 2 focuses on understanding how viruses have driven the evolution of immunity in coral and demonstrates how viral populations can be used to predict the structure of uncharacterized immune systems. In Chapter 3 the impact of viruses on the evolution of host genomes is explored and evidence for an ancient retroviral infection of a proto-cnidarian ancestor is presented. In Chapter 4 molecular techniques were employed to understand the origins of the antiviral response of programmed cell death (TNF-induced apoptosis) revealing over 550 million years of functional conservation. Finally in Chapter 5 a model is presented that describes the origin and development of the first metazoan immune system driven by microbial consortiums and viral-addiction mechanisms.
The Cambrian Explosion and the birth of metazoans: The Cambrian Explosion (CEx) occurred approximately 550 million years ago and is referred to as the “Big Bang” of life on Earth during which the majority of extant phyla first appeared in the fossil record. The CEx serves as a dividing landmark in the history of life on earth separating the Pre-Cambrian microscopic era from the period of visible life [1]. While the sudden increase in morphological complexity is well documented in the fossil record the triggers that led to the CEx are a major source of debate [1-3]. Environmental triggers that have been proposed include changes in atmospheric oxygen [4,5], ocean chemistry, and tectonic shifts [6], while intrinsic triggers include genetic exchange and predation pressure [1,7,8]. While many triggers have been proposed the genetic mechanisms that led to the CEx are not well understood. In Chapter 3 of this work, Cnidaria, an ancient phyla that radiated during the CEx, is used to propose a molecular mechanism driven by the expansion of viruses which contributed to the CEx and the diversity of metazoan life.

The unexpected complexity of cnidarians: The phylum Cnidaria consists of over 10,000 species that range in size from a few millimeters to over 75 meters [9]. Their simple body plan consists of two cell layers, an endoderm and ectoderm, held together by the jelly-like mesoglea. The phylum has been traditionally viewed as “primitive” and therefore the Cnidarian immune system has been posited to be less complex than those of “higher” [10]. However with the sequencing of the [11] Nematostella vectensis [12], Hydra magnipapillata [11], and Acropora digitifera [13]
genomes, as well as preliminary functional studies, the “primitive-immune system” hypothesis is no longer supported [14-19]. Despite their morphological simplicity Cnidarians are complex at the molecular level, even surpassing humans in some instances. For example genomic data suggests that corals possess more NOD-like receptors (NLR’s) than any organism described thus far (~500) with many novel domain combinations. In addition the extracellular detection and intracellular signaling of bacteria via Toll-Like Receptors is functionally conserved in Hydra [20]. This dissertation focuses on reef-building corals (order Scleractinia) responsible for supporting the most biologically diverse ecosystem on the planet: the coral reef [21].

Coral viruses: The diversity of viruses that infect corals is posited to include three distinct viral groups 1) those that infect the coral, 2) the endosymbiotic algae (Symbiodinium spp.), and 3) the other microorganisms associated with the holobiont. These viral groups commonly include representatives from Herpesviridae, Iridoviridae, Phycodnaviridae, archael viruses and bacteriophage. To date only 15 published studies have evaluated the viruses associated with coral, recently reviewed by Thurber and Correa [22]. These studies have focused on morphological classification/direct counts of Virus-Like Particles [23,24] and characterization of viral diversity through metagenomic analyses [25-27]. It has also been demonstrated that physiological stress increases viral production [25,28]. However the mechanistic interactions between coral viruses and host biology were unknown prior to the work presented here.
References


CHAPTER 2

Using viral populations to predict immune system structure in uncharacterized systems

Abstract

Viruses have influenced the evolution of metazoan life through direct interactions with host cell biology. Virally-expressed homologs to host proteins are able to manipulate the intracellular environment in favor of viral production or quiescence. Viral-host homologs can be classified into four functional groups based on their involvement with nucleic acid replication, protein synthesis, cell-cycle control, and immune modulation. Here we hypothesized that viral populations could be used as predictive tools of host immune structure. To test this hypothesis viral-host interactions were investigated in reef-building corals, a non-model organism that currently lacks traditional molecular tools such as transgenic animals, gene knock-in/out capabilities and in vitro cell cultures. DNA viruses were isolated from coral tissue using cesium chloride ultracentrifugation and nucleic acid was amplified using a Linker Amplified Shotgun Library (LASL) approach and sequencing. Viral gene segments were then compared to the coral proteome using BLASTx analyses revealing homologs to coral proteins predicted to be involved with nucleic acid replication, protein synthesis, cell-cycle control, and immune modulation including viral homologs to TRAF6, TRAF4, RIG-I, and RET-proto oncogene. To determine whether any of the host proteins with viral homology were actually expressed publicly available
transcriptomes were examined. Thirty-one genes that were naively predicted by the coral viral populations were confirmed to be differentially expressed under physiological stress. The predicted coral-viral interactome identified here provides a representative example of the utility in using viral populations to bioinformatically predict viral-host interactions. This pipeline provides a novel method to rapidly generate hypotheses relating to viral-host interactions and can be applied to virtually any system with the potential to discover novel immune components.
Introduction

The evolution of multicellular life on the Earth has been driven by intense competition and cooperation between organisms for approximately 2.1 billion years [1]. Survival depends on managing these selective pressures through both self-defense and persistence with non-self entities. This multi-billion year struggle between organisms contributed to both the morphological and immunological complexity of modern day animal life. Comparative immunology utilizes a variety of invertebrate and vertebrate species to better understand the evolution and origin of immunity. Many of these organisms have provided fundamental insights into evolutionarily conserved immune mechanisms. For example, the initial concept of self vs. non-self recognition was formed by experimentation with starfish and studies in chickens led to the identification of separate B and T cell lineages [2,3]. While these discoveries have been invaluable in understanding immune mechanisms the majority of comparative immunology is based on data from only 3 of the 30 extant animal phyla [2]. A broader representation of phyla is required to fully appreciate the complexity and origin of immunity [2].

Studies of *Drosophila melanogaster* began over a century ago with Thomas Morgan and the fruit fly remains one of the most studied organisms today. The continued success of *D. melanogaster* can be attributed to the availability of refined molecular techniques that support a rapid and reproducible experimental system. These techniques provide the ability to knock in and knock out genes of interest to determine overall protein function [4,5]. However, establishing these types of model
organisms can take years to develop and is simply not feasible for the entire metazoan tree of life. For example, the establishment of continuous cell lines has been a long-term goal of coral reef biologists for decades. While there have been successful reports of primary cell cultures standardized cell lines have yet to be established and genetic manipulation of any coral species remains unreported [6-9]. To investigate the immune systems of corals, as well as the remaining twenty-seven animal phyla a broad and rapid approach is required. Here we propose one such approach in which host viral populations are used to predict immune system structure.

Viruses are unable to replicate independently of a host cell therefore the persistence of a viral species is dependent on its ability to replicate and package their genomes while avoiding host immune detection. To establish and maintain control viruses will often express proteins that manipulate central cellular processes such as transcription, translation, cellular proliferation, and host immunity. The focus of this paper will be on the viral proteins that modulate the host immune response. As a representative example three host-modulating proteins expressed by Human Herpesvirus 8 (HHV-8) will be discussed [10]. Interferons are cytokines involved with the canonical antiviral response and their activation is dependent on transcription factors called Interferon Regulatory Factors (IRFs). HHV-8 encodes 4 IRFs (vIRF1-vIRF4) that block the transcription of Interferon-inducible genes resulting in the attenuation of the host antiviral response [11]. At the translational level HHV-8 expresses the endonuclease SOX that degrades all host mRNA within the cytoplasm providing the virus with the translational machinery necessary for viral protein
production [12]. Finally HHV-8 maintains viral control of host cell cycle using a homolog to Cyclin-D; a major driver of the G1/S phase transition [13,14]. Beyond these three specific examples over 50 % of all HHV-8 gene products are involved with host immune modulation and many of these modulators are homologs to host immune proteins [15]. Therefore, an in silico comparison of all HHV-8 gene products and the human proteome would produce a basic structure of the human antiviral response based on homology between viral and host proteins. If this in silico approach is expanded to include all viruses that infect a particular host a more robust prediction of host immune structure can be made without any a priori knowledge of host biology. A global comparison of host and viral gene products would be expected to identify two broad groups of homologous proteins; those involved with viral replication (i.e. transcription and translation) and those involved with host-modulation. The underlying hypothesis of this paper is that if a virus possesses a homolog to a host protein that is not involved with transcription or translation then that host protein is involved with viral control of cell biology. These host-modulators could be further divided into two functional groups; proteins that manipulate cellular proliferation and those that mimic host immune components.

Here we use viral populations taken from a reef-building coral as a representative example to apply a novel bioinformatic approach to characterize a relatively unknown immune system. This approach compares the predicted viral gene products to the host protein repertoire using BLASTx and requires only two components; first the isolation of virus-like particles (VLPs) from host tissue followed
by the sequencing of the host transcriptome. Methods to obtain these components are well established and can be successfully extracted from virtually any tissue type [16]. Taken together this method generates predicted host-virus interactions that can be further tested using traditional biochemical approaches with the potential to discover new types of immunology.

Results

*Extraction of viral DNA and bioinformatic analysis:* DNA was extracted from Virus-Like Particles taken from *Porites rus* tissue with minimal 16S rRNA and 18S rRNA contamination (Supplementary Figures 2.1 and 2.2). Following quality control with PRINSEQ and DeconSeq 495,953 segments of DNA remained with average size of 311 base pairs (Supplementary Table 2.1) [17,18]. Comparison of the combined virome to the predicted coral proteome [19] using BLASTx and an e-value cut off of < 1 E^{-04} resulted in 2,503 coral protein “hits” each matching a minimum of five viral sequence fragments comprising a total of 39,543 viral sequence fragments (Figure 2.1) [20].
Figure 2.1: Bioinformatic pipeline used to identify viral homologs to predicted coral immune proteins. Coral viromes extracted from Porites rus were compared to the Acropora digitifera proteome through BLASTx analysis. Viral sequence fragments gene segments that aligned to predicted coral protein with an e-value < 1E-04 were considered matches. Representative viral gene segments matching predicted coral immune proteins were further analyzed for GC-content. Coral proteins were classified as “hits” if greater than 4 unique viral sequences matched with e-value < 1E-04 and subsequently analyzed through BLASTp analysis against the human proteome as well as the Conserved Domain Database (CDD).

Construction of the predicted viral interactome: Predicted coral protein hits were further analyzed using the Conserved Domain Database (CDD) revealing multiple immune-related domains (Figures 2.2, and Supplementary Table 2) [21]. Compared to the coral proteome (23,677 proteins) the 2,503 coral proteins with matches to viral sequences were enriched for immune-related domains (Figure 2.3).
Figure 2.2: The predicted coral-viral interactome. Numbers located to the left of parentheses indicate the total number of predicted coral protein hits within a functional group that had viral sequence matches while numbers within parentheses represent the total number of viral sequences matches to that functional group of proteins. Functional groups of coral proteins include Interferon-Binding (IFN-Binding), Toll-Interacting Receptor (TIR), Tumor Necrosis Factor-Receptors (TNFR), Tumor Necrosis Factor (TNF), Lectins, Mannan-binding Serine Protease related (MASP), TNF-Receptor Associated Factors (TRAFs), Nod-Like Receptors (NLRs), Death Domain (DD), Caspase Activation and Recruitment Domain (CARD), and Retinoic Acid Inducible Gene –I (RIG-I). Other groups of proteins include those generally involved with transcription, protein processing, translation, nucleotide metabolism, DNA repair, cell cycle control and apoptosis. * = Proteins central to apoptosis (TRAF4 and TRAF6), viral nucleic acid detection (RIG-I), and cell cycle control (RET), that were used for additional bioinformatic analysis.
**Figure 2.3:** Viral sequences matching to coral immune proteins. Domain frequency refers to the total number of domain occurrences within the predicted coral proteome (23,677 proteins) and the total number of domain occurrences within all coral proteins that matched to viral sequences (2,503 proteins). (A) The apoptotic cascade involving (1) Receptor/ligand interactions, (2) Intracellular signaling and (3) Caspase activation. (B) Apoptotic-related domain frequency taken from total viral hits to the coral proteome. Numbers above bars indicate the stage of apoptosis they are involved with taken from (A). (C) Immune-related domain frequency taken from total viral hits and the coral proteome. (D) Combined immune domain frequency in viral hits and the coral proteome (p-value < 0.01, Students T-test). LRR = Leucine Rich Region, MATH = Meprin and TRAF-Homology, Horm_receptor = Hormone Receptor, TIR = Toll-interleukin receptor, C3_related = Complement component 3 related, FGF = Fibroblast Growth Factor, HLH = Helix Loop Helix, Interfer_Bind = Interferon Binding Domain, Pep_C14 = Caspase, CARD = Caspase Activation and Recruitment Domain.
In descending order predicted proteins that contained NACHT, LRR1, Lectin_C, MATH, and DEATH domains possessed the highest number of viral sequence matches. To determine whether any predicted coral protein hits were also homologous to human proteins the top 100 predicted coral proteins with the most viral sequence matches were compared to the predicted human proteome using BLASTp resulting in 64 predicted human homologs. Coral protein function was predicted based on human UniProt annotations and categorized into general cellular processes. Using the combined results of the Conserved Domain Database and human proteome analysis (BLASTp) the predicted coral-viral interactome was created (Figure 2). As expected translated viral gene segments were homologous proteins involved with immunity (predicted immune-mimics), transcription and translation (predicted viral replication) and cellular proliferation (predicted viral control of cell cycle).

**Bioinformatic analysis of select viral sequences matching to host proteins:**

From the predicted viral interactome viral hits to representative coral proteins were selected for additional analyses based on their central roles in apoptosis, cellular proliferation, and viral nucleic acid detection, including TRAF6, TRAF4, RET-Proto oncogene, and RIG-I, respectively [22-24]. Primary sequence alignments between representative viral gene segments and coral proteins were created revealing high conservation of multiple residues within domains that are central to protein function. Specifically, the translated viral sequences were homologous to the TNF-Receptor binding domain of TRAF4, the TRAF zinc-finger domain of TRAF6, the Helicase C-
terminal domain of RET-Proto oncogene and the Tyrosine Kinase C-terminal catalytic domain of RIG-I (Figure 2.4).

Figure 2.4: Primary sequence alignments and GC-content of translated viral gene segments matching to predicted coral immune proteins. (A) Primary sequence alignments of coral immune protein (upper panel) and translated viral homolog (lower panel). Descriptions in grey boxes above alignments indicate the domains that are present within the host protein for the region of homology between host and virus. TRAF = Tumor Necrosis Factor Receptor Associated Factor, RIG-I = Retinoic acid Inducible Gene-I, and RET = RET Proto-oncogene. (B) GC-content of all viral DNA segments that match to a specific host protein compared to the GC content of the entire coral genome (n = 4,172). All p-values were determined using students T-tests. Genome vs. v-TRAF4: n = 38, p-value = 0.0022; Genome vs. v-TRAF6: n = 54, p-value < 0.0001; Genome vs. v-RET: n = 13, p-value < 0.0001; Genome vs. v-RIG-I: n = 21, p-value < 0.0001.

Previous work has shown that some viruses vary their Guanine-Cytosine content (GC-content) with respect to their host genome as a mechanism to protect viral genome integrity [25]. To determine whether the GC-content of the translated gene segments homologous to TRAF6, TRAF4, RET-Proto oncogene, and RIG-I proteins was different compared to the coral genome a students T-test was performed. Compared to the entire coral genome the GC-content of the combined Gene segments matching to RET-proto oncogene was significantly lower with respect to the host genome while the GC content of viral hits to TRAF6, RIG-I, and TRAF4 was
significantly higher (Figure 2.4). Figures 1-4 provide DNA evidence supporting a viral origin of gene segments that are homologous to host proteins however the expression of predicted host genes has yet to be investigated.

*Expression of predicted immune proteins from virome analysis:* To determine whether any of the 2,503 host proteins with predicted viral homologs are actually expressed publicly available coral transcriptomes were examined [26]. Heat-stress is a well documented modulator of host immunity therefore we hypothesized that proteins predicted to be involved with host immunity by the viral populations would be differentially expressed under heat shock conditions [27]. In total 36 genes predicted by this study were also differentially expressed in heat-stressed corals compared to untreated controls including proteins involved with G-protein signaling, transcription factors, and cell cycle control (Figure 2.5).
Figure 2.5: Differentially expressed genes predicted from viral populations. Genes that were predicted to be involved with host cell biology from viral populations in this study and confirmed to be differentially expressed under heat stress conditions using publicly available transcriptomes. Coral transcriptions were extracted from untreated and heat-stressed *Acropora hyacinthus* fragments sample from American Samoa [26].

Using references databases to annotate unknown biological samples creates an associated bias in the results based on the biological information that was selected for the original database. For example, the Conserved Domain Database cannot identify domains that are unknown; only those have already been described. Therefore to identify completely novel proteins that may be important for cellular functions a
domain-independent approach is required. One of the underlying hypotheses of this paper states that if a virus possesses a homolog to a host protein than that protein is important for host biology. To test this hypothesis using a domain independent approach two proteins were selected based on two criteria; first both proteins were in the top 2% of all 2,503 host proteins based on total number of viral hits, second both proteins contained predicted transmembrane regions (aug_v2a.12072 and aug_v2a.03218). It has been previously established that viruses manipulate host intracellular signaling therefore we predicted that viral gene segments targeting host transmembrane proteins would localize to intracellular regions [10]. To predict the folding structure of aug_v2a.12072 and aug_v2a.03218 the TMPred prediction server was utilized resulting in three and four transmembrane regions, respectively (Supplementary Table 2.3). Distribution of viral gene segments matching host proteins revealed a strong localization to predicted intracellular regions, supporting our prediction (Figure 2.6).
Figure 2.6: Localization of viral hits to unknown coral proteins. (A) and (B) represent coral proteins with viral “hits” that possess no-known domains based on analysis with the Conserved Domain Database (CDD). Transmembrane domains were predicted with the TMPred server and the folding structure of each protein across a phospholipid bilayer is indicated. Each “*” represents 5 viral “hits” localized to the region of protein where the viral gene segment matches to the coral protein.

