

2021

Structure, activity, and biology of transcription factor NF-kappaB in evolutionarily basal organisms: insights into the origins of immune regulation

<https://hdl.handle.net/2144/43027>

"Downloaded from OpenBU. Boston University's institutional repository."

BOSTON UNIVERSITY
GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

**STRUCTURE, ACTIVITY, AND BIOLOGY OF TRANSCRIPTION
FACTOR NF-kappaB IN EVOLUTIONARILY BASAL ORGANISMS:
INSIGHTS INTO THE ORIGINS OF IMMUNE REGULATION**

by

LEAH MICHELE WILLIAMS

B.A., Wheaton College, Massachusetts, 2013
M.A., Boston University, 2019

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2021

Approved by

First Reader

Thomas D. Gilmore, Ph.D.
Professor of Biology
Professor of Biochemistry

Second Reader

Sarah W. Davies, Ph.D.
Assistant Professor of Biology

ACKNOWLEDGMENTS

I have had the privilege of being mentored by Dr. Thomas Gilmore. His talent, tact, and attention to detail have made my PhD not only successful, but incredibly enjoyable. I am forever grateful for his mentorship and support during my PhD training, and am lucky enough to consider him a mentor and friend.

I would also like to thank my committee members, Drs. Sarah Davies, John “Chip” Celenza, Sean Mullen, and Trevor Siggers, for their suggestions, support, and guidance throughout my research conducted at Boston University.

I owe a great debt of gratitude to numerous people who have contributed to this thesis project. Specifically, I would like to thank my lab mate Chris DiRusso for his close friendship and family-like support through nearly all of my graduate career. Life would never be as *good* without him in it. I would also like to thank Melissa Inge for her contribution to this research and her caring support both in and out of the lab, and making sure things are always going well. I greatly appreciate the support of past lab members Dr. Joseph Brennan, Dr. Kate Mansfield, Jon Messerschmidt, Nicole Carter, Milad Babaei, David Lui, and Ethan McCaslin the numerous undergraduate and high school students whom I worked with, including the BB522 classes of 2016-2020 and Sainetra Sridhar, who contributed to this work, and current lab members Mengrui Wang, Niharika Desai, Maria Valadez Ingersoll, and Wei Wang for their assistance and kindness.

Finally, and most importantly: everything I am, I owe to my family who grounds me. I would like to thank my parents, Don and Audrey, for their continued support and unconditional love, my sister, Ellie, for always being there for me, and for my brothers,

Ben and William, who also know when to pop in for a good laugh. I'd also like to thank my Grandma Lowell as my first motivation for becoming a scientist, and for telling me to trust my decisions. Without that advice, I would certainly be lost, and I wish she was still here to see me finish what success she inspired. I admire her, I miss her, and I wish she could have kept seeing me try to become half the scientist she was. Finally, I want to thank my partner Joshua for his incredible patience and for being my warmest advocate. His love and friendship are the favorite discovery I've made in my graduate career.

**STRUCTURE, ACTIVITY, AND BIOLOGY OF TRANSCRIPTION
FACTOR NF-kappaB IN EVOLUTIONARILY BASAL ORGANISMS:
INSIGHTS INTO THE ORIGINS OF IMMUNE REGULATION**

LEAH MICHELE WILLIAMS

Boston University Graduate School of Arts and Sciences, 2021

Major Professor: Thomas D. Gilmore, Professor of Biology, Professor of Biochemistry

ABSTRACT

Over the past 30 years, transcription factor nuclear factor kappa B (NF- κ B) has been extensively characterized in organisms ranging from flies to humans, where it is known to play key roles in developmental and immune-related processes. More recently, DNA sequencing approaches have identified homologs of NF- κ B and many upstream signaling components in basal phyla, including Cnidaria (sea anemones, corals, hydras, and jellyfish), Porifera (sponges), and single-celled protists, including *Capsaspora owczarzaki* and some choanoflagellates. However, little is known about the activity and regulation of NF- κ B proteins in these basal organisms. In this dissertation, the structure, activity, and biology of NF- κ B in three basal phyla is examined and the extent of conservation with more derived organisms as well as phylum-specific properties are investigated. In the coral *Orbicella faveolata* (Of) a simplified but nearly complete Toll-like receptor (TLR)-to-NF- κ B pathway exists, but basal to cnidarians, there are fewer upstream signaling molecules present. For example, in the poriferan *Amphimedon queenslandica* (Aq) and the protist *Capsaspora owczarzaki* (Co), singular NF- κ Bs and some upstream signaling proteins are encoded in their genomes, but no canonical TLRs

exist. In contrast, the expanded family of choanoflagellates, including the choanoflagellate *Acanthoecca spectabilis* (As), contains TLR-like and up to three NF- κ B-like homologs, although their domain structures differ from NF- κ B pathway members of higher organisms. Of-NF- κ B, Aq-NF- κ B, and Co-NF- κ B all resemble the mammalian NF- κ B protein p100 in that they contain an N-terminal DNA-binding domain, a C-terminal Ankyrin (ANK) repeat domain, and similar DNA binding-site profiles. C-terminal truncation results in translocation of these basal NF- κ Bs to the nucleus and increases their DNA-binding and transcriptional activation activities. Nevertheless, unlike mammalian NF- κ B p100, the C-terminal sequences of Aq-NF- κ B do not inhibit its DNA-binding activity. The three As-NF- κ B-like proteins all consist of primarily the N-terminal conserved Rel Homology domain sequences of NF- κ B, but lack C-terminal ANK repeats. All three As-NF- κ B proteins constitutively enter the nucleus of human and Co cells, but differ from one another in DNA-binding and transcriptional activation activities. Furthermore, all three As-NF- κ B proteins can form heterodimers, indicating that NF- κ B diversified into multi-subunit families at least two times during evolution. Expression of IKKs induce proteasome-dependent C-terminal processing of Of-NF- κ B and Aq-NF- κ B in human cells, and processing requires C-terminal serines. In contrast, C-terminal processing of Co-NF- κ B is not induced by co-expression of IKK in human cells and no IKK homolog exists in the Co genome, suggesting that IKK-mediated processing of NF- κ B is a mechanism that evolved solely in animals. Treatment of Of and sponge tissue with lipopolysaccharide (LPS), a ligand for mammalian innate immunity, results in gene expression changes consistent with NF- κ B pathway mobilization in Of and increases both

DNA-binding activity and processing of sponge NF- κ B. Furthermore, sponge tissue contains constitutive NF- κ B site DNA-binding activity, as well as nuclear and processed NF- κ B. Moreover, exogenously expressed Co-NF- κ B localizes to the nucleus in Co cells. Together, these data suggest that the mechanism as well as level of activation of NF- κ B in basal organisms is different from what is observed in higher organisms. Additionally, NF- κ B mRNA and DNA-binding levels differ across three life stages of *Capsaspora*, suggesting distinct roles for NF- κ B in these life stages. RNA-seq and GO analyses identify possible gene targets and biological functions of Co-NF- κ B. Overall, these data represent the first functional characterization of NF- κ B signaling proteins in an endangered coral, in any organisms basal to cnidarians (i.e., an evolutionary important sponge), and outside the Kingdom Animalia (protists). These findings suggest that these seemingly simple organisms contain conserved innate immune-like pathways that may be regulated by NF- κ B and provide information about the evolution and diversification of this biologically important transcription factor.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
TABLE OF CONTENTS.....	ix
LIST OF TABLES.....	xiii
LIST OF FIGURES	xv
LIST OF ABBREVIATIONS.....	xviii
CHAPTER ONE	1
INTRODUCTION	1
1.1 Introduction to transcription factor NF- κ B	1
1.2 Structure of basal NF- κ B proteins	5
1.3 Activity and regulation of basal NF- κ Bs	9
1.4 Basal NF- κ Bs likely have roles in development and immunity	14
1.5 Thesis rationale	20
CHAPTER TWO	29
MATERIALS AND METHODS.....	29
2.1 Phylogenetic analyses of conserved genes	29
2.2 Codon optimization of cDNA sequences.....	31
2.3 Recombinant DNA techniques	31
2.4 Cell culture.....	34
2.5 Whole-cell extract preparation.....	36
2.6 SDS-Polyacrylamide gel electrophoresis and Western blotting	38
2.7 Electrophoretic mobility shift assay (EMSA).....	39

2.8	GST-tagged protein purification	40
2.9	GST-pull down assays	41
2.10	Reporter gene assays.....	42
2.11	<i>In vitro</i> kinase assay.....	44
2.12	Indirect immunofluorescence.....	45
2.13	Treatment with LPS	47
2.14	Immunoprecipitations	48
2.15	<i>Cliona sp.</i> taxon identification by CO1 gene analysis.....	48
2.16	<i>Capsaspora owczarzaki</i> RNA-seq analysis	49
CHAPTER THREE		93
THE ENDANGERED CORAL <i>ORBICELLA FAVEOLATA</i> HAS A CONSERVED TOLL-LIKE RECEPTOR-TO-NF- κ B SIGNALING PATHWAY		93
3.1	Introduction.....	93
3.2	Conservation of Toll-like Receptor (TLR) pathway in <i>Orbicella faveolata</i>	96
3.3	A predicted TLR-to-NF- κ B pathway in <i>Orbicella faveolata</i>	97
3.4	The structure and phylogenetic analysis of the <i>O. faveolata</i> NF- κ B protein indicates that it is more similar to mammalian NF- κ B proteins than to Rel proteins.....	98
3.5	C-terminal truncation of the <i>O. faveolata</i> NF- κ B protein is required for nuclear localization, DNA binding, and transactivation.....	99
3.6	C-terminal truncation of Of-NF- κ B can be induced by an IKK-dependent mechanism in human cells	101

3.7	Chapter 3 summary	103
CHAPTER FOUR.....		133
PORIFERAN TRANSCRIPTION FACTOR NF- κ B HAS CONSERVED AND UNIQUE SEQUENCES, ACTIVITIES, AND REGULATION.....		133
4.1	Introduction.....	133
4.2	<i>A. queenslandica</i> NF- κ B resembles other metazoan NF- κ B proteins both structurally and phylogenetically	136
4.3	Aq-NF- κ B can bind DNA and activate transcription	137
4.4	C-terminal truncation of Aq-NF- κ B enables it to translocate to the nucleus and can be induced by an IKK-dependent mechanism in vertebrate cells	139
4.5	Sponge tissue expresses putative NF- κ B proteins and contains NF- κ B site DNA-binding activity that increases following treatment with bacterial lipopolysaccharide (LPS).....	142
4.6	Chapter 4 summary	144
CHAPTER FIVE		161
DIVERSIFICATION OF TRANSCRIPTION FACTOR NF- κ B IN PROTISTS ⁵		161
5.1	Introduction.....	161
5.2	Protist NF- κ B proteins vary in domain structure and choanoflagellates show evidence of gene duplication	164
5.3	DNA binding, nuclear translocation, and transactivation by Co-NF- κ B...	165
5.4	IKK-mediated processing of NF- κ B appears to have evolved with the rise of multicellularity	168

5.5	Exogenously expressed full-length and truncated versions of Co-NF- κ B localize primarily to the nucleus in <i>Capsaspora</i> cells	169
5.6	Co-NF- κ B mRNA levels and DNA-binding activity vary coordinately across different life stages and the identification of putative NF- κ B target genes	170
5.7	Choanoflagellate NF- κ Bs can form heterodimers and have different abilities to bind DNA and activate transcription	173
5.8	Chapter 5 summary	175
CHAPTER SIX		257
DISCUSSION		257
6.1	Summary	257
6.2	Chapter Three Discussion	258
6.3	Chapter Four Discussion	262
6.4	Chapter Five Discussion	269
6.5	Unanswered questions and future experimental goals	276
6.6	Conclusions and perspectives	281
BIBLIOGRAPHY		287
CURRICULUM VITAE		304

LIST OF TABLES

Table 1.1	Functions of NF- κ B in basal organisms.	22
Table 2.1	Oligonucleotides used in this thesis	51
Table 2.2	Plasmids used in this thesis.....	55
Table 2.3	Vertebrate cell lines used in this thesis	62
Table 2.4	Antibodies used in this thesis.....	63
Table 2.5	TLR TIR Domain sequences for Chapter 3 phylogenetic analysis.	65
Table 2.6	NF- κ B Rel Homology Domain sequences for Chapter 3 phylogenetic analysis.	73
Table 2.7	NF- κ B Rel Homology Domain sequences for Chapter 4 phylogenetic analysis.....	81
Table 2.8	Organisms RHD sequences used for Chapter 5 phylogenetic analysis.	89
Table 3.1	Conserved TLR signaling pathway components in <i>Orbicella faveolata</i>	104
Table 3.2	NF- κ B pathway genes differentially expressed following LPS treatment on <i>O. faveolata</i>	110
Table 4.1	Quantification of NF- κ B positive nuclei in DF-1 cells and sponge tissue	145
Table 4.2	TLR-to-NF- κ B pathway homologs encoded in the <i>A. queenslandica</i> genome.....	146
Table 5.1	Quantification of NF- κ B positive nuclei in DF-1 cells.....	176
Table 5.2	Genes that co-expression with NF- κ B in the three life stages of <i>Capsaspora</i> , and genes that are annotated within that group	177

Table 5.3	<i>Capsaspora</i> gene homologs related to development and immunity	228
Table 5.4	JASPER NF- κ B motif search results sites combined.	231
Table 6.1	Proposed immunity-related effector and regulatory molecules in basal organisms.	285

LIST OF FIGURES

Figure 1.1	The structures of NF- κ Bs in organisms basal to Bilateria are most similar to human NF- κ B proteins rather than Rel proteins.	23
Figure 1.2	Motif analysis of NF- κ B and Rel proteins.	25
Figure 1.3	Phylogenetic analysis places choanoflagellate NF- κ Bs as an outgroup of vertebrate and fly NF- κ Bs.	27
Figure 1.4	Alignment of putative caspase cleavage sites.	28
Figure 3.1	Homologs of proteins in the mammalian Toll-Like Receptor-to-NF- κ B pathway are present in <i>Orbicella faveolata</i>	112
Figure 3.2	Human codon optimization for Of-TLR.	117
Figure 3.3	Similarity of Of-TLR to human TLR4.	118
Figure 3.4	Human codon optimization of Of-NF- κ B.	123
Figure 3.5	Of-NF- κ B is most similar to other NF- κ B proteins and not to Rel proteins.	124
Figure 3.6	Characterization of the activity of the <i>O. faveolata</i> NF- κ B protein in cell-based assays.	127
Figure 3.7	IKK can phosphorylate conserved serines in Of-NF- κ B in vitro and induce processing of Of-NF- κ B in human cells.	129
Figure 3.8	Human codon optimization for Of-IKK.	132
Figure 4.1	The <i>A. queenslandica</i> NF- κ B protein resembles other metazoan NF- κ B proteins both structurally and phylogenetically.	148

Figure 4.2	Aq-NF- κ B contains sequences that can bind to DNA and activate transcription.	149
Figure 4.3	C-terminal sequences of Aq-NF- κ B inhibit nuclear translocation and can undergo IKK-dependent degradation and phosphorylation.	151
Figure 4.4	IKKs can induce processing of both WT Aq-NF- κ B and Aq-NF- κ B-Aiptasia serine mutant.	153
Figure 4.5	Demosponge tissue extract contains an NF- κ B-like protein that has DNA-binding activity and localizes to the nucleus.	154
Figure 4.6	The black encrusting sponge most closely resembles a demosponge of the genus <i>Cliona</i>	156
Figure 4.7	Treatment of sponge tissue with LPS results in increased NF- κ B protein processing and activity.	158
Figure 4.8	NF- κ B processing is induced upon LPS stimulation in black encrusting sponge tissue.	159
Figure 4.9	NF- κ B pathway proteins and upstream receptors identified in <i>A. queenslandica</i>	160
Figure 5.1	Protist NF- κ B proteins differ in domain structure and choanoflagellates show evidence of gene duplication.	243
Figure 5.2	DNA-binding and transcriptional activation activity of Co-NF- κ B.	246
Figure 5.3	IKK-mediated processing of NF- κ B arose with multicellularity.	247
Figure 5.4	MG132 inhibits constitutive processing of Co-NF- κ B.	248

Figure 5.5	Transfection of FLAG-Co-NF- κ B and FLAG-Co-RHD into <i>Capsaspora</i> cells results in nuclear localization.....	249
Figure 5.6	Full Co-NF- κ B localizes to the nucleus of <i>Capsaspora</i> cells.....	250
Figure 5.7	NF- κ B is differentially expressed during the different life stages of <i>Capsaspora</i> and its proposed roles in development and immunity.....	251
Figure 5.8	GO analysis of annotated <i>Capsaspora</i> genes co-expressing with Co-NF- κ B.....	253
Figure 5.9	Characterization of cellular and molecular properties of three choanoflagellate NF- κ Bs.	254
Figure 5.10	Schematic of the proteins found in the simplified TLR pathway of various organisms.	256
Figure 6.1	The Evolution of NF- κ B.	286

LIST OF ABBREVIATIONS

°C	degrees Celsius
aa	amino acid(s)
amp	ampicillin
ANK	ankyrin
APS	ammonium persulfate
Aq	<i>Amphimedon queenslandica</i>
Aq-NF-κB	NF-κB of <i>Amphimedon queenslandica</i>
As	<i>Acanthoecca spectabilis</i>
As-NF-κB	NF-κB of <i>Acanthoecca spectabilis</i>
ASW	artificial seawater
ATP	adenosine triphosphate
BB	blocking buffer
β-gal	β-galactosidase
BLAST	basic local alignment search tool
bp	base pair(s)
BSA	bovine serum albumin
C	carboxy
cam	chloramphenicol
cDNA	complimentary DNA
CFU	colony forming unit
cGAS	cyclic GMP-AMP synthase

cm	centimeter
Co	<i>Capsaspora owczarzaki</i>
Co-NF-κB	NF-κB of <i>Capsaspora owczarzaki</i>
Cterm	Carboxy-terminus
Δ	deletion
DAPI	4',6-diamidino-2-phenylindole
DD	Death Domain
DI	de-ionized
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleotide triphosphate
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediamine tetraacetic acid disodium salt
EGTA	ethylene glycol tetraacetic acid
EMSA	electromobility shift assay
FBS	fetal bovine serum
g	gram(s)
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GO	gene ontology
GRR	glycine-rich region

GST	glutathione-S-transferase
h	hour(s)
HEK	human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hpf	hours post fertilization
HRP	horseradish peroxidase
Hu	human
ID	identifier
I κ B	inhibitor of NF- κ B
IKK	I κ B kinase
IL-1R	interleukin-1 receptor
IP	immunoprecipitation
IPTG	isopropyl β -d-1-thiogalactopyranoside
IRF	interferon regulatory factor
KD	knockdown
kDa	kilodalton(s)
LB	Luria broth
LPS	lipopolysaccharide
LRR	leucine-rich repeat
Luc	luciferase
M	molar
mA	milliamp(s)

mg	milligram(s)
ml	milliliter(s)
MAL	MyD88 adapter-like protein
MAMP	microbial-associated molecular pattern
MAPK	mitogen-activated protein kinase
MAV	mitochondrial antiviral signaling protein
MHC	major histocompatibility complex
min	minute(s)
ml	milliliter(s)
mM	millimolar
MM	microbasic p-mastigophore
mRNA	messenger RNA
MYA	million years ago
MYC	myelocytomatosis
MyD88	myeloid differentiation primary response protein 88
μg	micrograms(s)
μl	microliter(s)
μm	micrometer
μM	micromolar
N	amino
NF-κB	nuclear factor-κB
ng	nanogram(s)

nm	nanometer(s)
Nv	<i>Nematostella vectensis</i>
Nv-NF- κ B	NF- κ B of <i>Nematostella vectensis</i>
Nv-TLR	TLR of <i>Nematostella vectensis</i>
OD	optical density
Of	<i>Orbicella faveolata</i>
Of-NF- κ B	NF- κ B of <i>Orbicella faveolata</i>
Of-TLR	TLR of <i>Orbicella faveolata</i>
ONPG	O-nitrophenyl- β -D-galactopyranoside
PAMP	pathogen-associated molecular pattern
PBM	protein binding microarray
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEI	polyethylenimine
phospho	phosphorylated
PIK3R	phosphoinositide 3-kinases
PMSF	phenylmethylsulfonyl fluoride
poly (I:C)	polyinosinic:polycytidylic acid
poly-dI/dC	poly(deoxyinosinic-deoxycytidylic) acid
ppt	parts per thousand
PTx	phosphate-buffered saline with Triton X-100
puro	puromycin

RHD	Rel Homology Domain
RIPK	Receptor-interacting serine/threonine-protein kinase
RNA	ribonucleic acid
RNase	ribonuclease
RPKM	reads per kilobase of exon per million mapped reads
rpm	revolutions per minute
RSV	Rous sarcoma virus
RT	reverse transcription
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
sec	second(s)
SEM	standard error mean
seq	sequencing
<i>sp.</i>	species
SS	supershift
STING	stimulator of interferon genes
TAK	transforming growth factor- β activated kinase
TANK	TRAF family member associated NF- κ B activator
TBK	TANK-binding kinase
TBS	Tris-buffered saline
TBS-Tween	TBS with Tween20

TE	Tris buffer with EDTA
TEMED	tetramethylethylenediamine
TGF- β	transforming growth factor beta
TIR	Toll/interleukin-1 receptor domain
TIRAP	TIR domain-containing adapter protein
TLR	Toll-like receptor
TM	transmembrane domain
TNF	tumor necrosis factor
TNFR	TNF receptor
Tollip	Toll-interacting protein
TRAF	TNFR-associated factor
TRAM	TRIF-related adapter molecule
TRIF	TIR-domain-containing adapter-inducing interferon β
Tris	1-[bis(2,3-dibromopropoxy)phosphinoyloxy]- 2,3-dibromo-propane
Tween-20	polyoxyethelene sorbitan monolaurate
U	units
v	volume
V	volts
w	weight
WB	Western blot

WT	wild-type
X	times

CHAPTER ONE

INTRODUCTION¹

For nearly a century, the field of immunology has been primarily concerned with the study of human wellness and disease. These studies have revealed the concerted complexity of the body's daily biological warfare, which relies on the proper function of signal transduction pathways to turn on genes important for defense and preservation. Until about a decade ago, it was unknown what types of genetic complexity early-branching organisms possessed for combatting disease. However, increased genomic and transcriptomic sequencing has revealed that these basal organisms contain ancient toolkits for many conserved signaling molecules and pathways, including those involved in immunity. This thesis will focus on one of those ancient pathway homologs, and how it has become apparent to the growing field of comparative immunology that a key immunity-related gene regulatory protein (transcription factor NF- κ B) likely plays biologically conserved and unique roles in basal organisms.

1.1 Introduction to transcription factor NF- κ B

The Nuclear Factor- κ B (NF- κ B) superfamily comprises a group of related transcription factors that have been intensively studied for their involvement in development and immunity since their near simultaneous discovery in a retrovirus, flies, and mice almost 35 years ago (Gilmore and Temin, 1986; Steward, 1987; Ghosh et al., 1990; Gilmore, 1990; Kieran et al, 1990). Indeed, there are now approximately 100,000

¹ Adapted from Williams and Gilmore, 2020

papers on NF- κ B, the vast majority focusing on mammals, flies, and viruses. However, with the proliferation of genomic and transcriptomic sequencing over the past decade, it has been discovered that many organisms ostensibly less complex than insects have NF- κ B-like genes (Putnam et al., 2007; Srivastaba et al., 2010; Shinzato et al., 2011; Gilmore and Wolenski, 2012; Suga et al., 2013; Baumgarten et al., 2015; Richter et al., 2018; Gold et al., 2019).

All NF- κ B proteins are related by an N-terminal DNA-binding and dimerization region called the Rel Homology Domain (RHD) containing a nuclear localization sequence (NLS), and the RHD allows them to enter the nucleus, bind to specific DNA sites (“ κ B sites”), and activate or repress transcription of target genes for specified biological outcomes (Gilmore, 2006). In vertebrates and flies, the NF- κ B superfamily can be divided into two subfamilies: the NF- κ B proteins that consist of vertebrate p100 and p105 and *Drosophila* Relish; and the Rel proteins, which include vertebrate RelA, RelB, and c-Rel, as well as *Drosophila* Dif and Dorsal (Gilmore, 2006). Thus, flies and vertebrates all contain multiple NF- κ B proteins that, for the most part, show complete combinatorial diversity by forming homodimers and heterodimers, which have distinct DNA target site specificities. The two subfamilies can be distinguished phylogenetically by sequence alignment of their RHDs, and by sequences C-terminal to the RHD. That is, NF- κ B proteins contain C-terminal inhibitory Ankyrin (ANK) repeat domains and Rel proteins contain C-terminal transactivation domains (Figure 1.1). The ANK repeats, either within the NF- κ B proteins themselves or in a separate family of NF- κ B inhibitors (I κ Bs), regulate the subcellular localization of NF- κ Bs by binding to the RHD and

sequestering them in the cytoplasm. Activation of the pathway by an appropriate upstream signal results in the degradation of the ANK repeat inhibitor, thus allowing the NF- κ B dimer to enter the nucleus and bind DNA (Gilmore, 2006; Hayden and Ghosh, 2008). NF- κ B p100 and p105 proteins also contain a C-terminal death domain (DD) that is important for protein-protein interactions with other members of the DD superfamily, which serve as adaptors in signaling pathways and/or to recruit other proteins into signaling complexes (Hayden and Ghosh, 2008).

Many evolutionarily conserved receptors can elicit downstream signals to activate NF- κ B translocation to the nucleus. These receptors, which include Toll-like receptors (TLRs) and interleukin-1 receptors (IL-1Rs), interact with cytoplasmic adaptor proteins (e.g., MYD88 and MAL) to initiate downstream signaling (Hayden and Ghosh, 2008). Once the pathway is activated, a series of phosphorylation events leads to the degradation of the ANK repeat inhibitor sequences, freeing NF- κ B to translocate from the cytoplasm to the nucleus. In vertebrates, activation of NF- κ B generally happens in one of two ways. In the canonical pathway, activation of an I κ B kinase β (IKK β) complex results in phosphorylation of an independent I κ B protein, and phosphorylated I κ B then undergoes ubiquitination and proteasomal degradation to liberate the NF- κ B dimer. In the non-canonical pathway, activation of the related kinase IKK α leads to phosphorylation of a serine cluster located C-terminal to the ANK repeats on the p100 protein (Sun, 2011). These phosphorylations result in proteasomal processing of the inhibitory ANK repeats on p100 until the proteasome reaches a glycine-rich region (GRR). At the GRR the proteasome falls off, generating p52, a C-terminally truncated version of p100 that can

enter the nucleus (Sun, 2011). Similar pathways exist in flies, except the p100 homolog Relish has no GRR and the truncated, active form of Relish is generated by a site-specific proteolytic cleavage that removes the ANK repeat region (Stöven et al., 2003).

Genomic and transcriptomic sequencing data strongly suggest that NF- κ B was also pervasive through early evolution, based on the presence of NF- κ B-like homologs in many extant organisms basal to flies and vertebrates. NF- κ B homologs have been identified in single-celled pre-metazoans such as the protists *Capsaspora owczarzaki* (Suga et al., 2013), which acts as a symbiont in the hemolymph of the fresh-water snail *Biomphalaria glabrata*, and choanoflagellates (Richter et al., 2018), which are aquatic unicellular colonial organisms that have been proposed to be the closest living relatives to multicellular animals. NF- κ Bs have also been identified in a variety of multicellular basal marine metazoans, including poriferans (sponges) and cnidarians (corals, sea anemones, hydras, and jellyfish) (Putnam et al., 2007; Srivastava et al., 2010; Shinzato et al., 2011; Baumgarten et al., 2015; Gold et al., 2019). In most of these basal organisms, there are single NF- κ B proteins, which, as discussed in this chapter, most closely resemble NF- κ B subfamily proteins such as Relish and p100. Curiously, no NF- κ B homologs have been found in nematodes (e.g., *Caenorhabditis elegans* and *Caenorhabditis briggsae*) or ctenophores, where the pathway appears to have been lost.

Only recently have several studies investigated how transcription factor NF- κ B and its upstream and downstream pathways function in early-branching organisms. This chapter aims to provide insight into how this highly conserved transcription factor and its regulatory pathway arose, describe domains that have appeared at pivotal points of

evolution and biological processes likely controlled by NF- κ B, and discuss areas of basal NF- κ B knowledge that remain unanswered.

1.2 Structure of basal NF- κ B proteins

Many basal eukaryotes have NF- κ B proteins that are structurally and phylogenetically most similar to vertebrate NF- κ B proteins rather than Rel proteins. In most cases, these organisms have single p100-like RHD-ANK repeat NF- κ B proteins that often possess C-terminal serine residues known in vertebrates to be required for processing of p100 via the non-canonical NF- κ B pathway (see Chapter 1.2, “Activity and regulation of basal NF- κ Bs”). For example, the protist *Capsaspora*, the sponge *Amphimedon queenslandica*, and several cnidarians all have NF- κ B proteins with an RHD and C-terminal ANK repeats encoded on the same transcript (Gauthier and Degnan, 2008; Stefanik et al., 2014; Mansfield et al., 2017; Voolstra et al., 2017; Cunning et al., 2018; Shinzato et al., 2018; Williams et al., 2018). Many ANK repeat-containing proteins are found throughout the genomes of archaea, bacteria, and all plants and animals (Al-Khodori et al., 2010; Jernigan and Bordenstein, 2014). Accordingly, it is likely that during the evolution of eukaryotes there arose a primitive RHD-only protein, which developed the ability to interact with a pre-existing ANK repeat-containing protein. At some point, these interacting RHD and ANK-repeat genes fused to create the more modern single RHD-ANK-repeat protein. Nevertheless, some basal organisms have NF- κ B-like proteins without C-terminal ANK repeat domains. Among protists, several choanoflagellates have NF- κ B-like proteins that do not have C-terminal ANK repeat

domains (Richter et al., 2018).

Choanoflagellates comprise a sister group to all metazoans, and are part of a diverse group of single-celled, colony-forming organisms that live in waters around the world. Although the genomes of the classically studied choanoflagellates *Monosiga brevicollis* and *Salpingoeca rosetta* do not contain NF- κ B homologs (Hoffmeyer and Burkhardt, 2016), a large-scale transcriptomic analysis of 19 choanoflagellates (Richter et al., 2018) showed that 12 had transcripts that encode RHD-containing NF- κ B-like sequences (Richter et al., 2018). However, none of these choanoflagellate NF- κ B-like proteins appears to have C-terminal ANK repeat domains (Figure 1.1). Moreover, several choanoflagellate NF- κ B-like proteins also have extended N-terminal domains that are not present in any other NF- κ B proteins (Figure 1.1, Table 1.1). Overall, these 19 choanoflagellate transcriptomes encode zero, one, two, or three NF- κ B-like proteins, none of which has C-terminal ANK repeats. Indeed, it is not yet known whether choanoflagellate NF- κ Bs are regulated by an ANK repeat-containing I κ B-like inhibitor. The lack of ANK repeats in choanoflagellate NF- κ B proteins and the wide diversification of NF- κ Bs among choanoflagellates are likely a result of 600 million years of independent evolution among the diverse species of choanoflagellates. Furthermore, this gene expansion mirrors the increasing number and complexity of NF- κ Bs seen throughout metazoans. In contrast to what is seen in choanoflagellates, the protist *Capsaspora* does contain a bipartite RHD-ANK protein (Suga et al., 2013). Thus, the earliest NF- κ B with an RHD-ANK fusion likely arose in the protist lineage that led to *Capsaspora*, but was excluded or lost in the lineage that led to choanoflagellates.

Although the choanoflagellate and *Capsaspora* NF- κ Bs differ from one another in their overall structures (i.e., especially the absence of a C-terminal ANK-repeat region in choanoflagellates), they share two motifs within their RHDs (Figure 1.2, Motifs 7 and 10 in teal and yellow) that are not present in any metazoan NF- κ B protein. Among choanoflagellate NF- κ Bs, there is a great deal of sequence diversity, which is consistent with the overall genetic differences among choanoflagellates wherein the average phylogenetic distance between any two choanoflagellate species is greater than the phylogenetic distance between sponges and mammals (Richter et al., 2018). Nevertheless, a Bayesian tree analysis of the choanoflagellate RHDs identified to date clusters them as an outgroup to several fly and vertebrate NF- κ Bs and Rels, which demonstrates a clear divergence of the choanoflagellate NF- κ Bs from the metazoan NF- κ B superfamily (Figure 1.3A). In general, the NF- κ B proteins within a single choanoflagellate species that has multiple NF- κ B proteins (e.g., *Salpingoeca helianthica*) are highly related (Figure 1.3A), suggesting that they arose by gene duplication events. However, it is important to note that some RHD subdomains are not present in the single-celled pre-metazoan lineages. For example, *Capsaspora* and choanoflagellate NF- κ B proteins have low homology in the dimerization sequence, and choanoflagellates do not have an NF- κ B-like NLS. On the other hand, all metazoan and pre-metazoan NF- κ B sequences contain a highly conserved sequence that is important for DNA binding (Figure 1.3B).

Among poriferans, the fully sequenced genome of the demosponge *Amphimedon queenslandica* encodes a single full-length NF- κ B protein that contains clear features of

the human p100 NF- κ B protein: that is, it has an RHD, NLS, GRR, six ANK repeats, and a DD (Gauthier and Degnan, 2008) (Figure 1.1). However, between the GRR and the ANK repeats, *A. queenslandica* NF- κ B contains a region that has no known function and no homology to any other protein (Gauthier and Degnan, 2008) (Figure 1.1, yellow). More limited sequencing data indicate that other sponges also contain NF- κ B-like transcripts, which, to date, only contain RHD sequences (Riesgo et al., 2014). Many of these sponge NF- κ B-like proteins appear to be much shorter than the prototypical NF- κ B proteins, but this may simply be due to lack of sequence coverage; for example, *Corticium candelabrum* is reported to have a transcript for NF- κ B that consists of only 86 amino acids that are similar to the beginning of the RHD (Riesgo et al., 2014).

All characterized cnidarians appear to contain single NF- κ B proteins, but the overall structures of these NF- κ B proteins have diverged, in some cases, likely by gene-splitting events. Thus, many cnidarians (the anemones *Edwardsiella lineata* (Stefanik et al., 2014) and *Exaiptasia pallida* (Mansfield et al., 2017) and the corals *Acropora digitifera*, *Stylophora pistillata*, *Orbicella faveolata*, and *Pocillopora damicornis* (DeSalvo et al., 2010; Voolstra et al., 2017; Cuning et al., 2018) have prototypical RHD-ANK repeat bipartite proteins, whereas others have separate RHD and ANK repeat proteins. In the best studied example of the latter, the sea anemone *Nematostella vectensis* has a single 440 amino acid NF- κ B-like protein that is most similar to the processed vertebrate p50/p52 NF- κ B proteins based on phylogenetic analysis of RHD amino acid sequences, intron-exon structure of the RHD, and the presence of a GRR after the RHD (Sullivan et al., 2007). Moreover, a separate *N. vectensis* gene encodes

an I κ B-like ANK-repeat protein with substantial homology to C-terminal sequences of the ANK-repeat domains of vertebrate p100 and p105 (Sullivan et al., 2007). Four lines of evidence support the hypothesis that separate *N. vectensis* NF- κ B and I κ B genes arose due to a gene splitting event: 1) the existence of RHD-ANK repeat bipartite proteins in more basal organisms (e.g., *Capsaspora*, *A. queenslandica*); 2) the presence of remnant GRR sequences at the C terminus of *N. vectensis* NF- κ B; 3) the homology of the separate *N. vectensis* I κ B protein to mammalian NF- κ B protein C-terminal ANK repeats; and 4) the presence of an intact RHD-ANK protein in the closely related anemone *E. lineata* (Stefanik et al., 2014). Similar to *N. vectensis*, the hydras *Hydractinia symbiolongicarpus* (Zárate-Potes et al., 2019) and *Hydra magnipapillata* (Fanzenburg et al., 2012) and the jellyfish *Aurelia* (Gold et al., 2019) have separate RHD-only NF- κ B proteins and I κ B-like genes, all of which likely came about due to gene-splitting events, which are common in cnidarians (Chapman et al., 2010; Gacesa et al., 2015).

1.3 Activity and regulation of basal NF- κ Bs

The activity and regulation of NF- κ B proteins in vertebrates are now known in great detail. Many of these properties are conserved in basal NF- κ Bs, however, there are clear exceptions, and many things about NF- κ B proteins from organisms basal to cnidarians are unknown.

Although not formally shown, either biochemically or structurally, it is likely that all basal NF- κ B-like proteins bind as dimers to DNA sites that are quite similar to those

bound by vertebrate NF- κ B proteins p50 and p52. Nevertheless, the absence of a clearly defined dimerization sequence in the pre-metazoan NF- κ Bs does raise the possibility that they bind DNA as monomers, similar to what has been found for the NFAT proteins (Stroud and Chen, 2003). However, consistent with the basal NF- κ Bs binding DNA as dimers, the NF- κ B protein/ κ B site DNA complexes seen in electrophoretic mobility shift assays using two sea anemone NF- κ B proteins migrate in a manner suggesting that the NF- κ B proteins are dimers (Wolesnki et al., 2011; Mansfield et al., 2017). Furthermore, based on protein-binding microarrays (PBMs), the NF- κ B proteins of *Capsaspora*, the sponge *A. queenslandica*, and two sea anemones bind to a set of κ B sites that are more similar to the sites bound by mammalian NF- κ B proteins than by Rel proteins (Ryzhakov et al., 2013; Mansfield et al., 2017), a biochemical finding that is consistent with their overall structural organization and phylogenetic data.

Most basal NF- κ B proteins appear to be activators of κ B site-containing promoters/enhancers. That is, the RHD sequences of sea anemone cnidarian NF- κ B proteins function as activators of transcription when expressed in reporter gene assays in yeast and human cells (Wolenski et al., 2011; Mansfield et al., 2017). The ability of these basal NF- κ Bs to function as independent activators of transcription is more similar to Relish, which can act as a homodimeric transcriptional activator (Stöven et al., 2003), than to mammalian p50/p52 proteins, which are generally activators of transcription only when in heterodimers with a Rel protein (Gilmore, 2006; Hayden and Ghosh, 2008). No DNA-binding or transcriptional regulatory studies have been performed with NF- κ B proteins from corals, sponges, or any protists (such as *Capsaspora* or choanoflagellates),

prior to the research presented in this thesis.

The most extensive research on basal NF- κ B activity has been done with the NF- κ B protein of the sea anemone *N. vectensis* (Nv-NF- κ B). Curiously, it was found that wild populations of these anemones have two major alleles of Nv-NF- κ B that differ at 10 residues, six of which are in the RHD (Sullivan et al., 2009). Two of these variable residues are ones that are predicted to contact DNA, and, as a consequence, one of the allelic Nv-NF- κ B proteins (Nv-NF- κ B-C) binds DNA with an approximately two-fold higher affinity than the other variant (Nv-NF- κ B-S) (Wolenski et al., 2011). Despite its reduced DNA-binding activity, Nv-NF- κ B-S activates transcription in reporter gene assays in human cells more effectively than Nv-NF- κ B-C (Wolenski et al., 2011). Overall, the PBM-based binding site profiles of both Nv-NF- κ B alleles still largely resemble that of mammalian p50 (Ryzhakov et al., 2013; Mansfield et al., 2017). Interestingly, the DNA-binding site profiles of Nv-NF- κ B-C and Nv-NF- κ B-S are as different from each other as each is from human p50 (Mansfield et al., 2017), suggesting that within a given species (i.e., *N. vectensis*) there can be considerable flexibility in DNA binding site recognition that does not have an obvious effect on the organism's overall developmental phenotype. However, these results do not exclude the possibility that anemones harboring distinct NF- κ B alleles have differences in other properties, such as immunity.

In the mammalian non-canonical pathway, human p100 is phosphorylated by IKK α at a cluster of three serine residues that are located C-terminal to the ANK repeats, and this phosphorylation event induces proteasomal processing of the C-terminal

sequences up to the GRR (Sun, 2011). This cluster of serine residues is conserved in many sequenced cnidarian NF- κ Bs, including those of the corals *Pocillopora damicornis*, *Stylophora pistillata*, and *Acropora millipora*, and the sea anemones *Actinia tenebrosa*, *Aulactinia veratra*, and *Aiptasia* (Gauthier and Degnan, 2008; Baumgarten et al., 2015; Anderson et al., 2016; Mansfield et al., 2017; Voolstra et al., 2017; Cunning et al., 2018; Ying et al., 2019). Furthermore, phosphorylation and proteasomal processing of NF- κ B from a sea anemone (*Aiptasia*) can be induced by co-expression of human IKK α and IKK β or the single *Aiptasia* IKK (Mansfield et al., 2017), suggesting that IKK-dependent processing of cnidarian NF- κ Bs can occur in their natural settings. On the other hand, the *Capsaspora* NF- κ B protein clearly has C-terminal ANK repeats and a GRR, but it does not have any apparent C-terminal IKK target site serine residues and no IKK can be found in the *Capsaspora* genome. Intriguingly, *Capsaspora* does have at least one possible caspase cleavage site similar to what is seen in the *Drosophila* protein Relish (Figure 1.1 and Figure 1.4). Overall, the emergence of IKK proteins appears generally coincident with the presence of regulatory serine residues downstream of the ANK repeat domains of basal NF- κ Bs. Of note, the *Drosophila* Relish protein also lacks C-terminal serine residues and a GRR and is processed by a site-specific protease cleavage event (Stöven et al., 2003), and in some conditions, shortened human p105 NF- κ B proteins can be generated by ribosomal stalling (Lin et al., 1998).

For the *N. vectensis* NF- κ B protein, which lacks C-terminal ANK repeats, it was shown that the independent Nv-I κ B could directly bind to Nv-NF- κ B and that this Nv-I κ B could be phosphorylated by an Nv-IKK-related kinase (Wolenski et al., 2011). Thus,

aspects of canonical signaling may also be present in cnidarians.

Proteasomal-mediated processing has not been directly demonstrated to induce nuclear translocation of any basal NF- κ B in its natural setting. Furthermore, while the ubiquitin-proteasome system is present in basal phyla, regulated processing has not been experimentally demonstrated for any protein (Fort et al, 2015). However, in a reconstituted mammalian cell system, IKK-induced processing of the anemone *Aiptasia* NF- κ B can be blocked by a proteasome inhibitor (Mansfield et al., 2017). In addition, when expressed in chicken tissue culture cells, the full-length RHD-ANK protein of the anemone *Aiptasia* is located in the cytoplasm, and deletion of the ANK repeat domains of this NF- κ B causes it to localize to the nucleus (Mansfield et al., 2017), suggesting that C-terminal processing would do the same in the host organism. Similarly, in these same assays, the naturally truncated anemone Nv-NF- κ B protein localizes to the nucleus, but it is sequestered in the cytoplasm when Nv-I κ B is co-expressed (Wolenski et al., 2011). Prior to the research in this thesis, nothing was known about the subcellular localization or regulation of sponge or protist NF- κ B proteins. Indeed, choanoflagellate NF- κ Bs lack a clear NLS as well as C-terminal ANK repeats.

Notwithstanding the above-described reconstitution experiments conducted with cnidarian NF- κ Bs in vertebrate cell systems, it is not known whether or how regulated processing of basal NF- κ Bs occurs in any natural setting. Indeed, in the anemone *Aiptasia* (Mansfield et al., 2017), most NF- κ B protein is largely processed and nuclear in the absence of any known stimulus. Furthermore, certain treatments (e.g., bleaching in *Aiptasia*) may induce NF- κ B activation by inducing transcriptional upregulation of NF-

κ B pathway genes rather than C-terminal processing (Mansfield et al., 2017). Prior to this thesis, the sea anemone *Aiptasia* was the only basal organism where this phenomenon had been seen, but it was suggested that this may be the case in other organisms (Mansfield et al., 2017; Mansfield and Gilmore, 2019).

1.4 Basal NF- κ Bs likely have roles in development and immunity

One of the most important questions regarding NF- κ B in basal metazoans is what biological processes are regulated by this transcription factor. At least in cnidarians and sponges, there is evidence that NF- κ B has roles in both early development and/or adult immunity (Table 1.1).

1.4.1 Role of NF- κ B in sponge and cnidarian development

In the sponge *A. queenslandica* and several cnidarians, NF- κ B transcripts have been shown to be expressed in the early embryo. In *A. queenslandica*, fluorescent *in situ* staining has shown that NF- κ B transcripts are broadly expressed throughout the embryo after cleavage, and are particularly strong in granular cells, which eventually become the outer layer of the developing embryo (Gauthier and Degnan, 2008). Similarly, some cnidarians express NF- κ B transcripts in the early embryo and juvenile larvae (Helm et al., 2013; Tulin et al., 2013; Wolenski et al., 2013; Fischer et al., 2014; Siboni et al., 2014; Warner et al., 2018). In the anemone *N. vectensis*, NF- κ B mRNA expression is seen as early as one hour post fertilization (hpf) (Helm et al., 2013; Tulin et al., 2013; Fischer et al., 2014; Warner et al., 2018), and nuclear NF- κ B protein was detected as early as the

late gastrula stage (Wolenski et al., 2013). Morpholino-based knockdown of NF- κ B in the developing *N. vectensis* embryo led to a failure to develop cnidocytes at the juvenile polyp stage (Wolenski et al., 2013). Cnidocytes are a Cnidaria-specific cell type that is involved in a variety of sensing, prey capture, and perhaps defense roles (Babonis and Martindale, 2014). In addition, morpholino knockdown of the single NF- κ B-inducing Toll-like receptor (TLR) transcript also led to an early developmental defect in *N. vectensis* (Brennan et al., 2017).

1.4.2 Role of NF- κ B in juvenile and adult sponges and cnidarians

NF- κ B transcripts and proteins have also been detected in several cell types in differentiated sponge tissues. Whole-mount *in situ* hybridization of juvenile sponge tissue showed that NF- κ B transcripts are present in flask cells, which are large ciliated cells that express a range of genes whose orthologues play roles in eumetazoan neurons (Gauthier and Degnan, 2008). Two recent studies (Musser et al., 2019; Sogabe et al., 2019) performed single cell sequencing on different adult sponge cell types, and they reported that NF- κ B transcripts were primarily expressed in two cell types: choanocytes, which are cells that reside in the chambers that allow the sponge to filter feed, and archaeocytes, which are motile phagocytic cells that are inside the mesohyl (a gelatinous matrix between the internal and external layers of sponges). Treatment of two Mediterranean species of sponges, *Aplysina aerophoba* and *Dysidea avara* with an immunogenic cocktail of lipopolysaccharide (LPS, a potent TLR-to-NF- κ B inducer in mammals) and peptidoglycan led to the upregulation of immune-related receptors involved in signaling to NF- κ B (Pita et al., 2018).

Taken together, the above studies suggest that NF- κ B plays a role in a select number of cells involved in sponge immunity. However, explicit studies on the protein level, for example on the processing or activation of sponge NF- κ B, have not been previously reported.

Several studies have suggested that NF- κ B also has a role in immunity in adult cnidarians. In the anemone *N. vectensis*, cytoplasmic NF- κ B and I κ B are expressed in a subset of cnidocytes in the body column of juvenile and adult anemones (Siboni et al., 2014). NF- κ B protein is also highly expressed in cnidocytes that are present in circulating multicellular bodies called nematosomes, which also express high levels of TLR and c-GAS-STING innate immune signaling components (Babonis and Martindale, 2014; Brennan et al., 2017), which are upstream of NF- κ B in vertebrates. Those results, along with the ability of nematosomes to take up bacteria (Babonis and Martindale, 2014; Brennan et al., 2017), suggest that the nematosome is a primitive circulating immune organ in some anemones.

Aiptasia is a tropical anemone that has been used as a model for cnidarian symbiosis with algal dinoflagellates in the family Symbiodiniaceae (Wolfowicz et al., 2016). Colonization of both larval and adult *Aiptasia* with algal symbionts of certain strains results in down-regulation of NF- κ B transcripts, protein, and DNA-binding activity (Burns et al., 2017; Mansfield et al., 2017; Mansfield et al., 2019). Conversely, induction of loss-of-symbiosis with either chemical or heat treatment results in increased levels of NF- κ B expression and activity (Mansfield et al., 2017), and *Aiptasia* lacking symbionts are more resistant to bacterial infection than ones with algal symbionts

(Mansfield et al., 2019). These results have led to the hypothesis that suppression of NF- κ B-directed immunity is required for the establishment of algal symbiosis in some cnidarians, similar to what has been found in amphibians (Burns et al., 2017). It is noteworthy that the symbiont-modulated changes in NF- κ B activity in *Aiptasia* occurred primarily at the level of protein and mRNA expression (Mansfield et al., 2017), and not by post-translational processing as generally seen in vertebrate systems. That is, both symbiotic and aposymbiotic *Aiptasia* have nuclear NF- κ B-staining cells, but the number of NF- κ B-positive cells is substantially increased in *Aiptasia* in which symbiont loss has occurred (Mansfield et al., 2017). In addition, most NF- κ B protein is in its processed form in both symbiotic and aposymbiotic *Aiptasia* (Mansfield et al., 2017).

The research on NF- κ B in *Aiptasia* is of high interest because corals host the same family of algal symbionts and because ocean warming-induced loss of symbiosis (“bleaching”), as well as microbial pathogen infections, are causing large-scale loss of coral reef health (Wolfowicz et al., 2016). Several studies have also investigated NF- κ B and NF- κ B pathway gene expression in corals, and such studies further suggest a role for NF- κ B in immunity and symbiosis/dysbiosis (Table 1.1). For example, transcripts encoding NF- κ B and its signaling components have been identified in transcriptomes from bleached *O. faveolata* (DeSalvo et al., 2010). Furthermore, NF- κ B mRNA was substantially upregulated in the coral *Acropora palmata* following extended exposure to elevated water temperature (Traylor-Knowles et al., 2017), while another study found that NF- κ B transcripts were only transiently induced shortly after heat treatment in the coral *Acropora hyacinthus* (Fuess et al., 2017). Thus, NF- κ B may be affected in different

ways in different cnidarians undergoing heat stress-induced changes. Additionally, one study has shown an increase in transcripts encoding members of the NF- κ B pathway in corals with microbial diseases (Wenger et al., 2014), further suggesting that NF- κ B has an immune-related role in cnidarians.

Several lines of evidence suggest that TLR and TLR-like pathways are upstream activators of NF- κ B in cnidarians. First, homologs of TLR-to-NF- κ B pathway signaling are present in most cnidarians (Sullivan et al., 2007; Chapman et al., 2010; Shinzato et al., 2011; Wolenski et al., 2011; Gilmore and Wolenski, 2012; Baumgarten et al., 2015; Voolstra et al., 2017; Gold et al., 2019). Second, *Hydra* and *N. vectensis* TLR-like proteins have a conserved ability to activate NF- κ B when ectopically expressed in human cells (Franzenburg et al., 2012; Brennan et al., 2017). Third, as mentioned above, TLR and NF- κ B proteins are expressed in many of the same cnidocytes in adult *N. vectensis* (Brennan et al., 2017). Indeed, TLR-to-NF- κ B signaling has been proposed to play a role in embryonic development (Siboni et al., 2014; Brennan et al., 2017), immunity (DeSalvo et al., 2010; Franzenburg et al., 2012; Wenger et al., 2014; Brennan et al., 2017), and regeneration (Woznica et al., 2017) in cnidarians. Nevertheless, TLR is not the sole receptor that can signal to NF- κ B in mammals that is also present in cnidarians. For example, cnidarians also possess homologs to tumor necrosis factor (TNF) receptors and their downstream components (Sullivan et al., 2007), which are extensively characterized activators of NF- κ B in vertebrates. However, the ability of cnidarian TNF receptors to activate NF- κ B in cnidarians has not been investigated. Finally, activation of the cGAS-STING pathway has recently been shown to increase expression and nuclear localization

of NF- κ B in *Nematostella* (Margolis et al., 2021).

The biological role of NF- κ B in single-celled organisms remains unknown (Table 1.1). However, one can speculate that biological processes controlled by NF- κ B in protists are quite different than in multicellular animals, given that protists have no innate or adaptive immunity in the traditional sense of specialized immune cells. Thus, if NF- κ B plays a role in immunity in protists, it may do so in an unexpected manner. Or, in the case of *Capsaspora*, which is a symbiont of the snail *B. glabrata*, NF- κ B may modulate either protist or host immunity to facilitate symbiosis, similar to what has been proposed for *Aiptasia*-Symbiodiniaceae symbiosis. Alternatively, NF- κ B could play a “developmental” role in the solitary vs. colony states seen in many protists, including choanoflagellates. Finally, NF- κ B may be involved in some protist-specific process. For example, one species of choanoflagellate has been shown to be induced to sexually reproduce upon stimulation with bacterial components (Brook et al., 2020).

Currently, little is known about the evolution of immunity, and the most ancient roles of conserved transcription factors involved in invertebrate immunity. Therefore, understanding how NF- κ B operates in the context of basal organisms will likely provide insights into our own evolution as well as allowing researchers to uncover molecular mechanisms for understanding of basal organism disease and early immunological roles of NF- κ B.

1.5 Thesis rationale

This thesis describes the first molecular and biochemical characterization of NF- κ B signaling in four organisms that represent three branches of basal evolution. This research describes the first characterization of a TLR-to-NF- κ B pathway in an endangered coral, and the first characterization of NF- κ B in two groups basal to cnidarians (sponges and protists). Therefore, this work provides a framework for our understanding of the structure, signaling, biology, and evolution of basal NF- κ B proteins.

Studies of NF- κ Bs in more derived organisms, such as flies and humans, have revealed that these NF- κ Bs have conserved biological roles in immunity and development. The primary goal of this research was to explore the evolution of the structure, activity and regulation, and biological roles of NF- κ B proteins from early branching organisms. This thesis describes NF- κ Bs from an endangered coral *Orbicella faveolata* (Of), an evolutionarily significant sponge *Amphimedon queenslandica* (Aq), and the protists *Capsaspora owczarzaki* (Co) and *Acanthoea spectabilis* (As). It is shown that the structure, DNA-binding profiles, and cellular localization of basal NF- κ Bs are most highly conserved to the human NF- κ B p100/p105 proteins. Furthermore, the regulation of NF- κ B processing, at least when reconstituted in human cells, is conserved even in the most basal animals, in that Of-NF- κ B and Aq-NF- κ B can be truncated to their nuclear form with IKK-mediated processing through conserved residues. However, this method of regulation is not seen outside of animals, as demonstrated in the protists characterized here. Finally, evidence is provided into the biological roles in these basal organisms, in that these NF- κ B proteins likely play immunological roles in some basal

organisms, but the level of activation of NF- κ B in basal organisms is likely different than what is seen in higher organisms. Taken together, this research provides insight into the ancient roles of NF- κ B signal transduction and its potential origins in innate immunity.

Table 1.1 Functions of NF- κ B in basal organisms.

Organism	NF-κB structure	Biological Function	Data Available	References
<i>Aiptasia</i>	RHD-ANK	Immunity	protein, transcriptomic, mRNA, DNA binding	Burns et al., 2017; Mansfield et al., 2017; Jacobovitz et al., 2019; Mansfield et al., 2019
<i>N. vectensis</i>	RHD-only	Immunity (?) and Development	protein, transcriptomic, mRNA, DNA-binding	Weis, 2008; Wolenski et al., 2011; Helm et al., 2013; Wolenski et al., 2013
<i>Hydra</i>	RHD-only	Immunity	transcriptomic, protein	Chapman et al., 2010; Franzenburg et al., 2012; Zárate-Potes et al., 2018
Corals	RHD-ANK	Immunity	transcriptomic	DeSalvo et al., 2010; Anderson et al., 2016; Fuess et al., 2017; Traylor-Knowles et al., 2017
Sponge	RHD-ANK	Immunity and Development	protein, transcriptomic, mRNA, DNA-binding	Gauthier and Degnan, 2008; Pita et al., 2018; Musser et al., 2019
<i>Capsaspora</i>	RHD-ANK	?	genomic	Suga et al., 2013
Choanoflagellates	RHD-only	?	transcriptomic	Richter et al., 2018

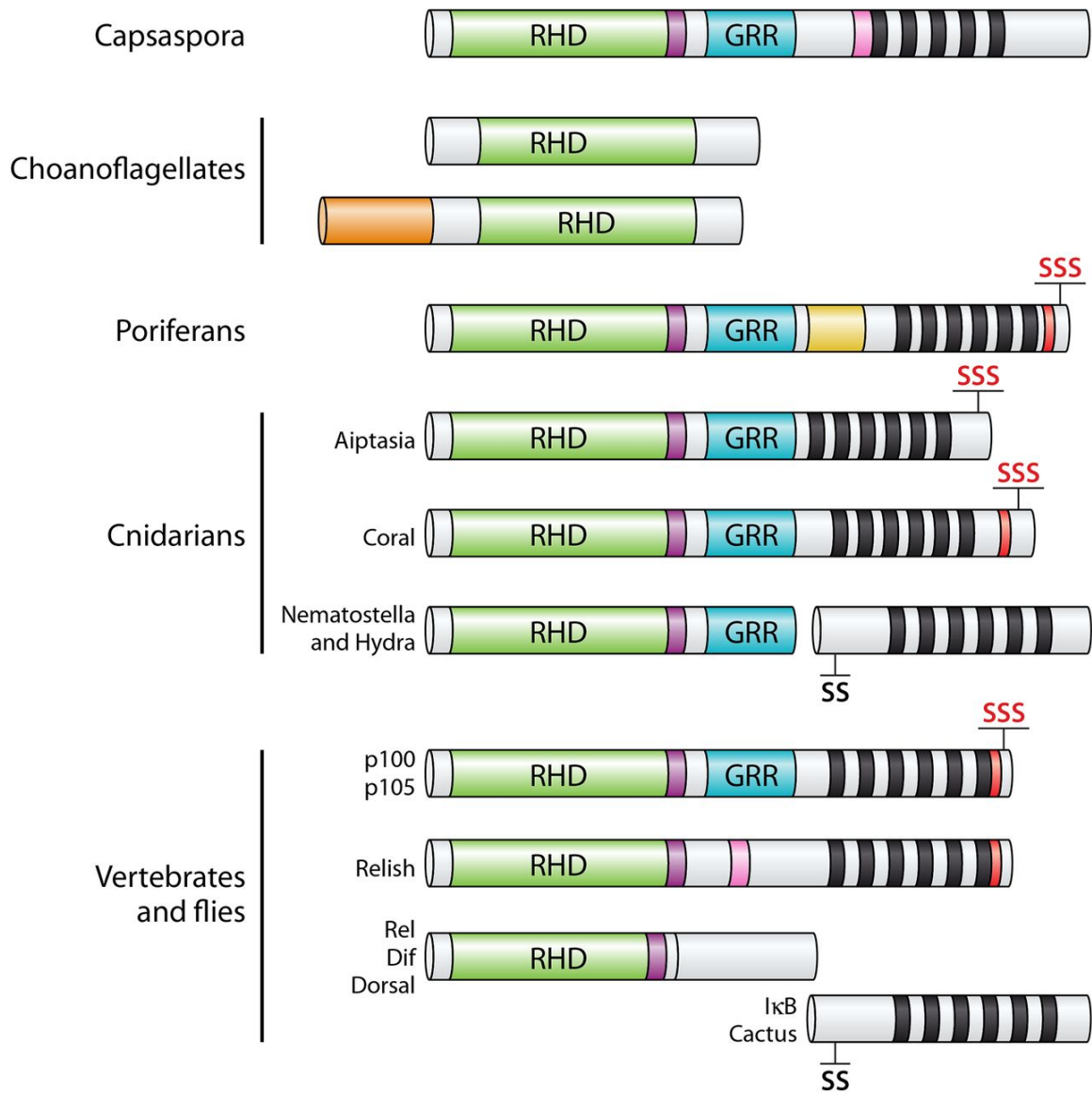


Figure 1.1 The structures of NF-κBs in organisms basal to Bilateria are most similar to human NF-κB proteins rather than Rel proteins.

The basic domain structures of homologs of NF-κB in *Capsaspora*, choanoflagellates, poriferans (sponges), and cnidarians are compared to the fly and vertebrate NF-κBs (Relish, p100/p105), Rels (Dif, Dorsal, RelA, RelB, c-Rel), and IκB (Cactus). Green, Rel Homology Domain (RHD); purple, nuclear localization signal (NLS); blue, Glycine-rich Region (GRR); pink, caspase cleavage site; black bars, Ankyrin repeats; red, Death

Domain; SSS/SS, conserved serines important for phosphorylation and degradation of the protein (red SSS indicate homology to p100); and orange, N-terminal domains of choanoflagellates not seen in other NF- κ B proteins.

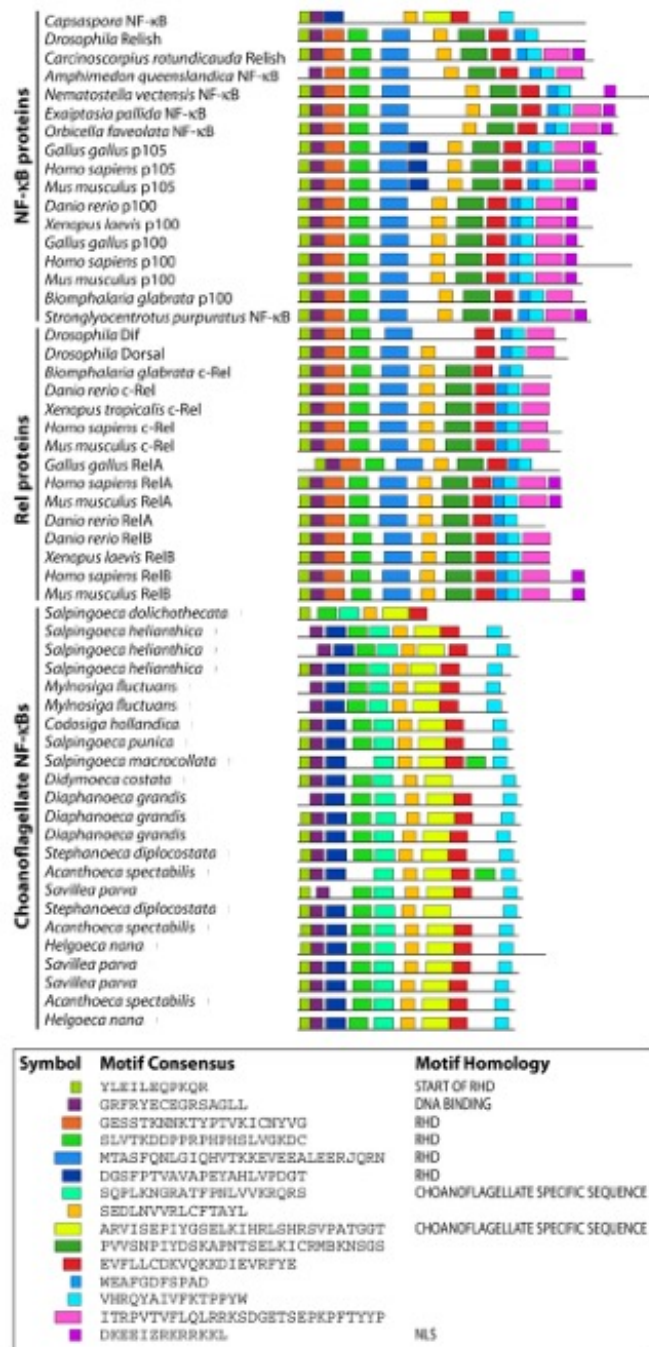


Figure 1.2 Motif analysis of NF-κB and Rel proteins.

MEME analysis was performed on each sequence to identify shared motifs. Conserved motifs are highlighted and two choanoflagellate specific sequences are labeled and shown in lime green and yellow.

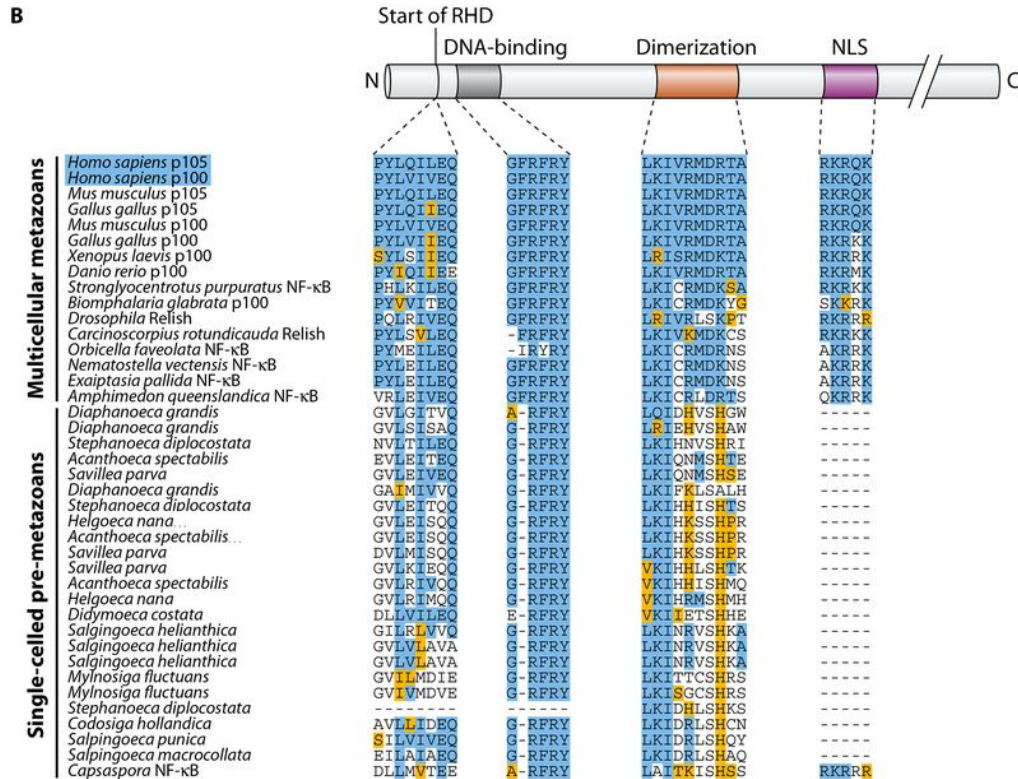
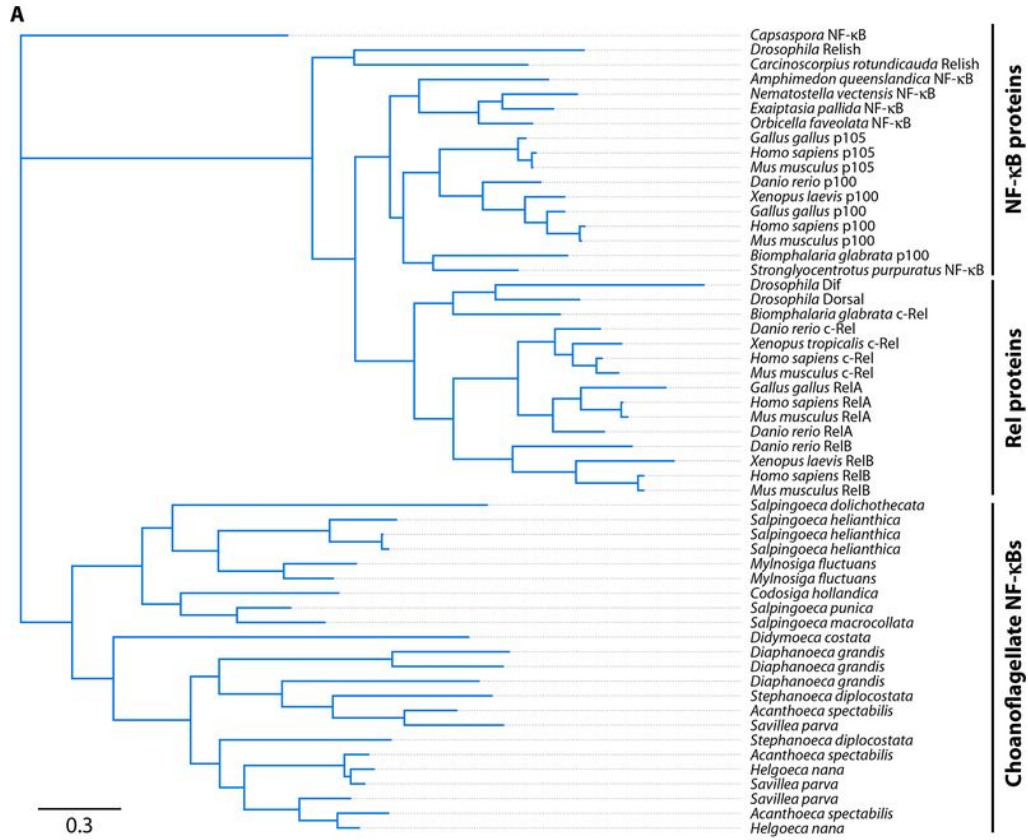


Figure 1.3 Phylogenetic analysis places choanoflagellate NF- κ Bs as an outgroup of vertebrate and fly NF- κ Bs.

A) Bayesian analysis of holozoan RHDs, including the recently identified choanoflagellate NF- κ Bs. MEME analysis was performed on each sequence to identify shared motifs. NF- κ B proteins and Rel proteins cluster separately from each other, and choanoflagellate NF- κ Bs cluster as a single outgroup. B) A multiple sequence alignment of important NF- κ B subdomains across multicellular metazoans and single-celled pre-metazoans. Exact amino acid matches to either p100/p105 are highlighted in blue across phyla, and residues with conserved functional groups are highlighted in yellow. All multicellular metazoans contain a defined start to the Rel Homology Domain (RHD), dimerization sequence, and nuclear localization sequence (NLS). Single-celled pre-metazoans have a less defined RHD start, low homology in the dimerization sequence, and no canonical NLS (with the exception of *Capsaspora*). All metazoans and pre-metazoan NF- κ B sequences contain a highly conserved DNA-binding sequences.

Caspase consensus sequence	D-X-X-D-  A/G/S/T
Capsaspora	aa635-E-N-G-D-S-aa639
Relish	aa541-L-Q-H-D-G-aa546

Figure 1.4 Alignment of putative caspase cleavage sites.

The consensus sequence for caspase cleavage is aligned with the known site of DREDD cleavage in *Drosophila* Relish and with a putative caspase site in *Capsaspora* NF- κ B.

The predicted site of cleavage is marked by a red triangle.

CHAPTER TWO

MATERIALS AND METHODS²

2.1 Phylogenetic analyses of conserved genes

For comparative analysis of Toll-like receptors, the predicted TIR domains of *Orbicella faveolata* (Of) TLR, *Nematostella vectensis* (Nv) TLR, *Drosophila melanogaster* (Dm) Toll, and an *Amphimedon queenslandica* (Aq) TLR-like protein (Gauthier et al., 2010) were analyzed along with the ten human TLR proteins. The TIR domains of Of-TLR, Nv-TLR, and Aq-TLR were identified through MEME analysis (Bailey et al., 2006), and sequences were trimmed to contain only the TIR domains based on motif prediction and known human TIR domains (Table 2.5). Human TLR and Dm-Toll amino acid sequences, along with their annotated TIR domains, were obtained from the UniProt database. Clustal Omega (Sievers et al., 2011) was then used to align the trimmed and culled TIR sequence dataset. The tree was rooted with the Aq-TLR, and phylogenetic comparison was performed using neighbor-joining analysis bootstrapped 1000 times using PAUP* (Swofford, 2001).

For phylogenetic analysis of the NF- κ B and Rel proteins in Chapter 3, the RHD sequences of *O. faveolata* (Of) NF- κ B (from NCBI) *Aiptasia* sp. NF- κ B (NCBI), *Actinia tenebrosa* NF- κ B (NCBI), *N. vectensis* NF- κ B (UniProt), *D. melanogaster* Relish, Dorsal, and Dif (UniProt), and *Homo sapiens* p100, p105, RelA, RelB, and c-Rel

² Adapted from Williams et al., 2018; Williams et al., 2020; Williams et al., 2021; Williams, Aguirre Carrión, and Gilmore, 2021

(UniProt) were obtained. The tree was rooted with an NF- κ B-like protein from *Capsaspora owczarzaki* (NCBI). As described above, conserved motifs from MEME analysis (Bailey et al., 2006) were truncated based on motif predictions (Table 2.6), were aligned by Clustal Omega (Sievers et al., 2011), and were inputted into PAUP* (Swofford, 2001) for maximum likelihood analysis bootstrapped 1000 times.

For phylogenetic analysis of the NF- κ B and Rel proteins in Chapter 4, the RHD sequences of NF- κ B from *A. queenslandica* (Aq) (Gauthier and Degnan, 2008) were compared phylogenetically to *Aiptasia pallida* (Ap) NF- κ B (AIPGENE8848) from the ReefGenomics database, *Orbicella faveolata* NF- κ B (Williams et al., 2018), and *N. vectensis* (Nv), *Drosophila* Dif, Dorsal, and Relish, and *Homo sapiens* and Mouse p100, p105, RelA, RelB, and c-Rel from the UniProt database. *Carcinoscorpius rotundicauda* (Horseshoe crab) Relish, and *Danio rerio*, *Xenopus laevis*, and *Strongylocentrotus purpuratus* NF- κ Bs were from NCBI (Accession numbers: ABC75034.1, NP_001001840.2, NP_001081181.1, and NP_999819.1, respectively). The tree was rooted with RHD of *Capsaspora owczarzaki* NF- κ B (from UniProt database). Conserved motifs from MEME analysis were truncated based on motif predictions (Table 2.7), were aligned by Clustal Omega (Sievers et al., 2011), and were inputted into PAUP* (Swofford, 2001) for maximum likelihood analysis bootstrapped 1000 times.

For phylogenetic analysis of the NF- κ B and Rel proteins in Chapter 5, the RHD sequences of NF- κ B from *C. owczarzaki* were compared phylogenetically to the NF- κ B-like sequences present in the transcriptomes of sequenced choanoflagellates. Details on databases and sequence acquisition can be found in Table 2.8. Sequences were aligned by

Clustal Omega (Sievers et al., 2011). A maximum likelihood phylogenetic tree was created using PAUP* (Swofford, 2001) and was bootstrapped 1000 times. Bayesian phylogenetic analyses showed similar results.

2.2 Codon optimization of cDNA sequences

The cDNA of Of-, Aq-, Co-NF- κ Bs and As-NF- κ B1-3 were optimized using GenScript's services before synthesizing through GenScript for optimal expression in mammalian cells.

2.3 Recombinant DNA techniques

Molecular cloning techniques were used to create bacterial and mammalian plasmids for the expression of full-length and deletion mutants of basal NF- κ Bs and other relevant proteins.

2.3.1 Identification and synthesis of basal NF- κ Bs, IKKs, and Of-TLR

The cDNA for Of-NF- κ B, Aq-NF- κ B, Of-IKK, Of-TLR and Co-NF- κ B were identified by BLAST homology on NCBI or through Pinzón et al. (2015) or Gauthier and Degnan (2008), and As-NF- κ B1-3 were identified from Richter et al. (2018). The human cell codon-optimized cDNA corresponding to the predicted aa sequence of each were synthesized by GenScript (Piscataway, NJ).

2.3.2 Digestion of DNA

DNA was digested with endonucleases and endonuclease buffers (New England Biolabs) according to the manufacturer's protocol. Generally, restriction digestion reactions contained 1 µg of DNA, 1 µl of restriction enzyme(s), and 2 µl of the appropriate endonuclease buffer in a total volume of 20 µl. Unless specified otherwise, samples were incubated for 1 h at 37°C in 1.5-ml microcentrifuge tubes. Digested products were analyzed by gel electrophoresis as described previously (Wolenski, 2012).

2.3.3 Polymerase chain reaction (PCR)

PCRs contained 10-20 ng of template plasmid DNA, 1.25 µM forward and reverse primers, 0.2 µM dNTPs, 0.02 Units/µl Q5 Hot Start DNA polymerase (New England Biolabs), and 5 µl of 5X Q5 reaction buffer in a total volume of 25 µl. The PCR mixture was mixed by pipetting and subjected to the following thermocycler conditions: 98°C for 30 sec, 35 cycles of 1) 98°C for 10 sec, 2) 58-65°C for 30 sec (optimized for each primer pair), and 3) 72°C for 30 sec (per kilobase of product), followed by a single extension at 72°C for 5 min. PCR products were analyzed by gel electrophoresis. When introducing multiple sites of mutations into DNA, overlapping PCR was performed as previously described (Wolenski, 2012).

2.3.4 Oligonucleotides

Oligonucleotides used in this thesis are listed in Table 2.1. Many oligonucleotides were designed to include start and stop codons as well as restriction endonuclease sites.

2.3.5 DNA ligation

DNA was run on a 1.5% low melt agarose gel at 60 volts (V) and correctly sized DNA bands were excised from the gel and placed in 1.5-ml microcentrifuge tubes. Excised DNA-gel slices were directly used or purified using phenol and chloroform for ligation reactions. Ligation reactions using T4 DNA ligase were performed as described previously (Haery, 2016).

2.3.6 Preparation of chemically competent DH5 α cells

Chemically competent *Escherichia coli* DH5 α cells were used for all cloning protocols. The preparation of chemically competent *E. coli* DH5 α bacterial cells was performed as previously described (Wolenski, 2012).

2.3.7 Transformation of plasmid DNA

Transformation of plasmid DNA was performed essentially as previously described (Wolenski, 2012; Haery, 2016). For transformation of DNA fragment assemblies using T4 DNA ligase and HiFi Assembly, the entire ligation/DNA assembly reaction volume (20-25 μ l) was combined with 60 μ l of chemically competent *E. coli* DH5 α , 170 μ l TCM [10 mM Tris-HCl (pH 7.5), 10 mM CaCl₂, 10 mM MgCl₂], and incubated on ice for 1 h in a 1.5-ml microcentrifuge tube. Samples were then heat-shocked at 42°C in a water bath for 2 min, then supplemented with 1 ml of LB [1% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, pH 7.4]. After 1 h at 37°C shaking at 225 rpm, cells were pelleted at 3,000 rpm for 3 min, resuspended in 100 μ l LB and plated on LB agar plates [1% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract,

1% (w/v) NaCl, pH 7.4; 1.5% (w/v) bacto-agar (Difco)] containing 100 µg/ml ampicillin and 50 µg/ml chloramphenicol (amp/cam), or 100 µg/ml ampicillin (amp) only. These plates were incubated at 37°C overnight. Bacterial colonies were picked the next day to inoculate 5 ml of LB-amp or LB-amp/cam for overnight growth at 37°C shaking at 225 rpm.

2.3.8 Plasmid constructions

Details about the source and construction of plasmids used in this thesis are described in Table 2.2.

2.3.9 Small-scale plasmid preparation

Small-scale preparations (minipreps) of plasmid DNA were performed as described previously (Wolenski, 2012; Haery, 2016).

2.3.10 Large-scale plasmid preparation

Large-scale plasmid isolation was performed by the cesium preparation method as described previously (Wolenski, 2012) or by using the ZymoPURE Midiprep kit (Zymo Research) following the manufacturers protocol.

2.4 Cell culture

Cell lines used in this thesis are listed in Table 2.3.

2.4.1 Vertebrate cell culture

DF-1 chicken fibroblasts and human HEK 293 or 293T cells were grown in Dulbecco's modified Eagle's Medium (Invitrogen) supplemented with 10% fetal bovine

serum (Biologos), 50 units/ml penicillin, and 50 µg/ml streptomycin as described previously (Wolenski, 2012).

2.4.2 *Capsaspora* cell culture

Capsaspora cell cultures (strain ATCC ®30864) were grown axenically in 25 cm² culture flasks (ThermoScientific Nunclon) with 10 ml ATCC medium 1034 (modified PYNFH medium) at 23°C. Cultures for each *Capsaspora* life stage were generated as previously described (Sebé-Pedrós et al., 2013) and as instructed by ATCC. Filopodic cells were maintained in an actively dividing adherent state by scraping and passaging 1/40-1/50 of the cultures every 6-8 days, before floating cells appeared. Floating cystic cells were collected from 14-day-old filopodic cultures. Aggregative cells were created by actively scraping dividing filopodic cells and seeding them into a 25 cm² culture flask, which was gently agitated at 60 RPM for 4-5 days, and grown axenically at 23°C.

2.4.3 Polyethylenimine-based cell transfection

Transfection of cells with expression plasmids was performed using polyethylenimine (PEI) (Polysciences, Inc.) essentially as described previously (Wolenski, 2012). Briefly, on the day of transfection, cells were incubated with plasmid DNA and PEI at a DNA:PEI (1 mg/ml) ratio of 1:6. Media was changed 24 h post-transfection, and whole-cell lysates were prepared 24 h later. If cells were used for immunofluorescence, they were passaged onto glass coverslips on the day prior to fixation.

Although a method for transfection of *Capsaspora* cells has been previously been described (Parra-Acero et al., 2018), we developed a new PEI-based method for transfection of these cells. That is, transfection of *Capsaspora* cells with expression plasmids was performed using polyethylenimine (PEI) (Polysciences, Inc.). Briefly, the day before transfection, actively dividing *Capsaspora* filopodic cells were plated at about 80-90% confluency. The next day, cells were transfected by incubation with 5 µg of plasmid DNA and 25 µl of 1 mg/ml PEI. Media was changed ~20 h post-transfection, and *Capsaspora* cells to be analyzed by immunofluorescence were passaged onto poly-D-lysine (ThermoFisher)-treated glass coverslips on the day prior to fixation.