This study demonstrates that unknown viral populations can used to naively predict host immune components. To determine whether this approach can naively identify host immune proteins that have already been experimentally tested the proteomes of eight Human Herpes viruses were compared to the human proteome using BLASTp analysis revealing multiple host immune components including proteins involved with T-cell signaling, chemokine signaling, and apoptosis (Supplementary Figure 2.4). Based on the pipeline utilized in this study a general experimental approach is proposed in Figure 2.7 to characterize unknown immune systems using viral populations. Viral nucleic acid is isolated from host tissue and compared to the host transcriptome to identify gene candidates involved with host immunity. Host proteins are then selected for further in vitro and vivo experimentation and verification based on homology to viral gene segments.
Figure 2.7: General pipeline to characterize unknown immune systems. Viruses are extracted from the organism of interest using a Virus-Like Particle protocol of choice and nucleic acid is sequenced. If a gene model is not available total RNA is also extracted from host tissue and an assembled transcriptome is prepared. Following quality control (PRINSEQ) and removal of contaminating DNA (DeconSeq) viral gene segments are compared to either the proteome or transcriptome through tBLASTn and predicted viral homologs to host immune proteins are identified. Immune proteins of interest as well as predicted viral-host interactions can be directly tested using In vitro and In vivo experimentation.
Discussion

_Viral communities manipulate host biology:_ Historically the field of virology has been biased towards viruses that cause disease [28]. However, with the rise of metagenomics it has become clear that viral populations or ‘viromes’ associated with hosts include many viruses that do not cause any known pathologies. For example the virome of apparently healthy humans includes members of _Herpesviridae_, _Polymoviridae, Papillomaviridae, Adenoviridae, Anelloviridae, Parvoviridae_ [29]. Once a virus infects a host it may pursue two basic life strategies for survival: persistence or acute infection. The persistent life cycle involves a reduced but consistent production of viral particles over time while acute infection is rapid and commonly associated with disease [30]. To maintain viral infection over time persistent viruses will often express multiple homologs to host proteins involved with transcription, translation, cell cycle control, and immunity [10,28]. The predicted coral-viral interactome presented in Figure 2.2 suggests that the viral populations in coral are employing a similar life strategy by manipulating central cellular processes. In addition to selfishly maintaining persistent infection viruses can also provide the host with immune protection against pathogenic invasion [31]. For example, murine cytomegalovirus protects mice against the bacterial pathogens _Yersinia pestis_ and _Listeria monocytogenes_ while the Hepatitis A virus can reduce Hepatitis C viral infection in humans [32,33]. Beyond individual viral species changes in viral communities are also associated with pathologies such as human inflammatory bowel disease [34]. Taken together these data indicate that competition between viral
populations is involved with the regulation of host health and immunity. Most corals lack the ability to move in response to non-self interactions therefore expanding this study to include corals that are in direct competition could directly test the proposed “virome versus virome” hypothesis of immune competition.

Viral manipulation of apoptosis and intracellular pathogen sensing: While the predicted viral interactome provides an expected network of viral-host interactions involved with protein translation, cell proliferation, and immunity two groups of proteins related to Tumor Necrosis Factor (TNF) signaling and Nod-Like Receptors are of particular interest within the coral system (Figure 2.2). The TNF signaling pathway acts as a central mediator of programmed cell death or apoptosis and appears to be functionally conserved from corals to humans [35,36]. TNF-Receptor Associated Factors (TRAFs) are critical adaptor proteins that bind to the intracellular portion of TNF Receptors regulating cell survival and cytokine production [37]. Based on the total number of viral sequences matching to coral proteins homologs of TRAF6 and TRAF4 were both in the top 10% of all host proteins suggesting that the TNF-signaling pathway is a major target of the coral virome. In humans viruses belonging to Herpesviridae, Adenoviridae, Reoviridae, and Retroviridae have been shown to interact with TNF-mediated apoptosis [38]. Future work in corals should focus on the assembly of whole virus genomes to determine which viruses are producing TRAF-related proteins. These analyses may provide insight into whether specific TNF-signaling strategies are shared by virus families infecting disparate metazoan hosts.
Nod-Like Receptors (NLRs) are intracellular signaling molecules that play a central role in the detection of pathogen-associated molecular patterns [39]. The genome of the reef-building coral *Acropora digitifera* encodes 500 predicted NLRs compared to only 22 found within the human repertoire [40,41]. In total 413 viral hits were distributed across 22 NLRs including many containing a Glycosyl Transferase domain (Figure 2.2, Supplementary Figure 2.3, and Supplementary Table 2.2). Mucins are glycosylated proteins produced by glycosyl transferases and act as a primary form of immune defense within the epithelial surface of corals [42]. While the function of any coral NLRs has yet to be determined the well-established connection between mucins and coral health combined with the high abundance of NLRs containing glycosyl transferase domains indicates a link between intracellular pathogen sensing and mucin production. The predicted manipulation of this system by viruses suggests it is an important factor in the maintenance of persistent viral infection in corals.

The discovery of novel immune components: The experimental workflow used in this study identified a variety of proteins with predicted immune function based on domain characterizations in other systems (Figures 2.2-2.5) [21]. Viral populations are predicted to manipulate these proteins based on homology and the high abundance of immune-related domains further supports the general validity of the experimental approach presented in Figure 7. The application of this pipeline to eight fully sequenced Herpes virus genomes demonstrates its ability to naively identify viral
immune homologs that have also been experimentally verified (Supplementary Figure 2.4). Viral populations provide a domain-independent approach to identify novel host immune components based on the hypothesis that if a virus expresses a protein that is similar to a host protein then that host protein is important to host cell biology. This population-based approach can be applied to virtually any system to characterize the “dark matter” of immunology.

Caveats and future work: The method presented here provides a novel high throughput approach to bioinformatically predict the immune system structure of virtually any animal. However, it is important to remember that this method generates hypotheses that require direct biochemical validation. This validation would ideally be performed within the organism of interest however the molecular techniques will remain unavailable for the majority of systems. Therefore a hybrid approach may be taken to provide a more rapid turnaround of hypothesis testing. Specifically, selecting a developed model system to test predicted viral-host interactions that were initially created within an undeveloped system. For example, the established model organism *Nematostella vectensis* (sea anemone) has well developed molecular methods and is a close relative of coral. Both organisms share many immune components therefore host-virus interactions predicted here could be tested using the molecular tools of *N. vectensis* [43,44].

Here we provide a representative example of how viral populations can be used to predict immune system structure in a non-model system. This method can be
applied to virtually any animal to predict host-virus interactions and potentially discover novel immune components.

**Materials/Methods**

*Isolation of VLPs from coral tissue and sequencing of viral DNA*: Virus-Like Particles (VLPs) were isolated from four individual colonies of *Porites rus* taken from two locations on the island of Mo’orea, French Polynesia. Sampling locations included the back reef of the Richard Gump Research Station (PorLT1 and PorLT2) and Tema’e beach (Por1TA and Por2TA). Permission to conduct this research was granted by the Haut-commissariat de la République en Polynésie Française (DRRT) (Protocole d’Accueil 20011-2012, Moorea). Briefly, coral tissue was homogenized in the field with a mortar and pestle and 12 mL of 0.02 µm filtered seawater was added until all coral tissue was removed from the skeleton. Next the coral-tissue slurry was placed in a 15 mL falcon tube followed by the addition of 1 mL of chloroform and stored at 4 °C for three weeks. To isolate VLPs from coral tissue established ultracentrifugation protocols were utilized [16,45]. Briefly, 600 µl of 50 nmol L⁻¹ dithiothreitol was mixed with 8 mL of coral slurry by vortexing, incubated at 37 °C for 1 hr, and centrifuged for 15 min at 3,000 rpm. Seven milliliters of the supernatant was then transferred to the top of a cesium chloride step gradient containing 1.0 mL of each CsCl solution at densities of 1.7 g/mL, 1.5 g/mL, and 1.35 g/mL. The step gradient containing the coral slurry was then spun for 2 hr at 22,000 g using an SW-41 Ti rotor (Beckman Coulter) and 1.5 mL was extracted from the 1.35 - 1.5 interface using a 24-
gauge needle with an upward facing tip. Next the sample was treated with DNase at a final concentration of 100 units/mL and incubated at 37 °C for 1 hr. Finally DNA was extracted from VLPs using standard CTAB/Formamide methods [16]. Two quality control approaches were taken to ensure viral purity of the final DNA product. First VLPs were visualized before and after ultracentrifugation using standard SYBR-gold staining techniques followed by a 16S and 18S PCR to ensure the isolated DNA was not contaminated with cellular genetic material [16].

To obtain sufficient nucleic acid for library preparation viral DNA was amplified using a modified Linker Amplified Shotgun Library (LASL) method [46]. Briefly, initial DNA concentrations were determined using the Quibit Fluorometric Assay (Life Technologies) and 5 ng of total DNA was combined with PCR-grade water to a final volume of 50 µl. Diluted DNA samples were briefly vortexed, aliquoted into 50 µl Covaris tubes, and sheared using the Covaris M220 Focused Ultrasonicator (40 sec at 9 Watts). Sheared DNA was then end-repaired and ligated to LASL Linker A (Linker A FWD: 5’-P-GTATGCTTCGTGATCTGTGTGGGTGT-3’, Linker A REV: 5’ CCACACAGATCACGAAGCATAC-3’) followed by size selection with a target size of 500 base pairs using Pippin Prep (Sage Science) [46]. For each biological sample four PCR reactions were prepared using a barcoded primer and seventeen cycles of PCR were performed (PfuTurbo Cx Hotstart DNA Polymerase, Agilent). Next replicates were combined, purified, and four additional PCR reactions per biological sample were prepared for three cycles of reconditioning PCR yielding the final DNA product. Finally, all barcoded samples were combined
into a single sample used for library preparation with the MiSeq Reagent Kit v3 and sequenced on the MiSeq platform (Illumina).

**Quality control of sequence data:** Preceding downstream analysis quality control of sequence data was performed using a variety of bioinformatic tools. Paired end reads were first joined using COPE [47], low quality sequences ( < 100 base pairs in length, mean quality score < 25, or containing > 10% N’s) were removed using PRINSEQ and any remaining bacterial or human sequences were removed with DeconSeq [17,18] (See Supplementary Table 2.1 for the number of sequences removed at each pre-processing step).

**Bioinformatic pipeline to analyze viral populations:** Viromes were compared to the predicted coral proteome [48] using standalone BLASTx analysis [20]. Predicted coral proteins were classified as “hits” if more than five unique viral sequence fragments matched a coral protein with an e-value < 1 E-04. Predicted coral protein hits were further analyzed using the Conserved Domain Database [21] and grouped according to related domains. Coral proteins were ranked based on total number of viral hits to a specific protein. To determine whether any coral protein hits were homologous to characterized human proteins BLASTp analysis was performed against the human proteome [21]. A predicted coral protein was considered to be homologous to a human protein if they shared > 40% amino acid identity with an e-value < 1 E-20. Based on the annotations and domains of the characterized human
proteins the biological function of the coral homolog was predicted. Four representative coral proteins and the associated viral hits were selected for additional analysis based on their established involvement with apoptosis (TRAF6, TRAF4), viral nucleic acid detection (RIG-I) and cellular proliferation (RETproto oncogene) in humans [22-24]. First global sequence alignments of predicted coral proteins and translated viral gene segments were constructed in Geneious using MUSCLE [49]. To investigate whether the GC-content of viral gene segments was different from the host genome total GC-content of viral hits was determined using the publicly available get_gc_content.pl script [19,48].

**Acknowledgments**

We gratefully acknowledge the support of the National Science Foundation (OCE 12-36905 and earlier awards) to the Moorea Coral Reef LTER site. We also acknowledge Jennifer Meneghin for the creation of the get_gc_content.pl script. This work was supported by an NSF Graduate Research Fellowship awarded to SDQ, an NIH grant (1 R21 AI094534-01) to FLR and the CIFAR Integrated Microbial Diversity Fellowship (IMB-ROHW-141679) to FLR.
References


Acknowledgments

Chapter 2, in full, is in preparation for submission. Steven D. Quistad, Yanwei Lim, Genivaldo Gueiros Z. Silva, Craig E. Nelson, Andreas F. Haas, Linda-Wegley Kelly, Robert Edwards and Forest Rohwer; 2015. The dissertation author was the primary investigator and author of this paper.
Appendix

Supplementary Figure 2.1: *Porites rus* tissue used for virome extraction.

A. 18S

B. 16S

Supplementary Figure 2.2: (A) 18S and (B) 16S PCR of viromes before library preparation. Lanes refer to coral sample numbers in supplementary figure 1 from A-D.
**Supplementary Figure 2.3:** Coral Nod-Like Receptors (NLRs) with the total number of viral hits indicated to the left of each protein. GlycTr = Glycosyl transferase, RVT_1 = Reverse Transcriptase, Rve = Integrase, AAA = ATPase Associated with diverse cellular Activities.

**Supplementary Figure 2.4:** Herpes proteins with human homologs categorized into cellular function.
**Supplementary Table 2.1:** Pre-processing viromes.

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Pre-Prinseq</th>
<th>Post-Prinseq</th>
<th>Post-Deconseq</th>
<th>Avg. Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Por1LT</td>
<td>50,512</td>
<td>44,989</td>
<td>36,668</td>
<td>314</td>
</tr>
<tr>
<td>Por2LT</td>
<td>128,714</td>
<td>111,546</td>
<td>58,299</td>
<td>321</td>
</tr>
<tr>
<td>Por1TA</td>
<td>84,419</td>
<td>74,957</td>
<td>66,140</td>
<td>297</td>
</tr>
<tr>
<td>Por2TA</td>
<td>446,715</td>
<td>361,838</td>
<td>313,193</td>
<td>297</td>
</tr>
<tr>
<td>Por1Fe</td>
<td>39,194</td>
<td>36,189</td>
<td>21,653</td>
<td>299</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>749,554</td>
<td>629,519</td>
<td>495,953</td>
<td>311</td>
</tr>
</tbody>
</table>

**Supplementary Table 2.2:** Total number of viral sequences that match to coral proteins possessing immune-related domains with an e-value < 1 E^{-04}.

<table>
<thead>
<tr>
<th>Viral Hits</th>
<th>Protein ID</th>
<th>Domain</th>
<th>Viral Hits</th>
<th>Protein ID</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>adi v1.12125</td>
<td>C3_related</td>
<td>19</td>
<td>adi v1.09747</td>
<td>MASP</td>
</tr>
<tr>
<td>17</td>
<td>adi v1.04638</td>
<td>CARD</td>
<td>53</td>
<td>adi v1.18053</td>
<td>MATH</td>
</tr>
<tr>
<td>23</td>
<td>adi v1.04662</td>
<td>DEATH</td>
<td>36</td>
<td>adi v1.20266</td>
<td>MATH</td>
</tr>
<tr>
<td>20</td>
<td>adi v1.21146</td>
<td>DEATH</td>
<td>50</td>
<td>adi v1.17699</td>
<td>NACHT</td>
</tr>
<tr>
<td>18</td>
<td>adi v1.03252</td>
<td>DEATH</td>
<td>34</td>
<td>adi v1.08270</td>
<td>NACHT</td>
</tr>
<tr>
<td>19</td>
<td>adi v1.07835</td>
<td>FGF</td>
<td>23</td>
<td>adi v1.13080</td>
<td>NACHT</td>
</tr>
<tr>
<td>18</td>
<td>adi v1.14925</td>
<td>HLH</td>
<td>26</td>
<td>adi v1.23549</td>
<td>NACHT</td>
</tr>
<tr>
<td>14</td>
<td>adi v1.15907</td>
<td>HLH</td>
<td>28</td>
<td>adi v1.12772</td>
<td>NACHT</td>
</tr>
<tr>
<td>10</td>
<td>adi v1.19818</td>
<td>Homeobox</td>
<td>22</td>
<td>adi v1.13065</td>
<td>NACHT</td>
</tr>
<tr>
<td>33</td>
<td>adi v1.21535</td>
<td>Hormone_R</td>
<td>12</td>
<td>adi v1.09204</td>
<td>NACHT</td>
</tr>
<tr>
<td>12</td>
<td>adi v1.22685</td>
<td>Interfer_Bin</td>
<td>18</td>
<td>adi v1.00397</td>
<td>NACHT</td>
</tr>
<tr>
<td>50</td>
<td>adi v1.16703</td>
<td>Lectin C</td>
<td>8</td>
<td>adi v1.21506</td>
<td>NACHT</td>
</tr>
<tr>
<td>33</td>
<td>adi v1.18953</td>
<td>Lectin C</td>
<td>15</td>
<td>adi v1.18050</td>
<td>NACHT</td>
</tr>
<tr>
<td>18</td>
<td>adi v1.20019</td>
<td>Lectin C</td>
<td>8</td>
<td>adi v1.04548</td>
<td>NACHT</td>
</tr>
<tr>
<td>41</td>
<td>adi v1.05619</td>
<td>LRR_1</td>
<td>10</td>
<td>adi v1.23372</td>
<td>NACHT</td>
</tr>
<tr>
<td>17</td>
<td>adi v1.01290</td>
<td>LRR_1</td>
<td>8</td>
<td>adi v1.14184</td>
<td>NACHT</td>
</tr>
<tr>
<td>18</td>
<td>adi v1.21260</td>
<td>LRR_1</td>
<td>14</td>
<td>adi v1.05709</td>
<td>NACHT</td>
</tr>
<tr>
<td>15</td>
<td>adi v1.21981</td>
<td>LRR_1</td>
<td>15</td>
<td>adi v1.24718</td>
<td>Pep_14</td>
</tr>
<tr>
<td>13</td>
<td>adi v1.01774</td>
<td>LRR_1</td>
<td>20</td>
<td>adi v1.04513</td>
<td>RIG-I</td>
</tr>
<tr>
<td>15</td>
<td>adi v1.01530</td>
<td>LRR_1</td>
<td>24</td>
<td>adi v1.23782</td>
<td>TIR</td>
</tr>
<tr>
<td>24</td>
<td>adi v1.01113</td>
<td>MASP</td>
<td>12</td>
<td>adi v1.12827</td>
<td>TNFR</td>
</tr>
</tbody>
</table>
**Supplementary Table 2.3**: Prediction of Transmembrane regions of coral proteins using TMpred. i = in, o = out.