2.5 Whole-cell extract preparation

Whole-cell extracts were prepared, in general, in AT Lysis buffer, as detailed below.

2.5.1 Tissue culture cells

Non-denaturing whole-cell protein lysates of transfected HEK 293 cells and HEK 293T cells were made using AT Lysis Buffer (20 mM HEPES, pH 7.9, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 20% wt/vol glycerol, 1% [wt/vol] Triton X-100, 20 mM NaF, 1 mM $\text{Na}_4\text{P}_2\text{O}_7 \cdot 10\text{H}_2\text{O}$, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, 1 µg/ml leupeptin, 1 µg/ml pepstatin A, 10 µg/ml aprotinin) essentially as described previously (Wolenski, 2012). Generally, cells were harvested 48 h post-transfection. Tissue culture plates containing HEK 293T cells were scraped with a cell scraper and media containing the cells was transferred to 15-ml conical tubes and samples were centrifuged at 3,000 rpm for 5 min. The media supernatant was removed by aspiration,

cells were washed once with 4°C phosphate buffered saline (PBS; GIBCO), cells were removed by gentle scraping with a rubber scraper, and then transferred to a 1.5-ml microcentrifuge tube. For lysis of a 100-mm tissue culture plate, cells were resuspended in 160 µl of AT buffer by pipetting. For lysis of a 60-mm tissue culture plate, cells were resuspended in 100 µl of AT buffer by pipetting. Cell samples were then passed five times through a 27.5- gauge needle to ensure complete lysis. NaCl was then added to a final concentration of 150 mM and protein lysates were clarified by centrifugation at 13,000 rpm for 25 min at 4°C. Supernatants were then transferred to fresh 1.5-ml microcentrifuge tubes and stored at -80°C. Protein concentration was determined using the Bio-Rad Protein Assay dye reagent (Bio-Rad) as described previously (Wolenski, 2012).

2.5.2 *Cliona* sp.

Black encrusting sponge tissue was obtained from the Boston University Marine Program lab by scraping tissue off a rock substrate, immediately flash freezing the tissue, and storing it at -80 °C until use. To prepare a sponge lysate, a piece of tissue approximately 2 cm by 2 cm was placed into a glass dounce tissue grinder (Wheaton) with 1 ml of AT Lysis Buffer and protease inhibitors (described previously in this chapter). Tissue was ground approximately ten times by hand, and then the sample was transferred to a 1.5-ml microcentrifuge tube and gently rotated for 1 h at 4 °C. The sample in buffer was then stored at -80 °C until use.

2.5.3 *Capsaspora*

Whole-cell lysates of cells from each life stage were prepared in AT Lysis Buffer (20 mM HEPES, pH 7.9, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 20% w/v glycerol, 1% w/v Triton X-100, 20 mM NaF, 1 mM $\text{Na}_4\text{P}_2\text{O}_7 \cdot 10 \text{ H}_2\text{O}$, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, 1 $\mu\text{g/ml}$ leupeptin, 1 $\mu\text{g/ml}$ pepstatin A, 10 $\mu\text{g/ml}$ aprotinin). On the day of lysis, cells from five 25 cm^2 culture flasks (ThermoScientific Nunclon) were washed once with PBS and then scrapped with a rubber scraper with 1 ml of PBS for each and transferred into five 1.7-ml microcentrifuge tube. The cells were pelleted at 3000 rpm for 5 min at 4°C and then combined and resuspended in 250 μl of AT buffer by pipetting. For lysis of a 60-mm tissue culture plate, cells were resuspended in 100 μl of AT buffer by pipetting. Cell samples were then passed five times through a 27.5-gauge needle to ensure complete lysis. NaCl was then added to a final concentration of 150 mM and protein lysates were clarified by centrifugation at 13,000 rpm for 25 min at 4°C. Supernatants were then transferred to fresh 1.5-ml microcentrifuge tubes and stored at -80°C. Protein concentration was determined using the Bio-Rad Protein Assay dye reagent (Bio-Rad) as described previously (Wolenski, 2012).

2.6 SDS-Polyacrylamide gel electrophoresis and Western blotting

Western blotting was performed as described previously (Wolenski et al., 2011; Mansfield et al., 2017). Briefly, cell and tissue extracts were first separated on SDS-polyacrylamide gels. Proteins were then transferred to nitrocellulose at 4°C at 250 mA for 2 h followed by 160 mA overnight. The membrane was blocked in TBST (10 mM

Tris-HCl [pH 7.4], 150 mM NaCl, 0.1% v/v Tween 20) containing 5% powdered milk (Carnation) for 1 h at room temperature. Filters were incubated at 4°C with primary antiserum diluted in 5% milk TBST as follows: FLAG antiserum (1:1000, Cell Signaling Technology), α -tubulin antiserum (1: 1000, Cell Signaling Technology) or Ap-NF- κ B antiserum (1:1000, (Mansfield et al., 2017)) (see Table 2.4 for details on antibodies). After extensive washing in TBST, filters were incubated with anti-rabbit horseradish peroxidase-linked secondary antibody (1: 4000, Cell Signaling Technology) or anti-mouse horseradish peroxidase-linked secondary antibody (1:3000, Cell Signaling Technology). Immunoreactive proteins were detected with SuperSignal West Dura Extended Duration Substrate (Pierce) after exposure to X-ray film, or using a Sapphire Biomolecular Imager (Azure Biosystems).

2.7 Electrophoretic mobility shift assay (EMSA)

EMSAs were performed using a ^{32} P-labeled κ B-site probe (GGGAATTCCC, see Table 2.1) and 293 whole-cell extracts, as described previously (Wolenski et al., 2011). For gel analysis of the EMSA, a 10X TGE buffer was prepared (1.9 M glycine, 250 mM, 10 mM EDTA pH 8.3) and chilled to 4°C. A 50 ml non-denaturing 5% polyacrylamide gel was cast [5% (v/v) polyacrylamide, 2.5% (v/v) glycerol, 400 μ l 10% (w/v) APS, and 80 μ l TEMED in 1X TGE] and allowed to polymerize for at least 1 h. The gel was pre-run for 30 min at 50 mA, then radioactive samples were loaded and the gel was electrophoresed at 30 mA for 2-3 h, until the 1X blue dye [6X: 50% (w/v) sucrose, 0.15% (w/v) bromophenol blue, 10 mM EDTA] ran through ~75% of the gel. The gel

was then dried under vacuum and either exposed overnight to a phosphor plate at room temperature or X-ray film at -80°C, and developed or imaged on a Sapphire Biomolecular Imager (Azure Biosystems).

2.8 GST-tagged protein purification

Isolation of GST proteins was performed essentially as previously described (Wolenski, 2012; Brennan, 2017). Glycerol stocks of BL21 DE3 *E. coli* containing the pGEX vector, BL21 DE3 pLysE *E. coli* containing the GST-Of-TIR or GST-Nv-TIR fusion, were used to inoculate 5 ml of LB-amp or LB-amp/cam, respectively, and were grown overnight at 37°C with shaking at 225 rpm. For expression of pGEX vector control, one overnight culture was inoculated. For expression of GST-Nv-TIR or GST-Of-TIR, six or 12 overnight cultures were inoculated, respectively. Approximately 18 h later, cultures were spun at 3,000 rpm, the supernatant discarded, and the cell pellets were resuspended in 1 ml of fresh LB with the required antibiotics. The resuspended culture was then transferred to a 50 ml flask of LB-amp or LB-amp/cam and grown in a shaking incubator for 3 h at 37°C. To induce protein expression, 0.1 mM of isopropyl-1-thio- β -D-galactoside (IPTG; Sigma) was added to each culture and cells were grown for an additional 3 h at 37°C in a shaking incubator.

Bacteria were then pelleted using a Sorvall SH-3000 rotor in a Sorvall RC-5 centrifuge at 5,000 rpm for 10 min at 4°C. The cell pellet was washed once with 50 ml PBS and centrifuged as described above. The supernatant was discarded, and the bacterial pellet was resuspended in 2 ml of PBS with protease inhibitors (1 mM PMSF,

1% aprotinin, and 2 µg/ml pepstatin-A). The resuspended pellet was lysed in a 15-ml conical tube by sonication using a Sonic Dismembrator 550 for six rounds of 30 sec at setting 4, with 2 min on ice in between each round of sonication. The lysate was transferred to a 2-ml microcentrifuge tube and each tube was supplemented with 200 µl of 10% Triton X-100 and rocked on a nutator for 10 min at 4°C. To clarify the lysate, each 2-ml tube was centrifuged at 14,000 rpm for 15 min at 4°C.

The GST-Of-TIR or GST-Nv-TIR lysates were combined in a 15-ml conical tube and 60 µl of 50% glutathione-agarose bead slurry (Sigma) was added. 60 µl of 50% glutathione-agarose bead slurry was also added to the pGEX vector clarified lysate. These lysates were rocked on a nutator for 18 at 4°C. Beads were then gently pelleted at 3,000 rpm for 3 min and washed five times with ice cold PBS to remove unbound proteins. After the final PBS wash, the beads were resuspended in 1 ml of PBS with protease inhibitors and 5% of the resuspended bead slurry was electrophoresed on an SDS-polyacrylamide gel and stained with Coomassie blue to determine the relative amounts of expressed GST-tagged proteins as described previously (Wolenski, 2012).

2.9 GST-pull down assays

GST and GST-TIR domain fusion proteins were expressed in BL21 and BL21 pLyse bacterial cells, respectively, and were purified from extracts using glutathione beads as described previously (Garbati et al., 2010). To assess the amounts and sizes of the purified GST proteins 1% of GST and 10% of the GST samples were electrophoresed on a 10% SDS-polyacrylamide gel, which was then stained with Coomassie blue (Bio-

Rad). The remaining portions of the GST proteins on beads were incubated with extracts from 293 cells expressing FLAG-MYD88. The beads were then washed four times with cold PBS, and were boiled in 2X SDS sample buffer to release proteins from the beads. Proteins were electrophoresed on a 7.5% SDS-polyacrylamide gel, and the gel was subjected to anti-FLAG Western blotting as described elsewhere (Wolenski et al., 2011) and in this chapter.

2.10 Reporter gene assays

2.10.1 Luciferase reporter assay

Luciferase reporter gene assays were performed in 293 cells as previously described (Sullivan et al., 2009; Wolenski, 2012). For TLR activation luciferase assays, the 293 cells were grown in 6-well 35-mm plates to 60% confluence then transfected with the following using PEI: 1.5 ug 3 $\times$ - κ B, 1.5 ug RSV- β GAL, and basal NF- κ Bs or NF- κ B mutants. The NF- κ B-site luciferase reporter contains three copies of the major histocompatibility complex (MHC) class I κ B element upstream of the gene encoding luciferase (Wolenski, 2012). Media was changed 20 h later, and cells were lysed 48 h post transfection and analyzed according to the Luciferase Assay System (Promega) manufacturer's instruction.

Luciferase measurements were first divided by the respective β -galactosidase values. If the assay was done in triplicate, the average value (luciferase/ β -galactosidase) was calculated. All values were normalized to a vector control that was set to 1.0. Standard error was calculated for the normalized values of all conditions.

2.10.2 GAL4 reporter assay

Two separate 3 ml cultures of selective Trp⁻ media were inoculated with two independent yeast colonies from GB-plasmids and then grown at 30°C for 24 h. 200 µl of cells from each culture was transferred into a 96-well micro plate for subsequent OD₅₉₅ measurements; 200 µl of Trp⁻ media was used as a blank control. One mL of each culture was then transferred to a 1.5-ml microcentrifuge tube, centrifuged for 1 min, decanted, and then resuspended in 90 µl of lysis buffer (100 mM Tris [pH 7.6]; 0.05% Triton X-100) with vortexing. Samples were freeze/thawed five times by repeated immersion in a dry ice/ethanol bath (until frozen) and then a 37°C water bath (until completely thawed). To each sample 450 µl of Z/ONPG solution (60 mM Na₂HPO₄, 40 mM NaH₂PO₄, 10 mM KCl, 1 mM MgSO₄, 50 mM 2-mercaptoethanol, 0.8 mg/ml orthonitrophenyl-β-D-galactopyranoside [ONPG], 1 mM DTT, 0.005% SDS, pH 7) was added. All samples were incubated at 30°C until appearance of a yellow color (or 35 min elapsed without any change in color), at which point the reaction was stopped by addition of 225 µl of 1 M Na₂CO₃ (time of reaction from start to termination was recorded for each sample). Samples were centrifuged for 2 min and 200 µl of the supernatant was transferred into a 96-well plate for measurement of β-galactosidase activity by measuring OD₄₁₅; as a blank control 200 µl of supernatant from a centrifuged sample containing 90 µl yeast lysis buffer, 450 µl of Z/ONPG, and 225 µl of 1 M Na₂CO₃ was used.

2.11 *In vitro* kinase assay

In vitro kinase assays were performed as described previously (Wolenski et al., 2011). Briefly, human 293T cells were transfected with pcDNA-FLAG-IKK β and pcDNA-FLAG-Of-IKK constructs, lysed two days later, and the kinases were immunoprecipitated with anti-FLAG beads (Sigma). These immunoprecipitates were then incubated with approximately 4 μ g of GST alone, GST-Of-NF- κ B or GST-Of-NF- κ B-3X-Ala C-terminal peptides and 5 μ Ci [γ - ^{32} P]ATP (Perkin Elmer) in kinase reaction buffer (25 mM Tris-HCl, pH 7.5, 20 mM β -glycerophosphate, 10 mM NaF, 10 mM MgCl₂, 2 mM DTT, 500 μ M Na₃VO₄, 50 μ M ATP) for 30 min at 30°C. Samples were then boiled in 2X SDS sample buffer and electrophoresed on a 10% SDS-polyacrylamide gel. The ^{32} P-labeled GST-Of-NF- κ B peptides were detected by phosphorimaging. As a control for protein input, 4 μ g of GST alone, GST-Of-NF- κ B or GST-Of-NF- κ B-3X-Ala was electrophoresed on a 10% SDS-polyacrylamide gel, and proteins were detected by staining with Coomassie blue (Bio-Rad).

In vitro kinase assays were performed as described previously (Wolenski et al., 2011; Mansfield et al., 2017). Briefly, human 293T cells were transfected with pcDNA-FLAG-Hu-IKK β , lysed two days later, and the tagged kinase was immunoprecipitated with anti-FLAG beads (Sigma). The immunoprecipitate was then incubated with approximately 4 μ g of GST alone, GST-Aq-NF- κ B, or GST-Aq-NF- κ B-ALA C-terminal peptides and 5 μ Ci [γ - ^{32} P] ATP (Perkin Elmer) in kinase reaction buffer (25 mM Tris-HCl, pH 7.5, 20 mM β -glycerophosphate, 10 mM NaF, 10 mM MgCl₂, 2 mM DTT, 500 μ M Na₃VO₄, 50 μ M ATP) for 30 min at 30°C. Samples were then boiled in 2X SDS

sample buffer and electrophoresed on a 10% SDS-polyacrylamide gel. The ^{32}P -labeled GST-Aq-NF- κB peptides were detected by phosphorimaging. As a control for protein input, 4 μg of GST alone, GST-Aq-NF- κB , or GST-Aq-NF- κB -ALA was electrophoresed on a 10% SDS-polyacrylamide gel, and these proteins were detected by staining with Coomassie blue (Bio-Rad).

2.12 Indirect immunofluorescence

2.12.1 Indirect immunofluorescence of tissue culture cells

Indirect immunofluorescence of transfected DF-1 cells was performed to detect the subcellular localization of basal NF- κB s and mutants. DF-1 cells were transfected with 3 μg of expression vector and were passaged onto UV sterilized coverslips 2 days post-transfection. Immunofluorescence of transfected DF-1 cells was performed as described previously (Wolenski et al. 2011; Wolenski, 2012). Information about antibodies used for immunofluorescence of DF-1 cells is included in Table 2.4.

2.12.2 Indirect immunofluorescence of sectioned *Cliona sp.* tissue

Flash-frozen black encrusting sponge tissue was placed into ice-cold 4% paraformaldehyde in full-strength filtered artificial sea water (ASW) (Instant Ocean) to fix overnight at 4 °C. Fixed tissue was then washed three times with filtered ASW, and then dehydrated in 30% (w/v) sucrose in ASW at 4 °C overnight. Samples were then embedded in Optimal Cutting Temperature Compound (Sakura Tissue-Tek), cryosectioned on a microtome into 45 μm slices, placed onto Superfrost Plus Microscope Slides (Fisher Scientific), and kept at -20 °C until subjected to preparation for

immunofluorescence. Tissue slices were rehydrated with room temperature PBS, and the tissue slices were placed in individual wells of a 48-well plate using forceps. Tissue slices were then washed with agitation three times with 1X PBS, and then blocked and permeabilized with 0.3% Triton X-100 + 5% goat serum (Gibco) in PBS for 1 h at room temperature. Slices were then incubated with anti-Ap-NF- κ B antibody (Mansfield et al., 2017) (1:5000) diluted in PBS containing 0.3% Triton X-100 plus 5% goat serum for an overnight period at 4°C in a humidified chamber, with rotation. Slides were then washed with agitation three times in PBS containing 0.3% Triton X-100, and were incubated with secondary antibody (goat-anti-rabbit Alexa Fluor 488 (1:500) diluted in 5% goat serum/0.3% Triton X-100 in PBS. Slices were then washed three times with PBS/0.3% Triton X-100, and on the last wash Hoechst (1:5000 of 20 μ M) was added for 10 min. Slices were then washed three more times with PBS/0.3% Triton X-100, and then mounted onto Superfrost Plus Microscope Slides (Fisher Scientific) with Prolong Gold (Molecular Probes by Life Technologies) with coverslips. Samples were imaged on a confocal microscope (Nikon C2 Si).

2.12.3 Indirect immunofluorescence of *Capsaspora owczarzaki* cells

Coverslips were treated by pipetting approximately 500 μ l of Poly-L-lysine (Sigma P4707-50mL) onto coverslips and placing in 37°C incubator for 2 h or overnight. Lysine was washed from coverslips three times with PBS. Coverslips were then placed in 35-mm tissue culture dish and UV sterilized by putting each coverslip in the open dish for about 40-45 min in the tissue culture hood. Approximately 24-30 h after fluid changing transfected Co cells, cells were passed onto UV-sterilized and Poly-L-lysine

treated coverslips. Cells were allowed to grow another 24 h and then subjected to indirect immunofluorescence exactly like tissue culture cells (described in this chapter), except cells were fixed using 4% paraformaldehyde.

2.13 Treatment with LPS

2.13.1 Treatment of *Orbicella faveolata* tissue with LPS

Collected *O. faveolata* were exposed to lipopolysaccharide (LPS) from *Escherichia coli* 0127:B8 (Sigma) at a final concentration of 10 µg/ml for 30 min, and samples were then aerated with seawater for 4 h before being removed and frozen in liquid nitrogen. Samples were lysed in a 50 mM Tris-base, pH 7.8, with 0.05 mM dithiothreitol, and then RNA was extracted, made into cDNA, and sequenced.

2.13.2 Treatment of *Cliona sp.* tissue with LPS

Black encrusting sponge tissue of approximately 1 cm x 1 cm was obtained from the Boston University Marine Program lab by collecting tissue off a rock substrate, transferring it immediately into a glass dish with ASW, and allowing it to acclimate at 25°C with light provided by Sylvania Gro-Lux (GRO/Aq/RP) fluorescent bulbs at approximately 20 µmol photons/m²/sec for 1 h. The sample was then cut in half with an 70% ethanol- cleaned razor and each piece was placed into a 6-well plate with 10 ml of ASW and lipopolysaccharide (LPS) from *Escherichia coli* 0111:B4 (Sigma) at a final concentration of 5 µg/ml (or an equal volume of sterile distilled H₂O, in which the LPS was dissolved). The LPS or water was applied directly above the surface of the tissue, and samples were incubated for 30 min. The tissue samples were then immediately

placed into AT buffer and lysed as described above, before being subjected to Western blotting or EMSA.

2.14 Immunoprecipitations

HEK 293T cells were transfected with MYC-As-NF- κ B2 and either a pcDNA FLAG empty vector, FLAG-As-NF- κ B1, or FLAG-As-NF- κ B3 as above (See “Vertebrate cell culture and transfection”). Lysates were prepared 48 h later and were incubated with 50 μ l of a 1X PBS-washed anti-FLAG bead slurry (Sigma) overnight at 4 °C with gentle rocking. The next day, the beads were washed three times with 1X PBS. The pellet was then boiled in 2X SDS sample buffer, and the supernatant was electrophoresed on a 7.5% SDS-polyacrylamide gel, as above (see “Western blotting”). The membrane was then probed with a rabbit anti-MYC (Cell Signaling Technologies, 1:1000) antiserum, then with anti-rabbit horseradish peroxidase-linked secondary antiserum (1: 4000, Cell Signaling Technology), and reacted with SuperSignal West Dura substrate (Pierce). An image was then obtained on a Sapphire Biomolecular Imager (Azure Biosystems). The membrane was stripped and probed with rabbit anti-FLAG antiserum as above. These same transfection and IP experiments were repeated with MYC-As-NF- κ B3 and either pcDNA FLAG empty vector, FLAG-As-NF- κ B1, or FLAG-As-NF- κ B2.

2.15 *Cliona* sp. taxon identification by CO1 gene analysis

Frozen black encrusting sponge tissue was lysed according to the RNAqueous

Total RNA Isolation Kit (ThermoFisher) protocol for extracting total nucleic acid. Total genomic DNA was assessed on an agarose gel. To sequence the mitochondrial cytochrome oxidase subunit 1 (CO1) gene, we used CO1 degenerate primers (see Table S2) and the PCR conditions from Vargas et al. (2012) with slight modifications. Briefly, we used standard three-step PCR conditions with an initial denaturation step of 3 min at 94 °C, followed by 35 cycles of 30 s at 94 °C, 30 s at 40 °C, and 1 min at 72 °C, and a final extension step of 5 min at 72 °C. PCR products were visualized on 1.7% agarose gels and then purified with the Qiagen MinElute PCR purification kit (Qiagen). Purified PCR products were sent for Next-Generation Sequencing using the Complete Amplicon Sequencing method at the Massachusetts General Hospital Center for Computational & Integrative Biology DNA Core (<https://dnacore.mgh.harvard.edu/new-cgi-bin/site/pages/index.jsp>). To identify poriferan sequences, the sequencing results were compared by BLAST analysis to the NCBI database.

2.16 *Capsaspora owczarzaki* RNA-seq analysis

RNA-sequencing data from each life stage of *Capsaspora* was obtained from Seb  -Pedr  s et al. (2013). To identify potential target genes of NF-  B, we sorted all sequencing data with genes that were differentially expressed in the same manner as NF-  B for each life stage (lowest expression in aggregative, medium expression in filopodic, highest expression in cystic). We discarded genes if they were not expressed at a given life stage (i.e., had an RPKM of 0). Of the 8674 total genes, 1348 genes were differentially expressed in a manner similar to NF-  B. Of these 1348 genes, 389 genes

had annotated homologs, and 305 had human homologs (Table 5.2). We then performed GO analysis (<http://pantherdb.org>) by entering the UniProt ID of each gene and selecting *Homo sapiens*. We then created and analyzed the Biological Processes that were present in this gene list.

To identify genes with potential upstream NF- κ B binding sites among these 1348 genes, we also extracted the 500 base pair sequence upstream of each gene. We then imported these 1348 upstream regions into MEME-FIMO and scanned the sequences for NF- κ B motifs (but not Rel motifs) extracted from JASPAR (<http://jaspar.genereg.net>) (Table 5.4).

Table 2.1 Oligonucleotides used in this thesis

Primers used for subcloning

Primer Name	Primer Sequence
pcDNA-FLAG-Of-RHD-F	5'- GCGGAATTCAGCTACAAACAGTGAGAGACAG-3'
pcDNA-FLAG-Of-RHD-R	5'-ATAGCGGCCGCTCATTATCCCAACAGGCCCAGTGGAGCG -3'
pcDNA-FLAG-Of-Cterm-F	5'- GCGGAATTCATTCTCTGATTACTTCTCCGGA -3'
pcDNA-FLAG-Of-Cterm-R	5'- ATAGCGGCCGCTCATTAGAACACATCCTGGTGC -3'
GB-Of-RHD-F	5'-GCGGAATTCGCTACAAACAGTGAGAGACAG-3'
GB-Of-RHD-R	5'-ATAGTCGACTCATTATCCCAACAGGCCCAGTGGAGCG-3'
GB-Of-Cterm-F	5'-GCGGAATTCTTCTCTGATTACTTCTCCGGA-3'
GB-Of-Cterm-R	5'-ATAGTCGACTCATTAGAACACATCCTGGTGC-3'
Of-NF-κB-3X-Ala-F-1	5'-CCTGGCTGAAAAATCTGCAGCCG-3'
Of-NF-κB-3X-Ala-F-2	5'-GACGCAGGACTGGGTGCCCAAGCTTTGTC-3'
Of-NF-κB-3X-Ala-R-3	5'-CAAAGCTTGGGCACCCAGTCCTGCGTCGAAG-3'
Of-NF-κB-3X-Ala-R-4	5'-ATAGCGGCCGCTCATTAGAACACATCCTGGTGC-3'
Of-NF-κB-F-Cterm	5'-GCGCGCGAATTCTAGTGCTGGAAATCTTG-3'
Of-NF-κB-R-Cterm	5'-GCGCGCTCGAGCTATGATCTGTAAGCGGACAA-3'
Of-TLR-F-TIR	5'-GCGCGCGAATTCTACAGAGCATTACGAAA-3'
Of-TLR-R-TIR	5'-GCGCGCTCGAGTCACTCCAAAATGTCC-3'

pcDNA- FLAG-Aq- RHD-F	5'- CGAAGGATCCCTGCTTTTAATGGTATTGATCCC-3'
pcDNA- FLAG-Aq- RHD-R	5'-ATATCTCGAGCTAGTTCCCACCGGAGAAC -3'
pcDNA- FLAG-Aq- Cterm-F	5'- ATCTAGAATTCAACTGGAGGCGGGACCACG -3'
pcDNA- FLAG-Aq- Cterm-R	5'- CGAGGGATCCCTAAACTTGACTAGAAGG -3'
GB-Aq- RHD-F	5'-CGAAGGATCCCTGCTTTTAATGGTATTGATCCC-3'
GB-Aq- RHD-R	5'-ATATCTCGAGCTAGTTCCCACCGGAGAAC-3'
GB-Aq- Cterm-F	5'-ATCTAGAATTCACTGGAGGCGGGACCACG-3'
GB-Aq- Cterm-R	5'-CGAGGGATCCCTAAACTTGACTAGAAGG-3'
Aq-NF-κB- ALA-F-1	5'-GCACTCCTGGATATCAATCGTA-3'
Aq-NF-κB- ALA-F-2	5'- CTTATGATGCTGGCATCGCTGCAAATCCTGCAGCTCTCTCTAAT GAACAAG-3'
Aq-NF-κB- ALA-R-3	5'- CTTGTTCAATTAGAGAGAGCTGCAGGATTTGCAGCGATGCCAGC ATCATAAG-3'
Aq-NF-κB- ALA-R-4	5'- GCTCGAGCGGCCGCGCTCGAGGGATCCTAAACTTGACTAGAAG GAATGG-3'
Aq-NF-κB- DEL-F	5'-CGCGGCTGCAGGACAGCAGTACAGAATC-3'
Aq-NF-κB- DEL-R	5'-CGGGCCTGCAGTCCTCCCATTCCTC-3'
GST-Aq- NF-κB- Cterm-F	5'-CGCGCGAATTCTAAATATAATTGATCGATATATGCAAG-3'
GST-Aq- NF-κB- Cterm-R	5'-GCGCGCTCGAGCTACTTCCCTACAACCTGTTC-3'

Aq-NF-κB-SER-F-2	5'-GGCTTCGGTTCACAATCTGCAGAGCTCTCTAATGAACAAGTTGTAGG-3'
Aq-NF-κB-SER-R-3	5'-GAGCTCTGCAGATTGTGAACCGAAGCCAGAATCATAAGTCTGTTAC-3'
F-Gib-RHD-1	5'-CTGGTGGGGTCGTGAAACGGCGTTAAGACTCAGGG-3'
R-Gib-RHD-1	5'-GGTACCATGGACTACAAGGACG -3'
F-Gib-Cterm-2	5'- CCGTTTTTCACGACCCCACCAGAATC -3'
R-Gib-Cterm-2	5'- CTAGCATTTAGGTGACACTATAGAATAGGG -3'
F-Gib-FLAG-3	5'-CTATTCTATAGTGTACCTAAATGCTAGAGC-3'
R-Gib-FLAG-3	5'-CCTTGTAGTCCATGGTACCAAGC-3'
F-FLAG-Co-NF-κB	5'-CGCGC <u>GAAATTC</u> AGACCTTTCTGAACTTTCCGGATGGG-3'
R-FLAG-Co-RHD	5'-GCGCG <u>CTCGAGT</u> TACCCGCCTTTACTTGAAGAACTCCC-3'
F-FLAG-Co-ANK	5'-CGCGC <u>GAAATTC</u> AGATGGGTCCGACGGTGGAAATG-3'
R-FLAG-Co-NF-κB	5'-GCGCG <u>CTCGAGT</u> TAAATCAACGGAGTAAAGGGCATGGG-3'
F-GBT9-Co-NF-κB	5'-CGCGC <u>GAAATTC</u> GACCTTTCTGAACTTTCCGGATGGG-3'
R-GBT9-Co-NF-κB	5'-GCGCG <u>CTCGAGT</u> TAAATCAACGGAGTAAAGGGCATGGGT-3'
R-GBT9-Co-RHD	5'-GCGCG <u>CTCGAGT</u> TACCCGCCTTTACTTGAAGAACTCCC-3'
F-GBT9-Co-Cterm	5'-CGCGC <u>GAAATTC</u> GATGGGTCCGACGGTGGAAATG-3'
F-GBT9-As-NF-κB1-EcoRI	5'-CGCGC <u>GAAATTC</u> GAAGCCAGATGGCGGCGGGCCCCCTGGCCTGG-3'
R-GBT9-As-NF-κB1-SalI	5'-CGCGC <u>GTCGACT</u> CACATCAGGGCAGGGCCAGGGGCGGCATCAC TAGC-3'
F-GBT9-As-NF-κB2-EcoRI	5'-CGCGC <u>GAAATTC</u> GAGCCCTGGCGAACACGCCAATACTCAC-3'

R-GBT9- As-NF- κB2-SalI	5'- CGCGC <u>GTCTGACT</u> TAAACGGAAGTCTCGAAGGACCCGTATACCC TCC-3'
F-GBT9- As-NF- κB3-EcoRI	5'- CGCGC <u>GAATTCC</u> CAAATGATTTCGAGGCTGTTCTGACGTCTGGG G-3'
R-GBT9- As-NF- κB3-SalI	5'-CGCGC <u>GTCTGACT</u> CATTTCGGGTGGTGGTCTGCGCCTTC-3'
F-MYC- As-NF- κB2-EcoRI	5'- <u>GAATTCT</u> GAGCCCTGGCGAACAC-3'
R-MYC- As-NF- κB2-XbaI	5'- <u>TCTAGAT</u> TAAACGGAAGTCTCGAAGGAC-3'
F-MYC- As-NF- κB3-EcoRI	5'- <u>GAATTCT</u> CCAAATGATTTCGAGGCTGTTC-3'
R-MYC- As-NF- κB3-XbaI	5'- <u>TCTAGAT</u> CATTTCGGGTGGTGGTC-3'
EMSA NF- κB- Consensus	5'- TCGAGAGGTCGGG <u>GGAATTCCCCCCCC</u> CG -3'
	5'- TCGACGGGGGGGGGAATTCCCCGACCTC -3'
COI Degenerate Primers for Sponge Sequencing:	
dgLCO149 0	5'-GGTCAACAAATCATAAAGAYATYGG-3'
dgHCO219 8	5'-TAAACTTCAGGGTGACCAAARAAYCA-3'

Table 2.2 Plasmids used in this thesis

Plasmid Name	Plasmid Description
pcDNA-FLAG	pcDNA with a 5' FLAG Tag (Sullivan et al., 2009).
pcDNA-FLAG-N ν -NF- κ B	Wolenski et al., 2011
pcDNA-FLAG-Aq-NF- κ B	BamHI-BamHI fragment containing amino acids 2-1096 of Aq-NF- κ B was PCR-amplified from pMD5-Aq-NF- κ B and subcloned into Bam HI-digested pcDNA-FLAG (Mansfield et al., 2017).
pcDNA-FLAG-Aq-NF- κ B-ALA	EcoRV-Aq-NF- κ B-1037-1082-ALA and Aq-NF- κ B-1066-1096-ALA-NotI PCR fragments were used as a template for assembly PCR of an EcoRV-NotI fragment containing full-length pcDNA-FLAG-Aq-NF- κ B. Primers used for amplification were Aq-NF- κ B-ALA-F-1, Aq-NF- κ B-ALA-F-2, Aq-NF- κ B-ALA-R-3, Aq-NF- κ B-ALA-R-4.
pcDNA-FLAG-Aq-NF- κ B- Δ	BamHI-PstI digested pcDNA-FLAG-Aq-NF- κ B PCR product containing codons 2-511 and PstI-BamHI digested pcDNA-FLAG-Aq-NF- κ B product containing codons 681-1096 were used as a template for assembly PCR of the two PCR products. Primers used for amplification were pcDNA-FLAG-Aq-RHD-F, pcDNA-FLAG-Aq-Cterm-R, Aq-NF- κ B-DEL-F, Aq-NF- κ B-DEL-R
pcDNA-FLAG-Aq-NF- κ B- Δ ALA	BamHI-PstI digested pcDNA-FLAG-Aq-NF- κ B-ALA PCR product containing codons 2-511 and PstI-BamHI digested pcDNA-FLAG-Aq-NF- κ B-ALA product containing codons 681-1096 were used as a template for assembly PCR of the two PCR products. Primers used for amplification were Aq-NF- κ B-ALA-F-1, Aq-NF- κ B-ALA-F-2, Aq-NF- κ B-ALA-R-3, Aq-NF- κ B-ALA-R-4.
pcDNA-FLAG-Aq-RHD	BamHI-XhoI digested Aq-RHD PCR product containing codons 2-452 was subcloned into BamHI-XhoI digested pcDNA-FLAG. Primers: pcDNA-FLAG-Aq-RHD-F and pcDNA-FLAG-Aq-RHD-R. PCR- amplified from pcDNA-FLAG-Aq-NF- κ B.

pcDNA-FLAG-Aq-Cterm	EcoRI-BamHI digested Aq-Cterm PCR product containing codons 442-1095 was subcloned into EcoRI-BamHI digested pcDNA-FLAG. Primers: pcDNA-FLAG-Aq-Cterm-F and pcDNA-FLAG-Aq-Cterm-R. PCR- amplified from pcDNA-FLAG-Aq-NF- κ B.
pcDNA-FLAG-Ap-IKK	Mansfield et al., 2017
pcDNA-FLAG-Aq-TBK	BamHI-BamHI fragment containing Aq-TBK (codons 2-759) was excised from pUC57-Aq-TBK and subcloned into BamHI digested pcDNA-FLAG. The cloned vector was then cut with HindIII to ensure correct orientation into pcDNA-FLAG vector.
HA-IKK β -SS-EE	Starczynowski et al., 2007
pUC57-Aq-TBK	pUC57-Simple with Aq-TBK cDNA codon-optimized for expression in human cells. Has a 5' BamHI site and 3' BamHI site for excision. Synthesized by GenScript.
pGBT9	Wolenski et al., 2011
pGB-Nv-NF- κ B	Wolenski et al., 2011
pGB-Aq-RHD	BamHI-XhoI digested Aq-RHD PCR product containing codons 2-452 was subcloned into BamHI-XhoI digested pGBT9. Primers: GB-Aq-RHD-F and GB-Aq-RHD-R. PCR-amplified from pcDNA-FLAG-Aq-NF- κ B.
pGB-Aq-Cterm	EcoRI-BamHI digested Aq-Cterm PCR product containing codons 442-1095 was subcloned into EcoRI-BamHI digested pGBT9. Primers: GB-Aq-Cterm-F and GB-Aq-Cterm-R. PCR- amplified from pcDNA-FLAG-Aq-NF- κ B.
pcDNA-FLAG-Aq-NF- κ B-Aiptasia serine mutant	EcoRV-Aq-NF- κ B-1037-1082-SER and Aq-NF- κ B-1066-1096-SER-NotI PCR fragments were used as a template for assembly PCR of an EcoRV-NotI fragment containing full-length pcDNA-FLAG-Ap-NF- κ B. Primers used for amplification were Aq-NF- κ B-ALA-F-1, Aq-NF- κ B-SER-F-2, Aq-NF- κ B-SER-R-3, Aq-NF- κ B-ALA-R-4.
pGEX-KG	Bacterial expression plasmid containing a 5' GST tag.

pGEX-KG-Aq-Cterm	EcoRI-XhoI fragment containing Aq-NF- κ B codons 1048-1086 was subcloned into EcoRI-XhoI digested pGEX-KG. Primers GST-Aq-NF- κ B-Cterm-F and GST-Aq-NF- κ B-Cterm-R were used to PCR amplify the fragment from pcDNA-FLAG-Aq-NF- κ B.
pGEX-KG-Aq-Cterm-ALA	EcoRI-XhoI fragment containing Aq-NF- κ B-ALA codons 1048-1086 was subcloned into EcoRI-XhoI digested pGEX-KG. Primers GST-Aq-NF- κ B-Cterm-F and GST-Aq-NF- κ B-Cterm-R were used to PCR- amplify the fragment from pcDNA-FLAG-Aq-NF- κ B-ALA.
pcDNA-FLAG-Co-NF- κ B	EcoRI-XhoI fragment containing amino acids 2-1224 of Co-NF- κ B was excised from pUC57-Co-NF- κ B and subcloned into EcoRI-XhoI digested pcDNA-FLAG.
pcDNA-FLAG-Co-NF- κ B-ALA	Gibson primers were used to amplify off pcDNA FLAG-Co-NF- κ B and pcDNA FLAG Aiptasia NF- κ B-AAA (Mansfield et al., 2017) to create three fragments which were assembled using the Gibson Assembly Cloning Kit (New England BioLabs). Primers F-Gib-RHD-1 and R-Gib-RHD-1 were used to create a fragment that contained the FLAG-tag and Co-NF- κ B aa 2-889; Primers F-Gib-Cterm-2 and R-Gib-Cterm-2 were used to create a fragment that contained the C-terminus of pcDNA FLAG Aiptasia NF- κ B-SSS; Primers F-Gib-FLAG-3 and R-Gib-FLAG-3 were used to create the pcDNA FLAG backbone.
pcDNA-FLAG-Co-NF- κ B-SER	Gibson primers were used to amplify off pcDNA FLAG-Co-NF- κ B and pcDNA FLAG Aiptasia NF- κ B (Mansfield et al., 2017) to create three fragments which were assembled using the Gibson Assembly Cloning Kit (New England BioLabs). Primers F-Gib-RHD-1 and R-Gib-RHD-1 were used to create a fragment that contained the FLAG-tag and Co-NF- κ B aa 2-889; Primers F-Gib-Cterm-2 and R-Gib-Cterm-2 were used to create a fragment that contained the C-terminus of pcDNA FLAG Aiptasia NF- κ B; Primers F-Gib-FLAG-3 and R-Gib-FLAG-3 were used to create the pcDNA FLAG backbone.

pcDNA-FLAG-Co-RHD	EcoRI-XhoI digested Co-RHD PCR product containing aa 2-582 was subcloned into EcoRI-XhoI digested pcDNA-FLAG. Primers: F-FLAG-Co-NF- κ B and R-FLAG-Co-RHD. PCR-amplified from pcDNA-FLAG-Co-NF- κ B.
pcDNA-FLAG-Co-Cterm	EcoRI-XhoI digested Co-Cterm PCR product containing aa 542-1224 was subcloned into EcoRI-XhoI digested pcDNA-FLAG. Primers: F-FLAG-Co-ANK and R-FLAG-Co-NF- κ B. PCR-amplified from pcDNA-FLAG-Co-NF- κ B.
pUC57-Co-NF- κ B	pUC57-Simple with Co-NF- κ B cDNA codon-optimized for expression in human cells. Has a 5' EcoRI site and 3' XhoI site for excision. Synthesized by GenScript.
pcDNA-FLAG-Ap-IKK	Mansfield et al., 2017
pcDNA-FLAG-IKK β -SS-EE	Mansfield et al., 2017
HA-NIK	Mansfield et al., 2017
pGB-Co-NF- κ B	EcoRI-XhoI digested Co-NF- κ B PCR product containing codons 2-1224 was subcloned into EcoRI-SalI digested pGBT9 vector. Primers: F-GBT9-Co-NF- κ B and R-GBT9-Co-NF- κ B. PCR-amplified from pcDNA-FLAG-Co-NF- κ B.
pGB-Co-RHD	EcoRI-XhoI digested Co-RHD PCR product containing codons 2-582 was subcloned into EcoRI-SalI digested pGBT9 vector. Primers: F-GBT9-Co-NF- κ B and R-GBT9-Co-RHD. PCR-amplified from pcDNA-FLAG-Co-NF- κ B.
pGB-Co-Cterm	EcoRI-XhoI digested Co-Cterm PCR product containing codons 542-1224 was subcloned into EcoRI-SalI digested pGBT9 vector. Primers: F-GBT9-Co-Cterm and R-GBT9-Aq-NF- κ B. PCR-amplified from pcDNA-FLAG-Co-NF- κ B.
pUC57-As-NF- κ B1	pUC57-Simple with As-NF- κ B1 cDNA codon-optimized for expression in human cells. Has a 5' BamHI site and 3' BamHI site for excision. Synthesized by GenScript.
pUC57-As-NF- κ B2	pUC57-Simple with As-NF- κ B2 cDNA codon-optimized for expression in human cells. Has a 5' EcoRI site and 3' EcoRI site for excision. Synthesized by GenScript.
pUC57-As-NF- κ B3	pUC57-Simple with As-NF- κ B3 cDNA codon-optimized for expression in human cells. Has a 5'

	EcoRI site and 3' EcoRI site for excision. Synthesized by GenScript.
pcDNA FLAG-As-NF- κ B1	BamHI-BamHI fragment containing amino acids 2-409 of As-NF- κ B1 was excised from pUC57-As-NF- κ B1 and subcloned into BamHI-digested pcDNA-FLAG.
pcDNA FLAG-As-NF- κ B2	EcoRI-EcoRI fragment containing amino acids 2-502 of As-NF- κ B2 was excised from pUC57-As-NF- κ B2 and subcloned into EcoRI-digested pcDNA-FLAG.
pcDNA FLAG-As-NF- κ B3	EcoRI-EcoRI fragment containing amino acids 2-412 of Co-NF- κ B was excised from pUC57-As-NF- κ B3 and subcloned into EcoRI-digested pcDNA-FLAG.
pGB-As-NF- κ B1	EcoRI-SalI digested As-NF- κ B1 PCR product containing codons 2-409 was subcloned into EcoRI-SalI digested pGBT9 vector. Primers: F-GBT9-As-NF- κ B1-EcoRI and R-GBT9-As-NF- κ B1-SalI. PCR- amplified from pcDNA-FLAG-As-NF- κ B1.
pGB-As-NF- κ B2	EcoRI-SalI digested As-NF- κ B2 PCR product containing codons 2-502 was subcloned into EcoRI-SalI digested pGBT9 vector. Primers: F-GBT9-As-NF- κ B2-EcoRI and R-GBT9-As-NF- κ B2-SalI. PCR- amplified from pcDNA-FLAG-As-NF- κ B2.
pGB-As-NF- κ B3	EcoRI-SalI digested As-NF- κ B3 PCR product containing codons 2-412 was subcloned into EcoRI-SalI digested pGBT9 vector. Primers: F-GBT9-As-NF- κ B3-EcoRI and R-GBT9-As-NF- κ B3-SalI. PCR- amplified from pcDNA-FLAG-As-NF- κ B3.
pcDNA MYC vector	Wolenski et al., 2011
pMYC-As-NF- κ B2	EcoRI-XbaI fragment containing amino acids 2-502 of As-NF- κ B2 was PCR-amplified from pcDNA FLAG-As-NF- κ B2 using primers F-MYC-As-NF- κ B2-EcoRI and R-MYC-As-NF- κ B2-XbaI and subcloned into EcoRI-XbaI digested pcDNA-MYC vector.
pMYC-As-NF- κ B3	EcoRI-XbaI fragment containing amino acids 2-412 of As-NF- κ B3 was PCR-amplified from

	pcDNA FLAG-As-NF- κ B3 and subcloned into EcoRI-XbaI digested pcDNA-MYC vector.
pcDNA-FLAG-Of-NF- κ B	EcoRI-NotI fragment containing full-length Of-NF- κ B was excised from pUC57-Of-NF- κ B and subcloned into EcoRI-NotI digested pcDNA-FLAG.
pcDNA-FLAG-Of-NF- κ B-3X-Ala	SgrAI-Of-NF- κ B-1434-2607-3X-Ser-Ala and Of-NF- κ B- 2581-2763-3x-Ser-Ala-NotI PCR fragments were used as a template for assembly PCR of an SgrAI-NotI fragment containing full-length Ap-NF- κ B-3x-Ser-Ala. Primers used for amplification were Of-NF- κ B-3X-Ala-F-1, Of-NF- κ B-3X-Ala-F-2, Of-NF- κ B-3X-Ala-R-3, Of-NF- κ B-3X-Ala-R-4.
pcDNA-FLAG-Of-RHD	EcoRI-NotI digested Of-RHD PCR product containing codons 2-425 was subcloned into EcoRI-NotI digested pcDNA-FLAG. Primers: pcDNA-FLAG-Of-RHD-F and pcDNA-FLAG-Of-RHD-R. PCR- amplified from pUC57-Of-NF- κ B.
pcDNA-FLAG-Of-Cterm	EcoRI-NotI digested Of-Cterm PCR product containing codons 389-920 was subcloned into EcoRI-NotI digested pcDNA-FLAG. Primers: pcDNA-FLAG-Of-Cterm-F and pcDNA-FLAG-Of-Cterm-R. PCR-amplified from pUC57-Of-NF- κ B.
pcDNA-FLAG-IKK α	Starczynowski et al., 2007
pcDNA-FLAG-IKK β	Starczynowski et al., 2007
pcDNA-FLAG-Of-IKK	BamHI fragment containing Of-IKK (codons 2-728) was excised from pUC57-Of-IKK and subcloned into BamHI digested pcDNA-FLAG. The cloned vector was then cut with BamHI and EcoRI to ensure correct orientation into pcDNA-FLAG vector.
pUC57-Of-IKK	pUC57-Simple with Of-IKK cDNA codon-optimized for expression in human cells. Has a 5' BamHI site and 3' BamHI site for excision. Synthesized by GenScript.
pUC57-Of-NF- κ B	pUC57-Simple with Of-NF- κ B cDNA codon-optimized for expression in human cells. Has a 5' EcoRI site and 3' NotI site for excision. Synthesized by GenScript.

pUC57-Of-TLR	pUC57-Simple with Of-TLR cDNA codon-optimized for expression in human cells. Has a 5' EcoRI site and 3' EcoRI site for excision. Synthesized by GenScript.
pGB-Of-RHD	EcoRI-SalI digested Of-RHD PCR product containing codons 2-425 was subcloned into EcoRI-SalI digested pGBT9. Primers: GB-Of-RHD-F and GB-Of-RHD-R. PCR- amplified from pUC57-Of-NF- κ B.
pGB-Of-Cterm	EcoRI-SalI digested Of-Cterm PCR product containing codons 389-920 was subcloned into EcoRI-SalI digested pGBT9. Primers: GB-Of-Cterm-F and GB-Of-Cterm-R. PCR-amplified from pUC57-Of-NF- κ B.
pGEX-KG	Expression plasmid containing a 5' GST tag.
pGEX-KG-Of-NF- κ B-Cterm	EcoRI-XhoI fragment containing Of-NF- κ B codons 843-874 was subcloned into EcoRI-XhoI digested pGEX-KG. Primers Of-NF- κ B-F-C-term and Of-NF- κ B-R-C-term were used to PCR-amplify the fragment from pcDNA-FLAG-Of-NF- κ B.
pGEX-KG-Of-NF- κ B-Cterm-3X-Ala	EcoRI-XhoI fragment containing Of-NF- κ B-3X-Ala codons 843-874 was subcloned into EcoRI-XhoI digested pGEX-KG. Primers Of-NF- κ B-F-C-term and Of-NF- κ B-R-C-term were used to PCR-amplify the fragment from pcDNA-FLAG-Of-NF- κ B-3X-Ala.
pGEX-KG-Of-TIR	EcoRI-XhoI fragment containing the predicted intermembrane and TIR domain of Of-TLR were subcloned into EcoRI-XhoI digested pGEX-KG. Primers Of-TLR-F-TIR and Of-TLR-R-TIR were used to PCR-amplify the fragment from pUC57-Of-TLR.

Table 2.3 Vertebrate cell lines used in this thesis

Cell line	Description/cell type	Culture media
A293	Human embryonic kidney	DMEM-10
A293T	Human embryonic kidney expressing the SV40 large T-antigen	DMEM-10
DF-1	Immortalized chicken fibroblast cell line	DMEM-10

Table 2.4 Antibodies used in this thesis

Name ¹	Protocol ²	Host ³	Dilution	Source ⁴	Blocking Buffer ⁵	Duration ⁶
Primary Antisera						
FLAG	WB	R	1:1000	CST #2368	TBS, 5% milk, 0.1% Tw	Overnight 4°C
FLAG	IF	R	1:50	CST #2368	PBS, 3% calf serum	1 h 37°C
HA (DF-1 cells only)	IF	R	1:50	Abcam #ab9110	PBS, 3% calf serum	1 h 37°C
Ap-NF-κB	WB	R	1:1000	Pierce (Custom)	TBS, 5% milk, 0.1% Tw	Overnight 4°C
Ap-NF-κB (Sponge tissue)	IF	R	1:50	Peirce (Custom)	PBS, 3% calf serum	1 h 37°C
MYC-Tag	WB	M	1:1000	CST #2276	TBS, 5% milk, 0.1% Tw	Overnight 4°C
Secondary Antisera						
488-α-R	IF	G	1:160	TF #A11008	PBS, 5% NGS, 1% BSA, 0.2% Tx	1.5 h 37°C
488-α-R (DF-1 cells only)	IF	G	1:80	TF #A11008	PBS, 3% calf serum	1 h 37°C
HRP-α-R	WB	G	1:4000	CST #7074S	TBS, 5% milk, 0.1% Tw	1 h RT
HRP-α-M	WB	H	1:3000	CST #7076S	TBS, 5% milk, 0.1% Tw	1 h RT

Table is adapted from Wolenski (2012).

¹Antisera conjugations: FITC, fluorescein isothiocyanate; 488, Alexa fluor 488; TR, Texas red; HRP, horseradish peroxidase.

²Protocols and respective blocking buffers and antibody dilutions: WB, Western blot; IF, Immunofluorescence.

³Antibody host animal: G, goat; GP, guinea pig; H, horse; M, mouse; R, Rabbit.

⁴Antibody source: CST, Cell Signaling Technology; TF, ThermoFisher; VL, Vector Labs.

Corresponding catalog number is listed below company number, except for custom antibodies Ap-NF- κ B (Mansfield et al., 2017).

⁵Blocking buffer used to block nitrocellulose membrane or cells/tissues and dilute primary and secondary antisera. TBS, Tris-buffered saline; PBS, phosphate-buffered saline; BSA, bovine serum albumen; NGS, normal goat serum; Tw, Tween-20; Tx, Triton-X-100. All concentrations are measured in v/v except for milk and BSA (w/v).

⁶Duration of primary antisera incubation: h, hours; RT, room temperature. Blocking was performed overnight at 4°C on a nutator.

Table 2.5 TLR TIR Domain sequences for Chapter 3 phylogenetic analysis.

Taxa: protein	Full-length TLR sequence with TIR domain in bold
<i>Homo sapiens: TLR 1</i>	MTSIFHFAIIFMLILQIRIQLSEESEFLVDRSKNGLIHVPKDLSQK TTILNISQNYISELWTSILSLSKLRILIISHNRIQYLDISVFKFNQ ELEYLDLSHNKLVKISCHPTVNLKHLDLSFNAFDALPICKEFGN MSQLKFLGLSTTHLEKSSVLPPIAHLNISKVLLVLGETYGEKEDP EGLQDFNTESLHIVFPTNKEFHFILDVSVKTVANLELSNIKCVLE DNKCSYFLSILAKLQTNPKLSNLTNNIETTWSNFIRILQLVWH TTVWYFSISNVKLQGGQLDFRDFDYSGTSLKALSIHQVVSVDVFG FPQSYIYEIFSNNMIKNFTVSGTRMVHMLCPSKISPFLHLDIFSNN LLTDTVFENCGLTELETLILQMNQLKELSKIAEMTTQMKSLQ QLDISQNSVSYDEKKGDCSWTKSLLSLNMSSNILTDTIFRCLPP RIKVLDLHSNKIKSIPKQVVKLEALQELNVAFNSLTDLPGCGSF SSLSVLIIDHNSVSHPSADFFQSCQKMRSIKAGDNPFOCTCELG EFVKNIDQVSSEVLEGWPDYSYKCDYPESYRGTLTKDFHMSLS CNITLLIVTIVATMLVLAVTVTSLCSYLDLPWYLRMVCQWTQT RRRARNIPLEELQRNLQFHAFISYSGHDSFWVKNEILLPNLEK EGMQICLHERNFVPGKSIVENIITCIEKSYKSIFVLSPNFVQS EWCHYELYFAHHNLFHEGSNSLILILLEPIPQYSIPSSYHKL KSLMARRTYLEWPKEKSKRGLFWANLRAAINIKLTEQAKK
<i>Homo sapiens: TLR 2</i>	MPHTLWMVWVLGVIIISLSKEESSNQASLSCDRNGICKGSSGSL NSIPSGLTEAVKSLDLSNNRITYISNSDLQRCVNLQALVLTNGI NTIEEDSFSSLGSLEHLDLSYNYLSNLSSSWFKPLSSLTFLNLLG NPYKTLGETSLFSLTKLQILRVGNMDTFTKIQRKDFAGLTFLE ELEIDASDLQSYEPKSLKSIQNVSHLILHMKQHILLLEIFVDVTS SVECLELRDTDLDTFHFSELSTGETNSLIKKFTFRNVKITDESLE QVMKLLNQISGLLELEFDDCTLNGVGNFRASDNDRVIDPGKVE TLTIRRLHIPRFYLFYDLSTLYSLTERVKRITVENSKVFLVPCLLS QHLKSLEYLDLSENLMVEEYLKNSACEDAWPSLQTLILRQNLH ASLEKTGETLLTLKNLTNIDISKNSFHSMPETCQWPEKMKYLN LSSTRIHSVTGCIPKTLEILDVSNNNLNLFSNLNPQLKELYISRN KLMTLPDASLLPMLLVLKISRNAITTSKEQLDSFHTLKTLEAG GNNFICSCEFLSFTQEQQALAKVLIDWPANYLCDSPSHVRGQQ VQDVRLSVSECHRTALVSGMCCALFLILLTGVLCHRFHGLWY MKMMWAWLQAKRKPRKAPSRNICYDAFVSYSERDAYWVEN LMVQELENFNPPFKLCLHHRDFIPGKWIIDNIIDSIEKSHKT VFVLSSENFVKSEWCKYELDFSHFRLFDENNDAAAILILLEPIE KKAIPQRFCKLRKIMNTKTYLEWPMDEAQREGFWVNLRA AIKS

<i>Homo sapiens: TLR 3</i>	MRQTLPCIYFWGGLLPFGMLCASSTTKCTVSHEVADCSHLKLT QVPDDLPTNITVLNLTHNQLRRLPAANFTRYSQLTSLDVGFNTI SKLEPELCQKLPMLKVLNLQHNELSQLSDKTFAFCTNLTELHL MSNSIQKIKNNPFVKQKNLITLDLSHNGLSSTKLGTVQVLENL QELLSNNKIQALKSEELDIFANSSLKKLELSSNQIKEFSPGCFH AIGRLFGLFLNNVQLGPSLTEKLCLELANTSIRNLSLSNSQLSTT SNTTFLGLKWTNLTMLDLSYNNLNVVGNDFAWLPQLEYFFL EYNNIQHLFSLHGLFNVRYNLKRSTKQSSISLASLPKIDDFS FQWLKCLEHLNMEDNDIPGIKSNMFTGLINLKYLSLSNSFTSLR TLTNETFVSLAHSPLHILNLTKNKISKIESDAFSWLGHLEVLDLG LNEIGQELTGQEWRGLENIFEIYLSYNKYQLTRNSFALVPSLQ RLMLRRVALKNVDSSPSPFQPLRNLTILDLSNNNIANINDDMLE GLEKLEILDQLHNNLARLWKHANPGGPYFLKGLSHLHILNLE SNGFDEIPVEVFKDLFELKIIDLGLNNLNTLPASVFNNQVSLKSL NLQKNLITSVEKKVFGPAFRNLTELDMRFPDCTCESIAWFVN WINETHTNPELSSHLYCNTPPHYHGFPVRLFDTSCKDSAPFE LFFMINTSILLIFIFIVLLIHFEGWRISFYWNVSVHRVLGFEIDR QTEQFEYAAYIIHAYKDKDWVWEHFSSMEKEDQSLKFCLEER DFEAGVFELEAIVNSIKRSRKIIFVITHHLLKDPLCKRFKVH HAVQQAIEQNLDSIILVFLEEIPDYKLNHALCLRRGMFKSH CILNWPVQKERIGAFRHKLQVALGSKNSVH
<i>Homo sapiens: TLR 4</i>	MMSASRLAGTLIPAMAFSLSCVRPESWEPCVEVVPNITYQCMEL NFYKIPDNLPFSTKNLDLSFNPLRHLGYSFFSFPELQVLDLRC EIQTIEDGAYQSLSHLSTLITGNPIQSLALGAFSGLSSLQKLAV ETNLASLENFPIGHLKTLKELNVAHNLIQSFKLPEYFSNLTNLE HLDLSSNKIQSIYCTDLRVLHQMPLLNLSLDLSLNPMNFIQPGA FKEIRLHKLTLRNNFDSLNVMKTCIQGLAGLEVHRLVLGEFRN EGNLEKFDKSALEGLCNLTIEEFRLAYLDYYLDIIDLFCNCLTN VSSFSLVSVTIERVKDFSYNFGWQHLELVNCKFGQFPTLKLKSL KRLTFTSNKGGNAFSEVDLPSLEFLDLNRNGLSFKGCCSQSDFG TTSLKYLDLSFNNGVITMSSNFLGLEQLEHLDFQHSNLKQMSEF SVFLSLRNLIYLDISHTHTRVAFNGIFNGLSSLEVLMAGNSFQE NFLPDIFTELRLNLTFLDLSQCQLEQLSPTAFNSLSSLQVLNMSH NNFFSLDTFPYKCLNSLQVLDYSLNHIMTSKKQELQHFPSSLAF LNLTONDFACTCEHQSFLQWIKDQRQLLEVERMECATPSDKQ GMPVLSLNTCQMNKTIIGVSVLSVLVVSVAVLVYKFYFHLM LLAGCIKYGRGENIYDAFVIYSSQDEDWVRNELVKNLEEGV PPFQLCLHYRDFIPGVAIAANIIHEGFHKSARKVIVVVSQHFI QSRWCIFEYEIAQTWQFLSSRAGIIFIVLQKVEKTLLRQQV ELYRLLSRNTYLEWEDSVLGRHIFWRRLRKALLDGKSWNP EGTVGTGCNWQEATSI

<i>Homo sapiens: TLR 5</i>	MGDHLDLLLGVVLMAGPVFGIPSCSFDGRIAFYRFCNLTQVPQVLN TTERLLLSFN YIRTVTASSFPFLEQLQLLELGSQYTPLTIDKEAFRNL PNLRILD LGSSKIYFLHPDAFQGLFHLFELRLYFCGLSDAVLKDG YF RNLKALTRLDLSKNQIRSLYLHPSFGKLNSLKSIDFSSNQIFLVCEHE LEPLQGKTL SFFSLAANSLSRVSDWVGKCMNPFNMVLEILDVSG NGWTVDITGNFSNAISKSQAFSLILAHHIMGAGFGFHNKDPDQNT FAGLARSSVRHLDLSHG FVFSLSNSRVFETLKD LKVLNLAYNKINKI ADEAFYGLDNLQVLNLSYNLLGELYSSNFYGLPKVAYIDLQKNHIA IIQDQTFKFLEKLQTLDLRDNALTTIHFIPIPDIFLSGNKLVTLPKINL TANLIHLSENRLENLDILYFLLRVPHLQILILNQNRFS SSCSGDQTPSE NPSLEQLFLGENMLQLAWETELCWDVFEGLSHLQVLYLNHN YLNS LPPGVFSHLTALRGLSLNSNRLTVLSHNDLPANLEILDISR NQLLAPN PDV FVSLSVLDITHNKFICECELSTFINWLNHTNVTIAGPPADIYCVY PDSFSGVSLFSLSTEGCDEEEVLKSLKFSLFIVCTVTTLTFLMTILTV TKFRGFCFICYKTAQRLVFKDHPQGTEPD MYKYDAYLCFSSKDFT WVQNALLKHLDTQYSDQNRFNLCFEERDFVPGENRIANIQDAI WNSRKIVCLVSRHFLRDGWCLEAFSYAQGRCLSDLNSALIMVV VGSLSQYQLMKHQ SIRGFVQKQQYL RWPEDFQDVGWFLHKL SQQILKKEKEKKKDNNIPLQTVATIS
<i>Homo sapiens: TLR 6</i>	MTKDKEPIVKS FHFVCLMIIIVGTRI QFSDGNEFAVDKSKRGLIHVPK DLPLKTKVLDMSQNYIAELQVSDMSFLSELT VLRLSHNRIQLLDLS VFKFNQDLEYLDLSHNQLQKISCHPIVSFRHLDLSFNDFKALPICKE FGNLSQLNFLGLS AMKLQKLDLLPIAHLHLSYILLDLRNYYIKENE TESLQILNAKTLHLVFHPTSLFAIQVNISVNTLGCLQLTNIKLNDDNC QVFIKFLSELTRGSTLLNFTLNHIETTWKCLVRVFQFLWPKPVEYLN IYNLTIIESIREEDFTYSKTTLKALTIEHITNQVFLFSQTALYTVFSEM NIMMLTISDTPFIHMLCPHAPSTFKFLNFTQNVFTDSIFEK CSTLVKL ETLILQKNGLKDLFKVGLMTKDMPSLEILDVSWNSLESGRHKENC TWVESIVVLNLSSNMLTDSVFRCLPPRIKVLDLHSN KIKSVPKQVV KLEALQELNVAFNSLTDLP GCGSFSSLSVLIIDHNSVSHPSADFFQSC QKMRSIKAGDNP FQCTCELREFVK NIDQVSSEVLEGWPD SYKCDY PESYRGSP LKDFHMSELSCNITLLIVTIGATMLVLAVTVTSLCIYLDL PWYLRMVCQWTQTRRRARNIPLEELQRNLQFHAFISYSEHDSAW VKSELVPYLEKEDIQICLHERNFVPGKSIVENIINCIEKSYKSIFV LSPNFVQSEWCHYELYFAHHNLFHEGSNNLILILLEPIPQNSIPN KYHKLKALMTQRTYLQWPKEKSKRGLFWANIRAAFNMKLTL VTENN DVKS

<i>Homo sapiens: TLR 7</i>	MVFPMWTLKRQILILFNIILISKLLGARWFPKTLPCDVTLDVPKNHVIV DCTDKHLTEIPGGIPTNTTNLTINHIPDISPASFHRLDHLVEIDFRNC VPIPLGSKNNMCIKRLQIKPRSFSGLTYLKSLYLDGNQELLEIPQGLPSSL QLLSLEANNIFSIRKENLTELANIEILYLQNCYRNPCYVSYSIEKDAF LNLTKLKVLSKDNNTAVPTVLPSTLTELYLYNNMIAKIQEDDFNNL NQLQILDLSGNCPRCYNAPFPCAPCKNNSPLQIPVNAFDALTELKVLRL HSNSLQHVPFRWFKNINKLQELDLSQNFLAKEIGDAKFLHFLPSLIQLD LSFNELQVYRASMNLSQAFSSLKSLKILRIRGYVFELKSFNLSPLHN LQNLEVLDLGTNFIKIANLSMFKQFKRLKVIDLSVNKISPSGDSSEVGF CSNARTSVESYEPQVLEQLHYFRYDKYARSCRFKNKEASFMSVNESE YKYGQTLDSLKNSIFFVKSSDFQHLKCLNLSGNLISQTLNGSEFQP LAELRYLDFSNNRLDLLHSTAFEELHKLEVLDISSNSHYFQSEGITHML NFTKNLKVQLKMMNDNDISSSTRTESESLRTLEFRGNHLDVLWR EGDNRYLQLFKNLLKLEELDISKNSLSFLPSGVFDGMPNPNLKNLSLAK NGLKSFSWKKLQCLKNLETLDLSHNQLTTVPERLSNCSRLKNLILKN NQIRSLTKYFLQDAFQLRYLDLSSNKIQMIQKTSFPENVLNNLKMLLL HHNRFLCTCDVWFWVWNHTEVTIPYLATDVTCVGPAGHKQSVI SLDLYTCELDLTNLILFSLSSVSLFLMVMMTASHLYFWDVWYIYHFC KAKIKGYQRLSPDCCYDAFIVYDTKDPVTEWVLAELVAKLEDPRE KHFNLCEERDWLPGQPVENLSQSIQLSKKTVMVMTDKYAKTE NFKIAFYLSHQRLMDEKVDVILIFLEKPFQKSKFLQLRKRLCGSS VLEWPTNPQAHFYFWQCLKNALATDNHVAYSQVFKETV
<i>Homo sapiens: TLR 8</i>	MENMFLQSSMLTCIFLLISGSCELCAEENFSRSYPCEKKQNDSVIAEC SNRRLQEVQTVGKYVTELDLSDNFITHITNESFQGLQNLTKINLHNPN NVQHONGNPGIQSNGLNITDGAFLNLKNLRELLLEDNQLPQIPSGLPES LTELSLIQNNIYNITKEGISRLINLKNLYLAWNCYFNKVCEKTNIEDGVF ETLTNLELLSLSFNSLSHVPPKLPSSLRKLFLSNTQIKYISEEDFKGLINL TLLDLSGNCPRCFNAPFPCVPCDGGASINIDRFAFQNLTLRLYNLSSTS LRKINAAWFKNMPHLKVLDLEFNLYLVEIASGAFLTMLPRLEILDLSF NYIKGSYPQHINISRNFSKLLSLRALHLRGYVFQELREDDFQPLMQLPN LSTINLGINFIKQIDFKLFQNFNLEIYYLSENRIPLVKDTRQSYANSSSF QRHIRKRRSTDFEFDPHSNFYHFTRPLIKPQCAAYGKALDLSLNSIFFIG PNQFENLPDIACNLANSNAQVLSGTEFSAIPHVKYLDLTNNRLDFD NASALTELSDEVLDSLNSHYFRIAGVTHHLEFIQNFTNLKVLNLSH NNIYTLTDKYNLESKSLVELVFSGNRLDILWNDDDNRYISIFKGLKNLT RLDLSLNRKHIPNEAFLNLPASLTEHINDNMLKFFNWTLTLLQGFPRLE LLDLRGNKLLFLTDSLSDFTSSLRTLLSHNRISHLPSGFLSEVSSLKHL DLSSNLLKTINKSALETKTTTKLSMLELHGPNPECTCDIGDFRRWMDE HLNVKIPRLVDVICASPGDQRGKSIVSLELTTCVSDVTAVILFFFTFFITT MVMLAALAHHLFYWDVWFIYNVCLAKVKGYRSLSTSQTIFYDAYISY DTKDASVTDWVINELRYHLEESRDKNVLLCLEERDWDPLAIDN LMQSINQSKKTVMVLTKKYAKSWNFKTAFYLALQRLMDENMDVII FILLEPVLQHSQYLRLRQRICKSSILQWPDNPKAEGFLFWQTLRNV VLTENDSRYNMYVDSIKQY

<i>Homo sapiens</i> : TLR 9	MGFCRSALHPLSLLVQAIMLAMTLALGTLPAFLPCELQPHGLVNCNW LFLKSVPHFMSMAAPRGVNTSLSLSSNRIHHLHDSDFAHLPRLHLNLK WNCPPVGLSPMHFPCMTIEPSTFLAVPTLEELNLSYNNIMTVPALPKS LISLSLSHTNMLMLDSASLAGLHALRFLFMDGNCYYKNPCRQALEVAP GALLGLGNLTHLSLKYNNTLVVPRNLPSSLEYLLLSYNRIVKLAPEDL ANLTALRVLDVGGNCRRCDHAPNPCMECPRHFPQLHPDTFSHLSRLEG LVLKDSSLSWLNASWFRGLGNLRVLDLSENFYKCIKTAKAFQGLTQL RKLNLSTFYQKRVSAHLSLAPSFGLVALKELDMHGIFRSLDETTLR PLARLPMLQTLRLQMNFINQAQLGIFRAFPGLRYVDLSDNRISGASELT ATMGEADGGEKVWLQPGDLAPAPVDTPSSEDFRPNCTLNFTLDLSR NNLVTVQPEMFAQLSHLQCLRLSHNCISQAVNGSQFLPTGLQVLDLS HNKLDLYHEHSFTELPRLEALDLSYNSQPFMGQGVGHNFSAHLRTL RHLSLAHNNIHSQVSQQLCSTSLRALDFSGNALGHMWAEGDLYLHFF QGLSGLIWLDSLQNRLHTLLPQTLRNLPKSLQVLRRLRDNYLAFFKWW SLHFLPKLEVLDLAGNQLKALTNGSLPAGTRLRRLDVSCNSISFVAPGF FSKAKELRELNLSANALKTVDHWSFGPLASALQILDVSANPLHCACG AAFMDFLLEVQAAVPGLPSPRVKCGSPGQLQGLSIFAQDLRLCLDEALS WDCFALLAVALGLGVPMHLHLCGWDLWYCFHLCLAWLPWRGRQ SGRDEDALPYDAFVVFDKTQSAVADWVYNELRGQLEECRGRWAL RLCLEERDWLPGKTLFENLWASVYGSRKTLFVLAHTDRVSGLLR ASFLLAQQRLLEDKDVVVLVILSPDGRRSRYVRLRQRLCRQSVL LWPHQPSGQRSFWAQLGMALTRDNHHFYNRNFCQGPTAE
<i>Homo sapiens</i> : TLR 10	MRLIRNIYIFCSIVMTAEGDAPELPEERELMTNCSNMSLRKVPADLTPA TTTTLDLSYNLLFQLQSSDFHSVSKLRVLILCHNRIQQDLKTFEFNKEL RYLDLSNNRLKSVTWYLLAGLRYLDLSFNDFDTMPICEEAGNMSHLEI LGLSGAKIQKSDFQKIAHLHLNTVFLGFRTLPHYEEGSLPILNTTKLHIV LPMDTNFWVLLRDGIKTSKILEMTNIDGKSQFVSYEMQRNLSLENAK TSVLLLNKVDLLWDDLFLILQFVWHTSVEHFQIRNVTFGGKAYLDHN SFDYSNTVMRTIKLEHVHFRVFIYQQDKIYLLLTCKMDIENLTISNAQMP HMLFPNYPTKFQYLNFAANNLTDELFKRTIQLPHLKTILNGNKLETLS LVSCFANNTPLEHLDLSQNLLQHKNDENCWPETVVNMNLSYNKLS SVFRCLPKSIQILDNNNQIQTVPKETIHLMALRELNIAFNFLTDLPGCS HFSRLSVLNIEMNFILSPSLDFVQSCQEVKTLNAGRNPFRCTCELKNFI QLETYSEVMMVGWSDSYTCEYPLNLRGTRLKDVHLHELSCNTALLIV TIVVIMLVGLAVAFCCCLHFDLPWYLRMLGQCTQTWHRVRKTTQEQL KRNVRFHAFISYSEHDSLWVKNELIPNLEKEDGSILICLYESYFDPG KSISENIVSFIEKSYKSIFVLSPNFVQNEWCHYEFYFAHNLFHENS DHIILILLEPIPFYCIPTRYHKLKALLEKKAYLEWPKDRRKCGLFW ANLRAAINVNVLATREMYELQTFTELNEESRGSTISLMRTDCL

<i>Nematostella vectensis:</i> TLR	MKGSILRQVTQCFCNAWVSVLFLRALLIAGAPSRDENCKKECANLRM PSTFRSPTGQIPMFFTCKCEIKGGAECSDLGPWIQPLANSSTVQYYLAI VCRSSTRIIFCNSPEVKRKNVILFYQMAGPCSVTVHDVSVLGNATDYR VQLFTHGAELLYADTESITGLRNIGTFSLQSSGTGIPRILTGFEPWMAE VLLSNLSITEIPEQFKTAMPRLQALDLNNNSLTRPPDFPWSHKPLSLPR NLSRLPVFNHHYQEGSVVQPRLYRRFLVLDYNQIRNLSQYPFTGHLQK LSIKGNGLRVIGGSCFSNLSGVNIIDLSNNEIRDFPEQLFRGQGSMLER FNHNFLSTLPNRVFTDMKRLKRLYLNNNRLQRLQAGLLYGNEEITLT LNDNDLTEIENNALPENSNTLKTTLQNRNRLTRVPRAVFLLRNLESADL SSNAITFGGILDVLDVSTADQLFYNLRRSASSSDNQLKSTKVELNLAN NGISSIDIGSLNKTQLGKLKVLRVYHIDLRDNPLICNCKLTALFRLLKR LTADYDPDVTHAQFDSWICSQPTRLRNVALLRVPENQFQCIMDLENCPR ECTCAVREIDQTVLVDCSERGLHRLPFKMPAGELEVNLRGNAIRELPW RHYLGNITVLELSNNEIKELNMTFVDSLARVVNLAINDNKLKYLPRGV TNLTAREGFRSLSISHNFFVDCYASWMRDWLANNTDKIEDTSSILCAS GRLEGLPIISVPLSDFNC SAYRPPVPGPITDGLSLLLAIVLAVLLVLSVVA FVMTYCFRWEMKILMYTHFNWHPFDRVDDTDVSKIYDTFISYSSQDA SWVRET QRTLESHVPPYRLCIHDRD FEIGASIH NILNSVRLSKR MIMVLSNHFIASEWCRLEFRAAHQKVLEDRTNYLIHILFDDVDPST LDDETKLYLRTNTYLSVSNKFWFWQKLFYALPKPLAPPQSYEGHVE MSKV
<i>Amphimedon</i> <i>queenslandica:</i> TLR	MFLLFALFCYQLLAKTNSQSYCTQSGVDQTFICSREINQSLTIPCGSSGD PTRNFTYTIIRPFEGSIETIQTVSFLTNLQTRDQMASISCTIDGRIEANY KINVTIPPYSIKFGSSPDVYVSGPIRLSCPIEGNPRPHFVWHRYNDIDRS KELSLPTELEFDVNSLNKTWTIQNWQEYMNGFYTCCAENYLGQHCYI NGATFRLFAKICDGSTDTDGYIKTASNELTYPIKSTISITCIAFNAVHNQP PNWFFKSFYETRPTRIPTLKPGPSYSIRYRYDDRKCLWNSTLIHSNFSVS LAGSYYSYGLNKVHSLDIDLQACHKLSSENISLTVSPTQRAVGENGSL LLVLSCMVNSTPHDAPPVWCKLNDNEDCEIINSTSNGRKNCWSISTVE ILVSDWTIGQYQCSHLDVVRQVEVTATKRKSPSSSLAIGAVSLIAAIIIVI IMTAVTALLRCYKATRYARTKEPPYEGF QWDAFLSYHTEAE EEFVIRHI KQVLEERGYKIFWHHSDFIGGKTITDNINDAVNESRKVIFVFTRNF VNSEYCM MELHSTLDRLQRTTRCMIPIALEDET TVPIELKSTVTY WPVLTEEE MNTEKLICLIGSSIRNKLQ EEKK
<i>Drosophila melanogaster:</i> Toll	MSRLKAASELALLVILQLLQWPGSEASFGRDACSEMSIDGLCQCAPIM SEYEIICPANAENPTFRLTIQPKDYVQIMCNLTDTTDDYQQLPKKLRIGEV DRVQMRRRCMLPGHTPIASILDYLGIVSPTTLIFESDNLGMNITRQHLDRL LHGLKRFRFTTRRLTHIPANLLTDMRNLSHLELRANIEEMP SHLFDDLE NLESIEFGSNKLRQMPRGIFGKMPKLKQLNLWSNQLHNLTKHDFEGAT SVLGIDIHDNGIEQLPHDVFAHLTNVTDINLSANLFRSLPQGLFDHNKH LNEVRLMNNRVPLATLPSRLFANQPELQILRLRAELQSLPGDLFEHSTQ ITNISLGDNLLKTLPATLLEHQVNLLSLDLSNNRLTHLPDSLFAHTTNTLT DLRLEDNLLTGISGDIFSNLGNLVTLMVSRNRLRTIDSRAFVSTNGLRH LHLDHNDIDLQOPLLDIMLQTOI

<i>Drosophila melanogaster</i> : Toll (con't)	NSPFGYMHGLLTNLNRNSIIFVYNDWKNTMLQLRELDLSYNNISLG YEDLAFLSQNRLHVNMTNTHNKIRRIALPEDVHLGEGYNNNLVHVDLND NPLVCDCTILWFIQLVRGVHKPQYSRQFKLRDRLVCSQPNVLEGTPV RQIEPQTLICPLDFSDDPREKRCPRGCNCHVRTYDKALVINCHSGNLTH VPRLPNLHKNMQLMELHLENNTLLRLPSANTPGYESVTSHLAGNNL TSIDVDQLPTNLTHLDSWNHLQMLNATVLGFLNRTMKWRSVKLSGN PWMCDCTAKPLLLFTQDNFERIGDRNEMMCVNAEMPTRMVELSTND ICPAEKGVFIALAVVIALTGLLAGFTAALYYKFQTEIKIWLYAHNLLLWF VTEEDLDKDK KKFD AFISY SHKDQSFIEDYLVPQLEHGPQKFQLCVH ERDWLVGGHIPENIMRSVADSRRTIIVLSQNF IKSEWARLEFRAAH RSALNEGRSRIIVIIYSDIGDVEKLDEELKAYLKMNTY LKWGDPWF WDKLR FALPHRRPVGNIGNGALIKTALKGSTDDKLELIKPSPTPLT TPPAEATKNPLVAQLNGVTPHQAIMIANGKNGLTNLYTPNGKSHGNH INGAFIINTNAKQSDV
<i>Orbicella faveolata</i> : TLR	MFVKSQTALIQSNLHILVLDQRHTVLLATSLVFDTKGSLIIFSCLVLFAIE ATGQQRRRFLHCGASFCLLKRSMSFFETGNGTLIKEQAVICVLKSDK VRDRCVVDISLILKTITTPQDVVLHFAAVCLTPMEIAFYNSLNATKKNA IFYLQIKGHCSFSADGISHWGKATDFRVFYLMENSTLLEGNKTLANN RRALENIGTLMIDKSKLRTLPKMFSSTKVWPRMAEVVFSKLQLTSIPPE LNTTMPFLQSLELANNKLTTPPPFPWCKATLKLPRGLQRTPTGNHHYQ FGTNVRPNYRRFLDLSYNNIEDLSTHNFRGFLNKLTLLEGNGLKVIGTS CFRNLKGIHVISLSKNKLKSLPSELFQGGQDSLLELRDHNHNSIIPNDLF KTVTQIKRIDLHSNKLSCIPQELFSKLKNIKILHLEDNHITQVHDKAFSI DSSSLQNIYLQKNKISRIPLTLLQRHAVKIDLSFNQLTFQDFNRLIQEL DLETFLYHHRHTASSSQMRLQESLKSISFAHNKFTTINIEAFNRTEELTF EYLLRVYEIDMSGNLLLCDCKILLRSRWLRALVQRHTRIRNEQFQTWK CAAPTELKGPILSVDENRFKCQRNLENCPHGCLCLVRALDGTVIDC KGRNLTAIPPKVPSGRIELKLEDNNIREIPPYPYPMENVTALYLTHNKIQV LNKSTVRRFTRIKVLFIDSNKLTLYLPKNIENLNFTSLALHHNFFKCDCT TLWIKHWLQRKQSKILHIKNVLCNSEGSTQGKAIYTLNPEEFVCKKNK KDIPTTQSITKDKTFKIIALTGGALVLTFFIAFIVAYKYRGEMKVLMYTH FNWHPFDRVDDSDPRKIYDAFVSYSGSDHQWVNTLQERLEHHD PYKLCIHHRDFVVGAPIQENILNSVDQSKRMLMVLSRNLKSEWC LLEFRAAHRKVLEDRMNYLIILFDGINMDELDD EMKLYMRTNTY LSVSYKWFWEKLYYAMPQSTDRQFRARDLSSISSNAGMQYSTETVF KNEAYLKVTDILE

Full-length amino acid sequences of TLR proteins from the organisms used for MEME analysis to identify the conserved motifs. TLR sequences from *Orbicella faveolata* was obtained from the NCBI database; *Amphimedon queenslandica* TLR was obtained from (Gauthier et al., 2010); and *Nematostella vectensis* *Drosophila*, and *Homo sapiens* TLRs were obtained from UniProt database. All sequences were truncated to their predicted (or known) TIR domains using MEME analysis and are shown in bold. The predicted TIR

domains were based on MEME predictions and by comparison to known *Homo sapiens* TIR domains. The truncated and culled TIR sequence dataset allowed Clustal Omega to successfully align TIR sequences to be input into PAUP* for neighbor-joining phylogenetic analysis that was bootstrapped 1000 times (Figure 3.3A).

Table 2.6 NF- κ B Rel Homology Domain sequences for Chapter 3 phylogenetic analysis.

Taxa: protein	Full-lenth NF- κ B with truncated RHD domain in bold
<i>Homo sapiens</i> : p100	MESCYNPGLDGIIEYDDFKLNSSIVEPKEPAPETADGPYLVIV EQPKQRGFRFRYGCEGPSHGGLPGASSEKGRKTYPTVKI CNYEGPAKIEVDLVTHSDPPRAHAHSLVGKQCSELGICAV SVGPKDMTAQFNNLGVLVHTKKNMGMGTMIQKLQRQRL RSRPQGLTEAEQRELEQEAKELKKVMDLSIVRLRFS AFL RASDGSFSLPKPVISQPIHDSKSPGASNLIKISRMDKTAGS VRGGDEVYLLCDKVQKDDIEVRFYEDDENGWQAFGDFS PTDVHKQYAIVFRTPPYHKMKIERPVTVFLQLKRKRGGD VSDSKQFTYYPLVEDKEEVQRKRRKALPTFSQPFGGGSH MGGGSGGAAGGYGGAGGGGSLGFFPSSLAYSPLYQSGAGPM GCYPGGGGGAQMAATVPSRDSGEEAAEPSAPSRTPOCEPQA PEMLQRAREYNARLFGLAQRSARALLDYGV TADARALLAG QRHLLTAQDENGDTPLHLAIIHGQTSVIEQIVYVIHHAQDLG VVNLTNHLHQTPHLAVITGQTSVVSFLLRVGADPALLDRHG DSAMHLALRAGAGAPELLRALLQSGAPAVPQLLHMPDFEGL YPVHLAVRARSPECLDLLVDSGAEVEATERQGGRTALHLATE MEELGLVTHLVTKLRANVNARTFAGNTPLHLAAGLGYPTLT RLLLKAGADIHAENEEPLCPLPSPPTSDSDSDSEGPEKDTRSS FRGHTPLDLTCSTKVKTLLL NAAQNTMEPPLTPPSPAGPGLSL GDTALQNLEQLLDGPEAQGSWAELAERLGLRSLVD TYRQTT SPSGSLLRSYELAGGDLA GLLEALSDMGLEEGVRLLRGPETR DKLPSTA EVKEDSAYGSQSVEQAEKLGPPPEPPGGLCHGHP QPQVH

<i>Homo sapiens</i> : p105	MAEDDPYLGRPEQMFHLDPSLTHITFNPEVFQPQMALPTDG PYLQILEQPKQRGFRFRYVCEGPSHGGLPGASSEKNKKS YPQVKICNYVGPAAKIVVQLVTNGKNIHLHAHSLVGKHCE DGICTVTAGPKDMVVGFANLGLHVTKKKVFETLEARM TEACIRGYNPGLLVHPDLAYLQAEGGGDRQLGDREKELI RQAALQQTKEMDLSVVRLMFTAFLPDSTGSFTRRLEPVV SDAIYDSKAPNASNLKIVRMDRTAGCVTGGEIYLLCDKV QKDDIQIRFYEEEEENGWEGFGDFSPDVRHQFAIVFK TPKYKDINITKPASVVFQLRRKSDLETSEPKPFLYYPEIKD KEEVQRKRQKLMPNFSDSFSGGSGAGAGGGGMFGSGGG GGGTGSTGPGYSFPHYGFPTYGGITFHPGTTKSNAGMKHGT MDTESKKDPEGCDKSDDKNTVNLFGKVIETTEQDQEPSEAT VGNGEVTLYATGTKEESAGVQDNLFLEKAMQLAKRHANA LFDYAVTGDVKMLLAVQRHLTAVQDENGDSVLHLAIIHLHSQ LVRDLLEVTSGLISDDIINMRNDLYQTPLHLAVITKQEDVVED LLRAGADLSLLDRLGNSVLHLAAKEGHDKVLSSILLKHKHAA LLLDHPNGDGLNAIHLAMMSNSLPCLLLLVAAGADVNAQEQ KSGRTALHLAVEHDNISLAGCLLLEGDAHVDSTTYDGTTPH IAAGRGSTRLAALLKAAGADPLVENFEPLYDLDDSWENAGE DEGVVPGTTPLDMATSWQVFDILNGKPYEFTSDDLLAQG DMKQLAEDVKLQLYKLEIPDPDKNWATLAQKLGLGILNNA FRLSPAPSKTMDNYEVSGGTVRELVEALRQMGYTEAIEVIQ AASSPVKTTSQAHSPLSPASTRQQIDELRDSVCDSGVETS FRKLSFTESLTSGASLLTLNKMPHDYGQEGPLEGKI
<i>Homo sapiens</i> : RelA	MDELFPLIFPAEPAQASGPYVEHIEQPKQRMFRYKCEGR SAGSIPGERSTDTTKTHPTIKINGYTGPQTVRISLVTKDPP HRPHPHELVGKDCRDGFYEAELCPDRCIHSFQNLGIQCV KKRDLEQAISQRIQTNNNPFQVPIEEQRGDYDLNAVRLCF QVTVRDPGRPLRLPPVLSHPFDNRAPNTAELKICRVNRN SGSCLGGDEIFLLCDKVQKEDIEVYFTGPGWEARGSFSA DVHRQVAIVFRTPPYADPSLQAPVRVSMQLRRPSDRELSE PMEFQYLPDTDDRHRIEEKRKRTYETFKSIMKKSPFSGPT DPRPPPRRIAVPSRSSASVPKPAPQYPFTSSLSTINYDEFPTMV FPSGQISQASALAPAPPQVLPQAPAPAPAPAMVSALAQAPAPV PVLAPGPPQAVAPPAPKPTQAGEGTLSEALLQLQFDDDLGA LLGNSTDPVFTDLASVDNSEFQQLNQGIPVAPHTTEPMLM EYPEAITRLVTGAQRPPDPAPAPLGAPGLPNGLLSGDEDFSSIA DMDFSALLSQISS

<i>Homo sapiens</i> : RelB	MLRSGPASGPSVPTGRAMPSRRVARPPAAPELGALGSPDLSSL SLAVSRSTDELEIIDEYIKENGFGLDGGQPGPGGLPRLVSRG AASLSTVTLGPVAPPATPPPWGCPGLRLVSPAPGPGPQPHLVI TEQPKQRGMRFRYECEGRSAGSILGESSTEASKTLPALIEL RDCGGLREVEVTACLWWDWPHRVHPSLVGKDCTDGI CRVRLRPHVSPRHSFNNLGIQCVRKKEIEAAIERKIQLGID PYNAGSLKNHQEVDNMNVVRICFQASYRDQQGQMRRMDP VLSEPVYDKKSTNTSELRICRINKESGPCTGGEELYLLCD KVQKEDISVVFSTRASWEGRADFSQADVHRQIAIVFKTPPY EDLEIVEPVTNVFLQRLTDGVCSEPLPFTYLPRDHDSYG VDKKRKRGMPPDVLGELNSSDPHGIESKRRKKKPAILDHFLP NHGSGPFLPPSALLPDPDFSGTVSLPGLEPPGGPDLLDDGFA YDPTAPTFTMLDLLPPAPPHASAVVCSGGAGAVVGETPGPEP LTLDSYQAPGPGDGGTASLVGSNMFPNHYREAAFGGGLSPG PEAT
<i>Homo sapiens</i> : c-Rel	MASGAYNPYIEIIEQPRQRGMRFYKCEGRSAGSIPGEHS TDNNRTYPSIQIMNYYGKGKVRITLVTKNDPYKPHPHDLV GKDCRDGYEAEFGQERRPLFFQNLGIRCVKKKEVKEAI ITRIKAGINPFNVPEKQLNDIEDCDLNVVRLCFQVFLPDEH GNLTTALPPVVSNNPIYDNRAPNTAELRICRVNKNCGSVRG GDEIFLLCDKVQKDDIEVRFVLNDWEAKGIFSQADVHRQ VAIVFKTPPYCKAITEPVTVMQLRRPSDQEVSESMDFRY LPDEKDTYGNKAKKQKTLLFQKLCQDHVETGFRHVDQD GLELLTSGDPPTLASQSAGITVNFPERPRPGLLSIGEGRYFKK EPNLFSDHAVVREMPGTGVSSQAESYYPSPGPISGLSHHASMA PLPSSSWSSVAHPTPRSGNTNPLSSFSTRTLPSNSQGIPFLRIP VGNDLNASNACIYNNADDIVGMEASSMPADLYGISDPNML SNCSVNMMTTSSDSMGETDNPRLLSMNLENPSCNSVLDPRD LRQLHQMSSSSMSAGANSNTTVFVSQSDAFEGSDFSCADNS MINESGPSNSTNPNSHGFVQDSQYSGIGSMQNEQLSDSFPYEF FQV
<i>Nematostella vectensis</i> : NF-κB	MAQSEQQVGSALTESMLNEIIRPGYLPDISALHVPLGTNAEEP SYTEPYLEILEQPKPRGFRFRYPSEGPSHGGLPGQFSTSKS KSYPVQVNNYQGPCRIVVTLVTKDEPYMLHAHSLTGKN ANEEGVVTVQVGPDQHMTASFPNLGIQHVTKKNVVKVL MDRFIKWQTLQNAFALKSEGIKDGVDLSLFGVNTAINSN KLGFDKNVALSVANQEAAKSREYAKQQAAMDL SAVRLC FQAYLPDQDGNFTRPLKPVYSDAVLDSKEPSASQLKICRM DKNSGCVTGGDEIYLLCDKVQKDDIEIHFYEMDDITGKY TWEDLGKFSPCDVHRQFAIVFKTPPYWNIAIERPANVLVE LRRKKNGETSEPQFTYQPQLFDKEAIGAKRRKTVPHF TEFLSGGSSGATGGGGSSVSGFNFSADFLQQGVFLTQNPSNM

<i>Drosophila melanogaster:</i> Relish	MNMNQYYDLNKGKVMFMNDASSTSGYSSSTSPNSTNRSFSPAHS SPKTMELQTDFAANLNLPGGNSPHQPPMANSFYQNQLLNNGGICQ LGATNLINSTGVSGVANVTSGFNMMDHQYFVPAPATVPPSQNF GYHQNGLASDGD~~IKHVPQLRIVEQPVEKFRFRYKSEMHGTHG~~ ~~SLNGANSKRTPKTFPEVTLCNYDGPVIRCSLFQTNLDSPHSH~~ ~~QLVVVKDDRDVCDPHDLHVS~~KERGYVAQFINMGIIHTAKKYIF EELCKKKQDRLVFQMNRRREL~~SHKQLQELHQETEREAKDMN~~ LNQVRLCFEAFKIEDNGAWVPLAPPVYSNAINNRKSAQTGELR IVRLSKPTGGVMGNDELILLVEKVSKKNIKVRFFEEDEGETVWE AYAKFRESDVHHQYAIVCQTPPYKDKDQVDRENVYIELIRPSDDE RSFPALPFRYKPRSVIVSRKRRRTGSSANSSSGTESSNNSLDLPKT LGLAQPPNGLPNLSQHDQTISEEFGREKHLNEFIASEDFRKLIEHNS SDLEKICQLDMGELQHDGHNRAEVP~~SHRNRTIKCLDDLFEIYKQD~~ RISPIKISHHKVEKWFIEHALNNYNRDTLLHEVISHKKDKLKLAIQ TIQVMNYFNLKDVVNSTLNADGDSALHVACQQDRAHYIRPLLGM GCNPNLKNNAGNTPLHVAVKEEHLSCVESFLNGVPTVQDLDSLNT DDGLTPLHMAIRQNKYDVAKKLISYDRTSISVANTMDGNNALHM AVLEQSVELLVLILDAQENLTDILQAQNAAGHTPLELAERKAND RVVQLLKNVYPEKGELAMTWIPCKVKEEIDSSSESSDAGQLEIK SEEMDIETKDEDSVELDLSSGPRRQKDESSRDTEMDNNKLQLLLK NKFIYDRLCSLLNQPLGHGSDPQDRKWMQLARQTHLKQFAFIWL GAEDLLDHVKRKGASVEFSTFARALQAVD PQAYALLVNPT
<i>Drosophila melanogaster:</i> Dorsal	MFPNQNNGAAPGQGPVVDGQQSLNYNGLPAQQQQQLAQSTKNV RKKPYVKITEQPAGKALRFRYECEGRSAGSIPGVNSTPENKTY PTIEIVGYKGRAVVVSCVTKDTPYRPHPHNLVGKEGCKKGV CTLEINSETMRAVFSNLGIQCVKKKDIEAALKAREEIRVDPFK TGFSHRFQPSIDLNSVRLCFQVFMESEQKGRFTSPLPPVVSEP IFDKKAMSDLVICRLCSCSATVFGNTQIILLCEKVAKEDISVRFF EEKNGQSVWEAFGDFQHTDVHKQTAITFKTPRYHTLDITEPA KVFIQLRRPSDGV~~TSEALPFEYVPMDSGKHTFWNLHRHLK~~RK PDEDLFQQILRLDAKREVQPPTIEVIDLDTPKIDVQREIPSEMEFNH EESQQSEPALEQE~~QSVQQEQYTQEQLQQEQYTQEQLQQEQYL~~ QQLEQQQSFQLEPMQQDQELPAQQSFDQAIDHLPDHTSDHIPED MEAADAHAEAEAHRLRSEQEKEIDTIIDEKVRELEQLDLGQQLEP RPLTANDKITEWMKSSEIEQQVHEPSPTAEADVLD~~SALEISKADKT~~ LDELLETVAELDEIYTD~~FKVQRD~~TYKNTIQNELAGLQGRAPLQVE DSFDDAATY~~TS~~LQIAFKNPVLIPMDDIMPPTPPMSQCAPEDA~~HQH~~ YDPVEVNSQARKPETPMRPVPPVPPAILTIQYPPEEDKLPLPPKRI RKQDSNAENRSIEANTVQTKPSTGESPLNKRLPPAPKNPNFNTLPR QKKPGFFSKLFSRRKSKPDLAQGQENSSILDSKANSREPSIGHFNM QDPMRASLRSSKSAAPFISNPAPAKSSPVKAKKPGSKLT~~KPVGRSV~~ SSVSGKRPAYLNADV~~VHIPLK~~GDSVNSLPQQQRTEGYQSSTISVG AGLDRRTASALQLADIPIS~~EGGMELVAIADRQSLHNLVSSIEGHFN~~ VQLDPNLDL~~TEAEHFALYTSIPPLAAASEFDETSAYYAPVDAGEILT~~ PDEVAKRLAAANGI

<i>Drosophila melanogaster</i> : Dif	MFEEAFGDIQEIINASMELNNGGATGGGSVAGAVGGGGAAHHI LSQSTSLPVMPSHIPLHLQNNMNQNLPEPSARSGPHLRIVE EPTSNIIRFRYKCEGRTAGSIPGMNSSSETGKTFTTIEVCNY DGPVHVVSCVTSDEPFRQHHPHWLVSKEEADACKSGIYQK KLPPEERRLVLQKVGIQCAKKLEMRDSLVERERRNIDPFN AKFDHKDQIDKINRYELRLCYQAFITVGNISKVPLDPIVSSP IYGKSELTTITRLCSCAATANGGDEIIMLCEKIAKDDIEVRF YETDKDGRETWFANAEFQPTDVFKQMAIAFKTPRYRNTE ITQSVNVELKLVRPSDGATSAPLPFEYYPNPELLTKHNRRV AQKTVESLKRSLMSTNLHPSKQVKTSSQYTIFSKPQIATTPQ TQVSPGMPLMFPGGSPNFVQDIKMENGFMVDVDSQSSQCPSV ERNFASPRSNCSSTVDSIPPMQMGQNQTHLYLPDATNFTFNGN FASPSSNCSTVDSIPPFQIGQRNNHMYLPENSFPVNGCSPTHF SGGSMTPINNNNNVLINNNNNDFLSQKMSAISIPPQGNFGIKQ VYQQTQQFLPQLQPESIPYLAQSHPEQSQYQQQQQPQEQQPP ADEPTQSFSDLISSIGMAPIDTSELIQDIEAELNS LGIQPFK
<i>Actinia tenebrosa</i> : NF- κ B	MIATMTQSEQQVGSTLTESMLHEIMTPGFLPDISALNVQMKG TYDGPYLEILEQPKSRGFRFRYPCEGPHGGLPGEFSSQK TKSYPSVQLNNYRGPCRIVVSLVTEEEPYPMPHAHSLTGKN ASKEGIVTVQIGPEQGMASAFPNLGIQHVTRKAVAKTLLE RYTKMQELQSLTMSVMSTTHNDPASLFGISGLGAQHAAG TDNKPFDKNLALSVADEEAKKIRKMVDDQAKQMNLSKV RLCFQAYLPDDNGNFTKPLKPCISNAVYDSKAPSACQLKI CRMDKNSGCVTGGDEVYLLCDRVQKDDIEVHFLEYDSEG KVNWEDLGTFAPSDVHRQFAIVFKTPPYWNVAVEKPVKV FCELRRKSDKETSEPMEFTYTPQLFDQEIQIGAKRRKKIPH FSEYFQPPPGGPAGGAGGIGGGDATGYFNFGGGFTLAGIGLP STTSDQTQGGTGGGMQQATASPGTQPSQEMLKELAWAIAQH TSSAMRDYASTGDTCYLLAIQRQLTAVQNDDGDALHLAVIN CQFNATESLSVSMKDLPGDLVNEYNYLRQTPHLAVLTKQPC AAECLMKGNATKATSCDRHGNTPVHICAQGDIGCLKVLLNK KLRKESEEFPEIHWQNYNGFTPLHLAVIRGNREIHKMILSVGA DVEAKDGTGRTPLHLAVENNNLAIAGFLILEAKCYVDSYTF DDNTPLHLAAGRGLEGLTALLVAAGADTMETNSEDETPYSL ASTAEVKKILADEDEVPTLDDIKIPGKTVVKSSSCSPFMSLH QSVRVQNDRCRNDWIIAGAIFFMEI

<i>Capsaspora owczarzaki</i>: NF-κB	MDLSELSGWDPNLSLQEHTANLLAMDDSTMAAMILHSDGLS LFRDMGLYNSVSTSLDISGIPKFPPPPQQPQLAQAPRRRGHNQ SSSSDSHSTPSPGSLVLFSPSPASQDMSLQSPGLVGLTGNSRL ASTSEADDILLANILGHPVRSVSANTSMVGLPDDLGFSPTTG MDLTITATSPATADSSASATAFPAASPIASPSVSTSSGPVTVGGT SAALAALTPDRINRLLSAAEAAAEGAATLVEDLLMVTEEPA QFARFRYMSEQRERSLAGENSFPTLMVNPKYARVVPEMA LVTAVLVTKMPDPHTGRQQKHWHHLGGIPAAPLEGPQRI ARFDNIAVIMDKANNKDKDKSKAPVRSKDDQRCVRIMFE LVFVSGNTQFYGRAISQPIYNAKLAITKISHSSGPVTGGNEVI MLCSKIRKGVTVRMTDPTQWSVQAPSGSAWELNPQTLKA DCNVPGANLFFHHQYAVVLTLPYHTQTITAPVTVRISILDTD DETESQYVEYTYLPAAEAVRNAELAARKRRRDDSMDRDFMD RFDGSDGGNGSGSGRGNNGGHDGSDANNNGRGGGGGSSSS KGGDEPFNFNSLIPMHQHLHQLALSTVRAVQGFAASGDAR YLLALHRQLLAAPNENGDSPLHTAVAQGNLRSTMALLPLLA AEDLQSVNDMGETVLHSAVIEKRAAIARLLLAVAGADLGQSN ARNFNRLNSLHYLARHGDRATAMAVFGVFGSAQAPPANTNTP AQAPAGETKPKPADLRLRLARIQAQAIKALLACELETGATPAH LAIRGGHWHVFEACAKLAASAPIKAAGSLLSMVAEKSSGHS LLHSCVLANNEQAVRLLINLGASGNARDFGKNTPLHLAARQ GHIGIAALLVEAGATLSLNAVSTPLDVLTSSEGLSRDQLRA LVAVLRGELKYADMGRPTLRMPHTAELHSTAAALTSASPGA VSLADFYAGKKASRSPAPLGASSLLSSTGASAAGASAPTIAA VHAASATPVERTSMNNDDDYVLEKDAPYPVEQQPHGKRN KHSHHRFTRSSHGSQDKDELKKDKDDPKKEKEPKELSKFTL KEAFVDGTNFWELTRKFAGKKKMASASTGEMEPLSPERPLSP TNAGSGAASPFNQAKEQVSPGAVPPTGLEKLVNKLMDASEA TLSSQPAAEAVTPEQKLAEKLEKLGLAPASTTSAPPPHPKVAAL NAQSVEDARKTSTHALYSVD
<i>Aiptasia sp.</i>: NF-κB	MTHSEQQVGGTLTDSMLHEIMTPGFLPDISSLNVQMEMGYE GPYLEILEQPKSRGFRFRYPCEGPPSHGGLPGEFSDSKNKS YPSVQVCNYQGPCRIVVSIVTEDEPHMPHAHSLTGKHAN NDGIVTVQIGTEQGMTASFPNLGIQHVTKKKVAKTLTERY TKMQALQNATLTALATNNSTPSSFMNFGSVAREQVMASQ GPFDRNLAAAVAGEETKKILKLVQEQSKTMNLSAVRLCF QAYLPDENGNTKPLKPCISNPVYDSKAPASCQLKICRMD KNSGCVTGGDEIYLLCDRVQKDDIEIRFYENNDGDKPIWE DTGKFAPADVHRQFAIVFKTPAYHNIAIERPVEVLLELRRK SDKETSEPFFTYSPQMFDTQIGAKRRKKVPHFSDYYPP GGPPGAAGGGGGGFNFGSGFLDFGFGGIPGVGASTSTSTQQ SQGSSSGTTQNAMSSTQPSQEMLKELAWAIAKHTSSAMQDY

<i>Aiptasia sp.</i> : NF-κB (con't)	AATGDVRYLLSVQRQITAVQNDDGDTALHIAVINCQFTAIEGL VSVMKDLQGDFFINTFNLYLRQTPLQLATITKQALATECLLRGN ADATLRDRHGNTPVHTACAQGDVHCLRVLLDTKLRKEKDGF PELHWQNYDGYTPLHLAVIKGNREIIQILLSEGANVESKDGTC GRSPLHLAIEHDNLAIAGYLILEARCDVDSLTYDDNTPLHLA AGLGLVGETALLVAAGADTMATNSEDETPYSLATTAEVKKIL GDDEGPSDSITTSNTDPVITDIADKVNKNVQRFTHPPESLSV NSSNYGDYKHWQNRQEMDREDSGFGSQAERDNDLSLLSLT THPDKIMRPITDQEPFRDIHHGQPMGRTN
<i>Orbicella faveolata</i> : NF-κB	MATNSERQIGETLTDSLMDIMTPGYLPDISALQVPTATYSGP YMEILEQPKQRGFRFRYPCEGSPHGGPLPGQYSEKGGKSY PSVQLCNYHGPARIIVSLVTVDEPPMPHAHSLIGKNSNNG AVTVQIGPEHGMTASFPNLGIQHVTKKSVGKVLMEYIK MQTLHTATLHALTADSKGFDMELVGDQALADGETATFN TMAEAVAAEESQKVRQMVEDQKQSMNLNAVRLCFQAYL PDDGGCFTKALPPCISNPVYDSKAPSASNLKICRMDRNSG CVTGGDEVYLLCDKVQKEDIDVMFYEIDVETGKKTWEA GGVFAPTDVHRQVAIVFKTPAYWNIATERPVKVHLELRRK SDQETSEPVFTYQPQLFDKEQIGAKRRKKIPHFSDYFSG GGGGGGPPGMGGAGAGGGGGFGTFAPLGLLSALWFANPGTS QSNNGGGSGGAGTSQQGQSHSSGQSASGQSHQADLSELA NLAEKSAAMRDYAATGDMRYLLAAQRHLTAVQDDSGDTA LHLAVINSQQEVVQCLIDIMAGLPESYISEYNFLRQSPLHLAI TKQPRMLDCLLRASANVRSRDRHGNTAVHIACMHGDAVCLK ALLNYNVSKTVLWQNYQGVTVPVHLAVLAGSKDVLKLLNS AGANMSAQDGTSGKTPLHHAVEQDNLAVAGFLILEANCDVD ATTFDGNTPLHIAAASGLKGQTALLVAAGADTTLQNSDEETA FDLANVAEVQEILDEDEALSTDPSQDDELTAAGLTGLTLGQGD MDKLDPYVRRKMAQRLDPSIGADWRELAKRLGLGTLESFA IHSSPTTQVLAQYEAADGSIKTLRQVLRDMRRGDVLEILDGR KSPLSHDSGFDGLGSQSL SAYRSEELASYEAGSSSVSSSSLQ KGTTSLSTRKAGGYRPIRQHQDVF

Full-length amino acid sequences of NF- κ B proteins from the organisms used as input for MEME analysis to identify conserved motifs. NF- κ B sequences for *Orbicella faveolata* and *Actinia tenebrosa* were obtained through a BLASTp on NCBI using *Aiptasia pallida* NF- κ B (from NCBI); *Capsaspora owczarzaki* and *Nematostella vectensis* NF- κ B, *Homo sapiens* p100, p105, RelA, RelB, c-Rel, and *Drosophila* Relish, Dorsal, and Dif were obtained from UniProt database. All sequences were truncated to the Rel Homology Domain using MEME analysis and are shown in bold. The start of the RHD was highly conserved in all taxa and was predicted using the start of the known *Homo sapiens* RHD for both the NF- κ B and Rel proteins. The end of the RHD was highly conserved in all taxa and was predicted by using the MEME identified motif immediately before the predicted glycine-rich region. The trimmed and culled RHD sequence dataset allowed Clustal Omega to successfully align RHD sequences to be input into PAUP* for a maximum likelihood analysis bootstrapped 1000 times (Figure 3.5B).

Table 2.7 NF- κ B Rel Homology Domain sequences for Chapter 4 phylogenetic analysis.

Taxa: protein	Full-length NF-κB with truncated RHD domain bolded
<i>Homo sapiens:</i> p100	MESCYNPGLDGIIEYDDFKLNSSIVEPKEPAPETADGPYLVIVEQPKQRGFR FRYGCEGSPSHGGLPGASSEKGRKTYPTVKICNYEGPAKIEVDLVTHSDP PRAHAHSLVGKQCSELGICAVSVGPKDMTAQFNNLGVLVHTKKNNMG TMIQKLQRQLRSRPQGLTEAEQRELEQEAKELKKVMDLSIVRLRFS FLRASDGSFSLPLKPVISQPIHDSKSPGASNLKISRMDKTAGSVRGGDEV YLLCDKVQKDDIEVRFYEDDENGWQAFGDFSPTDVHKQYAIVFRTPPY HKMKIERPVTVFLQLKRKRGGDVSDSKQFTYYPLVEDKEEVQRKRRK ALPTFSQPFGGGSHMGGGSGGAAGGYGGAGGGGSLGFFPSSLAYSPYQSG AGPMGCYPGGGGGAQMAATVPSRDSGEEAAEPSAPSRTPOCEPAPEMLQ RAREYNARLFGLAQRSARALLDYGVTADARALLAQRHLLTAQDENGDT LHLAIHGQTSVIEQIVYVIHHAQDLGVVNLTNHLHQTPHLAVITGQTSVVS FLLRVGADPALLDRHGDSAMHLALRAGAGAPELLRALQSGAPAVPQLLH MPDFEGLYPVHLAVRARSPECLDLLVDSGAEVEATERQGGRTALHLATEM EELGLVTHLVTKLRANVNARTFAGNTPLHLAAGLGYPTLTRLLKAGADIH AENEEPLCPLSPPTSDSDSDSEGPEKDTRSSFRGHTPLDLTCSTKVKTLLN AAQNTMEPPLTPSPAGPGLSLGDTALQNLEQLLDGPEAQGSAELAERLG LRSLVDTYRQTTSPSGSLLRSYELAGGDLAAGLLEALSDMGLEEGVRLRGPE TRDKLPSTA EVKEDSAYGSQSVEQEAELGPPPEPPGGLCHGHPQPQVH
<i>Homo sapiens:</i> p105	MAEDDPYLGRPEQMFHLDPSLTHTIFNPEVFQPMALPTDGPYLQILEQPK QRGFRFRYVCEGSPSHGGLPGASSEKNKKSYQVKICNYVGPAAKIVVQL VTNGKNIHLHAHSLVGKHCEDGICTVTAGPKDMVVGAFANLGHVTKK KVFETLEARMTEACIRGYNPGLLVHPDLAYLQAEAGGGDRQLGDREKE LIRQAALQQTKEMDLSVRLMFTAFLPDSTGSFTRRLEPVVSDAIYDSK APNASNLKIVRMDRTAGCVTGGEEIYLLCDKVQKDDIQRIFYEEEEENG VWEGFGDFSPTDVHRQFAIVFKTPKYKDINITKPASVVFQLRRKSDLETS EPKPFLYYPEIKDKEEVQRKRQKLMPNFSDSFSGGSGAGAGGGGMFGSG GGGGGTGSTGPGYSFPHYGFPTYGGITFHPGTTKSNAGMKHGTMDTESKKD PEGCDKSDDKNTVNLFGKVIETTEQDQEPSEATVGNGEVTLTYATGTKEES AGVQDNLFLEKAMQLAKRHANALFDYAVTGDKMMLAVQRHLTAVQDE NGDSVLHLAIIHLHSQLVRDLLEVTSGLISDDIINMRNDLYQTPHLAVITKQ EDVVEDLLRAGADLSLLDRLGNSVLHLAAKEGHDKVLISLLKHKKAALLLD HPNGDGLNAIHLAMMSNSLPCLLLLVAAGADVNAQEQQSGRTALHLAVEH DNISLAGCLLLEGDAHVDSTTYDGTTPHLIAAGRGSTRLAALLKAAGADPL VENFEPLYDLDDSWENAGEDEGVVPGTTPDMATSWQVFDILNGKPYEPEF TSDDLAAQGDMKQLAEDVKLQLYKLEIPDPDKNWATLAQKLGLGILNNA FRLSPAPSKTLMNDYEVSGGTVRELVEALRQMGYTEAIEVIAQAASSPVKTT QAHSLPLSPASTRQQIDELRDSVCDSDSGVETSFRKLSFTESLTSGASLLTLN KMPHDYQGEGPLEGKI
<i>Homo sapiens:</i> RelA	MDELFPFLIFPAEPAQASGPYVEIEEQPKQRGMRFRYKCEGRSAGSIPGERS TDTTKTHPTIKINGYTGPPTVRISLVTKDPPHRPHPHLVGKDCRDGFY EAELCPDRCIHSFQNLGIQCVKKRDLEQAISQRIQTNNNPQVPIEEQRG DYDLNAVRLCFQVTVRDPGRPLRLPPVLSHPIFDNRAPNTAELKICRVN RNSGSCLGDEIFLLCDKVQKEDIEVYFTGPGWEARGSFQADVHRQV AIVFRTPPYADPSLQAPVRVSMQLRRPSDRELSEPMFQYLPDTRRHRI EEKRKRITYETFKSIMKKSFPFSGPTDPRPPPRRIAVPSRSSASVPKPAPQYPF TSSLSTINYDEFPTMVFPSPGQISQASALAPAPPQVLPQAPAPAPAPAMVSALA QAPAPVPVLAPGPPQAVAPPAPKPTQAGEGTLSEALLQLQFDDDEDLGALLG

	NSTDPVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGA QRPPDPAPAPLGAPGLPNGLLSGDEDFSSIADMDFSALLSQISS
<i>Homo sapiens:</i> RelB	MLRSGPASGPSVPTGRAMPSRRVARPPAAPELGALGSPDLSSLSLAVSRSTD ELEIIDEYIKENGFGLDGGQPGPGEGLPRLVSRGAASLSTVTLGPVAPPATPP PWGCPLGRLVSPAPGPGPQPHLVITEQPKQQRGMFRFYECEGRSAGSILGE SSTEASKTLPALIELRDCGGLREVEVTACLWVKDWPHRVHPHSLVGKDC TDGICRVRLRPHVSPRHSFNNLGIQCVRKKEIEAAIERKIQLGIDPYNAG SLKNHQEVDNMNVVRICFQASYRDQQGQMRRMDPVLSEPVYDKKSTNT SELRICRINKESGPCTGGEELYLLCDKVQKEDISVVFSRASWEGRADFSQ ADVHRQIAIVFKTPPYEDLEIVEPVTNVFLQRLTDGVCSEPLPFTYLPR DHDSYGVDDKKRKRGMPPDVLGELNSSDPHGIESKRRKKKPAILDHFLPNHG SGPFLPPSALLPDPDFSGTVSLPGLEPPGGPDLLDDGFAYDPTAPTFTMLD LLPPAPPHASAVVCSGGAGAVVGETPGPEPLTDSYQAPGPGDGGTASLVG SNMFPNHYREAAFGGGLSPGPEAT
<i>Homo sapiens:</i> c- Rel	MASGAYNPYIEHIEQPRQRGMFRFYKCEGRSAGSIPGEHSTDNNRTYPSI QIMNYYGKGKVRITLVTKNDPYKPHPHDLVGKDCRDGYEEAEFGQER RPLFFQNLGIRCVKKKEVKEAHTRIKAGINPFNVPEKQLNDIEDCDLNV VRLCFQVFLPDEHGNLTALPPVVSNNPIYDNRAPNTAELRICRVNKNCG SVRGGDEIFLLCDKVQKDDIEVRFLNDWEAKGIFSQADVHRQVAVIFK TPPYCKAITEPVTVKMQLRRPSDQEVSESMDFRYLPDEKDTYGNKAKK QKTLLFQKLCQDHVETGFRHVDQDGLELLTSGDPPTLASQSAGITVNFPE RPRPGLLSIGEGRYFKKEPNLFSHDAVVREMPGTGVSSQAESYYPSPGPISG LSHHASMAPLPSSSWSSVAHPTPRSGNTNPLSSFSTRTLPSNSQGIPPFLRIPV GNDLNASNACIYNNADDIVGMEASSMPSADLYGISDPNMLSNCVNMMTT SSDSMGETDNPRLLSMNLENPSCNSVLDPRDLRQLHQMSSSSMSAGANSNT TVFVSQSDAFEGSDFSCADNSMINESGPSNSTNPNSHGFVQDSQYSGIGSMQ NEQLSDSFPEFFQV
<i>Capsaspora</i> <i>owczarzaki</i> : NF- κ B	MDLSELSGWDPNLSLQEHTANLLAMDDSTMAAMILHSDGLSLFRDMGLYN SVSTSLDISGIPKFPQQPQLAQAPRRGHNQSSSSDSHSTSPSGSVLFSPSP ASQDMSLQSPGLVGLTGNSRLASTSEADDILLANILGHPVRSVSANTSMV GLPDDLGFSPTTGMDLTITATSPATADSSASATAFPAASPIASPSVSTSSGPVT VGGTSAALAALTPDRINRLLSAAEAAAEGAATLVEDLLMVTEEPAQFARF RYMSEQRERSLAGENSFPTLMVNPKYARVVPEMALVTAVLVTKMPDP HTGRQQKHWHHLGGIPAAPLEGPRQRIARFDNIAVIMDKANNKDKDKSK APVRSKDDQRCVRIMFELVFVSGNTQFYGRAISQPIYNAKLAITKISHSSG PVTGGNEVIMLCISKIRKGVTVRMTDPTQWSVQAPSGSAWELNPQTLKAD CNVPGANLFFHHQYAVVLTLPYHTQTITAPVTVRISILDTDETESQYVEY TYLPAEAAVRNAELAARKRRRDDSMDRDFMDRFDGSDGGNGSGSGRGNNG GHDGSDANNNGRGGGGGSSSSKGGDEPFNFNSLIPMHQHKHLQLALSTVR AVQGFAASGDARYLLALHRQLLAAPNENGDSPLHTAVAQGNLRSTMALLP LLAAEDLQSVNDMGETVLHSAVIEKRAAIARLLLAVAGADLGQSNARNFNR NSLHYLARHGDRATAMAVFGVFGSAQAPPANTNTPAQAPAGETKPKPADL RLLARIQAQAIKALLACELETGATPAHLAIRGGHWHVFEACAKLAASAPIK AAGSLLSMVAEKSSGHSLLHSCVLANNEQAVRLLINLGASGNARDFGKNT LHLAARQGHIGIAALLVEAGATLSLNAVVSQTPLDVLTSEGSGLSRDQLRALV AVLRGELKYADMGRPTLRMPHTAELHSTAAALTSASPGAVSLADFYAGK KASRSPAPLGASSLLSSTGASAAAGASAPTIAAVHAASATPVERTSMNND YVLEKDAPYPVEQQPHGKRKNKSHHRFTRSSHGSQDKDELKKDKDDPKK EKEPKELSKFTLKEAFVDGTNFWELTRKFAGKKKMASASTGEMEPLSPERP LSPTNAGSGAASPFNAKEQVSPGAVPPTGLEKLVNKLMDASEATLSSQPA EAVTPEQKLAEKLEKLGLAPASTTSAPPPHPKVAALNAQSVEDARKTSTHA LYSVD

<i>Amphimedon queenslandica</i> : NF-κB	MAFNGIDPSSLPEALIRDTMNTLPTHIVNSNPDSLNSSPYSSLTQVRLEIVEQ PKSRGFRFRYDCEGQSHGGLPGENSEKNRRQKTYPTVHLKGYRGRAR VMVSLVTSDPAMPHAHSIVGKNAIDGRCVVEIGPETDMYAQFTSLGIL HVTKKKVPEVLTRRLLQQTTPRGQMVDQMEVVDVDMTTAQLTSEEQ DEIHQQAQTLAKSMNLSVRLCFQAFLPDENGRYTIPIDPVFSNKVYDS KAPSAGTLKICRLDRTSGSVKGGDDVFLLCDKVQKNDIEVVFYEDKQE TTGGMQLQPWMAKGRFGPNDVHHQYAIVFQTPTFYNQAIHPVQVWI ALKRPSDHETSEPKPFLYLPQEFDEERIGQKRRKKITHFNNFFEGPGGG GGAGGGAGGAGGNFFSRDFNYGSGGGYNSGFNFFGGSSGGGGSSGGGSAN NAGTGGGGTTFSGGNTSAANMPVSVDSLSTLPPSSNQHIFAATATNPHYPH MRQQQPHGLSFSNNGGGRGMGGTMYSHAMDQQFMSTASGHLVSRTGAGG VTVKREPPDYMDVERDNVQPPLPALSEEGGTGGGNIMRPKDQLPPSQRGP GDSGKLLDVSSNIEDVESGYVEAERMDSGLPTSMAAEPSEQSTSSSETEAQQA LQALKDRQQMAFEVCDRMFNALLAWATTKDIRYLLAAQRSLTAVQNQEG DTALHLAIIHNHQDVVLQLLDVLPQLPPTETPVVDCLNNFKQSPVHLAVITR QHKVVQYLLKANANPLVSDRNGDTPLHLACKYGFLLQIVPLLRSTRINTE GCRIPELVMRNNDGLTPLHLAAACGNPDCFKELVKAHADVNVAQDSKSGKS ALHYLIEKGDLPLTGFLITESETNIECTDFSGNTPLHCAAALGNVAIVSLLIAA GANLVCQNQEGELPLVLAIEYGGHEEVVKVLKDSL VKAGLDKPEEQLSTQM KSVSLTEEDKALAALRANSSEGDLSKLDLFRPRISLALILDPINEGCDWKALA KCLSLSHLEAGLEAMTSPTKELLTMYEACDGTIAKLRLQALLDINRSDAVNII DRYMQUEEKGIVTSKQTYDSGISSNPSSLSNEQVVVGKHSTIPSSQV*
<i>Nematostella vectensis</i> : NF-κB	MAQSEQQVGSALTESMLNEIRPGYLPDISALHVPLGTNAEEPSYTEPYLEIL EQPKPRGFRFRYPSEGPSHGGLPGQFSTSKSKSYPSVQVNNYQGPCRIV VTLVTKDEPYMLHAHSLTGKNANEEGVVTVQVGPDQHMTASFPNLGI QHVTKKNVVKVLMDFRIKWQTLQNAATFAKLSEGKIDGVDLSLFGVNT AINS NKLGF DKNVALSVANQEA AKSREYAKQQAAMDLSAVRLCFQAY LPDQDGNFTRPLKPVYSDAVLDSKEPSASQLKICRMDKNSGCVTGGDEI YLLCDKVQKDDIEIHFYEMDDITGKYTWEDLGKFSPCDVHRQFAIVFKT PPYWNIAIERPANVLVELRRKKNGETSEPVQFTYQPPQLFDKEAIGAKR RKTVPHFTEFLSGGSSGATGGGGSSVSGFNFSADFLQQGVFLTQNPSNM
<i>Aiptasia pallida</i> : NF-κB	MTHSEQQVGGTLTDSMLHEIMTPGFLPDISSLNVQMEMGYEGPYLEILEQPK SRGFRFRYPCEGPSHGGLPGEFSDSKNKSYPVQVCNYQGPCRIVVSLVTE EPHMPHAHSLTGKHANNDGIVTVQIGTEQGMTASFPNLGIQHVTKKKVAKT LTERYTKMQALQNAATLTALATNNSTPSSFMNFGSVAREQVMASQGPFDNR LAAAVAGEETKKILKLVEQSKTMNLSAVRLCFQAYLPDENGNFTKPLKPC ISNPVYDSKAPASCQLKICRMDKNSGCVTGGDEIYLLCDRVQKDDIEIRFYE NNDGKPIWEDTGKFAPADVHRQFAIVFKTPAYHNIAIERPVEVLLELRRKS DKETSEPF TFTYSPQMF DTEQIGAKRQKKVPHFSDYYPGGPPGAAGGGGG GFNFGSGFLDFDFGFGGIPGVGASTSTSTQSQSGSSGTTQNAMSSTQPSQEM LKELAWAIAKHTSSAMQDYAATGDVRYLLSVQRQITAVQNDGD TALHIAI INCQFTAIEGLVSVMKDLQGD FINTFN YLRQTPLQLATITKQALATECLLRG NADATLRDRHGNTPVHTACAQGDVHCLRVLLDTKL RKEKDGFP ELHWQN YDGYTPLHLAVIKGNREIIQILLSEGANVESKDGTCGRSPLHLAIEHDNLAI GYLILEARCDVDSLTYDDNTPLHLAAGGLVGETALLVAAGADTMATNSE DETPYSLATTAEVKKILGDDEGPSDSITTSNTDPVITDIAEKVNKNVQRF TTP PESESLVNSSNYGDYKHWQNRQEMDREDSGFGSQAERDND SLLSLTTHP DKIMRPITDQEP RFDIHHGQPMGR TN
<i>Drosophila (Fly)</i> : Dif	MFEEAFGDIQEIIINASMELNGGATGGGGSVAGAVGGGGAAHHILSQSTSLPV MPSHIPLHLQNQNMNQNLPESARSGPLRIVEEPTSNIIRFRYKCEGRT AGSIPGMNSSSETGKTFTPTIEVCNYDGPVHVVSCVTSDEPFRQHPHWLV SKEEADACKSGIYQKKLPPEERRLV LQKVG IQCAKKLEMRDSLVERER RNIDPFNAKFDHKDQIDKINRYELRLCYQAFITVGNSKVPLDPIVSSPIYG

	KSSELTITRLCSCAATANGGDEIIMLCEKIAKDDIEVRFYETDKDGRETW FANAEFQPTDVFQMAIAFKTPRYRNTETQSVNVELKLVRPSDGATSA PLPFEYYPNPELLTKHNRRVAQKTVESLKRSLMSTNLHPSKQVKTSSQYTI FSKPQIATTTTPQTQVSPGMPLMFPGGSPNFVQDIKMENGFMVDVSQSSQCPS VERNFA SPRSNCSTVDSIPPMQMGQNQTHLYLPDATNFTFNGNFAS PSSNCS TVDSIPPFQIGQRNNHMYLPENS NFPVNGCSPTHFSGGSMTPINNNNNVLINN NNNDFLSQKMSAISIPPQGNFGIKQVYQQTQQFLPQLQPESIPYLAQSHPEQS QYQQQQQPQEQQPPADEPTQSFSDLISSIGMAPIDTSELIQDIEAELNSLGIQ PFK
<i>Drosophila (Fly): Dorsal</i>	MFPNQNNGAAPGQGPVVDGQQSLNYNGLPAQQQQQLAQSTKNVRKKPYV KITEQPAGKALRFRYECEGRSAGSIPGVNSTPENKTYPTIEIVGYKGRAV VVVSCVTKDTPYRPHPHNLVGKEGCKKGVC TLEINSETMRAVFSNLGI QCVKKKDIEAALKAREEIRVDPFKTGFSHRFQPSIDLNSVRLCFQVFM ESEQKGRFTSPLPPVSEPIFDKKAMSDLVICRLCSCSATVFGNTQIILLC EKVAKEDISVRFFEEKNGQSVWEAFGDFQHTDVHKQTAITFKTPRYHT LDITEPAKVFIQLRRPSDGV TSEALPFEYVPMDSGKHTFWNLHRHLKRK PDEDLFQQILRLDAKREVQPPTIEVIDLDTPKIDVQREIPSEMEFNHEESQQSE PALEQEQSVQQEQYTQEQSLQQEQYTQEQSLQQEQYLQQLEQQQSFQLEEP MQQDQELPAQQSFDQAIDHLPDHTSDHIPEDMEAADAHAEAEAHRLRSEQE KEIDTIIDEKVRELEQLDLGQQLEPRPLTANDKITEWMKSEIEQQVHEPSPT AEADVLD SALEISKADKTLDELLETVAELDEIYTD FKVQRD TYKNTIQNELA GLQGRAPLQVEDSFDDAATY TSLQIAFKNPVLIPMDDIMPPTPPMSQCAPED AHQHYDPVEVNSQARKPETPMRPVPPVPPAILTIQYPPEEDKLPLPPKRIRK QDSNAENRSIEANTVQTKPSTGESPLNKRLPPAPKNPNFNTLPRQKKPGFFSK LFSRRKSKPD LAQQQENSSILDSKANSREPSIGHFNMQDPMRASLRSSKSAA PFISNPAPAKSSPVKAKKPGSKLTKPVGRSVSSVSGKRPA YLNADV VHIPLK GDSVNSLPQQQRTEGYSQSSTISVGAGLDRRTASALQLADIPISEGGMELVAI ADRQSLHNLVSSIEGHFNVQLDPNLDL TEAEHFALYTSIPPLAAASEFDE TSA YYAPVDAGEILTPDEVAKRLAAANGI
<i>Orbicella faveolata: NF-κB</i>	MATNSERQIGETL TDSLLMDIMTPGYLPDISALQVPTATYSGPYMEILEQPK QRGFRFRYPCEGPSHGGLPGQYSEK GKKSYP SVQLCNYHGPARI VVSL VTVDEPPMPHAHSLIGKNSNNGAVTVQIGPEHGMTASFPNLGIQHVTK KSVGKVL MERYIKMQTLHTATLHALTADSKGDFMELVGDQALADGET ATFNRTMAEVA AAEESQKVRQMVEDQKQSMNLNAVRLCFQAYLPDDG GCFTKALPPCISNPVYDSKAPSASNLKICRMDRNSGCVTGGDEVYLLCD KVQKEDIDVMFYEIDVETGKKTWEAGGVFAPTDVHRQVAIVFKTPAY WNIATERPVKVHLELRRKSDQETSEPV EFTYQPQLFDKEQIGAKRRKKI PHFSDYFSGGGGGGGPPGMGGAGAGGGGGGFTAPLGLLGSALWFANPGTS QSNNQGGSGGAGTSQQGQSHSSGQSQASGQSHQADLSELAWNLA EKSA A MRDYAATGDMRYLLAAQRHLTAVQDDSGDTALHLAVINSQQE VVQCLIDI MAGLPESYISEYNFLRQSPLHLAAITKQPRMLDCLLRASANVRSRDRHGNT AVHIACMHGDAVCLKALLNYNVSKTVLWNQNYQGVTPVHLAVLAGSKDV LKLLNSAGANMSAQDGTSGKTPLH HAVEQDNLAVAGFLILEANCVDV DATT FDGNTPLHIAAASGLKGQTALLVAAGADTTLQNSDEETA FDLANVAEVQEI LDEDEALSTDPSQDDELTA GLTGLTGQGDMDKLDPYVRRKMAQR LDP SIG ADWRELAKRLGLGTLES AFAIHSSPTTQVLAQYEAADGSIKTLRQVLRDMR RGDVLEILDGRKSPLSHDSGFD SGLGSQSLSAYRSEELAS YEAGSSSV PSSSS LQKGTTSL ESTRKAGGYRPIRQH QDVF
<i>Mus Musculus (Mouse): p100</i>	MDNCYDPGLDGIPEYDDFEFSPSIVEPKDPAPETADGPYL VIVEQPKQRGF RFRYGCEGPSHGGLPGASSEKGRKTYPTVKICNYEGPAKIEVDLVTHSD PPRAHAHSLVGKQCESELGVC AVSVGPKDMTAQFNNLGVLHVTKKNM MEIMIQKLQRQRLRSKPQGLTEAERRELEQEAKELKKVMDLSIVRLRF SAFLRASDGSFSLPLKPVISQPIHDSKSPGASNLKISRMDKTAGSVRGGDE

	VYLLCDKVQKDDIEVRFYEDDENGWQAFGDFSPTDVHKQYAI VFRTPPYH KMKIERPVTVFLQLKRKRGGDVSDSKQFTYYPLVEDKEEVQRKRRKALPTF SQPFGGGSHMGGGSGGSAGGYGGAGGGGSLGFFSSSLAYNPYQSGAAPMG CYPGGGGGAQMAGSRRD TDAGEGAEERTPPEAPQGE PQALDTLQRAREY NARLFGLAQRSARALLDYGVTADARALLAGQRHLLMAQDENGDTPLHLAI IHGQTGVIEQIAHVYHAQYLGVINLTNHLHQTPLHLAVITGQTRVVSFLLQV GADPTLLDRHGDSALHLALRAGAAPELLQALLRSGAHAVPQILHMPDFEG LYPVHLAVHARSPECLDLLVDCGAEVEAPERQGGRTALHLATEMEELGLVT HLVTKLHANVNARTFAGNTPLHLAAGLGSP TLRLLKAGADIHAENEEPL CPLPSPSTSGSDSDSEGPERDTQRNFRGHTPLDLTCSTKVKTLLLNAAQNTTE PPLAPPSPAGPGLSLGDAALQNLEQLLDGPEAQGSWAELAERLGLRSLVDT YRKTPSPSGSLLRSYKLAGGDLVGLLEALS DMGLHEGVRLKGPETRD KLP STEVKEDSA YGSQSVEQEA EKLCPPEPPGGLCHGHPQPVH
<i>Mus Musculus</i> (Mouse): p105	MADDDPYGTGQMFHLNTALTHSIFNAELYSPEIPLSTDGPY LQILEQPKQR GFRFRYVCEGPSHGG LP GASSEKNKKSYPQVKICNYVGP AKVIVQLVTN GKNIHLHAHSLVGKHCEDGVCTV TAGPKDMVVGFANLGILHVTKKKV FETLEARMTEACIRGYNPGLLVHSDLAYLQAEGGGDRQLDREKEIIRQ AAVQQTKEMDLSVVRMLMFTAFLPDSTGSFTRRLEPVVSDAIYDSKAPNA SNLKIVRMDRTAGCVTGGEIYLLCDKVQKDDI QIRFYEEENG GVWE GFGDFSPTDVHRQFAIVFKTPKYKDVNITKPASVFVQLRRKSDLETSEPK PFLYYPEIKDKEEVQRKRQKLMPNFSDSFGGGSGAGAGGGGMFGSGGGG GSTGSPGPYGYSNYGFPPYGGITFHPGVTKSNAGVTHGTINTKFKNGPKDC AKSDDEESLTLPEKETEGEGPSLPMACTKTEPIALASTMEDKEQDMGFQDN LFLEKALQLARRHANALFDYAVTGDVKMLLAVQRHLTAVQDENGDSVLH LAIHHLHAQLVRDLLEVTSGLISDDIINMRNDLYQTPHLAVITKQEDVVEDL LRVGADLSLLDRWGNSVLHLAAKEGHDRILSILLKSRKAAPLIDHPNGEGLN AIHIAVMSNSLPCLLLLVAAGAEVNAQE QKSGRTALHLAVEYDNISLAGCL LLEGDAHVDSTTYDGTTP LHIAAGRGSTRLAALLKAAGADPLVENFEPLYD LDDSWEKAGEDEGVVPGTTPLDMAANWQVFDILNGKPYEPVFTSDDILPQG DMKQLTEDTRLQLCKLLEIPDPDKNWATLAQKLGLGILNNAFRLSPAPSKTL MDNYEVSGGTIKELMEALQQMGYTEAIEVIQAAFRTPATTASSPVTTAQVH CLPLSSSSSTRQHIDELRDSDSVCDSGVETSFRKLSFTESLTGDSPLLSLNKMP HGYGQEGPIEGKI
<i>Mus Musculus</i> (Mouse): RelA	MDDLFP LIFPSEPAQASGPYVEIIEQPKQRMFRYKCEGRSAGSIPGERS TDTTKTHPTIKINGYTGP GTVRISLVTKDPPHRPHPHELVGKDCRDGY EADLCPDRSIHSFQNLGIQCVKKRDLEQAISQRIQTNNNPFHVPIEEQRG DYDLNAVRLCFQVTVRDPAGRPLLLTPVLSHPIFDNRAPNTAELKICRV NRNSGSCLG GDEIFLLCDKVQKEDIEVYFTGPGWEARGSFQSADVHRQ VAIVFRTPPYADPSLQAPVRVSMQLRRPSDRELSEPMEFQYLPD TDDRH RIEEKRKRTYETFKSIMKKSPFNGPTEPRPPTRRIAVPTRNSTSVPKPAPQPY TFPASLSTINFDEFSPMLLPSGQISNQALALAPSSAPVLAQTMVPSSAMVPLA QPPAPAPVLTGPPQSL SAPVPKSTQAGEGTLSEALLHLQFDAEDL GALLG NSTDPGVFTDLASVDNSEFQQLLNQGVSM SHSTAEPMLMEYPEAITRLVTG SQRPPDPAPTPLGTSGLPNGLSGDEDFSSIADMDFSALLSQISS
<i>Mus Musculus</i> (Mouse): RelB	MPSRRAARESAPELGALGSSDLSSLTVSRTTDELEIIDEYIKENGFLDGT QLSEMPRLVPRGPASLSSVTLGPAAPPPATPSWSCTLGRLVSPGPCRPYLV ITEQPKQRMFRYCEGRSAGSILGESSTEASKTLPAIELRDCGGLREV EVTACL VWKDWPHRVHPHSLVGKDCTDGVCRVRLRPHVSPRHSFNNL GIQCVRKKEIEAAIERKIQLGIDPYNAGSLKNHQEVD MNVVRICFQASY RDQQGHLHRMDPILSEPVYDKKSTNTSELRICRINKESGPCTGGEELYL LCDKVQKEDISVVFSTASWEGRADFSQADVHRQIAIVFKTPPYEDLEISE PVTNVNVLQRLTDGVCSEPLPFTYLPRDHDSYGVDKKRKRGLPDVLGE LSSSDPHGIESKRRKKKPVFLDHFLPGHSSGLFLPPSALQPADSDFFPASISL

	PGLEPPGGPDLLDDGFAYDPSAPTLFTMLDLLPPAPPLASAVVGSGGAGATV VESSGPEPLSLDSFAAPGPGDVGTASLVGSNMFPNQYREAAFGGGLSPGPE AT
<i>Mus Musculus</i> (Mouse): c-Rel	MASSGYNPYVEIIEQPRQRGMRFYKCEGRSAGSIPGERSTDNNRTYPS VQIMNYYGKGKIRITLVTKNDPYKPHPHDLVGKDCRDGYEAEFGPER RPLFFQNLGIRCVKKKEVKGAHLRISAGINPFNVGEQQLLDIEDCDLNV VRCVFMFFLPDEDGNFTTALPPIVSNPIYDNRAPNTAELRICRVNKNCGS VRGGDEIFLLCDKVQKDDIEVRVFLNDWEARGVFSQADVHRQVAIVFK TPPYCKAILEPVTVKMQLRRPSDQEVSESMDFRYLPDEKDAYGNKSKK QKTTLIFQKLLQDCGHFTEKPRTAPLGSTGEGRFIKKESNLFSHGTVLP RSGVPGQAEPYYSSCGSISSGLPHHPAIPSVAHQPTSWPPVTHPTSHPVST NTLSTFSAGTLSSNSQGILPFLEGPVSDLSASNSCLYNPDDLARMETPSMSP TDLYSISDVNMLSTRPLSVMAPSTDGMDGTDNPRLV SINLENPSCNARLGPR DLRQLHQMSPASLSAGTSSSSVFVSQSDAFDRSNFSCVDNGLMNEPGLSDD ANNPTFVQSSHYSVNTLQSEQLSDPFTYGFFKI
<i>Drosophila (Fly):</i> Relish	MNMNQYYDLNNGKNVFMFNDASSTSGYSSSTSPNSTNRSFSPAHSPTME LQTDFA NLNLPGGNSPHQPPMANSFYQNQLLNNGGICQLGATNLINSTGVSF GVANVTSFGNMYMDHQYFVPAPATVPPSQNFQYHQNGLASDGDIKHVPQL RIVEQPVEKFRFRYKSEM HGTHGSLNGANSKRTPKTFPEVTLCNYDGP AVIRCSLFQTNLDSPHSHQLVVRKDDRDVCDPHDLHVS KERGYYAQFIN MGIHTAKKYIFEELCKKKQDRLVFMNRRREL SHKQLQELHQETERE KDMNLNQVRLCFEAFKIEDNGAWVPLAPPVYSNAINNRKSAQTGELRIV RLSKPTGGVMGNDELILLVEKVS KKNIKVRFFEEDEGETVWEAYAKF RESDVHHQYAIVCQTPPYKDKDVDREVNVYIELIRPSDDERSFPALPFRY KPRSVIVSRKRRRTGSSANSSSGTESSNNSLDLPKTLGLAQPPNGLPNLSQ HDQTISEEFGREKHLNEFIASEDFRKLIEHNSSDLEKICQLDMGELQHDGHN AEVPSHRNRTIKCLDDLFEIYKQDRISPIKISHHKVEKWFIEHALNNYNRDTL LHEVISHKKDKLKLAIQTIQVMNYFNLKDVVNSTLNADGDSALHVACQD RAHYIRPLLGMGCNPNLKNAGNTPLHVAVKEEHLSCVESFLNGVPTVQLD LSLTNDDGLTPLHMAIRQNKYDVAKKLISYDRTSISVANTMDGNNALHMA VLEQSVELLVLILDAQNENLTDILQAQNAAGHTPLELAERKANDRVVQLLK NVYPEKGELAMTWIPCKVKEEIDSSDESSDAGQLEIKSEEMDIETKDEDSV ELDLSSGPRRQKDESSRDTEMDNNKLQLLLKNKFIYDRLCSSLNQPLGHGS DPQDRKWMQLARQTHLKQFAFIWLGAEDLLDHVKRK GASVEFSTFARALQ AVDPQAYALLVNPT
<i>Carcinoscorpius</i> <i>rotundicauda</i> (Horseshoe crab): Relish	MATYAQCFFPATRPNQEGTSSFFINPFCEPVKGRYMPSVMGEPSSLLLPEQ ETPGLSAIVSDCVSMSNTESFQHSFPITNCSPASPLQPTTPMAQPVGVSSSP PLHQIQNGISPSISPEHIHTLQPPCPDHPYLSVLEQPTNRIRYRYKSEKGS GLTGEFSSSSKKTYP TVKLENYRPQNNQVFIKATLYTVDDQKPHVHK LMGKHCQDGVCSVILGEDMVASFQNLGILFVGKKEVPDILYKKKLEDQ HLLRCLMQNGSLHITEADKNQLKKDAEQEAKNMDLNRVKIRFEAFVAVS QGHLYPICDAVFSNIIANQKCPDVGELKIVKMDKCSGVCTGSDEVFLLC EKVNKKDVKVIFFEEDENGMIQWQDFGSFTEADVHHQVAIVFKTPPYR DLMIKQPVVKVLQLYRYRDGECSKPKDFVYFPIDHHRDGIERKRKKLN QCQNFSGFDGEYFRSTQSFSGSGSGSIRNTSFYDHGQDSVGFQSKKML NASQKKGTEVRTNLKDDRQGNVETVKGFEPSTDSGKWSEVNNNQELP SDKSVSTDKPLENSFDRLCISKLKEEDENTNVSKSRNSRSEILCGKSTLASQV PQNKRANATKTKCTELHMSTSCSTQGDLELNSRTELSYEQNNVFSPTAE QDYQQNKLCKSSGPWLVEKLALRMSLSGREFVCSGDLAKILMTMHHLVAI QDSNGDNLFHLAMIHSGSQADHLELVRCLLNALKEETRDAINQCNNLKQT PLMLAVLTRNPYVVQELLFHGANLNVADAEGNTPLHIATQIGDDYCLSILLD SKMYEAQQSPISPNNALNAGYAALHLAVRHNHSDCVMVLCARGADINV MDGTSGHTPLHLAVEWNPQIVQFITKISHVNINVNQNFAGNTPLHLACAHRD

	ENVVRILINAHANPLVENYDVYSSSKRHERDIEVLKNKGKTPLDFAKNKKQ LRCILAGIVTSPKAQSQDLKSSFSISKPSHSSSTHNDKKVSRKRSEDEIKYSEG KHHLNNEHLSTNKEAVSNEFIPFDSGYASGEIKTEDKNRMDHDLCTKLDKKQ QWKQLGMALGFENSVINSLERLCGKEKSPAKAIIDRFLQENPDCDLMTLKN NALVAAELQEPELEVIQMIIEV
<i>Strongylocentrotus purpuratus</i> (Sea Urchin): NF-κB	MEEKSDSVQSVEIPEDMLQQLIHQNGQVGLPMSGAEENDFDSNDLMHLPVM KGLQEEVSKPHLKILEQPRQGRFRFRYGCEGSPSHGGLPGQNSQGRKRSF PSVEICNYKGSARIVVSLVTNEETPRPHAHSLVGKHCKDGLCTVQVGPK DMTASFPNLGILHVTRKDVVPTLKTTRILAQHRLYKDLINNSTPGESHW EPSDAEIEKKAKEMAKDMDLSVRLCFQTYLPDISGHFTRPLDPVISVPV FDSKAPNATTLKICRMDKSAGCCTGGEEVYLLCDKVQKEDIQVKFFEIS ADGQMVWQSLAEFGPTDVHRQYAIVFKTPAYKDINIDKPVYVHVQLKR KSDNETSDPKPFTFHPQVPDREGILRKRKKHLAHFNEYSSTYQQGGLGG SNGMGGGGGSGTSFNFNPNPMGYGANTYNFSGNQMTSSTSQQPSVPQQIPSH QSGVVATGNSSQQMVYTTSTGAQNHQPLPSIATLARKPQEQQLQQQFQLQQ QLQQQQQQQQQQQQQQQQQQQQQQQQQQPMQMDFMPPPIQGSAMGGQAGGDR VKMESESKYCGNMDTAEGFQLDSNTLRDCGPLDLKELQDTGVLDLIPKGI SRPKDMVEKNIQTDDITIDSDSMAWHVAQVTANALHDYAAATGDIKTILTQV RHLIAVEDDNGDTALHAAIINKKYDVTHALLSAVIKIPDQIIVNQTNHLKQTP LHLAVITNQSKMVEVLLRCGANPNLDCHEGNTPLHLATMMGMTEGVNFLV RGPKAKAAIKPIKTDINPTNYEGLAPVHLAVIAKNLDILKALVSSGADVNV DGKTGRALHYAVEVESFPILGYLLIEAKVDINAVTFCGDSALHLASSDLR AVATLLIAAGADPKLENADLGDDSDSDVDGEDDDGEKDGEEKKEEGVEEEE KEKHGKTP
<i>Danio rerio</i> (Zebrafish): NF-κB	MAGALRMDDTHYQIGFVVNEPMTFPSAYDDLPMPEAFEVKEHPFVQETV DGPYIQIIEEPKQGRGFRFRYECEGSPSHGGLPGASSERNRRTYPTVKVLNF VGNARVEVQLVTHTDPPRVHAHSLVGRHCNESGVCSVDVGPSDFTAQF SNLILHVTGRGVVEVLTKRLKEEKRKVKGPYKFSDAEENALMQEA KELGKNMDLNIVRLKFTAYLQDSNGSYTRALKPVVSNPIYDSKSPNASN LKISRMDKTSGLVGGEEVFLCDKVQKDDIDIRFYEEDEWEALGDFS PTDVHKQYAIVFKTPPYRCTNIDRPVTVFLQLKRKKGGDCSEPQKQFTYV PHNQDKEEVQRKRMKALPDHYGGWRGGPQGGAGGFGGPGSGTGGGGG MGGGFPFNPMDNSSFFVGGCGGFGGGAQMSDSPQTGDANTQSQSNSQIRQ QLIQIALHNANLAALKCNARALLQYSSTGDARPLALQRHLCGVQDENGDT PLHLAIIHQKTVAEQLIKALNSSPQQKFFNKLNLKSQAPLHLAVITKQPKLV EMLMKSGADPSLLDREGRTVLHLAAHTGDDVILRLVLSLLGEHHSHLINS DFTGQYPVHLAVKKDGERCLRLLEAGAKINMPEQKSGCTALHLAVRDNL LKLACNLITELKADVNACTYGGNSPLHLAASQGSPLHCSMLIAAGADKRIE NDEPLFLSSSSSDEDEESEKKLSEDMVQQVHCLSISSERGVFNPRKPAAGH TPFDLAKSQKVRDLLDGRKGPKPNSLSPKRTKISTEEVNQDLNEVLSKLKG ILIHDDVPWRELAELGMLTLTHLYQDSASPCQKLLESYQLSGGPVDGLVE ALQALGVREGVRLLRDSQPREDKQSSDSTEDSGFGSQSIGEEMANPAVGNH
<i>Xenopus laevis</i> : NF-κB	MMSVLKIENFDPYSCNGIEDRNGMGYSTALLNPVILGQDLLMSYLSHIEQPK QGRGFRFRYVCEGSPSHRGLPGASSEKGGKTFPTVKIFNYVGMARIEVDLV THTDPPRVHAHSLVGKHSNKTGNCIVTVGPEDMTAQFNNLGIVHVTKK SQTEILKEKMKRNILRNTGRNTLTEVEERKIEQEVKDLKKVTDLSIVRL KFTAYLPDSNGAYTLALPPVISDPIHDSKSPGASNLISRMDKTAGSVKG GDEVYLLCDKVQKDDIEVQFYEDDENGWHAFGDFAPTVDVHKQYAIVFR TPPYHTQKIDRPVTVFLQLKRKKGGDVSDSKQFTYYPLEQDKEEVERK RRKDLPTFNNHFGGSGPMGGAPPGSSFGQGGGSNNYQYTMNSAFYM SSPAGGGYHSSGHMMKHCSATNSSEKNQQPSISIKKEGEEASACSQTDSATT AQKEAQCMIMRQANLRMLSLTQRTSRALLDYATTADPRMLLAVQRHLTA TQDENGDTPLHLAVIHGQSSVIEQLVQIILSIPNQILNMSNHLQQTPLHLGVI

	TKQYSVVAFLKAGADPTILDRYGNSVLHLAVQSEDDKMLGVLLKYPSVG QKNLINMPDYHGLSPVHWSVKMKNEKCLVLLVKAGANVNSAERKSGKSPL HIAVEMDNLNLAFLVKKLHADINAKTYGGNTPLHLAASRGSPMLTRMLV NEGANVLSENDEPVNKLPCNSDTSESDDVQMDTSDHHGSDSDTSSTAV DSECEHSAEEMHRREQRNIRPHCAMKRRYSGHTAVDLTKSQKVRDILSKHT PGSASWKQKGPEPVNLALETNTVQRLEKLLNEGQTGADWTELASRLRLQ SLVETYKNTSSPTESLLRNYELAGGNLKEINTLQSMGLNEGVELLCKSETY AKHHSPAESKND SAYESQSMEVDQSSGNLMDDSQKQTIPVSAAELCPTTEP TIGQ

Full-length amino acid sequences of NF- κ B proteins from the organisms used as input for MEME analysis to identify conserved motifs. NF- κ B sequences from *A. queenslandica* (Aq) (Gauthier and Degnan, 2008) were compared phylogenetically to *Aiptasia pallida* (Ap) NF- κ B (AIPGENE8848) from the ReefGenomics database, *Orbicella faveolata* NF- κ B (Williams et al., 2018), and *N. vectensis* (Nv), *Drosophila* Dif, Dorsal, and Relish, and *Homo sapiens* and Mouse p100, p105, RelA, RelB, and c-Rel from the UniProt database. *Carcinoscorpius rotundicauda* (Horseshoe crab) Relish, and *Danio rerio*, *Xenopus laevis*, and *Strongylocentrotus purpuratus* NF- κ Bs were from NCBI (Accession numbers: ABC75034.1, NP_001001840.2, NP_001081181.1, and NP_999819.1, respectively). All sequences were truncated to the Rel Homology Domain using MEME analysis and are shown in bold. The start of the RHD was highly conserved in all taxa and was predicted using the start of the known *Homo sapiens* RHD for both the NF- κ B and Rel proteins. The end of the RHD was highly conserved in all taxa and was predicted by using the MEME identified motif immediately before the predicted glycine-rich region. The trimmed and culled RHD sequence dataset allowed Clustal Omega to successfully align RHD sequences to be input into PAUP* for a maximum likelihood analysis bootstrapped 1000 times (Figure 4.1C).

Table 2.8 Organisms RHD sequences used for Chapter 5 phylogenetic analysis.

Organism	Predicted RHD Sequence from MEME analysis	Sequence obtained from: (Reference or database with accession number)
<i>Diaphanoeca grandism.</i> 53837	GVLGITVQPAPYARFRYLKEGRRAALQGRNNMY PTVRLSKEWMDKVVGDLVVPALVTRHDDASGE PKRHYHTLECVDDDGHLERVLDKAGFAVFDNMV VTRTPFKHAATVGLPHSQRDDQMVVRLMFSVAF QCTDGTRVQSRVISEPVYGSELQIDHVSFGWIPAD ESTSVTLTTHKVAKKRVAVVLTDTSQIPQTTELTE EAKSHGWNVVQGYPTLQIDNLLVHMQYALVGTI PPYWDQTITTA	Richter et al., 2018
<i>Diaphanoeca grandism.</i> 68592	GAIMIVVQPAEHGRFRYRKEGRKTSLOGRVEGRF PTVQVSPKFRQHIPPDDTMIQASVVTRHDDQNGKPI PHWHVLEGRNNTSPALPVKNGVVEFKNLVVNRN TFEVHGIRTRTHTDDQQVIRIMFEVHFIMDDNLVV SRVITEPVYGNELKIFKLSALHVASECETVMILTS KIKKRNTSLRLTDPTPMAYAHEPLGLGWTLDSQN RPMFNIKNLVYHHQHAIMATFPPFWNKALQTT	Richter et al., 2018
<i>Diaphanoeca grandism.</i> 73238	EGVLSISAQAPYGRFRYRKEGRRTPLQGRCDGLF PTVIVAPEWIDRASSPIVITPSLVTRSNDLEGEPIPH YHSLECIQENVQLERPIMHGVAVFHNLVVNRVPF KEAAVLGLPGSTRDDQMVRVLFTRFRCTDGSH VEARVISEPIFGSDLRIEHSVHAWVAVDEATSITLL TRKVAKKRMAIQITDSIQIHMLDTNAVAAGWFLE HGYPTVLIEDVDVHMQYACTIKIPPYWDQTITTA	Richter et al., 2018
<i>Acanthoeca spectabilis</i> <i>m.114880</i>	DGVLEISQQAAGHGRFRYSKEGRKTPLPGAIEGSF PTVQLSPAYRHMVPDGTMTVSVTITKYHDDRGA VRHWHTLEVKEGGAASRPLSMGLVQFPNLVTR AAMPADLDDAGAARNPEDQHVIRLMFTMMFRDS AGVIFKTSVISNPIYGIELKIHKSSHPRVPVAGQLD VFFLTSTKVKKNTLIKVREVYPAPFDPGPNSGWE LDERGRLTFTVDNLHVHYQYAVVATVPPYWDQT ITTS	Richter et al., 2018
<i>Acanthoeca spectabilis</i> <i>m.431975</i>	GVLRIVQQPASHGRFRYSKEGRKTPLHGREDGSY PTVAIADRYRHLVEEGTQVDVTLVTKHNDEHGSP IQHWHVLEGKEGGPVSRPLKDGVAATFPNLVVTRT TAEKEGGARNVEDQHVIRLMYTMRFQDEKRRSV LARAISEPIFGQEVKIHRIHQIPATGNIEVFFLTS KIKRKNTILTFTETTPSSFDPGPRSKWQLDKNMCL TYAMKDLTVHYQYAVVARVPPYWDQTITSP	Richter et al., 2018
<i>Acanthoeca spectabilis</i> <i>m.49642</i>	EVLEITEQPTERGRFRYAKEKRRTPLPGRRDGGFP TVRVADRYRDVLPDGTTHIQASVVTRQDGEDGIPR PHWHRLEGREGESVSQPLSLGAATFSNLVVIRSDR NAAENHWGPRPTEDQQVIRIMFSVLFRLPTGEMA RSWVASEPIYGCCLKIQNMSHTEVPMLSGAEVTL LTSKIRKQSIALLMLVDEFTDHDFAFGTKIDPANGW	Richter et al., 2018

	ECDQDGKICCTIQPSYVHHQYALVAQIPQYWDLT LES DR	
<i>Salpingoeca_punicam.1 0159</i>	SILVIVEQPAQFGRFRYEAEGRQNCLEGRSSDTFPT VAVNPKYAASIPDGT LVRVSLVKRYDDEQGNPVP HWHMLTSKDNGDTSQPLRNGRAEFPNLVVKRQK SSVRHSSDQRAIRLKFSISVCAGSVPREVHVISAP VFNADLKIDRLSHQYGYANEPTTVILLCTKVQKK TIALRISTDPSSDYMPPGGGWHQDPQCGGLVAVL RSDELETHHQA VIADIPAFVDPFTKEP	Richter et al., 2018
<i>Salpingoeca_macrocolla tam.102663</i>	EILAI AEQPAFGRFRYEAEGRQNCLEGRRPGSV AVVVNPKYAQIIPDGT PVHVSLVKRHDDEQGPV PHWHILSSRDGGDTSQLLAQGRAEFPGLVVKRQK SSVRHNSEDQRAVRLKFSIIVYCSNAPREIHVVSNP VFNAELKIDRLSHAQGHANEITNLIVLCTKVQKKT VALRIFDDVNSEFVPRDSGWQHD PNGCGCVCML RANELETHHQA VIADIPAYQDPCIKEPV	Richter et al., 2018
<i>Stephanoeca_diplocostat am.262780</i>	TNVL TILEQPT EHGFRFRYKKEKRRTPLNGRKEGTF PAVGLQGEWQH KVRDGTVIYASV VTRQND SRGIP MPHWHGLEGTGESATQRVVQGKATFSNLVVVR NERNAHRCAEDHQVIRIMFQVKFDDPETGEHYYS FVVSEPIYGT ELKIHNVSHRIATERETEVIFLTSKI KKQNI ALKLTDSAPNADWRPPTN QKDDGAGWHL DEQQKVAFL LQPLYVHHQYALVAKLP PHFFPRSK VC	Richter et al., 2018
<i>Stephanoeca_diplocostat am.557985</i>	GVLEITQQPAERGRFRYSKERRNTALHGREGSSP AVAVTAAYRHLVPEGTTVDVTVVTKDDDKNGNP VQHWHTLDGKNGDSIARPLDKDKQATFSQLVVT RGTC DRASSQKTSEDQQVIRLLFTMMFYTTEGRP AYARVLSDP IHGRDLKLHHISHTSVHVLGGSVIILT SKVKRRTVRLKITDMQPHSWRDGPIHTLQRTLDM RGWKLDEDNRP TIYLADLHVHYQYGLVVNMPPF WDQNITT	Richter et al., 2018
<i>Helgoeca_nanam.33607</i>	GVL RIMQQPASHGRFRYSKEGRKTPLHGRADGTF PAVAVAGRYRRLIEDGTHIDVALVTKHDDERGHP VQHWHTLVLEGKEGGPVSRLKSGVATFPNLVVTR TTAEKEGGARNVEDQHVIRLMYTM LFRDESGRL VQAQAISEPIYGLEVKIHRMSHMHIPATGNIEVFFL TSKIKRKNILKFTEVTPSAFDPGPRSMWQLDEDN CLTYSMKDLTVHYQYAVVVRVPPFWDQTITA	Richter et al., 2018
<i>Helgoeca_nanam.58794</i>	EGVLEISQQPAAHGRFRYSKEGRKTS LPGSIDGSFP TVQLTPAYRQLVPDGT MVAVSVITKYRDDD GAP VRHWHTLEVKEGGAASRPLSMG SVQFPNLVVTR AAMPTEPDDSGAAPNPEDQHVIRLMFTMMFRDST GAFFQSRIISDPIY GIELKIHKSSHPRVSVTGQLDVF FLTSKVKKKNTLIK VREVYPAPFDPGPHSGWELD DRGRLTFTVDNLHVHYQYAVVAPLPPYWDQTITT SRVLDVCLVDTA QGLESNCVQIEYCPSSGA	Richter et al., 2018
<i>Didymoeca_costatam.57 667</i>	DLLVILEQPAHYERFRYKENRSNCLGGRQDGSFP TVAVNPKHPHAEYLTNGTLIN VSLVKKDIKNGLE PHWHTLEGKDGPISRKITGGRAVFPNLVVKRERQ EEQQGRKVEKRLKEDTKALRLRFFADYYTSDGQ RHTASAVSDIVGNDEVKIIETSHHEGFESVKQRVL	Richter et al., 2018

	MLTAKFDKNHVVLITIDRNTRLDPQMKHLLQNE WVLDQENHPKYMFKQDKVFGHHQSGVAFNMPR YWRPLDAP	
<i>Savillea_parvam.25201</i>	GVLEIVEQPTERGRFRYAKERRRTPLPGRSEGSFP TVRVSKAYRDIVSDGVQVQASVVTRTDGEDGRP VPHWHRLEGRNGESVSQPIACGTATFCNLVVIRS DRNAYETYGGGAQPPLDAQVIRILFEISFRMPSPG ELARSWVTSDSIYGCDLKIQNMSHSEVSMSKGQE VILLTSKIKKQHINVMLVDSTTDHGYQPGSVVNA DNGWKCDDGKIYCMVQPTHVHHQYALVADLP PYWEQTKETR	Richter et al., 2018
<i>Savillea_parvam.44316</i>	GVLKIEQQPASHGRFRYSKEGRKTPHGRVDGSF PSVSVTPEYRSLIEDGTHVNVSLVTKNMDEQGGP VPHWHVLEGKEGGPVSRLTNGKATFPNLVVTRT TGEEGGGVRNAEDQHIVRLMYTMMFRDERGRLV QACASAPIYGLEVKIHHLSHTKVPATGDVDIIVLT SKIKKNTILRFTELTPSAFDPGPRSAWRLDEDNC LYFDMTDLIVHYQYAVVARVPPFWDPHTIAPR	Richter et al., 2018
<i>Savillea_parvam.68229</i>	DVLMISQQPAAHGRFRYSKEGRKTPHGRVDGSF PTVQLTPAYRHLVPDGTMSVSLITKYHDDHGNP VRHWHVLEVKEGGAASRPLTMGSVQFPNLVVTR AAMPTDPDESGAARNPEDQHIVRLMFTMVFRDPT GALFKTRVISDPIYGIELKIHKSSHPRVPVTGQLDV FFLTSKVKKKNTLIRVREVYPAPFDPGPNSGWELD DRGRLTFTVDNLHVHYQYAVVATVPPYWDQTIT TSR	Richter et al., 2018
<i>Salpingoeca_dolichothec atam.27377</i>	MKPFVMSFKTTDWLQWTSIIRLVFSYTGQPIKHW HVLVLKDGSKPQRRMKYHAVEFPDLVIRRRSTV KHSSDDQRSVRLRFALKGPNGENLAQVISRPIFNC DLKIDHLSHKSGPSSGGNDVILLVSKVRCGGGCV VCVCL	Richter et al., 2018
<i>Codosiga_hollandicam.6 61230</i>	AVLLIDEQPAQFGRFRYLAEGRQTCLEGVHPGTY PSIRVNPFAHLLPDGTMVHVKLVTFRDAPDGSP VDHWHVLTGKDETQVSQVLVNGRCTFRNLVVQR QKSDEKHDAADQRAVRLLFSAQYSIDNTTRTASA ISRPVFNADLKIDRLSHCNSPAQGGQTIFILCSKVQ KKTIALRISDCRYGSFVSKGNDWEFDSETQRHVIV LRADQLDTHHQQYVVAIEIPAYHNQSITAST	Richter et al., 2018
<i>Salpingoeca_helianthica m.70197</i>	RGAASDALEGVLVLAVAPSQFGRFRYEAEGRQN CVEGEEKGTFTVEIAPEWAPLCPDGTYYNVSLV RRDDFRVHHHILSSKDGGQTQQPLINRRATFSNLV VKRQRSEIRYPSEDQRAVRLLFVSFPRDGVITQA YAVSQPVFNAELKINRVSHKAGPMAAYTDVMLF CSKVQKKSGLITDNKHLGGIMPDSADAWMVT EGAHVFLVQHGDVHHQYGLAFKFPYFDQSHAA EGVS	Richter et al., 2018
<i>Salpingoeca_helianthica m.70211</i>	GVLVLAVAPSQFGRFRYEAEGRQNCIEGEERGTY PTVAIAQDWAALCPDGTIVTVSLVRRDDMRPHHH VLAAKDGGPTAQPLVGRATFSNLVVKRQRSDV RYPSEDQRAVRLLFVSFPRDGSVMVHAQVVSPIF NSDLKINRVSHKAGPMHAPTDMVLMFCSKVQKKS	Richter et al., 2018

	VGVLTDDQSLSFLEVADTWVRTADGGHLFLVQ HGIEVHHQYGLAFRFPFCFETARAGVSV	
<i>Salpingoeca_helianthica</i> <i>m.70222</i>	GILRLVVQPSQFGRFRYEAEGRQNCVEGEEKGTFP TVEIAPEWAPLCPDGTYYNVSLVRRDDFRVHHHI LSSKDGGQTQQPLINRRATFSNLVVKRQRSEIRYP SEDQRAVRLLFTVSFPRDGVITQAYAVSQPVFNAE LKINRVSHKAGPMAAYTDVMLFCSKVQKKS VGIL ITDNKHLGGIMPDSADAWMVTPEGAHVFLVQHGI DVHHQYGLAFKFPPYFDQSHAAEGVS	Richter et al., 2018
<i>Mylnosiga_fluctuansm.2</i> <i>4231</i>	GVILMDIEPAQFGRFRYEAEGRQNCLEGCEQNSFP TICVNPEWASYPDGTVISVSLVRRHDLKPHHHV LATKDGSDTRQVLRDGRASFPNLVVKRHRSAIRY SPEDQRAVRLLFTMTFSKDGSTDAACVSRPIFNS DLKITTCSHRSGPAHQATDAILLCSKVQKKSIAIRI AENMETDFTPEVEGWHQGPDGSWVCLIRSGFEV HHQYAIIFRVPPYHEALQVS	Richter et al., 2018
<i>Mylnosiga_fluctuansm.2</i> <i>58041</i>	GVIVMDVEPAQFGRFRYEAEGRQNSLEGEKGGFP AVRVSAEWAGLIPDGTIPINVTLIRRHDFKPHHHVL MSKDSTDIRQFLRNGRAAFPNLVVKRQRSALRYP PEDQRAVRLLFTMTFVKDGTLEAFVCVSKPIFNSD LKISGCSHRSGPANDWTDAILLCSKVQKKSIAIRIA ENTETDYVPPDAEGWQQAPDGSFVCFVFRAGFEV HHQYAI AFRLPPFHDQSISVSR	Richter et al., 2018
<i>Capsaspora owczarzaki</i>	DLLMVTEEPAQFARFRYMSEQRERSLAGENSFPT LMVNPKYARVVP EMALVTAVLVTKMPDPHTGR QQKHWHHLGGIPAAPLEGPQRIARFDNIAVIMDK ANNKDKDKSKAPVRSKDDQRCVRIMFELVFVSG NTQFYGRAISQPIYNAKLAIKISHSSGPVTGGNEV IMLCSKIRKGVTVRMTDPTQWSVQAPSGSAWEL NPQTLKADCNVPGANLFFHHQYAVVLTLPYHTQ TITAPVTVRISILDTDDETESQYVEYTYLPAAEA VR NA	UniProt Database (F1ARL5)

Amino acid sequences of NF- κ B-like proteins from the organisms used as input for MEME analysis to identify conserved motifs. The species used (first column), predicted RHD sequences from these organisms from MEME analysis (middle column), and where the sequence came from (last column) are shown. The trimmed and culled RHD sequence dataset allowed Clustal Omega to successfully align RHD sequences to be input into PAUP* for a maximum likelihood analysis bootstrapped 1000 times (Figure 5.1B).