<table>
<thead>
<tr>
<th>Viral Hits</th>
<th>Protein</th>
<th>Size</th>
<th>Start</th>
<th>Stop</th>
<th>Score</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>aug_v2a.12072</td>
<td>1231</td>
<td>652</td>
<td>672</td>
<td>717</td>
<td>i-o</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>761</td>
<td>780</td>
<td>507</td>
<td>o-i</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1111</td>
<td>1131</td>
<td>1310</td>
<td>i-o</td>
</tr>
<tr>
<td>64</td>
<td>aug_v2a.03218</td>
<td>999</td>
<td>337</td>
<td>356</td>
<td>620</td>
<td>o-i</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>476</td>
<td>494</td>
<td>661</td>
<td>o-i</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>748</td>
<td>768</td>
<td>679</td>
<td>i-o</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>781</td>
<td>799</td>
<td>663</td>
<td>o-i</td>
</tr>
</tbody>
</table>
CHAPTER 3

Evidence of an ancient retroviral infection of a proto-cnidarian ancestor

Abstract

During the Cambrian Explosion the majority of modern day animal phyla first appeared in the fossil record. The genetic mechanisms that led to the diversification of metazoan body plans are not well understood. Endogenous Retroviruses (ERVs) have played a significant role in the evolution of metazoans through their donation of genetic material to the host genome, which can subsequently be inherited by host progeny. Specifically, ERVs can directly influence the expression of host genes through the donation of Long Terminal Repeats (LTRs), which act as promoter elements. Here we bioinformatically identify and characterize a novel ERV family found in the genomes of phyla that predate the Cambrian Explosion, representing, to the best of our knowledge, the oldest ERV infection described to date. The LTR region of the novel ERV family contains eight predicted transcription factor binding sites including PBX1, FOXD1, NFATC2, and RREB1. In addition, the LTR is associated with Fibroblast Growth Factor Receptors (FGFRs) in all three cnidarian species investigated here, which suggests it is involved with cnidarian-specific body plan design. By analogy to mammalian systems we discuss a general model of animal evolution mediated by the expansion of ERV promoter elements (LTRs) and modulation of host Gene Regulatory Networks (GRNs).
Introduction

The Cambrian Explosion occurred approximately 550 million years ago and refers to the period during which most modern-day animal phyla (Metazoa) first appeared in the fossil record. While the origin of specific animal lineages was likely much earlier, the emergence of complex skeletal fauna demonstrated remarkable complexity never before seen in the history of life on Earth [1]. One of the major factors that led to this complexity was the evolution of Gene Regulatory Networks (GRNs), which regulate development and therefore control animal body plan design [2]. While this increase in morphological complexity is observable in the fossil record, one major question remains: what was the trigger? Proposed extrinsic triggers include changes in ocean chemistry, atmospheric oxygen, and tectonic shifts, while intrinsic triggers include predation pressure and genetic exchange [3]. The most successful vehicle for genetic transfer between organisms can be found in the most abundant biological entities on the planet: viruses [5].

Viruses provide the most diverse pool of genetic material on Earth and transfer an estimated $10^{28}$ base pairs of DNA per year in the oceans alone [6,7]. They are directly involved in a broad range of biological processes from global carbon cycling to global pandemics [8,9]. Retroviridae is a unique family of RNA viruses that replicate through a DNA intermediate using a virally encoded reverse transcriptase. Once transcribed, the viral DNA is integrated into the host genome where it becomes an endogenous retrovirus (ERV) [10]. As long as the ERV maintains genes for a functional envelope protein it can continue to produce infectious virions. However,
loss of envelope function through mutations or deletions will transform the ERV into a retroelement that retains the ability to replicate internally but is no longer able to infect neighboring cells. ERV infection of differentiated host cell types usually results in the termination of that ERV lineage when the infected cell dies [11,12] but occasionally an ERV will infect an undifferentiated germ cell, resulting in the inheritance of all ERV genomic material by subsequent host progeny [10]. Once integrated into the host germline this genetic material can result in the emergence of new traits [13-16].

One example of ERV-driven evolution involves the split between marsupial and placental mammals, which occurred approximately 130 million years ago [1]. This event appears to have been driven by an ERV infection of the last common placental ancestor, which resulted in the cooption of a viral envelope gene by the host that is expressed during development [17]. This envelope gene, known as Syncytin in humans, is required for placenta formation directly involved within placental morphogenesis, suggesting a mechanism by which ancient marsupials gained the genetic machinery to produce a placenta via an ERV infection [18].

In addition to evolutionary transitions, ERV infections impact host gene expression through the donation of Long Terminal Repeats (LTRs). LTRs are repetitive sequences of viral DNA located at both ends of an integrated provirus and serve as transcriptional promoters to both viral and host genes [19,20]. Depending on the location of viral integration LTRs can have widespread impacts on transcriptional networks. An example of such an element is the transcription factor p53, which is referred to as the “guardian of the genome” due to its critical role in cell cycle control
Recently it has been demonstrated that greater than 1500 binding sites for p53 are attributed to the LTR region of class I ERVs and the expression of adjacent genes is dependent on p53 binding to the LTR element [22]. Therefore ERV replication results in LTR-mediated expansion and innovation of Gene Regulatory Networks (GRNs) leading to the acquisition of novel biological traits [23-25].

While ERV-mediated expansion of GRNs has been examined within certain vertebrate systems it has yet to be placed within the context of general animal evolution, nor the Cambrian Explosion [23,26]. Here we identify an ancient ERV lineage present in the genomes of multiple ancient phyla including three cnidarians (corals, sea anemones, and hydra) suggesting a pre-split origin. Within all three cnidarians the LTR region of this ERV family is located on the same genomic scaffolds as the genes encoding for Fibroblast Growth Factor Receptor (FGFRs) which play a critical role in cnidarian developmental processes [27]. Taken together this suggests that the viral-derived LTR element identified here may be involved with cnidarian-specific developmental processes [27,28]. Using these data we propose a general model of animal evolution involving the modulation of GRNs through ERV expansion resulting in both evolutionary innovation and viral addiction.
Results

Identification of a novel ERV family in Acropora digitifera: A representative genome of an ancient reef-building coral (class Anthozoa) was used to identify Precambrian-era ERV infections. First, coral genomic scaffolds were compared to metagenomes of coral-associated viruses (henceforth referred to as viromes) to identify potential regions of similarity, which would suggest the presence of an integrated virus. This analysis identified a novel LTR in the A. digitifera genome (see supplementary methods) based on the presence of canonical LTR components including a TATA box promoter and two polyadenylation signals (Supplementary Table 3.1) [29]. The LTR was associated with In silico translation of this element and a primary sequence alignment with the original coral protein revealed that the LTR contained a highly conserved 54 amino acid motif (Supplementary Figure 3.1). To determine the prevalence of the LTR within the coral genome, the 54 amino acid motif taken from the LTR region was compared against the A. digitifera genome. In total one hundred and eleven occurrences of the LTR were distributed across ninety-six unique genomic scaffolds (Figure 3.1 and Supplementary Table 3.2).
Figure 3.1: Bioinformatic pipeline used to identify novel ERV family. The LTR element was initially identified through BLASTx of the coral proteome against metagenomes of coral-associated viruses. A portion of the LTR was then used to search the genomes of *N. vectensis*, *H. magnipapillata*, *S. purpuratus* and *S. kowalevskii* as well as metagenomes of coral and hydra-associated viruses using tBLASTn.
Previous work has indicated that the G + C content and Open Reading Frame (ORF) density of viruses differ compared to the host genome [30-33]. To investigate whether the LTR-containing scaffolds differ in both of these characteristics ninety-six genomic scaffolds were randomly selected from the *A. digitifera* genome and compared against the ninety-six scaffolds containing the LTR. Scaffolds containing the LTR element had a significantly higher G + C content and ORF density compared to random genomic scaffolds (Figures 3.2; *p*-value < 0.0001 and 0.00001 respectively, students T-test, and Supplementary Table 3.3), suggesting an ERV origin. ORFs extracted from the LTR scaffolds also contained multiple viral-related domains including reverse transcriptases, integrases, RNase H, and RT_pep17 peptidases (Figure 2C and Supplementary Table 3.4) [34]. In addition a representative ERV genome possesses thirteen predicted ORFs containing helicase, reverse transcriptase and RT_pepA17 peptidase domains located within two LTRs (Figure 3.2D and Supplementary Table 3.4).
Figure 3.2: ERV family characterization in *Acropora digitifera*. Ninety-six scaffolds containing the LTR were compared against ninety-six randomly selected genomic scaffolds. (A) GC content of LTR-scaffolds versus random scaffolds (T-test, *p*-value < 0.0001, effect size 0.0055 +/- 0.00067. (B) Open Reading Frame (ORF) density per 10 Kilobases of DNA in LTR-scaffolds versus random scaffolds (T-test, *p*-value < 0.00001, effect size 2.0 +/- 0.36). (C) Number of retroviral-related domains found in ORFs greater than 100 amino acids were extracted from the LTR scaffolds. (D) A representative ERV genome with ORFs, LTR, and viral-related domains indicated. * = the location of the 54 amino acid motif used for LTR searches.
**Distribution of the ERV family across phyla and viromes:** To determine whether the LTR is found in other organisms the genomes of *Hydra magnipapillata* (hydra), *Nematostella vectensis* (sea anemone), *Saccoglossus kowalevskii* (acorn worm) and *Strongylocentrotus purpuratus* (purple sea urchin), were queried for the 54 amino acid motif taken from the LTR region, which revealed 110, 541, 82, 13, and 17 occurrences respectively (Supplementary Table 3.2). The LTR was also found within virus communities isolated from hydra and coral tissue (Supplementary Figure 3.2). Phylogenetic relationships between all occurrences of the LTR were examined using the 54 amino acid motif extracted from the LTR region and demonstrated tight clustering based on the source genome or virome (Figure 3.3). The high abundance and distribution of the LTR throughout the genomes of three cnidarians and apparent lack of the LTR in all other phyla investigated (excluding sea urchins and acorn worms) suggests a founding ERV infection that occurred in the last common cnidarian ancestor (Figure 3.4).
Figure 3.3: Phylogenetic tree of the LTR element in genomes and viral metagenomes. The tree was constructed using the highly-conserved 54 amino acid motif located within the LTR element. Genomes include hydra (H.mag), sea anemone (N.vec), acorn worm (S.kow), purple sea urchin (S.purp), coral (A.dig). Metagenomes were extracted from viruses associated with two corals Acropora tenuis (Vir_A.ten) and Pocillipora damicornis (Vir_P.dam), and one hydra species: Hydra magnipappillata (Vir_H.mag). Bootstrap values indicate 10,000 sampling events.
**Figure 3.4:** A proposed model for an ERV infection of the proto-cnidarian ancestor. Total number of individual hits to 54 amino acid motif within the genomes of hydra (*H. magnipapillata*), sea anemone (*N. vectensis*), acorn worm (*S. kowalevskii*), and purple sea urchin (*S. purpuratus*) placed within their predicted evolutionary origins. X’s indicate representative genomes that were analyzed with no apparent evidence for the presence of the conserved motif. Estimated divergence times are in millions of years (Myr) and based on Shu et al (2014) [41]
**LTR as a promoter element:** Within five representative *A. digitifera* LTRs eight DNA motifs contained known transcription factor binding sites, which included PBX1, FOXD1, NFATC2, RREB1, YNR063W, SWI5, RUNX2, and BAP (Figure 3.5A, Supplementary Table 3.4). Further investigation of the PBX1 protein revealed high conservation from humans to corals and an apparent absence of PBX1 from the transcription factor repertoires of *A. queenslandica* (sponge), *Saccharomyces cerevisiae* (yeast), Fungi (taxid:4751) and Bacteria (taxid:2) (Figure 3.5B).
Figure 3.5: Predicted PBX1 binding site within LTR region and primary sequence alignment of the PBX1 protein across metazoans. (A) Five representative LTRs were extracted from the *A. digitifera* genome and analyzed with MEME [47]. Bottom regions of each row designate a specific portion of the LTR region with predicted transcription factors placed above their predicted bind sites. Letter size indicates nucleotide abundance. (B) Primary sequence alignment of PBX1 proteins taken from *H. sapiens, D. rerio, S. purpuratus N. vectensis, H. vulgaris,* and *A. digitifera.*
Given evidence that host transcription factors can bind to viral LTRs, the surrounding genomic space of LTR-scaffolds was probed for associated genes that may be affected by LTR presence. This search yielded Fibroblast Growth Factor Receptors (FGFRs), which were found to be associated with at least one LTR located on the same genomic scaffold in all three cnidarians (Figure 3.6A-3.6C and Supplementary Table 3.6). To determine which specific FGFRs might be affected by the presence of an LTR presence BLASTp analysis of cnidarian FGFRs was performed against the human proteome. In addition to FGFRs, the LTR element was also associated with proteins involved with extracellular matrix production (Fibrillin 2), membrane trafficking (Rab 1A) and development (Notch-2) (Supplementary Table 3.6). A general model for metazoan evolution mediated via ERV expansion is presented in Figure 3.7.
Figure 3.6: The location of the LTR element within genomic scaffolds containing Fibroblast Growth Factor Receptors (FGFRs) in three cnidarians. (A) *Acropora digitifera* (coral), (B) *Hydra magnipappillata* (hydra), (C) *Nematostella vectensis* (sea anenome). Numbers following “FGFR” represent the top BLASTp hit to all human proteins and numbers below vertical lines indicated base pairs of a specific genomic scaffold.
Figure 3.7: Proposed model of ERV-mediated metazoan evolution. Donation of promoter (LTR) by founder retrovirus affects local gene expression of “Gene A”. ERV replication expands the number of host genes affected by LTRs (Genes B-G) modulating host Genetic Regulatory Networks (GRNs) and leading to evolutionary innovations as well as host addiction to an ERV family.
Discussion

Summary: This study finds evidence of a Precambrian ERV infection based on the identification and characterization of an LTR region within the genome of all available cnidarians. Multiple genes associated with this LTR code for known transcription-factor binding sites and receptors involved in critical developmental processes, suggesting this virally-donated LTR plays a role in body plan design. The presence of the LTR in Cnidaria and more recent sister lineages, but absence in other more distant lineages highlights its potential specificity to cnidarian body-plan design, and perhaps suggests it played a role in the diversification of these taxa. To the authors’ knowledge, this is the most ancient ERV infection as yet to be identified and likely predates the Cambrian Explosion.

Fibroblast Growth Factor Receptors (FGFRs): Fibroblast Growth Factor Receptors (FGFRs), which were found in close proximity to the virally-donated LTR, are involved with body plan formation across animal phyla [35] (Figure 3.6). They consist of three extracellular Immunoglobulin-like domains that signal through an intracellular tyrosine kinase [36]. FGFR signaling is involved with the formation of neural tissue and the coordination of cell movements during gastrulation [35]. Recently it has been proposed that the involvement of FGFR signaling with both of these developmental processes pre-dates the Cambrian Explosion and can be traced to the last common Eumetazoan ancestor [27,37]. In Cnidaria FGFR signaling during development has been well characterized [38]. For example, in N. vectensis an
examination of two FGFRs (NvFGFRa and NvFGFRb) during development demonstrated that both FGFRs are involved with endoderm invagination and the formation of a chemosensory organ [27,28]. NvFGFRa and NvFGFRb are 100% identical to N. vectensis proteins 98571 and 31143 respectively, and are both located on genomic scaffolds containing the LTR identified here (Supplementary Table 3.6 and Figure 3.6C). The proximity of the LTR to FGFR genes in N. vectensis, H. magnipappilata, and A. digitifera suggests that the LTR may affect gene expression however functional studies are still needed. In addition the co-occurrence of the LTR with FGFRs in all three cnidarians and apparent absence of the LTR from FGFR genomic scaffolds taken from S. purpuratus and S. kowalevskii further suggests the LTR is involved with cnidarian-specific traits.

PBX1 transcription factors and the Cambrian Explosion: This study identified eight known transcription factor binding sites within gene motifs of the LTR. Among them, the Pre-B-Cell Leukemia Factor-1 (PBX1) proteins serve as cofactors for Hox genes and have been implicated in a variety of developmental processes from muscle cell differentiations to organogenesis [39]. The canonical Hox gene family encodes transcription factors that control body plan design and their function is highly conserved across metazoan phyla [40]. In addition, PBX1 has been directly linked to FGFR signaling during nervous system development [41]. The established role of PBX1 in development combined with the high conservation of PBX1 from corals to humans and apparent absence from bacteria, fungi, and supports the hypothesis that
PBX1 was involved with the emergence of complex body plans observed during the Cambrian Explosion (Figure 5B). Taken together the donation of a predicted PBX1 binding site to FGFR-encoding scaffolds further supports the involvement of the novel ERV family described here with cnidarian-specific body plan design.