CHAPTER THREE

THE ENDANGERED CORAL *ORBICELLA FAVEOLATA* HAS A CONSERVED TOLL-LIKE RECEPTOR-TO-NF- κ B SIGNALING PATHWAY³

3.1 Introduction

The Toll-Like Receptor (TLR)-to-NF- κ B signaling pathway is a prominent innate immune pathway in higher metazoans. Specific pathogen-associated molecular patterns (PAMPs) are detected by membrane-bound TLRs, which then initiate intracellular signaling cascades. One main TLR pathway leads to activation of transcription factor NF- κ B to induce changes in the expression of genes encoding innate immune effector molecules such as cytokines and anti-microbial peptides, (Akira et al., 2006; Kawai and Akira, 2007). TLR-to-NF- κ B pathways have been intensively studied for their roles in immunity in many model systems from flies to humans (Aderem and Ulevitch, 2000; Silverman and Maniatis, 2001; Vasselon and Detmers, 2002; Minakhina and Steward, 2006; Kawai and Akira, 2007). Recently, genome and transcriptome sequencing has revealed that many basal metazoans and some pre-metazoans also have homologs of the TLR-to-NF- κ B pathway (Gilmore and Wolenski, 2012). However, the biological roles of TLR and NF- κ B in these basal organisms are not well understood (Bosch et al., 2009).

In mammals, TLRs are single-pass transmembrane proteins with an N-terminal extracellular leucine-rich repeat domain (LRR), a central transmembrane domain (TM), and a C-terminal intracellular Toll/interleukin-1 receptor (TIR) domain. Ligand

³Adapted from Williams et al., 2018 *Developmental and Comparative Immunology*

recognition by the LRR domain promotes engagement of the intracellular TIR domain with other TIR domain-containing adapter proteins, which initiates downstream signaling cascades (Kawai and Akira, 2007). These adapter proteins include myeloid differentiation primary response protein 88 (MYD88), TIR domain-containing adapter protein (TIRAP/MAL), TIR domain-containing adapter inducing IFN β (TRIF), and Trif-related adapter protein (TRAM) (Aderem and Ulevitch, 2000; Kawai and Akira, 2007). The number of TLRs varies widely among organisms, with ten TLRs in humans, 13 in mice, nine in fruit flies, and over 200 in sea urchins (Vasselon and Detmers, 2002; Valanne et al., 2011; Buckley and Rast, 2012). In more basal organisms, such as sponges and cnidarians (which include hydras, jellyfish, sea anemones, and corals), there are two types of TLR-like proteins. Most often, these basal TLR-like proteins contain only a TM domain and a TIR domain; however, some basal animals have genes encoding full-length TLRs with LRR, TM, and TIR domains, similar to mammalian TLRs. Within the phylum Cnidaria, there are animals having no complete TLRs (*Aiptasia pallida*, *Hydra vulgaris*), ones having one complete TLR (*Nematostella vectensis*), and ones having multiple complete TLR and TLR-like proteins (*Acropora digitifera*) (Miller et al., 2007; Shinzato et al., 2011; Poole and Weis, 2014).

In insects and mammals, the NF- κ B superfamily comprises multiple related transcription factors that bind distinct DNA sequences known as κ B sites (Hayden and Ghosh, 2004). Based on DNA binding-site preference and sequence similarity, the NF- κ B RHDs are more related to each other than to the RHDs of Rel proteins (Siggers et al., 2011; Finnerty and Gilmore, 2015). Where characterized, basal organisms (e.g.,

cnidarians, sponges, pre-metazoans) have single NF- κ B family proteins, which are phylogenetically most similar to the vertebrate and insect NF- κ B subclass (Finnerty and Gilmore, 2015). In the mammalian non-canonical NF- κ B signaling pathway, I κ B kinase (IKK)-mediated phosphorylation of p100 at three serine residues located C-terminal to the ANK repeats promotes its processing to p52 by the proteasome (Sun, 2011). On the other hand, the *Drosophila melanogaster* NF- κ B protein Relish is activated by a discrete caspase-mediated proteolytic event that removes the C-terminal ANK repeat domain (Stöven et al., 2003; Valanne et al., 2011). Overall, it is not known how NF- κ B proteins are regulated and activated in organisms basal to *Drosophila*.

The mountainous star coral *Orbicella faveolata* (Of), previously known as *Montastraea faveolata*, is an endangered reef-building coral found in the Caribbean Sea and the Gulf of Mexico. *O. faveolata* forms a symbiotic relationship with a dinoflagellate of the genus *Symbiodinium* (Steele et al., 2011; Davy et al., 2012). Like most reef-building corals, *O. faveolata* is susceptible to an environmentally induced loss of symbiosis, commonly referred to as “bleaching” because this process gives the coral tissue a white appearance (Gleason and Wellington, 1993; Brown, 1997; Hoegh-Guldberg et al., 2007; Hughes et al., 2017; Weis, 2008). Bleaching often causes coral death, but in some cases, corals can recover from a bleaching event and re-establish a symbiotic relationship with *Symbiodinium*. Nevertheless, recovered corals often show increased susceptibility to microbial diseases such as yellow band disease, black band disease, and plague (Kushmaro et al., 1996; Pinzón et al., 2015).

Recent reports of transcriptional changes in immune-related molecules in

bleached and pathogen-infected corals have suggested that the cnidarian innate immune system plays a role in coral diseases (Pinzón et al., 2015; Anderson et al., 2016; Fuess et al., 2016, 2017; Zhou et al., 2017). To gain deeper insights into molecular processes important for coral immunity and health, we have had an ongoing interest in characterizing cnidarian homologs of mammalian immunoregulatory molecules and pathways (Wolenski et al., 2011, 2013). In this chapter, we use phylogenetic, biochemical, and cell-based assays to characterize the structure, activity, and regulation of TLR and NF- κ B proteins from *O. faveolata*. We also show that lipopolysaccharide (LPS) treatment of *O. faveolata* tissue can induce a gene expression profile consistent with induction of the NF- κ B pathway. These results represent the first characterization of proteins in the conserved immunoregulatory TLR-to-NF- κ B pathway of a critically endangered coral.

3.2 Conservation of Toll-like Receptor (TLR) pathway in *Orbicella faveolata*

To determine whether proteins in the vertebrate TLR-to-NF- κ B innate immune pathway (Silverman and Maniatis, 2001) are present in the coral *O. faveolata*, we scanned transcriptomic data (Pinzón et al., 2015). Homologous transcripts of many of the receptor and downstream signaling components of the mammalian TLR pathway were present in *O. faveolata* (Figure 3.1). Nevertheless, in many cases, there were reduced numbers of signaling components in each family (Table 3.1). For example, there are five NF- κ B/Rel proteins in humans, but there was a single NF- κ B-like protein in *O. faveolata*. Similarly, there are ten complete TLRs in humans, but in *O. faveolata* there was only a

single predicted full-length TLR (Figure 3.2) and four TLR-like proteins containing only TM and TIR domains.

3.3 A predicted TLR-to-NF- κ B pathway in *Orbicella faveolata*

From transcriptomic analysis, the single full-length Of-TLR is predicted to contain *bona fide* LRR, TM, and TIR domains. By phylogenetic analysis using the neighbor-joining method, the TIR domain of Of-TLR was most closely related to the TIR domains of the single characterized *N. vectensis* (Nv) TLR (Brennan et al., 2017) and to the *D. melanogaster* Toll protein. Among the human TLRs, Of-TLR appeared to be most similar to the TIR domain of human TLR4 (Figure 3.3A).

Since the TIR domain of human TLR4 interacts with the intracellular adapter protein MYD88 as an early step in signaling to NF- κ B (O'Neill et al., 2003), we sought to determine whether the TIR domain of Of-TLR (Of-TIR) could also interact with human MYD88. We incubated bacterially expressed GST-Of-TIR with 293 cell extracts overexpressing FLAG-MYD88, and then performed a pull-down assay (Figure 3.3B). As shown in Figure 3.3C, MYD88 was detected by Western blotting in pull-down fractions containing GST-Of-TIR, but MYD88 was not seen with GST alone. As a positive control, we showed that MYD88 also interacted with GST-Nv-TIR, which we have recently shown can interact with human MYD88 (Brennan et al., 2017). Thus, the phylogenetic similarity of Of-TLR and human TLR4 proteins was reinforced functionally by Of-TLR's ability to interact with the human TLR4 adapter protein MYD88.

Having established that Of-TLR had similarities to human TLR4, we sought to

determine whether *O. faveolata* tissue would elicit a response to gram-negative lipopolysaccharide (LPS), an activator of the mammalian TLR4-to-NF- κ B pathway. Therefore, we treated *O. faveolata* tissue for 30 min with LPS, and followed that treatment with a 4-h washout period with seawater. RNA was then extracted from treated tissues and subjected to RNA-seq analysis. Ingenuity Pathway Analysis (IPA) of the RNA-seq data identified 39 pathways in *O. faveolata* that were significantly activated (as compared to control tissue), and 13 of these pathways had numerical z-scores. Of these 13 pathways, NF- κ B signaling was the most significantly activated pathway (IPA: z-score = 1.8, $p = 0.03$) (Fuess et al., 2017). A total of 16 genes in the NF- κ B signaling pathway were expressed in our samples, resulting in a $9.3E^{-02}$ pathway ratio (Table 3.2). Among the NF- κ B pathway components activated by LPS treatment were the following: TLR1- and 2-like proteins; members of the tumor necrosis factor receptor associated factor family (TRAF2, 3, 5), which are proteins involved with regulation of the pathway; kinases including RIPK1, MAP3K3, and PIK3R4; and the adapter protein MYD88. Nearly all genes in the NF- κ B signaling pathway were up- or down-regulated in a manner that was consistent with activation of the pathway as it known in mammalian systems (Table 3.2).

3.4 The structure and phylogenetic analysis of the *O. faveolata* NF- κ B protein indicates that it is more similar to mammalian NF- κ B proteins than to Rel proteins

Because IPA predicted that LPS could activate the NF- κ B pathway in *O. faveolata*, we sought to characterize NF- κ B-like proteins in *O. faveolata*. By scanning an

O. faveolata transcriptomic database (Pinzón et al., 2015), we identified transcripts encoding a single NF- κ B-like protein, but no Rel-like protein. The amino acid sequence of the Of-NF- κ B protein (Figure 3.4) indicated that it had an overall domain structure that was most similar to mammalian NF- κ B proteins, i.e., with an N-terminal RHD, followed by a glycine-rich region, and then a series of ANK repeats (Figure 3.5A). A maximum-likelihood analysis of Of-NF- κ B also indicated that it was most similar to other invertebrate NF- κ B proteins, most notably those from sea anemones (*N. vectensis*, *Aiptasia* sp., and *A. tenebrosa*) (Figure 3.5B). Moreover, Of-NF- κ B clustered with NF- κ B proteins from *D. melanogaster* (Relish) and humans (p100 and p105), and was distinct from the Rel proteins. This result is similar to the clustering of Nv-NF- κ B with NF- κ B proteins, but distinct from Rel proteins, as reported by Sullivan et al. (2007).

3.5 C-terminal truncation of the *O. faveolata* NF- κ B protein is required for nuclear localization, DNA binding, and transactivation

The mammalian NF- κ B proteins p100 and p105 require removal of their C-terminal ANK repeat domain in order to become active transcription factors. Due to the structural similarity of Of-NF- κ B to human NF- κ B p100 and p105, we sought to determine whether C-terminal truncation of Of-NF- κ B would unveil active transcription factor properties. Therefore, we created expression vectors for FLAG-tagged codon-optimized versions of full-length Of-NF- κ B, an N-terminal form (Of-RHD), and a protein containing only the C-terminal glycine-rich and ANK repeat domain (Of-Cterm) (Figure 3.6A). As a control, we used an expression vector for the 440 amino acid *N. vectensis*

NF- κ B protein (Nv-NF- κ B), which we have characterized previously as an active NF- κ B protein (Wolenski et al., 2011). Each of the Of-NF- κ B constructs directed the expression of an appropriately sized protein when transfected into HEK 293 cells (Figure 3.6B).

To determine the subcellular localization of the Of-NF- κ B proteins, we transfected the plasmids into chicken fibroblasts and later performed anti-FLAG indirect immunofluorescence. We found that full-length Of-NF- κ B and Of-Cterm were present in the cytoplasm of transfected chicken fibroblasts, whereas Of-RHD and the control Nv-NF- κ B protein were both located in the nucleus of these cells (Figure 3.6C).

We next analyzed the DNA-binding and transactivation properties of the full-length and truncated Of-NF- κ B proteins. To measure DNA-binding activity, 293 cells were transfected with the Of-NF- κ B expression plasmids, and extracts were analyzed for DNA-binding activity by an electrophoretic mobility shift assay (EMSA) using a κ B-site probe that we have previously shown can be avidly bound by Nv-NF- κ B (Wolenski et al., 2011). Lysates from cells transfected with expression plasmids for Of-RHD and the positive control Nv-NF- κ B both contained proteins that strongly bound to the κ B-site probe (Figure 3.6D). In contrast, extracts from cells transfected with the empty vector control, Of-NF- κ B or Of-Cterm showed only low or background κ B-site DNA-binding activity (Figure 3.6D). In addition, Of-RHD and Nv-NF- κ B both activated transcription of an NF- κ B-responsive luciferase reporter plasmid when overexpressed in 293 cells as compared to empty vector control-transfected cells (Figure 3.6E). As expected, full-length Of-NF- κ B and Of-Cterm did not activate the reporter above vector control levels (Figure 3.6E). Taken together, these results showed that removal of C-terminal ANK

repeat domain sequences of Of-NF- κ B enabled the protein to enter the nucleus, bind to DNA, and activate transcription in vertebrate cells.

To further demonstrate that Of-NF- κ B was a functional transcriptional activator, we assessed its ability to activate transcription in yeast, which do not have endogenous NF- κ B proteins. To accomplish this, we fused the Of-RHD to the GAL4 DNA-binding domain (GAL4-Of-RHD) and expressed this protein in yeast cells, which contain an integrated GAL4-site *LacZ* reporter locus. In these cells, GAL4-Of-RHD activated transcription of the GAL4-site reporter substantially (~135-fold) above control (GAL4 alone) levels. However, the GAL4-Cterm fusion protein did not activate transcription of the reporter in yeast (Figure 3.6F). As we have shown previously (Alshanbayeva et al., 2015; Wolenski et al., 2011), a GAL4-Nv-NF- κ B fusion protein strongly (~240-fold) activated transcription in yeast as compared to the GAL4 alone control. Therefore, the ability of the Of-RHD sequences to activate transcription appeared to be conserved from yeast to human cells, and appeared to be an intrinsic property of sequences within the N-terminal half of the protein.

3.6 C-terminal truncation of Of-NF- κ B can be induced by an IKK-dependent mechanism in human cells

Upon activation, the human NF- κ B p100 protein is converted to its active p52 form by proteasomal processing of C-terminal sequences up to the glycine-rich region (Sun, 2011). This processing is initiated by IKK-dependent phosphorylation of a cluster of three serine residues located C-terminal to the ANK repeats of p100 (Sun, 2011). The

Of-NF- κ B protein contains three serine residues with similar spacing and flanking sequences to the IKK target serines in human p100 and in an NF- κ B protein from the sea anemone *A. pallida* (Figure 3.7A). In an *in vitro* kinase assay (Figure 3.7B), a constitutively active form of human IKK β and the predicted Of-IKK protein (Figure 3.8) could each phosphorylate a bacterially expressed GST-fusion peptide containing the C-terminal serine residues of Of-NF- κ B (amino acids 843-874); however, neither of these IKKs phosphorylated the analogous GST-fusion protein containing alanine substitutions for the three serine residues or GST alone (Figure 3.7B).

To determine whether IKK-dependent phosphorylation could induce processing of Of-NF- κ B, we transfected human 293T cells with expression vectors for FLAG-Of-NF- κ B and either FLAG-Of-IKK or HA-Hu-IKK β , which we have recently shown can induce processing of the sea anemone *A. pallida* NF- κ B protein in similar experiments (Mansfield et al., 2017). As shown in Figure 3.7C, co-transfection of 293T cells with FLAG-Of-NF- κ B and either FLAG-Of-IKK or HA-Hu-IKK β resulted in the appearance of an Of-NF- κ B band of approximately 50 kDa. No such smaller band was seen when the Of-NF- κ B-3X-Ala mutant was co-transfected with FLAG-Of-IKK or HA-Hu-IKK β . Together, these results indicate that Of-NF- κ B can undergo IKK-induced processing in human cells, and that this processing requires conserved sites of IKK-directed phosphorylation.

3.7 Chapter 3 summary

The endangered, reef-building coral, *O. faveolata* expresses homologs to a conserved TLR-to-NF- κ B signaling pathway. Tissue exposed to LPS, a potent vertebrate activator of NF- κ B, expresses a gene expression profile consistent with activation of the NF- κ B pathway. Furthermore, the structure and phylogenetic relationships of Of-TLR suggest it is most homologous to Hu-TLR4, and structure and phylogenetic analyses Of-NF- κ B suggest it is most related to the NF- κ B proteins rather than the Rel proteins. Pull-down assays of Of-TIR domain reveal that it can bind MYD88 *in vitro*. Cellular and molecular analyses of Of-NF- κ B suggest that it is most homologous to human p100 protein, in that it requires C-terminal truncation to enter the nucleus, bind DNA, and activate transcription, and that this truncation can be induced by IKK-mediated processing of conserved C-terminal residues. These results suggest that *O. faveolata* contains homologs to a TLR-to-NF- κ B pathway and that LPS induces activation of the NF- κ B pathway.

Table 3.1 Conserved TLR signaling pathway components in *Orbicella faveolata*.

Human Gene Name	Present in TLR or TNF	Present in <i>O. faveolata</i>	Contig ID in <i>O. faveolata</i> transcriptome (Pinzón et al., 2015)
A20	TLR/TNF	+	comp268115_c0_seq1
Akt	TNF	-	
CD40	TNF	-	
cIAP1/2	TNF	-	
Cot	TNF	-	
CYLD	TNF	+	comp268162_c0_seq1 comp268162_c0_seq2
BR3	TNF	-	
ECSIT	TLR	+	comp253068_c0_seq1
HOIL	TNF	-	
HOIP	TNF	-	
I κ B α	TLR/TNF	-	
IKK α/β	TLR/TNF	+	comp262681_c0_seq3 comp262681_c0_seq2 comp262681_c0_seq6
IKK γ /NEMO	TLR/TNF	-	
IRAK-1	TLR	-	
IRAK-4	TLR	-	
IRAK-M	TLR	-	
ITCH	TNF	-	
LT β R	TNF	-	
MyD88	TLR	+	comp261294_c0_seq1
NF- κ B	TLR/TNF	+	comp266589_c1_seq2 comp266589_c1_seq4

			comp266589_c1_seq6 comp266589_c1_seq7 comp266589_c1_seq8 comp266589_c1_seq14 comp266589_c1_seq18 comp266589_c1_seq19 comp266589_c1_seq22
NIK	TNF	+	comp268939_c0_seq3 comp268939_c0_seq5
PDK1	TNF	+	comp256479_c0_seq1
PI3K	TNF	-	
RIPK	TLR/TNF	+	comp267015_c3_seq12 comp260996_c1_seq5 comp260996_c1_seq13
RNF11	TNF	+	comp263654_c0_seq1
SHARPIN	TNF	-	
TAB1	TLR	+	comp253554_c0_seq1
TAB2	TNF	-	
TAB3	TNF	-	
TAK1	TLR	-	
TAX1BP1	TNF	-	
TIRAP	TLR	-	
TLR1	TLR	+	comp265379_c0_seq1
TLR2	TLR	+	comp209859_c0_seq3 comp209859_c0_seq2 comp264804_c0_seq1 comp264804_c0_seq3 comp264804_c0_seq4 comp264804_c0_seq5 comp258786_c0_seq19 comp262700_c0_seq1 comp262700_c0_seq2

			comp260489_c0_seq2 comp260489_c0_seq3 comp260489_c0_seq4 comp269333_c1_seq3
TLR4	TLR	+	comp267159_c0_seq1
TLR5	TLR	+	comp249146_c0_seq2
TLR6	TLR	+	comp257063_c0_seq2 comp269420_c0_seq1 comp269420_c0_seq2 comp269420_c0_seq3 comp269420_c0_seq4 comp261273_c0_seq6 comp261273_c0_seq2 comp261273_c0_seq7
TLR11	TLR	-	
TNFR	TNF	+	comp171719_c1_seq1 comp268242_c0_seq3 comp268242_c0_seq15 comp266622_c0_seq2 comp266622_c0_seq3
TOLLIP	TLR	+	comp267143_c1_seq4
TRADD	TNF	-	
TRAF2	TNF	+	comp263023_c0_seq1
TRAF3	TNF	+	comp267070_c0_seq3 comp267070_c0_seq7 comp267070_c0_seq8 comp267070_c0_seq9 comp267070_c0_seq11 comp267070_c0_seq17 comp267070_c0_seq22 comp267070_c0_seq25 comp267070_c0_seq27 comp269526_c0_seq3 comp269526_c0_seq8

			comp268663_c0_seq7 comp268663_c0_seq15 comp268663_c0_seq20 comp268663_c0_seq21 comp268721_c0_seq1
TRAF5	TNF	+	comp262923_c1_seq1
TRAF6	TLR	+	comp268969_c1_seq2 comp268969_c1_seq3 comp268969_c1_seq1 comp268969_c1_seq4 comp267923_c1_seq10 comp267923_c1_seq11 comp245176_c0_seq1 comp245269_c0_seq1 comp267813_c1_seq1 comp260799_c1_seq14 comp258243_c1_seq1 comp249176_c0_seq11 comp249176_c0_seq13 comp249176_c0_seq14 comp249176_c0_seq16 comp249176_c0_seq21 comp262285_c0_seq1 comp269073_c0_seq1 comp269073_c0_seq2 comp269073_c0_seq4 comp269073_c0_seq5 comp269073_c0_seq7 comp269073_c0_seq8 comp294316_c0_seq1 comp262894_c1_seq1 comp249176_c0_seq1 comp267764_c1_seq3 comp267764_c1_seq7 comp263153_c0_seq2

			comp263153_c0_seq3 comp263153_c0_seq4 comp263153_c0_seq5 comp263153_c0_seq11 comp263153_c0_seq12 comp263153_c0_seq13 comp263153_c0_seq14 comp263153_c0_seq16 comp263153_c0_seq17 comp258243_c1_seq3 comp248322_c0_seq5 comp254506_c0_seq3 comp254506_c0_seq4 comp254420_c2_seq1 comp254420_c2_seq6 comp254506_c0_seq2 comp254506_c0_seq5 comp267764_c1_seq15 comp268258_c0_seq20 comp268258_c0_seq27 comp268258_c0_seq4 comp254420_c2_seq3 comp254420_c2_seq4 comp248322_c0_seq1 comp248322_c0_seq4 comp266296_c0_seq1 comp254420_c0_seq2 comp263924_c0_seq1 comp265647_c0_seq8
TRAM	TLR	-	
TRIF	TLR	-	
Ubc13	TLR/TNF	-	
UEV1A	TLR	-	

An overview of human proteins that have homologs based on the *O. faveolata* transcriptome from Pinzón et al. (2015). Listed are contig IDs for each sequence matching the corresponding protein, which can be searched in the supplemental material from Pinzón et al. (2015). Full reads can be found on the National Center for Biotechnology Information Short Read Archive under the SRP022773 accession number.

Table 3.2 NF- κ B pathway genes differentially expressed following LPS treatment on *O. faveolata*.

Gene	Number of Contigs	Average Log ₂ Fold Change	Standard Error	Function	Activation Status
TRAF2	1	2.7	N/A	Enzyme	Consistent
RIPK1	1	2.6	N/A	Kinase	Consistent
FGFRL1	1	1.8	N/A	Transmembrane receptor	Consistent
FGFR4	1	1.4	N/A	Kinase	Consistent
TRAF3	8	0.8	0.43	Enzyme	Consistent
TLR2-like	4	0.6	1.00	Transmembrane receptor	Not predicted
TRAF5	9	0.5	0.30	Transporter	Consistent
FGFR1	3	0.4	0.01	Kinase	Consistent
MAP3K3	3	0.4	0.51	Kinase	Consistent
BMP2	2	0.3	0.87	Growth factor	Consistent
MALT1	3	0.2	0.53	Peptidase	Consistent
IGF2R	1	-0.3	N/A	Transmembrane receptor	Not predicted
TLR1-like	2	-0.6	0.97	Transmembrane receptor	Not predicted
TDP2	2	-0.7	0.60	Transcription regulator	Inconsistent
MYD88	2	-0.7	0.24	Other	Inconsistent
PIK3R4	2	-1.4	1.08	Kinase	Consistent

O. faveolata tissue was exposed to *E. coli* LPS for 30 min and then returned to seawater for 4 h. RNA-seq was performed to identify transcripts whose expression was affected by this treatment, and Ingenuity Pathway Analysis was then used to identify pathways whose gene expression patterns were significantly changed by LPS treatment. The “NF- κ B pathway” was found to be the most significantly activated pathway overall. Listed are 16 *O. faveolata* genes with homologs in the NF- κ B pathway whose expression was affected by LPS treatment. Average Log₂-Fold Change is indicated as well as associated error, encoded protein function from IPA analysis (Function), and consistency with pathway activation (Activation Status). The complete RNA-seq dataset for differently

expressed genes can be found in Fuess et al. (2017).

in *O. faveolata* are in yellow. TLR1, 2, 5, and 6 homologs in *O. faveolata* are TIR domain-only proteins.

Original Sequence	Codon Optimized	Amino Acid Sequence
ATGTTTGTGAAATCACAAACAGCACTAATTCAATCAAATCTACACATCTTGGTCCTTGATCAACGTCACACAGTC	-----T--AGT--G--G--T--T--C--G--T-----G--T--C-----G-----A-A--T--T--	M F V K S Q T A L I Q S N L H I L V L D Q R H T V
CTGCTCGCCACATCTCTTGTGTTTCGACACGAAAGGCTCTCTGATTATATTCAGCTGCTTGGTTCTCTTTGCGATC	--CT-G--G--G--A--C--C--T--T-----G-----C--A--C-----A--T-----A--T	L L A T S L V F D T K G S L I I F S C L V L F A I
GAGGCTACTGGACAACAACGCAGAAGGTTTCTACACTGCGGTGCATCTTTCTGCCTGTTGAAAAGGTCTATGTCCG	-----G--G--C-----G--C--C-----CT-G--T--T--C--GAG--T-----C-----GC--AG---AGT	E A T G Q Q R R R F L H C G A S F C L L K R S M S
TCCTTCTTCGAAACAGGCAATGGTACTCTTATCAAAGAGCAAGCGGTAATATGTGTCTTGAAAAGCGACAAGGTT	-----T--G-----A-----A--C--C--T--G--A-----T--T--C-----AC-C--G--T--T-----	S F F E T G N G T L I K E Q A V I C V L K S D K V
CGCGACCGATGTGTCTGTGGACATCAGTCTGATATTGAAAACGATAACAACGCCTCAAGATGTCGTATTGCATTTT	A-G--T--G--C--G--A--T--ATCCT---TC-----C--C--T--A-----C--T--C-----	R D R C V V D I S L I L K T I T T P Q D V V L H F
GCTGCCGTGTGTTTAACCCCGATGGAAATTGCCTTCTACAACCTCTTTAAACGCCACAAAGAAAAACGCCATATTT	--C--A--A--G--A--C-----G--C--A--T-----T-----G--T--A-----G-----	A A V C L T P M E I A F Y N S L N A T K K N A I F
TACTTGCAAATCAAGGGACACTGCAGCTTTTCAGCCGATGGCATCTCTCACTGGGGAAAAGCTACAGACTTTTCGG	--TC-C--G-----T--TTC--C--C--G--C--G-----C--T-----C--G--C--C-----C--	Y L Q I K G H C S F S A D G I S H W G K A T D F R
GTTTTCTATCTCATGGAAACTCGACTTTACTAGAAGGAAATAAAACGCTGGCAAACAACCTCTCGTCGAGCATTG	--A-----T-----TAGC---C-T--G-----C-----T--C--T--AG--G-----T--	V F Y L M E N S T L L E G N K T L A N N S R R A L
GAAAACATTGGCACTCTTATGATAGACAAATCTAAACTACGAACCTTCCAAAAATGTTTAGTAGTACTAAAGTG	--G-----A--G--T--G-----C--T--GAGC-----T-----A--C--G--G-----CTC--TCC--G--G--C	E N I G T L M I D K S K L R T L P K M F S S T K V
TGGCCAAGAATGGCCGAGGTAGTATTCTCAAATTACAACCTCACATCGATACCTCCGGAGCTCAATACGACGATG	-----CC-G-----A-----C--C-----C-T--G--G--AGC-----A--C-----T--C--C--T--	W P R M A E V V F S K L Q L T S I P P E L N T T M
CCGTTCCTTGCACTCGCTTGAGCTTGCTAACAATAAACTAACAACACCACCTCCGTTTCCATGGTGTAAGCTACG	--C--TC-C-----CT-G-----C--C--T--C--G--C--C--T--G--A--A-----C--G-----	P F L Q S L E L A N N K L T T P P P F P W C K A T
CTCAAACCTGCCAAGAGGACTGCAAAGAATCCAACAGGAAATCATCACTACCAATTCGGCACTAACGTGCGTCCA	T-G---T---CC-G--C-----GC-----T--C--C-----T--T--G-----T-----T--T--G---	L K L P R G L Q R T P T G N H H Y Q F G T N V R P

AATATCTACCGACGATTTCTTGATTTATCGTACAACAACATCGAAGATTTATCGACGCACAACTTTAGAGGATTT
--C--T-----A-G--T-G--C-C--A--T-----T-----CC-C--C--T-----T--CC-----C--C
N I Y R R F L D L S Y N N I E D L S T H N F R G F

CTGAACAAGTTGACTCTCGAGGGAAACGGCCTAAAAGTCATTGGAACATCGTGTTTTCGCAACTTGAAAGGAATA
--T--T--C-C--A-G-----C-----T-G-----G-----T--GAGC-----C-----C-C-----T--T
L N K L T L E G N G L K V I G T S C F R N L K G I

CATGTGATCTCCCTCAGCAAAAAACAAGTTGAAAAGTTTACCATCAGAGCTGTTCCAAGGACAAGATAGTTTACTA
--C-----T--T-----AC-C--TC-C-C--C-----A-----T-----C--G--CTCA--G--G
H V I S L S K N K L K S L P S E L F Q G Q D S L L

GAATTACGACTTGACCACAATAACATTTTCGATAATACCAAACGATCTCTTCAAAACAGTGACCCAAATCAAAAGA
----G--T-G--T--T--C-----AAGT-----C--G--T--C--T--T--G-----A--A-----A--GC--
E L R L D H N N I S I I P N D L F K T V T Q I K R

ATTGACTTGCATAGCAACAACTGAGCTGTATTCCACAAGAATTGTTTCAGCAAGTTAAAAAATATCAAAATTCTA
--A--TC-----G--C-----C--C--G-----TTCT--AC-G-----T--G-----G
I D L H S N K L S C I P Q E L F S K L K N I K I L

CATTTAGAAGACAACCACATCACGCAAGTACACGACAAAGCTTTCTCCATAGATTCAAGTTCATTACAAAACATT
--C-T--G--T--T--T-----A-----G--T-----G--A--T--T-----AGCTCAAGCC-C-----T--
H L E D N H I T Q V H D K A F S I D S S S L Q N I

TATCTCCAAAAGAACAAAATAAGCCGATTCCATTAACTACTACTGCAACGGCACGCTGTTAAATTTGATTTG
----G-----A-----TTC-----TC-T--TC-CT-G-----G--A--T--G--C--G--A--CC--
Y L Q K N K I S R I P L T L L L Q R H A V K I D L

TCTTTTAATCAGTTGACGTTTCAAGACTTTAACAGGCTGATACAAGAATTAGACTTGGAGACATTTCTGTATCAC
--A-----AC-----A--C-----TC--T-----C--G--GC--T--TC-----T--C--C-----
S F N Q L T F Q D F N R L I Q E L D L E T F L Y H

CATCGCCATACGGCGAGCTCTTCTCAAATGAGACTGCAAGAAAGTTTAAAGTCTATCAGCTTCGCTCACAAACAG
--A-G-----C--TAGC--C--G--C--T-----G--TCCC-C-----C--T--T--T--G--T-----
H R H T A S S S Q M R L Q E S L K S I S F A H N K

TTCACCACGATCAACATAGAAGCATTCAACAGAACAGAGGAACCTAACGTTTCGAGTATCTTCTGCGGGTTTACGAG
----T-----A--T--T--G--T-----C-----C--A--G--G--A-----C--C--TA-----A--T--
F T T I N I E A F N R T E E L T F E Y L L R V Y E

ATAGACATGTCGGGGAATCTCCTCCTATGCGACTGCAAGATTCTGCTTCTGAGTCGCTGGCTTCGAGCTTTAGTT
--T--T-----T--T-----T-GT-G--T-----A-----CA-G-----GA-----C-C--C
I D M S G N L L L C D C K I L L L S R W L R A L V

CAAAGGCACACGCGCATCCGAAATGAACAGTTTCAAACCTTGAAATGCGCCGCTCCGACTGAACCTCAAAGGGAAA
--C-C-----A-G-----C--C-----G--G-----T--A-----G--T--G--A--
Q R H T R I R N E Q F Q T W K C A A P T E L K G K

CCGATCCTATCCGTGGACGAAAAATCGATTTAAATGCCAGAGAAATCTAGAAAACCTGCCCCACGGATGCTTGTGT
-----T-G--T-----A-----C--G--T--A--G--T--G--G-----T--T--T-----
P I L S V D E N R F K C Q R N L E N C P H G C L C

CTCGTGCGTGCTTTAGACGGTACTGTGGTCATCGATTGTAAAGGAAGAAACCTTACCGCAATACCTCCCAAAGTT
 -----TA-A--AC-G--T--A--C-----T--A-----TC-----T-G--A--T--T--G-----G--G
 L V R A L D G T V V I D C K G R N L T A I P P K V

CCAAGCGGTCGCATAGAGCTTAAGTTAGAGGACAACAACATCCGGGAGATTCCACCATATCCTTATATGGAAAAAT
 --CTCT-----A-----A-----AC-T--A--T--T-----A--C--A--A-----C--A--C-----
 P S G R I E L K L E D N N I R E I P P Y P Y M E N

GTGACGGCTCTGTACTTAACGCATAACAAGATACAAGTATTAAACAAGTCCACGGTACGAAGGTTTACGCGAATT
 --T--T--G--C-----G--A-----T--A--C--G--C--G--T--AAGT-----GA-G-----C-----C---
 V T A L Y L T H N K I Q V L N K S T V R R F T R I

AAAGTTTTGTTTCATCGATTCTGAACAAACTCACCTATTTGCCTAAGAATATCGAAAATTTAAACTTTACGAGTCTT
 -----CC-C-----T--AGT-----G--G--G-----A--A--C-----G--T-----TCA--C
 K V L F I D S N K L T Y L P K N I E N L N F T S L

GCGTTACACCACAATTTCTTCAAATGCGACTGTACGACATTATGGATAAAGCACTGGCTACAGAGAAAACAAAGT
 ---C-C-----C-----T--G--T-----C--C--CC-G-----C-----T-----G--A--G--G--GTCC
 A L H H N F F K C D C T T L W I K H W L Q R K Q S

AAAATATTACACATTAAAAATGTTCTTTGCAACTCCGAGGGAAGTACGCAGGGAAAAGCAATTTACACTCTGCCA
 --G---C-T-----A--G--C-----G--T-----C-----C--A--G--G-----C--T--AT----T
 K I L H I K N V L C N S E G S T Q G K A I Y T L P

AATGAAGAATTTGTTTGCAAAAAGAATAAGAAAGACATCCCTACTACACAATCTATAACAAAGGACAAGACTTTC
 -----C-----G-----A-----T-----C--A--G--GAGC--T--G--A--T--A--A--T
 N E E F V C K K N K K D I P T T Q S I T K D K T F

AAAATAATCGCGTTGACTCTTGAGGCGCGCTGGTTTTGACTTTTCATCGCTTTTATCGTGGCTTACAAATACCGC
 --G---A--AC---A--C--C--G--T--C--C--T-----C--C-----C-----G--T---
 K I I A L T L G G A L V L T F I A F I V A Y K Y R

GGAGAAATGAAAGTCCTCATGTACACTCATTTTAACTGGCACCCATTTGACCGTGTAGATGACTCGGATCCAAGA
 --T-----TT-G-----G--C-----T-----C--T--C-----C--C--TC-C
 G E M K V L M Y T H F N W H P F D R V D D S D P R

AAAATATACGACGCGTTTCGTATCGTACAGCGGGAGTGATCACCAGTGGGTGTCAACACCTTGCAAGAGCGACTC
 -----T--T-----T--T--CAGC--TTC--ATC--C--T--A-----C--G--T--TC-T--G--A--G---
 K I Y D A F V S Y S G S D H Q W V V N T L Q E R L

GAACATCACGACCCCCCTTACAAATTGTGCATTACCACCGCGACTTCGTAGTGGGTGCACCAATCCAGGAAAAC
 -----T-----T--A--T--GC-C-----C-----G-----G--A--A--C-----T--A--G---
 E H H D P P Y K L C I H H R D F V V G A P I Q E N

ATTCTGAACAGCGTTGACCAAAGTAAGCGCATGCTCATGGTGCTCTCTCGCAACTTTTTAAAGAGCGAATGGTGC
 --A-----T--T-----C--A-----T-G-----C--T--AA-A--T-----G--TC--G-----
 I L N S V D Q S K R M L M V L S R N F L K S E W C

TTGTTGGAGTTTCGCGCGGCACATCGTAAAGTGCTAGAAGATCGCATGAATTATCTGATAATTATTCTGTTTGAT
 -----C-----T--G--C--C-----C--T-----A-----C--C--T--C-----CT-----C
 L L E F R A A H R K V L E D R M N Y L I I I L F D

```

GGCATCAACATGGATGAGCTCGATGACGAAATGAAGCTATACATGCGCACTAACACATATCTCAGTGTCTCAGCTAC
--A--T--T-----T--C--T--G-----A--T-----A-G--A-----T-----TTC---G-----T
G   I   N   M   D   E   L   D   D   E   M   K   L   Y   M   R   T   N   T   Y   L   S   V   S   Y

AAGTGGTTCTGGGAGAAGTTGTATTACGCGATGCCACAGAGTACTGACAGGCAGTTTAGAGCGAGAGATCTCTCA
-----T-----A--AC-----T--T-----G--ATCA--C--TC-----G--T--G--CT-G---
K   W   F   W   E   K   L   Y   Y   A   M   P   Q   S   T   D   R   Q   F   R   A   R   D   L   S

AGCATTTCTCCTCAAATGCTGGCATGCAATATTCCACTGAGACAGTGTTCAAGAATGAAGCATACCTCAAAGTCACT
TC---AAGTAGC-----G--A-----G-----T--C-----C--C--T-----T---T-G-----T--G
S   I   S   S   N   A   G   M   Q   Y   S   T   E   T   V   F   K   N   E   A   Y   L   K   V   T

GATATTCTGGAGTGA
--C---T-----
D   I   L   E   *

```

Figure 3.2 Human codon optimization for Of-TLR.

Original cDNA sequence (top line, red), nucleotide changes made in the human codon-optimized cDNA sequence (middle, green), and corresponding amino acid sequence (bottom, black).

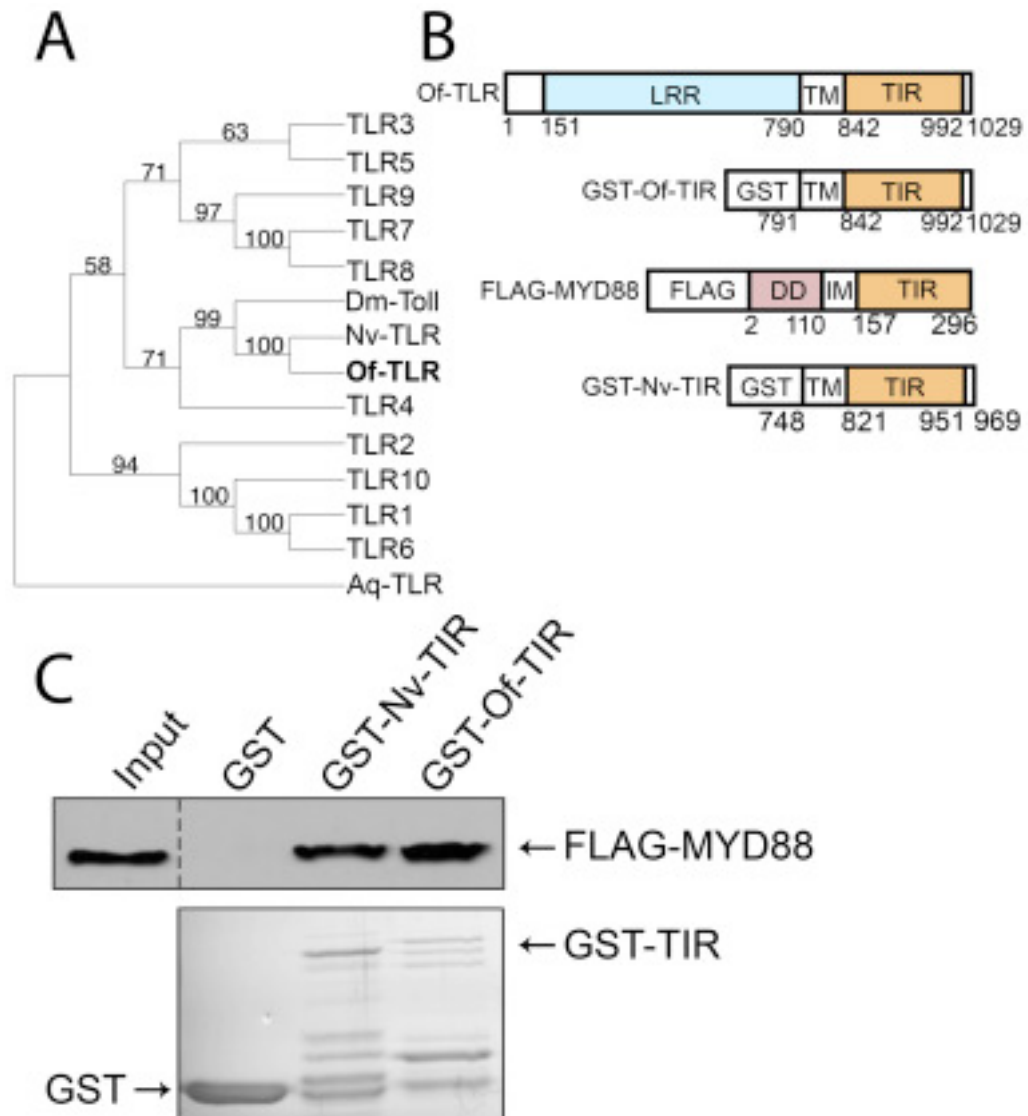


Figure 3.3 Similarity of Of-TLR to human TLR4.

(A) Phylogenetic analysis of the TIR domains of the indicated TLR proteins was performed using neighbor-joining tree analysis bootstrapped 1000 times. The phylogeny was rooted with the TIR domain of a TLR-like protein of the sponge *A. queenslandica*. Branches indicate bootstrap support values. Of, *O. faveolata*; Nv, *N. vectensis*; Aq, *A. queenslandica*. TLR1-10 are the human TLR proteins. (B) Schematic structures of Of-TLR (1029 amino acids) and the GST-TIR and FLAG-MYD88 constructs used in

pull-down assays. Highlighted are conserved structural components shared with other TLR and adapter proteins: LRR, leucine-rich region; TM, transmembrane domain; TIR, Toll/interleukin-1 receptor; DD, death domain, IM, intermediary domain. (C) Pull-down assays with GST and GST-TIR, and FLAG-tagged MYD88. The top panel shows an anti-FLAG Western blot of pull-downs using GST, GST-Nv-TIR or GST-Of-TIR. The input lane contains 0.7% of the FLAG-MYD88 293 cell lysate used in the pull-downs. Expression of bacterially expressed GST proteins was confirmed by SDS-PAGE followed by Coomassie blue staining (bottom). Dashed line represents the merge of Western blot lanes from a single gel. (Note: so that the amounts of GST proteins analyzed in the pull-downs were more similar, one-half of the sample from the GST alone pull-down was loaded on the gel used in the Western blot).

Original Sequence

Codon Optimized

Amino Acid Sequence

```
ATGGCCACCAACAGCGAGAGGCAGATCGGCGAGACCCTGACCGACAGCCTGCTGATGGACATCATGACCCCCGGC
----T--A-----T-----A-----T--T--A---T-----A---TC-----T--T-----T---
M  A  T  N  S  E  R  Q  I  G  E  T  L  T  D  S  L  L  M  D  I  M  T  P  G

TACCTGCCCCGACATCAGCGCCCTGCAGGTGCCACCGCCACCTACAGCGGCCCTACATGGAGATCCTGGAGCAG
---T-----ATC---A-----T-----T--TTCA--T--A-----A-----
Y  L  P  D  I  S  A  L  Q  V  P  T  A  T  Y  S  G  P  Y  M  E  I  L  E  Q

CCCAAGCAGAGGGGCTTCAGGTTCAGGTACCCCTGCGAGGGCCCCAGCCACGGCGGCCTGCCCGGCCAGTACAGC
----A--A--A--T--T--A-----A-----A--T-----T--T-----T--T-----A-----TTCT
P  K  Q  R  G  F  R  F  R  Y  P  C  E  G  P  S  H  G  G  L  P  G  Q  Y  S

GAGAAGGGCAAGAAGAGCTACCCAGCGTGCAGCTGTGCAACTACCACGGCCCCGCCAGGATCGTGGTGAGCCTG
--A-----A--TCA-----T-----A-----T--T-----A-----TTC--T--
E  K  G  K  K  S  Y  P  S  V  Q  L  C  N  Y  H  G  P  A  R  I  V  V  S  L

GTGACCGTGAGAGCCCCCATGCCCCACGCCACAGCCTGATCGGCAAGAACAGCAACAACGGCGCCGTGACC
-----T--T-----A--T-----T--T--A---TCT-----T--A--TCT-----T-----A
V  T  V  D  E  P  P  M  P  H  A  H  S  L  I  G  K  N  S  N  N  G  A  V  T

GTGCAGATCGGCCCCGAGCACGGCATGACCGCCAGCTTCCCCAACCTGGGCATCCAGCACGTGACCAAGAAGAGC
--T--A-----T-----A--T-----T-----T--A-----T-----T--A---TCT
V  Q  I  G  P  E  H  G  M  T  A  S  F  P  N  L  G  I  Q  H  V  T  K  K  S

GTGGGCAAGGTGCTGATGGAGAGGTACATCAAGATGCAGACCCTGCACACCGCCACCCTGCACGCCCTGACCGCC
----A--A--T-----A--T-----A--TT-----T--A--A-----AT-----A---
V  G  K  V  L  M  E  R  Y  I  K  M  Q  T  L  H  T  A  T  L  H  A  L  T  A

GACAGCAAGGGCTTCGACATGGAGCTGGTGGGCGACCAGGCCCTGGCCGACGGCGAGACCGCCACCTTCAACAGG
--TTCT-----T--T-----A-----T-----T-----T-----A-----T-----A-----T--A
D  S  K  G  F  D  M  E  L  V  G  D  Q  A  L  A  D  G  E  T  A  T  F  N  R

ACCATGGCCGAGGCCGTGGCCGCCGAGGAGAGCCAGAAGGTGAGGCAGATGGTGGAGGACCAGAAGCAGAGCATG
--T-----A--A-----T--A---TC-----A-----A--A-----T--A--T-----ATCA---
T  M  A  E  A  V  A  A  E  E  S  Q  K  V  R  Q  M  V  E  D  Q  K  Q  S  M

AACCTGAACGCCGTGAGGCTGTGCTTCCAGGCCTACCTGCCCCACGACGGCGGCTGCTTCACCAAGGCCCTGCCC
--T-----A--T--AT-----T--T-----A--T-----A--T--T-----T--A--T-----
N  L  N  A  V  R  L  C  F  Q  A  Y  L  P  D  D  G  G  C  F  T  K  A  L  P

CCCTGCATCAGCAACCCCGTGTACGACAGCAAGGCCCCCAGCGCCAGCAACCTGAAGATCTGCAGGATGGACAGG
--A-----TTCA-----T-----T--T-----A--T-----ATCT--T-----T--A-----A
P  C  I  S  N  P  V  Y  D  S  K  A  P  S  A  S  N  L  K  I  C  R  M  D  R
```

AACAGCGGCTGCGTGACCGGCGGCGACGAGGTGTACCTGCTGTGCGACAAGGTGCAGAAGGAGGACATCGACGTG
---TC---T-----A-----A--T-----T-----T-----A-----A-----A--T--T-----T
N S G C V T G G D E V Y L L C D K V Q K E D I D V

ATGTTCTACGAGATCGACGTGGAGACCGGCAAGAAGACCTGGGAGGCCGCGGCGGTGTTGCCCCCACCACGTG
-----T-----A--A-----A-----A--T-----T-----A--T--T
M F Y E I D V E T G K K T W E A G G V F A P T D V

CACAGGCAGGTGGCCATCGTGTTCAAGACCCCCGCTACTGGAACATCGCCACCGAGAGGCCCGTGAAGGTGCAC
--T--A-----T--T-----A--A--T-----A-----A--T-----A--T--
H R Q V A I V F K T P A Y W N I A T E R P V K V H

CTGGAGCTGAGGAGGAAGAGCGACCAGGAGACCAGCGAGCCCGTGGAGTTCACCTACCAGCCCCAGCTGTTTCGAC
-----AT-----A--A--T--T--T--A--A--TC-----A--T-----T--T-----T--A-----
L E L R R K S D Q E T S E P V E F T Y Q P Q L F D

AAGGAGCAGATCGGCGCCAAGAGGAGGAAGAAGATCCCCACTTCAGCGACTACTTCAGCGGCGGCGGCGGCGGCGG
--A--A-----T--T-----A--A-----A--T--A--T--TCT--T-----TC--A--T-----A--T
K E Q I G A K R R K K I P H F S D Y F S G G G G G

GGCGGCCCCCGGCATGGGCGGCGCGGCGCGGCGGCGGCGGCTTCGGCACCTTCGCCCCCTGGGCCTGCTG
-----A--T-----A-----T-----T--T--A--A--T-----A--T--A--A-----T--A-----
G G P P G M G G A G A G G G G F G T F A P L G L L

GGCAGCGCCCTGTGGTTCGCCAACCCCGGCACCAGCCAGAGCAACAACCAGGGCGGCAGCGGCGGCGCCGGCACC
--A-----TT-----T--A--T--T--T--ATC-----TCA--T-----A--T--TCT--A--T--A-----
G S A L W F A N P G T S Q S N N Q G G S G G A G T

AGCCAGCAGGGCCAGAGCCACAGCAGCGGCCAGAGCCAGGCCAGCGGCCAGAGCCACCAGGCCGACCTGAGCGAG
TCA-----A--A--A-----TTCTTC-----TCT--A--TTC--A--TCA-----A--A--T-----
S Q Q G Q S H S S G Q S Q A S G Q S H Q A D L S E

CTGGCCTGGAACCTGGCCGAGAAGAGCGCCGCCGCCATGAGGGACTACGCCGCCACCGGCGACATGAGGTACCTG
T-----T--A--ATCT--A-----T-----A--T-----A-----A--T--
L A W N L A E K S A A A M R D Y A A T G D M R Y L

CTGGCCGCCCAGAGGCACCTGACCGCCGTGCAGGACGACAGCGGCGACACCGCCCTGCACCTGGCCGTGATCAAC
T-----T--A-----A--T-----T-----T--A--T--TC--A-----A--A-----T-----T--T
L A A Q R H L T A V Q D D S G D T A L H L A V I N

AGCCAGCAGGAGGTGGTGCAGTGCCTGATCGACATCATGGCCGGCCTGCCCCGAGAGCTACATCAGCGAGTACAAC
TCA-----A--A-----T-----T-----T--T-----T--T-----T-----T--TTCT--A-----
S Q Q E V V Q C L I D I M A G L P E S Y I S E Y N

TTCTGAGGCAGAGCCCCCTGCACCTGGCCGCCATCACCAAGCAGCCCAGGATGCTGGACTGCCTGCTGAGGGGCC
-----A--TCT-----TT-----T-----A--A--A--T-----T-----A--
F L R Q S P L H L A A I T K Q P R M L D C L L R A

AGCGCCAACGTGAGGAGCAGGGACAGGCACGGCAACACCGCCGTGCACATCGCCTGCATGCACGGCGACGCCGTG
 TCT--T--T-----ATC---A-----A-----T-----T--T--T--T-----A--T-----T-----
 S A N V R S R D R H G N T A V H I A C M H G D A V

TGCCTGAAGGCCCTGCTGAACCTACAACGTGAGCAAGACCGTGCTGAACCTGGCAGAACTACCAGGGCGTGACCCCC
 -----A--T--T-----T-----TC-----TT--T-----T--A--A-----T--T
 C L K A L L N Y N V S K T V L N W Q N Y Q G V T P

GTGCACCTGGCCGTGCTGGCCGGCAGCAAGGACGTGCTGAAGCTGCTGAACAGCGCCGGCGCCAACATGAGCGCC
 --T--T-----TT--T--TTCA--A--T-----TCT--A--T-----TC--T
 V H L A V L A G S K D V L K L L N S A G A N M S A

CAGGACGGCACCAGCGGCAAGACCCCCCTGCACCACGCCGTGGAGCAGGACAACCTGGCCGTGGCCGGCTTCCTG
 -----T--T--TTCA-----A--A-----T-----A--T-----A-----TT--T--A--T--T--
 Q D G T S G K T P L H H A V E Q D N L A V A G F L

ATCTTGAGGGCAACTGCGACGTGGACGCCACCACCTTCGACGGCAACACCCCCCTGCACATCGCCGCCGCCAGC
 --TT--A--T--T--T--T--T--T-----A--T--A-----T--A-----A-----T--
 I L E A N C D V D A T T F D G N T P L H I A A A S

GGCTGAAGGGCCAGACCGCCCTGCTGGTGGCCGCCGGCGCCGACACCACCTGCAGAACAGCGACGAGGAGACC
 --AT-----T-----A-----T-----A-----T--T-----T-----A--TTCT--T--A-----
 G L K G Q T A L L V A A G A D T T L Q N S D E E T

GCCTTCGACCTGGCCAACGTGGCCGAGGTGCAGGAGATCCTGGACGAGGACGAGGCCCTGAGCACCGACCCCAGC
 -----T--T--A-----T-----A--T-----T--A-----T--TCA-----A--
 A F D L A N V A E V Q E I L D E D E A L S T D P S

CAGGACGACGAGCTGACCGCCGGCCTGACCGGCCTGACCCTGGGCCAGGGCGACATGGACAAGCTGGACCCCTAC
 --A--T-----A-----T-----T-----A-----TT-----A--T-----T--T--
 Q D D E L T A G L T G L T L G Q G D M D K L D P Y

GTGAGGAGGAAGATGGCCCAGAGGCTGGACCCCAGCATCGGCGCCGACTGGAGGGAGCTGGCCAAGAGGCTGGGC
 --T--A--A-----A--AT--T--TC--T--TG-----A-----T--A--A-----T
 V R R K M A Q R L D P S I G A D W R E L A K R L G

CTGGGCACCCTGGAGAGCGCCTTCGCCATCCACAGCAGCCCCACCACCCAGGTGCTGGCCCAGTACGAGGCCGCC
 T-----AT--ATCA-----T--T-----TTCA-----T--A-----T--A--T-----T--A
 L G T L E S A F A I H S S P T T Q V L A Q Y E A A

GACGGCAGCATCAAGACCCTGAGGCAGGTGCTGAGGGACATGAGGAGGGGCGACGTGCTGGAGATCCTGGACGGC
 -----TTCT--T-----T-----A-----TT--A--T-----A--A--T-----A--T-----T--
 D G S I K T L R Q V L R D M R R G D V L E I L D G

AGGAAGAGCCCCCTGAGCCACGACAGCGGCTTCGACAGCGGCCTGGGCAGCCAGAGCCTGAGCGCCTACAGGAGC
 --A--TC--T--TCA-----T-----TCA--A-----T-----ATCTT--TC--T-----ATCA
 R K S P L S H D S G F D S G L G S Q S L S A Y R S

```

GAGGAGCTGGCCAGCTACGAGGCCGGCAGCAGCAGCGTGCCCAGCAGCAGCAGCCTGCAGAAGGGCACCACCAGC
--A--T-----T-----T---TCTTC-TCA-----TCTTC-TCA-----A--T--ATCT
E E L A S Y E A G S S S V P S S S S L Q K G T T S

CTGGAGAGCACCAGGAAGGCCGGCGGCTACAGGCCCATCAGGCAGCACCAGGACGTGTTC
T----ATCT--A-----A--A--A--A-----A--A--A--A-----T-----TAATGA
L E S T R K A G G Y R P I R Q H Q D V F * *

```

Figure 3.4 Human codon optimization of Of-NF-κB.

Original cDNA sequence (top line, red), nucleotide changes made in the human codon-optimized cDNA sequence (middle, green), and corresponding amino acid sequence (bottom, black).

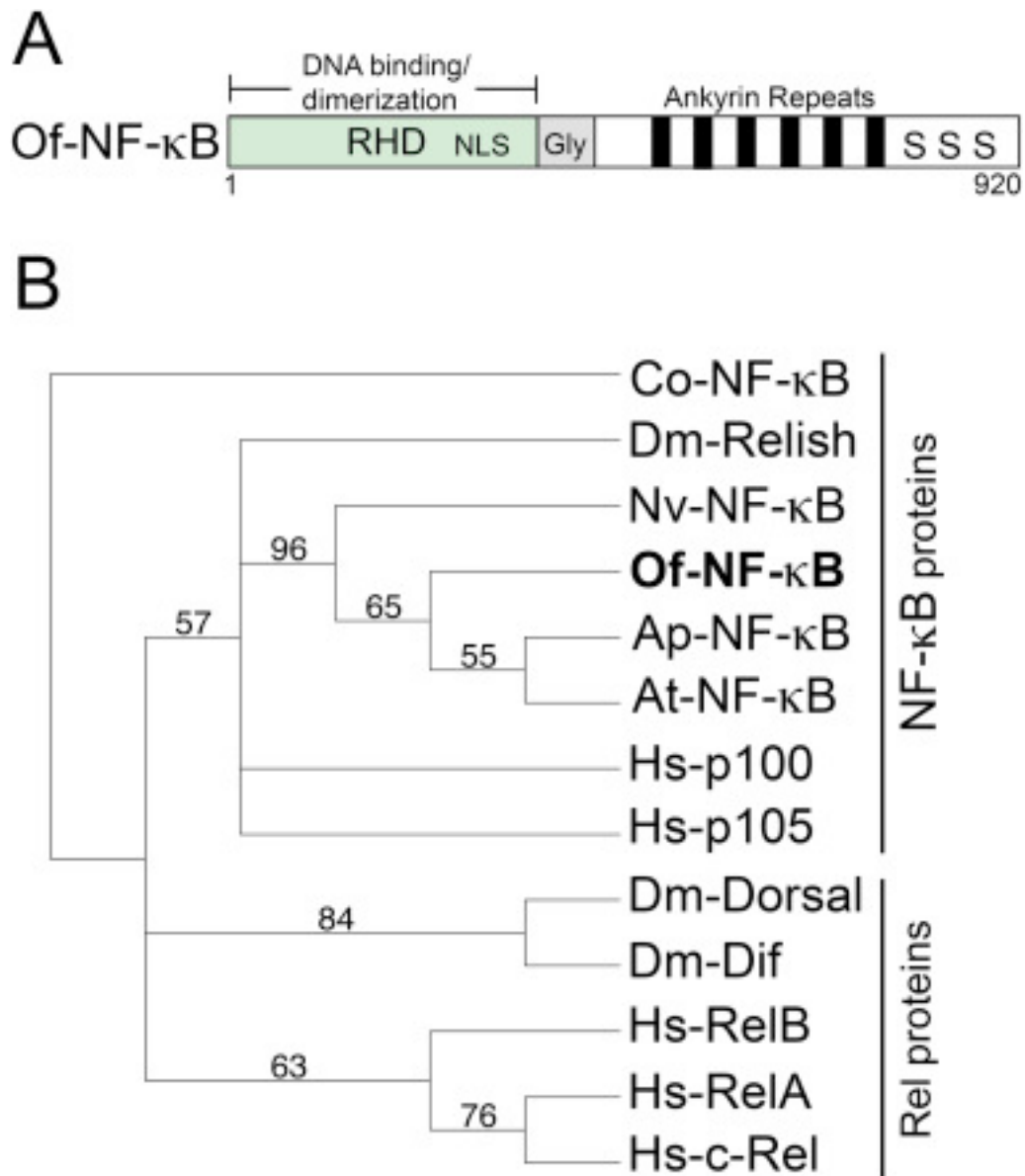


Figure 3.5 Of-NF- κ B is most similar to other NF- κ B proteins and not to Rel proteins.

(A) Schematic linear structure of the Of-NF- κ B protein (920 amino acids). Highlighted are functional domains shared with mammalian NF- κ B proteins: RHD, Rel Homology Domain; NLS, Nuclear Localization Sequence; Gly, glycine-rich region; S, serine; vertical black bars, Ankyrin repeats. (B) Phylogenetic analysis of the RHDs of the

indicated NF- κ B and Rel proteins using maximum-likelihood analysis bootstrapped 1000 times. The phylogeny was rooted with the NF- κ B protein of the single-celled eukaryote *Capsaspora owczarzaki*. Branches indicate bootstrap support values. Co, *Capsaspora owczarzaki*; Of, *Orbicella faveolata*; Nv, *Nematostella vectensis*; Ap, *Aiptasia pallida*; Hs, *Homo sapiens*.

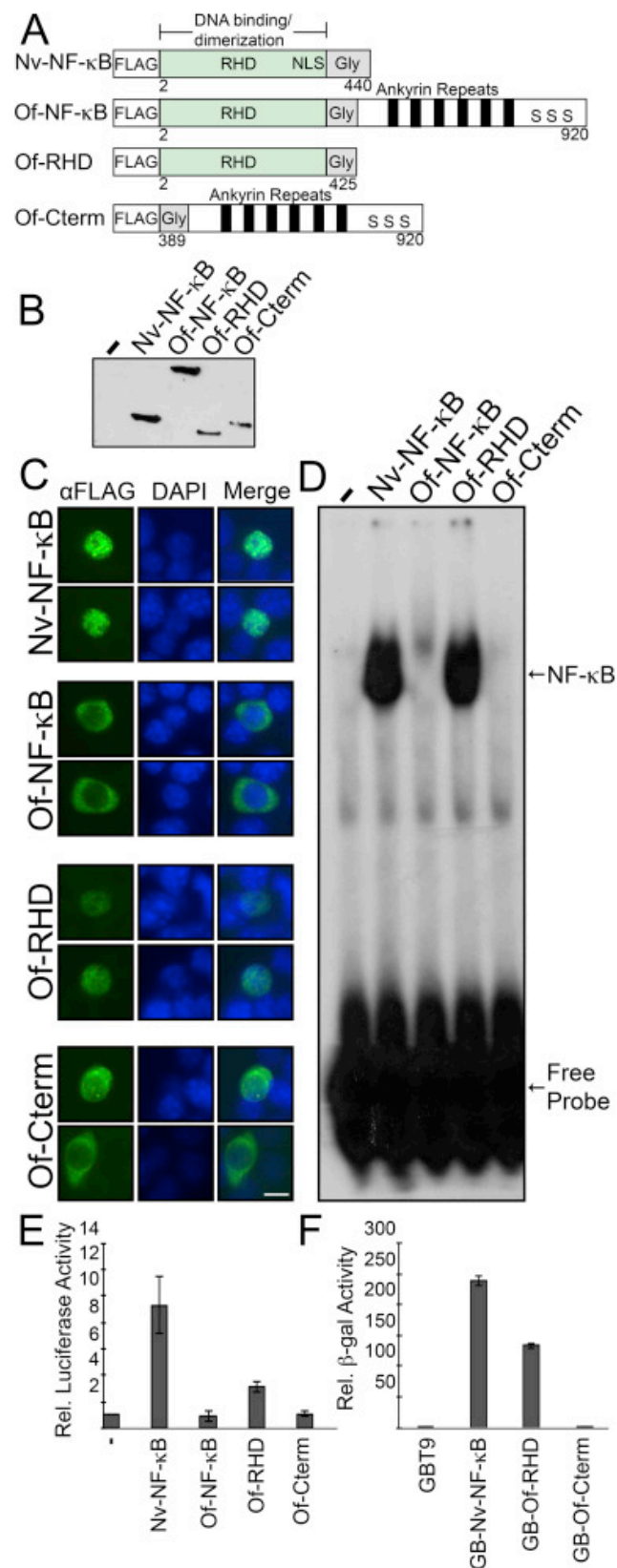


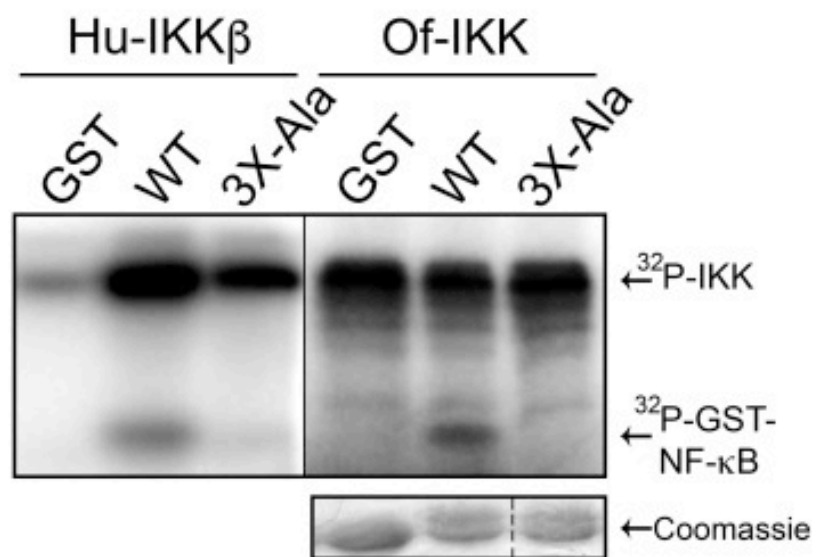
Figure 3.6 Characterization of the activity of the *O. faveolata* NF- κ B protein in cell-based assays.

(A) FLAG-tagged expression vectors used in these experiments. Shown from top to bottom are generalized structures of the naturally shortened Nv-NF- κ B, full-length Of-NF- κ B, a C-terminal truncation mutant of Of-NF- κ B (Of-RHD), and an N-terminal deletion mutant of Of-NF- κ B possessing only the glycine-rich region and the Ankyrin repeats (Of-Cterm). (B) An anti-FLAG Western blot of lysates from 293 cells transfected with the indicated plasmid expression vectors. (C) Chicken DF-1 fibroblasts were transfected with pcDNA-FLAG vectors containing the indicated proteins, and three days later cells were analyzed by anti-FLAG (α FLAG) indirect immunofluorescence (left, green) and DAPI staining (middle, blue). Merged images are shown in the right panels. Scale bar = 10 μ m. (D) Top: A κ B-site EMSA using lysates from 293 cells expressing the indicated FLAG-tagged proteins, as shown in (B). Arrows indicate the free probe and the shifted NF- κ B-DNA complex. (E) A κ B-site luciferase reporter gene assay was performed with the indicated proteins in 293 cells. Luciferase activity is relative (Rel.) to that seen with the empty vector control (1.0), and values are the averages of three assays performed with triplicate samples with standard error. (F) A GAL4-site *LacZ* reporter gene assay was performed in yeast Y190 cells. β -galactosidase (β -gal) reporter gene activity is relative (Rel.) to the GAL4 (aa 1-147) control (1.0), and values are the averages of seven assays performed with duplicate samples with standard error.

A

<i>H. sapiens</i> p100	865-D	S	A	Y	G	S	Q	S	V	-873
<i>A. pallida</i> NF- κ B	799-D	S	G	F	G	S	Q	S	A	-807
<i>O. faveolata</i> NF- κ B	861-D	S	G	L	G	S	Q	S	L	-869

B



C

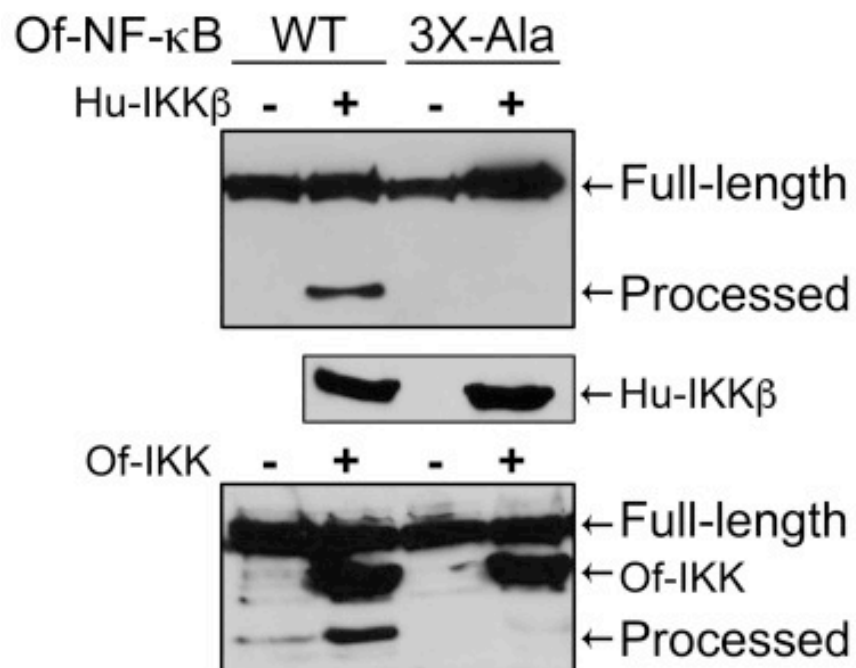


Figure 3.7 IKK can phosphorylate conserved serines in Of-NF- κ B in vitro and induce processing of Of-NF- κ B in human cells.

(A) An alignment of relevant regions of NF- κ B from human p100, the sea anemone *A. pallida*, and *O. faveolata*. Conserved serine residues are boxed in red. (B) *In vitro* kinase assay (top) using a constitutively active FLAG-Hu-IKK β protein and FLAG-Of-IKK with bacterially expressed substrates of GST alone or GST fusion proteins containing the wild-type (WT) serine residues of Of-NF- κ B (amino acids 843-874), or the same C-terminal residues with conserved serines mutated to alanines (3X-Ala). The GST proteins were electrophoresed on an SDS-polyacrylamide gel and stained with Coomassie blue (bottom). Dashed lines indicate the respective Coomassie blue stain for each protein. (C) A Western blot of lysates from 293T cells co-transfected with Of-NF- κ B wild-type (WT) or 3X-Ala and with HA-Hu-IKK β (top two panels) or FLAG-Of-IKK (bottom panel). Full-length and processed forms of Of-NF- κ B are indicated. Shown also are the HA-Hu-IKK β (middle panel) and FLAG-Of-IKK proteins (bottom panel) from the same lysates.

Original Sequence	Codon Optimized	Amino Acid Sequence
ATGGAATATGGAAGAGAAAGGAAAAATCAAGGGAGGGGTGGGTGGAGGAAAGATGTATTGGGTCCGGTGGT	-----C--C--G--G--A--G--GAGC--A--A--C-----A--G--G--C--C-----A-----C--C	M E Y G R E R K K S R E G W V E E R C I G S G G F
GGAACAGTTACTGTATGAGAACAAGATTACCAAGCAACAGATTGCCCTTAAGAGATGTGCGATGGATT	---G-----G--CT-----A-----A--C--G--A-----A--A-----G--AC-----C-----C--T--C	G T V T L Y E N K I T K Q Q I A L K R C R M D L N
GCTCAAAACAAGAAGCTATGGCGACAGGAAGTTGACATAATGAAGAGACTGGATCACCCAAACCTTGTGT	---G--G-----A--A--C---A--G--A-----A--T--C-----G-----C--T--C--T--C--T-----G	A Q N K K L W R Q E V D I M K R L D H P N L V S A
AAAGATGTTCCAGCACCCTGGATGTCAGTGTGATGAACCTCCTCTGTTGGCCATGGAGTTTTGCTCTGGAGGA	-----G--G-----T---C--TTCA-----A--TC-----A--C--T--C--G--G	K D V P A P L D V S V D E L P L L A M E F C S G G
GACTTGAGAAAGGTACTAAACTCCCCAGAAAGTTGCTGTGGGCTGAAAGAAAAGACAGTTTTGAAGATTGCCTCA	---T-----A--C--C-----T---GTC---T--C--C-----G--G--A--G---C--T-----G--C	D L R K V L N S P E S C C G L K E K T V L K I A S
GATGTTGCTGCTGCAGTGAATTTCTTCATGGCCACCGCATCATTACAGAGATCTAAAACCTGAAAACATTGTT	----G--A--G-----T---C--C--C--T--T--A--T--C-----C--C--G--C-----A--	D V A A A V E F L H G H R I I H R D L K P E N I V
ATAAGTCATGTGGATAATAAGGTAGTTTACAACTGATAGATCTTGGCTATGCAAAGAAGTGGATCAAGGAAGT	---T--C--C-----C-----G--C--T--G--T-----T--G--G--C--C-----GT-----G--GTCC	I S H V D N K V V Y K L I D L G Y A K E L D Q G S
CTGGCTACCACTTTTGTCTGGGACTCTAAAATATCTGGCCCCTGAACTGCTTGACAAGGTGCAATACACAAAAACG	T---C-----C--A--C-----G-----C--C-----T---C--T--A-----A	L A T T F V G T L K Y L A P E L L D K V Q Y T K T
GTGGATTATTGGAGTTTGGCACACTCTTGTTTGAATGCATTACTGGAATGCGGCCTTTTCTTACCGAACTATCT	---A-----C-----C--C-----G--TC---C-----C-----G--C--C-----G--	V D Y W S F G T L L F E C I T G M R P F L T E L S
CCAGTCCAGTGGTATAATAAAGTTTCGAGCAAAGGACCGGATGACATCTGTGCCTACTTTGATCTAAAAGAGGAA	--C--A-----C-----CA-GTC---G--T--C--C--T-----C--T-----T--G-----	P V Q W Y N K V R S K G P D D I C A Y F D L K E E
ATACAGTTCTCTTCTGTTCTGCCGACTCCCAATTCCTCTGTAGGCAACTTCAAGACAAGTTTGTGTCCTTCTCT	---T--A-----AGC-----C--C--G--CAG-----CC-A--GT-G--G--T-----C--A--A--C--G	I Q F S S V L P T P N S L C R Q L Q D K F V A L L

CGACTGCTGTTGTTGTGGGACCTCAAGCACGTGGAGGATCAATTCTTGAAAGTGGCAAGAAGGAGTGCTTTGAT
 A---T--CC-CC-T-----G---C--G-----T--C--A--G-----A--T--C--C
 R L L L L W D P Q A R G G S I L E S G K K E C F D

GTGCTTGAAAAGATCATCAATACTGTGATAGTCCGTGTATTCTTAGTCAGCTCAGCAGTGGCGATTCCATTTGAA
 --C-----T--T--C--G--C--C--T--G--T-----G--ATC-----C--T--C-----G--C--G
 V L E K I I N T V I V R V F L V S S A V A I P F E

GTATTACCAAGTGAAACGATAGAGGATCTGCAAGTTAAATTATCTGAGGCAACTGGAGTCGATGTTAAGGAGCAG
 --TC-T--TTCC--G--C-----CT---G--A--C-G-----T-----A---G--A-----
 V L P S E T I E D L Q V K L S E A T G V D V K E Q

GAATTGTTGTTGGCTTCAGGGCAGGCTGTTAAGCCAAATACACTTGTTCATGGATTGCATCAAGGATGATCAGATT
 ---C--C-TC-T-----T-----C--G--A--G-----A-----T-----C-----A---
 E L L L A S G Q A V K P N T L V M D C I K D D Q I

GGAGATGATTGCGCTCTGTTCCTTTTGCCAACAAGTGTTGAACCAAATCCACACGGCCATTGTACACTTTGCCA
 --T-----C---GC---C--G--C--C---G-----T--TA-A--GC-C--T--AC-C--C
 G D D C A L F L L P T S V E P K S T R P L Y T L P

CCTACAGTTCAAGGAATGGTCACTGAGCAGAAAACGCTGATGTCGTACGAGGAATTGAAAAGGGCCTTTGCTCAG
 --C--T--A--G--C-----T--A-----A-----A-----A---C---GC-C-----C--A
 P T V Q G M V T E Q K T L M S Y E E L K R A F A Q

GGCCTGAACTTTTGCCGTGAACAGACAAAGATAAATCAGCTGCTGATTGAAGGCTTTAGGGCGGCACAGATTTCA
 --G--A--T-----A-G-----C-----C--A--TT---C-----T--C--A--C--T--A-----
 G V N F C R E Q T K I N Q L L I E G F R A A Q I S

TTGTTAGTTCTTAATTCCAACTTGACAGCTCAAGTCTGAAGTTCTAACTGAATCCAACAACTGGAGGCAAAA
 ---C-T--G--G--AG-----T-G--A--A--G--C--G--C--G--T--T-----T--A--T---
 L L V L N S K L G Q L K S E L L T E S N K L E A K

ATAGACTTTTTCAACTCCAGTTTACAGACTGATTTAGAAATTTATGCTGATAATATCAACCTTGTAaaaaatgat
 -----T--T-----C-C--A--A--CC-C--G-----C--C-----T-----C--G-----
 I D F F N S S L Q T D L E I Y A D N I N L V K N D

AAACGCTCAGAACTGGAACGGGCGCAGAAAAAGGTGGAGTCGTTTCAGAATAATGATGTAACACAACCTGAAA
 --G---T-GC-G--T-----G--C--T--A-----A--AAGC-----T--G-----C---
 K L L R N W K R A Q K K V E S F Q N N D V T Q L K

GAGTTCACAGCAGCAGTACAGACTAAAGTTGTTGAGCTGCAGAAGAGTCCATACACAAGTGGAACGCAATCGAAA
 -----T--T--T-----G--C--G-----T-----ATCC--G--T--G--C--T--T-----C--G
 E F T A A V Q T K V V E L Q K S P Y T S G T Q S K

TCTAAACCTGATGGAATGAGGTACGACTGTGCTGTCAAACCTTTACGAGAAGCTGAAACAAAGCGACAGACGACCC
 AG---G--A--C-----C-----T--C--A--T--G-----A--T--G--GTCT--TC-CA-G--A
 S K P D G M R Y D C A V K L Y E K L K Q S D R R P

```

GCGGACAGTCAGCCAATGGCCACGATTATCTTACAGATTCTTGACAGAGAGAGAAAAGTGCACAGGGATATTTGT
-----G-----A--C--C--TC-C--A-----C--A-----TC-----A---
A  D  S  Q  P  M  A  T  I  I  L  Q  I  L  A  Q  R  E  K  L  H  R  D  I  C

ACTCAGTTAAGTGAAGTGATGTCCTGCAGAAGAGAGATCCGCAAAGTCATGTATCGTGTTATACAGGTCAAGGAG
--A--AC-GTCC-----C-----T--C-G-----A-----A-----G-----C--G-----G-----
T  Q  L  S  E  V  M  S  C  R  R  E  I  R  K  L  M  Y  R  V  I  Q  V  K  E

AAACTTATCACCAGAGCAGATGAGCTCACAGAGATGCAGAAGCAAAGACAAGTTGACCTGTGGAGACTTCTCCAA
--G--C-----AC-----T--C-----A-----A-----G-----G--G-----C--T---
K  L  I  T  R  A  D  E  L  T  E  M  Q  K  Q  R  Q  V  D  L  W  R  L  L  Q

TCAAAGAGGCCAAAGAGAACAGAATCGTTCGTCAAGCAGTCCACATACATCAATTTCTTCGCTCACAAGTTCTTGT
AGT---C-----GC-----G--A-----C--TAGTTC---C-----C--G--T-----AGCT-G--G--C-----
S  K  R  Q  R  E  Q  N  R  S  S  S  S  P  H  T  S  I  S  S  L  T  T  S  C
CTCAGTACACAATCCGTTGTTCTAATTGAGGAGAGCTCTAGATCCTTTGAAAAGTTTGCCGGATTGATAAACAAA
-----G--G--T--C--G--C--C--A---TCAAG-C-GAG-----G-----GC-----T-----G
L  S  T  Q  S  V  V  L  I  E  E  S  S  R  S  F  E  K  F  A  G  L  I  N  K

TTGAAGATCGATCAGATCAGTGAGACGGAGAATTTGACTGGGAATTCCTTGAGGAAGAAAATAGCTCAACTGAC
C-C-----A-----A--T--A--C-----T-----G--T-----A-----CTCTAGC--A---
L  K  I  D  Q  I  S  E  T  E  N  F  D  W  E  F  L  E  E  E  N  S  S  T  D

CAAGAGAGCGTT
--G--ATCT--CTGA
Q  E  S  V  *

```

Figure 3.8 Human codon optimization for Of-IKK.

Original cDNA sequence (top line, red), nucleotide changes made in the human codon-optimized cDNA sequence (middle, green), and corresponding amino acid sequence (bottom, black).

CHAPTER FOUR

PORIFERAN TRANSCRIPTION FACTOR NF- κ B HAS CONSERVED AND UNIQUE SEQUENCES, ACTIVITIES, AND REGULATION ⁴

4.1 Introduction

Sponges are the sole members of the phylum Porifera, with about 8,500 known extant species (Van Soest et al., 2012). Poriferans are widely considered the sister group to all other multicellular animals (Simion et al., 2017). Sponges are estimated to have first appeared approximately 600 million years ago, prior to the Cambrian explosion (Yin et al., 2015). Sponges are found in a variety of aquatic habitats throughout the world, with the majority of species living in marine environments (Van Soest et al., 2012).

Sponges have simple body plans that resemble a sac of cells with an extracellular matrix and pores. Nevertheless, the simplicity of the sponge body plan belies the complex mechanisms that must be coordinated to allow for their existence. Indeed, due to their water filtration system and body plan, sponges have an extremely high exposure to a diverse array of microorganisms as compared to other multicellular organisms (Wiens et al., 2005). Therefore, it is not surprising that sponges have a robust innate immune system to monitor and counteract their continuous exposure to microbes. As with more advanced metazoans, the immune system of sponges has the ability to distinguish between friendly and foreign matter, such as the recognition of cells from heterologous sponges, the uptake of bacteria by phagocytosis through macrophage-like cells (Wehrl et al., 2007), the production of

⁴Adapted from Williams et al., 2020 *Developmental and Comparative Immunology*

secondary metabolites for defense (Proksch, 1994), and the recognition of bacterial cell wall-derived lipopolysaccharide (LPS) (Böhm et al., 2001). The molecular mechanisms by which these defense processes are regulated are largely unknown; however, the presence of homologs to higher metazoan innate immune molecules in sponges suggests a conserved and complex level of immune defense and regulation (Müller and Müller, 2003).

Because of their phylogenetic position as the earliest lineage of living animals, sponges are useful for addressing questions regarding the origin of metazoan-specific gene families and their evolution (Riesgo et al., 2014). The Great Barrier Reef demosponge *Amphimedon queenslandica* (Aq) is perhaps the most extensively studied poriferan at the cellular and molecular level, including having its genome sequenced and being the subject several studies focused on investigating the evolution of metazoan development and multicellularity (Larroux et al., 2006; Adamska et al., 2007; Tompkins-MacDonald et al., 2009; Srivastava et al., 2010).

One of the most prominent proteins involved in the regulation of immunity in complex metazoans is transcription factor NF- κ B, whose activity is regulated by many upstream factors (Gilmore, 2006). Many homologs of the NF- κ B pathway are conserved across diverse phyla, from vertebrates to the single-celled holozoan *Capsaspora owczarzaki* (Gilmore and Wolenski, 2012). NF- κ B pathway components and transcriptional responses have been intensively studied in vertebrates and flies for their involvement in immunity, development, and disease. However, this pathway and genes regulated by NF- κ B are, for the most part, less well-characterized in more evolutionarily basal organisms (Gilmore and Wolenski, 2012).

Overall, the amino acid (aa) sequence and structural organization of the single NF- κ B proteins that have been identified in cnidarians (Wolenski et al., 2011; Gilmore and Wolenski, 2012; Mansfield et al., 2017; Williams et al., 2018), poriferans (Gauthier and Degnan, 2008), the protist *Capsaspora owczarzaki* (Sebé-Pedrós et al., 2011), and some choanoflagellates (Richter et al., 2018) more closely resemble the NF- κ B subfamily of proteins found in vertebrates and flies.

Activation of NF- κ B pathways can be initiated by various upstream ligands that bind to their specific receptors (e.g., Toll-like receptors (TLRs), Interleukin-1 receptors (IL-1Rs), tumor necrosis factor receptors (TNFRs)), which then direct appropriate downstream effects. In some cases in humans, Tank-binding Kinase (TBK) acts with other signaling complexes that feed into a NIK-IKK cascade, which leads to non-canonical NF- κ B activation (Pomerantz and Baltimore, 1999). Ultimately, non-canonical processing of p100 leads to generation of the RHD-only protein p52 in a heterodimer with the Rel-like protein RelB, and the p52-RelB heterodimer can then translocate to the nucleus, bind DNA, and activate target genes (Gilmore, 2006).

Transcriptomic and genomic sequencing has revealed that NF- κ B and homologs of many of its upstream regulators are present in most eukaryotes from protists to vertebrates (Gilmore and Wolenski, 2012). However, the numbers and structures of these signaling proteins vary across species, and generally become more complex and numerous through evolutionary time.

In this chapter, we use phylogenetic, biochemical, and cell- and tissue-based approaches to characterize the structure, activity, and regulation of sponge NF- κ B. Our

results indicate that sponge NF- κ B has many of the same properties as other metazoan NF- κ B proteins, but also has sequences and methods of regulation that may be specific to poriferans.

4.2 *A. queenslandica* NF- κ B resembles other metazoan NF- κ B proteins both structurally and phylogenetically

Gauthier and Degnan (2008) reported that the sole NF- κ B protein in the demosponge sole *A. queenslandica* (Aq-NF- κ B) has an overall structure (RHD-ANK repeat domain) and aa sequence phylogeny that are more closely related to mammalian NF- κ B proteins (p100/105) than Rel proteins (RelA, RelB, c-Rel). That is, Aq-NF- κ B contains an N-terminal RHD, a nuclear localization sequence (NLS), a glycine-rich region (GRR), and ANK repeats (Figure 4.1A). However, Aq-NF- κ B has an additional 171-aa sequence between the GRR and the ANK repeats (Figure 4.1A, grey domain) that is not found in any other NF- κ B protein and has no sequence similarity to any other protein in BLAST databases. Nevertheless, the general structure of Aq-NF- κ B resembles other basal NF- κ Bs discovered to date, such as those in cnidarians like *Aiptasia* NF- κ B and *Nematostella* NF- κ B (Wolenski et al., 2011; Mansfield et al., 2017; Williams et al., 2018) and *Capsaspora owczarzaki* (Sebé-Pedrós et al., 2011) (Figure 4.1B). Our own phylogenetic comparison of RHD sequences confirmed that the Aq-RHD is more similar to the RHDs of other NF- κ B proteins as compared to Rel proteins across phyla, and that the Aq-RHD sequences are most closely related to NF- κ B-like proteins in Cnidaria, the closest metazoan phylum to Porifera (Figure 4.1C and Table 2.7). Taken together, these

results are consistent with other studies indicating that NF- κ B proteins are more ancient than Rel proteins (Wolenski et al., 2011; Gilmore and Wolenski, 2012; Mansfield et al., 2017; Williams et al., 2018).

4.3 Aq-NF- κ B can bind DNA and activate transcription

To investigate the overall DNA binding-site specificity of Aq-NF- κ B, we assessed the DNA-binding profile of bacterially expressed, affinity purified RHD sequences of Aq-NF- κ B in a protein binding microarray (PBM) consisting of 2592 κ B sites and 1159 random background sequences (for array probe sequences, see Mansfield et al., 2017). We (Siggers et al., 2011; Mansfield et al., 2017) and others (Ryzhakov et al., 2013) have previously used this type of analysis on both mammalian and cnidarian NF- κ B proteins. By analysis of z-scores of binding to DNA sites on the PBM, the DNA-binding profile of Aq-NF- κ B is similar to NF- κ B from the sea anemone *N. vectensis* and human NF- κ B p52, but is distinct from human c-Rel and RelA (Figure 4.2A).

In mammals and cnidarians, NF- κ B proteins require post-translational processing for nuclear translocation, active DNA binding, and transcriptional activation activity (Sun, 2011; Mansfield et al., 2017; Williams et al., 2018). To investigate corresponding properties of Aq-NF- κ B, we first created FLAG-tagged expression vectors for full-length Aq-NF- κ B, a mutant containing the RHD and GRR, and a mutant containing only the sequences C-terminal to the RHD (Aq-Cterm, i.e., primarily the ANK repeat domain) (Figure 4.2B). As a control, we used an expression vector for the sea anemone *N. vectensis* NF- κ B protein (Nv-NF- κ B), which we have characterized previously as an active, naturally

truncated NF- κ B protein (Wolenski et al., 2011). As shown by anti-FLAG Western blotting (Figure 4.2C), each construct expressed a protein of the appropriate size when transfected into HEK 293T cells.

To analyze the transactivation properties of full-length and truncated Aq-NF- κ B proteins, we performed an NF- κ B-site luciferase reporter assay in 293 cells. Aq-RHD and Nv-NF- κ B both activated transcription of the reporter gene (4.8-fold and 21.9-fold, respectively) as compared to the empty vector control and to full-length Aq-NF- κ B (Figure 4.2D). Furthermore, a GAL4-Aq-NF- κ B RHD fusion protein activated transcription approximately 15-fold above control levels in a GAL4-site reporter assay in yeast (Figure 4.2E). As we have shown previously (Wolenski et al., 2011; Alshanbayeva et al., 2015; Williams et al., 2018), a GAL4-Nv-NF- κ B fusion protein strongly activated transcription in yeast (approximately 600-fold) as compared to the GAL4 DNA-binding domain alone control. Therefore, the ability of the Aq-RHD sequences to activate transcription appears to be an intrinsic property of sequences within the N-terminal half of the protein, and the ability of the Aq-RHD sequences to activate κ B site-dependent transcription in cells is not seen with the full-length Aq-NF- κ B protein.

To assess the DNA-binding ability of Aq-NF- κ B, whole-cell detergent extracts from 293T cells transfected with the Aq-NF- κ B expression plasmids were analyzed in an electrophoretic mobility shift assay (EMSA) using a κ B-site probe, which we have previously shown can be avidly bound by cnidarian NF- κ B proteins (Wolenski et al., 2011; Mansfield et al., 2017; Williams et al., 2018). Lysates from cells transfected with expression plasmids for Aq-RHD and the positive control Nv-NF- κ B both contained an

activity that strongly bound to the κ B-site probe (Figure 4.2F). As expected, lysates from cells transfected with the empty vector or the Aq-Cterm had little basal DNA-binding activity. Somewhat surprisingly, full-length Aq-NF- κ B bound the κ B-site probe to approximately the same extent as the Aq-RHD protein, and showed two prominent DNA-protein complexes that migrated more slowly than the Aq-RHD-DNA complex.

Taken together, these results demonstrate that removal of C-terminal ANK repeat domain sequences of Aq-NF- κ B enables the protein to activate transcription in vertebrate cells, but the presence of C-terminal sequences of Aq-NF- κ B does not substantially inhibit its DNA-binding activity when Aq-NF- κ B is expressed in human cells.

4.4 C-terminal truncation of Aq-NF- κ B enables it to translocate to the nucleus and can be induced by an IKK-dependent mechanism in vertebrate cells

In non-canonical NF- κ B pathway activation, the human NF- κ B p100 protein is converted to an active form by proteasomal processing of C-terminal sequences up to the GRR, which allows for the shortened p52 protein to translocate to the nucleus (Sun, 2011). To investigate the subcellular localization properties of Aq-NF- κ B, we expressed the above-described FLAG-tagged Aq-NF- κ B proteins in DF-1 chicken fibroblast cells and assessed their subcellular localization by anti-FLAG indirect immunofluorescence (Figure 4.3A). Full-length Aq-NF- κ B showed cytoplasmic staining, whereas the C-terminally truncated Aq-RHD protein was located exclusively in the nucleus (as judged by overlap with DAPI staining of the nucleus). The N-terminally deleted Aq-Cterm protein, which consists of essentially the C-terminal ANK repeats, showed cytoplasmic

staining. Consistent with our previous results (Wolenski et al., 2011), the naturally shortened *N. vectensis* Nv-NF- κ B protein localized to the nucleus in these cells.

We have previously shown that two cnidarian NF- κ B proteins (from the sea anemone *Aiptasia* and the coral *Orbicella faveolata*) can be induced to undergo C-terminal processing in human 293T cells by overexpression of human and cnidarian IKKs (Mansfield et al., 2017; Williams et al., 2018). Furthermore, IKK-induced processing of these cnidarian NF- κ Bs requires a cluster of C-terminal serine residues with similar sequence and spacing to ones found in human NF- κ B p100 (Sun, 2011; Mansfield et al., 2017; Williams et al., 2018). Previously, results from our lab indicated that co-transfection of IKKs with Aq-NF- κ B did not induce its C-terminal processing (Mansfield et al., 2017). This lack of processing was proposed to be due to the absence of obvious sequence homology of sites of IKK phosphorylation in the C-terminal sequences of Aq-NF- κ B as compared to human p100 and *Aiptasia* NF- κ B. Based on that previous result, we hypothesized that the native spacing of serines near the C terminus of Aq-NF- κ B was not allowing IKK-induced C-terminal processing of the protein. Therefore, in initial experiments, we created a FLAG-tagged Aq-NF- κ B mutant that had the same spacing and number of serines as *Aiptasia* NF- κ B (Figure 4.4). WT Aq-NF- κ B and the Aq-NF- κ B mutant containing the *Aiptasia*-like serine cluster were then co-transfected with vectors for expression of Hu-IKK β or Ap-IKK, or the corresponding empty vector. In these experiments, WT Aq-NF- κ B and the Aq-NF- κ B-*Aiptasia* serine mutant were both processed when co-expressed with the IKKs, but neither was processed in cells transfected with the empty vector control (Figure 4.4C, top and bottom, lanes 1 and 2,

and Figure 4.4B). This result demonstrated that the WT Aq-NF- κ B can indeed be processed in 293T cells by co-expression of these IKKs, and that there was likely an error in analysis in the previous (Mansfield et al., 2017) experiments. We believe that the previously published Western blot (Mansfield et al., 2017) was run too long and the lower processed band was run off the gel or that the blot was not exposed long enough to detect processing of Aq-NF- κ B.

With this new information, we sought to identify sequences important for IKK-induced C-terminal processing of Aq-NF- κ B. In a re-analysis, we found that Aq-NF- κ B contains a cluster of serine residues that are somewhat similar to the serines required for processing of human p100 and the sea anemone *Aiptasia* NF- κ B protein (Figure 4.4A, serines in red). Additionally, the initial discovery of Aq-NF- κ B identified extra sequences after the GRR (aa 511-681) (Gauthier and Degnan, 2008) that contained several serines, but the function of this region is not known (Figure 4.1A). To determine which sequences were required for processing of Aq-NF- κ B, we created three mutants: 1) with five C-terminal serine residues (Figure 4.4A serines in red) changed to phospho-resistant alanines (ALA), 2) with a deletion of the additional region between the GRR and the ANK repeats (Δ), and 3) a double mutant (Δ ALA) that had both the serine-to-alanine mutations and the deletion of aa 511-681 region. The Aq-NF- κ B-ALA mutant showed substantially reduced IKK-induced processing as compared to wild-type Aq-NF- κ B in 293T cells (Figure 4.3C, top and bottom, lanes 3 and 4). The Aq-NF- κ B Δ mutant showed a level of IKK-induced processing similar to WT Aq-NF- κ B (Figure 4.3C, top and bottom, lanes 5 and 6), and the double mutant Aq-NF- κ B- Δ ALA showed reduced

processing as compared to the Aq-NF- κ B Δ mutant (Figure 4.3C, top and bottom, lanes 7 and 8). Overall, these results demonstrate that IKK-induced C-terminal processing can also occur with Aq-NF- κ B, and that this processing requires a cluster of C-terminal serine residues but does not require the additional sequences (aa 511-681) in Aq-NF- κ B.

To determine whether these Aq-NF- κ B serine residues could be directly phosphorylated by human IKK β , we first generated bacterially expressed GST-fusion peptides containing either the wild-type or phospho-resistant alanine sequences (aa 1048-1086) near the C terminus of Aq-NF- κ B. These GST-fusion proteins were then incubated with FLAG-Hu-IKK β in a radioactive *in vitro* kinase assay. As shown in Figure 4.3D, the GST-tagged peptides containing the wild-type Aq-NF- κ B serine residues, but not the Ala residues, were phosphorylated by IKK β .

4.5 Sponge tissue expresses putative NF- κ B proteins and contains NF- κ B site DNA-binding activity that increases following treatment with bacterial lipopolysaccharide (LPS)

To determine whether sponge tissue expresses active NF- κ B, we acquired tissue from an aquarium grown black encrusting demosponge from the Caribbean that is likely of the genus *Cliona* (Figure 4.5A, and Figure 4.6). This sponge was chosen because of the difficulty of growing *A. queenslandica* in captivity (Degnan et al., 2008) and because preliminary experiments showed that our anti-sea anemone Aiptasia (Ap) NF- κ B antiserum (Mansfield et al., 2017) cross-reacted with our aquarium sponge NF- κ B protein but not with Aq-NF- κ B (not shown). Anti-NF- κ B Western blotting of extracts from the black

encrusting sponge tissue detected one high (~130 kDa) and one lower (~65 kDa) molecular weight band, with the majority of the protein being in the lower band (Figure 4.5B). As expected, the anti-Aiptasia NF- κ B antibody detected two bands of appropriate sizes in a 293T cell extract containing overexpressed Aiptasia NF- κ B (Figure 4.5B).

When this same sponge tissue lysate was analyzed by an EMSA using an NF- κ B-site probe, a band was detected that migrated similar to the prominent band seen with extracts from 293T cells overexpressing FLAG-Aq-RHD (Figure 4.5C).

To analyze the organismal pattern of NF- κ B expression, we performed anti-NF- κ B immunofluorescence on cryosections of black encrusting sponge tissue. A subset of cells expressed immunoreactive proteins that closely coincided with Hoechst-stained nuclei in the tissue (Figure 4.5D, for quantitation, Table 4.1).

Overall, the results in this section indicate that the black encrusting sponge tissue has proteins that are similar to full-length and processed NF- κ B proteins from other species, contains NF- κ B-site binding activity, and expresses NF- κ B in the nucleus of some but not all cells.

To determine whether a known activator of mammalian NF- κ B could affect sponge NF- κ B, we treated black encrusting sponge tissue with *E. coli* lipopolysaccharide (LPS) for 30 min and then analyzed the tissue by Western blotting or EMSA. After treatment with LPS, the amount of processed protein (lower band, Figure 4.7A) increased by approximately 15% and the amount of unprocessed protein decreased by approximately 10% (normalized to the α -tubulin loading control, Figure 4.8). Additionally, LPS treatment resulted in increased κ B-site DNA-binding activity as assessed by EMSA (Figure 4.7B).

Overall, these results demonstrate that treatment of the isolated sponge tissue with *E. coli*-derived LPS results in an increase in NF- κ B processing and DNA-binding activity.

4.6 Chapter 4 summary

In this chapter, it is demonstrated that at least two sponges have active NF- κ B proteins. First, it is shown that the *A. queenslandica* NF- κ B protein is phylogenetically and structurally similar to other metazoan NF- κ B proteins, and that its DNA-binding profile, transactivation properties, and subcellular regulation by C-terminal sequences are similar to NF- κ B proteins found in other species. It is also demonstrated that a black encrusting demosponge contains a protein that can bind a conserved κ B site in an EMSA and two proteins that can be detected by an anti-NF- κ B antibody in Western blots and tissue immunofluorescence. Furthermore, treatment of this sponge tissue with LPS leads to further activation of its putative NF- κ B protein, as shown by both increased processing and increased DNA binding. Finally, it is shown that *A. queenslandica* contains homologs to many components of a TLR-to-NF- κ B pathway (Figure 4.9 and Table 4.2). Taken together, these results suggest that sponges have a highly conserved NF- κ B pathway, which is expressed in a subset of cells that likely are involved in an innate immune response.

Table 4.1 Quantification of NF- κ B positive nuclei in DF-1 cells and sponge tissue

Sponge Tissue

Number of nuclei counted	Number of NF- κ B positive nuclei	Percentage NF- κ B positive nuclei/total nuclei
233	111	44%

Localization of the FLAG-tagged protein (NF- κ B positive/total cells counted)

<u>DF-1 cells</u>			
Transfected Plasmid	Nucleus	Cytoplasm	Cytoplasm/Nucleus
FLAG-Nv-NF- κ B	79/79 (100%)	0	0
FLAG-Aq-NF- κ B	0	238/238 (100%)	0
FLAG-Aq-RHD	449/452 (99.3%)	0	3/452 (0.7%)
FLAG-Aq-Cterm	1/170 (0.6%)	168/170 (98.8%)	1/170 (0.6%)

Table 4.2 TLR-to-NF- κ B pathway homologs encoded in the *A. queenslandica* genome.

Protein Name	Blast Accession Number	Homolog Present <i>A. queenslandica</i>	Accession Number of <i>A. queenslandica</i> Homolog	Homologs in <i>A. queenslandica</i> Homolog	Comments
TLR	NP_003254.2	–			
IL-1R		+	XP_003383414.2	1	TIR-only protein with Immunoglobulin-like domains
MYD88	NP_002459.3	+	NP_001266224.1	1	
IRAK	NP_066961.2	+	XP_019855709.1 XP_019855710.1	2	Only has Kinase domain (missing death domains)
TRAF	NP_066961.2	+	XP_019855421.1 XP_019861351.1 XP_003383545.3 XP_019858615.1 XP_019858616.1 XP_019858613.1 XP_003386185.1 XP_003387035.1 XP_019852875.1 XP_019852874.1 XP_003385751.1 XP_019860776.1 XP_019852876.1	13	Most have ring domains and MATH domains, while some have Zn-coil domains and Zn-finger domains
NEMO/IKK γ	NP_003630.1	–			
IKK α	NP_001269.3	–			
IKK β	NP_001547.1	–			
TBK/IKK ϵ	NP_037386.1	+	XP_019854889.1 XP_003387126.1	2	
I κ B	NP_113607.1 NP_004547.2 NP_002494.2	+	XP_011404515.1 XP_019853663.1 XP_019853661.1	3	Epsilon-like Cactus-like
NF- κ B	NP_001070962.1	+	NP_001266238.1	1	
RelB	NP_006500.2	–			

Listed are select proteins in the TLR–NF- κ B pathway (see Figure 4.9) and their Accession numbers on NCBI. “Blast Accession Number” indicates the protein sequence used to identify the *A. queenslandica* homolog. Comments provide details of domain structures of these various sequences.

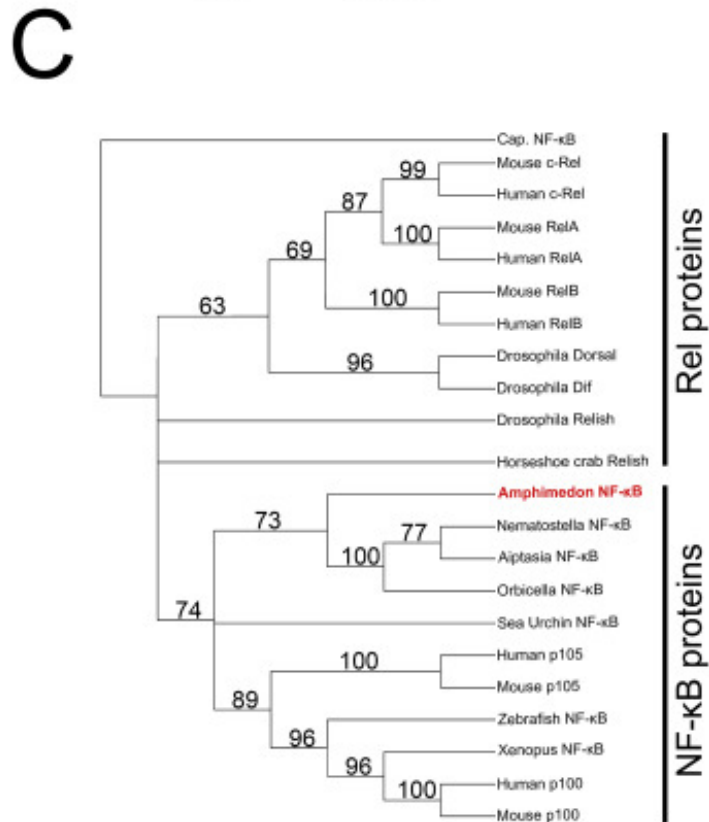
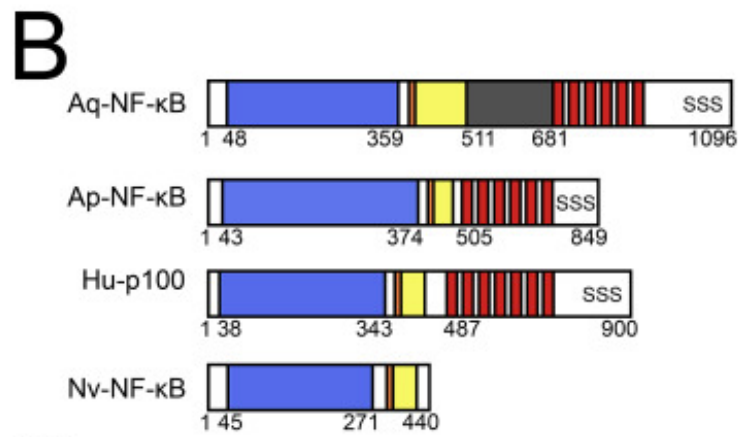
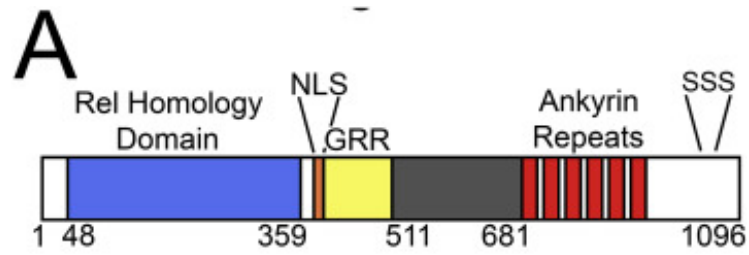


Figure 4.1 The *A. queenslandica* NF- κ B protein resembles other metazoan NF- κ B proteins both structurally and phylogenetically.

(A) Schematic of Aq-NF- κ B domains: Rel Homology Domain (RHD, blue), Nuclear Localization Sequence (NLS, orange), Glycine-Rich Region (GRR, yellow), extra sequences after GRR (grey), Ankyrin Repeats (ANK, red), and conserved C-terminal serines (SSS). (B) Schematic comparing domain structures of Aq-NF- κ B (1096 amino acids), the previously characterized sea anemone *Aiptasia* NF- κ B protein (Ap-NF- κ B), human p100, and the naturally truncated *N. vectensis* NF- κ B (Nv-NF- κ B). Functional domains shared between these proteins are marked by colors, as in (A). (C) Phylogenetic analysis of RHD sequences of the indicated NF- κ B proteins was performed using Maximum Likelihood analysis. The phylogeny was rooted with the predicted RHD of the single-celled protist *C. owczarzaki*. Branches indicate bootstrap support values.

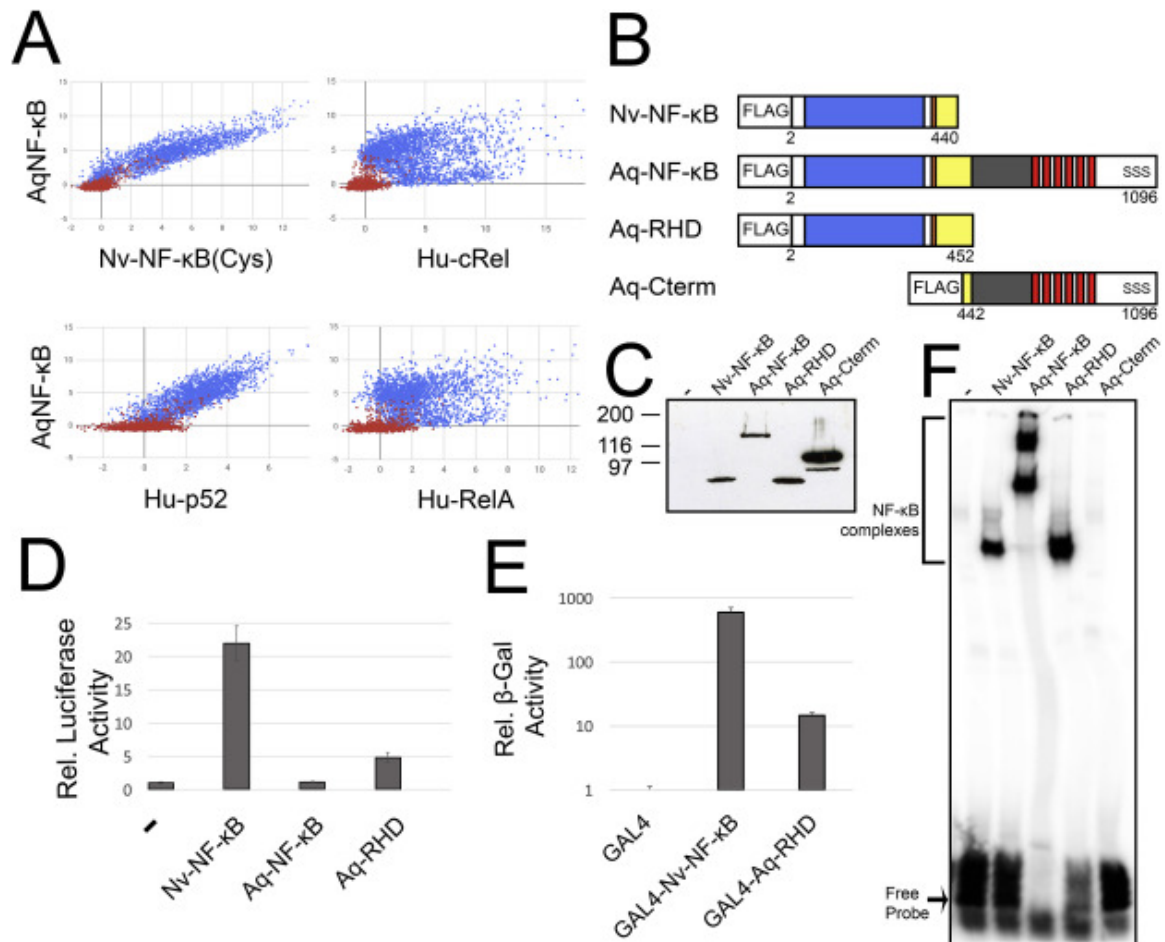


Figure 4.2 Aq–NF–κB contains sequences that can bind to DNA and activate transcription.

(A) Comparison of the DNA-binding profile of Aq–NF–κB in a protein binding microarray (PBM) to the *N. vectensis* (Nv) NF-κB (cysteine (Cys) allele) (top, left), human c-Rel (top, right), p52 (bottom, left), and RelA (bottom, right). Red dots represent random background sequences and blue dots represent κB binding sites. Z-scores are compared. (B) FLAG-tagged expression vectors used in these experiments. From top to bottom, the drawings depict the naturally shortened Nv–NF–κB, full-length Aq–NF–κB, an N-terminal-only mutant containing the RHD and glycine-rich region (Aq-RHD), and a

C-terminal-only mutant containing the ANK repeats and other C-terminal sequences (Aq-Cterm). (C) Anti-FLAG Western blot of 293T cell lysates transfected with the indicated expression vectors. (D) A κ B-site luciferase reporter gene assay was performed with the indicated proteins in 293 cells. Luciferase activity is relative (Rel.) to that seen with the empty vector control (1.0), and values are averages of three assays performed with triplicate samples with standard error. (E) A GAL4-site *LacZ* reporter gene assay was performed in yeast Y190 cells. β -galactosidase (β -gal) reporter gene activity is relative (Rel.) to the GAL4 (aa 1–147) control (1.0) and is presented on a log-scale. Values are averages of seven assays performed with duplicate samples with standard error. (F) A κ B-site electromobility shift assay (EMSA) using each of the indicated lysates, as in (C). The NF- κ B complexes and the free probes are indicated.

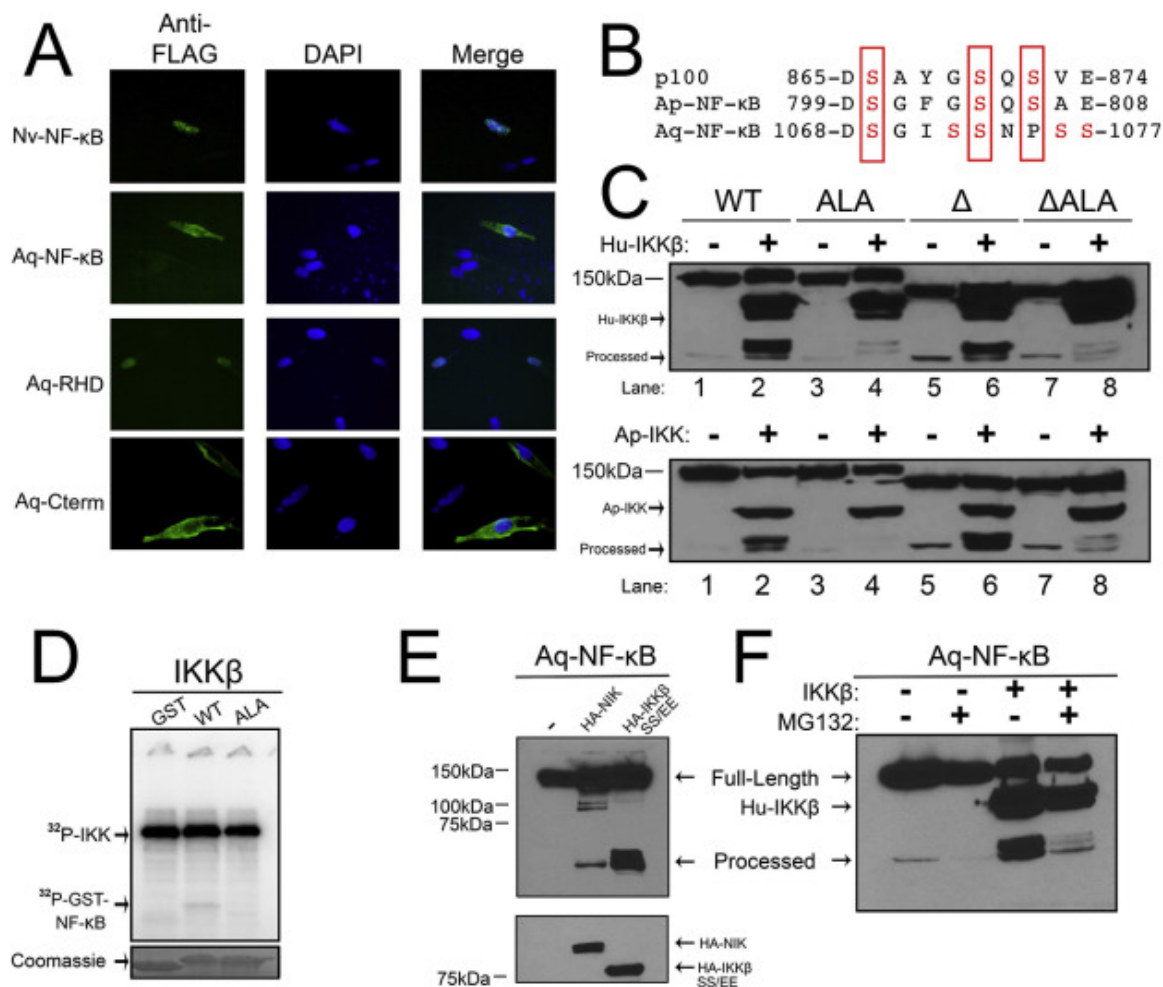


Figure 4.3 C-terminal sequences of Aq-NF- κ B inhibit nuclear translocation and can undergo IKK-dependent degradation and phosphorylation.

(A) DF-1 chicken fibroblasts were transfected with the indicated FLAG vectors and subjected to anti-FLAG indirect immunofluorescence. Anti-FLAG, green (left); DAPI-stained nuclei, blue (middle); and merged images (right). For quantitation of staining see Table 4.1 (B). An alignment of relevant sequences of NF- κ B from human p100, the sea anemone *Aiptasia*, and *A. queenslandica*. Conserved serine residues are indicated in red. (C) Western blot of lysates from 293T cells co-transfected with Aq-NF- κ B wild-type

(WT), C-terminal alanine mutant (ALA), aa 511–681 deletion mutant (Δ), or double mutant (Δ ALA) and with empty vector (–), FLAG-Hu-IKK β (+, top) or FLAG-Ap-IKK (+, bottom). Full-length and processed forms of Aq–NF- κ B, and FLAG-Hu-IKK β are indicated. Lanes are numbered at the bottom. (D) *In vitro* kinase assay (top) using FLAG-Hu-IKK β protein with bacterially expressed substrates of GST alone or GST fusion proteins containing the wild-type (WT) serine residues of Aq–NF- κ B (aa 1048–1086), or the same C-terminal residues with conserved serines mutated to alanines (ALA). The GST proteins were electrophoresed on an SDS-polyacrylamide gel and stained with Coomassie blue (bottom). (E) Anti-FLAG Western blot of 293T cells co-transfected with FLAG-Aq–NF- κ B and with empty vector (–), HA-NIK, or HA-IKK β SS/EE. The full-length and processed forms of Aq–NF- κ B are indicated (top). The same membrane was stripped and probed using an anti-HA antiserum to detect NIK and IKK β SS/EE (bottom). (F) Anti-FLAG Western blot of 293T cells co-transfected with FLAG-Aq–NF- κ B and either empty vector (–) or FLAG-Hu-IKK β . Cells were untreated (–) or pre-treated (+) with MG132. Full-length and processed Aq–NF- κ B and IKK β are indicated.

A

Ap-NF- κ B	799-D	S	G	F	G	S	Q	S	A	E-808
Aq-NF- κ B	1068-D	S	G	I	S	S	N	P	S	S-1077
Aq-NF- κ B Aiptasia serine mutant	1068-D	S	G	F	G	S	Q	S	A	E-1077

B

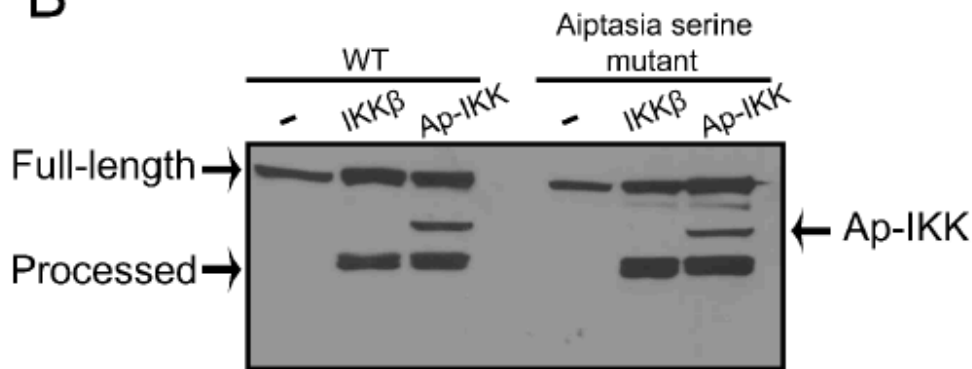


Figure 4.4 IKKs can induce processing of both WT Aq-NF- κ B and Aq-NF- κ B-Aiptasia serine mutant.

(A) Amino acid (aa) sequences for IKK-induced phosphorylation and processing in Ap-NF- κ B (top) aligned with Aq-NF- κ B aa sequence (middle). Aq-NF- κ B-Aiptasia serine mutant (bottom) was created by changing Aq-NF- κ B aa 1068-1077 to aa 799-808 from Ap-NF- κ B. Shown in red are relevant serines. (B) Western blot of lysates from 293T cells co-transfected with Aq-NF- κ B wild-type (WT) or Aq-NF- κ B-Aiptasia serine mutant and with empty vector (-) or HA-Hu-IKK β or FLAG-Ap-IKK (bottom). Full-length and processed forms of Aq-NF- κ B, and FLAG-Ap-IKK are indicated.

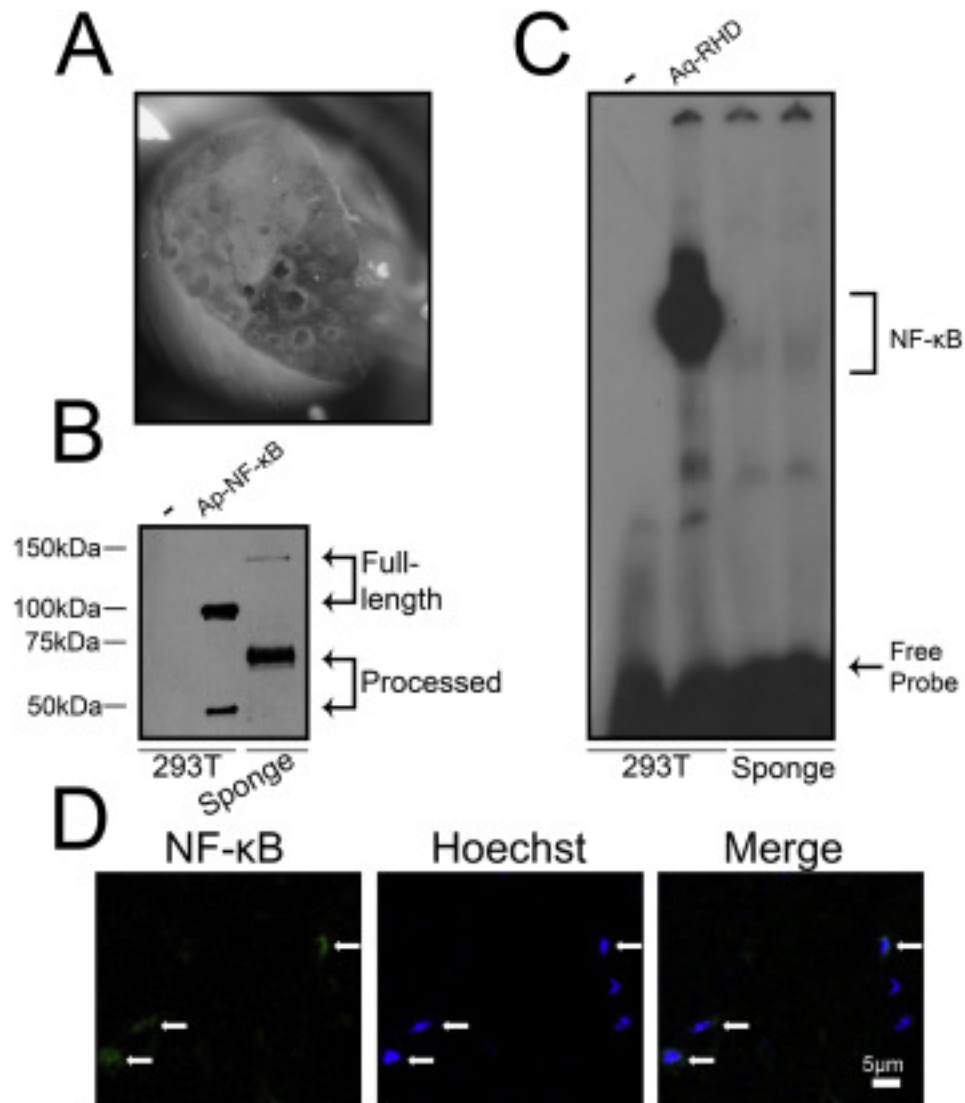


Figure 4.5 Demosponge tissue extract contains an NF- κ B-like protein that has DNA-binding activity and localizes to the nucleus.

(A) Black encrusting demosponge (*Cliona sp.*) tissue under a light microscope. (B) Anti-Ap-NF- κ B Western blot of lysates from 293T cells transfected with either empty vector (first lane) or FLAG-Ap-NF- κ B (middle lane), and a lysate from black encrusting sponge tissue (third lane). Molecular weight markers are indicated to the left of the blot in kilodaltons. (C) A κ B-site EMSA with lysates from 293T cells transfected with either

empty vector (–) or FLAG-Aq-RHD, or with tissue extracts from black encrusting sponge tissue (Sponge). The shifted protein-DNA complexes and free probes are indicated. (D) Confocal microscopy of cryosectioned sponge tissue identifies cells that express NF- κ B, which is primarily localized to the nucleus. NF- κ B is in green (left panel), Hoechst-stained nuclei are shown in blue (middle panel), and a merged image showing co-localization (right panel). For quantitation of staining see Table S4.

A

Top five BLAST hits	Description	E value	Percent Identity	Accession
1	Biemna saucia voucher QM:G303281 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial	0.002	100%	JF773146.1
2	Cliona celata voucher MNRJ 741 cytochrome oxidase subunit I (cox1) gene, partial cds; mitochondrial	0.002	100%	HM999030.1
3	Cliona chilensis voucher MNRJ 9268 cytochrome oxidase subunit I (cox1) gene, partial cds; mitochondrial	0.002	100%	HM999022.1
4	Cliona chilensis voucher MNRJ 10414 cytochrome oxidase subunit I (cox1) gene, partial cds; mitochondrial	0.002	100%	HM999018.1
5	Cliona chilensis voucher MNRJ 10412 cytochrome oxidase subunit I (cox1) gene, partial cds; mitochondrial	0.002	100%	HM999017.1

B

Sponge	Color	Geographic Location	Morphology	Sequencing Data Available
Black Encrusting Sponge	Black/brown	Caribbean	Encrusting/ papillated	?
Biemna saucia	yellow/brown	Australia	Fibrous/ stringy	COI
Cliona celata	yellow/orange	Bahamas, Western Atlantic ocean	Encrusting/ papillated	COI, Ribosomal
Cliona chilensis	orange	Southwest Atlantic and Southeast Pacific: Argentina and Chile	Encrusting/ papillated	COI, Ribosomal
Clinoa caribbaea	Black/brown	Western Central Atlantic: Belize	Encrusting/ papillated	none

Figure 4.6 The black encrusting sponge most closely resembles a demosponge of the genus *Cliona*.

(A) The top five BLAST hits from the black encrusting sponge COI PCR product (see text and Materials and Methods). (B) A comparison of the morphologies and geographical locations of three of the top five BLAST hits from the black encrusting sponge COI gene analysis (in A). Although *Clinoa caribbaea* is not in the NCBI

databases, our black encrusting sponge is most likely this sponge based on color, morphology, geographical location, and its relationship to other *Cliona* sp.

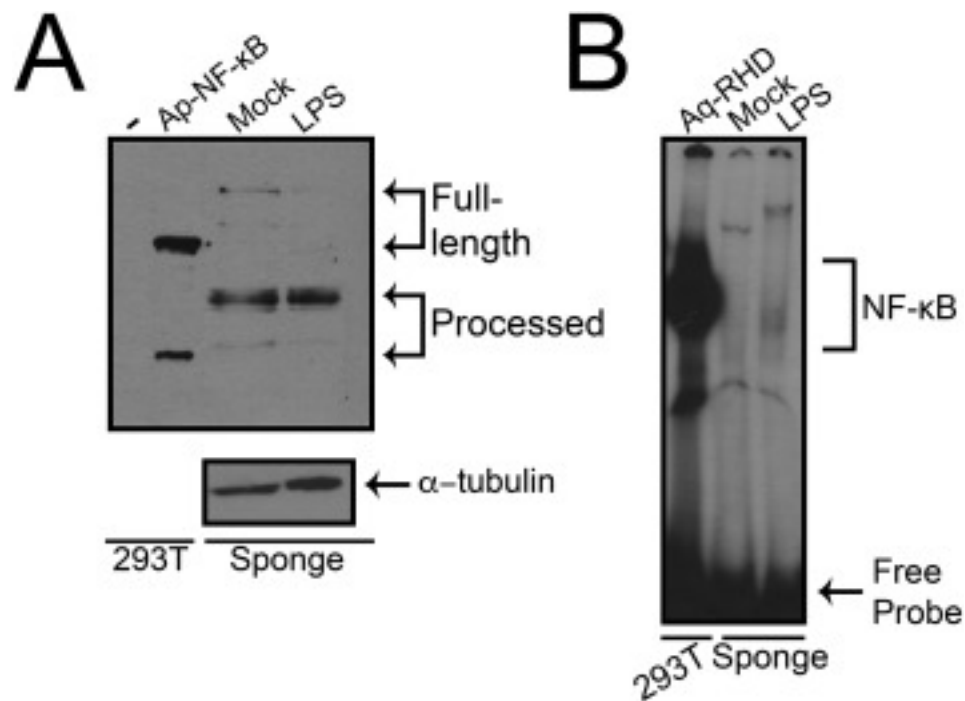


Figure 4.7 Treatment of sponge tissue with LPS results in increased NF-κB protein processing and activity.

Black encrusting sponge tissue was treated with control water (mock) or *E. coli*-derived lipopolysaccharide (LPS) for 30 min. Lysates were then prepared and analyzed by anti-NF-κB Western blotting (A) or EMSA (B). (A) The left-most two lanes contain lysates from 293T cells transfected with empty vector (-) or FLAG-Ap-NF-κB, respectively. Lanes three and four contain black encrusting sponge tissue extracts that were from either control treated (mock) or LPS-treated (LPS) tissue. α-tubulin Western blotting is shown below for normalization purposes. (B) A κB-site EMSA with a 293T cell lysate

expressing FLAG-Aq-RHD (first lane), or with extracts from mock-treated or LPS-treated black encrusting sponge tissue (third and fourth lanes, respectively). The NF-κB-DNA complexes and the free probes are indicated.

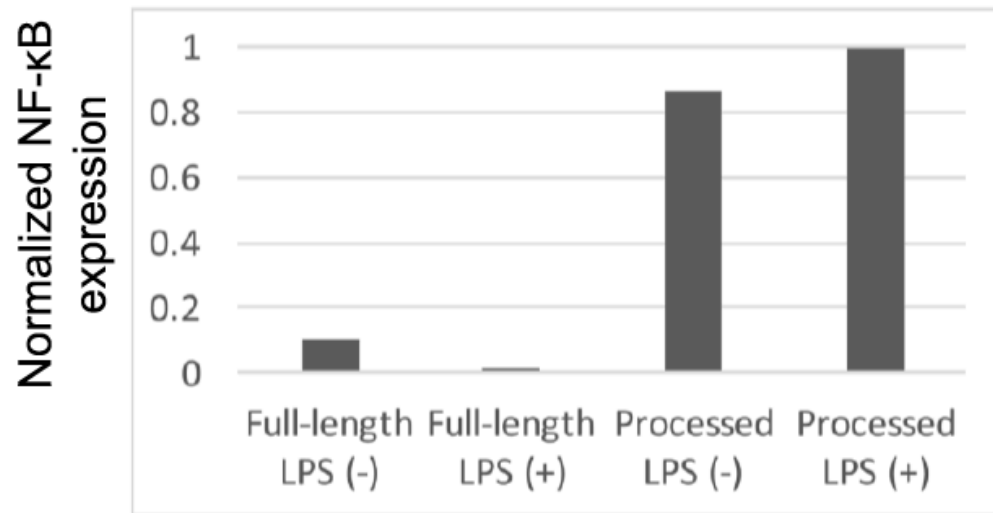


Figure 4.8 NF-κB processing is induced upon LPS stimulation in black encrusting sponge tissue.

The amount of NF-κB protein as compared to α -tubulin protein, as a percentage of processed NF-κB following control treatment or treatment with LPS.

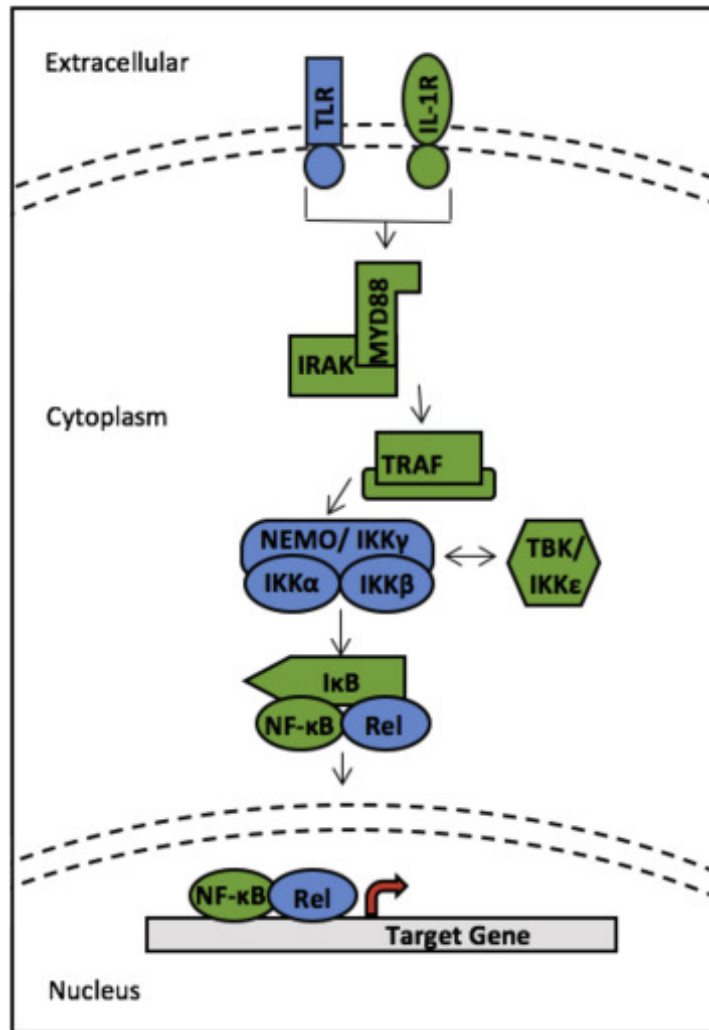


Figure 4.9 NF- κ B pathway proteins and upstream receptors identified in *A. queenslandica*.

NF- κ B pathway members that are present in humans are colored in blue, and homologs that are also encoded in the *A. queenslandica* genome (see Table 4.2) are colored in green.

CHAPTER FIVE

DIVERSIFICATION OF TRANSCRIPTION FACTOR NF- κ B IN PROTISTS⁵

5.1 Introduction

Transcription factor NF- κ B (Nuclear Factor- κ B) has been extensively studied for its roles in development and immunity in animals from sponges to humans (Gilmore, 2006; Ghosh and Hayden, 2012; Williams and Gilmore, 2020). Only within the last few years has it been discovered that the appearance of NF- κ B pre-dates metazoan life; that is, that certain single-cell eukaryotes--namely some choanoflagellates and the holozoan *Capsaspora owczarzaki*--also contain genes encoding NF- κ B-like proteins (Suga et al., 2013; Richter et al., 2018). In this chapter, the first functional characterization of this important transcription factor in single-celled protists is presented.

Protists comprise a large group of eukaryotes that are either unicellular or multicellular with poorly differentiated tissue, and they make up one of the six major Kingdoms (Archibald et al., 2017). Presumably, protists have a common ancestor, but they are now known to be an extensively diverse collection of organisms with several major supergroups. Two protists that have been studied reasonably well are *Capsaspora* and the taxonomic class of choanoflagellates, both of which are in the Opisthokonta subgroup.

Capsaspora is a single-celled eukaryote that is thought to be among the closest unicellular relatives to animals (i.e., basal to sponges) and is the sister group to the Filozoa (the clade comprising metazoans and choanoflagellates) (Ferrer-Bonet and Ruiz-

⁵Adapted from Williams et al., 2020 *bioRxiv*

Trillo, 2017). *Capsaspora* was originally discovered as an amoeba-like symbiont in the hemolymph of the fresh-water snail *Biomphalaria glabrata*. *Capsaspora* kills sporocysts of the flatworm *Schistosoma mansoni*, the causative agent of schistosomiasis in humans, which also inhabits *B. glabrata* (Stibbs et al., 1979). More recently, the life cycle of *Capsaspora* has been shown to contain three different cell configurations (Stibbs et al., 1979). Under *in vitro* culture conditions, *Capsaspora* grow primarily as filopodial cells, which attach to the substrate and undergo active replication until the end of the exponential growth phase. Then, cells start to detach, retracting their branching filopodia and encysting. During this cystic phase, cell division is stopped. Alternatively, filopodia cells can form a multicellular aggregative structure by secreting an unstructured extracellular matrix that promotes aggregation but prevents direct cell-cell contact (Sebé-Pedrós et al., 2013). The genome of *Capsaspora* contains many genes involved in metazoan multicellular processes including integrins, protein tyrosine kinases, and transcription factors, including NF- κ B (Suga et al., 2013). Furthermore, RNA-sequencing has revealed that each life stage contains distinct transcriptomic profiles (Sebé-Pedrós et al., 2013). However, the details of how these life-stage processes and transitions are carried out on the molecular level in a unicellular organismal context remain unclear.

A second class of protists of interest for evolutionary biologists includes the choanoflagellates. These flagellated eukaryotes comprise over 125 species of free-living unicellular and colonial organisms distributed in nearly every aquatic environment (King, 2005). Choanoflagellates are also widely regarded as being close living relatives to the animals, and they are capable of asexual and sexual reproduction (King, 2005). The

feeding of choanoflagellates on bacteria provides a critical ecological role within the global carbon cycle by linking trophic levels. Until somewhat recently, little was known about the genomic diversity of choanoflagellates, with only two published genomes of *Monosiga brevicollis* and *Salpingoeca rosetta* (King et al., 2008; Fairclough et al., 2013). Neither genome of these species contains homologs to NF- κ B, and it was thought for nearly a decade that this transcription factor had been lost in the evolutionary branch containing choanoflagellates. However, in 2018, Richter et al. (2018) reported the transcriptomes of 19 additional choanoflagellates. Of these 19 species of choanoflagellates, 12 expressed one-to-three NF- κ B-like transcripts. Amazingly, sequence comparisons have revealed that choanoflagellates are generally as genetically distant from each other as a mouse is from a sea sponge, a testimony to the modern-day diversity among these taxa (Richter et al., 2018). Among basal organisms -including cnidarians, poriferans, and some protists- only NF- κ B-like proteins have been found (Williams and Gilmore, 2020). Indeed, no Rel proteins have been identified in any organism basal to flies.

Transcriptomic and genomic sequencing has revealed that NF- κ B and homologs of many of its upstream regulators are present in most eukaryotes from protists to vertebrates (Wolenski and Gilmore, 2012; Williams and Gilmore, 2020). However, the numbers and structures of these signaling proteins vary across species, and generally become more complex and numerous through evolutionary time (Wolenski and Gilmore, 2012). Furthermore, in the most basal groups of metazoans (cnidarians and sponges), only single homologs to NF- κ B exist within their genomes (Williams and Gilmore,

2020).

In this chapter, the molecular functions of transcription factor NF- κ B in two unicellular protists are characterized using phylogenetic, cellular, and biochemical techniques. As with the human p100 protein, it is shown that some unicellular NF- κ B proteins require removal of C-terminal ANK repeats to enter the nucleus and bind DNA. However, Co-NF- κ B does not undergo IKK-mediated processing, and homologs to IKK do not exist in *Capsaspora* or choanoflagellates. Furthermore, it is shown that the multiple NF- κ Bs of a single choanoflagellate can form heterodimers, a first finding in an organism outside of the kingdom Animalia, suggesting that choanoflagellates contain their own subclasses of NF- κ Bs, much like in vertebrates and flies. These results are the first functional characterization of NF- κ B in a taxonomic kingdom outside Animalia.

5.2 Protist NF- κ B proteins vary in domain structure and choanoflagellates show evidence of gene duplication

Suga et al. (2013) reported the presence of a single gene encoding an NF- κ B-like protein in *C. owczarzaki* (Co). The overall protein structure of Co-NF- κ B is similar to most other basal NF- κ B proteins known to date, in that it has an N-terminal RHD, followed by a glycine-rich region (GRR), and five C-terminal ANK repeats (Wolenski et al., 2013; Mansfield et al., 2017; Williams et al., 2018; Williams et al., 2020). However, Co-NF- κ B is larger than other NF- κ B homologs, due primarily to additional residues C-terminal to the ANK repeats (Figure 5.1A). Recently, Richter et al. (2018) showed that the transcriptomes of several choanoflagellates had NF- κ B-like genes, even though NF-

κ B-like genes are not present in two commonly studied choanoflagellates (*M. brevicollis* and *S. rosetta*) (King et al., 2008; Fairclough et al., 2013). Overall, 12 of the 21 choanoflagellates are now known to contain NF- κ B-like genes, and among those 12, there are one, two, or three NF- κ B transcripts present.

A phylogenetic comparison suggests that many of the choanoflagellate NF- κ Bs arose from gene duplications within a given species because the multiple NF- κ Bs from a given species often cluster closely to each other (for example, *Diaphanoeca grandis* and *Salpingoeca helianthica*) (Figure 5.1B, Table 5.1). Nevertheless, there are some choanoflagellates that have multiple NF- κ Bs that cluster separately with the NF- κ Bs of other choanoflagellate (e.g., *Acanthoeca spectabilis* and *Savillea parva*).

In contrast to what is seen in most basal metazoans, these choanoflagellates express transcripts that primarily encode RHD sequences, with no C-terminal GRRs or ANK repeats. However, some choanoflagellate NF- κ Bs do contain extended N termini with homology to sequences not normally associated with NF- κ Bs in vertebrates (Figure 5.1A, pink bar).

5.3 DNA binding, nuclear translocation, and transactivation by Co-NF- κ B

To investigate the overall DNA binding-site specificity of Co-NF- κ B, we first characterized the activity of a bacterially expressed Co-NF- κ B RHD-only protein by protein binding microarray (PBM) analysis on an array containing 2592 κ B-like sites and 1159 random background sequences (for complete array probe sequences, see Mansfield et al., 2017). By comparison of the z-scores for binding to DNA sites on the PBM, the

DNA-binding profile of Co-NF- κ B is most similar to NF- κ Bs from the sea anemone *N. vectensis* and human p52, and it is distinct from human c-Rel and RelA (Figure 5.2A).

To investigate properties of Co-NF- κ B in cells, we created pcDNA-FLAG vectors for full-length Co-NF- κ B and two truncation mutants, one (FLAG-Co-RHD) containing the N-terminal RHD sequences including the NLS and the GRR, and a second (FLAG-Co-Cterm) consisting of the C-terminal ANK repeat sequences and downstream residues (Figure 5.2B). As a control, we used the active, naturally truncated *N. vectensis* (Nv) FLAG-tagged Nv-NF- κ B protein that we have characterized previously (Wolenski et al., 2011) (Figure 5.2B). As shown by anti-FLAG Western blotting, each plasmid expressed a protein of the appropriate size when transfected into HEK 293T cells (Figure 5.2C).

In sponge and some cnidarian NF- κ Bs, removal of C-terminal ANK-repeat sequences are required for nuclear localization in vertebrate cell-based assays (Mansfield et al., 2017; Williams et al., 2018; Williams et al., 2020). Based on those results, we next transfected each FLAG expression plasmid into DF-1 chicken fibroblast cells and performed indirect immunofluorescence using anti-FLAG antiserum (Figure 5.2D; Table 5.1). Full-length Co-NF- κ B and Co-Cterm were both located primarily in the cytoplasm of these cells (99.9% and 94%, respectively). In contrast, the Co-RHD and control Nv-NF- κ B proteins were both primarily nuclear, as evidenced by co-localization with the Hoechst-stained nuclei (Figure 5.2D; Table 5.1). Thus, the removal of the ANK repeats allows Co-NF- κ B to enter the nucleus, consistent with what is seen with other metazoan RHD-ANK bipartite NF- κ B proteins.

To further assess the DNA-binding activity of Co-NF- κ B proteins, whole-cell

extracts from 293T cells transfected with each of the FLAG constructs were analyzed in an electrophoretic mobility shift assay (EMSA) using a high affinity κ B-site probe. Extracts containing overexpressed Nv-NF- κ B and Co-RHD bound the κ B site avidly, whereas extracts containing full-length Co-NF- κ B and Co-Cterm showed essentially no κ B site-binding activity (Figure 5.2E).

We also assessed the ability of Co-NF- κ B proteins to activate transcription in reporter gene assays in HEK 293 cells using a κ B-site reporter. Co-RHD and Nv-NF- κ B activated transcription well above control levels (i.e., Co-RHD was ~60-fold above the negative control; Figure 5.2F). In contrast, full-length or Cterm Co-NF- κ B proteins showed little to no ability to activate transcription. From these data, the ability to activate transcription of a κ B site gene locus appears to be a property of sequences within the N-terminal half of Co-NF- κ B. We also assessed the ability of Co-RHD to activate transcription in reporter gene assays in yeast cells using a GAL4-site reporter. Indeed, the N-terminal half (RHD) of Co-NF- κ B activated transcription strongly, nearly 1000-fold above the GAL4 (aa 1-147) alone negative control. The transactivation ability of the GAL4-RHD sequences of Co-NF- κ B in yeast suggests that this is an intrinsic property of these sequences.

Taken together, the results in this section show that Co-RHD can bind DNA, activate transcription, and localizes primarily to the nucleus, unlike the inactive full-length Co-NF- κ B protein, consistent with findings with most NF- κ Bs from sponges to humans when assayed in vertebrate cells (Wolenski et al., 2011; Mansfield et al., 2017; Williams et al., 2018; Williams et al., 2020).

5.4 IKK-mediated processing of NF- κ B appears to have evolved with the rise of multicellularity

As is described above for Co-NF- κ B, vertebrate NF- κ B p100 requires the removal of its C-terminal ANK repeats to enter the nucleus and activate transcription (Sun, 2011). This proteasome-mediated processing of p100 is initiated by phosphorylation of a C-terminal cluster of serine residues by an I κ B kinase (IKK) (Sun, 2011). We have previously shown that some basal organisms, including NF- κ B proteins from two cnidarians and one sponge, contain homologous C-terminal serines that can be phosphorylated by IKKs to initiate proteasome-mediated processing in human cell culture assays (Mansfield et al., 2017; Williams et al., 2018; Williams et al., 2020). Examination of the C-terminal sequences of Co-NF- κ B failed to identify any C-terminal serine clusters similar to other NF- κ Bs that undergo IKK-initiated processing. Nevertheless, we performed a series of experiments that examined the ability of IKK to induce processing of Co-NF- κ B by co-transfecting HEK 293T cells with Co-NF- κ B and several IKK proteins, including two from humans and one from a sea anemone (Mansfield et al., 2017; Williams et al., 2018; Williams et al., 2020). In all cases, co-expression of the IKK did not induce processing of Co-NF- κ B (Figure 5.3A), beyond the small amount of constitutive processing of Co-NF- κ B that occurs even in the absence of IKK (Figure 5.3A and Figure 5.2C). Of note, the lower Co-NF- κ B band seen in these extracts was roughly the same size as the predicted RHD (Figure 5.2C), and incubation of transfected cells with the proteasome inhibitor MG132 reduced the appearance of the lower band, suggesting that it arises by proteasomal processing of full-length Co-NF- κ B (Figure 5.4).

To determine whether Co-NF- κ B *could be* processed by an IKK-dependent mechanism, we created a mutant in which we replaced C-terminal sequences Co-NF- κ B (downstream of the ANK repeats) with C-terminal sequences of the sea anemone *Aiptasia* (Ap)-NF- κ B that contain conserved serines which can facilitate IKK-induced processing of Ap-NF- κ B (Mansfield et al., 2017). We termed this mutant Co-NF- κ B-SER, and also created the analogous protein (Co-NF- κ B-ALA) in which the relevant serines were replaced by alanines. Co-expression of Co-NF- κ B-SER with constitutively active human IKK β (IKK β SS/EE) resulted in increased amounts of the lower band, which was not seen with Co-NF- κ B-ALA (Figure 5.3B). Thus, the Co-NF- κ B protein (consisting of the RHD, GRR, and ANK repeats) can undergo IKK-induced processing if supplied with a C terminus containing the IKK target serine residues. However, the native Co-NF- κ B protein does not appear to be susceptible to IKK-induced processing, which is also consistent with the lack of any IKK sequences in the genome of *Capsaspora*.

5.5 Exogenously expressed full-length and truncated versions of Co-NF- κ B localize primarily to the nucleus in *Capsaspora* cells

We were next interested in examining the subcellular localization of NF- κ B in *Capsaspora* cells. For these experiments, we transfected *Capsaspora* cells with our FLAG-tagged Co-NF- κ B constructs (Co-NF- κ B, mutant Co-RHD, and mutant Co-Cterm, see Figure 5.2D) and then performed anti-FLAG indirect immunofluorescence.

Consistent with results seen in DF-1 chicken fibroblast cells (Figure 5.2D), FLAG-Co-

RHD and FLAG-Co-Cterm localize to the nucleus and cytoplasm, respectively (Figure 5.5, middle and bottom rows). Surprisingly, full-length FLAG-Co-NF- κ B also appeared to be fully nuclear, as judged by its co-localization with the Hoechst-stained nuclei (Figure 5.5, top row, and Figure 5.6). We noted that the FLAG-tagged RHD and full-length Co-NF- κ B appeared as puncta within the nucleus of transfected *Capsaspora* cells, which may be indicative of active transcriptional sites (Brasch and Ochs, 1992).

5.6 Co-NF- κ B mRNA levels and DNA-binding activity vary coordinately across different life stages and the identification of putative NF- κ B target genes

Capsaspora has been shown to have three different life stages: aggregative, filopodia, and cystic, and RNA-Seq of each life stage has been reported (Sebé-Pedrós et al., 2013). We were interested in determining whether *Capsaspora* NF- κ B protein was active at different levels in these three life stages and whether we could use the previous RNA-Seq data to identify genes whose expression may be controlled by NF- κ B. We first examined previous mRNA expression data (Sebé-Pedrós et al., 2013) for NF- κ B mRNA, and found that NF- κ B was expressed at the lowest level in the aggregative stage and 2.3- and 5-fold higher in the filopodic and cystic stages, respectively (Figure 5.7A). We next generated cultures of *Capsaspora* at each life stage (Figure 5.7A), made protein extracts, and performed an EMSA using the κ B-site probe that we showed can be bound by Co-NF- κ B expressed in HEK 293T cells (Figure 5.2E) and by bacterially expressed Co-RHD in our PBM assays (see Mansfield et al., 2017 for details on PBM arrays). Consistent with the mRNA expression data, the κ B-site probe was bound progressively stronger in

aggregative, filopodic, and cystic stages (Figure 5.7B). To determine whether the EMSA band indeed included Co-NF- κ B, we incubated our protein extracts with 10- and 25-fold excesses of unlabeled κ B-site probe, and we saw a substantial decrease in binding in the putative NF- κ B band. In contrast, incubation with a 25-fold excess of an unlabeled IRF-site probe did not decrease the putative Co-NF- κ B complex, indicating that the binding was specific for the κ B-site probe.

We next sought to identify genes that might be influenced by the expression of NF- κ B in order to gain insight into potential functional roles for NF- κ B in these life stages. We examined the existing RNA-Seq data (Sebé-Pedrós et al., 2013), which contains the mRNA expression of 8674 genes of *Capsaspora* during its three life stages. We first narrowed our gene list to those genes that were differentially expressed in a manner similar to NF- κ B mRNA levels and DNA-binding activity during each life stage (i.e., progressively increased in expression from aggregative, filopodic, and cystic stages). From this analysis, we identified 1348 mRNAs that were expressed at low levels in the aggregative stage, and progressively higher levels in the filopodic and cystic stages.

Of the 1348 genes that we identified with life-stage expression patterns similar to Co-NF- κ B, 389 genes were annotated (which is consistent with approximately 1/4 of *Capsaspora*'s total predicted protein-encoding genes being annotated, Table 5.2), and 305 of these 389 annotated genes had human homologs (Table 5.2). We then performed GO analysis to look at Biological Processes overrepresented in these 305 genes. That analysis showed that this set of 305 genes was predicted to be involved in several

biological processes, including 14 genes that encode proteins associated with developmental and immune system processes (Figure 5.7C), which are biological processes regulated by NF- κ B in many more complex organisms and suggested to be regulated by NF- κ B in several basal organisms (Steward, 1987; Roth et al, 1989; Kappler et al., 1993; Reichhart et al., 1993; Mansfield et al., 2017; Williams et al., 2018; Williams et al., 2020). Although 14 genes may seem low, the total database for human GO analysis of immune system and developmental processes genes is approximately 2200 genes, but the number of annotated homologs that exist in *Capsaspora* in these two GO categories is only 66 genes (Table 5.3). Thus, about 20% (14/66) of the *Capsaspora* genes in the GO category for the developmental and immune processes subset are among the 305 homologous genes coordinately regulated with NF- κ B. Other broad categories overrepresented in these 305 genes included Signaling, Metabolic Process, and Locomotion (Figure 5.8).

We then looked for NF- κ B binding sites within 500 bp upstream of the transcription start sites (TSS) for each of the 1348 genes with expression profiles that were similar to NF- κ B. 192 of these 1348 gene upstream regions contained one, two or three κ B sites within 500 bp of the TSS, with the majority of these genes containing 1 κ B site (Table 5.4). Two of the 14 genes that encode protein homologs associated with GO developmental and immune system processes contained a κ B site within the 500 bp upstream of their TSS (Figure 5.7C). Of the 192 genes with κ B sites, 60 are annotated. The GO terms within these 60 genes include similar processes, including Development and Immune System (GO:0032502 and GO:0002376, respectively).

5.7 Choanoflagellate NF- κ Bs can form heterodimers and have different abilities to bind DNA and activate transcription

Richter et al. (2018) sequenced the transcriptomes of 19 choanoflagellates and identified RHD-containing NF- κ B-related proteins in 12 of these species. We chose to characterize the NF- κ B proteins from *Acanthoeca spectabilis* (As) because it has three NF- κ B-like proteins, which separated into multiple branches when phylogenetically compared to all choanoflagellate NF- κ Bs (Figure 5.1B). These three As-NF- κ B proteins contained ostensibly complete DNA-binding sequences, which were similar to other NF- κ B proteins and a putative NLS (Williams and Gilmore, 2020). These three proteins contained extended sequences N-terminal to the RHD, but they contained few C-terminal residues beyond the RHD (and no GRRs or ANK repeats). As a first step in characterizing these proteins, we created pcDNA FLAG vectors for As-NF- κ B1, As-NF- κ B2, and As-NF- κ B3 (Figure 5.9A) and transfected them into HEK 293T cells. As assessed by anti-FLAG Western blotting, each vector directed the expression of appropriately sized FLAG-tagged proteins (Figure 5.9B).

To determine the subcellular localization properties of the three As-NF- κ Bs, we performed indirect immunofluorescence on DF-1 chicken fibroblasts and *Capsaspora* cells transfected with each FLAG-tagged vector. All three As-NF- κ B proteins co-localized with Hoechst-stained nuclei in both DF-1 cells (Figure 5.9C) and *Capsaspora* cells (Figure 5.9D). As with FLAG-tagged Co-NF- κ B proteins, the FLAG-tagged As-NF- κ B proteins appeared as puncta within the nucleus of transfected *Capsaspora* cells.