**Precambrian origin of ERV family:** The presence of the LTR within the genomes of all available cnidarians and apparent absence in the genomes of yeast, protozoans, choanoflagellates, and poriferans suggests an ancient retroviral infection, which occurred in a proto-cnidarian ancestor of *N. vectensis*, *H. magnipappillata*, and *A. digitifera*. Interestingly the LTR was also found in a sea urchin (Echinodermata) and acorn worm (Hemichordata) genomes while apparently absent from Protostomia and Chordata (Figure 3.4). Echinoderms and hemichordates are hypothesized to have diverged from the last common ancestor (LCA) Ambulacraria approximately 680 Mya ago during the Ediacaran period [4,42]. Therefore, two competing hypothesis exist regarding the appearance of this ERV lineage in Ambulacrarians; one lateral gene transfer event from a Cnidarian to Ambulacraria followed by independent genomic expansion or at least two independent losses of the ERV family from the LCA of Protostomia and Vertebrata. If viral expansion occurred before the divergence of Protosomia and Vertebrata ancestor’s more than two events would also need to be explained. Given this, we conclude that the former hypothesis is the most parsimonious explanation of viral expansion across these disparate phyla. As more genomes become available from previously unrepresented phyla the presence of this
novel viral motif should be investigated to achieve a more comprehensive hypothesis of how ERV expansion occurred.

**ERVs, GRNs, and metazoan evolution:** While the impact of ERV expansion on host evolution has been well documented within humans it has yet to placed within the context of general metazoan evolution [22,25,26]. To the authors’ knowledge this work provides the first evidence of a Precambrian-era ERV infection. Multiple lines of evidence suggest that the LTR associated with this novel ERV family is involved with phyla-specific developmental processes (Figure 3.5 and Figure 3.6). Combining previous work in humans with the novel ERV family described here, a general model of metazoan evolution mediated by ERV expansion of GRNs is proposed as follows and depicted in Figure 3.7. Following a founder retroviral infection, an ERV donates a transcription factor-binding module (LTR) in proximity to a host gene resulting in the modulation of gene expression. As the ERV continues to replicate within the host genome the network of affected genes increases in parallel until a “point of no return” is reached with a particular GRN (i.e. development of a specific body part). ERV expansion ultimately creates novel GRNs resulting in the fixation of both biological traits and ERV lineages within the host genome.

**Predictions of ERV-mediated metazoan evolution:** Based on the model presented in Figure 7, the distribution and conservation of specific ERV families and their impact on GRNs is expected to vary depending on the biological trait with which
they are involved with. In general GRNs involved with developmental processes are more conserved at the inter-phyla level compared to GRNs involved with peripheral functions such as the creation of a specialized cell-type. The most central and conserved family of GRNs are referred to as “kernals” while the GRNs associated with more flexible traits are termed “gene-batteries”. We expect the impact of ERVs on developmental GRNs to be ancient and fixed with respect to “kernals” while more recent and flexible with respect to “gene-batteries”. If a kernal is viral in origin then the surrounding sequence space should have maintained a viral signature due to its fixation into developmental processes. This viral signature should be conserved at the intra-phyla level and divergent between phyla. Based on the cnidarian-specific association of the LTR described here with FGFRs and PBX1 (Figures 3.5 and 3.6) we hypothesize that the ERV family described here is involved with a novel cnidarian kernel. The active role of ERVs operating at the gene battery level has already been observed during nervous system development. For example in vitro studies of the LINE-1 retroelement have demonstrated that it is able to retrotransposase in neuronal precursor cells which directly affects gene expression and neuronal cell fate [43]. Furthermore insertions of LINE-1 during neurogenesis appear to be non-random. Given this, it has been proposed that LINE-1 retrotransposition serves as a mechanism to direct neuronal diversification and brain development [43,44].

Viral addiction and metazoan evolution: From the perspective of a virus, the transition from a retrovirus to an ERV ensures that as long as reproduction occurs the
virus will be maintained in perpetuity. Therefore, any mechanism that increases the probability of undergoing an endogenous transition would increase viral fitness. LTRs are traditionally thought of as viral promoters that are co-opted by host genes by chance. However, if a virus possess an LTR that matches the transcription factor repertoire of a host that virus will be labeled by host protein and integrate into host GRNs more quickly compared to retroviruses that lack LTR specificity. This ready-to-use promoter hypothesis was first posited using evidence taken the ERV-L LTR promoter of human β1,3-galactosyltransferase 5 (β3GAL-T5) [45]. Genome-wide analysis of multiple primate genomes suggested that the human β3GAL-T5 LTR promoter possessed its current transcription factor binding sites at the time of insertion.

The existence of the novel LTR described here and the presence of transcription factor binding sites within the viral populations of hydra and coral provides further support for the ready-to-use promoter hypothesis (Figure 3.3C and 3.3D). It is important to recognize that the viromes analyzed here were taken from DNA viruses not RNA viruses. While viral endogenization is often discussed within the context of retroviruses possessing RNA genomes the process has also been reported for a variety of DNA viruses [46]. Therefore why DNA viruses are harboring an LTR promoter remains unknown. Further experimentation is required to determine which viruses are maintaining the LTR within their genomes and whether the viral LTRs can act as ready-to-use promoters following genomic insertion.
Conclusions: We provide evidence for an ERV infection that occurred within the LCA of cnidarians during the Precambrian-era. To the authors knowledge this ERV infection represents the oldest infection described to date. The association of the LTR with FGFRs suggests a role in development and its phylogenetic distribution further supports a role in cnidarian-specific traits. Using these data we have proposed a general model of metazoan evolution involving the donation of promoters (LTRs) to host genomes affected local gene expression. Expansion of ERV populations subsequently expands the number of affected host genes leading to the modulation of GRNs and evolutionary innovation.

Acknowledgments

We thank Drs. James Sikes (University of San Francisco) and Bob Zeller (San Diego State University) for helpful comments with the manuscript.

Funding

This work was supported by an NSF Graduate Research Fellowship awarded to SDQ, the NIH (1 R21 AI094534-01) to FLR and the CIFAR Integrated Microbial Diversity Fellowship (IMB-ROHW-141679) to FLR.
References


Acknowledgments

Chapter 3, in full, is under review in the Proceedings of the National Academy of Sciences. Steven D. Quistad, Aaron C. Hartmann, Linda Wegley-Kelly, and Forest Rohwer; 2015. The dissertation author was the primary investigator and author of this paper.
Appendix

Supplementary Methods

The LTR was initially identified within the genome of *A. digitifera* [28] by standalone tBLASTx [29] analysis of viromes extracted from *Porites compressa* [23] against the *A. digitifera* proteome. From this analysis *A. digitifera* protein aug.11951 was identified as a top hit to the viromes. Further tBLASTn analysis of aug.11951 against the *A. digitifera* genome revealed a conserved 54 amino acid (162 base pair) motif within the LTR region. To investigate whether specific transcription factor binding sites might exist within the LTR region five representative 162 base pair DNA sequences from *A. digitifera* were analyzed using MEME (v 4.9.1) [30]. A representative 54 amino acid motif taken from *A. digitifera* (adi_Scaffold1125) was used to query the genomes of *Hydra magnipilalata* (Hydra), *Acropora digitifera* (Coral), *Nematostella vectensis* (Sea Anenome), *Strongylgous purpatus* (Sea Urchin), and *Saccoglossus kowalevskii* (Acorn Worm), *D. melanogastor*, *C. elegans*, *S. cervisiae*, *D. rerio*, *Monosiga brevicollis* (choanoflagellate), *Amphimedon queenslandica* (sponge), *Plasmodium* (protists), and *H. sapiens* using tBLASTn. Sequence that matched with 54 amino acid motif with greater than 75% coverage and 60% identity to the 54 amino acid motif were classified as positive.
Supplementary Figure 3.1: Representative primary sequence alignments of 54 amino acid motif extracted from the LTR element. (A) hydra (H. magnipapillata), (B) acorn worm (S. kowalevskii), (C) sea anemone (N. vectensis), and (D) purple sea urchin (S. purpuratus) with consensus sequences indicated above each alignment.
Supplementary Figure 3.2: Abundance of the LTR element among host genomes and viromes. (A) Abundance of the LTR element within the genomes of *N. vectensis* (sea anenome), *H. magnipapillata* (hydra) *S. purpuratus* (purples sea urchin) and *S. kowalevskii* (acorn worm). (B) Abundance of the LTR element normalized per 10 Megabases of genomic DNA. (C) Abundance of the LTR element in metagenomes of viruses isolated from coral and hydra. A.ten = *Acropora tenuis*, P.dam = *Pocillopora damicornis*, H.vul = *Hydra vulgaris*, H.mag = *Hydra magnipapillata*, H.olg = *Hydra olgactis*. (D) Abundance of LTR normalized per 10 Kilobases of viral DNA.
Supplementary Table 3.1: Transcription features of the novel LTR located within scaffold 1125 of the *Acropora digitifera* genome.

<table>
<thead>
<tr>
<th>Base Pair Start</th>
<th>Base Pair Stop</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>24934</td>
<td>24939</td>
<td>TATAAA</td>
</tr>
<tr>
<td>24958</td>
<td>24963</td>
<td>AGUAAA</td>
</tr>
<tr>
<td>24075</td>
<td>25080</td>
<td>AAUAAA</td>
</tr>
<tr>
<td>25058</td>
<td>25219</td>
<td>Motif used for searches</td>
</tr>
</tbody>
</table>

Supplementary Table 3.2: Distribution of LTR across viromes and genomes

<table>
<thead>
<tr>
<th>Sample</th>
<th># LTR occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A. digitifera</em> (Genome)</td>
<td>111</td>
</tr>
<tr>
<td><em>N. vectensis</em> (Genome)</td>
<td>82</td>
</tr>
<tr>
<td><em>H. magnipapillata</em> (Genome)</td>
<td>541</td>
</tr>
<tr>
<td><em>S. purpuratus</em> (Genome)</td>
<td>17</td>
</tr>
<tr>
<td><em>S. kowalevskii</em> (Genome)</td>
<td>32</td>
</tr>
<tr>
<td><em>A. tenuis</em> (Virome)</td>
<td>123</td>
</tr>
<tr>
<td><em>P. damicornis</em> (Virome)</td>
<td>59</td>
</tr>
<tr>
<td><em>P. damicornis</em> (Virome)</td>
<td>1</td>
</tr>
<tr>
<td><em>P. damicornis</em> (Virome)</td>
<td>25</td>
</tr>
<tr>
<td><em>P. damicornis</em> (Virome)</td>
<td>30</td>
</tr>
<tr>
<td><em>H. vulgaris</em> (Virome)</td>
<td>7</td>
</tr>
<tr>
<td><em>H. magnipapillata</em> (Virome)</td>
<td>10</td>
</tr>
<tr>
<td><em>H. oligactis</em> (Virome)</td>
<td>5</td>
</tr>
<tr>
<td><em>H. oligactis</em> (Virome)</td>
<td>2</td>
</tr>
</tbody>
</table>
**Supplementary Table 3.3**: Representative ERV genome located within scaffold 1125 of the *Acropora digitifera* genome.

<table>
<thead>
<tr>
<th>ORF</th>
<th>Start</th>
<th>Stop</th>
<th>Size (aa)</th>
<th>Domain</th>
<th>ID</th>
<th>e-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,016</td>
<td>2,531</td>
<td>171</td>
<td>DEAD-like helicase</td>
<td>cd00046</td>
<td>2.37E-07</td>
</tr>
<tr>
<td>2</td>
<td>3,936</td>
<td>4,745</td>
<td>269</td>
<td>RT_like super family</td>
<td>cl02808</td>
<td>3.89E-16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RVT_1</td>
<td>pfam00078</td>
<td>1.59E-12</td>
</tr>
<tr>
<td>3</td>
<td>4,839</td>
<td>5,435</td>
<td>198</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6,091</td>
<td>6,669</td>
<td>193</td>
<td>unknown function</td>
<td>cl13449</td>
<td>8.55E-08</td>
</tr>
<tr>
<td>5</td>
<td>9,766</td>
<td>10,131</td>
<td>121</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>12,744</td>
<td>13,154</td>
<td>136</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>13,151</td>
<td>13,786</td>
<td>211</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>14,276</td>
<td>14,809</td>
<td>177</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>14,865</td>
<td>15,257</td>
<td>130</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>15,282</td>
<td>16,553</td>
<td>423</td>
<td>RT_pepA17</td>
<td>cd01644</td>
<td>3.46E-79</td>
</tr>
<tr>
<td>11</td>
<td>17,348</td>
<td>18,145</td>
<td>265</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>20,344</td>
<td>20,646</td>
<td>100</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>20,861</td>
<td>21,517</td>
<td>218</td>
<td>unknown function</td>
<td>cl19219</td>
<td>9.10E-04</td>
</tr>
<tr>
<td>14</td>
<td>24,845</td>
<td>25,324</td>
<td>160</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>25,556</td>
<td>25,921</td>
<td>122</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Supplementary Table 3.4:** MEME analysis of the 54 amino acid motif located within the novel LTR.

<table>
<thead>
<tr>
<th>Binding Site</th>
<th>Base Start</th>
<th>Base End</th>
<th>Transcription Factor</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28</td>
<td>35</td>
<td>YNR063W</td>
<td>5.9E-05</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>22</td>
<td>PBX1</td>
<td>3.3E-03</td>
</tr>
<tr>
<td>3</td>
<td>42</td>
<td>49</td>
<td>SWI5</td>
<td>4.0E-03</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>11</td>
<td>BAP</td>
<td>2.8E-04</td>
</tr>
<tr>
<td>5</td>
<td>31</td>
<td>38</td>
<td>FOXD1</td>
<td>9.5E-03</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>25</td>
<td>RUNX2</td>
<td>1.8E-03</td>
</tr>
<tr>
<td>7</td>
<td>42</td>
<td>48</td>
<td>NFATC2</td>
<td>2.4E-06</td>
</tr>
<tr>
<td>8</td>
<td>21</td>
<td>40</td>
<td>RREB1</td>
<td>7.5E-03</td>
</tr>
</tbody>
</table>

**Supplementary Table 3.5:** Cnidarian genome scaffolds that contain both the LTR identified here and genes encoding for Fibroblast Growth Factor Receptors (FGFRs).

<table>
<thead>
<tr>
<th>Species</th>
<th>Protein ID</th>
<th>Best Blast Hit (humans)</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>H. magnipappilata</em></td>
<td>XP_002164661.2</td>
<td>FGFR 1</td>
<td>2.0E-51</td>
</tr>
<tr>
<td><em>N. vectensis</em></td>
<td>98570</td>
<td>FGFR 1</td>
<td>2.0E-170</td>
</tr>
<tr>
<td><em>N. vectensis</em></td>
<td>98616</td>
<td>FGFR1</td>
<td>7.0E-113</td>
</tr>
<tr>
<td><em>N. vectensis</em></td>
<td>31143</td>
<td>FGFR 3</td>
<td>0</td>
</tr>
<tr>
<td><em>N. vectensis</em></td>
<td>81258</td>
<td>FGFR 3</td>
<td>4.0E-138</td>
</tr>
<tr>
<td><em>N. vectensis</em></td>
<td>94990</td>
<td>FGFR1</td>
<td>2.0E-112</td>
</tr>
<tr>
<td><em>N. vectensis</em></td>
<td>94997</td>
<td>FGFR 2</td>
<td>4.0E-113</td>
</tr>
<tr>
<td><em>A. digitifera</em></td>
<td>adi v1.00236</td>
<td>FGFR 3</td>
<td>1.0E-179</td>
</tr>
<tr>
<td><em>A. digitifera</em></td>
<td>adi v1.00239</td>
<td>FGFR 18</td>
<td>1.0E-31</td>
</tr>
<tr>
<td><em>A. digitifera</em></td>
<td>adi v1.06066</td>
<td>FGFR 1</td>
<td>2.0E-170</td>
</tr>
<tr>
<td><em>A. digitifera</em></td>
<td>adi v1.06067</td>
<td>FGFR 3</td>
<td>2.0E-133</td>
</tr>
<tr>
<td><em>A. digitifera</em></td>
<td>adi v1.06069</td>
<td>FGFR 2</td>
<td>3.0E-82</td>
</tr>
<tr>
<td><em>A. digitifera</em></td>
<td>adi v1.22552</td>
<td>FGFR 5</td>
<td>2.0E-33</td>
</tr>
</tbody>
</table>
**Supplementary Table 3.6:** Additional genes that are associated with the novel LTR identified here in all three cnidarians investigated.

<table>
<thead>
<tr>
<th>Coral Protein</th>
<th>Human Homolog</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>adi v1.24006</td>
<td>cationic amino acid transporter 2</td>
<td>2.0E-154</td>
</tr>
<tr>
<td>adi v1.21722</td>
<td>Plasminogen</td>
<td>6.0E-44</td>
</tr>
<tr>
<td>adi v1.21685</td>
<td>EIF3A</td>
<td>0</td>
</tr>
<tr>
<td>adi v1.17963</td>
<td>fibrillin 2</td>
<td>1.0E-90</td>
</tr>
<tr>
<td>adi v1.06066a</td>
<td>Fibroblast growth factor receptor 1</td>
<td>2.0E-170</td>
</tr>
<tr>
<td>adi v1.06069</td>
<td>Fibroblast growth factor receptor 2</td>
<td>3.0E-82</td>
</tr>
<tr>
<td>adi v1.00236</td>
<td>Fibroblast growth factor receptor 3</td>
<td>1.0E-179</td>
</tr>
<tr>
<td>adi v1.06067</td>
<td>Fibroblast growth factor receptor 3</td>
<td>2.0E-133</td>
</tr>
<tr>
<td>adi v1.06066b</td>
<td>Fibroblast growth factor receptor 5</td>
<td>1.0E-14</td>
</tr>
<tr>
<td>adi v1.08137</td>
<td>MATN2</td>
<td>1.0E-36</td>
</tr>
<tr>
<td>adi v1.07797</td>
<td>Notch homolog 2</td>
<td>2.0E-99</td>
</tr>
<tr>
<td>adi v1.03173</td>
<td>ovochymase-2 precursor</td>
<td>3.0E-46</td>
</tr>
<tr>
<td>adi v1.06782</td>
<td>ovochymase-2 precursor</td>
<td>4.0E-42</td>
</tr>
<tr>
<td>adi v1.00577</td>
<td>E3 ubiquitin-protein</td>
<td>0</td>
</tr>
<tr>
<td>adi v1.01189</td>
<td>ras-related protein Rab-1A</td>
<td>1.0E-70</td>
</tr>
<tr>
<td>adi v1.09733</td>
<td>transmembrane protease serine 11D</td>
<td>4.0E-45</td>
</tr>
<tr>
<td>adi v1.03220</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
CHAPTER 4

Evolution of TNF-induced apoptosis reveals 550 million years of functional conservation

Abstract

The Precambrian explosion led to the rapid appearance of most major animal phyla alive today. It has been argued that the complexity of life has steadily increased since that event. Here we challenge this hypothesis through the characterization of apoptosis in reef-building corals, representatives of some of the earliest animals. Bioinformatic analysis reveals that all of the major components of the Death Receptor Pathway are present in coral with high-predicted structural conservation with H. sapiens. The Tumor Necrosis Factor (TNF) Receptor-Ligand Superfamilies (TNFRSF/TNFSF) are central mediators of the Death Receptor Pathway and the predicted proteome of Acropora digitifera contains more putative coral TNFRSF members than any organism described thus far, including humans. This high abundance of TNFRSF members, as well as the predicted structural conservation of other Death Receptor Signaling proteins led us to wonder what would happen if corals were exposed to a member of the Human TNFSF (HuTNFα). HuTNFα was found to bind directly to coral cells, increase Caspase activity, cause apoptotic blebbing and cell death and finally induce coral bleaching. Next immortalized human T-cells (Jurkats) expressing a functional Death Receptor Pathway (Wild-Type) and a corresponding FADD-knockout cell line were exposed to a coral TNFSF member (AdTNF1) purified here. AdTNF1 treatment resulted in significantly higher cell death...
(p<0.0001) in Wild-Type Jurkats compared to the corresponding FADD-knockout demonstrating that coral AdTNF1 activates the *H. sapien* Death Receptor Pathway. Taken together these data show remarkable conservation of the TNF-induced apoptotic response representing 550 million years of functional conservation.
Introduction

Discoveries from model organisms have significantly influenced the field of human immunology. For example, the original concept of self versus non-self recognition was discovered from observations in echinoderms, while the discovery of Toll-like receptors in humans stemmed from investigations into the response of insects to pathogens. Despite the impact of these studies, the majority of our understanding of immune function remains based on data from a select few taxa, mainly Chordata, Arthropoda, and Nematoda, which represent only 3 of the 30 extant animal phyla [1]. While these models have provided valuable insight into the molecular basis of immune defense, we are overlooking a significant and potentially informative portion of metazoan biology. With the rise of the genomic revolution, an increasing number of genomes from basal phyla are revealing the evolution of immunity to be a non-linear process, involving multiple instances of gene gain and loss [2]. Therefore, the investigation of non-traditional phyla will provide a deeper understanding of the evolution of immunity, including the potential for the discovery of novel immune reactions.

The phylum Cnidaria diverged from Bilateria 550 million years ago and contains over 10,000 species that range in size from a few millimeters to over 75 meters [3]. Their body plan consists of two cell layers, an endoderm and ectoderm, held together by the jelly-like mesoglea [4]. Stony corals (Order Scleratina) are colonial cnidarians, and are responsible for supporting the most biologically diverse ecosystem on the planet: the coral reef. Reefs support economically important
industries such as fishing and tourism, and provide coastal protection to hundreds of millions of people worldwide. Recent global surveys have indicated that 19% of coral reefs have been destroyed, 15% are under imminent risk of collapse, and a further 20% are under long-term threat of collapse [5]. Anthropogenic impacts such as overfishing and nutrient runoff have been implicated in increased coral death and bleaching [6]. However, while many of the environmental factors leading to coral mortality are well established, the biological mechanisms behind coral death remain poorly understood [7,8]

One common route of coral death on reefs around the world occurs through a process called coral bleaching. During bleaching, the coral’s intracellular symbiotic zooxanthellae are expelled from the host [9]. Programmed cell death or apoptosis has been observed during the bleaching process, however the components of the apoptotic pathway have yet to be fully identified and functionally tested [10]. In humans apoptosis can be activated through either intrinsic or extrinsic pathways. The intrinsic pathway is initiated by cell stress, while the extrinsic pathway is initiated by death-ligand/death-receptor interactions. Both pathways converge on the members of the Bcl-2 family, which can ultimately lead to Caspase activation and the release of signaling molecules such as cytokines to neighboring cells. This process directly links the apoptotic pathway with the innate immune system [11]. In corals apoptosis has been observed in response to hyperthermic oxidative stress, disease, and as a post-phagocytic removal mechanism of zooxanthellae during the onset of symbiosis [10,12]. The recent *Acropora digitifera* genome suggests that coral possess homologs
to the human intrinsic and extrinsic apoptotic pathways [13]. While the activation of
the intrinsic apoptotic pathway in corals has been observed in response to
environmental stress, the extrinsic pathway has yet to be investigated in this system.
Specifically, no work to date has focused on the upstream receptor-ligand families
involved with apoptotic signaling and activation.

The Tumor Necrosis Factor (TNF) receptor-ligand superfamily is a central
mediator of the extrinsic apoptotic pathway. It is known to be involved in a variety of
chronic human diseases such as multiple sclerosis, rheumatoid arthritis and type-2
diabetes [14]. The TNF-Ligand Superfamily (TNFSF) is characterized by a ligand
trimerization interface and TNF-Receptor binding domain. Members of the TNF-
Receptor Superfamily (TNFRSF) are defined by the presence of Cysteine-Rich
Domains (CRD), which are important for receptor oligomerization [15]. Crystal
structure characterization of TNFRSF-TNFSF interactions have also revealed a 50s
and 90s loop structure which are important for ligand binding and specificity,
respectively [15]. Upon ligand binding, the TNFR-1 recruits the TNF-R1-Associated
Death Domain protein (TRADD), RIP1 and TRAF2 creating complex I, which
disassociates from the receptor. Complex I can then activate either the NFκB
transcription factor (among others) leading to cell survival, or bind to FADD resulting
in Caspase recruitment and apoptosis [16,17].

Phylogenetic analysis indicates a deep evolutionary origin of the TNFSF and
TNFRSF that precedes the divergence of vertebrates and invertebrates. The most
ancient and well-defined invertebrate TNF ligand-receptor system that has been
described to date is that of the fruit fly *Drosophila melanogaster* [18]. *D. melanogastor* possesses just one member of both the TNFRSF/TNFSF, in contrast to humans which have 18 and 29 respectively [19]. This difference has led to the widely accepted hypothesis that the TNF ligand-receptor superfamily expanded after the divergence of invertebrates and vertebrates [20,21].

In this paper we describe the annotation of 40 members of the TNFRSF and 13 members of the TNFSF in the reef building coral *Acropora digitifera*, suggesting key parts of the TNF receptor-ligand superfamily have been lost in *D. melanogastor* but maintained in coral [22]. Comparison of these coral TNFSF/TNFRSF members to those of *Homo sapiens* reveals high genetic and predicted structural conservation. Exposure of coral to Human TNFα (HuTNFα) results in apoptotic cellular blebbing, caspase activation, cell death, and finally coral bleaching. Further we show that exposure of human T-cell lymphocytes to a coral TNFSF member purified here (AdTNF1) directly activates the death receptor pathway in humans. Taken together these data demonstrate functional conservation of TNF-induced apoptosis across 550 million years of evolution. This work also identifies the first ligand-receptor signaling pathway to be directly involved in the activation of bleaching and apoptosis in coral. Since coral bleaching events are expected to increase in frequency with future climate change, improving our understanding of the molecular mechanisms involved is prudent for reef conservation as well as our understanding of the general evolution of apoptosis [23].
Results

Bioinformatic Analysis of the Acropora digitifera Apoptotic Repertoire Reveals High-Predicted Conservation: To elucidate the complexity of the coral apoptotic repertoire we utilized the recently published genome of Acropora digitifera [13]. Thirty-one putative TNF-receptor associated factors (TRAF’s) with an average length of 458 aa were found to have high conservation with Homo sapien TRAF1 within the Tumor Necrosis Factor Receptor Associated Factor Family Domain. (Supplementary Figure 4.1A). Twenty putative Caspases with an average length of 533 aa were found to have high conservation within the α/β fold regions of H. sapien Caspase-3 including residues located within the Caspase-3 active site (Supplementary Figure 4.1B). Thirteen members of the TNFSF with an average length of 228 aa were found to have high conservation with H. sapien TNFα (HuTNFα) with the TNF-ligand Superfamily domain (Figure 4.1A). Forty putative members of the TNFRSF with an average length of 508 aa we also identified within the genome of A. digitifera. All AdTNFR’s contained a minimum of one 50s TNF-binding loop (ligand binding specificity), and one 90s binding loop (receptor oligimerization). The total number of cysteine rich domains (CRD) ranged from 0 to 4. Eleven of the putative TNFRSF’s proteins contained death domains (AdTNFR1-AdTNFR11) while five contained Ig-domains (AdTNFR18-AdTNFR22). Structural threading of AdTNF1/AdTNFR1 with two members of the human TNFSF/TNFRSF, CD40L and CD40 respectively, suggests high-predicted homology (Figure 4.1B and 4.1C). Compared to previously published work on members of the TNFRSF corals contain the most diverse TNFRSF
repertoire of any organism described to date, including humans (Supplementary Table 4.1). A. digitifera also possess other canonical apoptotic proteins including Bcl-2 members (8), Inhibitors of Apoptosis (4), APAF-1, and Cytochrome c (Figure 4.1D). See Supplementary Figure 4.2 for phylogenetic relationships between AdTNF’s, AdTNFR’s, AdCasp’s, and AdTRAF’s and H. sapien homologs.
Figure 4.1: Bioinformatic analysis of coral and human proteins involved with Death Receptor Signaling. (A) Primary sequence alignment of putative Acropora digitifera TNF-ligands with Homo sapiens TNFα. (B) Predicted structural alignment of AdTNF1 (residues 25-161) and Human CD40L. Light blue and orange represent high predicted structural homology, dark red = CD40L, dark blue = AdTNF1. (C) Predicted structural alignment of AdTNFR1 (residues 19-79) with Human CD40. Light orange and light blue represent high-predicted structural homology, dark red = CD40, dark blue = AdTNFR1. (D) The putative TNF receptor repertoire of Acropora digitifera (top) with Death Domain (DD), Cysteine Rich Domain (CRD), Immunoglobulin Domain (Ig), 50s loop-TNF binding site (red dot), and 90s loop TNF binding site (blue dot) indicated. Members of the Death Receptor Signaling Pathway (bottom) found in the Acropora digitifera genome with number of proteins within a specific protein family indicated for both Acropora digitifera and Homo Sapiens including TNF-Receptor Associated Factors (TRAFs), B-Cell Lymphoma family members (Bcl-2), Inhibitor of Apoptosis proteins (IAP’s), and Caspases.
HuTNFα causes apoptosis in Acropora yongeii: To investigate whether HuTNFα affects coral protein expression we utilized Human Explorer Antibody Arrays (Full Moon Biosystems) and found that HuTNFα led to dynamic changes of unknown coral proteins bound to human antibodies to Bcl-XL, Fas, CD40, and multiple CD-receptors (Supplementary Figures 4.3 and 4.4). To characterize the cellular response of coral to HuTNFα we first performed immunohistochemistry to demonstrate that HuTNFα binds directly to coral cells (Figures 4.2A and 4.2B). Next we exposed a 20 µm cultured coral cell to HuTNFα under live confocal microscopy and found evidence of apoptotic blebbing within 10 minutes of treatment (Figure 4.2C). Quantification of a second 7-µm cell-type extracted from adult coral tissue revealed an increase in the number of visible apoptotic cells after 90 minutes of HuTNFα exposure (Figure 4.2D). HuTNFα caused a shift in the percentage of apoptotic cells from ~15% in the untreated coral cells to ~75% in the HuTNFα treated cells (n=200 cells counted) (Figure 2E). HuTNFα exposure was also found to significantly increase (p<0.0001) Caspase activity of extracted coral cells relative to a negative control inhibitor (Z-FA-FMK, BD Pharmigen) (Figure 4.2F). Furthermore, 4 hrs of HuTNFα exposure resulted in a significant (P<0.001) increase in the total number of dead coral cells (Figure 4.2G). Taken together these data support the hypothesis that HuTNFα causes apoptosis in the reef-building coral *Acropora yongeii*. 
Figure 4.2: Effect of HuTNFα on Acropora yongei cell populations. (A). Representative coral cells incubated with HuTNFα and HuTNFα-Antibody, stained with DAPI white bar = 5 μm. (B) Representative coral cells incubated with HuTNFα-Antibody and stained with DAPI, white bar = 5 μm. (C) Live confocal microscopy of a GFP autofluorescent coral cell exposed to HuTNFα at 0, 10, and 60 minutes. White bar = 10 μm, white/green arrow indicates regions of apoptotic body formation. (D) Representative images of apparently healthy coral cells (upper panel) and apoptotic cells (lower panel) stained with DAPI at 0 and 90 minutes. Black bar = 5 μm, white arrows indicate apoptotic bodies. (E) Relative percentages of apoptotic/non-apoptotic coral cell populations left untreated or incubated with HuTNFα for 90 minutes (n = 200 cells). (F) Caspase activity of coral cells stimulated with HuTNFα for 30 minutes with negative control inhibitors (Z-FA-FMK) and pan-caspase inhibitors (Z-VAD-FMK) indicated, (**P = 0.0024; unpaired T-Test with +/- SEM indicated). (G) Total number of dead cells per image (n=32 images) under untreated conditions and HuTNFα exposure for 4 hr with interquartile ranges (boxes), and whiskers (10-90 percentiles) indicated (**P = 0.0011; unpaired T-test).


*HuTNFα causes myosin fragmentation and results in an acidic shift in the coral proteome.* To further examine the effect of HuTNFα on the coral proteome, protein was extracted from untreated coral and coral exposed to HuTNFα and analyzed with 2D-gel electrophoresis and Liquid Chromatography Mass Spectrometry (LC-MS). Following 30 minutes of HuTNFα stimulation, 92 spots were found to be significantly different (P<0.05) between the untreated and HuTNFα gels (Figure 3A), with an observed isoelectric shift towards more acidic proteins (Figure 3B). Mass spectra data was compared to a custom-built coral protein database created from the predicted proteome[13]. Four separate spots were identified as coral myosin (adi_v1.18643), which increased upon HuTNFα exposure (Figure 3A, spots 2,3,5,6). The predicted molecular mass of coral myosin is ~316 kDa while the four identified myosin spots ranged from 70-84 kDa suggesting cleavage. In addition the banding pattern observed in Figure 3C is suggestive of myosin phosphorylation. The initial induction of apoptosis also affects calcium signaling [24] and following HuTNFα exposure two proteins that contain an EF-hand calcium binding site as well as Calreticulin were found to be upregulated (Supplementary Table 4.6), suggesting that HuTNFα affects calcium signaling in corals. Finally the Zooxanthallae-specific protein Peridinin was also upregulated providing initial evidence that HuTNFα affects the coral-algal symbiosis (Figure 4.3C and Supplementary Table 4.2).
**Figure 4.3:** Proteomic analysis of coral exposed to HuTNFα using Two-Dimensional Gel Electrophoresis. (A) 2-D gel electrophoresis of extracted protein from untreated coral (left) and Human TNFα stimulated coral (right) for 30 minutes. Numbered circles indicate proteins that were identified through mass spectrometry while blue circles specifically indicate fragments of myosin. (B) 92 proteins that were significantly different between the untreated and Human TNFα treated gels grouped by their respective isoelectric points. Red bar indicate the number of proteins in a specific pI range that were downregulated in response to HuTNFα stimulation. Blue bars represent the number of proteins in a specific pI range that were upregulated in response to HuTNFα stimulation. (C) Representative proteins that were upregulated in response to HuTNFα including myosin, the zooxanthallae-specific protein Peridinin, and Calreticulin.
**HuTNFα causes bleaching of Acropora yongei:** To test whether HuTNFα-induced apoptosis is also involved in coral bleaching we utilized a flow cytometric approach. Coral tissue was found to expel zooxanthellae in the presence and absence of treatment with HuTNFα (Supplementary Figure 4.5A-B). Untreated coral initially expelled more zooxanthellae after 1 hr than HuTNFα treated coral; however, after 7 hr the HuTNFα treated coral expelled ~200% more algae than the untreated coral (Supplementary Figure 4.5C). When both untreated and HuTNFα treated corals were exposed to 100 mg mL⁻¹ of Ampicillin (AMP), the untreated control initially released more zooxanthallae than the HuTNFα treated coral. However, at 10 hr the +HuTNFα/+AMP treated coral had ~400% more expelled algae which increased to ~500% by 12 hr (Supplementary Figure 4.5D). From 6-12 hr the zooxanthellae released by the -HuTNFα/+AMP coral remained relatively constant at ~2,000 expelled algal cells (Supplementary Figure 4.5D).