We then performed a κ B-site EMSA on whole cell extracts from HEK 293T cells

overexpressing each As-NF- κ B, using Co-RHD as a positive control. As-NF- κ B2 and 3 bound the κ B-site probe to nearly the same extent as Co-RHD, but As-NF- κ B1 only weakly bound the probe (Figure 5.9E). We also assessed the transactivating ability of each As-NF- κ B protein in a κ B-site reporter assay in HEK 293T cells, using the strongly activating Co-RHD protein as a positive control. Both As-NF- κ B2 and 3 were able to activate transcription of the luciferase reporter above vector control levels (~1.6- and 2.7-fold, respectively) but As-NF- κ B1 did not (Figure 5.9F). We also assessed the intrinsic transactivating ability of each As-NF- κ B protein using a GAL4-fusion reporter assay in yeast cells. In this assay, all three As-NF- κ Bs activated transcription over vector control levels, although As-NF- κ B1 and 3 activated to a much lesser degree than As-NF- κ B2 (Figure 5.9G). From these data, we hypothesized that the homodimeric version of As-NF- κ B1 was not capable of binding DNA and could not activate transcription using κ B sites, but likely contained some intrinsic ability to activate transcription (as a GAL4-fusion protein), whereas homodimeric As-NF- κ B2 and 3 could activate transcription both in human cell-based and yeast GAL4-fusion reporter assays.

Since As-NF- κ B1 did not substantially bind DNA or activate transcription when transfected alone, we hypothesized that As-NF- κ B1 ordinarily acts as a heterodimer with the other As-NF- κ Bs. Therefore, we performed a series of co-immunoprecipitation experiments to determine whether As-NF- κ B1 could interact with As-NF- κ B2 or As-NF- κ B3. To do this, we subcloned As-NF- κ B2 and 3 into MYC-tagged vectors, co-transfected each with FLAG-As-NF- κ B1 in HEK 293T cells, and first immunoprecipitated cell extracts using anti-FLAG beads. We then performed anti-FLAG

and anti-MYC Western blotting on the immunoprecipitates to assess whether these NF- κ Bs could interact. MYC-As-NF- κ B2 and MYC-As-NF- κ B3 were both co-immunoprecipitated with FLAG-As-NF- κ B1, as well as with each other (Figure 5.9H, IP). The MYC-As-NF- κ B proteins were not seen when they were co-transfected with the empty vector control (Figure 5.9H, IP).

From these data, it appears that all three As-NF- κ Bs can enter the nucleus when expressed in vertebrate and protist cells, but they bind DNA and activate transcription to varying degrees. Furthermore, all three As-NF- κ Bs can form heterodimers with the other As-NF- κ Bs. As we discuss below, we think that the reduced ability of As-NF- κ B1 to bind DNA and activate a reporter gene is due to a limited ability to form homodimers.

5.8 Chapter 5 summary

In this chapter, we have functionally characterized and compared, for the first time, NF- κ B proteins from two protists. Taken together, these results demonstrate that although functional DNA-binding and transcriptional-activating NF- κ B proteins exist in these protists (Figure 5.10), the overall structures and regulation of these proteins varies considerably, both among protists and when compared to animal NF- κ Bs.

Table 5.1 Quantification of NF- κ B positive nuclei in DF-1 cells

Localization of the FLAG-tagged protein in DF-1 cells (NF- κ B positive/total cells counted)

Transfected Plasmid	Nucleus	Cytoplasm	Cytoplasm/Nucleus
FLAG-Nv-NF- κ B	79/79 (100%)	0	0
FLAG-Co-NF- κ B	1/702 (0.1%)	701/702 (99.9%)	0
FLAG-Co-RHD	1874/2114 (88.7%)	19/2114 (0.9%)	221/2114 (10.5%)
FLAG-Co-Cterm	3/570 (0.5%)	536/570 (94%)	31/570 (5.4%)

Table 5.2 Genes that co-expression with NF- κ B in the three life stages of *Capsaspora*, and genes that are annotated within that group

Genes expressing with NF-κB	Protein product	Length	Protein Name
CAOG_02491	XP_004349241.1	459	1-acyl-sn-glycerol-3-phosphate acyltransferase
CAOG_04763	XP_004347514.1	350	2-acylglycerol O-acyltransferase 2-A
CAOG_04187	XP_004348012.1	577	24-dehydrocholesterol reductase
CAOG_05231	XP_004346916.2	372	2OG-Fe(II) oxygenase
CAOG_02841	XP_004348654.2	470	3-oxoacyl-[acyl-carrier-protein] synthase 2, variant 2
CAOG_00541	XP_004365412.1	82	40S ribosomal protein S21
CAOG_00263	XP_004365134.1	56	40S ribosomal protein S29
CAOG_05035	XP_004346720.2	744	5-methyltetrahydrofolate-homocysteine methyltransferase reductase
CAOG_02304	XP_004349054.1	1066	5'-AMP-activated protein kinase catalytic subunit alpha-1
CAOG_05898	XP_004345488.1	513	5'-AMP-activated protein kinase catalytic subunit alpha-2
CAOG_08769	XP_011270399.1	779	6-phosphofructokinase
CAOG_02333	XP_004349083.2	945	AAA family ATPase
CAOG_05007	XP_004346692.1	1529	ABC transporter
CAOG_05822	XP_004345412.1	758	ABC transporter
CAOG_07365	XP_004343224.1	209	Abca3 protein
CAOG_04125	XP_004347950.1	322	ADP-ribosyl cyclase
CAOG_03774	XP_004363502.1	184	ADP-ribosylation factor-like protein 8B
CAOG_00796	XP_004365667.1	455	alkane-1 monooxygenase
CAOG_02472	XP_004349222.2	520	Amidophosphoribosyltransferase

CAOG_07181	XP_004343905.2	845	ankyrin repeat domain-containing protein 13
CAOG_07387	XP_004343246.1	392	aquaporin 10
CAOG_05051	XP_004346736.2	829	archipelago beta form
CAOG_05267	XP_004346952.1	436	ataxin-3
CAOG_01086	XP_004365957.1	985	AtMMH-1
CAOG_00397	XP_004365268.1	221	ATP synthase subunit delta
CAOG_04594	XP_004347341.1	658	ATP-binding cassette
CAOG_00533	XP_004365404.2	1769	ATP-binding cassette sub-family A member 1
CAOG_00237	XP_004365108.2	616	ATP-binding cassette sub-family B member 10
CAOG_03871	XP_004363599.2	734	ATP-binding cassette sub-family B member 7
CAOG_04819	XP_004347570.2	621	ATP-binding cassette sub-family E member 1
CAOG_06637	XP_004344258.2	767	ATP-binding cassette transporter sub-family G member 2c
CAOG_06442	XP_004345191.1	257	ATP-dependent Clp protease proteolytic subunit
CAOG_03162	XP_004364001.2	838	ATP-dependent DNA helicase
CAOG_02749	XP_004349499.2	1149	ATP-dependent protease La
CAOG_07773	XP_004342846.1	986	ATP-dependent protease La
CAOG_00916	XP_004365787.1	352	ATPase
CAOG_05593	XP_004346266.1	861	ATPase
CAOG_04235	XP_004348060.1	1249	ATPase type 13A
CAOG_09828			Atypical/RIO/RIO1 protein kinase, variant 4
CAOG_06184	XP_004345774.2	433	B3GNTL1 protein
CAOG_07333	XP_004343192.2	544	Beclin
CAOG_00500	XP_004365371.1	533	betaine aldehyde dehydrogenase
CAOG_00760	XP_004365631.1	338	branched-chain amino acid aminotransferase
CAOG_00481	XP_004365352.2	455	Bystin
CAOG_07134	XP_004343858.1	293	C-4 methylsterol oxidase
CAOG_09996			Ca ²⁺ transporting ATPase, plasma membrane
CAOG_04295	XP_004348120.1	256	calcium-dependent cysteine protease
CAOG_01208	XP_004349705.1	992	calcium-translocating P-type ATPase

CAOG_10057			CAMK/CAMKL/BRSK protein kinase
CAOG_09592			CAMK/CAMKL/PASK protein kinase
CAOG_02108	XP_004348858.2	1562	cAMP-dependent protein kinase catalytic subunit
CAOG_07847	XP_004342925.1	249	carbonate dehydratase
CAOG_01084	XP_004365955.2	325	carnitine/acylcarnitine carrier protein
CAOG_00701	XP_004365572.1	355	carrier protein
CAOG_03609	XP_004363337.2	352	carrier protein
CAOG_04242	XP_004348067.1	473	carrier protein
CAOG_02768	XP_004349521.1	308	casein kinase
CAOG_05925	XP_004345515.2	493	cathepsin C
CAOG_04515	XP_004347262.1	324	cathepsin L2
CAOG_07286	XP_004343145.1	313	ccr4-associated factor
CAOG_02644	XP_004349394.1	300	cell motility mediator
CAOG_03528	XP_004348433.1	1194	ceramide kinase
CAOG_02756	XP_004349506.2	548	cereblon
CAOG_08208	XP_004342463.2	583	cholesterol acyltransferase
CAOG_00569	XP_004365440.2	574	cholinesterase
CAOG_02652	XP_004349402.1	365	chorismate mutase/prephenate dehydratase
CAOG_01398	XP_004349918.1	205	chromatin modifying protein 1b
CAOG_07361	XP_004343220.1	213	chromatin modifying protein 1b
CAOG_04220	XP_004348045.2	223	chromatin modifying protein 2a
CAOG_03362	XP_004364201.2	394	CHY zinc finger family protein
CAOG_07281	XP_004343140.2	477	COBW domain-containing protein
CAOG_00712	XP_004365583.1	591	collagen type IV alpha-3-binding protein
CAOG_04276	XP_004348101.2	318	copine-9
CAOG_04972	XP_004346657.1	1180	copper-transporting ATPase
CAOG_01314	XP_004349834.2	480	coproporphyrinogen oxidase
CAOG_06354	XP_004345103.1	471	cre
CAOG_00172	XP_004365043.2	497	crk-like protein
CAOG_01275	XP_004349795.2	837	Cullin 4
CAOG_01396	XP_004349916.2	416	cyclin-box carrying protein isoform
CAOG_04469	XP_004348297.1	450	cystathionine beta-lyase
CAOG_03984	XP_004347809.1	586	cystathionine gamma-synthase

CAOG_06008	XP_004345598.2	488	cysteine desulfurylase
CAOG_06337	XP_004345086.1	805	cytochrome b5 reductase 4
CAOG_05595	XP_004346268.1	547	cytochrome P450
CAOG_05678	XP_004346351.1	484	cytochrome P450
CAOG_04035	XP_004347860.1	314	cytosolic Fe-S cluster assembling factor nbp35
CAOG_05420	XP_004346093.2	588	Dcun1d3 protein
CAOG_04252	XP_004348077.1	793	DDHD domain containing 2
CAOG_06774	XP_004344395.2	598	DEAD box ATP-dependent RNA helicase
CAOG_02709	XP_004349459.2	1324	DEAH helicase isoform 6
CAOG_00393	XP_004365264.1	207	decarboxylase
CAOG_08183	XP_004342784.1	336	delta-aminolevulinic acid dehydratase
CAOG_03778	XP_004363506.1	340	deoxyhypusine synthase
CAOG_04524	XP_004347271.1	634	deoxyribodipyrimidine photo-lyase
CAOG_00128	XP_004364999.2	294	derlin-2 like protein
CAOG_06433	XP_004345182.1	334	DHHC zinc finger domain-containing protein
CAOG_05176	XP_004346861.1	396	diacylglycerol acyltransferase type 2A
CAOG_07051	XP_004343775.2	1542	diacylglycerol kinase 1
CAOG_01121	XP_004365992.2	363	diaphorase
CAOG_08167	XP_004342768.2	411	dihydroderamide delta-4 desaturase
CAOG_05518	XP_004346191.2	395	diphosphomevalonate decarboxylase
CAOG_02451	XP_004349201.1	515	diphthamide biosynthesis protein
CAOG_07910	XP_004342995.2	116	DNA-directed RNA polymerase II polypeptide
CAOG_03769	XP_004363497.1	838	dolichyl-phosphate-mannose-protein mannosyltransferase
CAOG_05442	XP_004346115.2	398	DREV methyltransferase
CAOG_00863	XP_004365734.1	633	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1B
CAOG_04903	XP_004347654.1	2072	E1A binding protein p300
CAOG_09787			E3 ubiquitin-protein ligase NEDD4
CAOG_03032	XP_004363871.1	797	efflux ABC transporter
CAOG_08089	XP_004342690.1	1739	ephrin type-A receptor 4a
CAOG_03281	XP_004364120.1	542	eukaryotic peptide chain release factor subunit 1

CAOG_03741	XP_004363469.2	2217	eukaryotic translation initiation factor 2-alpha kinase
CAOG_03792	XP_004363520.2	1048	eukaryotic translation initiation factor 2-alpha kinase 1
CAOG_07494	XP_004343368.1	601	F-box/WD repeat-containing protein pof1
CAOG_04544	XP_004347291.1	360	FAM72A protein
CAOG_05028	XP_004346713.1	418	farnesyl-diphosphate farnesyltransferase 1
CAOG_03987	XP_004347812.1	372	fatty acid elongase
CAOG_01458	XP_004364326.2	268	ferredoxin 1-like protein
CAOG_05797	XP_004345387.1	251	ferredoxin 1-like protein
CAOG_05336	XP_004347021.1	587	ferrous ion membrane transporter DMT1
CAOG_03584	XP_004363312.2	4679	G protein-coupled receptor
CAOG_00499	XP_004365370.1	476	gamma-butyrobetaine dioxygenase
CAOG_05023	XP_004346708.1	259	gamma-butyrobetaine dioxygenase
CAOG_05409	XP_004346082.1	188	general stress protein
CAOG_02947	XP_004363786.1	314	geranylgeranyl pyrophosphate synthetase
CAOG_07422	XP_004343281.1	186	glucosamine-6-phosphate N-acetyltransferase
CAOG_05976	XP_004345566.1	316	glutamate receptor Gr2
CAOG_03218	XP_004364057.2	231	glutaredoxin-3
CAOG_07508	XP_004343382.2	398	glutathione S-transferase
CAOG_02422	XP_004349172.2	589	glycerol-3-phosphate acyltransferase
CAOG_02549	XP_004349299.2	472	glycine hydroxymethyltransferase
CAOG_00165	XP_004365036.2	823	glycosyltransferase-like protein LARGE1
CAOG_01947	XP_004364815.2	685	gphn protein
CAOG_05140	XP_004346825.2	933	growth hormone-regulated TBC protein 1
CAOG_07977	XP_004342578.2	588	growth-inhibiting protein 1
CAOG_03775	XP_004363503.1	640	GTP binding protein 4
CAOG_02329	XP_004349079.1	205	GTP-binding protein yptV3
CAOG_05271	XP_004346956.1	191	GTPase
CAOG_04766	XP_004347517.2	2032	guanine nucleotide-exchange protein
CAOG_00281	XP_004365152.1	568	heat shock factor 2
CAOG_07347	XP_004343206.1	155	HIT domain-containing protein

CAOG_04678	XP_004347425.1	578	HLA-B associated transcript 5
CAOG_04537	XP_004347284.1	573	hsp70-like protein
CAOG_01499	XP_004364367.1	523	hydroxymethylglutaryl-CoA synthase 1
CAOG_05029	XP_004346714.1	1125	hydroxymethylglutaryl-coenzyme A reductase
CAOG_04522	XP_004347269.1	382	hydroxysteroid dehydrogenase 7
CAOG_06367	pseudo gene		hypothetical protein
CAOG_09260			hypothetical protein
CAOG_09288			hypothetical protein
CAOG_09291			hypothetical protein
CAOG_09295			hypothetical protein
CAOG_09302			hypothetical protein
CAOG_09303			hypothetical protein
CAOG_09304			hypothetical protein
CAOG_09305			hypothetical protein
CAOG_09317			hypothetical protein
CAOG_09318			hypothetical protein
CAOG_09321			hypothetical protein
CAOG_09322			hypothetical protein
CAOG_09324			hypothetical protein
CAOG_09335			hypothetical protein
CAOG_09336			hypothetical protein
CAOG_09342			hypothetical protein
CAOG_09343			hypothetical protein
CAOG_09359			hypothetical protein
CAOG_09360			hypothetical protein
CAOG_09361			hypothetical protein
CAOG_09364			hypothetical protein
CAOG_09366			hypothetical protein
CAOG_09367			hypothetical protein
CAOG_09372			hypothetical protein
CAOG_09379			hypothetical protein
CAOG_09390			hypothetical protein
CAOG_09396			hypothetical protein

CAOG_09398			hypothetical protein
CAOG_09399			hypothetical protein
CAOG_09400			hypothetical protein
CAOG_09410			hypothetical protein
CAOG_09412			hypothetical protein
CAOG_09421			hypothetical protein
CAOG_09426			hypothetical protein
CAOG_09428			hypothetical protein
CAOG_09459			hypothetical protein
CAOG_09460			hypothetical protein
CAOG_09463			hypothetical protein
CAOG_09478			hypothetical protein
CAOG_09490			hypothetical protein
CAOG_09521			hypothetical protein
CAOG_09531			hypothetical protein
CAOG_09532			hypothetical protein
CAOG_09534			hypothetical protein
CAOG_09548			hypothetical protein
CAOG_09561			hypothetical protein
CAOG_09563			hypothetical protein
CAOG_09574			hypothetical protein
CAOG_09580			hypothetical protein
CAOG_09581			hypothetical protein
CAOG_09583			hypothetical protein
CAOG_09584			hypothetical protein
CAOG_09594			hypothetical protein
CAOG_09595			hypothetical protein
CAOG_09602			hypothetical protein
CAOG_09603			hypothetical protein
CAOG_09623			hypothetical protein
CAOG_09624			hypothetical protein
CAOG_09629			hypothetical protein
CAOG_09634			hypothetical protein
CAOG_09635			hypothetical protein

CAOG_09674			hypothetical protein
CAOG_09689			hypothetical protein
CAOG_09690			hypothetical protein
CAOG_09692			hypothetical protein
CAOG_09697			hypothetical protein
CAOG_09698			hypothetical protein
CAOG_09699			hypothetical protein
CAOG_09702			hypothetical protein
CAOG_09703			hypothetical protein
CAOG_09709			hypothetical protein
CAOG_09711			hypothetical protein
CAOG_09712			hypothetical protein
CAOG_09720			hypothetical protein
CAOG_09725			hypothetical protein
CAOG_09738			hypothetical protein
CAOG_09739			hypothetical protein
CAOG_09740			hypothetical protein
CAOG_09749			hypothetical protein
CAOG_09750			hypothetical protein
CAOG_09759			hypothetical protein
CAOG_09775			hypothetical protein
CAOG_09777			hypothetical protein
CAOG_09778			hypothetical protein
CAOG_09786			hypothetical protein
CAOG_09790			hypothetical protein
CAOG_09801			hypothetical protein
CAOG_09802			hypothetical protein
CAOG_09808			hypothetical protein
CAOG_09814			hypothetical protein
CAOG_09818			hypothetical protein
CAOG_09841			hypothetical protein
CAOG_09845			hypothetical protein
CAOG_09846			hypothetical protein
CAOG_09863			hypothetical protein

CAOG_09864			hypothetical protein
CAOG_09865			hypothetical protein
CAOG_09868			hypothetical protein
CAOG_09869			hypothetical protein
CAOG_09872			hypothetical protein
CAOG_09878			hypothetical protein
CAOG_09887			hypothetical protein
CAOG_09888			hypothetical protein
CAOG_09895			hypothetical protein
CAOG_09899			hypothetical protein
CAOG_09902			hypothetical protein
CAOG_09911			hypothetical protein
CAOG_09915			hypothetical protein
CAOG_09916			hypothetical protein
CAOG_09926			hypothetical protein
CAOG_09927			hypothetical protein
CAOG_09935			hypothetical protein
CAOG_09939			hypothetical protein
CAOG_09942			hypothetical protein
CAOG_09943			hypothetical protein
CAOG_09963			hypothetical protein
CAOG_09964			hypothetical protein
CAOG_09967			hypothetical protein
CAOG_09969			hypothetical protein
CAOG_09976			hypothetical protein
CAOG_09977			hypothetical protein
CAOG_09990			hypothetical protein
CAOG_10000			hypothetical protein
CAOG_10013			hypothetical protein
CAOG_10014			hypothetical protein
CAOG_10016			hypothetical protein
CAOG_10017			hypothetical protein
CAOG_10024			hypothetical protein
CAOG_10045			hypothetical protein

CAOG_10049			hypothetical protein
CAOG_10051			hypothetical protein
CAOG_10054			hypothetical protein
CAOG_10056			hypothetical protein
CAOG_10062			hypothetical protein
CAOG_10074			hypothetical protein
CAOG_10075			hypothetical protein
CAOG_10078			hypothetical protein
CAOG_10084			hypothetical protein
CAOG_10085			hypothetical protein
CAOG_10094			hypothetical protein
CAOG_10106			hypothetical protein
CAOG_10108			hypothetical protein
CAOG_10109			hypothetical protein
CAOG_10116			hypothetical protein
CAOG_10122			hypothetical protein
CAOG_10133			hypothetical protein
CAOG_10145			hypothetical protein
CAOG_10159			hypothetical protein
CAOG_10160			hypothetical protein
CAOG_10168			hypothetical protein
CAOG_10174			hypothetical protein
CAOG_10240			hypothetical protein
CAOG_00002	XP_004364873.1	1261	hypothetical protein CAOG_00002
CAOG_00003	XP_004364874.1	182	hypothetical protein CAOG_00003
CAOG_00024	XP_004364895.2	2264	hypothetical protein CAOG_00024
CAOG_00041	XP_004364912.2	485	hypothetical protein CAOG_00041
CAOG_00061	XP_004364932.2	369	hypothetical protein CAOG_00061
CAOG_00065	XP_004364936.2	1582	hypothetical protein CAOG_00065
CAOG_00066	XP_004364937.2	662	hypothetical protein CAOG_00066
CAOG_00072	XP_004364943.1	623	hypothetical protein CAOG_00072
CAOG_00074	XP_004364945.2	1419	hypothetical protein CAOG_00074
CAOG_00100	XP_004364971.1	208	hypothetical protein CAOG_00100
CAOG_00103	XP_004364974.1	884	hypothetical protein CAOG_00103
CAOG_00110	XP_004364981.2	1361	hypothetical protein CAOG_00110

CAOG_00111	XP_004364982.2	1032	hypothetical protein CAOG_00111
CAOG_00116	XP_004364987.1	566	hypothetical protein CAOG_00116
CAOG_00118	XP_004364989.1	353	hypothetical protein CAOG_00118
CAOG_00126	XP_004364997.2	399	hypothetical protein CAOG_00126
CAOG_00127	XP_004364998.1	72	hypothetical protein CAOG_00127
CAOG_00136	XP_004365007.2	96	hypothetical protein CAOG_00136
CAOG_00138	XP_004365009.1	460	hypothetical protein CAOG_00138
CAOG_00183	XP_004365054.1	602	hypothetical protein CAOG_00183
CAOG_00195	XP_004365066.1	942	hypothetical protein CAOG_00195
CAOG_00207	XP_004365078.1	359	hypothetical protein CAOG_00207
CAOG_00209	XP_004365080.1	683	hypothetical protein CAOG_00209
CAOG_00213	XP_004365084.1	1022	hypothetical protein CAOG_00213
CAOG_00224	XP_004365095.1	147	hypothetical protein CAOG_00224
CAOG_00228	XP_004365099.1	1263	hypothetical protein CAOG_00228
CAOG_00230	XP_004365101.1	655	hypothetical protein CAOG_00230
CAOG_00236	XP_004365107.1	528	hypothetical protein CAOG_00236
CAOG_00242	XP_004365113.1	66	hypothetical protein CAOG_00242
CAOG_00275	XP_004365146.2	759	hypothetical protein CAOG_00275
CAOG_00280	XP_004365151.1	305	hypothetical protein CAOG_00280
CAOG_00295	XP_004365166.2	478	hypothetical protein CAOG_00295
CAOG_00298	XP_004365169.2	970	hypothetical protein CAOG_00298
CAOG_00303	XP_004365174.2	854	hypothetical protein CAOG_00303
CAOG_00333	XP_004365204.2	1075	hypothetical protein CAOG_00333
CAOG_00340	XP_004365211.1	435	hypothetical protein CAOG_00340
CAOG_00345	XP_004365216.1	842	hypothetical protein CAOG_00345
CAOG_00362	XP_004365233.1	155	hypothetical protein CAOG_00362
CAOG_00387	XP_004365258.1	573	hypothetical protein CAOG_00387
CAOG_00395	XP_004365266.2	714	hypothetical protein CAOG_00395
CAOG_00412	XP_004365283.2	715	hypothetical protein CAOG_00412
CAOG_00466	XP_004365337.1	331	hypothetical protein CAOG_00466
CAOG_00502	XP_004365373.1	366	hypothetical protein CAOG_00502
CAOG_00504	XP_004365375.1	55	hypothetical protein CAOG_00504
CAOG_00510	XP_004365381.1	138	hypothetical protein CAOG_00510
CAOG_00526	XP_004365397.1	176	hypothetical protein CAOG_00526
CAOG_00528	XP_004365399.2	1347	hypothetical protein CAOG_00528
CAOG_00537	XP_004365408.1	341	hypothetical protein CAOG_00537
CAOG_00538	XP_004365409.2	343	hypothetical protein CAOG_00538

CAOG_00540	XP_004365411.1	347	hypothetical protein CAOG_00540
CAOG_00542	XP_004365413.1	330	hypothetical protein CAOG_00542
CAOG_00548	XP_004365419.2	730	hypothetical protein CAOG_00548
CAOG_00562	XP_004365433.1	534	hypothetical protein CAOG_00562
CAOG_00574	XP_004365445.1	644	hypothetical protein CAOG_00574
CAOG_00581	XP_004365452.1	195	hypothetical protein CAOG_00581
CAOG_00585	XP_004365456.1	555	hypothetical protein CAOG_00585
CAOG_00623	XP_004365494.2	126	hypothetical protein CAOG_00623
CAOG_00651	XP_004365522.1	104	hypothetical protein CAOG_00651
CAOG_00652	XP_004365523.2	908	hypothetical protein CAOG_00652
CAOG_00710	XP_004365581.2	552	hypothetical protein CAOG_00710
CAOG_00715	XP_004365586.2	619	hypothetical protein CAOG_00715
CAOG_00737	XP_004365608.2	870	hypothetical protein CAOG_00737
CAOG_00752	XP_004365623.1	442	hypothetical protein CAOG_00752
CAOG_00758	XP_004365629.1	2958	hypothetical protein CAOG_00758
CAOG_00780	XP_004365651.2	2048	hypothetical protein CAOG_00780
CAOG_00781	XP_004365652.1	442	hypothetical protein CAOG_00781
CAOG_00794	XP_004365665.2	453	hypothetical protein CAOG_00794
CAOG_00799	XP_004365670.2	408	hypothetical protein CAOG_00799
CAOG_00805	XP_004365676.1	878	hypothetical protein CAOG_00805
CAOG_00833	XP_004365704.1	206	hypothetical protein CAOG_00833
CAOG_00839	XP_004365710.1	311	hypothetical protein CAOG_00839
CAOG_00842	XP_004365713.1	132	hypothetical protein CAOG_00842
CAOG_00851	XP_004365722.1	263	hypothetical protein CAOG_00851
CAOG_00866	XP_004365737.1	566	hypothetical protein CAOG_00866
CAOG_00879	XP_004365750.1	136	hypothetical protein CAOG_00879
CAOG_00883	XP_004365754.1	443	hypothetical protein CAOG_00883
CAOG_00896	XP_004365767.2	593	hypothetical protein CAOG_00896
CAOG_00912	XP_004365783.2	1721	hypothetical protein CAOG_00912
CAOG_00924	XP_004365795.1	603	hypothetical protein CAOG_00924
CAOG_00933	XP_004365804.2	386	hypothetical protein CAOG_00933
CAOG_00946	XP_004365817.2	653	hypothetical protein CAOG_00946
CAOG_00951	XP_004365822.1	458	hypothetical protein CAOG_00951
CAOG_00967	XP_004365838.2	1184	hypothetical protein CAOG_00967
CAOG_00973	XP_004365844.1	378	hypothetical protein CAOG_00973
CAOG_00979	XP_004365850.2	169	hypothetical protein CAOG_00979
CAOG_00982	XP_004365853.2	668	hypothetical protein CAOG_00982

CAOG_01001	XP_004365872.1	519	hypothetical protein CAOG_01001
CAOG_01017	XP_004365888.2	859	hypothetical protein CAOG_01017
CAOG_01030	XP_004365901.2	473	hypothetical protein CAOG_01030
CAOG_01033	XP_004365904.1	222	hypothetical protein CAOG_01033
CAOG_01038	XP_004365909.1	1045	hypothetical protein CAOG_01038
CAOG_01070	XP_004365941.1	1606	hypothetical protein CAOG_01070
CAOG_01083	XP_004365954.1	697	hypothetical protein CAOG_01083
CAOG_01127	XP_004365998.2	245	hypothetical protein CAOG_01127
CAOG_01129	XP_004366000.2	1413	hypothetical protein CAOG_01129
CAOG_01136	XP_004366007.1	720	hypothetical protein CAOG_01136
CAOG_01150	XP_004366021.2	590	hypothetical protein CAOG_01150
CAOG_01167	XP_004366038.1	565	hypothetical protein CAOG_01167
CAOG_01175	XP_004366046.1	217	hypothetical protein CAOG_01175
CAOG_01191	XP_004349688.2	144	hypothetical protein CAOG_01191
CAOG_01250	XP_004349770.2	221	hypothetical protein CAOG_01250
CAOG_01256	XP_004349776.2	429	hypothetical protein CAOG_01256
CAOG_01257	XP_004349777.2	706	hypothetical protein CAOG_01257
CAOG_01276	XP_004349796.2	650	hypothetical protein CAOG_01276
CAOG_01283	XP_004349803.2	1064	hypothetical protein CAOG_01283
CAOG_01284	XP_004349804.2	1677	hypothetical protein CAOG_01284
CAOG_01286	XP_004349806.1	397	hypothetical protein CAOG_01286
CAOG_01310	XP_004349830.1	82	hypothetical protein CAOG_01310
CAOG_01315	XP_004349835.2	435	hypothetical protein CAOG_01315
CAOG_01328	XP_004349848.2	339	hypothetical protein CAOG_01328
CAOG_01351	XP_004349871.2	531	hypothetical protein CAOG_01351
CAOG_01353	XP_004349873.2	751	hypothetical protein CAOG_01353
CAOG_01365	XP_004349885.1	531	hypothetical protein CAOG_01365
CAOG_01379	XP_004349899.2	231	hypothetical protein CAOG_01379
CAOG_01425	XP_004349945.1	968	hypothetical protein CAOG_01425
CAOG_01434	XP_004349954.1	138	hypothetical protein CAOG_01434
CAOG_01450	XP_004364318.2	465	hypothetical protein CAOG_01450
CAOG_01456	XP_004364324.1	100	hypothetical protein CAOG_01456
CAOG_01476	XP_004364344.1	229	hypothetical protein CAOG_01476
CAOG_01483	XP_004364351.1	219	hypothetical protein CAOG_01483
CAOG_01487	XP_004364355.2	1613	hypothetical protein CAOG_01487
CAOG_01497	XP_004364365.1	261	hypothetical protein CAOG_01497
CAOG_01500	XP_004364368.2	132	hypothetical protein CAOG_01500

CAOG_01515	XP_004364383.1	531	hypothetical protein CAOG_01515
CAOG_01542	XP_004364410.2	906	hypothetical protein CAOG_01542
CAOG_01543	XP_004364411.1	410	hypothetical protein CAOG_01543
CAOG_01615	XP_004364483.2	313	hypothetical protein CAOG_01615
CAOG_01616	XP_004364484.1	119	hypothetical protein CAOG_01616
CAOG_01617	XP_004364485.1	236	hypothetical protein CAOG_01617
CAOG_01624	XP_004364492.1	702	hypothetical protein CAOG_01624
CAOG_01629	XP_004364497.1	519	hypothetical protein CAOG_01629
CAOG_01631	XP_004364499.2	603	hypothetical protein CAOG_01631
CAOG_01637	XP_004364505.1	585	hypothetical protein CAOG_01637
CAOG_01653	XP_004364521.2	1263	hypothetical protein CAOG_01653
CAOG_01656	XP_004364524.1	347	hypothetical protein CAOG_01656
CAOG_01657	XP_004364525.1	561	hypothetical protein CAOG_01657
CAOG_01710	XP_004364578.1	556	hypothetical protein CAOG_01710
CAOG_01713	XP_004364581.2	915	hypothetical protein CAOG_01713
CAOG_01718	XP_004364586.2	783	hypothetical protein CAOG_01718
CAOG_01728	XP_004364596.1	406	hypothetical protein CAOG_01728
CAOG_01736	XP_004364604.2	404	hypothetical protein CAOG_01736
CAOG_01741	XP_004364609.1	173	hypothetical protein CAOG_01741
CAOG_01767	XP_004364635.1	394	hypothetical protein CAOG_01767
CAOG_01771	XP_004364639.1	862	hypothetical protein CAOG_01771
CAOG_01795	XP_004364663.1	214	hypothetical protein CAOG_01795
CAOG_01801	XP_004364669.1	403	hypothetical protein CAOG_01801
CAOG_01821	XP_004364689.1	445	hypothetical protein CAOG_01821
CAOG_01823	XP_004364691.1	166	hypothetical protein CAOG_01823
CAOG_01852	XP_004364720.1	219	hypothetical protein CAOG_01852
CAOG_01853	XP_004364721.1	1092	hypothetical protein CAOG_01853
CAOG_01871	XP_004364739.1	229	hypothetical protein CAOG_01871
CAOG_01882	XP_004364750.1	231	hypothetical protein CAOG_01882
CAOG_01886	XP_004364754.1	304	hypothetical protein CAOG_01886
CAOG_01893	XP_004364761.2	258	hypothetical protein CAOG_01893
CAOG_01900	XP_004364768.1	117	hypothetical protein CAOG_01900
CAOG_01902	XP_004364770.2	200	hypothetical protein CAOG_01902
CAOG_01906	XP_004364774.2	464	hypothetical protein CAOG_01906
CAOG_01922	XP_004364790.2	607	hypothetical protein CAOG_01922
CAOG_01923	XP_004364791.1	141	hypothetical protein CAOG_01923
CAOG_01924	XP_004364792.1	85	hypothetical protein CAOG_01924

CAOG_01941	XP_004364809.1	3082	hypothetical protein CAOG_01941
CAOG_01958	XP_004364826.2	143	hypothetical protein CAOG_01958
CAOG_01971	XP_004364839.1	387	hypothetical protein CAOG_01971
CAOG_01977	XP_004364845.2	855	hypothetical protein CAOG_01977
CAOG_01986	XP_004364854.2	1068	hypothetical protein CAOG_01986
CAOG_02015	XP_004348765.1	829	hypothetical protein CAOG_02015
CAOG_02019	XP_004348769.1	484	hypothetical protein CAOG_02019
CAOG_02032	XP_004348782.1	346	hypothetical protein CAOG_02032
CAOG_02054	XP_004348804.1	664	hypothetical protein CAOG_02054
CAOG_02074	XP_004348824.1	592	hypothetical protein CAOG_02074
CAOG_02077	XP_004348827.1	197	hypothetical protein CAOG_02077
CAOG_02087	XP_004348837.1	154	hypothetical protein CAOG_02087
CAOG_02088	XP_004348838.1	127	hypothetical protein CAOG_02088
CAOG_02122	XP_004348872.1	737	hypothetical protein CAOG_02122
CAOG_02148	XP_004348898.2	593	hypothetical protein CAOG_02148
CAOG_02156	XP_004348906.2	681	hypothetical protein CAOG_02156
CAOG_02158	XP_004348908.1	930	hypothetical protein CAOG_02158
CAOG_02167	XP_004348917.1	1138	hypothetical protein CAOG_02167
CAOG_02187	XP_004348937.2	300	hypothetical protein CAOG_02187
CAOG_02190	XP_004348940.2	3356	hypothetical protein CAOG_02190
CAOG_02192	XP_004348942.1	215	hypothetical protein CAOG_02192
CAOG_02193	XP_004348943.1	2772	hypothetical protein CAOG_02193
CAOG_02202	XP_004348952.1	385	hypothetical protein CAOG_02202
CAOG_02206	XP_004348956.1	649	hypothetical protein CAOG_02206
CAOG_02209	XP_004348959.1	248	hypothetical protein CAOG_02209
CAOG_02210	XP_004348960.1	71	hypothetical protein CAOG_02210
CAOG_02216	XP_004348966.1	447	hypothetical protein CAOG_02216
CAOG_02227	XP_004348977.2	3131	hypothetical protein CAOG_02227
CAOG_02233	XP_004348983.2	838	hypothetical protein CAOG_02233
CAOG_02237	XP_004348987.1	1632	hypothetical protein CAOG_02237
CAOG_02247	XP_004348997.1	880	hypothetical protein CAOG_02247
CAOG_02257	XP_004349007.1	1314	hypothetical protein CAOG_02257
CAOG_02265	XP_004349015.2	69	hypothetical protein CAOG_02265
CAOG_02275	XP_004349025.1	2013	hypothetical protein CAOG_02275
CAOG_02276	XP_004349026.1	116	hypothetical protein CAOG_02276
CAOG_02281	XP_004349031.2	741	hypothetical protein CAOG_02281
CAOG_02285	XP_004349035.2	98	hypothetical protein CAOG_02285

CAOG_02286	XP_004349036.1	69	hypothetical protein CAOG_02286
CAOG_02288	XP_004349038.2	679	hypothetical protein CAOG_02288
CAOG_02290	XP_004349040.1	192	hypothetical protein CAOG_02290
CAOG_02303	XP_004349053.2	807	hypothetical protein CAOG_02303
CAOG_02345	XP_004349095.1	428	hypothetical protein CAOG_02345
CAOG_02355	XP_004349105.2	377	hypothetical protein CAOG_02355
CAOG_02363	XP_004349113.2	561	hypothetical protein CAOG_02363
CAOG_02369	XP_004349119.1	720	hypothetical protein CAOG_02369
CAOG_02372	XP_004349122.1	1153	hypothetical protein CAOG_02372
CAOG_02382	XP_004349132.2	626	hypothetical protein CAOG_02382
CAOG_02423	XP_004349173.2	366	hypothetical protein CAOG_02423
CAOG_02441	XP_004349191.1	221	hypothetical protein CAOG_02441
CAOG_02442	XP_004349192.1	106	hypothetical protein CAOG_02442
CAOG_02449	XP_004349199.1	445	hypothetical protein CAOG_02449
CAOG_02503	XP_004349253.1	443	hypothetical protein CAOG_02503
CAOG_02527	XP_004349277.2	1328	hypothetical protein CAOG_02527
CAOG_02546	XP_004349296.1	557	hypothetical protein CAOG_02546
CAOG_02568	XP_004349318.2	299	hypothetical protein CAOG_02568
CAOG_02613	XP_004349363.1	688	hypothetical protein CAOG_02613
CAOG_02645	XP_004349395.1	101	hypothetical protein CAOG_02645
CAOG_02665	XP_004349415.1	288	hypothetical protein CAOG_02665
CAOG_02668	XP_004349418.1	696	hypothetical protein CAOG_02668
CAOG_02670	XP_004349420.2	213	hypothetical protein CAOG_02670
CAOG_02683	XP_004349433.2	151	hypothetical protein CAOG_02683
CAOG_02691	XP_004349441.2	1812	hypothetical protein CAOG_02691
CAOG_02715	XP_004349465.1	126	hypothetical protein CAOG_02715
CAOG_02729	XP_004349479.1	106	hypothetical protein CAOG_02729
CAOG_02735	XP_004349485.1	826	hypothetical protein CAOG_02735
CAOG_02741	XP_004349491.1	2316	hypothetical protein CAOG_02741
CAOG_02744	XP_004349494.2	332	hypothetical protein CAOG_02744
CAOG_02761	XP_004349512.1	376	hypothetical protein CAOG_02761
CAOG_02778	XP_004349531.2	618	hypothetical protein CAOG_02778
CAOG_02788	XP_004349541.1	488	hypothetical protein CAOG_02788
CAOG_02794	XP_004349547.2	219	hypothetical protein CAOG_02794
CAOG_02800	XP_004348613.1	727	hypothetical protein CAOG_02800
CAOG_02818	XP_004348631.1	175	hypothetical protein CAOG_02818
CAOG_02854	XP_004363693.2	449	hypothetical protein CAOG_02854

CAOG_02857	XP_004363696.1	968	hypothetical protein CAOG_02857
CAOG_02864	XP_004363703.2	900	hypothetical protein CAOG_02864
CAOG_02891	XP_004363730.1	231	hypothetical protein CAOG_02891
CAOG_02900	XP_004363739.2	1050	hypothetical protein CAOG_02900
CAOG_02908	XP_004363747.2	476	hypothetical protein CAOG_02908
CAOG_02919	XP_004363758.1	1901	hypothetical protein CAOG_02919
CAOG_02926	XP_004363765.1	290	hypothetical protein CAOG_02926
CAOG_02927	XP_004363766.1	304	hypothetical protein CAOG_02927
CAOG_02934	XP_004363773.1	162	hypothetical protein CAOG_02934
CAOG_02935	XP_004363774.1	1392	hypothetical protein CAOG_02935
CAOG_02943	XP_004363782.2	361	hypothetical protein CAOG_02943
CAOG_02951	XP_004363790.2	1004	hypothetical protein CAOG_02951
CAOG_02960	XP_004363799.2	609	hypothetical protein CAOG_02960
CAOG_02969	XP_004363808.2	696	hypothetical protein CAOG_02969
CAOG_02970	XP_004363809.2	234	hypothetical protein CAOG_02970
CAOG_02973	XP_004363812.1	696	hypothetical protein CAOG_02973
CAOG_02980	XP_004363819.1	98	hypothetical protein CAOG_02980
CAOG_03031	XP_004363870.2	1526	hypothetical protein CAOG_03031
CAOG_03050	XP_004363889.1	412	hypothetical protein CAOG_03050
CAOG_03051	XP_004363890.2	646	hypothetical protein CAOG_03051
CAOG_03053	XP_004363892.1	522	hypothetical protein CAOG_03053
CAOG_03057	XP_004363896.1	386	hypothetical protein CAOG_03057
CAOG_03072	XP_004363911.1	3668	hypothetical protein CAOG_03072
CAOG_03081	XP_004363920.1	1398	hypothetical protein CAOG_03081
CAOG_03096	XP_004363935.1	500	hypothetical protein CAOG_03096
CAOG_03101	XP_004363940.1	300	hypothetical protein CAOG_03101
CAOG_03124	XP_004363963.2	209	hypothetical protein CAOG_03124
CAOG_03125	XP_004363964.2	314	hypothetical protein CAOG_03125
CAOG_03129	XP_004363968.1	175	hypothetical protein CAOG_03129
CAOG_03135	XP_004363974.1	834	hypothetical protein CAOG_03135
CAOG_03144	XP_004363983.1	176	hypothetical protein CAOG_03144
CAOG_03166	XP_004364005.1	602	hypothetical protein CAOG_03166
CAOG_03168	XP_004364007.1	431	hypothetical protein CAOG_03168
CAOG_03186	XP_004364025.2	815	hypothetical protein CAOG_03186
CAOG_03190	XP_004364029.1	1180	hypothetical protein CAOG_03190
CAOG_03206	XP_004364045.2	928	hypothetical protein CAOG_03206
CAOG_03242	XP_004364081.2	254	hypothetical protein CAOG_03242

CAOG_03257	XP_004364096.2	486	hypothetical protein CAOG_03257
CAOG_03266	XP_004364105.1	330	hypothetical protein CAOG_03266
CAOG_03304	XP_004364143.2	1655	hypothetical protein CAOG_03304
CAOG_03306	XP_004364145.2	412	hypothetical protein CAOG_03306
CAOG_03307	XP_004364146.2	232	hypothetical protein CAOG_03307
CAOG_03311	XP_004364150.2	361	hypothetical protein CAOG_03311
CAOG_03315	XP_004364154.2	872	hypothetical protein CAOG_03315
CAOG_03335	XP_004364174.2	659	hypothetical protein CAOG_03335
CAOG_03345	XP_004364184.1	69	hypothetical protein CAOG_03345
CAOG_03349	XP_004364188.1	85	hypothetical protein CAOG_03349
CAOG_03392	XP_004364231.1	364	hypothetical protein CAOG_03392
CAOG_03420	XP_004364259.2	495	hypothetical protein CAOG_03420
CAOG_03422	XP_004364261.2	810	hypothetical protein CAOG_03422
CAOG_03434	XP_004364273.2	961	hypothetical protein CAOG_03434
CAOG_03448	XP_004364287.2	231	hypothetical protein CAOG_03448
CAOG_03463	XP_004364302.2	832	hypothetical protein CAOG_03463
CAOG_03467	XP_004348372.2	805	hypothetical protein CAOG_03467
CAOG_03475	XP_004348380.1	334	hypothetical protein CAOG_03475
CAOG_03503	XP_004348408.1	395	hypothetical protein CAOG_03503
CAOG_03525	XP_004348430.2	1529	hypothetical protein CAOG_03525
CAOG_03535	XP_004348440.1	112	hypothetical protein CAOG_03535
CAOG_03548	XP_004348459.1	650	hypothetical protein CAOG_03548
CAOG_03558	XP_004363286.1	929	hypothetical protein CAOG_03558
CAOG_03581	XP_004363309.1	170	hypothetical protein CAOG_03581
CAOG_03592	XP_004363320.1	440	hypothetical protein CAOG_03592
CAOG_03611	XP_004363339.1	1750	hypothetical protein CAOG_03611
CAOG_03618	XP_004363346.2	236	hypothetical protein CAOG_03618
CAOG_03652	XP_004363380.1	961	hypothetical protein CAOG_03652
CAOG_03659	XP_004363387.1	377	hypothetical protein CAOG_03659
CAOG_03677	XP_004363405.1	384	hypothetical protein CAOG_03677
CAOG_03686	XP_004363414.2	191	hypothetical protein CAOG_03686
CAOG_03695	XP_004363423.1	6039	hypothetical protein CAOG_03695
CAOG_03696	XP_004363424.1	264	hypothetical protein CAOG_03696
CAOG_03707	XP_004363435.2	359	hypothetical protein CAOG_03707
CAOG_03726	XP_004363454.1	433	hypothetical protein CAOG_03726
CAOG_03761	XP_004363489.1	585	hypothetical protein CAOG_03761
CAOG_03767	XP_004363495.2	663	hypothetical protein CAOG_03767

CAOG_03781	XP_004363509.2	522	hypothetical protein CAOG_03781
CAOG_03783	XP_004363511.1	1055	hypothetical protein CAOG_03783
CAOG_03809	XP_004363537.1	539	hypothetical protein CAOG_03809
CAOG_03823	XP_004363551.1	271	hypothetical protein CAOG_03823
CAOG_03833	XP_004363561.2	1057	hypothetical protein CAOG_03833
CAOG_03837	XP_004363565.1	795	hypothetical protein CAOG_03837
CAOG_03869	XP_004363597.2	249	hypothetical protein CAOG_03869
CAOG_03870	XP_004363598.1	322	hypothetical protein CAOG_03870
CAOG_03883	XP_004363611.1	499	hypothetical protein CAOG_03883
CAOG_03900	XP_004363628.1	115	hypothetical protein CAOG_03900
CAOG_03901	XP_004363629.1	186	hypothetical protein CAOG_03901
CAOG_03912	XP_004363640.1	317	hypothetical protein CAOG_03912
CAOG_03928	XP_004363656.2	494	hypothetical protein CAOG_03928
CAOG_03931	XP_004363659.2	1027	hypothetical protein CAOG_03931
CAOG_03941	XP_004363669.2	866	hypothetical protein CAOG_03941
CAOG_03952	XP_004363680.1	289	hypothetical protein CAOG_03952
CAOG_03986	XP_004347811.2	616	hypothetical protein CAOG_03986
CAOG_04002	XP_004347827.1	638	hypothetical protein CAOG_04002
CAOG_04003	XP_004347828.1	329	hypothetical protein CAOG_04003
CAOG_04007	XP_004347832.1	258	hypothetical protein CAOG_04007
CAOG_04023	XP_004347848.1	837	hypothetical protein CAOG_04023
CAOG_04029	XP_004347854.2	665	hypothetical protein CAOG_04029
CAOG_04034	XP_004347859.1	167	hypothetical protein CAOG_04034
CAOG_04036	XP_004347861.1	419	hypothetical protein CAOG_04036
CAOG_04039	XP_004347864.1	371	hypothetical protein CAOG_04039
CAOG_04056	XP_004347881.2	342	hypothetical protein CAOG_04056
CAOG_04135	XP_004347960.2	634	hypothetical protein CAOG_04135
CAOG_04150	XP_004347975.1	1082	hypothetical protein CAOG_04150
CAOG_04155	XP_004347980.1	481	hypothetical protein CAOG_04155, partial
CAOG_04202	XP_004348027.1	241	hypothetical protein CAOG_04202
CAOG_04207	XP_004348032.2	699	hypothetical protein CAOG_04207
CAOG_04209	XP_004348034.1	1079	hypothetical protein CAOG_04209
CAOG_04232	XP_004348057.1	454	hypothetical protein CAOG_04232
CAOG_04240	XP_004348065.1	188	hypothetical protein CAOG_04240
CAOG_04248	XP_004348073.2	359	hypothetical protein CAOG_04248
CAOG_04251	XP_004348076.1	416	hypothetical protein CAOG_04251

CAOG_04272	XP_004348097.2	959	hypothetical protein CAOG_04272
CAOG_04279	XP_004348104.2	427	hypothetical protein CAOG_04279
CAOG_04309	XP_004348137.1	535	hypothetical protein CAOG_04309
CAOG_04318	XP_004348146.1	413	hypothetical protein CAOG_04318
CAOG_04324	XP_004348152.1	730	hypothetical protein CAOG_04324
CAOG_04329	XP_004348157.1	212	hypothetical protein CAOG_04329
CAOG_04364	XP_004348192.1	986	hypothetical protein CAOG_04364
CAOG_04373	XP_004348201.1	421	hypothetical protein CAOG_04373
CAOG_04382	XP_004348210.1	363	hypothetical protein CAOG_04382
CAOG_04388	XP_004348216.1	225	hypothetical protein CAOG_04388
CAOG_04391	XP_004348219.2	225	hypothetical protein CAOG_04391
CAOG_04405	XP_004348233.1	225	hypothetical protein CAOG_04405
CAOG_04412	XP_004348240.1	827	hypothetical protein CAOG_04412
CAOG_04413	XP_004348241.2	123	hypothetical protein CAOG_04413
CAOG_04419	XP_004348247.1	182	hypothetical protein CAOG_04419
CAOG_04430	XP_004348258.2	968	hypothetical protein CAOG_04430
CAOG_04437	XP_004348265.1	323	hypothetical protein CAOG_04437
CAOG_04444	XP_004348272.2	2157	hypothetical protein CAOG_04444
CAOG_04445	XP_004348273.1	190	hypothetical protein CAOG_04445
CAOG_04448	XP_004348276.2	1486	hypothetical protein CAOG_04448
CAOG_04451	XP_004348279.2	222	hypothetical protein CAOG_04451
CAOG_04466	XP_004348294.1	323	hypothetical protein CAOG_04466
CAOG_04470	XP_004348298.1	146	hypothetical protein CAOG_04470
CAOG_04503	XP_004348331.1	228	hypothetical protein CAOG_04503
CAOG_04516	XP_004347263.1	260	hypothetical protein CAOG_04516
CAOG_04525	XP_004347272.1	962	hypothetical protein CAOG_04525
CAOG_04534	XP_004347281.1	236	hypothetical protein CAOG_04534
CAOG_04547	XP_004347294.1	1257	hypothetical protein CAOG_04547
CAOG_04563	XP_004347310.2	357	hypothetical protein CAOG_04563
CAOG_04569	XP_004347316.1	975	hypothetical protein CAOG_04569
CAOG_04576	XP_004347323.1	185	hypothetical protein CAOG_04576
CAOG_04579	XP_004347326.1	471	hypothetical protein CAOG_04579
CAOG_04582	XP_004347329.1	1148	hypothetical protein CAOG_04582
CAOG_04584	XP_004347331.2	111	hypothetical protein CAOG_04584
CAOG_04588	XP_004347335.1	383	hypothetical protein CAOG_04588
CAOG_04642	XP_004347389.1	669	hypothetical protein CAOG_04642
CAOG_04648	XP_004347395.1	529	hypothetical protein CAOG_04648

CAOG_04666	XP_004347413.2	153	hypothetical protein CAOG_04666
CAOG_04717	XP_004347464.1	1055	hypothetical protein CAOG_04717
CAOG_04726	XP_004347473.1	519	hypothetical protein CAOG_04726
CAOG_04781	XP_004347532.1	518	hypothetical protein CAOG_04781
CAOG_04786	XP_004347537.2	704	hypothetical protein CAOG_04786
CAOG_04796	XP_004347547.1	486	hypothetical protein CAOG_04796
CAOG_04800	XP_004347551.1	103	hypothetical protein CAOG_04800
CAOG_04812	XP_004347563.1	236	hypothetical protein CAOG_04812
CAOG_04830	XP_004347581.2	217	hypothetical protein CAOG_04830
CAOG_04835	XP_004347586.1	667	hypothetical protein CAOG_04835
CAOG_04845	XP_004347596.1	383	hypothetical protein CAOG_04845
CAOG_04847	XP_004347598.2	165	hypothetical protein CAOG_04847
CAOG_04849	XP_004347600.2	716	hypothetical protein CAOG_04849
CAOG_04859	XP_004347610.1	459	hypothetical protein CAOG_04859
CAOG_04863	XP_004347614.1	177	hypothetical protein CAOG_04863
CAOG_04872	XP_004347623.2	699	hypothetical protein CAOG_04872
CAOG_04879	XP_004347630.2	801	hypothetical protein CAOG_04879
CAOG_04884	XP_004347635.1	265	hypothetical protein CAOG_04884
CAOG_04889	XP_004347640.1	802	hypothetical protein CAOG_04889
CAOG_04892	XP_004347643.2	189	hypothetical protein CAOG_04892
CAOG_04895	XP_004347646.1	84	hypothetical protein CAOG_04895
CAOG_04908	XP_004347659.1	109	hypothetical protein CAOG_04908
CAOG_04918	XP_004347669.1	407	hypothetical protein CAOG_04918
CAOG_04926	XP_004347677.1	554	hypothetical protein CAOG_04926
CAOG_04941	XP_004347692.1	1747	hypothetical protein CAOG_04941
CAOG_04999	XP_004346684.1	360	hypothetical protein CAOG_04999
CAOG_05003	XP_004346688.1	1349	hypothetical protein CAOG_05003
CAOG_05022	XP_004346707.1	280	hypothetical protein CAOG_05022
CAOG_05027	XP_004346712.2	284	hypothetical protein CAOG_05027
CAOG_05042	XP_004346727.1	392	hypothetical protein CAOG_05042
CAOG_05056	XP_004346741.2	503	hypothetical protein CAOG_05056
CAOG_05075	XP_004346760.2	625	hypothetical protein CAOG_05075
CAOG_05076	XP_004346761.2	209	hypothetical protein CAOG_05076
CAOG_05083	XP_004346768.1	1890	hypothetical protein CAOG_05083
CAOG_05085	XP_004346770.2	420	hypothetical protein CAOG_05085
CAOG_05108	XP_004346793.2	143	hypothetical protein CAOG_05108
CAOG_05132	XP_004346817.1	438	hypothetical protein CAOG_05132

CAOG_05142	XP_004346827.2	714	hypothetical protein CAOG_05142
CAOG_05163	XP_004346848.1	363	hypothetical protein CAOG_05163
CAOG_05168	XP_004346853.2	583	hypothetical protein CAOG_05168
CAOG_05169	XP_004346854.1	364	hypothetical protein CAOG_05169
CAOG_05183	XP_004346868.1	687	hypothetical protein CAOG_05183
CAOG_05184	XP_004346869.1	117	hypothetical protein CAOG_05184
CAOG_05189	XP_004346874.1	2742	hypothetical protein CAOG_05189
CAOG_05190	XP_004346875.1	156	hypothetical protein CAOG_05190
CAOG_05226	XP_004346911.2	1618	hypothetical protein CAOG_05226
CAOG_05233	XP_004346918.2	266	hypothetical protein CAOG_05233
CAOG_05238	XP_004346923.2	632	hypothetical protein CAOG_05238
CAOG_05246	XP_004346931.1	94	hypothetical protein CAOG_05246
CAOG_05248	XP_004346933.2	612	hypothetical protein CAOG_05248
CAOG_05251	XP_004346936.1	732	hypothetical protein CAOG_05251
CAOG_05254	XP_004346939.1	620	hypothetical protein CAOG_05254
CAOG_05259	XP_004346944.1	416	hypothetical protein CAOG_05259
CAOG_05288	XP_004346973.1	1106	hypothetical protein CAOG_05288
CAOG_05289	XP_004346974.1	459	hypothetical protein CAOG_05289
CAOG_05296	XP_004346981.1	353	hypothetical protein CAOG_05296
CAOG_05308	XP_004346993.1	877	hypothetical protein CAOG_05308
CAOG_05309	XP_004346994.1	1101	hypothetical protein CAOG_05309
CAOG_05312	XP_004346997.1	555	hypothetical protein CAOG_05312
CAOG_05323	XP_004347008.2	211	hypothetical protein CAOG_05323
CAOG_05339	XP_004347024.2	461	hypothetical protein CAOG_05339
CAOG_05381	XP_004347066.1	142	hypothetical protein CAOG_05381
CAOG_05435	XP_004346108.1	766	hypothetical protein CAOG_05435
CAOG_05459	XP_004346132.1	648	hypothetical protein CAOG_05459
CAOG_05460	XP_004346133.2	397	hypothetical protein CAOG_05460
CAOG_05461	XP_004346134.2	647	hypothetical protein CAOG_05461
CAOG_05472	XP_004346145.2	275	hypothetical protein CAOG_05472
CAOG_05485	XP_004346158.1	110	hypothetical protein CAOG_05485
CAOG_05497	XP_004346170.1	494	hypothetical protein CAOG_05497
CAOG_05502	XP_004346175.1	1026	hypothetical protein CAOG_05502
CAOG_05508	XP_004346181.2	690	hypothetical protein CAOG_05508
CAOG_05520	XP_004346193.1	1390	hypothetical protein CAOG_05520
CAOG_05539	XP_004346212.1	338	hypothetical protein CAOG_05539
CAOG_05554	XP_004346227.1	381	hypothetical protein CAOG_05554

CAOG_05559	XP_004346232.1	1329	hypothetical protein CAOG_05559
CAOG_05567	XP_004346240.1	560	hypothetical protein CAOG_05567
CAOG_05576	XP_004346249.1	163	hypothetical protein CAOG_05576
CAOG_05589	XP_004346262.2	1269	hypothetical protein CAOG_05589
CAOG_05603	XP_004346276.1	459	hypothetical protein CAOG_05603
CAOG_05638	XP_004346311.2	740	hypothetical protein CAOG_05638
CAOG_05647	XP_004346320.2	179	hypothetical protein CAOG_05647
CAOG_05652	XP_004346325.1	712	hypothetical protein CAOG_05652
CAOG_05661	XP_004346334.1	711	hypothetical protein CAOG_05661
CAOG_05682	XP_004346355.1	499	hypothetical protein CAOG_05682
CAOG_05708	XP_004346381.1	159	hypothetical protein CAOG_05708
CAOG_05709	XP_004346382.1	1199	hypothetical protein CAOG_05709
CAOG_05719	XP_004346392.2	787	hypothetical protein CAOG_05719
CAOG_05726	XP_004346399.2	203	hypothetical protein CAOG_05726
CAOG_05736	XP_004346409.2	114	hypothetical protein CAOG_05736
CAOG_05777	XP_004346450.1	789	hypothetical protein CAOG_05777
CAOG_05779	XP_004346452.1	749	hypothetical protein CAOG_05779
CAOG_05781	XP_004346454.1	210	hypothetical protein CAOG_05781
CAOG_05791	XP_004346464.2	744	hypothetical protein CAOG_05791
CAOG_05794	XP_004346467.1	124	hypothetical protein CAOG_05794
CAOG_05795	XP_004345385.1	100	hypothetical protein CAOG_05795
CAOG_05809	XP_004345399.2	522	hypothetical protein CAOG_05809
CAOG_05819	XP_004345409.1	164	hypothetical protein CAOG_05819
CAOG_05844	XP_004345434.1	602	hypothetical protein CAOG_05844
CAOG_05852	XP_004345442.2	522	hypothetical protein CAOG_05852
CAOG_05856	XP_004345446.1	99	hypothetical protein CAOG_05856
CAOG_05864	XP_004345454.2	119	hypothetical protein CAOG_05864
CAOG_05871	XP_004345461.1	837	hypothetical protein CAOG_05871
CAOG_05876	XP_004345466.1	437	hypothetical protein CAOG_05876
CAOG_05901	XP_004345491.2	1430	hypothetical protein CAOG_05901
CAOG_05984	XP_004345574.1	188	hypothetical protein CAOG_05984
CAOG_05998	XP_004345588.1	102	hypothetical protein CAOG_05998
CAOG_06011	XP_004345601.1	199	hypothetical protein CAOG_06011
CAOG_06066	XP_004345656.1	172	hypothetical protein CAOG_06066
CAOG_06077	XP_004345667.1	2246	hypothetical protein CAOG_06077
CAOG_06086	XP_004345676.1	624	hypothetical protein CAOG_06086
CAOG_06087	XP_004345677.2	424	hypothetical protein CAOG_06087

CAOG_06100	XP_004345690.1	1733	hypothetical protein CAOG_06100
CAOG_06115	XP_004345705.2	279	hypothetical protein CAOG_06115
CAOG_06119	XP_004345709.2	336	hypothetical protein CAOG_06119
CAOG_06121	XP_004345711.1	744	hypothetical protein CAOG_06121
CAOG_06129	XP_004345719.1	466	hypothetical protein CAOG_06129
CAOG_06137	XP_004345727.1	2286	hypothetical protein CAOG_06137
CAOG_06145	XP_004345735.1	1623	hypothetical protein CAOG_06145
CAOG_06150	XP_004345740.1	108	hypothetical protein CAOG_06150
CAOG_06161	XP_004345751.2	993	hypothetical protein CAOG_06161
CAOG_06186	XP_004345776.1	530	hypothetical protein CAOG_06186
CAOG_06198	XP_004345788.1	423	hypothetical protein CAOG_06198
CAOG_06202	XP_004345792.1	216	hypothetical protein CAOG_06202
CAOG_06204	XP_004345794.1	295	hypothetical protein CAOG_06204
CAOG_06209	XP_004345799.1	60	hypothetical protein CAOG_06209
CAOG_06227	XP_004344976.1	711	hypothetical protein CAOG_06227
CAOG_06232	XP_004344981.2	384	hypothetical protein CAOG_06232
CAOG_06236	XP_004344985.1	1380	hypothetical protein CAOG_06236
CAOG_06242	XP_004344991.1	1184	hypothetical protein CAOG_06242
CAOG_06260	XP_004345009.2	2217	hypothetical protein CAOG_06260
CAOG_06268	XP_004345017.2	442	hypothetical protein CAOG_06268
CAOG_06269	XP_004345018.2	704	hypothetical protein CAOG_06269
CAOG_06271	XP_004345020.1	246	hypothetical protein CAOG_06271
CAOG_06285	XP_004345034.1	712	hypothetical protein CAOG_06285
CAOG_06290	XP_004345040.1	409	hypothetical protein CAOG_06290
CAOG_06293	XP_004345042.2	1592	hypothetical protein CAOG_06293
CAOG_06294	XP_004345043.2	536	hypothetical protein CAOG_06294
CAOG_06308	XP_004345057.2	341	hypothetical protein CAOG_06308
CAOG_06339	XP_004345088.1	185	hypothetical protein CAOG_06339
CAOG_06351	XP_004345100.1	771	hypothetical protein CAOG_06351
CAOG_06355	XP_004345104.1	910	hypothetical protein CAOG_06355
CAOG_06371	XP_004345120.1	136	hypothetical protein CAOG_06371, partial
CAOG_06383	XP_004345132.1	1457	hypothetical protein CAOG_06383
CAOG_06392	XP_004345141.1	699	hypothetical protein CAOG_06392
CAOG_06396	XP_004345145.1	69	hypothetical protein CAOG_06396
CAOG_06403	XP_004345152.1	544	hypothetical protein CAOG_06403
CAOG_06415	XP_004345164.2	599	hypothetical protein CAOG_06415

CAOG_06434	XP_004345183.2	869	hypothetical protein CAOG_06434
CAOG_06455	XP_004345204.1	175	hypothetical protein CAOG_06455
CAOG_06461	XP_004345210.1	272	hypothetical protein CAOG_06461
CAOG_06475	XP_004345224.2	435	hypothetical protein CAOG_06475
CAOG_06480	XP_004345229.2	948	hypothetical protein CAOG_06480
CAOG_06523	XP_004345272.2	407	hypothetical protein CAOG_06523
CAOG_06525	XP_004345274.1	869	hypothetical protein CAOG_06525
CAOG_06547	XP_004345296.1	338	hypothetical protein CAOG_06547
CAOG_06549	XP_004345298.1	2104	hypothetical protein CAOG_06549
CAOG_06552	XP_004345301.2	579	hypothetical protein CAOG_06552
CAOG_06560	XP_004345309.1	946	hypothetical protein CAOG_06560
CAOG_06609	XP_004344230.1	1342	hypothetical protein CAOG_06609
CAOG_06636	XP_004344257.1	268	hypothetical protein CAOG_06636
CAOG_06641	XP_004344262.1	1000	hypothetical protein CAOG_06641
CAOG_06646	XP_004344267.1	680	hypothetical protein CAOG_06646
CAOG_06647	XP_004344268.2	306	hypothetical protein CAOG_06647
CAOG_06653	XP_004344274.1	186	hypothetical protein CAOG_06653
CAOG_06657	XP_004344278.2	334	hypothetical protein CAOG_06657
CAOG_06659	XP_004344280.2	657	hypothetical protein CAOG_06659
CAOG_06661	XP_004344282.2	349	hypothetical protein CAOG_06661
CAOG_06684	XP_004344305.2	447	hypothetical protein CAOG_06684
CAOG_06709	XP_004344330.2	311	hypothetical protein CAOG_06709
CAOG_06724	XP_004344345.1	1541	hypothetical protein CAOG_06724
CAOG_06726	XP_004344347.1	552	hypothetical protein CAOG_06726
CAOG_06759	XP_004344380.1	301	hypothetical protein CAOG_06759
CAOG_06761	XP_004344382.1	276	hypothetical protein CAOG_06761
CAOG_06767	XP_004344388.1	946	hypothetical protein CAOG_06767
CAOG_06776	XP_004344397.1	423	hypothetical protein CAOG_06776
CAOG_06780	XP_004344401.1	109	hypothetical protein CAOG_06780
CAOG_06781	XP_004344402.1	288	hypothetical protein CAOG_06781
CAOG_06789	XP_004344410.1	112	hypothetical protein CAOG_06789
CAOG_06805	XP_004344426.2	589	hypothetical protein CAOG_06805
CAOG_06807	XP_004344428.2	415	hypothetical protein CAOG_06807
CAOG_06808	XP_004344429.1	170	hypothetical protein CAOG_06808
CAOG_06810	XP_004344431.1	63	hypothetical protein CAOG_06810
CAOG_06820	XP_004344441.1	550	hypothetical protein CAOG_06820
CAOG_06836	XP_004344457.1	392	hypothetical protein CAOG_06836

CAOG_06852	XP_004344473.1	250	hypothetical protein CAOG_06852
CAOG_06854	XP_004344475.1	581	hypothetical protein CAOG_06854
CAOG_06862	XP_004344483.1	985	hypothetical protein CAOG_06862
CAOG_06865	XP_004344486.1	429	hypothetical protein CAOG_06865
CAOG_06866	XP_004344487.2	786	hypothetical protein CAOG_06866
CAOG_06869	XP_004344490.1	131	hypothetical protein CAOG_06869
CAOG_06896	XP_004344517.2	692	hypothetical protein CAOG_06896
CAOG_06911	XP_004344532.1	273	hypothetical protein CAOG_06911
CAOG_06912	XP_004344533.1	243	hypothetical protein CAOG_06912
CAOG_06928	XP_004344549.2	695	hypothetical protein CAOG_06928
CAOG_06938	XP_004343662.1	117	hypothetical protein CAOG_06938
CAOG_06952	XP_004343676.1	156	hypothetical protein CAOG_06952
CAOG_06957	XP_004343681.2	496	hypothetical protein CAOG_06957
CAOG_06970	XP_004343694.2	848	hypothetical protein CAOG_06970
CAOG_06975	XP_004343699.1	499	hypothetical protein CAOG_06975
CAOG_06976	XP_004343700.2	334	hypothetical protein CAOG_06976
CAOG_06980	XP_004343704.1	563	hypothetical protein CAOG_06980
CAOG_06999	XP_004343723.1	888	hypothetical protein CAOG_06999
CAOG_07015	XP_004343739.1	491	hypothetical protein CAOG_07015
CAOG_07036	XP_004343760.1	354	hypothetical protein CAOG_07036
CAOG_07045	XP_004343769.1	619	hypothetical protein CAOG_07045
CAOG_07062	XP_004343786.1	540	hypothetical protein CAOG_07062
CAOG_07084	XP_004343808.2	298	hypothetical protein CAOG_07084
CAOG_07099	XP_004343823.2	162	hypothetical protein CAOG_07099
CAOG_07117	XP_004343841.1	541	hypothetical protein CAOG_07117
CAOG_07124	XP_004343848.2	584	hypothetical protein CAOG_07124
CAOG_07125	XP_004343849.1	498	hypothetical protein CAOG_07125
CAOG_07128	XP_004343852.1	545	hypothetical protein CAOG_07128
CAOG_07130	XP_004343854.1	325	hypothetical protein CAOG_07130
CAOG_07136	XP_004343860.2	267	hypothetical protein CAOG_07136
CAOG_07155	XP_004343879.2	548	hypothetical protein CAOG_07155
CAOG_07160	XP_004343884.1	268	hypothetical protein CAOG_07160
CAOG_07180	XP_004343904.2	334	hypothetical protein CAOG_07180
CAOG_07191	XP_004343915.1	805	hypothetical protein CAOG_07191
CAOG_07192	XP_004343916.2	645	hypothetical protein CAOG_07192
CAOG_07202	XP_004343926.1	176	hypothetical protein CAOG_07202
CAOG_07206	XP_004343930.2	271	hypothetical protein CAOG_07206

CAOG_07207	XP_004343931.1	3184	hypothetical protein CAOG_07207
CAOG_07214	XP_004343073.1	181	hypothetical protein CAOG_07214
CAOG_07217	XP_004343076.2	1611	hypothetical protein CAOG_07217
CAOG_07223	XP_004343082.2	396	hypothetical protein CAOG_07223
CAOG_07229	XP_004343088.1	1511	hypothetical protein CAOG_07229
CAOG_07235	XP_004343094.1	521	hypothetical protein CAOG_07235
CAOG_07246	XP_004343105.1	409	hypothetical protein CAOG_07246
CAOG_07266	XP_004343125.1	1561	hypothetical protein CAOG_07266
CAOG_07269	XP_004343128.1	989	hypothetical protein CAOG_07269
CAOG_07271	XP_004343130.2	2985	hypothetical protein CAOG_07271
CAOG_07284	XP_004343143.2	1141	hypothetical protein CAOG_07284
CAOG_07334	XP_004343193.1	452	hypothetical protein CAOG_07334
CAOG_07343	XP_004343202.1	814	hypothetical protein CAOG_07343
CAOG_07348	XP_004343207.1	79	hypothetical protein CAOG_07348
CAOG_07354	XP_004343213.2	223	hypothetical protein CAOG_07354
CAOG_07370	XP_004343229.2	661	hypothetical protein CAOG_07370
CAOG_07388	XP_004343247.2	185	hypothetical protein CAOG_07388
CAOG_07418	XP_004343277.1	333	hypothetical protein CAOG_07418
CAOG_07429	XP_004343288.1	1880	hypothetical protein CAOG_07429
CAOG_07434	XP_004343293.2	860	hypothetical protein CAOG_07434
CAOG_07446	XP_004343305.1	839	hypothetical protein CAOG_07446
CAOG_07451	XP_004343310.1	612	hypothetical protein CAOG_07451
CAOG_07453	XP_004343312.1	360	hypothetical protein CAOG_07453
CAOG_07458	XP_004343317.1	271	hypothetical protein CAOG_07458
CAOG_07459	XP_004343318.1	385	hypothetical protein CAOG_07459
CAOG_07463	XP_004343337.1	641	hypothetical protein CAOG_07463
CAOG_07468	XP_004343342.2	326	hypothetical protein CAOG_07468
CAOG_07481	XP_004343355.1	108	hypothetical protein CAOG_07481
CAOG_07503	XP_004343377.1	166	hypothetical protein CAOG_07503
CAOG_07507	XP_004343381.1	64	hypothetical protein CAOG_07507
CAOG_07516	XP_004343390.1	101	hypothetical protein CAOG_07516
CAOG_07517	XP_004343391.1	431	hypothetical protein CAOG_07517
CAOG_07563	XP_004343437.1	100	hypothetical protein CAOG_07563
CAOG_07590	XP_004343464.2	428	hypothetical protein CAOG_07590
CAOG_07620	XP_004343494.1	155	hypothetical protein CAOG_07620
CAOG_07622	XP_004343496.2	283	hypothetical protein CAOG_07622
CAOG_07642	XP_004343516.1	246	hypothetical protein CAOG_07642

CAOG_07652	XP_004343526.1	472	hypothetical protein CAOG_07652
CAOG_07674	XP_004343548.1	177	hypothetical protein CAOG_07674
CAOG_07675	XP_004343549.1	1044	hypothetical protein CAOG_07675
CAOG_07697	XP_004343571.1	306	hypothetical protein CAOG_07697
CAOG_07755	XP_004342828.2	452	hypothetical protein CAOG_07755, partial
CAOG_07760	XP_004342833.1	842	hypothetical protein CAOG_07760
CAOG_07766	XP_004342839.2	6160	hypothetical protein CAOG_07766
CAOG_07771	XP_004342844.1	382	hypothetical protein CAOG_07771
CAOG_07777	XP_004342850.1	308	hypothetical protein CAOG_07777
CAOG_07805	XP_004342878.1	221	hypothetical protein CAOG_07805
CAOG_07806	XP_004342879.1	2199	hypothetical protein CAOG_07806
CAOG_07815	XP_004342888.1	1484	hypothetical protein CAOG_07815
CAOG_07821	XP_004342894.1	86	hypothetical protein CAOG_07821
CAOG_07834	XP_004342907.1	111	hypothetical protein CAOG_07834
CAOG_07839	XP_004342912.1	1656	hypothetical protein CAOG_07839
CAOG_07844	XP_004342918.1	1615	hypothetical protein CAOG_07844
CAOG_07856	XP_004342941.2	813	hypothetical protein CAOG_07856
CAOG_07858	XP_004342943.1	2627	hypothetical protein CAOG_07858
CAOG_07861	XP_004342946.1	506	hypothetical protein CAOG_07861
CAOG_07894	XP_004342979.1	467	hypothetical protein CAOG_07894
CAOG_07899	XP_004342984.1	478	hypothetical protein CAOG_07899
CAOG_07901	XP_004342986.1	703	hypothetical protein CAOG_07901
CAOG_07903	XP_004342988.1	199	hypothetical protein CAOG_07903
CAOG_07909	XP_004342994.1	508	hypothetical protein CAOG_07909
CAOG_07914	XP_004342999.2	628	hypothetical protein CAOG_07914
CAOG_07930	XP_004343015.1	224	hypothetical protein CAOG_07930
CAOG_07942	XP_004343030.1	295	hypothetical protein CAOG_07942
CAOG_08008	XP_004342609.1	436	hypothetical protein CAOG_08008
CAOG_08011	XP_004342612.1	4326	hypothetical protein CAOG_08011
CAOG_08013	XP_004342614.1	518	hypothetical protein CAOG_08013
CAOG_08014	XP_004342615.1	351	hypothetical protein CAOG_08014
CAOG_08055	XP_004342656.1	147	hypothetical protein CAOG_08055
CAOG_08074	XP_004342675.1	237	hypothetical protein CAOG_08074
CAOG_08099	XP_004342700.1	116	hypothetical protein CAOG_08099
CAOG_08109	XP_004342710.2	733	hypothetical protein CAOG_08109
CAOG_08112	XP_004342713.1	579	hypothetical protein CAOG_08112

CAOG_08126	XP_004342727.1	1593	hypothetical protein CAOG_08126
CAOG_08139	XP_004342740.2	255	hypothetical protein CAOG_08139
CAOG_08144	XP_004342745.2	513	hypothetical protein CAOG_08144
CAOG_08163	XP_004342764.1	180	hypothetical protein CAOG_08163
CAOG_08166	XP_004342767.2	763	hypothetical protein CAOG_08166
CAOG_08169	XP_004342770.1	233	hypothetical protein CAOG_08169
CAOG_08173	XP_004342774.1	449	hypothetical protein CAOG_08173
CAOG_08185	XP_004342786.1	796	hypothetical protein CAOG_08185
CAOG_08209	XP_004342464.2	694	hypothetical protein CAOG_08209
CAOG_08212	XP_004342467.1	86	hypothetical protein CAOG_08212
CAOG_08225	XP_004342480.1	62	hypothetical protein CAOG_08225
CAOG_08251	XP_004342420.1	430	hypothetical protein CAOG_08251
CAOG_08259	XP_004342428.1	1777	hypothetical protein CAOG_08259
CAOG_08261	XP_004342430.1	438	hypothetical protein CAOG_08261
CAOG_08270	XP_004342439.1	210	hypothetical protein CAOG_08270
CAOG_08283	XP_004342006.1	423	hypothetical protein CAOG_08283
CAOG_08337	XP_004340960.2	279	hypothetical protein CAOG_08337
CAOG_08338	XP_004340787.1	237	hypothetical protein CAOG_08338
CAOG_08346	XP_004340843.1	566	hypothetical protein CAOG_08346, partial
CAOG_08352	XP_004340771.1	184	hypothetical protein CAOG_08352
CAOG_08353	XP_004340833.1	155	hypothetical protein CAOG_08353
CAOG_08357	XP_004340773.2	972	hypothetical protein CAOG_08357
CAOG_08360	XP_004340463.1	135	hypothetical protein CAOG_08360
CAOG_08367	XP_004340491.2	448	hypothetical protein CAOG_08367
CAOG_08383	XP_011269953.1	1330	hypothetical protein CAOG_08383
CAOG_08384	XP_011269954.1	395	hypothetical protein CAOG_08384
CAOG_08391	XP_011269961.1	453	hypothetical protein CAOG_08391
CAOG_08418	XP_011269989.1	484	hypothetical protein CAOG_08418
CAOG_08425	XP_011269996.1	158	hypothetical protein CAOG_08425
CAOG_08426	XP_011269997.1	635	hypothetical protein CAOG_08426
CAOG_08429	XP_011270000.1	708	hypothetical protein CAOG_08429
CAOG_08434	XP_011270005.1	322	hypothetical protein CAOG_08434
CAOG_08443	XP_011270014.1	202	hypothetical protein CAOG_08443
CAOG_08455	XP_011270027.1	955	hypothetical protein CAOG_08455
CAOG_08462	XP_011270034.1	605	hypothetical protein CAOG_08462
CAOG_08463	XP_011270035.1	761	hypothetical protein CAOG_08463

CAOG_08470	XP_011270042.1	917	hypothetical protein CAOG_08470
CAOG_08475	XP_011270047.1	169	hypothetical protein CAOG_08475, partial
CAOG_08476	XP_011270048.1	695	hypothetical protein CAOG_08476
CAOG_08499	XP_011270081.1	759	hypothetical protein CAOG_08499
CAOG_08511	XP_011270093.1	114	hypothetical protein CAOG_08511
CAOG_08513	XP_011270095.1	102	hypothetical protein CAOG_08513
CAOG_08529	XP_011270110.1	191	hypothetical protein CAOG_08529
CAOG_08536	XP_011270117.1	157	hypothetical protein CAOG_08536
CAOG_08544	XP_011270126.1	88	hypothetical protein CAOG_08544
CAOG_08546	XP_011270128.1	505	hypothetical protein CAOG_08546
CAOG_08550	XP_011270132.1	211	hypothetical protein CAOG_08550, partial
CAOG_08560	XP_011270142.1	212	hypothetical protein CAOG_08560, partial
CAOG_08566	XP_011270148.1	535	hypothetical protein CAOG_08566
CAOG_08588	XP_011270188.1	80	hypothetical protein CAOG_08588
CAOG_08590	XP_011270190.1	128	hypothetical protein CAOG_08590, partial
CAOG_08598	XP_011270198.1	310	hypothetical protein CAOG_08598
CAOG_08605	XP_011270205.1	908	hypothetical protein CAOG_08605
CAOG_08623	XP_011270223.1	286	hypothetical protein CAOG_08623
CAOG_08630	XP_011270230.1	2056	hypothetical protein CAOG_08630
CAOG_08644	XP_011270255.1	592	hypothetical protein CAOG_08644
CAOG_08652	XP_011270263.1	86	hypothetical protein CAOG_08652
CAOG_08662	XP_011270273.1	328	hypothetical protein CAOG_08662
CAOG_08666	XP_011270277.1	283	hypothetical protein CAOG_08666
CAOG_08681	XP_011270292.1	321	hypothetical protein CAOG_08681
CAOG_08686	XP_011270297.1	97	hypothetical protein CAOG_08686
CAOG_08705	XP_011270325.1	1328	hypothetical protein CAOG_08705
CAOG_08710	XP_011270330.1	768	hypothetical protein CAOG_08710
CAOG_08712	XP_011270332.1	945	hypothetical protein CAOG_08712
CAOG_08716	XP_011270336.1	113	hypothetical protein CAOG_08716
CAOG_08720	XP_011270340.1	374	hypothetical protein CAOG_08720
CAOG_08727	XP_011270347.1	992	hypothetical protein CAOG_08727
CAOG_08775	XP_011270405.1	1030	hypothetical protein CAOG_08775
CAOG_08783	XP_011270413.1	2093	hypothetical protein CAOG_08783
CAOG_08785	XP_011270415.1	379	hypothetical protein CAOG_08785