**Crude AdTNF1 causes apoptosis in coral and is involved with bleaching:** To conduct preliminary investigations into the biological effects of one of the newly described coral TNF ligands (AdTNF1) we created a construct with a Prolactin signal sequence fused to AdTNF1 ensuring its secretion into the surrounding media (PBMN.i.mChAdTNF1) [25]. A diagram representing the constructs for the secretion of GFP and AdTNF1 is presented in Supplementary Figure 4.5A. As a control a construct with GFP fused to a Prolactin signal sequence was also created (pBMN.i.mChGFP). 293T cells transfected with pBMN.i.mChGFP displayed stable
expression of GFP localized to the Endoplasmic Reticulum/Trans-Golgi, as expected of secreted proteins (Supplementary Figure 4.6B). Exposure of coral cells to 10 μl of media from 293T cells transfected with pBMN.i.mChAdGFP control resulted in ~20% apoptotic cells, while cells exposed to media from the pBMN.i.mChAdTNF1 resulted in ~80% apoptotic cells (n=200, Supplementary Figure 4.6C). Adult coral tissue exposed to 250 μl of pBMN.i.mChAdTNF1 media resulted in a significant (p<0.05) reduction in total algae expelled at 4 hr post-treatment however there was no significant difference from the control by 6 hr (Supplementary Figure 4.6D). Finally the FLAG epitope was cloned into the C-terminal of AdTNF1 and evidence for direct binding of AdTNF1-FLAG to coral cells is presented in Supplementary Figures 4.6E and 4.6F.

*Purified coral AdTNF1 causes apoptosis in Human T-Lymphocytes:* To directly test whether coral AdTNF1 interacts with the human death receptor pathway we utilized Wild Type immortalized human T-lymphocytes (Jurkats) and a corresponding FADD-knockout cell line (ATCC CRL-2572™, Figure 4.4A). AdTNF1 was further purified through His-tag Nickel Affinity Chromatography (Figure 4.4B, Monserate Biotechnology Group) and used for subsequent experimentation. Bioinformatic analysis demonstrates high-predicted structural conservation between AdTNF1 and FasL [26] (Figures 4.4C). Propidium Iodide staining demonstrated that FasL negatively affects Wild-type cell viability in a dose-dependent manner while FADD-knockout cells viability are unaffected (Supplementary Figure 4.7A). Next we
exposed both Wild type and FADD-KO cells to AdTNF1. Following 48 hr of AdTNF1 treatment Wild-Type Jurkat cells exhibited a significant (p<0.0001) reduction in cell viability compared to the FADD-KO cells demonstrating that coral AdTNF1 directly interacts with the death receptor pathway in humans causing cell death (Figure 4.4D and Supplementary Figure 4.7B).
Figure 4.4: Effect of AdTNF1 on Human T-Lymphocyte cellular viability. (A) T-Cell lymphoma cell lines (Jurkat) utilized for experimentation. Wild-type (WT) cells express are sensitive to Fas-induced apoptosis while FADD-Knockout (FADD-KO) lack the FADD protein and are resistant to Fas-induced apoptosis. (B) Production and isolation of His-tagged AdTNF1 through Nickel-affinity chromatography with correct size of His-AdTNF1 indicated. (C) Predicted structural alignment of FasL and AdTNF1 Light orange and light blue represent high-predicted structural homology, dark red = FasL, dark blue = AdTNFR1. (F) The effect AdTNF1 on cellular viability of WT and FADD-KO Jurkat cell lines (****p<0.0001; unpaired T-Test).
Discussion

HuTNFα exposure led to increased Caspase activity, cellular blebbing, and cell death demonstrating that HuTNFα causes apoptosis in coral. Furthermore, AdTNF1 was found to directly interact with the *H. sapien* Death Receptor Pathway also resulting in cell death. This suggests that apoptotic signaling through TNFRSF/TNFSF proteins was fully functional at the time of the Pre-Cambrian explosion and remarkably the domains necessary to active apoptosis have been maintained from corals to humans.

A recent review by Weins et al [21] explores the origin and evolution of the TNF receptor-ligand superfamilies and concludes that their evolutionary origin could be traced back to single copy genes within arthropods. They posit that these founding genes underwent multiple duplication events following the divergence of invertebrates and vertebrates, which coincided with the development of the adaptive immune system [21]. However, this study failed to take into account the recently published Cnidarian genomes of *Nematostella vectensis* and *Acropora yongei* [2,13]. The existence of 40 putative coral TNF receptors (AdTNFR1-AdTNFR40) and 13 putative coral TNF ligands (AdTNF1-AdTNF13) identified here suggests a far more ancient origin of the TNF receptor-ligand superfamily that precedes arthropods. Our data demonstrate high conservation of TNFRSF/TNFSF members in corals suggesting a reduction of the superfamily in arthropods and other invertebrates. Corals possess more putative members of the TNFRSF of any organism described thus far, and possess a similar number of putative TNFSF proteins as many vertebrates.
While the function of AdTNR1-AdTNFR40 and AdTNF1-AdTNF13 still requires elucidation, the TNF receptor-ligand superfamily has clearly undergone dynamic changes throughout the various lineages of metazoan evolution, independent of a particular phylum’s structural complexity. Similar complexity has also been observed in the Cnidarian *Nematostella vectensis* within the Nme and Wnt gene families, a complexity that has also been lost in other model ecdyzoans [27,28]. These studies along with ours highlight the need to take into account the genomes of a broad range of animal phyla before we can draw broad conclusions about the evolution of gene families.

Beyond the specific TNF-ligand-receptor pathway, the general existence of cytokines in invertebrates has been argued to be the result of convergent evolution [29-31]. For example, in *D. melangastor* the Toll receptor pathway is involved in the response to microbial infection. Upon immune stimulation protease cascades lead to the activation of the cytokine Spatzle [32]. In humans the related pathway involves the Toll-like receptor (IL1-R1) and its respective ligand (IL-1) [33]. While both the *D. melangastor* Toll and the Human IL-1 pathways converge on the activation of NF-κB transcription factor homologs, IL-1 and Spatzle do not show any significant similarity at the amino acid level. Furthermore, the completed genome of *D. melangastor* failed to reveal any proteins homologous to human IL-1. From these data it was concluded that invertebrates lack any homologous pathway to the vertebrate IL-1/IL-1-R signaling. Beyond the specific IL-1 pathway, multiple vertebrate and invertebrate cytokines have been shown to exhibit similar biological functions, yet they lack any
It was therefore postulated that the similar biologically activities of these cytokines is a result of convergent evolution. However, the vast majority of data supporting this hypothesis is taken from the model systems of *D. melanogaster* and *C. elegans*, which as discussed above, have lost complexity in multiple gene networks. The high amino acid conservation between coral TNFRSF members and HuTNFα, as well as the activation of apoptosis in coral using a human cytokine, support the hypothesis of a divergent evolution of the TNF-ligand receptor superfamily. Future work should focus on the cytokine repertoire of other phyla to develop a comprehensive hypothesis of metazoan cytokine evolution.

The canonical apoptotic cascade is executed by a group of cysteine-dependent aspartate-directed proteases known as Caspases, which upon activation by the adaptor protein FADD cleave various cellular substrates leading to apoptotic body formation and eventual cell death. Within FADD two essential domains designated Death Domain (DD) and Death Effector Domain (DED) are required for apoptotic transduction [34] A putative FADD protein containing the DD and DED domains has been identified in both *Hydra* and the *A. digitifera* proteome (aug_v2a.04795)[35,36]. Upon activation of apoptosis the phosphorylation and cleavage of the myosin-light chain is critical for the subsequent morphological changes involved with cellular blebbing [37]. Fragments of coral myosin were found to significantly increase upon HuTNFα stimulation and the associated banding pattern in response to HuTNFα is suggestive of a phosphorylation event (Figure 4.3C). Furthermore, the acidic shift of the 92 proteins could also be the result of a larger phosphorylation cascade (Figure
Interestingly one of the most highly upregulated proteins in the HuTNFα-stimulated gel contained a Zona-Pellucida (ZP) domain, which has traditionally been studied within the context of fertilization [39]. While ZP-domain proteins have not been well studied within the context of TNF signaling in humans, pre-exposure of sperm to HuTNFα impairs sperm binding [40]. The role of ZP-domain proteins in the coral TNF signaling cascade should be a focus of future studies.

While previous studies have demonstrated apoptotic coral cells in whole-coral tissue, to the authors knowledge Figure 2 reveals the first images of an isolated coral cell undergoing cytokine-induced apoptosis [41]. We hypothesize that HuTNFα binds to one of the AdTNF-receptors containing a death domain (AdTNF1-AdTNF6), initiating the apoptotic cascade and Caspase activation (Figure 4.2F). A biochemical model of bleaching has been proposed in which Reactive Oxygen Species (ROS) production by the algal symbionts compromises the structural integrity of the mitochondrial membrane, stimulating the release of apoptotic factors and caspase activation [42]. With the identification of a diverse repertoire of 40 putative TNF receptors and 13 putative TNF ligands described here, as well as the potential involvement of the mitochondria upon HuTNFα exposure (Supplementary Figure 4.2D), we propose supplementing this model with further investigation into the specific members of the coral TNFRSF/TNFSF and their involvement in bleaching and apoptotic processes.

Previous investigations into the mechanism of coral bleaching have largely relied on thermal stress to induce zooxanthellae expulsion. While environmentally
relevant, the application of thermal stress causes dynamic changes to the coral holobiont making a determination of the specific signaling pathways directly involved in apoptosis and bleaching challenging if not impossible [43]. In this study we have induced both of these cellular processes through the application of a single protein to adult coral tissue and individual coral cells. While previous work has investigated the downstream effectors of apoptosis such as caspases [42] and Bcl-2 family members [41], this is the first examination of the upstream ligands and receptors involved with initiating apoptosis in coral. Recently published transcriptomic studies of corals exposed to various environmental stressors have implicated members of the TNFSF/TNFRSF as well as downstream proteins involved with apoptosis supporting the ecological relevance of the TNF pathway in coral [44-48].

This study reveals an ancient origin of the TNF-receptor-ligand superfamily. The activation of apoptosis in coral using a human TNF-ligand (Figure 4.2) in conjunction with the induction of apoptosis in humans using coral AdTNF1 (Figure 4.4), demonstrates remarkable evolutionary conservation that has been functionally maintained for 550 million years. While we demonstrate that AdTNF1 specifically interacts with the death receptor pathway in humans the mechanism remains unknown. Furthermore the existence of 12 additional coral TNF ligands (AdTNF2-AdTNF13) and the interactions of those TNF-ligands with human cell physiology create exciting possibilities for future research.
Materials and Methods

*Bioinformatic analysis of the TNFRSF and TNFSF in* *Acropora digitifera:*

Coral TNF ligands and receptors were extracted from the *Acropora digitifera* genome using the Pfam search option of TNF and TNFR_c6 respectively. In addition BLASTp was performed with extracted coral TNF’s against the predicted coral proteome. The DAS transmembrane prediction server [49], PROSITE database [50], and the Conserved Domain Database [51], were used to analyze the predicted proteins, while Geneious software was used for manual annotations and construction of phylogenetic trees [52]. *H. sapien* Caspase-3, *H. sapien* TRAF1, *H. sapien* Fas and *H. sapien* TNFα were downloaded from NCBI and phylogenetic trees were constructed using the Jukes-Cantor neighbor-joining method with no outgroups. Homology detection and structure prediction by HMM-HMM comparison was performed with AdTNF1/AdTNFR1 and two representatives of the human TNFSF/TNFRSF, CD40L/CD40 (GI: 4557433, 114053977) [53].

*Human antibody arrays: A. yongeii* coral nubbins were exposed to heat shock (HS), cold shock (CS), the known coral pathogen *Vibrio coralliilyticus* (V), HuTNFα and untreated (C1 and C2). Expression levels of unknown coral proteins bound to specific human antibodies were analyzed on a GenePix 4000B (Axon Instruments) and relative protein expression was determined. Interestingly an unknown coral protein bound the Fas human antibody had a relative expression of ~75%, 60%, 40% and 25% in C1, C2, HS, CS, and V treatments respectively, while HuTNFα resulted in a
complete disappearance of the protein (Supplementary Figure 4.3A). In addition the human antibody CD40 (also a member of the TNFRSF) was undetectable in C1, C2, HS, CS, and V however HuTNFα resulted in relatively high expression of an unknown coral protein bound to the Human CD40 antibody (Supplementary Figure 4.3B).

Finally a coral protein bound to the human Bcl-X\textsubscript{L} antibody maintained expression levels >95% in all treatments however, following HuTNFα stimulation the expression of this coral protein was downregulated by ~97% (Supplementary Figure 4.3C).

HuTNFα also affected the expression of unknown coral proteins bound to the antibodies of human proteins involved with mitochondrial protection, DNA damage response, cystoskeletal structure, and various CD receptors (Supplementary Figure 4.4A-4.4O). The heat shock and cold shock corals were maintained for a period of 3 hrs at 32 °C (6 °C above ambient) and 20 °C (6 °C below ambient), respectively, in a 2 L beaker with aeration. For the bacterial treatment \textit{V. coralliilyticus} was grown to a lawn on TCBS agar, scraped from the plate, and resuspended into 1 mL of seawater. 200 µl of this suspension was injected directly into the tissue of the coral with a 25-gauge needle. For the HuTNFα treatment, HuTNFα was added to the seawater at a final concentration of 1 pg mL\textsuperscript{-1} and incubated for 2 hr. To normalize for total protein added to the arrays individual protein hits were ranked by total fluorescence above background level and the following equation was used to assign a relative rank value.

\textit{2D-Gel Electrophoresis of \textit{A. yongei} stimulated with HuTNFα:} Aquaria-maintained \textit{Acropora yongei} was fragmented into two ~6 cm “nubbins” and allowed
to recover for 24 hr. Each fragment was placed into 100 mL of seawater. One fragment was incubated with HuTNF\(\alpha\) (Sigma-Aldrich, H8916) for 30 minutes at a final concentration of 1 pg mL\(^{-1}\) while the second nubbin was incubated with no additional treatment as a control. Coral nubbins were airbrushed with 0.22 \(\mu\)m filtered seawater (FSW) to remove the tissue from the skeleton and protease inhibitor cocktail (Sigma-Aldrich, P2714) was added to the blastate for a final concentration of 1X. Protein was then extracted as described by [54], resuspended in 7M Urea, 2M Thiourea, 30 mM Tris, 4% CHAPS buffer (Cell Signal), and stored at -80°C. 2D-Gel Electrophoresis was performed and significantly different spots between the treatment and control were selected for LC-MS (Aberdeen Proteomics and Samespots Software). Mass spectra data was searched against a custom-built *Acropora digitifera* database based on the predicted proteome [13].

*Production of crude AdTNF1 utilizing a Prolactin secretion scaffold*: A DNA sequence containing AdTNF1 flanked by BamHI and XhoI restriction sites was obtained (DNA 2.0), digested with BamHI and XhoI (Fermentas) and the extracted fragment was ligated into the retroviral vector pBMN.i.mCherry containing a psi promoter (stable expression), Kozak sequence (translational start site), and Prolactin signal sequence (protein secretion) [25], upstream of AdTNF1 (pBMN.i.mChAdTNF1). This vector was transformed into *Escherichia coli* XL-1 Blue amplified, and the ligated sequence was confirmed by primer extension sequencing (Eton Bioscience Inc.). A control vector containing GFP in place of
AdTNF1 was also created (pBMN.i.mChGFP). 3 μg of both vectors were separately transfected into 293T Cells (ATCC, Manassas, VA) by mixing the plasmids with 125 μl of FCS-free DMEM and 30 μg of Polyethylenimine (linear, MW 24000; Polysciences, Inc.). Cells were maintained for 48 hrs in DMEM media supplemented with 10% fetal bovine serum (Gemini Bio-Products), penicillin G (100 units mL\(^{-1}\)), glutamine (2 mM) and streptomycin (100 μg mL\(^{-1}\)). Culture media from pBMN.i.mChAdTNF1 and pBMN.i.mChGFP transfected cells was collected used for further experimentation.

**Direct stimulation of coral cells with HuTNFα and crude AdTNF1:** To initiate coral cell cultures methods were adapted from Helman et al. (2008). Briefly, coral fragments were treated with calcium-free seawater supplemented with 3% antibiotic-antimyotic solution (GIBCO) for 2.5 hr with gentle shaking followed by 30 minute incubation in 0.22 μm FSW, 25 mM HEPES buffer supplemented treated with Collagenase (Sigma-Aldrich) at a final concentration of 1.5 mg mL\(^{-1}\). Fragments were then incubated overnight in cell culture media. Coral tissue was passed through a custom build 20 μm-nylon cell strainer, washed 3 times with 0.22 μm FSW and resuspended in 3 mL of culture media in a 6 well cell culture plate. Cells were maintained under 12 hr light/dark for 1 week followed by trypsinization (Cellgro). After 48 hr an individual ~20 μm coral cell was identified and monitored in real time with confocal microscopy following HuTNFα exposure at final concentration of 1 pg mL\(^{-1}\).
A second method adapted from [55], was used to obtain a different coral cell type *A. yongei* nubbins were treated with three 15-minute incubations in a 0.22 µm FSW solution containing Ampicillin (270 µg mL⁻¹), Streptomycin (135 µg mL⁻¹), and Chloramphenicol (27 µg mL⁻¹) (Invitrogen). Corals were then submerged in cell culture media [56], in a 50 mL Falcon tube that was wrapped in aluminum foil for 48 hr. 50 µl of sloughed-off coral tissue was resuspended into 1 mL of seawater and 100 µl of coral-cell solution was placed onto a glass slide within a humidified chamber. Cells were allowed to settle for 30 minutes before direct stimulation with HuTNFα at a final concentration of 1 pg mL⁻¹ for 90 min, fixed with 2% paraformaldehyde and stained with DAPI (10 µg mL⁻¹). Slides were then visualized on a fluorescent microscope (Zeiss Axioplans Imaging). Visual apoptotic cell counts were corroborated with a Live/Dead Cell Stain Imaging Kit (Life Technologies) of coral cells exposed to HuTNFα for 4 hrs. For AdTNF1 stimulation 10 µl of media from 293T cells transfected with either pBMN.i.mChAdTNF1 or pBMN.i.mChGFP was added to 100 µl of resuspended coral cells.