CAOG_08789	XP_011270429.1	142	hypothetical protein CAOG_08789, partial
CAOG_08801	XP_011270441.1	103	hypothetical protein CAOG_08801
CAOG_08806	XP_011270446.1	389	hypothetical protein CAOG_08806
CAOG_08808	XP_011270448.1	829	hypothetical protein CAOG_08808
CAOG_08816	XP_011270456.1	157	hypothetical protein CAOG_08816
CAOG_08820	XP_011270460.1	313	hypothetical protein CAOG_08820
CAOG_08825	XP_011270465.1	412	hypothetical protein CAOG_08825
CAOG_08826	XP_011270466.1	508	hypothetical protein CAOG_08826
CAOG_08836	XP_011270477.1	418	hypothetical protein CAOG_08836
CAOG_08845	XP_011270499.1	432	hypothetical protein CAOG_08845
CAOG_08846	XP_011270500.1	887	hypothetical protein CAOG_08846
CAOG_08859	XP_011270513.1	669	hypothetical protein CAOG_08859
CAOG_08863	XP_011270518.1	803	hypothetical protein CAOG_08863
CAOG_08864	XP_011270519.1	2586	hypothetical protein CAOG_08864
CAOG_08865	XP_011270520.1	452	hypothetical protein CAOG_08865
CAOG_08878	XP_011270533.1	939	hypothetical protein CAOG_08878, partial
CAOG_08886	XP_011270548.1	62	hypothetical protein CAOG_08886
CAOG_08887	XP_011270549.1	573	hypothetical protein CAOG_08887
CAOG_08898	XP_011270560.1	249	hypothetical protein CAOG_08898
CAOG_08906	XP_011270568.1	461	hypothetical protein CAOG_08906
CAOG_08909	XP_011270571.1	386	hypothetical protein CAOG_08909
CAOG_08934	XP_011270605.1	337	hypothetical protein CAOG_08934
CAOG_08935	XP_011270606.1	1985	hypothetical protein CAOG_08935
CAOG_08939	XP_011270610.1	214	hypothetical protein CAOG_08939
CAOG_08940	XP_011270611.1	723	hypothetical protein CAOG_08940
CAOG_08946	XP_011270617.1	622	hypothetical protein CAOG_08946
CAOG_08962	XP_011270640.1	133	hypothetical protein CAOG_08962
CAOG_08974	XP_011270652.1	700	hypothetical protein CAOG_08974
CAOG_08985	XP_011270663.1	980	hypothetical protein CAOG_08985
CAOG_08994	XP_011270672.1	375	hypothetical protein CAOG_08994
CAOG_08996	XP_011270683.1	1611	hypothetical protein CAOG_08996
CAOG_09003	XP_011270690.1	2463	hypothetical protein CAOG_09003
CAOG_09004	XP_011270691.1	380	hypothetical protein CAOG_09004
CAOG_09006	XP_011270693.1	176	hypothetical protein CAOG_09006

CAOG_09010	XP_011270697.1	506	hypothetical protein CAOG_09010, partial
CAOG_09019	XP_011270706.1	921	hypothetical protein CAOG_09019
CAOG_09036	XP_011270724.1	377	hypothetical protein CAOG_09036
CAOG_09045	XP_011270733.1	1026	hypothetical protein CAOG_09045
CAOG_09048	XP_011270736.1	498	hypothetical protein CAOG_09048
CAOG_09056	XP_011270789.1	331	hypothetical protein CAOG_09056
CAOG_09077	XP_011270810.1	3369	hypothetical protein CAOG_09077
CAOG_09083	XP_011270816.1	839	hypothetical protein CAOG_09083
CAOG_09095	XP_011270752.1	394	hypothetical protein CAOG_09095
CAOG_09100	XP_011270757.1	211	hypothetical protein CAOG_09100
CAOG_09101	XP_011270758.1	333	hypothetical protein CAOG_09101, partial
CAOG_09116	XP_011270773.1	501	hypothetical protein CAOG_09116
CAOG_09120	XP_011270777.1	294	hypothetical protein CAOG_09120, partial
CAOG_09125	XP_011270835.1	1573	hypothetical protein CAOG_09125
CAOG_09133	XP_011270843.1	12732	hypothetical protein CAOG_09133
CAOG_09134	XP_011270844.1	350	hypothetical protein CAOG_09134
CAOG_09140	XP_011270850.1	797	hypothetical protein CAOG_09140
CAOG_09146	XP_011270859.1	725	hypothetical protein CAOG_09146
CAOG_09164	XP_011270877.1	244	hypothetical protein CAOG_09164, partial
CAOG_09172	XP_011270885.1	822	hypothetical protein CAOG_09172
CAOG_09187	XP_011270904.1	279	hypothetical protein CAOG_09187, partial
CAOG_09188	XP_011270905.1	592	hypothetical protein CAOG_09188, partial
CAOG_09192	XP_011270909.1	1420	hypothetical protein CAOG_09192
CAOG_09203	XP_011270920.1	127	hypothetical protein CAOG_09203
CAOG_09207	XP_011270924.1	287	hypothetical protein CAOG_09207
CAOG_09208	XP_011270944.1	241	hypothetical protein CAOG_09208, partial
CAOG_09223	XP_011270935.1	813	hypothetical protein CAOG_09223
CAOG_05909	XP_004345499.1	190	import inner membrane translocase subunit Tim22
CAOG_06578	XP_004345327.1	524	inosine monophosphate dehydrogenase 2
CAOG_01679	XP_004364547.1	541	insect-derived growth factor-B-like protein

CAOG_02408	XP_004349158.1	260	integral membrane protein
CAOG_07424	XP_004343283.1	309	iron-sulfur protein I
CAOG_08016	XP_004342617.1	467	iron/zinc purple acid phosphatase-like protein
CAOG_02772	XP_004349525.1	308	ketoheokinase
CAOG_07968	XP_004343056.1	832	krox-20
CAOG_03608	XP_004363336.1	978	krueppel-like factor 16
CAOG_05737	XP_004346410.1	476	L-gulono-gamma-lactone oxidase
CAOG_03827	XP_004363555.1	321	L-isoaspartyl protein carboxyl methyltransferase
CAOG_00854	XP_004365725.1	511	leucine-rich repeat containing protein
CAOG_03990	XP_004347815.2	1110	leucine-rich repeat-containing protein 28
CAOG_02239	XP_004348989.1	2004	leucine-rich repeat-containing protein 69
CAOG_06135	XP_004345725.1	521	lipase, family member N
CAOG_03075	XP_004363914.1	645	long-chain-fatty-acid-CoA ligase ACSBG2
CAOG_01478	XP_004364346.1	738	lss protein
CAOG_04352	XP_004348180.2	506	lung seven transmembrane receptor
CAOG_07225	XP_004343084.2	216	LYR domain-containing protein-containing protein 5
CAOG_04258	XP_004348083.2	711	lysophosphatidylcholine acyltransferase 2-B
CAOG_04992	XP_004346677.1	431	lysophosphatidylcholine acyltransferase 3
CAOG_00329	XP_004365200.1	786	lysosomal alpha-N-acetyl glucosaminidase
CAOG_07000	XP_004343724.1	217	lysozyme
CAOG_05009	XP_004346694.1	411	macrophage erythroblast attacher protein
CAOG_06815	XP_004344436.1	245	Maf1-pending-prov protein
CAOG_06529	XP_004345278.1	597	max-like protein X
CAOG_08090	XP_004342691.1	358	metal-binding domain in RNase L inhibitor
CAOG_08049	XP_004342650.1	426	metallo-beta-lactamase
CAOG_00476	XP_004365347.1	282	methyltransferase
CAOG_07700	XP_004343574.1	117	microtubule associated protein
CAOG_03874	XP_004363602.2	1911	mitogen-activated protein kinase kinase kinase 1

CAOG_04640	XP_004347387.1	2280	mitogen-activated protein kinase kinase kinase 1
CAOG_03065	XP_004363904.2	1156	mitogen-activated protein kinase kinase kinase 3
CAOG_08035	XP_004342636.2	531	mitogen-activated protein kinase kinase kinase
CAOG_02162	XP_004348912.2	657	molybdenum cofactor synthesis 1 isoform 1
CAOG_04319	XP_004348147.1	305	molybdenum cofactor synthesis protein 2 large subunit
CAOG_06543	XP_004345292.2	462	molybdopterin synthase sulfurylase
CAOG_00650	XP_004365521.2	240	Mrto4 protein
CAOG_07710	XP_004343584.1	546	methfr-prov protein
CAOG_02706	XP_004349456.2	1491	multi-sensor hybrid histidine kinase
CAOG_02063	XP_004348813.2	2168	myosin IXA
CAOG_00121	XP_004364992.2	1291	N-acetylglucosamine-1-phosphate transferase
CAOG_04189	pseudo gene		n-acetylglucosamine-6-phosphate deacetylase
CAOG_10025			NAD-dependent deacetylase sirtuin 7
CAOG_03222	XP_004364061.1	334	NAD-dependent epimerase/dehydratase
CAOG_03621	XP_004363349.1	127	NADH dehydrogenase 1 beta subcomplex subunit 9
CAOG_07923	XP_004343008.1	709	NADPH-cytochrome P450 reductase
CAOG_03423	XP_004364262.1	215	Nat9 protein
CAOG_00591	XP_004365462.1	257	natural killer cell-specific antigen KLIP1
CAOG_01134	XP_004366005.1	1447	neuropathy target esterase
CAOG_01936	XP_004364804.1	1063	neutral sphingomyelinase activation associated factor
CAOG_05761	XP_004346434.2	997	non-receptor protein kinase
CAOG_08366	XP_004340478.1	1380	non-receptor protein kinase, partial
CAOG_02457	XP_004349207.1	2031	NPR2 protein
CAOG_06310	XP_004345059.1	635	NTRK3 protein
CAOG_01270	XP_004349790.1	957	nuclear elongation and deformation protein 1
CAOG_07127	XP_004343851.2	343	nuclease Le3
CAOG_07597	XP_004343471.1	1042	nucleic acid binding protein

CAOG_04115	XP_004347940.1	274	nucleotide-binding protein 2
CAOG_06336	XP_004345085.1	700	oxidation resistance protein 1
CAOG_07035	XP_004343759.2	1349	patatin-like phospholipase domain-containing protein 7
CAOG_04074	XP_004347899.2	448	peptidase M19
CAOG_03679	XP_004363407.1	660	peptidase S8 and S53
CAOG_06448	XP_004345197.1	172	peptide methionine sulfoxide reductase
CAOG_01957	XP_004364825.1	583	periodic tryptophan protein
CAOG_00607	XP_004365478.1	462	phenylalanyl-tRNA synthetase
CAOG_07357	XP_004343216.1	547	phosphate transporter
CAOG_02967	XP_004363806.2	777	phosphatidylethanolamine N-methyltransferase
CAOG_01921	XP_004364789.1	251	phosphatidylinositol glycan anchor biosynthesis
CAOG_00043	XP_004364914.2	340	phospho-2-dehydro-3-deoxyheptonate aldolase
CAOG_04210	XP_004348035.2	1564	phosphodiesterase 1B
CAOG_04883	XP_004347634.1	719	phosphodiesterase 2A
CAOG_02238	XP_004348988.2	945	phosphoinositide-3-kinase
CAOG_06640	XP_004344261.1	164	phospholipid transfer protein
CAOG_00556	XP_004365427.1	249	phosphomannomutase
CAOG_00222	XP_004365093.1	303	phosphopantothenoylcysteine decarboxylase
CAOG_05987	XP_004345577.1	299	phosphopantothenoylcysteine synthetase
CAOG_04597	XP_004347344.1	299	phosphoribosylaminoimidazole-succinocarboxamide synthase
CAOG_07423	XP_004343282.2	1970	Pik4cb protein
CAOG_06417	XP_004345166.1	353	pirin domain-containing protein domain-containing protein
CAOG_03564	XP_004363292.2	884	polyadenylate-binding protein
CAOG_00164	XP_004365035.2	610	potassium channel tetramerisation domain-containing protein
CAOG_00596	XP_004365467.2	176	prefoldin
CAOG_05548	XP_004346221.2	242	programmed cell death protein
CAOG_04153	XP_004347978.1	950	protease m1 zinc metalloprotease
CAOG_05835	XP_004345425.2	643	protein arginine methyltransferase 3
CAOG_00972	XP_004365843.1	487	protein disulfide-isomerase ERp60
CAOG_01341	XP_004349861.2	504	protein kinase

CAOG_05528	XP_004346201.2	1200	protein kinase
CAOG_08111	XP_004342712.1	231	protein-S-isoprenylcysteine O-methyltransferase
CAOG_06163	XP_004345753.2	1446	protein-tyrosine-phosphatase
CAOG_02035	XP_004348785.2	308	pyridoxamine 5'-phosphate oxidase
CAOG_04392	XP_004348220.2	474	queuine tRNA-ribosyltransferase
CAOG_05584	XP_004346257.1	480	radical S-adenosyl methionine domain-containing protein 1
CAOG_01797	XP_004364665.2	1007	radixin isoform a
CAOG_00452	XP_004365323.2	1090	rap guanine nucleotide exchange factor GEF2
CAOG_00294	XP_004365165.1	187	ras family protein
CAOG_06717	XP_004344338.1	956	ras guanine nucleotide exchange factor, partial
CAOG_10093			Ras like enriched
CAOG_08182	XP_004342783.1	202	Ras-like protein Rab-21
CAOG_00778	XP_004365649.1	1052	ras-specific guanine nucleotide-releasing factor RalGPS1
CAOG_01664	XP_004364532.2	413	Rbm22 protein
CAOG_02716	XP_004349466.1	176	receptor accessory protein 5
CAOG_02461	XP_004349211.2	1533	response regulator receiver
CAOG_05966	XP_004345556.1	827	response regulator receiver domain-containing protein
CAOG_05260	XP_004346945.2	494	rho GTPase activating protein
CAOG_05230	XP_004346915.1	1876	rho GTPase-activating protein
CAOG_07539	XP_004343413.1	455	rho GTPase-activating protein 8
CAOG_02310	XP_004349060.2	361	ribonuclease P/MRP 30 subunit
CAOG_01734	XP_004364602.1	60	ribosomal protein L29e
CAOG_03983	XP_004347808.2	326	ribosomal RNA large subunit methyltransferase J
CAOG_07203	XP_004343927.1	173	ring finger protein, partial
CAOG_07739	XP_004343613.2	648	RIO kinase 3
CAOG_03214	XP_004364053.1	1179	RNA binding domain-containing protein
CAOG_07410	XP_004343269.2	302	RNA-binding protein PNO1
CAOG_00935	XP_004365806.2	428	saccharopine dehydrogenase
CAOG_00040	XP_004364911.1	215	SapB domain-containing protein
CAOG_08152	XP_004342753.1	629	SEC14-like protein

CAOG_03476	XP_004348381.1	557	selenocysteine-specific elongation factor
CAOG_05807	XP_004345397.2	1951	sensor protein GacS
CAOG_03468	XP_004348373.1	609	septrin/SUMO-specific protease 15
CAOG_02174	XP_004348924.1	252	sepiapterin reductase
CAOG_07938	XP_004343026.1	428	Ser/Thr kinase
CAOG_05563	XP_004346236.1	825	serine/threonine protein kinase
CAOG_08028	XP_004342629.1	340	serine/threonine protein kinase
CAOG_00905	XP_004365776.1	2948	serine/threonine protein kinase 15
CAOG_01911	XP_004364779.2	509	serine/threonine-protein kinase gad8
CAOG_03006	XP_004363845.2	1600	serine/threonine-protein kinase ppk15
CAOG_02400	XP_004349150.2	876	serine/threonine-protein kinase SAPK2
CAOG_05649	XP_004346322.2	1172	serine/threonine-specific protein kinase NAK
CAOG_04062	XP_004347887.2	1154	serologically defined colon cancer antigen 1
CAOG_00420	XP_004365291.2	559	serum/glucocorticoid regulated kinase 2
CAOG_04446	XP_004348274.1	723	sestrin 1
CAOG_03488	XP_004348393.1	607	SFRS protein kinase 1
CAOG_04695	XP_004347442.1	365	Sh3yl1-prov protein
CAOG_03052	XP_004363891.1	221	small GTP-binding protein Rab4
CAOG_04656	XP_004347403.1	177	small HspC2 heat shock protein
CAOG_07464	XP_004343338.2	447	SMYD5 protein
CAOG_04776	XP_004347527.1	199	SNARE Ykt6
CAOG_06772	XP_004344393.2	1785	snrk protein
CAOG_01682	XP_004364550.2	97	snRNP core protein SMX5
CAOG_04882	XP_004347633.1	466	snurportin-1
CAOG_00414	XP_004365285.1	988	sodium/calcium exchanger protein
CAOG_01804	XP_004364672.1	296	soluble NSF attachment protein
CAOG_07131	XP_004343855.2	574	solute carrier family 24
CAOG_02881	XP_004363720.1	330	solute carrier family 25
CAOG_04976	XP_004346661.1	310	solute carrier family 25
CAOG_07519	XP_004343393.1	381	solute carrier family 25
CAOG_00501	XP_004365372.1	405	solute carrier family 35
CAOG_06826	XP_004344447.1	435	solute carrier family 35 member C2
CAOG_02956	XP_004363795.2	385	solute carrier family 39

CAOG_04923	XP_004347674.2	499	solute carrier family 39
CAOG_07012	XP_004343736.1	754	solute carrier family 47
CAOG_04621	XP_004347368.2	815	sphingosine kinase SphK
CAOG_06412	XP_004345161.2	527	spinster like protein
CAOG_02700	XP_004349450.2	500	SPRY domain-containing protein
CAOG_07297	XP_004343156.1	527	squalene monooxygenase
CAOG_00304	XP_004365175.2	608	ST7 protein
CAOG_04243	XP_004348068.2	527	STAM binding protein
CAOG_09263			STE/STE7 protein kinase
CAOG_04722	XP_004347469.1	285	sterol desaturase
CAOG_05258	XP_004346943.1	363	sterol-4-alpha-carboxylate 3-dehydrogenase
CAOG_05285	XP_004346970.2	418	Sua5/YciO/YrdC/YwIc family protein
CAOG_06766	XP_004344387.1	542	sugar transporter
CAOG_05077	XP_004346762.1	740	syk-prov protein
CAOG_08354	XP_004340834.2	782	syk-prov protein
CAOG_07641	XP_004343515.1	334	syntaxin 5A
CAOG_05240	XP_004346925.1	289	syntaxin-6
CAOG_05526	XP_004346199.2	1260	T-box transcription factor TBX4
CAOG_01193	XP_004349690.1	878	TBC1 domain family protein
CAOG_05197	XP_004346882.1	235	testis enhanced transcript
CAOG_05847	XP_004345437.2	1008	testis-specific protein kinase 1
CAOG_02776	XP_004349529.2	260	TGF beta-inducible nuclear protein 1
CAOG_02459	XP_004349209.2	518	Tgs1 protein
CAOG_04267	XP_004348092.2	362	thiamin pyrophosphokinase 1
CAOG_00550	XP_004365421.1	538	ThiF family protein
CAOG_01239	XP_004349759.1	2246	thyroid hormone receptor interactor 12
CAOG_09258		670	TKL protein kinase
CAOG_09665			TKL protein kinase
CAOG_09920			TKL protein kinase
CAOG_05478	XP_004346151.2	830	tnf receptor-associated factor 7
CAOG_04891	XP_004347642.2	309	TPA regulated locus
CAOG_00603	XP_004365474.2	835	transcription factor 25
CAOG_04493	XP_004348321.2	236	transcription factor FET5
CAOG_05942	XP_004345532.2	385	translation factor pelota

CAOG_01960	XP_004364828.1	200	translation initiation factor SUI1
CAOG_03576	XP_004363304.1	111	translation initiation factor SUI1
CAOG_02863	XP_004363702.2	594	transmembrane 9 superfamily member 1
CAOG_00427	XP_004365298.2	425	transmembrane protein
CAOG_01822	XP_004364690.2	511	transmembrane protein
CAOG_00931	XP_004365802.2	278	transmembrane protein 41B
CAOG_03235	XP_004364074.2	325	Trip1-PA
CAOG_00199	XP_004365070.1	283	tRNA methyltransferase
CAOG_09256			tRNA-ALA
CAOG_02975	XP_004363814.1	741	tRNA-dihydrouridine synthase 3
CAOG_02807	XP_004348620.2	541	type I phosphodiesterase/nucleotide pyrophosphatase
CAOG_06007	XP_004345597.1	1378	tyrosine-protein kinase CSK
CAOG_08227	XP_004342482.1	1866	tyrosine-protein kinase CSK
CAOG_00226	XP_004365097.2	583	tyrosine-protein kinase Src42A
CAOG_01470	XP_004364338.2	797	tyrosine-protein kinase Srms
CAOG_05691	XP_004346364.2	830	tyrosine-protein kinase Srms
CAOG_04919	XP_004347670.2	740	tyrosine-protein kinase transforming protein Abl
CAOG_04506	XP_004348334.2	810	tyrosine-protein kinase transforming protein Fes
CAOG_07889	XP_004342974.1	730	tyrosine-protein phosphatase non-receptor type 11
CAOG_05870	XP_004345460.2	655	ubiquitin carboxyl-terminal hydrolase
CAOG_03638	XP_004363366.1	247	ubiquitin conjugating enzyme protein 26
CAOG_05967	XP_004345557.2	781	ubiquitin ligase E3A isoform 1
CAOG_07320	XP_004343179.1	451	ubiquitin specific peptidase 12
CAOG_05894	XP_004345484.1	1177	ubiquitin specific protease 41
CAOG_07247	XP_004343106.1	332	ubiquitin thioesterase OTU1
CAOG_08184	XP_004342785.1	412	ubiquitin-activating enzyme 5
CAOG_07990	XP_004342591.2	419	ubiquitin-conjugating enzyme E2
CAOG_06956	XP_004343680.2	289	ubiquitin-conjugating enzyme E2 R1
CAOG_04633	XP_004347380.1	156	ubiquitin-conjugating enzyme E2I
CAOG_02291	XP_004349041.2	152	ubiquitin-conjugating enzyme E2N
CAOG_07576	XP_004343450.1	151	ubiquitin-conjugating enzyme HR6A
CAOG_06327	XP_004345076.1	176	ubiquitin-fold modifier-conjugating enzyme 1

CAOG_04078	XP_004347903.1	184	ubiquitin-protein ligase
CAOG_03163	XP_004364002.1	444	UDP-N-acetylglucosamine 2-epimerase
CAOG_02196	XP_004348946.2	1261	Usp8 protein
CAOG_03371	XP_004364210.1	163	V-type ATPase
CAOG_02575	XP_004349325.2	620	V-type proton ATPase catalytic subunit A
CAOG_04328	XP_004348156.1	116	vacuolar ATP synthase subunit G
CAOG_03924	XP_004363652.1	428	vacuolar protein sorting-associated protein
CAOG_02427	XP_004349177.1	236	vacuolar protein sorting-associated protein 28
CAOG_01737	XP_004364605.1	1014	vacuolar proton ATPase
CAOG_01054	XP_004365925.1	110	VMA21 domain-containing protein
CAOG_05253	XP_004346938.1	516	voltage-dependent potassium channel protein
CAOG_04223	XP_004348048.1	315	WD repeat domain-containing protein
CAOG_08061	XP_004342662.1	1018	WD repeat domain-containing protein
CAOG_04821	XP_004347572.1	671	WD-repeat protein
CAOG_05929	XP_004345519.1	586	WD-repeat protein 50
CAOG_02049	XP_004348799.1	273	WD40 repeat-containing protein
CAOG_01873	XP_004364741.2	936	xenotropic and polytropic murine leukemia virus receptor xpr1
CAOG_00279	XP_004365150.1	432	XPA binding protein 1
CAOG_03653	XP_004363381.2	640	ZC3H3 protein
CAOG_06521	XP_004345270.1	1579	zinc finger CCCH domain-containing protein 37
CAOG_05935	XP_004345525.2	633	zinc finger family protein
CAOG_01096	XP_004365967.2	450	zinc finger protein
CAOG_07032	XP_004343756.2	479	zinc finger protein
CAOG_03804	XP_004363532.1	505	zinc finger RAD18 domain-containing protein C1, partial
CAOG_02096	XP_004348846.1	214	zinc-binding dehydrogenase
CAOG_01546	XP_004364414.2	1551	zipper protein kinase

The 389 annotated genes from the 1348 genes and their corresponding UniProt IDs:

Gene Name	UniProt ID
1-acyl-sn-glycerol-3-phosphate acyltransferase	GPAT3
2-acylglycerol O-acyltransferase 2-A	MOGAT1
24-dehydrocholesterol reductase	DHCR24
2OG-Fe(II) oxygenase	I14I2086
3-oxoacyl-[acyl-carrier-protein] synthase 2, variant 2	KASII
40S ribosomal protein S21	RPS21
40S ribosomal protein S29	RPS29
5-methyltetrahydrofolate-homocysteine methyltransferase reductase	MTR
5'-AMP-activated protein kinase catalytic subunit alpha-1	PRKAA1
5'-AMP-activated protein kinase catalytic subunit alpha-2	PRKAA2
6-phosphofructokinase	PFKL
AAA family ATPase	ATAD2
ABC transporter	ABCA2
ABC transporter	ABCA2
Abca3 protein	ABCA3
ADP-ribosyl cyclase	CD38
ADP-ribosylation factor-like protein 8B	ARL8B
alkane-1 monooxygenase	alkB
amidophosphoribosyltransferase	PPAT
ankyrin repeat domain-containing protein 13	ANKRD13B
aquaporin 10	AQP10
archipelago beta form	LW93_5577
ataxin-3	ATXN3
AtMMH-1	FPG1
ATP synthase subunit delta	ATP5F1D
ATP-binding cassette	ABCG2
ATP-binding cassette sub-family A member 1	ABCA2
ATP-binding cassette sub-family B member 10	ABCB9
ATP-binding cassette sub-family B member 7	ABCB7
ATP-binding cassette sub-family E member 1	ABCE1
ATP-binding cassette transporter sub-family G member 2c	ABCB9
ATP-dependent Clp protease proteolytic subunit	CLPP
ATP-dependent DNA helicase	DNA2
ATP-dependent protease La	LONP1

ATP-dependent protease La	LONP1
ATPase	
ATPase	
ATPase type 13A	ATP13A1
Atypical/RIO/RIO1 protein kinase, variant 4	PHSY_000610
B3GNTL1 protein	B3GNTL1
beclin	BECN1
betaine aldehyde dehydrogenase	ALDH7A1
branched-chain amino acid aminotransferase	BCAT2
bystin	BYSL
C-4 methylsterol oxidase	MSMO1
Ca ²⁺ transporting ATPase, plasma membrane	ACA8
calcium-dependent cysteine protease	CAPNS1
calcium-translocating P-type ATPase	TELCIR_04040
CAMK/CAMKL/BRSK protein kinase	SPPG_02241
CAMK/CAMKL/PASK protein kinase	PGTG_14851
cAMP-dependent protein kinase catalytic subunit	PRKX
carbonate dehydratase	CA1
carnitine/acylcarnitine carrier protein	SLC25A20
carrier protein	
carrier protein	
carrier protein	
casein kinase	CSNK1D
cathepsin C	CTSC
cathepsin L2	CTSV
ccr4-associated factor	CNOT7
cell motility mediator	MEMO1
ceramide kinase	CERK
cereblon	CRBN
cholesterol acyltransferase	LCAT
cholinesterase	BCHE
chorismate mutase/prephenate dehydratase	pheA
chromatin modifying protein 1b	CHMP1B
chromatin modifying protein 1b	CHMP1B
chromatin modifying protein 2a	CHMP2A
CHY zinc finger family protein	RZPF34
COBW domain-containing protein	CBWD1

collagen type IV alpha-3-binding protein	CERT1
copine-9	CPNE9
copper-transporting ATPase	ATP7B
coproporphyrinogen oxidase	hemN
cre	AHK4
crk-like protein	CRKL
Cullin 4	CUL4A
cyclin-box carrying protein isoform	CAOG_001396
cystathionine beta-lyase	<u>metC</u>
cystathionine gamma-synthase	CGS1
cysteine desulfurylase	102586162
cytochrome b5 reductase 4	CYB5R4
cytochrome P450	CYP2C8
cytochrome P450	CYP2C8
cytosolic Fe-S cluster assembling factor nbp35	K457DRAFT_140382
Dcun1d3 protein	DCUN1D3
DDHD domain containing 2	DDHD2
DEAD box ATP-dependent RNA helicase	DDX54
DEAH helicase isoform 6	DHX9
decarboxylase	GAD1
delta-aminolevulinic acid dehydratase	ALAD
deoxyhypusine synthase	DHPS
deoxyribodipyrimidine photo-lyase	PHR1
derlin-2 like protein	DERL2
DHHC zinc finger domain-containing protein	ZDHHC8
diacylglycerol acyltransferase type 2A	DGAT2
diacylglycerol kinase 1	DGKA
diaphorase	NQO1
dihydroderamide delta-4 desaturase	CAOG_008167
diphosphomevalonate decarboxylase	MVD
diphthamide biosynthesis protein	DPH1
DNA-directed RNA polymerase II polypeptide	POLR2B
dolichyl-phosphate-mannose-protein mannosyltransferase	POMT1
DREV methyltransferase	METTTL9
dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1B	
E1A binding protein p300	EP300
E3 ubiquitin-protein ligase NEDD4	NEDD4

efflux ABC transporter	ABCA1
ephrin type-A receptor 4a	epha4a
eukaryotic peptide chain release factor subunit 1	GSPT1
eukaryotic translation initiation factor 2-alpha kinase	EIF2AK1
eukaryotic translation initiation factor 2-alpha kinase 1	EIF2AK1
F-box/WD repeat-containing protein pof1	pof1
FAM72A protein	FAM72A
farnesyl-diphosphate farnesyltransferase 1	FDFT1
fatty acid elongase	ELOVL1
ferredoxin 1-like protein	FDX2
ferredoxin 1-like protein	FDX2
ferrous ion membrane transporter DMT1	SLC11A2
G protein-coupled receptor	GPBR1
gamma-butyrobetaine dioxygenase	BBOX1
gamma-butyrobetaine dioxygenase	BBOX1
general stress protein	mgsR
geranylgeranyl pyrophosphate synthetase	GGPS1
glucosamine-6-phosphate N-acetyltransferase	GNPNAT1
glutamate receptor Gr2	Gria2
glutaredoxin-3	GLRX3
glutathione S-transferase	GSTO1
glycerol-3-phosphate acyltransferase	GPAT3
glycine hydroxymethyltransferase	SHMT2
glycosyltransferase-like protein LARGE1	LARGE1
gphn protein	GPHN
growth hormone-regulated TBC protein 1	GRTP1
growth-inhibiting protein 1	ING1
GTP binding protein 4	GTPBP1
GTP-binding protein yptV3	YPTV3
GTPase	KRAS
guanine nucleotide-exchange protein	ARFGEF1
heat shock factor 2	HSF2
HIT domain-containing protein	ZNHIT1
HLA-B associated transcript 5	BAG6
hsp70-like protein	HSPBP1
hydroxymethylglutaryl-CoA synthase 1	HMGCS1
hydroxymethylglutaryl-coenzyme A reductase	lovC

hydroxysteroid dehydrogenase 7	HSD17B10
import inner membrane translocase subunit Tim22	TIMM22
inosine monophosphate dehydrogenase 2	IMPDH2
insect-derived growth factor-B-like protein	
integral membrane protein	GPR180
iron-sulfur protein I	NUBPL
iron/zinc purple acid phosphatase-like protein	papI
ketohexokinase	KHK
krox-20	EGR2
krueppel-like factor 16	KLF16
L-gulono-gamma-lactone oxidase	GULO
L-isoaspartyl protein carboxyl methyltransferase	PCMT1
leucine-rich repeat containing protein	LRRC1
leucine-rich repeat-containing protein 28	LRRC28
leucine-rich repeat-containing protein 69	LRRC69
lipase, family member N	LIPN
long-chain-fatty-acid-CoA ligase ACSBG2	ACSBG2
lss protein	LSS
lung seven transmembrane receptor	GPR107
LYR domain-containing protein-containing protein 5	ETFRF1
lysophosphatidylcholine acyltransferase 2-B	LPCAT2
lysophosphatidylcholine acyltransferase 3	LPCAT3
lysosomal alpha-N-acetyl glucosaminidase	
lysozyme	SPACA3
macrophage erythroblast attacher protein	MAEA
MafI-pending-prov protein	MAF1
max-like protein X	MLX
metal-binding domain in RNase L inhibitor	RNASEL
metallo-beta-lactamase	blaNDM-1
methyltransferase	DNMT1
microtubule associated protein	MAP1S
mitogen-activated protein kinase kinase kinase 1	MAP3K1
mitogen-activated protein kinase kinase kinase 1	MAP3K1
mitogen-activated protein kinase kinase kinase 3	MAP3K3
mitogen-activated protein kinase kinase kinase kinase	SLK
molybdenum cofactor synthesis 1 isoform 1	MOCS1
molybdenum cofactor synthesis protein 2 large subunit	MOCS2

molybdopterin synthase sulfurylase	MOCS3
Mrto4 protein	MRT04
methfr-prov protein	MTHFR
multi-sensor hybrid histidine kinase	Psta_0819
myosin IXA	MYO9A
N-acetylglucosamine-1-phosphate transferase	DPAGT1
n-acetylglucosamine-6-phosphate deacetylase	AMDHD2
NAD-dependent deacetylase sirtuin 7	SIRT7
NAD-dependent epimerase/dehydratase	azf
NADH dehydrogenase 1 beta subcomplex subunit 9	NDUFB9
NADPH-cytochrome P450 reductase	POR
Nat9 protein	NAT9
natural killer cell-specific antigen KLIP1	YIPF3
neuropathy target esterase	PNPLA6
neutral sphingomyelinase activation associated factor	NSMAF
non-receptor protein kinase	TNK1
non-receptor protein kinase, partial	TNK1
NPR2 protein	NPR2
NTRK3 protein	NTRK3
nuclear elongation and deformation protein 1	NED1
nuclease Le3	CAOG_007127
nucleic acid binding protein	PCBP1
nucleotide-binding protein 2	RACK1
oxidation resistance protein 1	OXR1
patatin-like phospholipase domain-containing protein 7	PNPLA7
peptidase M19	A1D17_16880
peptidase S8 and S53	<i>Peptidase_S8/S53_dom</i>
peptide methionine sulfoxide reductase	MSRA
periodic tryptophan protein	PWP1
phenylalanyl-tRNA synthetase	FARS2
phosphate transporter	SLC20A2
phosphatidylethanolamine N-methyltransferase	PEMT
phosphatidylinositol glycan anchor biosynthesis	PIGU
phospho-2-dehydro-3-deoxyheptonate aldolase	DHS1
phosphodiesterase 1B	PDE1B
phosphodiesterase 2A	PDE2A
phosphoinositide-3-kinase	PIK3AP1

phospholipid transfer protein	PLTP
phosphomannomutase	PMM1
phosphopantothenoylcysteine decarboxylase	PPCDC
phosphopantothenoylcysteine synthetase	PPCS
phosphoribosylaminoimidazole-succinocarboxamide synthase	PAICS
Pik4cb protein	PI4KB
pirin domain-containing protein domain-containing protein	PYDC1
polyadenylate-binding protein	PABPN1
potassium channel tetramerisation domain-containing protein	KCTD1
prefoldin	PFDN1
programmed cell death protein	PDCD1
protease m1 zinc metalloprotease	APM1
protein arginine methyltransferase 3	PRMT3
protein disulfide-isomerase ERp60	PDIA3
protein kinase	CHK1
protein kinase	CHK1
protein-S-isoprenylcysteine O-methyltransferase	ICMT
protein-tyrosine-phosphatase	PTPRU
pyridoxamine 5\\'-phosphate oxidase	PNPO
queuine tRNA-ribosyltransferase	QTRT1
radical S-adenosyl methionine domain-containing protein 1	RSAD2
radixin isoform a	SLC9A3R1
rap guanine nucleotide exchange factor GEF2	RAPGEF6
ras family protein	MRAS
ras guanine nucleotide exchange factor, partial	RASGRF2
Ras like enriched	<u>RHEB</u>
Ras-like protein Rab-21	RAB21
ras-specific guanine nucleotide-releasing factor RalGPS1	RALGPS1
Rbm22 protein	RBM22
receptor accessory protein 5	<u>REEP5</u>
response regulator receiver	Dace_1490
response regulator receiver domain-containing protein	BG023_111682
rho GTPase activating protein	ARHGAP1
rho GTPase-activating protein	ARHGAP1
rho GTPase-activating protein 8	ARHGAP8
ribonuclease P/MRP 30 subunit	RPP30
ribosomal protein L29e	RPL29

ribosomal RNA large subunit methyltransferase J	rlmJ
ring finger protein, partial	RNF11
RIO kinase 3	RIOK3
RNA binding domain-containing protein	PN01
RNA-binding protein PNO1	PN01
saccharopine dehydrogenase	AASS
SapB domain-containing protein	spp-1
SEC14-like protein	SEC14L1
selenocysteine-specific elongation factor	EEFSEC
sensor protein GacS	gacS
sentrin/SUMO-specific protease 15	SEN1
sepiapterin reductase	SPR
Ser/Thr kinase	CSNK1A1
serine/threonine protein kinase	STK32C
serine/threonine protein kinase	STK16
serine/threonine protein kinase 15	AURKA
serine/threonine-protein kinase gad8	gad8
serine/threonine-protein kinase ppk15	ppk15
serine/threonine-protein kinase SAPK2	MAPK11
serine/threonine-specific protein kinase NAK	AAK1
serologically defined colon cancer antigen 1	ENTR1
serum/glucocorticoid regulated kinase 2	SGK2
sestrin 1	SESN1
SFRS protein kinase 1	SRPK1
Sh3yl1-prov protein	SH3YL1
small GTP-binding protein Rab4	RAB4A
small HspC2 heat shock protein	HSP90AB1
SMYD5 protein	SMYD5
SNARE Ykt6	YKT6
snrk protein	SNRK
snRNP core protein SMX5	LSM2
snurportin-1	SNUPN
sodium/calcium exchanger protein	SLC8B1
soluble NSF attachment protein	NAPA
solute carrier family 24	SLC25A24
solute carrier family 25	SLC25A25
solute carrier family 25	SLC25A25

solute carrier family 25	SLC25A25
solute carrier family 35	SLC35B1
solute carrier family 35 member C2	SLC35C2
solute carrier family 39	SLC39A7
solute carrier family 39	SLC39A7
solute carrier family 47	SLC47A1
sphingosine kinase SphK	SPHK1
spinster like protein	SPNS1
SPRY domain-containing protein	FSD2
squalene monooxygenase	SQLE
ST7 protein	ST7
STAM binding protein	STAMPB
STE/STE7 protein kinase	
sterol desaturase	SC5D
sterol-4-alpha-carboxylate 3-dehydrogenase	NSDHL
Sua5/YciO/YrdC/YwlC family protein	US98_C0002G0004
sugar transporter	ESL1
syk-prov protein	SYK
syk-prov protein	SYK
syntaxin 5A	STX5
syntaxin-6	STX6
T-box transcription factor TBX4	TBX4
TBC1 domain family protein	TBC1D1
testis enhanced transcript	
testis-specific protein kinase 1	TSSK1B
TGF beta-inducible nuclear protein 1	TINP1
Tgs1 protein	TGS1
thiamin pyrophosphokinase 1	TPK1
ThiF family protein	thiF
thyroid hormone receptor interactor 12	TRIP12
TKL protein kinase	TNNI3K
TKL protein kinase	TNNI3K
TKL protein kinase	TNNI3K
tnf receptor-associated factor 7	TRAF7
TPA regulated locus	PLAT
transcription factor 25	TCF25
transcription factor FET5	<u>FET5</u>

translation factor pelota	ETF1
translation initiation factor SUI1	EIF1
translation initiation factor SUI1	EIF1
transmembrane 9 superfamily member 1	TM9SF1
transmembrane protein	
transmembrane protein	
transmembrane protein 41B	TMEM41B
Trip1-PA	EIF3I
tRNA methyltransferase	METTL1
tRNA-ALA	DUS3L
tRNA-dihydrouridine synthase 3	
type I phosphodiesterase/nucleotide pyrophosphatase	phoD
tyrosine-protein kinase CSK	CSK
tyrosine-protein kinase CSK	CSK
tyrosine-protein kinase Src42A	SRC
tyrosine-protein kinase Srms	SRMS
tyrosine-protein kinase Srms	SRMS
tyrosine-protein kinase transforming protein Abl	ABL1
tyrosine-protein kinase transforming protein Fes	ABL1
tyrosine-protein phosphatase non-receptor type 11	PTPN11
ubiquitin carboxyl-terminal hydrolase	UCHL1
ubiquitin conjugating enzyme protein 26	UBE2D2
ubiquitin ligase E3A isoform 1	UBE3A
ubiquitin specific peptidase 12	USP12
ubiquitin specific protease 41	<u>USP41</u>
ubiquitin thioesterase OTU1	YOD1
ubiquitin-activating enzyme 5	UBA5
ubiquitin-conjugating enzyme E2	CDC34
ubiquitin-conjugating enzyme E2 R1	CDC34
ubiquitin-conjugating enzyme E2I	UBE2I
ubiquitin-conjugating enzyme E2N	UBE2N
ubiquitin-conjugating enzyme HR6A	HR6A
ubiquitin-fold modifier-conjugating enzyme 1	UFC1
ubiquitin-protein ligase	RBX1
UDP-N-acetylglucosamine 2-epimerase	GNE
Usp8 protein	USP8
V-type ATPase	ATP6V0D1

V-type proton ATPase catalytic subunit A	ATP6V0A1
vacuolar ATP synthase subunit G	ATP6V1G1
vacuolar protein sorting-associated protein	VPS13A
vacuolar protein sorting-associated protein 28	VPS28
vacuolar proton ATPase	ATP6V0A1
VMA21 domain-containing protein	VMA21
voltage-dependent potassium channel protein	CACNA1D
WD repeat domain-containing protein	WDFY1
WD repeat domain-containing protein	WDFY1
WD-repeat protein	
WD-repeat protein 50	<u>WDR77</u>
WD40 repeat-containing protein	SMU1
xenotropic and polytropic murine leukemia virus receptor xpr1	XPR1
XPA binding protein 1	GPN1
ZC3H3 protein	ZC3H3
zinc finger CCCH domain-containing protein 37	HUA1
zinc finger family protein	
zinc finger protein	
zinc finger protein	
zinc finger RAD18 domain-containing protein C1, partial	RAD18
zinc-binding dehydrogenase	PTGR2
zipper protein kinase	MAP3K12

All 1348 genes that are expressed in the same way as NF- κ B across the three life stages of *Capsaspora*. The 389 annotated genes from the 1348 genes and their corresponding UniProt IDs are shown as well.

Table 5.3 *Capsaspora* gene homologs related to development and immunity

<u>Developmental Process Genes in <i>Capsaspora</i> (47 genes)</u>	<u>Immune System Process Genes in <i>Capsaspora</i> (26 genes)</u>	<u>Total Number of Unique Genes in Developmental and Immune System Process in <i>Capsaspora</i></u>
actin-like protein 2	ankyrin repeat domain- containing protein	66 genes
adenylyl cyclase-associated protein	ADP-ribosyl cyclase	
ankyrin repeat domain- containing protein	ATP-binding cassette	
Coronin	ATP-dependent RNA helicase	
cyclin-dependent kinase	cactin	
cytoplasmic FMR1-interacting protein 1	cathepsin L2	
E3 ubiquitin-protein ligase NEDD4	heat shock protein	
ephrin type-B receptor 3	hydrolase	
flightless-1	integral membrane protein	
G protein-coupled receptor	mitogen-activated protein kinase kinase kinase (SLK)	
GTP-binding protein	peroxiredoxin-2	
GTPase	phospholipase A2	
guanine nucleotide exchange factor	programmed cell death protein	
integral membrane protein	protein kinase C	
kinesin-1	protein tyrosine kinase	
leucine-rich repeat-containing protein	ring finger protein	
microtubule-associated protein	RNA helicase	
mitogen-activated protein kinase kinase kinase (SLK)	RNA-binding protein	
myosin-VI	SH3 domain-containing protein	
myosin-VIIa	transcriptional regulator	
Myotubularin	tyrosine-protein kinase CSK	
nucleosome assembly protein	tyrosine-protein kinase ITK/TSK	

nucleotide pyrophosphatase/phosphodiesterase	tyrosine-protein kinase Src42A
Paxillin	tyrosine-protein kinase Srms
phosphoenolpyruvate carboxykinase	tyrosine-protein kinase SYK
phosphoglucomutase	zinc finger protein
protein phosphatase 2	
rho GTPase-activating protein	
serine/threonine-protein kinase 4	
SPRY domain-containing protein	
T-box transcription factor TBX4	
Thrombospondin	
transcriptional regulator	
tubulin-beta chain	
twintin-1	
Tyrosine-protein kinase Fyn (SLK)	
tyrosine-protein kinase ITK/TSK	
tyrosine-protein kinase Src42A	
tyrosine-protein kinase Srms	
tyrosine-protein kinase SYK	
tyrosine-protein kinase transmembrane receptor Ror	
ubiquitin-protein ligase	
vasodilator-stimulated phosphoprotein	
vesicle-associated membrane protein	
Villin	
WD-repeat protein 50	
NT-3 growth factor receptor NTRK3	

All *Capsaspora* gene homologs that are annotated within the genome present in the Developmental and Immune system process GO categories. The genes highlighted in red are present in both lists.

Table 5.4 JASPER NF- κ B motif search results sites combined

Sequence Name	Strand	Start	End
CAOG_07458	-	373	382
CAOG_00002	-	233	243
CAOG_00199	-	334	344
CAOG_01030	-	96	106
CAOG_01542	+	45	55
CAOG_01771	+	473	482
CAOG_02202	-	430	440
CAOG_02706	-	111	121
CAOG_02716	+	402	412
CAOG_03349	-	81	91
CAOG_03422	+	16	25
CAOG_04633	-	29	39
CAOG_05576	+	293	302
CAOG_06780	+	317	327
CAOG_08652	-	75	85
CAOG_09305	+	270	279
CAOG_09674	+	376	385
CAOG_09720	-	283	292
CAOG_09786	-	195	204
CAOG_09996	-	98	107
CAOG_00003	+	410	420
CAOG_00061	-	204	213
CAOG_00172	+	81	90
CAOG_00207	-	202	211
CAOG_00226	-	220	232
CAOG_00281	-	147	159
CAOG_00387	-	171	181
CAOG_00420	+	266	275
CAOG_00476	+	357	366
CAOG_00499	+	98	108
CAOG_00500	-	43	53
CAOG_00526	+	68	78
CAOG_00542	+	409	418
CAOG_00574	+	92	102
CAOG_00710	-	83	92

CAOG_00758	-	156	165
CAOG_00760	+	135	145
CAOG_00982	-	111	121
CAOG_01083	-	327	337
CAOG_01134	-	191	200
CAOG_01284	+	194	203
CAOG_01365	-	214	223
CAOG_01398	+	112	122
CAOG_01458	+	342	352
CAOG_01736	-	371	381
CAOG_01767	-	233	243
CAOG_01886	-	23	33
CAOG_01906	-	460	470
CAOG_01960	+	42	54
CAOG_01971	+	385	395
CAOG_02015	-	21	30
CAOG_02032	-	100	112
CAOG_02148	-	427	437
CAOG_02167	-	40	50
CAOG_02196	+	141	151
CAOG_02233	+	483	495
CAOG_02239	-	226	236
CAOG_02290	-	195	205
CAOG_02369	+	17	26
CAOG_02372	-	83	93
CAOG_02408	+	285	294
CAOG_02441	-	226	236
CAOG_02442	+	347	357
CAOG_02575	-	77	87
CAOG_02668	-	108	118
CAOG_02709	+	210	220
CAOG_02749	-	461	470
CAOG_02772	+	242	252
CAOG_02776	-	123	132
CAOG_03051	-	181	191
CAOG_03135	-	168	177
CAOG_03168	-	281	293
CAOG_03281	-	154	163

CAOG_03315	+	487	497
CAOG_03476	+	268	278
CAOG_03528	+	130	140
CAOG_03609	-	79	89
CAOG_03621	-	79	89
CAOG_03679	+	414	423
CAOG_03774	-	195	204
CAOG_03781	+	9	19
CAOG_03987	-	30	40
CAOG_04035	+	67	77
CAOG_04056	+	317	326
CAOG_04062	-	252	261
CAOG_04153	-	324	333
CAOG_04223	-	408	418
CAOG_04295	-	164	174
CAOG_04329	+	61	71
CAOG_04391	+	16	26
CAOG_04516	+	84	93
CAOG_04563	-	69	79
CAOG_04594	+	65	77
CAOG_04642	-	282	291
CAOG_04819	+	323	333
CAOG_04849	-	65	75
CAOG_04882	-	155	165
CAOG_04892	-	274	284
CAOG_04908	+	425	435
CAOG_04992	+	436	448
CAOG_05007	-	466	476
CAOG_05042	-	175	185
CAOG_05075	+	89	99
CAOG_05168	-	160	172
CAOG_05184	+	103	113
CAOG_05190	+	109	118
CAOG_05197	-	161	171
CAOG_05238	+	16	25
CAOG_05251	-	314	324
CAOG_05258	+	202	212
CAOG_05289	-	102	114

CAOG_05336	-	458	468
CAOG_05381	-	95	104
CAOG_05508	-	417	426
CAOG_05528	-	337	347
CAOG_05649	+	67	76
CAOG_05652	+	127	136
CAOG_05719	-	80	90
CAOG_05726	+	323	333
CAOG_06135	+	91	100
CAOG_06137	-	143	153
CAOG_06290	+	21	30
CAOG_06339	-	229	238
CAOG_06392	+	180	189
CAOG_06415	-	89	99
CAOG_06578	-	456	465
CAOG_06657	+	59	69
CAOG_06709	+	146	156
CAOG_06774	-	384	393
CAOG_06776	-	146	156
CAOG_06810	-	357	367
CAOG_06815	+	488	500
CAOG_06865	-	232	242
CAOG_06896	-	378	390
CAOG_06938	-	67	76
CAOG_06970	+	38	48
CAOG_06975	+	85	95
CAOG_07084	+	362	372
CAOG_07134	+	222	232
CAOG_07207	-	46	56
CAOG_07214	-	411	420
CAOG_07348	-	362	372
CAOG_07365	-	289	299
CAOG_07463	+	320	330
CAOG_07597	+	440	450
CAOG_08008	+	413	423
CAOG_08035	+	39	48
CAOG_08055	+	181	191
CAOG_08163	-	330	340

CAOG_08167	-	30	40
CAOG_08182	+	159	168
CAOG_08185	+	450	460
CAOG_08270	+	154	164
CAOG_08353	-	279	289
CAOG_08354	-	269	279
CAOG_08360	-	55	65
CAOG_08384	-	434	444
CAOG_08463	+	459	468
CAOG_08550	-	451	461
CAOG_08705	+	31	41
CAOG_08859	-	447	457
CAOG_08864	-	427	436
CAOG_08909	+	102	112
CAOG_08935	+	126	138
CAOG_09048	+	322	331
CAOG_09134	+	9	21
CAOG_09172	-	448	457
CAOG_09187	-	60	70
CAOG_09324	+	49	58
CAOG_09343	-	225	235
CAOG_09359	-	316	325
CAOG_09379	-	278	290
CAOG_09595	+	455	465
CAOG_09692	-	415	425
CAOG_09703	-	257	266
CAOG_09709	+	288	298
CAOG_09738	+	104	113
CAOG_09828	+	110	120
CAOG_09865	-	207	217
CAOG_09878	+	406	416
CAOG_09887	-	345	355
CAOG_09902	-	128	138
CAOG_09920	-	352	362
CAOG_09939	-	486	496
CAOG_09943	+	296	306
CAOG_09964	+	366	376
CAOG_09976	-	183	193

CAOG_10017	-	156	165
CAOG_10024	-	211	221
CAOG_10056	+	475	485
CAOG_10074	-	45	55
CAOG_10108	-	479	489
CAOG_00002	+	323	333
CAOG_00061	-	203	213
CAOG_00061	-	203	213
CAOG_00199	-	333	343
CAOG_00199	-	62	72
CAOG_00226	+	220	232
CAOG_00281	+	147	159
CAOG_00420	-	266	275
CAOG_00420	-	266	276
CAOG_00420	+	266	276
CAOG_00420	-	265	275
CAOG_00476	+	357	367
CAOG_00476	+	356	366
CAOG_00476	+	357	367
CAOG_00526	+	67	77
CAOG_00542	+	409	419
CAOG_00542	+	408	418
CAOG_00574	-	92	102
CAOG_00574	-	91	103
CAOG_00574	-	90	102
CAOG_00574	+	91	103
CAOG_00574	+	90	102
CAOG_00574	+	90	102
CAOG_00574	-	90	102
CAOG_00710	-	82	92
CAOG_00710	+	83	93
CAOG_00710	-	82	94
CAOG_01030	+	123	133
CAOG_01030	+	96	106
CAOG_01030	-	123	133
CAOG_01030	-	122	134
CAOG_01030	+	95	107
CAOG_01030	+	122	134

CAOG_01030	-	95	107
CAOG_01030	-	122	134
CAOG_01030	+	122	134
CAOG_01030	-	95	107
CAOG_01030	+	95	107
CAOG_01134	-	190	200
CAOG_01134	+	190	200
CAOG_01134	+	189	201
CAOG_01542	+	344	356
CAOG_01542	+	44	56
CAOG_01542	-	44	56
CAOG_01542	-	344	356
CAOG_01542	+	44	56
CAOG_01771	-	147	157
CAOG_01771	+	473	483
CAOG_02148	+	428	438
CAOG_02196	-	141	151
CAOG_02196	+	140	152
CAOG_02202	-	259	269
CAOG_02233	-	483	495
CAOG_02369	+	17	27
CAOG_02369	+	16	26
CAOG_02408	+	285	295
CAOG_02706	-	357	367
CAOG_02709	-	210	220
CAOG_02716	-	251	261
CAOG_03135	-	167	177
CAOG_03135	-	167	177
CAOG_03168	+	281	293
CAOG_03281	-	153	163
CAOG_03281	-	153	163
CAOG_03281	+	152	164
CAOG_03281	-	152	164
CAOG_03281	+	152	164
CAOG_03349	+	385	395
CAOG_03422	-	462	471
CAOG_03422	+	16	26
CAOG_03528	+	129	141

CAOG_03774	-	194	204
CAOG_03774	-	194	204
CAOG_04062	+	251	261
CAOG_04062	+	252	262
CAOG_04062	+	251	261
CAOG_04062	-	250	262
CAOG_04153	+	324	334
CAOG_04153	-	325	335
CAOG_04223	+	408	418
CAOG_04223	-	407	419
CAOG_04223	+	407	419
CAOG_04223	+	407	419
CAOG_04329	-	60	70
CAOG_04329	+	59	71
CAOG_04329	-	59	71
CAOG_04329	+	59	71
CAOG_04516	+	84	94
CAOG_04594	-	65	77
CAOG_04633	-	50	60
CAOG_04642	-	281	291
CAOG_04642	-	281	291
CAOG_04992	-	436	448
CAOG_04992	-	436	448
CAOG_04992	+	436	448
CAOG_05007	-	465	477
CAOG_05007	+	465	477
CAOG_05007	-	465	477
CAOG_05168	+	160	172
CAOG_05190	+	108	120
CAOG_05238	+	16	26
CAOG_05381	-	94	104
CAOG_05381	-	95	105
CAOG_05381	-	94	104
CAOG_05508	-	416	426
CAOG_05508	+	417	427
CAOG_05576	-	293	302
CAOG_05576	+	333	342
CAOG_05576	+	293	303

CAOG_05576	-	292	302
CAOG_05576	+	333	343
CAOG_05576	+	293	303
CAOG_05576	-	292	302
CAOG_05576	+	333	343
CAOG_05576	-	332	342
CAOG_05576	+	291	303
CAOG_05576	-	292	304
CAOG_05652	-	127	137
CAOG_05652	-	126	136
CAOG_05652	+	126	138
CAOG_05652	-	126	138
CAOG_06135	+	91	101
CAOG_06290	-	22	31
CAOG_06290	+	21	31
CAOG_06290	-	21	31
CAOG_06290	+	21	31
CAOG_06290	+	22	32
CAOG_06290	-	20	32
CAOG_06290	+	20	32
CAOG_06290	+	20	32
CAOG_06290	-	20	32
CAOG_06339	+	228	237
CAOG_06339	+	229	238
CAOG_06339	-	228	237
CAOG_06339	-	228	238
CAOG_06339	+	228	238
CAOG_06339	+	229	239
CAOG_06339	-	228	238
CAOG_06339	+	227	239
CAOG_06339	-	227	239
CAOG_06339	-	226	238
CAOG_06339	-	227	239
CAOG_06339	+	227	239
CAOG_06339	+	226	238
CAOG_06339	-	226	238
CAOG_06392	-	181	190
CAOG_06392	+	181	190

CAOG_06392	+	180	190
CAOG_06392	-	180	190
CAOG_06392	+	180	190
CAOG_06392	-	179	191
CAOG_06392	+	179	191
CAOG_06392	+	179	191
CAOG_06392	-	179	191
CAOG_06709	-	145	157
CAOG_06780	+	356	366
CAOG_06938	-	66	76
CAOG_06938	-	67	77
CAOG_07084	-	361	373
CAOG_07084	+	361	373
CAOG_07207	-	45	57
CAOG_07207	+	45	57
CAOG_07207	+	45	57
CAOG_07214	-	410	420
CAOG_07214	-	410	420
CAOG_07214	+	411	421
CAOG_07458	-	372	382
CAOG_07458	-	484	494
CAOG_07458	+	446	456
CAOG_08035	+	39	49
CAOG_08035	-	39	49
CAOG_08035	+	38	50
CAOG_08035	-	38	50
CAOG_08463	+	459	469
CAOG_08652	+	52	62
CAOG_08705	-	30	42
CAOG_08864	-	426	436
CAOG_08864	-	426	436
CAOG_08864	+	425	437
CAOG_08864	-	425	437
CAOG_08864	-	425	437
CAOG_08935	-	126	138
CAOG_08935	+	126	138
CAOG_08935	-	126	138
CAOG_09048	+	322	332

CAOG_09048	+	321	331
CAOG_09048	+	322	332
CAOG_09048	-	321	333
CAOG_09134	-	9	21
CAOG_09134	-	9	21
CAOG_09134	+	9	21
CAOG_09305	+	270	280
CAOG_09305	-	270	280
CAOG_09305	+	223	233
CAOG_09359	+	315	325
CAOG_09379	+	278	290
CAOG_09674	+	376	386
CAOG_09674	+	67	77
CAOG_09703	+	256	265
CAOG_09703	+	256	266
CAOG_09703	-	256	266
CAOG_09703	-	255	265
CAOG_09703	+	255	265
CAOG_09703	+	256	266
CAOG_09703	+	255	267
CAOG_09703	-	255	267
CAOG_09703	+	254	266
CAOG_09703	-	254	266
CAOG_09703	-	255	267
CAOG_09703	+	255	267
CAOG_09703	+	254	266
CAOG_09703	-	254	266
CAOG_09720	+	430	439
CAOG_09720	+	430	440
CAOG_09720	-	282	292
CAOG_09720	+	430	440
CAOG_09720	-	282	292
CAOG_09720	+	429	441
CAOG_09786	-	68	78
CAOG_09902	-	127	137
CAOG_09996	-	19	29
CAOG_10017	-	154	166

The results of the JASPER NF- κ B motif search of genes that are co-regulated with NF- κ B. These genes contain one, two or three κ B sites within their putative promoters. Each gene is separated by color and includes the base pair coordinates of the κ B site in the promoter.

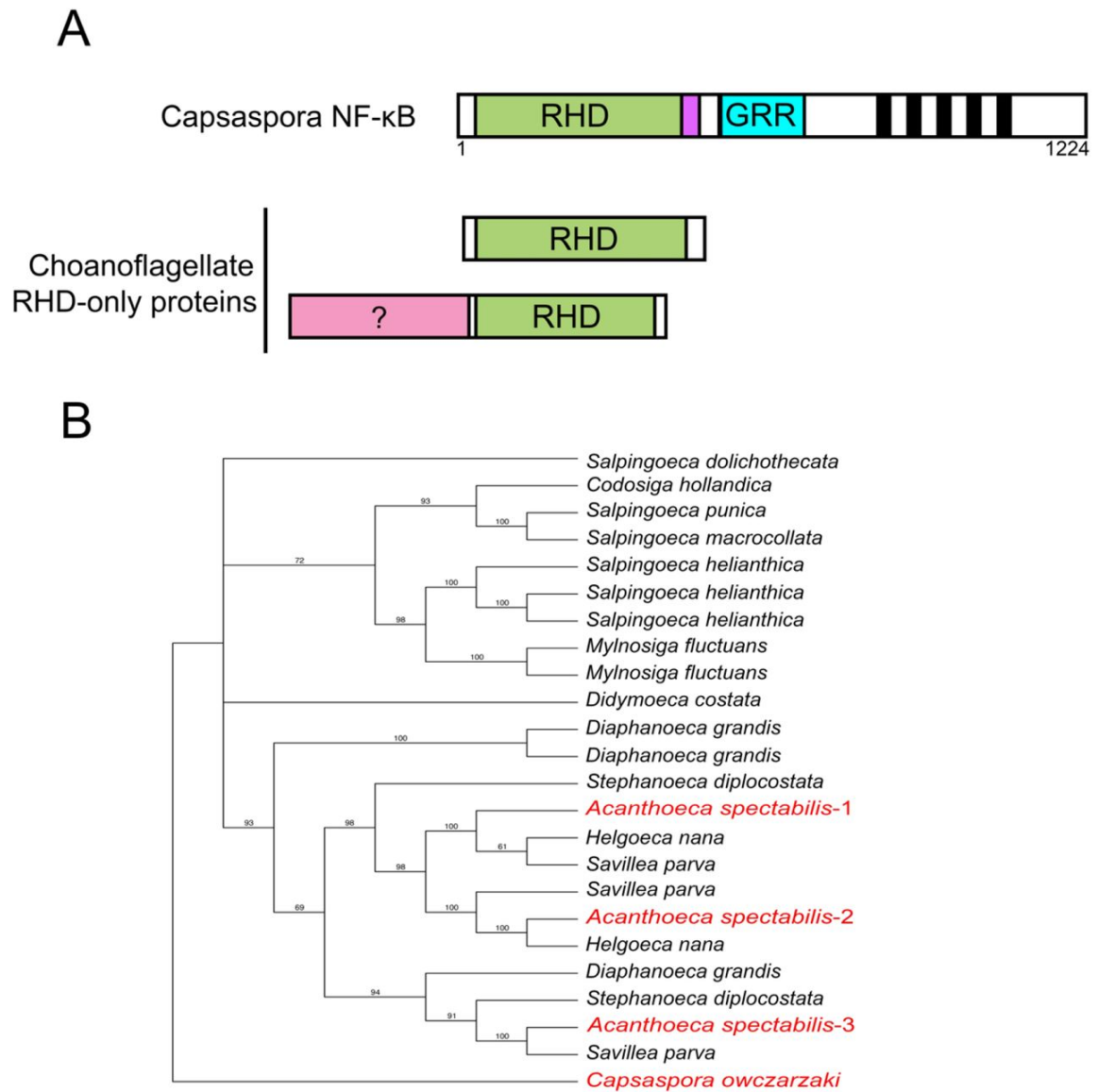


Figure 5.1 Protist NF- κ B proteins differ in domain structure and choanoflagellates show evidence of gene duplication.

(A) The general domain structures of both *Capsaspora* NF- κ B and the choanoflagellate RHD-only proteins. Green, RHD (Rel Homology Domain); Purple, nuclear localization sequence; Blue, GRR (glycine-rich region); Black bars, Ankyrin repeats; Pink, sequences in choanoflagellates that are not typically seen in other organisms. (B) A phylogenetic

estimation using maximum likelihood (bootstrapped 1000 times) of choanoflagellate and *Capsaspora* RHDs. The NF- κ B proteins used in these studies are highlighted in red.

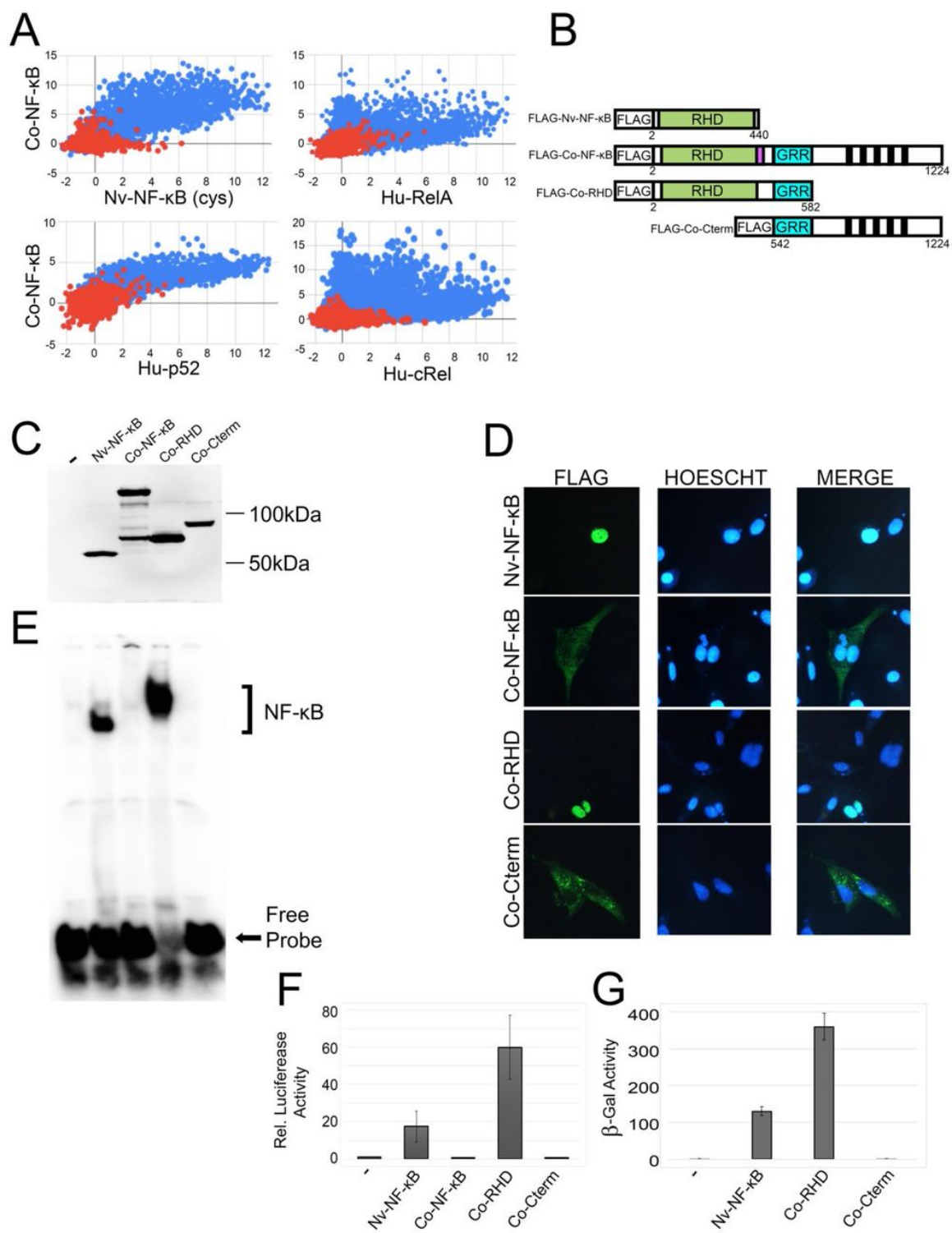


Figure 5.2 DNA-binding and transcriptional activation activity of Co-NF- κ B.

(A) Protein binding microarray (PBM) DNA-binding profiles of Co-NF- κ B as compared to *Nematostella vectensis* (Nv) NF- κ B cysteine (cys) allele (top, left), human (Hu) RelA (top, right), human p52 (bottom, left), and human cRel (bottom, right). The axes are z-scores. Red dots represent random background sequences, and blue dots represent NF- κ B binding sites. (B) FLAG-tagged expression vectors used in these experiments. From top to bottom, the drawings depict the naturally shortened Nv-NF- κ B, the full-length Co-NF- κ B protein, an N-terminal-only mutant containing the RHD and GRR (Co-RHD), and a C-terminal-only mutant containing the ANK repeats and other C-terminal sequences (Co-Cterm). (C) Anti-FLAG Western blot of lysates of HEK 293T cells transfected with the indicated expression vectors. (D) Indirect immunofluorescence of DF-1 chicken fibroblast cells transfected with the indicated expression vectors. Cells were then stained with anti-FLAG antiserum (left panels) and HOESCHT (middle panels), and then MERGED on the right panels. (E) A κ B-site electromobility shift assay (EMSA) using each of the indicated lysates from (C). The NF- κ B complexes and free probe are indicated by arrows. (F) A κ B-site luciferase reporter gene assay was performed with the indicated proteins in HEK 293 cells. Luciferase activity is relative (Rel.) to that seen with the empty vector control (1.0), and values are averages of three assays performed with triplicate samples with standard error. (G) A GAL4-site *LacZ* reporter gene assay was performed in yeast Y190 cells. Values are average units from seven assays performed with duplicate samples with standard error.

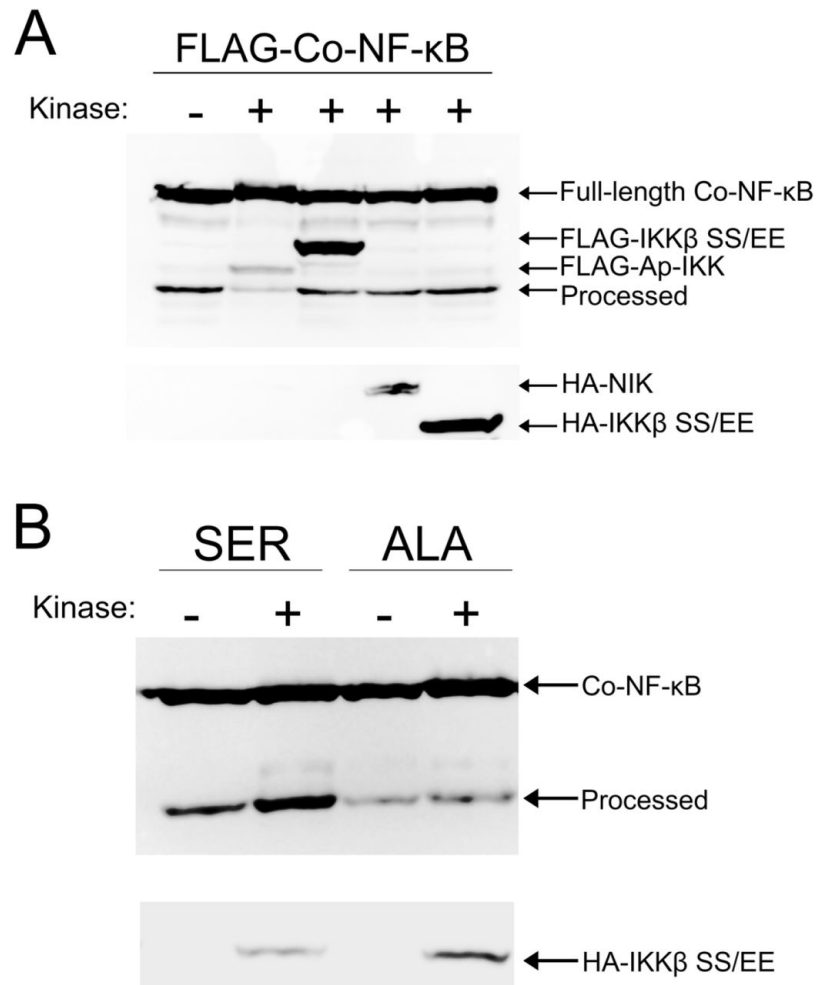


Figure 5.3 IKK-mediated processing of NF- κ B arose with multicellularity.

(A) Co-transfection with various kinases does not induce processing of FLAG-Co-NF- κ B in HEK 293T cells. Arrows indicate the various FLAG- or HA-tagged kinases used in these assays. Full-length Co-NF- κ B and processed Co-NF- κ B are also indicated. (B) Co-NF- κ B with C-terminal IKK target serines from *Aiptasia* NF- κ B (Co-NF- κ B-SER) or serine-to-alanine mutants (Co-NF- κ B-ALA) were co-transfected with constitutively active human HA-IKK β (SS/EE). Transfecting Co-NF- κ B-SER and HA-IKK β (SS/EE) resulted in the appearance of an increased amount of the lower band, but the alanine mutations (Co-NF- κ B-ALA) abolished that processing.

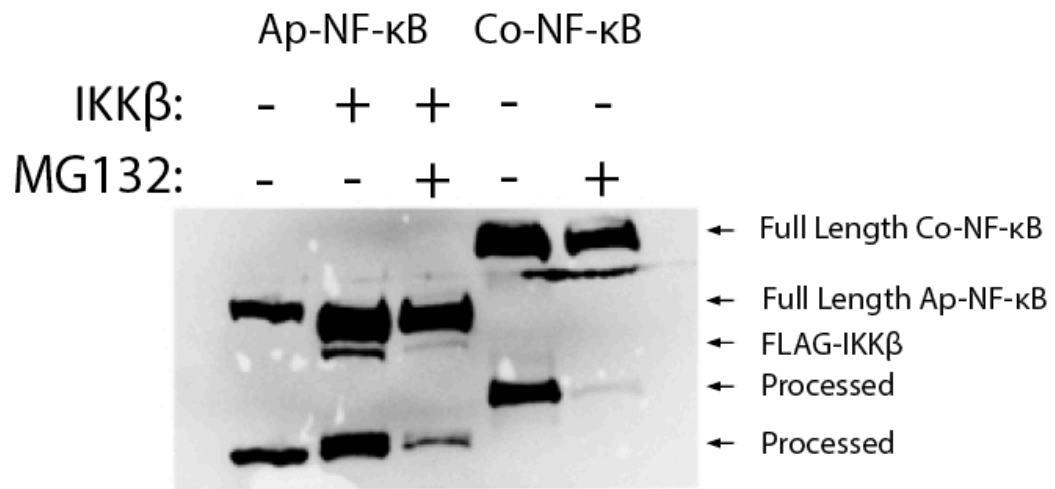


Figure 5.4 MG132 inhibits constitutive processing of Co-NF- κ B.

HEK 293T cells were transfected with FLAG-Ap-NF- κ B or FLAG-Co-NF- κ B and either a vector control or a FLAG-IKK β expression vector. Where indicated, MG132 (40 μ M) was added for 16 h. Cells were then lysed and anti-FLAG Western blotting was performed.

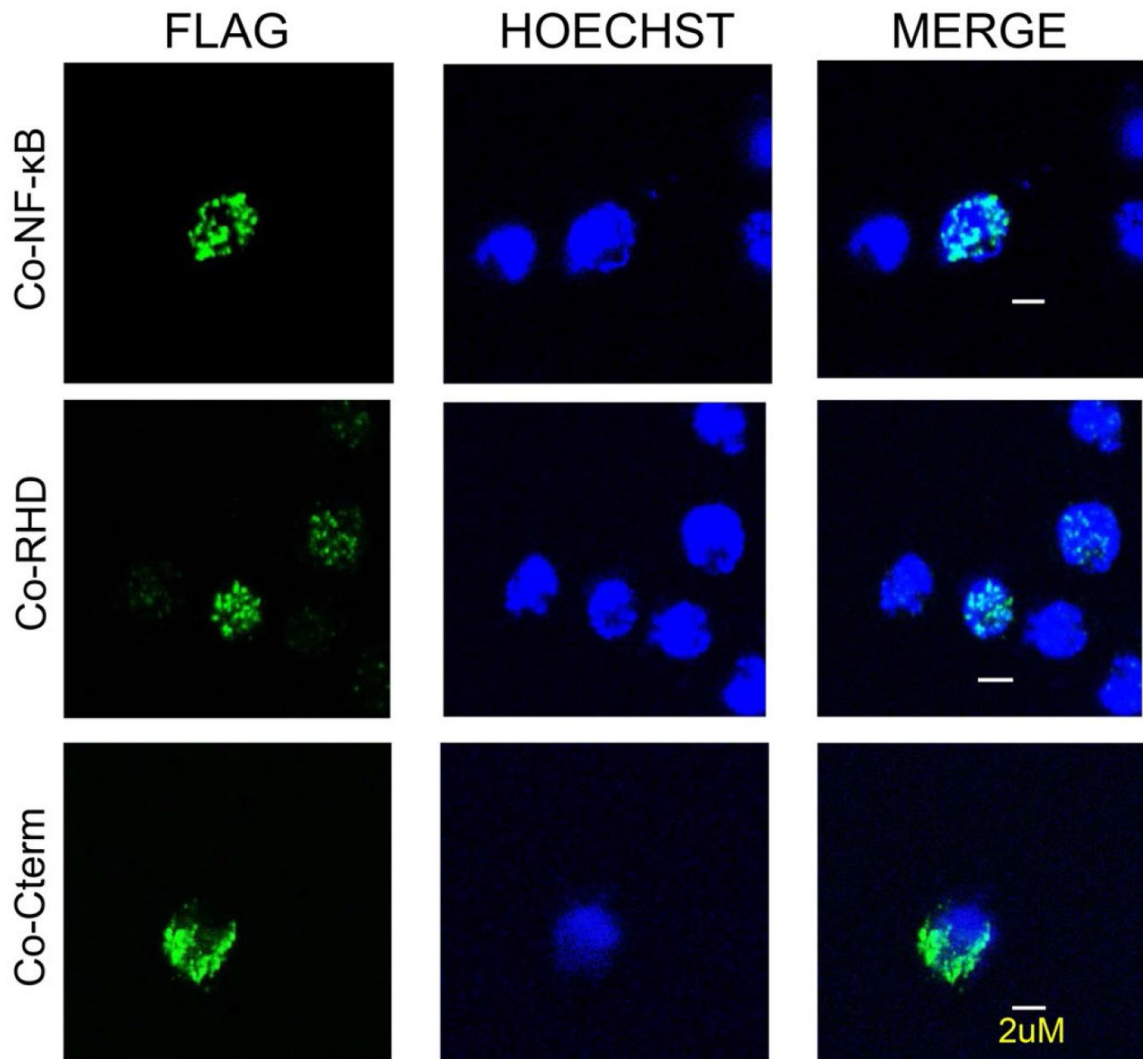


Figure 5.5 Transfected FLAG-Co-NF-κB and FLAG-Co-RHD proteins are located in the nucleus of *Capsaspora* cells.

Capsaspora cells were transfected with FLAG-tagged vectors for full-length Co-NF-κB, mutant Co-RHD, and mutant Co-Cterm. The cells were stained using anti-FLAG antiserum (left panels) and HOECHST (middle panels), and then MERGED (right panels). Scale bars are 2 μM.

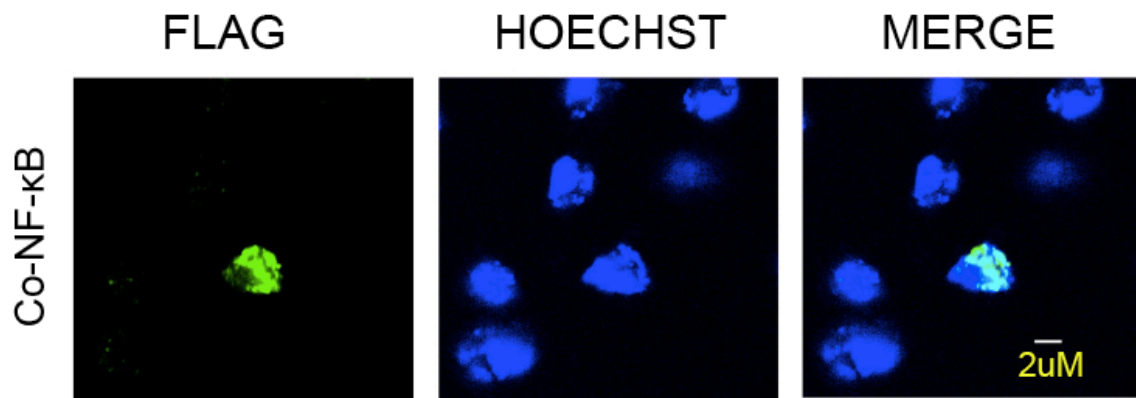


Figure 5.6 Full-length Co-NF-κB localizes to the nucleus of *Capsaspora* cells.