**Immunohistochemistry of coral cells exposed to HuTNFα and crude AdTNF1-FLAG:** The FLAG epitope (DYKDDDDK) was cloned into the C-Terminal of AdTNF1 and confirmed with primer extension sequencing (see supplementary information for sequence and primer sets). AdTNF1-FLAG was then cloned into the Prolactin signal sequence secretion scaffold, transfected into 293T cells, and media was collected for further experimentation as described above. To obtain coral cells
adult coral tissue was incubated for 24 hrs under dark conditions in cell media [55] and centrifuged for 15 minutes at 4 C° and 4,000 rpm. Cells were washed twice in ice cold PBS, resuspended in 1 mL of seawater, and aliquoted for further incubations. Cells were then treated with either HuTNFα, 200 ul of cell media from AdTNF1-FLAG transfected cells, or left untreated on ice for 45 minutes and fixed in 2% PFA. Following fixation cells were washed twice in ice cold PBS and stained with either FLAG antibody (Sigma) or HuTNFα (R and D Systems) antibody followed by further washes in PBS and staining with a FITC-mouse secondary antibody (Sigma). Cells were then visualized on a fluorescent microscope (Zeiiss Axioplans Imaging).

**Caspase activity of coral cells exposed to HuTNFα:** Coral cells were isolated as described above and allowed to equilibrate for 48 hr. Cells were treated with a negative control inhibitor for caspases Z-FA-FMK (BD Pharmigen), negative control inhibitor and HuTNFα, or the pan-caspase inhibitor Z-VAD-FMK (BD Pharmigen) and HuTNFα. Caspase assays were performed at 31°C and activity was quantified based on the linear slope of enzyme activity.

**Flow cytometric analysis of coral bleaching:** For HuTNFα stimulation ~ 4 cm coral nubbins from the same parent colony were fragmented and placed into separate 15 mL falcon tubes containing 6 mL of seawater +/- 80 pg mL⁻¹ and 400 µl of surrounding seawater was sampled for expelled algal cells, every 2 hr for 12 hrs total. For AdTNF1 stimulation, 6 coral nubbins were fragmented into ~2cm pieces and
randomly placed into six separate 15 mL falcon tubes containing 10 mL of 0.2 FSW also supplemented with Ampicillin (final concentration of 100 µg mL\(^{-1}\)). Three of the coral nubbins were treated with 200 µl of media from 293T cells transfected with pBMN.i.mChAdTNF1 while the other three were treated with cell media from untransfected cells. Surrounding seawater was sampled every 2 hr for 12 hr total. To enumerate concentrations of algae by flow cytometry a 96-well plate containing 200 µl of each sample was loaded onto a BDFACSCanto (BD Biosciences, San Jose CA) with a High Throughput Sampler (HTS). A 100 µl aliquot was mixed three times by syringe and run using the HTS at 3 µl/second. Between each sample the probe was rinsed with 400 µl of sheath to minimize carryover of samples. To quantify Zooxanthellae, a 633 nm laser line and a 660/20 Bandpass filter were used. Additionally, a 488 nm laser line and a 530/30 Bandpass filter was used to quantify the number of coral cells in the samples. During acquisition, the threshold was set to 200 for green (530/30BP) fluorescence OR 1000 for red (660/20BP). After acquisition, data was exported as FCS3.0 and analyzed with FlowJo 7.6.3 (Treestar, Inc Ashland, OR).

*Flow cytometric analysis of Jurkat cell viability*: To quantify cell viability Propidium Iodide (PI) staining was performed and analyzed by flow cytometry in a 96-well plate format using the BDFACSCanto (BD Biosciences, San Jose CA) (2 laser (4-2 color)). Cells were seeded at 6x10\(^6\) cells/mL in a final volume of 150 µL and grown for 48 hours under various treatment conditions. PI was added to each well at a
final concentration of 1X, which was excited by the blue (488 nm) laser and analyzed in the PerCP-Cy5.5 channel (670 long pass filter preceded by a 655 long pass filter). Using a High Throughput Sampler (HTS), 30 µL aliquots were mixed three times by the syringe and run at a flow rate of 2µL/second. Between each sample 800 µL of sheath fluid was used to rinse the syringe to ensure minimal carryover from sample to sample. During acquisition the threshold was set at forward scatter (FSC) 5,000 and 1x10^5 cells were recorded. After acquisition the data was exported as FCS3.0 and analyzed with FlowJo 7.6.3 (Treestar, Inc Ashland, OR).

Acknowledgments

We thank Leo Su and the Monserate Biotechnology Group for the purification of AdTNF1, Fernando Nosratpour and Vince Levesque at the Birch Aquarium for their generous donation of Acropora yongei, the San Diego State University FACS facility, Dr. Phillip Cash for 2D-Gel analysis and Nikki Traylor-Knowles for her helpful comments with the manuscript.

Funding

This work was funded by a National Science Foundation Graduate Research Fellowship (awarded to SDQ) and NIH (1 R21 AI094534-01) awarded to FLR.
References


Acknowledgments

Chapter 4, in full, is a reprint of the material as it appears in the Proceedings of the National Academy of Sciences. Steven. D. Quistad, Aleksandr Stotland, Katie L. Baroti, Cameron Smurthweites, Brett Hilton, Juris Grasis, Roland Wolkowicz, and Forest Rohwer; 2014. The dissertation author was the primary investigator and author of this paper.
Appendix

Supplementary Figures

Supplementary Figure 4.1: Primary sequence alignments of Acropora digitifera and Homo sapiens Death Receptor Pathway proteins. (A) Primary sequence alignment of putative A. digitifera TNF-Receptor Associated Factors (TRAFs) with H. sapiens TRAF1. (B) Primary sequence alignment of putative A. digitifera Caspases with H. sapiens Caspase-3. Amino acids conferring α- and β-secondary structures of H. sapiens Caspase-3 are indicated by solid black lines while side chains located in the active sites of Homo sapiens Caspase-3 are indicated by black dots.
Supplementary Figure 4.2: Phylogenetic trees of *Acropora digitifera* and *Homo sapiens* Death Receptor Pathway proteins. (A) Phylogenetic tree of the *A. digitifera* TNF repertoire and HuTNFα with branches values indicating the number of substitutions per site. (B) Phylogenetic tree of *A. digitifera* TNF receptor repertoire and *H. sapien* Fas with branch values indicating the number of substitutions per site. (C) Phylogenetic tree of *A. digitifera* TRAF repertoire and *H. sapien* TRAF1 with branches values indicating the number of substitutions per site. (D) Phylogenetic tree of *A. digitifera* TRAF repertoire and *H. sapien* Caspase-3 with branches values indicating the number of substitutions per site.
Supplementary Figure 4.3: HuTNFα affects the expression of unknown coral proteins bound to human antibodies involved with Death Receptor Signaling. The expression of coral proteins bound to human Fas (A), CD40 (B), and Bcl-X$_L$ (C), antibodies. C1=untreated coral 1, C2= untreated coral 2, HS= heat shocked coral, CS= cold shocked coral, V=Vibrio corallilyticus injected coral, HuTNFα= Human TNFα stimulated coral. (D). Alignment of an A. digitifera Bcl-X$_L$-like protein with H. sapiens Bcl-X$_L$. 
Supplementary Figure 4.5: The effect of HuTNFα on coral bleaching. (A) Representative images of expelled zooxanthallae at 0, 6, and 12 hours of untreated (upper panel) and HuTNFα-treated (lower panel) coral. Red = zooxanthallae autofluorescence. (B), Total number of expelled zooxanthallae at 1, 6, and 12 hours. (C) Total number of algae expelled from coral in untreated and HuTNFα exposure. (D) Total number algae expelled from coral exposed to Ampicillin (AMP) and HuTNFα + AMP.
Supplementary Figure 4.6: The physiological and cellular responses of coral to crude AdTNF1. (A) Scaffold inserted into pBMN.i.mCh with Kozak sequence and Prolactin Signal Sequence. Secreted GFP (pBMN.i.mChGFP) and AdTNF1 (pBMN.i.mChAdTNF1). (B) Merged image of 293T cells transfected with pBMN.i.mChGFP under stable expression of GFP. (C) Percent of apoptotic cells following exposure to media from pBMN.i.mChAdTNF1 transfected 293T cells and secreted GFP (pBMN.i.mChGFP) (n=200). (D) Total number of algae expelled from coral in media control (n=3) and AdTNF1 exposed (n=3). (E) Representative coral cells incubated with FLAG Antibody and stained with DAPI, white bar = 5 µm. (F) Representative coral cells incubated with AdTNF1-FLAG and FLAG-Antibody, stained with DAPI white bar = 5 µm.
Supplementary Figure 4.7: Live/dead assay of Jurkat cells in the presence of FasL and purified AdTNF1. (A) Dose-dependent response of Wild-type Jurkat cells to Fas Ligand (upper panel) and lack of cellular response of FADD-KO cells to FasL (lower panel). (B) FADD-DKO Jurkat cells (upper panel) and Wild-type jurkat cells (lower panel) incubated for 48 hours in buffer only (left panel) and AdTNF1 (right panel) shows a significant decrease in percent death of the death receptor knockdown cell line compared to the WT when analyzed.
Supplementary Tables

**Supplementary Table 4.1:** Previously Identified TNF receptor/ligand family members across evolution. TNFSF= TNF super family, TNFRSF= TNF Receptor Superfamily, (DD) designates the number of TNF Receptors that contain a death domain. *= This study

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Organism</th>
<th>TNFSF</th>
<th>TNFRSF (DD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Cnidaria</td>
<td><em>A. digitifera</em></td>
<td>14</td>
<td>40 (11)</td>
</tr>
<tr>
<td>Cnidaria</td>
<td><em>N. vectensis</em></td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>Mollusk</td>
<td>Abalone/Scallop</td>
<td>2</td>
<td>1(1)</td>
</tr>
<tr>
<td>Arthropod</td>
<td><em>D. melangastor</em></td>
<td>1</td>
<td>1(0)</td>
</tr>
<tr>
<td>Echinoderm</td>
<td><em>S. purpatus</em></td>
<td>4</td>
<td>7(3)</td>
</tr>
<tr>
<td>Urochordate</td>
<td>Sea squirts</td>
<td>2</td>
<td>2(1)</td>
</tr>
<tr>
<td>Telost</td>
<td><em>D. rerio</em></td>
<td>17</td>
<td>~33(8)</td>
</tr>
<tr>
<td>Amphibian</td>
<td><em>Xenopus</em></td>
<td>13</td>
<td>&gt;3(3)</td>
</tr>
<tr>
<td>Avian</td>
<td><em>Gallus</em></td>
<td>10</td>
<td>11(4)</td>
</tr>
<tr>
<td>Mammalia</td>
<td><em>H. sapiens</em></td>
<td>18</td>
<td>29(8)</td>
</tr>
</tbody>
</table>
**Supplementary Table 4.2:** Proteins identified through mass spectrometry with the best hit to either the *Acropora digitifera* or Symbiodinium proteome.

<table>
<thead>
<tr>
<th>Spot Number</th>
<th>Protein ID/GI</th>
<th>Domains Present</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>adi_v1.07627</td>
<td>Zonapellucida</td>
<td>Coral</td>
</tr>
<tr>
<td>2</td>
<td>adi_v1.18643</td>
<td>Myosin_Tail_1, Tropomyosin, Intermediate Filament</td>
<td>Coral</td>
</tr>
<tr>
<td>3</td>
<td>adi_v1.18643</td>
<td>Myosin_Tail_1, Tropomyosin, Intermediate Filament</td>
<td>Coral</td>
</tr>
<tr>
<td>4</td>
<td>adi_v1.04802</td>
<td>Zonapellucida</td>
<td>Coral</td>
</tr>
<tr>
<td>4</td>
<td>adi_v1.04802</td>
<td>Zonapellucida</td>
<td>Coral</td>
</tr>
<tr>
<td>4</td>
<td>adi_v1.04802</td>
<td>Zonapellucida</td>
<td>Coral</td>
</tr>
<tr>
<td>5</td>
<td>adi_v1.18643</td>
<td>Myosin_Tail_1, Tropomyosin, Intermediate Filament</td>
<td>Coral</td>
</tr>
<tr>
<td>6</td>
<td>adi_v1.18643</td>
<td>Myosin_Tail_1, Tropomyosin, Intermediate Filament</td>
<td>Coral</td>
</tr>
<tr>
<td>7</td>
<td>adi_v1.16781</td>
<td>Calreticulin</td>
<td>Coral</td>
</tr>
<tr>
<td>8</td>
<td>384383661</td>
<td>Peridinin-chlorophyll a-binding protein</td>
<td>Algae</td>
</tr>
<tr>
<td>9</td>
<td>adi_v1.02613</td>
<td>AspGlu race</td>
<td>Coral</td>
</tr>
<tr>
<td>10</td>
<td>adi_v1.02613</td>
<td>AspGlu race</td>
<td>Coral</td>
</tr>
<tr>
<td>11</td>
<td>adi_v1.04828</td>
<td>14-3-3</td>
<td>Coral</td>
</tr>
<tr>
<td>12</td>
<td>adi_v1.01410</td>
<td>AhpC-TSA, Redoxin, 1-cysPrx C</td>
<td>Coral</td>
</tr>
<tr>
<td>13</td>
<td>adi_v1.23158</td>
<td>efhand</td>
<td>Coral</td>
</tr>
<tr>
<td>14</td>
<td>adi_v1.05155</td>
<td>14-3-3</td>
<td>Coral</td>
</tr>
<tr>
<td>15</td>
<td>adi_v1.14836</td>
<td>efhand</td>
<td>Coral</td>
</tr>
<tr>
<td>16</td>
<td>adi_v1.10342</td>
<td>E1_DerP2_DerF2</td>
<td>Coral</td>
</tr>
</tbody>
</table>
CHAPTER 5

Viruses and the origin of metazoan immunity

Abstract

The Cambrian Explosion occurred approximately half a billion years ago and is widely considered to be amongst the most dramatic periods in the history of life on the earth. It was during this relatively short period of time that the majority of extant animal phyla first appeared in the fossil record. With the explosion of animal life came a concurrent increase of self vs non-self interactions resulting in phyla-specific immunity. Yet, mechanisms linking the Cambrian Explosion to the origin of immunity are lacking. Here we propose a model for the evolution of metazoan immunity mediated by microbial interactions.
Summary

First, an ancient endogenous viral infection of a proto-metazoan ancestor led to viral control of the apoptotic machinery and provided the host with pro-viral immunity against related virus species. Specifically this endogenization event provided the host with an anti-apoptotic protein (c-FLIP) that required constant expression to prevent cell death, thus ‘addicting’ the host to the virus and forming a viral addiction module consisting of a Tumor Necrosis Factor Receptor (TNFR), FADD protein, and c-FLIP. With multicellularity also came the first mucus-secreting epithelial cell surface that attracted founding bacterial populations via chemosensory mechanisms. These symbiotic bacterial communities were maintained via phage-predation within the surface mucus layer providing the host with immune protection against bacterial pathogens. Based on the ubiquity of the mucosal environment and the conservation of TNF-induced apoptosis we propose that both of components emerged during the Precambrian era and continue to drive the evolution of immunity in modern-day phyla (Figure 5.1).
**Figure 5.1:** A proposed model for the development of metazoan immunity. Single-cell eukaryotic organisms organized into multicellular metazoans consisting of two cell layers (ectoderm and endoderm). Viral infection of a proto-metazoan ancestor led to viral control of the TNFR apoptotic machinery and host addiction to the virus. Maintenance of the virus within the host provided pro-viral immunity against related viral species. The first multicellular metazoans also provided an energy-rich mucin environment that was colonized by microbial populations (phage and bacteria). Symbiotic bacterial species were maintained by phage predation providing the host with immune protection against invading bacterial pathogens.
Introduction

A functional immune system depends on the proper recognition and response to self and non-self interactions. Failure with either of these interactions can result in catastrophic loss of organismal integrity and ultimately death [1]. Many immune components have been discovered within the context of pathogenesis [2,3]. By understanding how a pathogen causes diseases we are able to better understand how the immune system works. This has led to the erroneous implication that pathogens exist and interact within a two-dimensional space: host-pathogen [4]. In reality all biological interactions including pathogenesis occur within the context of an ecological community; the prepared third party will always fill vacant niche space [5]. This two-dimensional bias is exemplified through the pervasive use of “anti-viral” and “anti-microbial” when describing host-microbe interactions [6,7]. These terms suggest that immune responses are targeted against a specific pathogen, and once that pathogen is eliminated the story is over. However any compound that targets a specific microbe will create novel niche space within that particular ecosystem once that microbe is destroyed [8]. Therefore we will utilize the terms “selective-viral” and “selective-microbial” to describe any mechanism that is involved with maintaining a specific microbial community.
Viruses and the origin of TNF-induced apoptosis

Apoptosis and multicellularity: The transition from single-cell eukaryotes to multicellular metazoans created an organism consisting of two cell layers; an ectoderm exposed to the extracellular environment and an inward facing endoderm. To maintain organismal integrity a multicellular organism must decide what to do with damaged cell populations. For example once a virus traverses the extracellular membrane and enters the intracellular space the host has two options; it can either fight off the virus or kill the infected cell [3]. The first option involves the detection and destruction of an intracellular pathogen while maintaining and restoring normal cellular function. However, if the host determines that a cell is not worth saving and the risk of spreading the infection is too great it can undergo programmed cell death or apoptosis [9]. This choice between cell survival and death is a complex decision that must take into account both the cost of losing a cell and the cost of resisting infection [10]. Any coordinated response to resist infection increases the total time of pathogen persistence within the host. Therefore fighting for the survival of a pluripotent stem cell responsible for gamete production may be worth more to a host compared to a differentiated skin cell. Once an apoptotic signal is received the cell undergoes a carefully coordinated sequence of events ultimately resulting in death. First, specialized proteases called executioner Caspases cleave host DNA and other complex macromolecules [11]. These digested cellular components are then packaged into spherical membrane structures called “apoptotic bodies” which are recycled by macrophages [12]. Apoptosis is a requirement for all multicellular life to exist however
metazans apoptosis appears to be unique through its use of the Tumor Necrosis Factor Receptors (TNFR’s), which are involved with the detection and signal transduction of apoptotic signals [13,14]. We propose that TNF-induced apoptosis emerged during the Precambrian as a selective-viral mechanism through which the host was provided immune protection from related viruses as long as persistent viral infection was maintained.

While viruses are commonly perceived as pathogens associated with disease recent metagenomic studies suggest that persistent, asymptomatic viral infections are more common than previously thought [15]. Persistent viral infection can provide a host with immune protection against related viral species that may normally cause acute infection leading to a the survival of the persistently infected [16]. For example, in mice a mother infected with Mouse Hepatitis Virus (MHV) will pass on Ig antibodies to her offspring through milk production. These Ig antibodies protect the young from lethal Central Nervous System MHV infection thus establishing viral persistence in the next generation [17]. Therefore, from an evolutionary perspective both the virus and the host can benefit from persistent infection. Another approach taken by viruses to ensure persistence within the host involves the production of toxin/antitoxin systems that create a “viral addiction module” [18]. These systems generally involve the production of a relatively stable but harmful protein that associates with a less stable, protective protein. Therefore the protective protein provides the host with immune protection however it must be continuously expressed due to its low stability. Any disruption to viral-host interactions will result in the
degradation of the protective protein resulting allowing the harmful protein to induce cell-death. The first addiction module described involved P1 phage and *Escherichia coli*. This addiction module involves the death-inducing protein doc which is sequestered by the protective protein phd as long as the prophage is maintained within the host [19,20]. In metazoans the major components of TNFR-mediated apoptosis create an ideal system to investigate viral addiction modules.

*TNFRs and viral addiction:* Members of the Tumor Necrosis Factor Receptor (TNFR) superfamily are type I transmembrane proteins that consist of an extracellular N terminus with a cytoplasmic C terminus [21]. The defining module of the extracellular region is a Cysteine Rich Domain (CRD) that is involved with ligand binding and receptor oligomerization [22]. Some TNFRs also contain a Death Domain (DD) that once activated will result in Caspase activation and eventual cell death [11,14]. Under normal cellular conditions Fas-Associated Death Domain protein (FADD) is bound to the DD of TNFRs, inhibiting apoptosis. However, upon receiving an apoptotic stimulus FADD disassociates from TNFRs and activates the apoptotic cascade. TNF-induced apoptosis via FADD signaling is apparently conserved from corals to humans suggesting an ancient Precambrian origin [13]. Given FADD’s central position within the apoptotic decision making process it is not surprising that it is targeted by an array of viral proteins. These viral proteins termed v-FLIPs are expressed by members of the *Herpesviridae* family and bind directly to the Death Inducing Signaling Complex, inhibiting apoptosis [23-25]. The cellular homolog to v-
FLIP, c-FLIP, has been demonstrated to interact directly with FADD and also acts as an antiapoptotic regulator [25,26]. Taken together the TNFR-FADD-c-FLIP signaling pathway possesses all of the components required for a viral addiction system. Constant expression of c-FLIP is required to prevent cell death and loss of c-FLIP expression will result in FADD-mediated apoptosis (Figure 5.2).

**Figure 5.2**: The TNFR-FADD-c-FLIP addiction module. Under normal cellular conditions the FADD protein is bound directly to the Death Domain (DD) of a TNFR preventing apoptosis. A proposed viral infection resulted in an endogenization event and donated the c-FLIP protein to the host. c-FLIP binds to FADD resulting in Death Domain signaling and apoptosis thus giving the endogenized virus control over the apoptotic decision making process.
To investigate a potential viral origin of c-FLIP the genomic scaffold that express a coral c-FLIP homolog was extracted and examined for viral-like characteristics. The c-FLIP-expressing scaffold was found to contain multiple reverse transcriptase domains and Long Terminal Repeats (LTRs) suggesting a viral origin (Figure 5.3).

**Figure 5.3:** Genomic scaffold taken from *Acropora digitifera* expressing a c-FLIP homolog with predicted viral-domains indicated. RT-like = reverse transcriptase-like, RT-nLTR = reverse-transcriptase non-long terminal repeat retrotransposon.

Furthermore, the existence of multiple v-FLIPs in modern-day Herpesviruses provides further evidence that the original c-FLIP protein could have been co-opted from an endogenized virus. While Herpes viruses are not commonly associated with endogenization multiple events have been reported [27]. Interestingly corals have been shown to harbor Herpes viruses and similar to vertebrates, the abundance of Herpes viruses increase during physiological stress [28]. Herpes viruses infect nerve cells and corals belong to the phyla proposed to have the first nervous system [29]. Therefore, co-evolution between Herpesviridae and Metazoans may have been occurring since the Precambrian era. Following the initial virus infection host survival became dependent on c-FLIP expression thus requiring the host to maintain persistent viral infection resulting in a viral addiction module. The conservation of TNF-induced apoptosis via FADD signaling suggests that this addiction module was rapidly fixed
into metazoan populations and has been maintained for over half a billion years. We speculate that the TNFR-FADD-c-FLIP addiction module may have provided the first metazoans with protection against related viruses thus acting as one of the first selective-virals in animal evolution.

**Phage: an ancient metazoan selective-microbial**

*The ubiquitous mucosal environment:* Mucosal environments can be found on the surface of epithelial cells in virtually all metazoans [30,31]. These environments are defined by the highly variable mucin macromolecule, which consists of an oligosaccaride backbone covalently bonded to protein side-chains [32]. Mucins can either be tethered to the cell surface or secreted by the host forming a dense gel matrix that physically separates the host from the surrounding environment [33]. Whether the system of interest is the human intestine or the tentacles of a reef-building coral the same processes and characteristics exist. These include a dynamic environment undergoing constant turnover of mucin molecules colonized by diverse and specific microbes. For example in one day the human gastrointestinal (GI) tract replaces the entire mucin pool while corals can release up to 4.8 liters of mucus per square meter of reef [34,35]. This pool of energy-rich glycoprotein serves as a food source for a variety of symbiotic bacteria that can provide the host with nutrients such as nitrogen and phosphorous [31,36]. The glycosylation patterns of the mucin pool are distinct throughout the GI tract suggesting that specific mucins may select for specific bacterial species [32]. Disturbance of these bacterial populations has been associated
with a variety of human and coral diseases therefore selection pressure maintains specific microbial-host associations and provides a first line of immunological defense to the host [37].

*Niche exclusion and immunology:* The distribution of microbes within the mucosal environment is controlled by the availability of niche space. The ecological concept of niche space has already been applied to the field of immunology within the context of immunosuppression and opportunistic infections. For example the HIV virus destroys T-cell surveillance creating novel niche space within the human lungs. This niche space can be filled by microbes such as *Tuburcle bacillus* resulting in Tuberculosis [38]. The lung niche space is a nutrient rich environment that a variety of microbes could easily thrive in however, that niche space is usually unavailable due to immune surveillance [39]. Competition for niche space is a major driver of community composition and in healthy individuals *Tuburcle bacillus* is out-competed for niche space by the host immune system. This microbe-host competition can be broadened to include the entire microbial community and how is competes with the host immune system for niche space. Perturbations of the microbial community can result in immune dysfunction and subsequent pathologies such as inflammatory bowel disease and colon cancer in humans [40].

All metazoans investigated thus far appear to harbor highly specific microbial communities that are associated with the maintenance of host health therefore the first major component of the metazoan immune system was probably a selective-microbial
The first metazoans emerged approximately 550 million years ago during the Cambrian Explosion and likely resembled modern-day cnidarians [43]. The cnidarian body plan is morphologically simple and consists of just two cell layers, an ectoderm and endoderm [43,44]. Therefore, the initial host-microbe interactions and subsequent selection for specific microbial populations likely occurred on a mucosal surface exposed to the marine environment. Bacteriophage (phage) are the most dynamic biological entity on the planet and are able to kill specifically, evolve rapidly, and transfer genes broadly [45]. Therefore, we propose phage as the initial selection pressure that structured the first metazoan microbial communities.

**BAM: an ancient metazoan selective microbial:** In addition to harboring distinct bacterial populations the mucosal surface also supports a dynamic and abundant community of phage [46,47]. Contact between a phage and host bacterial cell can result in either bacterial cell lysis or integration into the host genome through lysogeny [48]. In addition to carrying the necessary genes for replication phage can also provide the host with novel auxiliary genes that affect host fitness [49]. For example, cyanobacteria are estimated to account for 25% of global photosynthesis and the genomes of their associated phage demonstrate that they provide the host with functional photosynthesis genes [50]. Lysogeny can also provide the bacterial host with protection from other phage as well as maintain symbiotic relationships between the bacterial and eukaryotic hosts [51]. Taken together phage regulate bacterial populations through the donation of novel genetic material as well as the direct killing
of a specific bacterial species. The abundance of phage within the surface mucosa layer is significantly higher compared to the surrounding environment. For example, in corals the abundance of Virus-Like Particles (VLPs) within the surface mucosa can be 17-fold higher compared to the surrounding seawater and metagenomic analysis suggests the majority of those VLPs are phage [46,52]. The increased abundance of phage in mucus has also been demonstrated in a variety of organisms including sea anemones, worms, fish, and mice. This observation led Barr et al. to investigate how phage are able to maintain such large populations within an environment that is undergoing constant turnover [53]. Recent work by Minot et. al. demonstrated that phages in the human gut encode hypervariable regions and many of these regions encode proteins with Ig-like domains that are capable of accommodating $>10^{13}$ variations [54,55]. Mutations within these regions had little effect on viral growth therefore these Ig-like proteins were hypothesized to be involved with viral tropism [55]. To test this hypothesis Barr et al. used a T4 in vitro model system to demonstrate that phage stick to mucin molecules through Ig-like domains and protect human tissue culture cells from bacterial-induced cell death. Application of the BAM hypothesis to a bacterial community inhabiting a mucosal surface predicts that phage provide protection from pathogens and actively contribute to the maintenance of symbiotic populations (Figure 5.4).
Figure 5.4: The Bacterial Adherence to Mucus (BAM) model of immunity. The founding bacterial populations were attracted to secreted mucin molecules via chemosensory mechanisms. In addition to bacteria mucins also attracted phage, which bound directly to mucin molecules. These specific phage populations offered immune protection to invading pathogens.

The BAM model utilizes an anthropomorphic perspective to describe how phage protect the host from potential pathogens, providing a form of immunity. However if we take the microbial perspective the mucosal surface is simply a habitat that contains higher food abundance and physical stability compared to the surrounding aqueous environment. For example the nitrogen and carbon content of coral mucus is approximately twenty-three times higher compared to the surrounding seawater [56]. In addition the gel-like matrix created by mucin polymers acts as an efficient particle trap [35]. Due to the high nutrient concentrations available within mucus compared to the surrounding environment bacteria would have located mucosal surfaces through chemotaxis and brought with them their resident phages. Phages do not possess the flagellate structures that bacteria use for locomotion and therefore they rely on diffusion to find their prey. The high viscosity of mucus would increase the residence time of phage and selection pressure would favor phage with additional
mechanisms to stick around such as Ig-like domains. Once established within the mucosal environment selection pressure on the bacteria populations would favor maintaining the constant supply of food (mucins), which would translate to protection of the host epithelia serving as a form of immunity. From the microbial perspective the establishment of the first metazoan mucosal community was simply a colonization event of a novel environment. What made this event unique was that the microbial-mediated maintenance of the mucosal environment reduced the energy requirement of the host to maintain epithelial integrity. This previously unavailable pool of energy could then provide the basis for the development of morphologically complex and energy expensive structures (i.e. metazoan body plans).

Conclusions

Future Perspectives: Here we have proposed a model for the development of metazoan immunity via extracellular microbial population dynamics and intracellular viral addiction modules (Figure 1). In the last common metazoan ancestor mucins were released from epithelial tissue and attracted bacteria through chemosensory mechanisms. With the first bacteria populations came the predators of bacteria, phage, which selected for specific bacterial species that benefited the host. In addition to attracting bacteria and phage, mucins also increased the rate of contact with eukaryotic viruses. To counter this viral onslaught specific viral species provided the host with protection against related viruses through addiction modules such as the TNFR-FADD-c-FLIP described here. During the Precambrian era and into the present day
phylla-specific interactions with microbial communities continue to serve as major drivers in the evolution of metazoan immunity.

While the evolutionary progression of metazoan immunity cannot be reproduced within the laboratory the individual hypotheses proposed here can be directly tested. First, the BAM model has yet to be tested within the context of ecological communities and should be a major focus of future investigations. Specifically, demonstrating that the selection and maintenance of specific phage populations promotes the growth symbiotic bacterial populations. To investigate the proposed TNFR-FADD-c-FLIP viral addiction module additional genomes should be examined for viral-like characteristics associated with c-FLIP scaffolds. In addition functional studies are required to directly test the proposed addiction module via biochemical approaches.
References


Acknowledgments

Chapter 5, in full, is in preparation for submission. Steven D. Quistad and Forest Rohwer; 2015. The dissertation author was the primary investigator and author of this paper.
CHAPTER 6

Conclusions

Summary

Viruses have driven the evolution of metazoan life since the Precambrian era. Chapter 2 provides bioinformatic evidence that viral populations can successfully be used as tools to predict the structure of host immunity. In chapter 3 an ancient protocnidarian retroviral infection is identified and a general model of evolution mediated by the expansion of endogenous retroviruses is proposed. Chapter 4 explores the origins of apoptosis and provides molecular data demonstrating that TNF-induced apoptosis has been functionally conserved for over 550 million years. Finally chapter 5 a model is proposed for the development of metazoan immunity mediated by extracellular microbial populations and intracellular viral addiction modules. This final chapter outlines future challenges and provides approaches for future investigations.

Future Directions

Viral populations as immune predictors: Chapter 2 provides bioinformatic evidence that predicts viral-host interactions in coral however these interactions require direct experimentation. As a fellow anthozoan, the sea anemone is closely related to corals and provides well-established molecular techniques to test predicted viral-host interactions at the molecular level [1-3]. However, in-situ hybridization methods are established in corals therefore the localization of
specific viral transcript could provide insight into how that viral species affects host
physiology [4]. A recent study in American Samoa found that heat-tolerant corals
have significantly higher expression of TNFRs compared to heat-sensitive individuals
[5]. An interesting direction for future research could include investigating whether
specific viral genes are also associated with the corals ability to resist heat shock or
other physiological stressors.

ERV infection of the Proto-cnidarian ancestor: Chapter 3 identifies the oldest
retroviral infection described thus far and predicts that the LTR region of the novel
ERV family is involved with cnidarian-specific developmental traits via Fibroblast
Growth Factor Receptor (FGFR) signaling. The involvement of FGFRs has already
been investigated in sea anemones and therefore would be a good system to test the
proposed hypothesis [6-8]. To directly test whether the LTR is involved with FGF
signaling during development the LTR region could be deleted using the CRISPR-Cas
system and the resulting molecular or morphological phenotype observed [9].

The evolution of TNF-induced apoptosis: Chapter 4 demonstrates that TNF-
induced apoptosis has been functionally maintained for over half a billion years
however the specific components involved have yet to be investigated. For example
corals possess 40 TNFRs, 14 TNF ligands, 31 TRAFs, and 22 caspases. Which of
these components (if any) were directly involved with the TNF-induced apoptotic
response requires additional experimentation. To identify which coral receptor that the
human TNF ligand is actually binding to Co-immunoprecipitation should be performed. Further, why corals have the most TNFRs compared to any organism investigated thus far is a fascinating question. As described above specific TNFRs appear to be involved with heat resistance however the mechanism behind the observed TNFR-mediated heat resistance has yet to be investigated. To identify additional examples of evolutionary conservation between corals and humans mechanistic-based studies of specific TNFRs could be performed in-vitro using human tissue cell lines.

The future of cnidarian biology: Progress in the field of cnidarian biology requires an integrative approach that utilizes multiple systems. With the availability of the three cnidarian genomes including the sea anemone (*Nematostella vectensis*) [3], Hydra (*Hydra magnipapillata*) [10], and coral (*Acropora digitifera*) [11] each organism offers unique benefits depending on the types of questions being investigated. For example in both hydra and sea anemones the expression of genes of interest can be knocked down using RNAi methods and transgenic animals are also available [1,2,12,13]. In contrast to *H. magnipapillata* hydra and *N. vectensis* reef-building corals harbor a diverse assemblage of microorganisms including a symbiotic algae. The field of coral biology is also supported by large historical datasets that investigate the roles of oceanography, ecology, and microbiology in shaping the coral reef system. Taken together the available experimental tools of hydra and sea anemone provide a better system to test molecular based hypotheses while the coral system provides an excellent system to investigate ecologically-based questions. As additional
methods are developed in cnidarians it is important to not become too focused on a single model system. Cnidarians provide a unique evolutionary perspective based on their basal position within metazoan evolution however the majority of extant phyla have yet to be fully investigated [14]. Therefore broad evolutionary conclusions require data from a broad representation of phyla.

Conclusion

The primary goal of this dissertation was to examine how viruses have impacted the evolution of metazoan life. In chapter 2 a predicted viral-host interactome is presented and a general pipeline is proposed to rapidly predict the immune system of virtually any animal. Chapter 3 broadens the scope of how viruses impact host biology to include the evolution of host genomes and the origin of phyla-specific developmental characteristics. The proposed hypothesis of a proto-cnidarian retroviral infection can be directly tested using the available molecular tools of multiple cnidarian systems. Chapter 4 provides a focused investigation into the canonical antiviral response of programmed cell death or apoptosis with a proposed viral origin of TNF-induced apoptosis discussed in Chapter 5. Taken together this dissertation provides novel insight into the unexpected evolutionary conservation between corals and humans and further supports the significant role viruses have played in shaping the evolution of host biology.
References