Capsaspora cells were transfected with an expression plasmid for FLAG-Co-NF-κB and anti-FLAG indirect immunofluorescence was performed as in Figure 5.5.

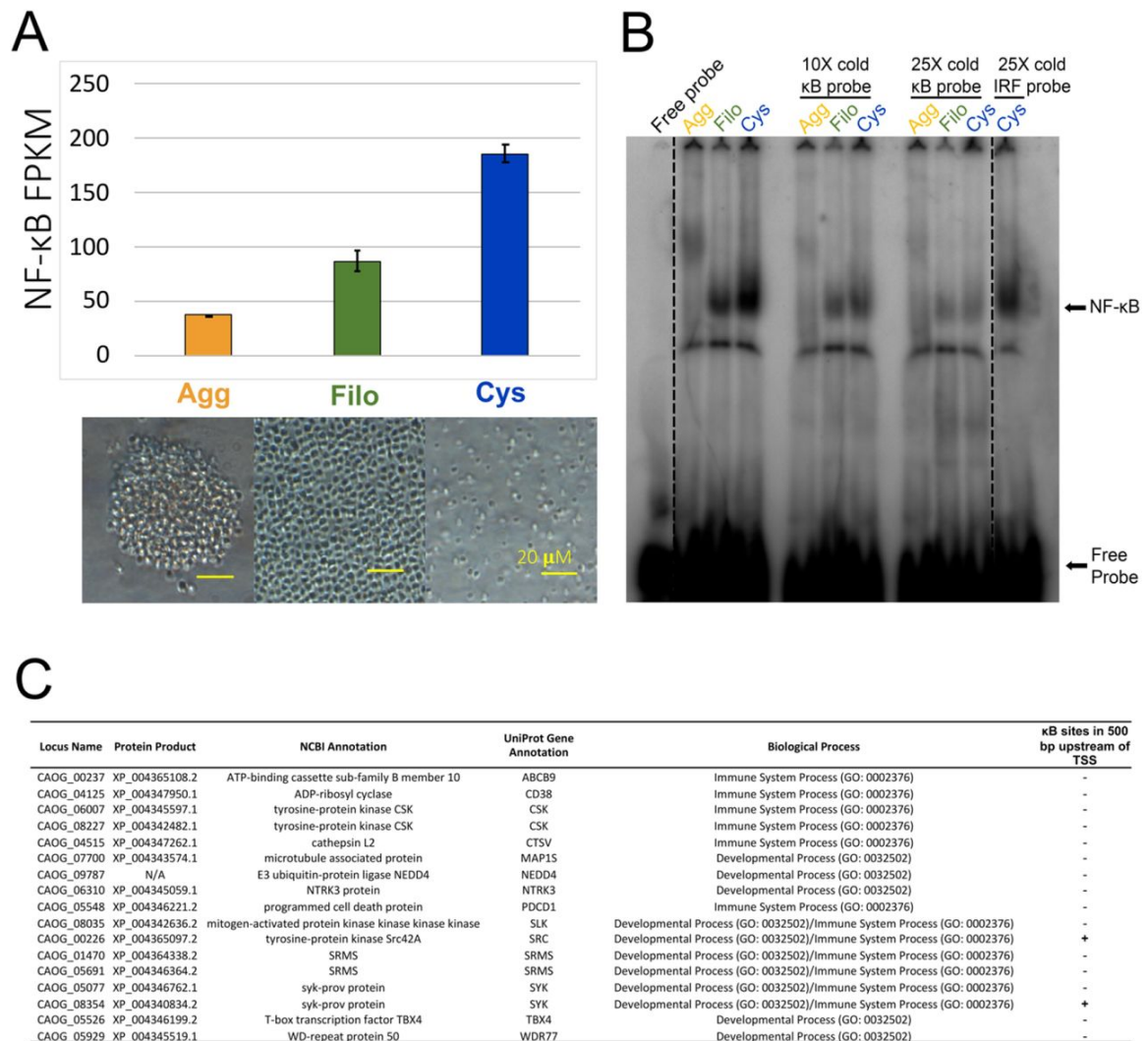


Figure 5.7 NF-κB mRNA levels and activity differ in different life stages of *Capsaspora* and proposed roles for NF-κB in *Capsaspora* development and immunity.

(A) Top: The FPKM values from Sebé-Pedrós et al. (2013) of NF-κB at each life stage, done in triplicate. Agg, Aggregative (yellow), Filo, Filopodic (green), Cys, Cystic (blue). Error bars are standard deviation. Bottom: Images taken with a light microscope of each life stage (Agg, Filo, and Cys from left to right). Yellow scale bar is 20 μM.

(B) *Capsaspora* whole-cell extracts were created from each life stage (see Methods). 70 µg of each extract was then used in an electromobility shift assay (EMSA). Lane 1 contains only free probe. Lanes 2-4 contain lysates from Agg, Filo, and Cys life stages incubated with a radioactive κB-site probe. Lanes 5 and 9 are empty. Lanes 6-8, and lanes 10-12 contain lysates from Agg, Filo, and Cys life stages as indicated, and were incubated with an excess (10X and 25X, respectively) of unlabeled κB-site probe. Lane 13 contains the Cys lysate incubated with 25X unlabeled IRF-site probe. NF-κB complexes and Free Probe are indicated with arrows. The dashed lines indicate where the gel was cut to remove excess lanes. (C) NF-κB expression may influence the expression of genes that are involved in developmental and immune system processes. The expression profiles of these genes correlate with NF-κB mRNA expression in each life stage (Agg, low; Filo, medium; Cys, high), and were identified via Biological Processes GO analysis. Two of the genes in this list (SYK and SRC) also contain κB sites in the 500 bp upstream of their TSS.

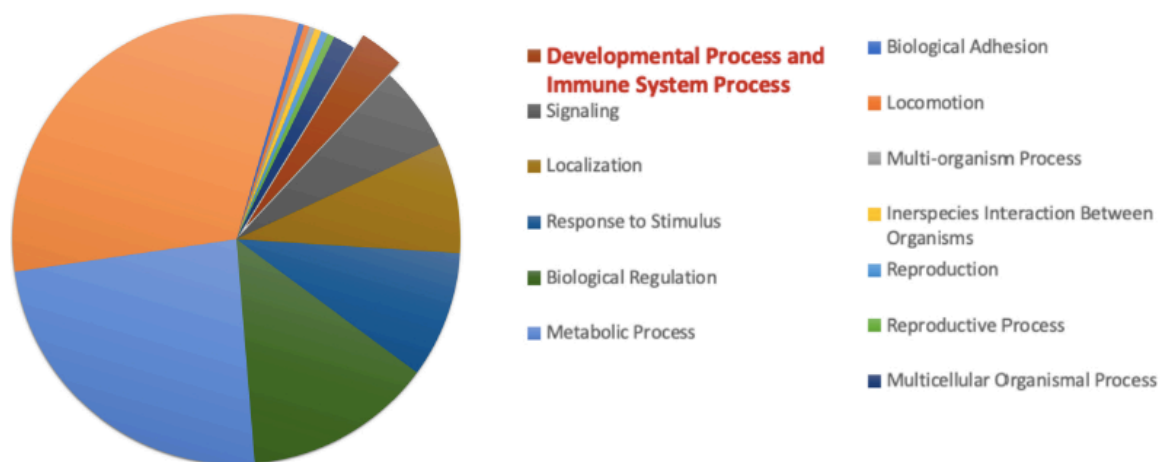


Figure 5.8 GO analysis of annotated *Capsaspora* genes co-expressing with Co-NF- κ B.

GO enrichment analysis was performed on the 305 annotated *Capsaspora* gene homologs that show the same expression pattern as Co-NF- κ B mRNA across the three *Capsaspora* life stages. The top 14 GO terms (right) and their relative abundance (left pie chart) are shown.

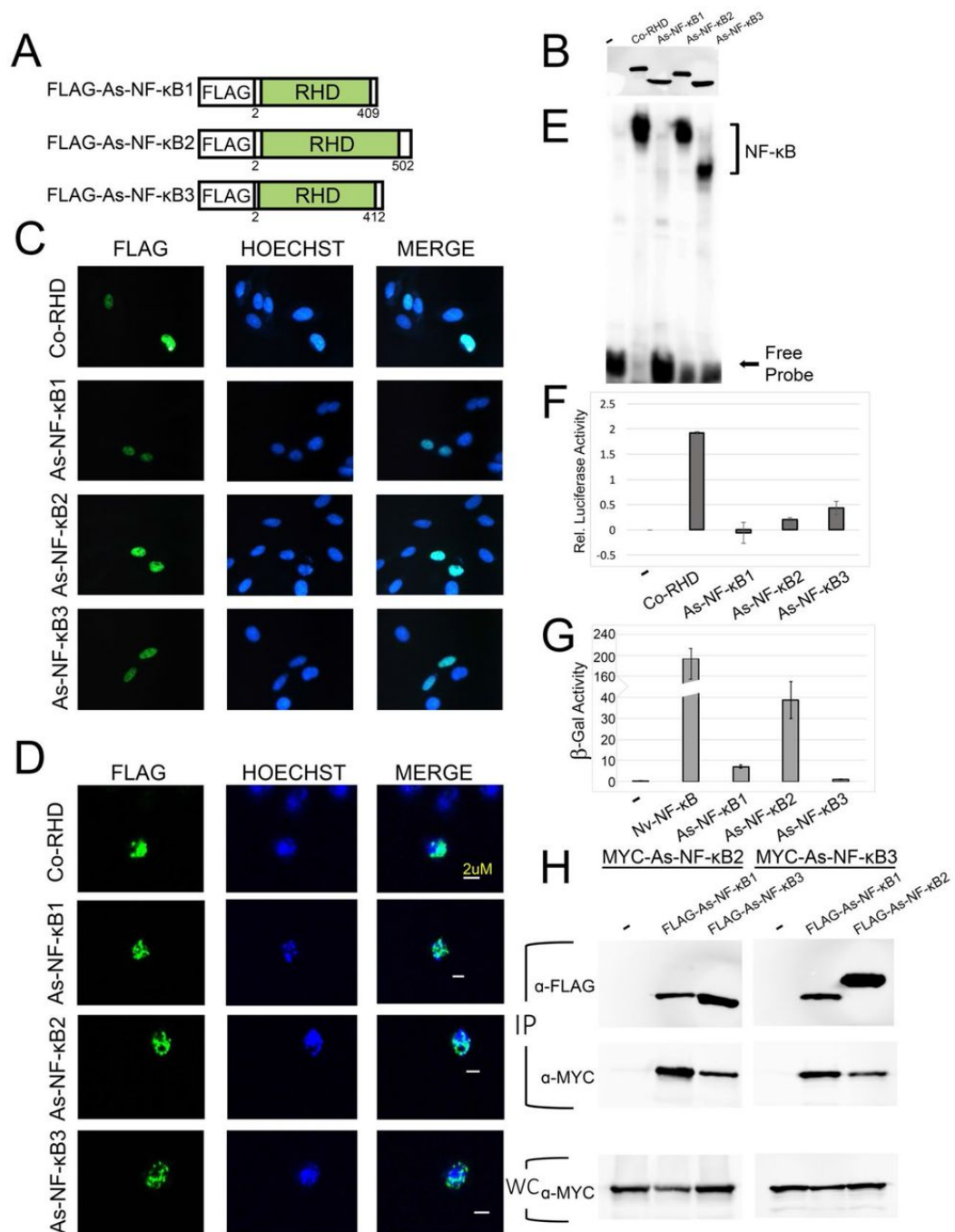


Figure 5.9 Characterization of cellular and molecular properties of three choanoflagellate NF-κBs.

(A) FLAG-tagged NF-κB proteins used in these experiments. From top to bottom the drawings depict the three NF-κB-like proteins from the transcriptome of *A. spectabilis*

(5). RHDs are in green (Rel Homology Domain). (B) Anti-FLAG Western blot of lysates from HEK 293T cells transfected with the indicated expression vectors. (C and D) Indirect immunofluorescence of DF-1 chicken cells (C) or *Capsaspora* cells (D) transfected with the indicated expression vectors. Cells were then stained with anti-FLAG antiserum (left panels) and HOESCHT (middle panels), and then MERGED in the right panels. Yellow scale bar in *Capsaspora* panels is 2 μ M. (E) A κ B-site electromobility shift assay (EMSA) using each of the indicated lysates from (B). The NF- κ B complexes and free probe are indicated by arrows. (F) A κ B-site luciferase reporter gene assay was performed with the indicated proteins in HEK 293 cells. Luciferase activity is relative (Rel.) to that seen with the empty vector control (1.0), and values are averages of three assays performed with triplicate samples with standard error. Values are shown on a log scale. (G) A GAL4-site *LacZ* reporter gene assay was performed with GAL4-fusion proteins in yeast Y190 cells. Values are average units of two assays performed with four samples with standard error. (H) Coimmunoprecipitation (IP) assays of MYC-tagged As-NF- κ B2 and As-NF- κ B3. In each IP assay, MYC-As-NF- κ Bs were co-transfected with pcDNA FLAG, FLAG-As-NF- κ B1, 2 or 3 as indicated. An IP using anti-FLAG beads was performed. Anti-FLAG (top) and anti-MYC (middle) Western blotting was then performed. An anti-MYC Western blot of the whole-cell (WC) lysates was also performed (bottom).

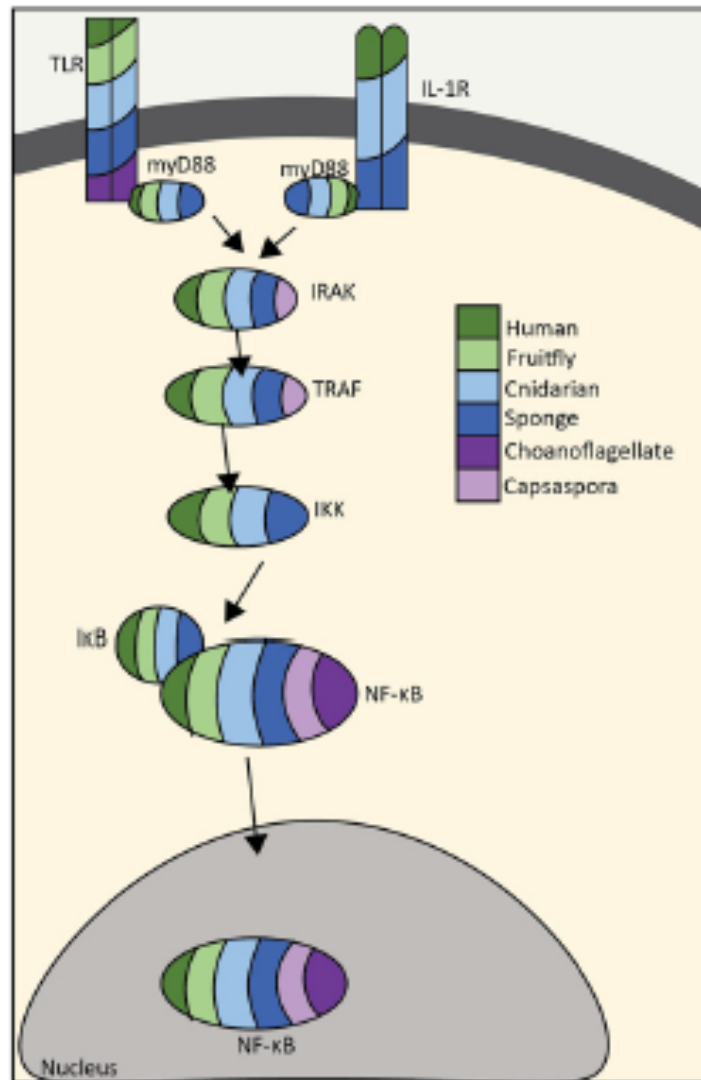


Figure 5.10 Schematic of the proteins found in the simplified TLR pathway of various organisms.

For each protein, the presence of a homolog is indicated according to the color legend on the right for the indicated organisms.

CHAPTER SIX

DISCUSSION

6.1 Summary

Prior to this thesis, the only information that was known about NF- κ B signaling in cnidarians were studies conducted in organisms such as *Aiptasia*, *Nematostella*, and *Hydra* (Franzenburg et al., 2012; Wolenski et al., 2013; Mansfield et al., 2017; Brennan et al., 2017). No information was known about NF- κ B signaling in more ecologically relevant cnidarians such as endangered corals or in organisms basal to cnidarians other than that homologs existed in the transcriptomes or genomes. As such, the molecular and functional characterization of NF- κ B signaling in corals, sponges, and protists required the development of novel tools and assays, and a deeper understanding of NF- κ B functions across basal phyla. The work presented in this thesis provides a foundation for future studies that seek to understand the role of NF- κ B signaling in basal organisms *in vivo*, including more biological information concerning this important, conserved signaling pathway.

In the coral *Orbicella faveolata*, and the sponge *Amphimedon queenslandica*, it appears that NF- κ B might play a role in pathogen response. That is, tissue that is stimulated with a potent activator of NF- κ B in higher organisms, LPS, results in pathway activation and increased expression of NF- κ B molecules in these organisms. While the way this activation appears to be induced in cell culture appears most consistent with the IKK-mediated activation seen in the p100 NF- κ B protein of vertebrates, the actual mechanism and level of activation is likely different *in vivo*. In protists, the function of

NF- κ B is unknown. However, in *Capsaspora*, NF- κ B likely plays different and unique roles in different life stages, given its differential expression in each life stage. How Co-NF- κ B behaves in the natural symbiotic setting of the snail could be different from what we have described in our in vitro experiments. In the choanoflagellates, up to three NF- κ B-like proteins exist in each organism. For the first time outside the Animal kingdom, it is demonstrated that NF- κ B can create heterodimers, demonstrating that subclasses of NF- κ B have evolved at least twice in evolution. These results suggest that in choanoflagellates such as *A. spectabilis*, there is unique and distinct DNA binding of these heterodimers that could allow for transcriptional flexibility and pathway activation.

6.2 Chapter Three Discussion

6.2.1 Summary

The data presented in Chapter three suggest that there are highly conserved elements of a TLR-to-NF- κ B innate immune pathway in the endangered coral *O. faveolata*. That is, we show that the intracellular TIR domain of the single Of-TLR protein is most similar by sequence analysis to human TLR4, and that the TIR domain of Of-TLR can directly interact with human MYD88 (Figure 3.3), which is an essential adapter protein for human TLR4-based signaling to NF- κ B. Moreover, the single Of-NF- κ B protein has sequence similarity, an overall structural organization, and transactivation properties that are more like human NF- κ B proteins than human Rel proteins (Figure 3.5). Furthermore, the Of-NF- κ B protein can be processed in response to overexpression of human and *O. faveolata* IKK proteins in a manner consistent with non-canonical

pathway processing of NF- κ B p100 in human cells (Figure 3.6 and 3.7). Finally, we show that a mammalian ligand of TLR4-to-NF- κ B activation, namely gram-negative bacterial LPS, activates a gene expression profile in isolated *O. faveolata* tissue that is consistent with overall activation of NF- κ B pathway genes (Table 3.2).

6.2.2 TLR-to-NF- κ B pathway components in cnidarians

To date, a handful of studies have identified TLR-to-NF- κ B pathway components in environmentally stressed cnidarians. Anderson et al. (2016) and Pinzón et al. (2015) identified transcripts encoding NF- κ B signaling components in their transcriptomes from bleached *O. faveolata*. In the coral *Acropora palmata*, transcripts for a putative tumor necrosis factor receptor-associated factor 3 (TRAF3) homolog and NF- κ B were shown to be upregulated when coral were exposed to increased water temperature (DeSalvo et al., 2010). The upregulation of NF- κ B pathway components at the transcriptional level has also recently been described in the sea anemone *A. pallida* following treatment with heat or chemicals that induce bleaching (Mansfield et al., 2017). However, the precise mechanisms by which these conserved signal transduction pathways elicit their responses, the gene targets regulated by NF- κ B in these organisms, and their biological roles in these cnidarians remain largely unknown.

6.2.3 LPS treatment of *O. faveolata* indicate an activation of the NF- κ B pathway

Treatment of *O. faveolata* tissue with LPS results in significant differential expression of genes in a manner indicating activation of the NF- κ B pathway. These results are consistent with other studies that have reported upregulation of NF- κ B pathway transcripts as a response to pathogen-induced stress in cnidarians. For example,

Franzenburg et al. (2012) demonstrated that microbial infection of *Hydra* elicits a TLR-like pathway response and bacterial flagellin can activate a TLR-to-NF- κ B signaling when the *Hydra* TLRs are expressed in mammalian cells (Bosch et al., 2009). Similarly, Brennan et al. (2017) have shown that a coral bacterial pathogen can activate the *N. vectensis* TLR. However, our results are the first to show that NF- κ B pathway components in an endangered coral are upregulated in response to treatment with a known mammalian inducer of NF- κ B, i.e., LPS. Whether this activation of NF- κ B components by LPS occurs through the Of-TLR that we characterize here is not known at this time. Interestingly, a recent paper demonstrated that about 40 gram-negative marine microbes, as well as their purified LPS molecules, evaded detection by mammalian immune systems, which might be due to the lack of co-evolution between mammals and marine microbes, as well as longer side-chains on the lipid portion of the LPS (Gauthier et al., 2021). It is tempting to speculate that this range of microbes gives rise to a receptor that has a promiscuous ability to recognize a variety of microbes, such as the TLR described here, or perhaps, a range of receptors capable of recognizing this diversity that is yet discovered in marine organisms. To support the former hypothesis, as noted in this thesis, there are significantly less immune-related receptors present in basal organisms.

6.2.4 Activation of the NF- κ B pathway is likely different mammalian and basal organisms

Although our results suggest a link between LPS treatment and activation of NF- κ B in *O. faveolata*, it is important to note the differences between our findings and the LPS-induced rapid activation of NF- κ B that is usually studied in mammalian systems.

That is, our results show that the levels of mRNAs encoding several NF- κ B pathway components are upregulated approximately 4 h after treatment of *O. faveolata* tissue with LPS. In contrast, LPS-induced activation of NF- κ B in mammalian systems is usually studied as the rapid nuclear localization of a latent cytoplasmic NF- κ B complex approximately 10-30 min following treatment with LPS. Thus, we have not directly shown that LPS treatment activates the NF- κ B protein in a conventional sense.

6.2.5 *O. faveolata* has non-canonical and canonical components of NF- κ B signaling

Our results suggest that the TLR-to-NF- κ B pathway in *O. faveolata* has characteristics of both non-canonical and canonical NF- κ B signaling in mammals. On the one hand, the Of-NF- κ B pathway is like the mammalian non-canonical pathway in that 1) the overall organization of Of-NF- κ B is similar to the mammalian NF- κ B protein p100, 2) C-terminal truncation leads to nuclear localization, enhanced DNA binding, and enhanced transactivation in vertebrate cells, 3) IKKs can induce partial processing of overexpressed Of-NF- κ B in human cells, 4) sites of IKK-dependent serine phosphorylation in p100 are conserved in Of-NF- κ B, and 5) these serine residues are required for processing of Of-NF- κ B in human cells. On the other hand, human IKK β , which is normally associated with canonical NF- κ B signaling in mammalian cells, is as active as, if not even more active than, Of-IKK towards Of-NF- κ B in our *in vitro* phosphorylation and human cell induction experiments. Moreover, TLR4 generally activates the canonical NF- κ B pathway in mammalian cells. Furthermore, although the TIR domain of the Of-TLR has properties of human TLR4, the extracellular LRR domain of Of-TLR is more similar to the *Drosophila* Toll protein (a multiple cysteine cluster

TLR) than to TLR4 (a single cysteine cluster TLR protein, similar to what is seen with Nv-TLR (Brennan et al., 2017). Finally, the clustering of the TIR domain of Of-TLR with human TLR4 is not seen with other methods of phylogenetic analysis (e.g., maximum-likelihood and maximum parsimony [data not shown]). Thus, the evolutionary processes by which cnidarian TLR and TLR-like proteins, as well as a single cnidarian IKK, diversified into the multiple member higher metazoan TLR and IKK families vis-à-vis canonical vs. non-canonical signaling to NF- κ B are not obvious.

6.3 Chapter Four Discussion

6.3.1 Summary

In Chapter four, we demonstrate that at least two sponges have active NF- κ B proteins. First, we show that the *A. queenslandica* NF- κ B protein is phylogenetically and structurally similar to other metazoan NF- κ B proteins (Figure 4.1), and that its DNA-binding profile, transactivation properties, and subcellular regulation by C-terminal sequences are similar to NF- κ B proteins found in other species (Figure 4.2-4.4). We also demonstrate that a black encrusting demosponge contains a protein that can bind a conserved κ B site in an EMSA and two proteins that can be detected by an anti-NF- κ B antibody in Western blots and tissue immunofluorescence (Figure 4.5-4.6). Furthermore, treatment of this sponge tissue with LPS leads to further activation of its putative NF- κ B protein, as shown by both increased processing and increased DNA binding (Fig 4.7). Finally, we also show that *A. queenslandica* contains homologs to many components of a TLR-to-NF- κ B pathway (Figure 4.9). Taken together, these results suggest that sponges

have a highly conserved NF- κ B pathway, which is expressed in a subset of cells that likely are involved in an innate immune response.

6.3.2 *A. queenslandica* expresses an NF- κ B protein that is most homologous to a human p100 NF- κ B protein

Consistent with previous results (Gauthier and Degnan, 2008), our results indicate that Aq-NF- κ B is structurally similar to the human NF- κ B p100 protein and to our previously characterized sea anemone *Aiptasia* and coral *O. faveolata* NF- κ B proteins (Sun, 2011; Mansfield et al., 2017; Williams et al., 2018). In our cell-based assays, removal of the ANK repeat domain allowed for localization of Aq-NF- κ B to the nucleus and liberation of its transcription activation activity.

The occurrence of bipartite NF- κ B proteins that contain both the RHD and the C-terminal ANK repeats in sponges (*A. queenslandica*) and the protist *C. owczarzaki* is different than what is seen in some cnidarians, such as *N. vectensis*, which contains separate RHD and I κ B-like ANK-repeat proteins. More specifically, we and others have demonstrated that the sponge and protist NF- κ B proteins are more closely related structurally to the human p100 and p105 NF- κ B proteins (Gauthier and Degnan, 2008; Mansfield et al., 2017; Williams et al., 2018), suggesting that genes such as the naturally shortened *N. vectensis* NF- κ B protein (Wolenski et al., 2011) are the result of organism-specific gene splitting events. Such gene-splitting events are not uncommon in cnidarians (Gilmore and Wolenski, 2012).

6.3.3 The structure of *A. queenslandica* NF- κ B might be responsible for its unique DNA-binding properties

Surprisingly, full-length Aq-NF- κ B bound to DNA to a similar extent as the Aq-RHD protein (deleted of the C-terminal ANK repeats) (Figure 4.2F). Therefore, the C-terminal ANK repeat domain of Aq-NF- κ B appears to inhibit nuclear translocation (Figure 4.3A), but not DNA binding. In addition, two protein-DNA complexes were seen in EMSAs when using extracts from cells overexpressing full-length Aq-NF- κ B. We suspect that these two complexes correspond to dimer and tetramer forms of Aq-NF- κ B. Of note, NF- κ B complexes larger than dimers have been reported for mammalian NF- κ B p100 and p105 proteins (Savinova et al., 2009). In all other NF- κ B proteins analyzed to date (as well as with Of-NF- κ B and Co-NF- κ B), the C-terminal ANK repeat domain inhibits the DNA-binding activity of RHD sequences in assays similar to those that we have performed in Chapter four. As we note, Aq-NF- κ B contains an unusually long sequence between its RHD and the start of its ANK repeats. It is possible that this extended sequence interferes with the ability of the ANK repeat domain to inhibit DNA binding by the RHD. In any case, the ANK repeat domain can block nuclear translocation (Figure 4.3A) and transcriptional activation by Aq-NF- κ B in reporter gene assays (Figure 4.2D), and inhibition of nuclear translocation would functionally block Aq-NF- κ B's ability to activate gene transcription in cells. Why there are extra sequences between the RHD and the ANK repeat domain in Aq-NF- κ B and whether these sequences are important for any NF- κ B-related processes in sponges are not known. However, the additional sequences do not appear to be involved in or required for proteolytic

processing of Aq-NF- κ B by known IKKs (Figure 4.3C).

6.3.4 Sponges have both canonical and non-canonical signaling components

A. queenslandica is similar to other evolutionarily basal organisms in that its genome encodes one NF- κ B protein and no Rel proteins. Additionally, Aq-NF- κ B contains serine residues C-terminal to the ANK repeats that can be phosphorylated *in vitro* by the human IKK β kinase (an activator of canonical NF- κ B signaling in humans), and these conserved serines are required for IKK-induced processing of Aq-NF- κ B in human cells in tissue culture. It is also of interest that activation of canonical signaling (by IKK β) in 293T cells can promote processing of Aq-NF- κ B (Figure 4.3C), as well as some cnidarian NF- κ B proteins (Mansfield et al., 2017; Williams et al., 2018). Given that cnidarians and poriferans lack Rel-like proteins, their NF- κ B proteins appear to be obligate homodimers, whereas human p100 is generally in a heterodimer with RelB. Thus, basal NF- κ B signaling may have aspects of both the canonical and non-canonical signaling pathways found in mammals.

6.3.5 Constitutively processed NF- κ B in the sponge

In any case, the majority of a putative NF- κ B protein in tissue from a black encrusting sponge appears to be processed, binds to DNA, and is located in the nucleus of cells. Thus, in this sponge, most of its NF- κ B protein appears to be constitutively in its active form. We have previously shown that most of the NF- κ B protein of the sea anemone *Aiptasia* is also a constitutively nuclear and processed protein in animals (Mansfield et al., 2017). Moreover, the majority of mouse NF- κ B p100 is in its processed form when analyzed in extracts taken directly from liver, lymph nodes, and bone marrow

(Senftleben et al., 2001). The factors that promote constitutive processing of NF- κ B proteins in these diverse organisms and tissues *in vivo* are not known. However, it is interesting to speculate that continual exposure to pathogens or other activating ligands induces NF- κ B processing in each of these situations. Further studies should be directed at determining the requirements for the production of the constitutively active sponge NF- κ B protein, whether this constitutive activity is also found in other poriferans, the trade-offs to this constitutive activity, and the situations in which this processing does not occur.

6.3.6 IKKs can process Aq-NF- κ B, but are not encoded in the *A. queenslandica* genome

Two IKK-like proteins are encoded in genomic sequences of *A. queenslandica* (NCBI Accession numbers XP_019854889.1 and XP_003387126.1), and both appear more similar to the TBK/IKK ϵ subclass of IKKs than to the IKK α/β subclass (data not shown). In general, the mammalian TBK/IKK ϵ kinases do not phosphorylate I κ Bs or C-terminal serines in p100. Therefore, although our results in 293T cells suggest that *A. queenslandica* has a kinase that is capable of inducing processing of Aq-NF- κ B, such a kinase has yet to be identified, either because there is a lack of genomic coverage to identify it or this kinase has no obvious sequence homology to other known metazoan IKKs.

6.3.7 A receptor for LPS exists in the sponge

We also found that treatment of fresh sponge tissue with LPS for 30 min resulted in the almost complete processing of the full-length NF- κ B protein, and an increase in

κ B-site binding activity. In the demosponge *Suberites domuncula* it has been shown that LPS can be bound by an LPS-interacting protein (Wiens et al., 2007), which is capable of binding and dimerizing with an Sd-MyD88 homologous protein. Currently, there is no identified LPS-interacting protein homolog in *A. queenslandica*, but it does contain a homolog of the TIR-domain adapter protein MyD88 (Table 4.2). Nevertheless, our data suggest that the black encrusting sponge possesses an LPS-interacting protein and/or a receptor capable of recognizing LPS.

6.3.8 Homologs to the NF- κ B pathway exist in the sponge

Although we have yet to identify a sponge kinase capable of processing Aq-NF- κ B, we have identified many other homologs of a likely NF- κ B pathway in *A. queenslandica* (Figure 4.9). *A. queenslandica* lacks a complete Toll-like receptor (TLR), but does express what appears to be a homolog of the Interleukin 1 receptor (IL-1R), which contains an intracellular Toll-interleukin receptor (TIR) domain as well as extracellular immunoglobulin domains. While mammalian TLR and IL-1R proteins bind different ligands due to their different extracellular domains, both receptors contain intracellular TIR domains that can interact with many of the same downstream signaling proteins (e.g., MYD88).

6.3.9 Cell-specific expression of NF- κ B in the sponge

Recently, it was shown that *A. queenslandica* NF- κ B transcripts are expressed in at least two cell types: choanocytes, which are cells that reside in the chambers that allow the sponge to filter feed, and to a lesser extent in archaeocytes, which are pluripotent cells that are motile inside the mesoglea (Sogabe et al., 2019). In the black encrusting sponge

that we examined, primarily nuclear NF- κ B protein was detected in a subset of cells that appeared to be within the tissue mesoglea. We did not detect any choanocyte chambers in our images. Therefore, we believe that the cells that are stained with our NF- κ B antiserum are likely either archaeocytes or another cell type that has yet to be identified. Archaeocytes are known to be amoeboid-like and to transport particulate matter taken-up by choanocytes into the sponge tissue. Therefore, it is possible that archaeocytes require NF- κ B-related immune functions due to their continuous exposure to water-borne microbes. Furthermore, archaeocytes are also known to be totipotent and capable of differentiation into many cell types (Buscema et al., 1980), and it is well established that NF- κ B plays a role in the differentiation of several cell types in vertebrates (Guttridge et al., 1999; Bottero et al., 2006; Zhang et al., 2012).

6.3.10 Other roles of NF- κ B in the sponge

The ability of *E. coli*-derived LPS to further activate NF- κ B in sponge tissue suggests that NF- κ B has role in poriferan anti-bacterial immunity. However, there is likely to also be an early developmental role for NF- κ B in poriferans. That is, Gauthier and Degnan (2008) showed that NF- κ B transcripts are present during early development of *A. queenslandica*, especially in granular cells and flask cells (Gauthier and Degnan, 2008). Dual roles in embryonic development and adult immunity have been shown for the NF- κ B protein Dorsal in *Drosophila* and have been postulated for the NF- κ B protein in *N. vectensis* (Wolenski et al., 2011).

6.4 Chapter Five Discussion

6.4.1 Summary

In Chapter five, we have functionally characterized and compared, for the first time, NF- κ B proteins from two protists. Taken together, these results demonstrate that although functional DNA-binding and transcriptional-activating NF- κ B proteins exist in these protists, the overall structures and regulation of these proteins vary considerably, both among protists and when compared to animal NF- κ Bs.

6.4.2 Proteasome-mediated processing without an IKK in *Capsaspora*

In the vertebrate NF- κ B proteins p100 and p105, C-terminal ANK repeats inhibit the DNA-binding activity of the RHD, and proteasome-mediated processing of the ANK repeats is terminated by and within the GRR (Sun, 2011). In the *Drosophila* Relish protein, the C-terminal ANK repeats also inhibit DNA binding, but the C-terminal ANK repeats are removed by an internal site-specific cleavage event, which does not involve the proteasome, and Relish has no GRR (Stöven et al., 2003). Thus, the presence of ANK repeats and a GRR in Co-NF- κ B suggests that proteasomal processing would lead to nuclear translocation and activation of its DNA-binding activity. Indeed, removal of the C-terminal residues of Co-NF- κ B does allow it to enter the nucleus of vertebrate cells and unleashes its DNA-binding activity (Figs. 5.2D and E). Moreover, treatment of cells with the proteasome inhibitor MG132 blocks the basal processing of Co-NF- κ B that is seen in transfected 293T cells (Figure 5.4). However, this basal processing of Co-NF- κ B in 293T cells is not enhanced by co-expression of IKK (unless C-terminal IKK target sequences are added, see Figure 5.3B) and there are no obvious IKK target serines in Co-NF- κ B nor

are there any IKK homologues in the *Capsaspora* genome. Furthermore, most of the Co-NF- κ B appears to be in the nucleus when it is overexpressed in *Capsaspora* cells (Figure 5.5 and 5.6), suggesting that it is constitutively processed to its RHD. If Co-NF- κ B does undergo a signal-induced proteasomal processing in *Capsaspora* cells, then it is unlikely to be dependent on an IKK-like kinase. Alternatively, but probably less likely, full-length Co-NF- κ B enters the nucleus of *Capsaspora* cells but not DF-1 chicken cells. In the absence of a Co-NF- κ B-specific antiserum, we cannot distinguish between these possibilities.

6.4.3 Regulation of choanoflagellate NF- κ Bs is unclear

In contrast to *Capsaspora*, the choanoflagellate NF- κ B proteins lack C-terminal ANK repeats and GRRs, and all three *As*-NF- κ B proteins are constitutively in the nucleus when overexpressed in vertebrate or *Capsaspora* cells (Figure 5.9). We have not been able to identify an I κ B-like protein in the *A. spectabilis* genome. Thus, it is unclear whether choanoflagellate NF- κ B is regulated by an ANK-dependent cytoplasmic retention mechanism, or whether, for example, choanoflagellate proteins are constitutively nuclear in their native settings. Nevertheless, it is clear that the regulation of both *Capsaspora* and choanoflagellate NF- κ B proteins is distinct from what is seen with NF- κ B proteins in higher metazoans.

6.4.4 Constitutive processing of NF- κ B in protists

Of interest, constitutively nuclear localization of NF- κ B proteins has also been seen in other settings. That is, we have previously shown that in both the sea anemone *Aiptasia* and sponges, most NF- κ B staining is nuclear and the proteins are mostly

processed in their native settings (Mansfield et al., 2017; Williams et al., 2020). Thus, we have argued previously (Williams and Gilmore, 2020) that these basal NF- κ B proteins may be constitutively in an active state, perhaps due to continual interaction with upstream activating ligands or pathogens. Of note, most NF- κ B p100 is also in its processed form in mouse spleen tissue (Senftleben et al., 2001).

6.4.5 NF- κ B likely has a specialized, rather than general function in choanoflagellates

Among the 21 choanoflagellates for which there is sufficient transcriptomic/genomic information, it appears that only 12 have any NF- κ B genes. Surprisingly, in seven of these 12 choanoflagellates there are multiple NF- κ B genes. Thus, it is clear that there have been gains and losses of NF- κ B genes among the choanoflagellates. We note that NF- κ B has also been lost in other organisms including *C. elegans* and ctenophores (Wolenski and Gilmore, 2012). The absence of NF- κ B in some choanoflagellates and its expansion in others (e.g., *A. spectabilis*) suggests that NF- κ B has a specialized, rather than a general, biological function in choanoflagellates.

6.4.6 Choanoflagellate NF- κ Bs likely dimerize in vivo for specialized transcriptional control

The presence of three NF- κ B-like heterodimerizing proteins in *A. spectabilis* is the first example of an organism in a lineage that predates flies with multiple NF- κ B family proteins that are capable of forming heterodimers. Thus, expansion of NF- κ B genes has occurred multiple times in evolution, i.e., at least once in the metazoan lineage and once within choanoflagellates. Furthermore, since each As-NF- κ B homodimer has a

different ability to bind DNA and activate transcription, it appears that there are subclasses of NF- κ B within *A. spectabilis* and likely also within other choanoflagellates that have multiple NF- κ Bs. It is interesting to note that As-NF- κ B1, 2, and 3 are phylogenetically separate and cluster most closely with NF- κ Bs from certain other choanoflagellate species that contain multiple NF- κ Bs. For example, *Savillea parva* contains three NF- κ Bs, each of which clusters with a separate As-NF- κ B (Figure 5.9B). Thus, we hypothesize that choanoflagellates, like vertebrates, have evolved a mechanism for differential transcriptional control of genes through the use of combinatorial dimer formation.

6.4.7 Regulation of NF- κ B in protists is different than vertebrates

The differential mRNA expression and DNA-binding activity of NF- κ B among different life stages of *Capsaspora* suggest that NF- κ B has life stage-specific functions. It is notable that the DNA-binding activity of NF- κ B in these different life stages correlates with differences in the levels of NF- κ B mRNA, rather than as differences in induced activity. That is, in most metazoans, the activity of NF- κ B is regulated at the post-transcriptional level, whereas in *Aiptasia* and corals, we have found that NF- κ B mRNA levels and DNA-binding activity appear to be coordinately regulated, suggesting transcriptional regulation. That is, in *Aiptasia*, thermal bleaching causes transcriptional upregulation of NF- κ B, which also results in increased protein expression of nuclear, DNA binding-active NF- κ B (Mansfield et al., 2017), which is similar to what we see with NF- κ B across the *Capsaspora* life stages. Similarly, treatment of the coral *O. faveolata* with lipopolysaccharide results in increased expression of NF- κ B target genes, rather than

increased post-translational activation of NF- κ B (Williams et al., 2018). Thus, it appears that in several basal organisms NF- κ B proteins are constitutively nuclear and that increases in their activity is the result of transcriptional upregulation of NF- κ B mRNA, rather than induced proteolysis, which occurs in most mammalian and fly systems.

6.4.8 NF- κ B in many basal organisms are constitutively processed

Activation of NF- κ B by signal-induced degradation of I κ B sequences is essentially dogma in vertebrates. Although several basal NF- κ Bs, including Co-NF- κ B, can be formulated to undergo IKK-induced processing when expressed in human cells in culture (Mansfield et al., 2017; Williams et al., 2018; Williams et al., 2020), there is much evidence that such regulation may not occur in the native organisms. For example, although sponge and sea anemone *Aiptasia* NF- κ Bs have C-terminal ANK repeats and GRRs, they are largely processed and nuclear in the animals themselves (Mansfield et al., 2017; Williams et al., 2018; Williams et al., 2020). Moreover, induction of NF- κ B DNA-binding activity and protein levels by loss of symbiosis in *Aiptasia* appears to be primarily a result of increased transcription of NF- κ B and not due to induced ANK repeat degradation. Similarly, the increased NF- κ B DNA-binding activity seen in different life stages of *Capsaspora* is paralleled by increased expression of NF- κ B mRNA in these life stages. Furthermore, when Co-NF- κ B is overexpressed in *Capsaspora* cells it is primarily in the nucleus, likely due to constitutive processing to the RHD form. We note here that the nuclear staining of transfected FLAG-Co-NF- κ B appears punctate, which may be due its localization at sites of active gene transcription, or “nuclear bodies” (Brasch and Ochs, 1992). Thus, a more relevant question for many basal NF- κ B proteins may be what

conditions stop them from being processed. Of note, the three As-NF- κ Bs do not have C-terminal ANK repeats and we have not been able to identify a putative I κ B in *A. spectabilis* transcriptomic databases. Therefore, it is possible that the activity of these constitutively nuclear choanoflagellate NF- κ Bs is fully regulated by transcriptional control of their genes.

6.4.9 Target genes of *Capsaspora* NF- κ B

Of the nearly 1350 genes whose expression correlated with NF- κ B expression across different *Capsaspora* life stages, almost 20% contained NF- κ B binding sites within 500 base pairs upstream their TSS (Figure 5.7). While this result might be an overestimate of NF- κ B gene targets or genes indirectly influenced by expression of NF- κ B, there are likely additional NF- κ B binding sites that could affect target gene expression. For example, ATAC-seq data have suggested that the regulatory sites in the *Capsaspora* genome are present in first introns, 5' UTRs, as well as the proximal intergenic regions (Sebé-Pedrós et al., 2016). Among the ~1350 genes that we identified with expression patterns similar to NF- κ B, there are 192 genes that contain NF- κ B binding sites in their proximal promoters. Two of these genes are homologs of SYK and SRC (CAOG_08354 and CAOG_00206), which are known in higher organisms to be involved in immunity and development (Mócsai et al., 2010; Lowell, 2011). However, the list of ~1350 genes most certainly contains genes that are controlled by other transcription factors or are regulated by signaling or developmental events that are partially or not at all affected by NF- κ B. Nevertheless, it is clear that NF- κ B might play several roles, and perhaps different roles, in each life stage of *Capsaspora*.

6.4.10 The biological function of *Capsaspora* NF- κ B remains unclear

Our studies suggest that the regulation and associated biology of NF- κ B in single-celled organisms are different from what is seen in multicellular vertebrates and flies. Perhaps, the concerted effort of aggregation in *Capsaspora* and the correlative decrease in NF- κ B in these aggregated cells reflect a need to suppress collective immunity to form a symbiotic group. Alternatively, *Capsaspora* is normally a symbiont in the hemolymph of the snail *B. glabrata*, where NF- κ B may play a role in maintaining symbiosis, which has been suggested as one function of NF- κ B in other organisms (Mansfield and Gilmore, 2019).

6.4.11 NF- κ B pathway signaling in protists

It is not clear what type of pathway might lead to activation of NF- κ B in protists. Across the animal kingdom, the binding of a ligand to receptors such as TLRs/IL-1Rs or TNFRs initiates signaling pathways that converge on an IKK complex which then activates NF- κ B (Gilmore, 2006). However, in basal organisms, many of these components are missing, few in number, or lack critical domains (Williams and Gilmore, 2020). For example, while some cnidarians contain homologs to TLRs, other cnidarians, some sponges, and choanoflagellates contain only TIR-domain proteins that lack the important extracellular components of the TLR (King et al., 2008; Brennan and Gilmore, 2018; Richter et al., 2018; Williams et al., 2018; Williams et al., 2020). Furthermore, *Capsaspora* does not contain any homologs to TLRs, ILR-1 or TNFRs (Suga et al., 2013).

6.5 Unanswered questions and future experimental goals

Although this research has studied the phylogeny, structure, and function of several basal NF- κ B proteins, these results beget many questions about the role of NF- κ B in basal organisms. Some of these unanswered questions are discussed below.

1. *Does TLR activate an NF- κ B pathway in corals?* We have shown that exposure of *O. faveolata* tissue to LPS induces an expression profile that implies that members of the TLR-to-NF- κ B pathway are activated. Nevertheless, these results do not explicitly answer whether TLR is the receptor responsible for activation of this pathway in corals. Our lab has previously shown that exogenous expression of the anemone Nv-TLR can activate canonical NF- κ B signaling in human cells by reporter assays and by induction of I κ B α phosphorylation, however, we have not been able to show activation of Nv-NF- κ B in anemones or isolated anemone structures such as nematosomes (Brennan et al., 2017). For example, AP-1 is also a transcription factor downstream of NF- κ B in vertebrates (Roy et al., 2016), and this could be the effector pathway for TLR-induced immunity in *Nematostella*.

Human TLRs typically activate canonical NF- κ B signaling, as we have demonstrated with Nv-TLR-dependent activation of NF- κ B in 293 cells (Brennan et al., 2017). However, cnidarian NF- κ B pathways appear to consist of elements that resemble both canonical and non-canonical NF- κ B pathways (Wolenski et al. 2011; Mansfield et al. 2017; Williams et al. 2018). Therefore, the molecular mechanism(s) by which Of-TLR activates Of-NF- κ B signaling in corals remains unclear.

2. *Are basal organisms constitutively processing NF- κ B, i.e., is NF- κ B constitutively activated in basal organisms?* In data from our lab and others, it has become apparent that the activation of NF- κ B in basal organisms may be different than what is seen in higher organisms. For example, in *Aiptasia* (Mansfield et al. 2017), as well as in one coral (Mansfield et al. 2019) and one sponge described here (Fig. 4.5), whole-cell extracts and immunofluorescence staining of tissue from basal organisms reveal that most of the endogenous NF- κ B protein in these animals is in its processed form and primarily in the nucleus (Mansfield et al., 2017; Figure 4.5). Exogenous FLAG-Co-NF- κ B is also primarily in the nucleus when overexpressed in *Capsaspora* (Figures 5.5-5.6). Furthermore, when undergoing an apparent immunological response (i.e., bleaching or stimulating with LPS), the amount of NF- κ B protein increases in a sea anemone and a sponge (Mansfield et al., 2017; Figure 4.7). Additionally, when *O. faveolata* is stimulated with *E. coli*-derived LPS, the overall expression of NF- κ B pathway members increases (Table 3.2). Taken together, the basal expression and activation of NF- κ B in basal organisms seems quite different from what is seen in higher animals. That is, in vertebrates, the activation of NF- κ B is carried out by a multi-component signal transduction process in response to a stimulus, which is followed by nuclear translocation of NF- κ B to activate its target genes. In contrast, it appears that NF- κ B in some, if not many, basal organisms, is constitutively processed and in the nucleus. Whether this phenomenon is due to a continual exposure to pathogens, a naturally heightened immunity, or some other reason remains to be determined.

3. *What mechanism may cause the constitutive processing of basal NF- κ B proteins?* If basal organisms constitutively process NF- κ B, this leads to further questions: 1) What is the mechanism by which this constitutive processing occurs, and 2) in what states are basal organisms *not* constitutively activating NF- κ B? It is possible that the activation and control of NF- κ B is carried out on a transcriptional level, rather than a processing level. For example, the expression of NF- κ B pathway members in *O. favoelata*, rather than the nuclear translocation of NF- κ B, occurs after exposure to a stimulus. Another possibility is that alternative splicing to generate different forms of the NF- κ B protein occurs during mRNA expression. Furthermore, most basal organisms encode a full-length bipartite (RHD-ANK) NF- κ B protein. In what setting then, does the organism express a full-length non-processed NF- κ B protein? Would constitutive processing occur in a completely sterile setting, or a setting that lacks certain external cues? At the present time, it is not known what promotes or inhibits the constitutive NF- κ B processing that occurs in most basal organisms that have been studied.

4. *What are the biological functions of NF- κ B in protists such as Capsaspora and choanoflagellates?* Although the data presented here suggest that NF- κ B plays a role in immunity and/or development in *Capsaspora*, there is no evidence to date that this is the actual biological role of NF- κ B. In fact, very little empirical data exist for the role of NF- κ B in organisms basal to flies. Therefore, future research should be directed at the study of the biological outcomes of basal NF- κ B. For example, the actual function of *Capsaspora* NF- κ B might be quite different from what is seen in multicellular animals. All previous data on NF- κ B has been in the context of a multicellular organism that

contains an immune system. How does NF- κ B work in the context of a single-celled organism? Perhaps NF- κ B plays a larger role in not only the immunological state of the single cell, but also to initiate signaling pathways to communicate between other single-celled organisms or other organisms that they inhabit. The GO enrichment analysis performed on *Capsaspora* RNA-seq data suggest that NF- κ B plays roles that have to do with signaling between organisms. For example, other top hits for the genes that co-migrated with NF- κ B were “Multi-organism process,” “Interspecies interaction between organisms,” and “Multicellular organismal process” (Figure 5.8). Indeed, *Capsaspora* naturally lives in the hemolymph of the snail *Biomphalaria glabrata* and thus, NF- κ B may regulate a symbiont-host interaction.

In choanoflagellates, the role of NF- κ B is unclear. No data to date have suggested a biological role for NF- κ B in choanoflagellates. However, these organisms are more closely related to animals than *Capsaspora*, contain multiple NF- κ B-like proteins, sometimes live colonially, and contain homologs to upstream members of the NF- κ B pathway (Richter et al., 2018). Therefore, the choanoflagellate NF- κ B proteins might play more conserved roles to animals than *Capsaspora*, i.e., related to immunity or development. Then again, choanoflagellates have undergone incredible divergence alongside the evolution of animals, including undergoing their own gene duplications and loss of NF- κ B. Further research should be directed at the functional understanding of the multiple NF- κ B protein sequences present in the transcriptomes of these organisms, and why NF- κ B has been lost in some lineages and conserved/duplicated in others.

5. *Does NF- κ B signaling actually have a role in immunity in basal organisms?*

Perhaps one of the most important implications of this thesis is that aspects of vertebrate innate immunity, specifically TLR-to-NF- κ B and NF- κ B signaling pathways, are present and are, functional, in basal phyla. As discussed in Chapter 6 (section 6.2.4), TLR-directed activation of NF- κ B has not yet been directly demonstrated in any cnidarian system, or basal system, to date. While MyD88-deficient *Hydra* show increased sensitivity to bacterial infection (Franzenburg et al. 2012), and reconstituted Nv-TLR in HEK 293T cells activates NF- κ B pathway (Brennan et al., 2017), studies have yet to directly implicate a NF- κ B signaling pathway in basal immunity. Furthermore, as discussed in section 6.3.7, in organisms like the sponge, there are no clear homologs to TLR, yet NF- κ B activation in response to a TLR-ligand (namely, *E. coli* LPS) appears to occur.

To directly address whether TLR-to-NF- κ B and NF- κ B signaling has a role in basal immunity, ongoing efforts should be directed at genetically manipulating (e.g., by CRISPR/Cas9) NF- κ B signaling in basal organisms. Currently, we are attempting to accomplish this in the model organism *N. vectensis* by generating Nv-NF- κ B knockout animals and testing their susceptibility to known pathogens, among other biological processes. Future experiments should take advantage of the increased genomic and transcriptomic sequencing efforts of basal organisms, and attempt genetic manipulation of NF- κ B in these organisms. One opportunity for genetically modifying NF- κ B would be using *Capsaspora*, since it is an amenable organism in several ways: the genome has an accurate sequence, it can grow easily in cell culture, it can be transfected, albeit at low

efficiency, and presumably transfected cells could be selected in culture and analyzed for genome modifications.

7. What are the important downstream effector genes of NF- κ B in basal organisms? To date, there is no direct evidence of target genes of NF- κ B in basal organisms. However, some efforts have been made. For example, in cnidarians, one gene has been suggested as a target gene of Nv-NF- κ B: Nv-TRAF3, in that it contains three κ B binding sites in its promoter and has been tested using luciferase assays. It was found that the reporter assay showed activation with Nv-NF- κ B, and reduced activation when the κ B binding sites were removed (Brennan et al., 2017). Nevertheless, no empirical evidence has been done beyond this work, and in no other basal organisms have any target genes been identified. In this thesis, we identified two potential target genes of Co-NF- κ B, and ongoing work is attempting to clone the putative promoters of these genes to assay for NF- κ B-responsiveness in reporter assays. SYK and SRC are known to be immune-related genes in higher organisms, but whether they are target genes in *Capsaspora* remain to be seen. Ideally, performing ChIP-seq using a sensitive antibody for NF- κ B would identify target genes of NF- κ B and help elucidate the biological roles of basal NF- κ Bs. These findings would validate the importance of comparative studies of immune signaling pathways in basal organisms, as has been performed in this thesis.

6.6 Conclusions and perspectives

Since its discovery, NF- κ B has been studied as a prominent player in many biological processes in vertebrate and fly systems. Nonetheless, until a decade ago, NF-

κ B was not known to be present in the genomes of organisms basal to insects. The discovery of NF- κ Bs in single-celled protists, and the data gathered from this dissertation has led to improved hypotheses for how and why this transcription factor evolved (Figure 6.1). That is, we propose that prior to the rise of holozoan life, ANK repeats were present in bacteria and archaea genomes. Some ancestral organism developed a primordial RHD-only protein that eventually fused to an ANK repeat protein, resulting in the modern-day bipartite NF- κ B protein. RHDs then diversified in choanoflagellates, while the metazoan lineage retained the full-length NF- κ B fusion protein and also developed separate NF- κ B/Rel and I κ B proteins, likely from gene-splitting and duplication events (Figure 6.1).

Overall, our results show that NF- κ B is generally structurally and functionally conserved in all animals, but lacks certain attributes in earlier branching organisms. That is, the most basal phyla of animals, sponges and corals, contain single NF- κ B proteins that have conserved, functional sequences, but earlier branching organisms like *Capsaspora* lack residues important for processing seen in higher organisms. Furthermore, the diversified group of choanoflagellates has exhibited gains and losses of NF- κ B. Several studies, including the data from this dissertation, have now demonstrated that the NF- κ B proteins of evolutionarily basal organisms have many of the same structural features, activities, modes of regulation, and biological effects that are found in the expanded set of NF- κ B and Rel proteins of more complex organisms. For example, ANK repeat regulation of the NF- κ B RHD is even found in protists, and NF- κ B is likely to have dual roles in development and immunity in organisms ranging from sponges to humans. Nevertheless, there are clear differences in basal NF- κ Bs. Particularly intriguing

is the finding that NF- κ B proteins appear to be constitutively processed and active in some sponges, anemones, and perhaps *Capsaspora*, and that induced activation of the NF- κ B pathway may be at the transcriptional level in some of the same basal organisms. It is interesting to speculate that continuous exposure to pathogens or other activating ligands induces NF- κ B processing in these situations and results in a constant heightened immunological state. We note that this concept of heightened immunity has been suggested in more complex organisms as well; for example, some species of bats appear to have constitutively active immune systems due to their microbe-rich habitats (Banerjee et al., 2017; Brook et al., 2020).

Overall, the research gathered in this dissertation contributes to our understanding of basal immunity. There are a number of proposed effector and immune regulatory molecules that have been identified in invertebrates (Table 6.1), but it is clear that research on the evolution of NF- κ B has only scratched the surface of what is yet to be discovered. Future studies will likely seek to identify the most primitive biological processes controlled by NF- κ B in protists. Moreover, the identification of NF- κ B target genes and the development of genetic systems to study gene function will certainly reveal much about the function of NF- κ B in cnidarians and sponges. In particular, the identification of the immune response gene targets of NF- κ B in basal organisms may provide insights into novel antimicrobials, as well as reveal information about the molecular processes underlying global ecological crises of marine invertebrates, such as coral bleaching and microbial pathogenesis. Looking backwards at the evolutionary history of NF- κ B will no doubt bring further knowledge to our understanding of extant

immunological systems, especially in organisms that are environmentally sensitive and that are situated at the base of multicellular life.

Table 6.1 Proposed immunity-related effector and regulatory molecules in basal organisms.

Effector molecules	Reference
Anti-microbial peptides	Tincu and Taylor, 2004; Margolis et al., 2021
Toll-like receptors	Brennan and Gilmore, 2018
Interleukin receptors	Huang et al., 2015
Complement	Nonaka and Kimura, 2006
Cytokines	Melillo et al., 2018
Lectins	Ocampo et al., 2015
cGAS/STING	Brennan and Gilmore, 2018
Scavenger receptors	Ocampo et al., 2015; Neubauer et al., 2016
NOD-like receptors	Yuen et al., 2014; Melillo et al., 2018; Parisi et al., 2020
Nitric Oxide (NO)	Parisi et al., 2020; Mansfield et al., 2019
TRAF	Anderson et al., 2016; Brennan and Gilmore, 2018; Seneca et al., 2020; Parisi et al., 2020
MYD88	Brennan and Gilmore, 2018; Williams et al., 2018
Regulatory molecules	Reference
NF- κ B	Wolenski and Gilmore, 2012; Mansfield and Gilmore, 2019; Williams et al., 2018, 2020, and 2021; Williams and Gilmore, 2020
IRF	Mansfield and Gilmore, 2019; Seneca et al., 2020
NFAT	Collins and Meyer, 2012
TGF β	Detournay et al., 2012
AP-1	Traylor-Knowles et al., 2017; Mansfield and Gilmore, 2019
Caspase-3	Mansfield and Gilmore, 2019
Elk	Mansfield and Gilmore, 2019

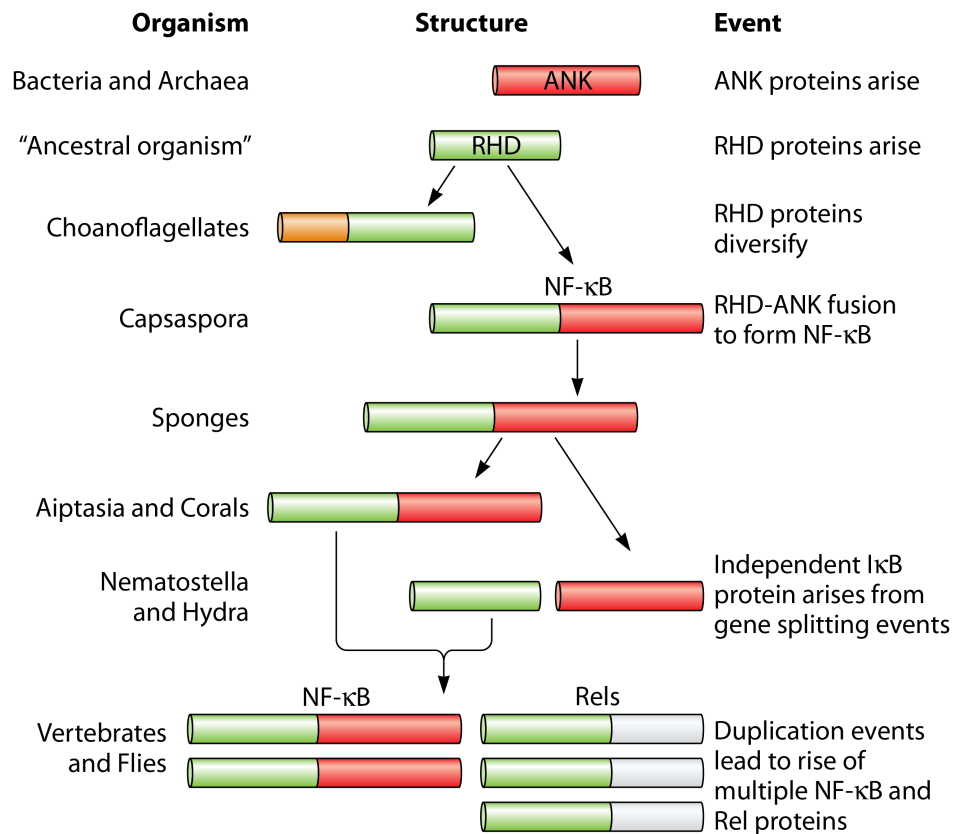


Figure 6.1 The Evolution of NF-κB.

Prior to the rise of holozoan life, Ankyrin repeats (ANK, red) were present in bacteria and archaea. The appearance of an RHD-containing protein (green) likely led to RHD-only proteins that diversified in choanoflagellates and to an RHD-ANK fusion in an ancestor of Capsaspora. Metazoans generally retained the full-length NF-κB, although in some cnidarians (e.g., *Hydra* and *Nematostella*) there were gene splitting events that created separate RHD and ANK repeat proteins. Eventually a series of duplication events gave rise to the multiple NF-κB and Rel proteins that are present in vertebrates and flies.

BIBLIOGRAPHY

- Adamska, M., Degnan, S.M., Green, K.M., Adamski, M., Craigie, A., Larroux, C., Degnan, B.M. 2007. Wnt and TGF- β Expression in the sponge *Amphimedon queenslandica* and the origin of metazoan embryonic patterning. PLoS One 2, e1031. <https://doi.org/10.1371/journal.pone.0001031>
- Aderem, A., Ulevitch, R.J. 2000. Toll-like receptors in the induction of the innate immune response. Nature 406, 782-787. doi:<http://dx.doi.org/10.1038/35021228>
- Aguirre Carrión, P.J., Williams, L.M., Gilmore, T.D. 2021. Molecular and biochemical approaches to study the evolution of NF- κ B signaling in basal metazoans. Methods in Molecular Biology, *in press*.
- Akira, S., Uematsu, S., Takeuchi, O. 2006. Pathogen recognition and innate immunity. Cell 124, 783–801. doi:10.1016/j.cell.2006.02.015
- Al-Khodori, S., Price, C.T., Kalia, A., Abu Kwaik, Y. 2010. Functional diversity of ankyrin repeats in microbial proteins. Trends in Microbiology 18, 132–139.
- Alshanbayeva, A., Thomas, A., Tremblay, M., Abbas, M., Abdurrob, F., Almojel, T., Aparicio, A., Asarpota, D., Ayers, A., Aziz, A., Bishop, J., Christie, T., Chua, M.J., Melissa, Chung, O., Dhar, N., Diedrich, A., Fortin, C., He, Q., Heerboth, S., Hok, R., Khedkar, A., Kitchloo, S., Lawlor, C., Leonard, B., Linderman, S., Maloyan, M., Miller, L., Pak, C., Pandita, A., Park, I., Patel, N., Ramachandran, J., Reynoso, M., Samaha, Y., Thole, G., Turnbull, J., Xia, L., Zhu, J., Navarro, C., Gilmore, T.D. 2015. N- and C-terminal non-conserved residues contribute to transactivation by a sea anemone (*Nematostella vectensis*) NF- κ B transcription factor. BIOS 86, 165–175. <https://doi.org/10.1893/0005-3155-86.4.165>
- Anderson, D.A., Walz, M.E., Weil, E., Tonellato, P., Smith, M.C. 2016. RNA-Seq of the Caribbean reef-building coral *Orbicella faveolata* (Scleractinia-Merulinidae) under bleaching and disease stress expands models of coral innate immunity. PeerJ 4, e1616.
- Babonis, L.S., Martindale, M.Q., Ryan, J.F. 2016. Do novel genes drive morphological novelty? An investigation of the nematosomes in the sea anemone *Nematostella vectensis*. BMC Evolutionary Biology 16, 114.
- Babonis, L.S., Martindale, M.Q. 2014. Old cell, new trick? Cnidocytes as a model for the evolution of novelty. Integrated and Comparative Biology 54, 714–722.
- Bailey, T.L., Williams, N., Mistleh, C., Li, W.W. 2006. MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Research. 34, W369–W373. doi:10.1093/nar/gkl198

- Banerjee, A., Rapin, N., Bollinger, T., Misra, V. 2017. Lack of inflammatory gene expression in bats: a unique role for a transcription repressor. *Scientific Reports* 7, 2232.
- Baumgarten, S., Simakov, O., Esherrick, L.Y., Liew, Y.J., Lehnert, E.M., Michell, C.T., Li, Y., Hambleton, E.A., Guse, A., Oates, M.E., Gough, J., Weis, V.M., Aranda, M., Pringle, J.R., Voolstra, C.R. 2015. The genome of *Aiptasia*, a sea anemone model for coral symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* 112, 11893–11898.
- Böhm, M., Hentschel, U., Friedrich, A., Fieseler, L., Steffen, R., Gamulin, V., Müller, I., Müller, W. 2001. Molecular response of the sponge *Suberites domuncula* to bacterial infection. *Marine Biology* 139, 1037–1045. <https://doi.org/10.1007/s002270100656>
- Bosch, T.C.G., Augustin, R., Anton-Erxleben, F., Fraune, S., Hemmrich, G., Zill, H., Rosenstiel, P., Jacobs, G., Schreiber, S., Leippe, M., Stanisak, M., Grötzinger, J., Jung, S., Podschun, R., Bartels, J., Harder, J., Schröder, J.-M. 2009. Uncovering the evolutionary history of innate immunity: the simple metazoan *Hydra* uses epithelial cells for host defense. *Developmental and Comparative Immunology* 33, 559–569. doi:10.1016/j.dci.2008.10.004
- Bottero, V., Withoff, S., Verma, I.M. 2006. NF- κ B and the regulation of hematopoiesis. *Cell Death & Differentiation* 13, 785–797. <https://doi.org/10.1038/sj.cdd.4401888>
- Brasch K, Ochs RL. 1992. Nuclear bodies (NBs): a newly "rediscovered" organelle. *Experimental Cell Research* 202, 211-223. doi: 10.1016/0014-4827(92)90068-j. PMID: 1397076.
- Brennan, J.J., Messerschmidt, J.L., Williams, L.M., Matthews, B.J., Reynoso, M., Gilmore, T.D. 2017. Sea anemone model has a single Toll-like receptor that can function in pathogen detection, NF- κ B signal transduction, and development. *Proceedings of the National Academy of Sciences of the United States of America* 114, E10122-E10131.
- Brennan, J.J. 2018. Role of a Sea Anemone (*Nematostella vectensis*) Toll-Like Receptor in Pathogen Detection, Development, and Activation of NF-Kappab Signaling. PhD dissertation, Boston University. <https://open.bu.edu/handle/2144/33241>
- Brennan, J.J., Gilmore, T.D. 2018. Evolutionary origins of Toll-like receptor signaling. *Molecular Biology and Evolution* 35, 1576-1587.
- Brook, C.E., Boots, M., Chandran, K., Dobson, A.P., Drosten, C., Graham, A.L., Grenfell, B.T., Müller, M.A., Ng, M., Wang, L.F., van Leeuwen, A. 2020.

- Accelerated viral dynamics in bat cell lines, with implications for zoonotic emergence. *eLife* 9, e48401.
- Brown, B.E. 1997. Coral bleaching: causes and consequences. *Coral Reefs* 16, S129–S138. doi:10.1007/s003380050249
- Buckley, K.M., Rast, J.P. 2012. Dynamic evolution of Toll-like receptor multigene families in echinoderms. *Frontiers in Immunology* 3, 136. doi:10.3389/fimmu.2012.00136
- Burns, J.A., Zhang, H., Hill, E., Kim, E., Kerney, R. 2017. Transcriptome analysis illuminates the nature of the intracellular interaction in a vertebrate-algal symbiosis. *eLife* 6, e22054.
- Buscema, M., De Sutter, D., Van de Vyver, G. 1980. Ultrastructural study of differentiation processes during aggregation of purified sponge archaeocytes. *Wilhelm Roux's Archives of Developmental Biology* 188, 45–53. <https://doi.org/10.1007/BF00848609>
- Chapman, J.A., Kirkness, E.F., Simakov, O., Hampson, S.E., Mitros, T., Weinmaier, T., Rattei, T., Balasubramanian, P.G., Borman, J., Busam, D., Disbennett, K., Pfannkoch, C., Sumin, N., Sutton, G.G., Viswanathan, L.D., Walenz, B., Goodstein, D.M., Hellsten, U., Kawashima, T., Prochnik, S.E., Putnam, N.H., Shu, S., Blumberg, B., Dana, C.E., Gee, L., Kibler, D.F., Law, L., Lindgens, D., Martinez, D.E., Peng, J., Wigge, P.A., Bertulat, B., Guder, C., Nakamura, Y., Ozbek, S., Watanabe, H., Khalturin, K., Hemmrich, G., Franke, A., Augustin, R., Fraune, S., Hayakawa, E., Hayakawa, S., Hirose, M., Hwang, J.S., Ikeo, K., Nishimiya-Fujisawa, C., Ogura, A., Takahashi, T., Steinmetz, P.R.H., Zhang, X., Aufschnaiter, R., Eder, M.-K., Gorny, A.-K., Salvenmoser, W., Heimberg, A.M., Wheeler, B.M., Peterson, K.J., Böttger, A., Tischler, P., Wolf, A., Gojobori, T., Remington, K.A., Strausberg, R.L., Venter, J.C., Technau, U., Hobmayer, B., Bosch, T.C.G., Holstein, T.W., Fujisawa, T., Bode, H.R., David, C.N., Rokhsar, D.S., Steele, R.E. 2010. The dynamic genome of *Hydra*. *Nature* 464, 592–596.
- Collins, S. R., Meyer, T. 2011. Evolutionary origins of STIM1 and STIM2 within ancient Ca²⁺ signaling systems. *Trends in Cell Biology*, 21, 202–211. <https://doi.org/10.1016/j.tcb.2011.01.002>
- Cunning, R., Bay, R.A., Gillette, P., Baker, A.C., Traylor-Knowles, N. 2018. Comparative analysis of the *Pocillopora damicornis* genome highlights role of immune system in coral evolution. *Scientific Reports* 8, 1–10.
- Davy, S.K., Allemand, D., Weis, V.M. 2012. Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiology and Molecular Biology Reviews* 76, 229–261. doi:10.1128/MMBR.05014-11

- Degnan, B.M., Adamska, M., Craigie, A., Degnan, S.M., Fahey, B., Gauthier, M., Hooper, J.N.A., Larroux, C., Leys, S.P., Lovas, E., Richards, G.S. 2008. The demosponge *Amphimedon queenslandica*: reconstructing the ancestral metazoan genome and deciphering the origin of animal multicellularity. Cold Spring Harbor Protocols 2008, pdb.emo108. <https://doi.org/10.1101/pdb.emo108>
- Degnan, S.M. 2015. The surprisingly complex immune gene repertoire of a simple sponge, exemplified by the NLR genes: A capacity for specificity? Developmental and Comparative Immunology 48, 269–274. <https://doi.org/10.1016/j.dci.2014.07.012>
- DeSalvo, M.K., Sunagawa, S., Voolstra, C.R., Medina, M. 2010. Transcriptomic responses to heat stress and bleaching in the elkhorn coral *Acropora palmata*. Marine Ecological Progress Series 17, 3952-3971.
- Detournay O, Schnitzler CE, Poole A, Weis VM. 2012. Regulation of cnidarian-dinoflagellate mutualisms: Evidence that activation of a host TGF β innate immune pathway promotes tolerance of the symbiont. Developmental and Comparative Immunology 38, 525-537. doi: 10.1016/j.dci.2012.08.008.
- Fairclough, S.R., Chen, Z., Kramer, E., Zeng, Q., Young, S., Robertson, H.M., Begovic, E., Richter, D.J., Russ, C., Westbrook, M.J., Manning, G., Lang, B.F., Haas, B., Nusbaum, C., King, N. 2013. Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. Genome Biology 14, R15.
- Ferrer-Bonet, M., Ruiz-Trillo, I. 2017. *Capsaspora owczarzaki*. Current Biology 27, PR829-PR830).
- Finnerty, J.R., Gilmore, T.D. 2015. Methods for analyzing the evolutionary relationship of NF- κ B proteins using free, web-driven bioinformatics and phylogenetic tools. Methods in Molecular Biology 1280, 631–646. doi:10.1007/978-1-4939-2422-6_37
- Fischer, A.H., Mozzherin, D., Eren, A.M., Lans, K.D., Wilson, N., Cosentino, C., Smith, J. 2014. SeaBase: a multispecies transcriptomic resource and platform for gene network interface. Integrative and Comparative Biology 54, 250-263.
- Fort, P., Kajava, A.V., Delsuc, F., Coux, O. 2015. Evolution of proteasome regulators in eukaryotes. Genome Biology and Evolution 7, 1363-1379.
- Franzenburg, S., Fraune, S., Künzel, S., Baines, J.F., Domazet-Lošo, T., Bosch, T.C.G. 2012. MyD88-deficient *Hydra* reveal an ancient function of TLR signaling in sensing bacterial colonizers. Proceedings of the National Academy of Sciences of the United States of America 109, 19374–19379.

- Fuess, L.E., Pinzón C, J>H., Weil, E., Grinshpon, R.D., Mydlarz, L.D. 2017. Life or death: disease-tolerant coral species activate autophagy following immune challenge. *Proceedings of the Royal Society. B: Biology* 284, 20170771.
- Fuess, L.E., Pinzón C, J.H., Weil, E., Mydlarz, L.D. 2016. Associations between transcriptional changes and protein phenotypes provide insights into immune regulation in corals. *Developmental and Comparative Immunology* 62, 17–28. doi:10.1016/j.dci.2016.04.017
- Gacesa, R., Chung, R., Dunn, S.R., Weston, A.J., Jaimes-Becerra, A., Marques, A.C., Morandini, A.C., Hranueli, D., Starcevic, A., Ward, M., Long, P.F. 2015. Gene duplications are extensive and contribute significantly to the toxic proteome of nematocysts isolated from *Acropora digitifera* (Cnidaria: Anthozoa: Scleractinia). *BMC Genomics* 16, 774.
- Garbati, M.R., Alço, G., Gilmore, T.D. 2010. Histone acetyltransferase p300 is a coactivator for transcription factor REL and is C-terminally truncated in the human diffuse large B-cell lymphoma cell line RC-K8. *Cancer Letters* 291, 237–245. doi:10.1016/j.canlet.2009.10.018
- Gauthier, A.E., Chandler, C.E., Poli, V., Gardner, F.M., Tekiau, A., Smith, R., Bonham, K.S., Cordes, E.E., Shank, T.M., Zaroni, I., Goodlett, D.R., Biller, S.J., Ernst, R.K., Rotjan, R.D., Kagan, J.C. 2021. Deep-sea microbes as tools to refine the rules of innate immune pattern recognition. *Science Immunology* 6, eabe0531. doi: 10.1126/sciimmunol.abe0531.
- Gauthier, M., Degnan, B.M. 2008. The transcription factor NF- κ B in the demosponge *Amphimedon queenslandica*: insights on the evolutionary origin of the Rel homology domain. *Developmental Genes and Evolution* 218, 23–32.
- Gauthier, M.E.A., Du Pasquier, L., Degnan, B.M. 2010. The genome of the sponge *Amphimedon queenslandica* provides new perspectives into the origin of Toll-like and interleukin 1 receptor pathways. *Evolution and Development* 12, 519–533. doi:10.1111/j.1525- 142X.2010.00436.x
- Ghosh, S., Gifford, A.M., Riviere, L.R., Tempst, P., Nolan, G.P., Baltimore, D. 1990. Cloning of the p50 DNA binding subunit of NF- κ B: homology to rel and dorsal. *Cell* 62, 1019–1029.
- Ghosh, S., Hayden, M.S. 2012. Celebrating 25 years of NF- κ B research. *Immunological Reviews* 246, 5–13.
- Gilmore, T.D. 1990. NF- κ B, KBF1, dorsal, and related matters. *Cell* 62, 841–843.

- Gilmore, T.D. 2006. Introduction to NF- κ B: players, pathways, perspectives. *Oncogene* 2, 6680–6684.
- Gilmore, T.D., Temin, H.M. 1986. Different localization of the product of the *v-rel* oncogene in chicken fibroblasts and spleen cells correlates with transformation by REV-T. *Cell* 44, 791–800.
- Gilmore, T.D., Wolenski, F.S. 2012. NF- κ B: where did it come from and why? *Immunological Reviews* 246, 14–35.
- Gleason, D.F., Wellington, G.M. 1993. Ultraviolet radiation and coral bleaching. *Nature* 365, 836–838. doi:10.1038/365836a0,
- Gold, D.A., Katsuki, T., Li, Y., Yan, X., Regulski, M., Ibberson, D., Holstein, T., Steele, R.E., Jacobs, D.K., Greenspan, R.J. 2019. The genome of the jellyfish *Aurelia* and the evolution of animal complexity. *Nature Ecology & Evolution* 3, 96–104.
- Guttridge, D.C., Albanese, C., Reuther, J.Y., Pestell, R.G., Baldwin, A.S. 1999. NF- κ B controls cell growth and differentiation through transcriptional regulation of cyclin D1. *Molecular and Cellular Biology* 19, 5785–5799.
- Haery, L. 2016. Effects of C-terminal truncations of the histone acetyltransferase P300 on the growth and gene expression patterns of human diffuse large B-cell lymphoma cell lines. PhD Dissertation, Boston University.
<https://open.bu.edu/handle/2144/14537>
- Hayden, M.S., Ghosh, S. 2008. Shared principles in NF- κ B signaling. *Cell* 132, 344–362.
- Hayden, M.S., Ghosh, S. 2004. Signaling to NF- κ B. *Genes & Development* 18, 2195–2224. doi:10.1101/gad.1228704
- Helm, R.R., Siebert, S., Tulin, S., Smith, J., Dunn, C.W. 2013. Characterization of differential transcript abundance through time during *Nematostella vectensis* development. *BMC Genomics* 14, 266.
- Hoegh-Guldberg, O., Mumby, P.J., Hooten, A.J., Steneck, R.S., Greenfield, P., Gomez, E., Harvell, C.D., Sale, P.F., Edwards, A.J., Caldeira, K., Knowlton, N., Eakin, C.M., Iglesias-Prieto, R., Muthiga, N., Bradbury, R.H., Dubi, A., Hatzioelos, M.E. 2007. Coral reefs under rapid climate change and ocean acidification. *Science* 318, 1737–1742. doi:10.1126/science.1152509
- Hoffmeyer, T.T., Burkhardt, P. 2016. Choanoflagellate models - *Monosiga brevicollis* and *Salpingoeca rosetta*. *Current Opinion in Genetics and Development* 39, 42–47.

- Huang X-D, Zhang, H, He, M-X. 2015. Comparative and evolutionary analysis of the interleukin 17 gene family in invertebrates. PLoS One 10, e0132802. <https://doi.org/10.1371/journal.pone.0132802>
- Hughes, T.P., Kerry, J.T., Álvarez-Noriega, M., Álvarez-Romero, J.G., Anderson, K.D., Baird, A.H., Babcock, R.C., Beger, M., Bellwood, D.R., Berkelmans, R., Bridge, T.C., Butler, I.R., Byrne, M., Cantin, N.E., Comeau, S., Connolly, S.R., Cumming, G.S., Dalton, S.J., Diaz-Pulido, G., Eakin, C.M., Figueira, W.F., Gilmour, J.P., Harrison, H.B., Heron, S.F., Hoey, A.S., Hobbs, J.-P.A., Hoogenboom, M.O., Kennedy, E.V., Kuo, C.-Y., Lough, J.M., Lowe, R.J., Liu, G., McCulloch, M.T., Malcolm, H.A., McWilliam, M.J., Pandolfi, J.M., Pears, R.J., Pratchett, M.S., Schoepf, V., Simpson, T., Skirving, W.J., Sommer, B., Torda, G., Wachenfeld, D.R., Willis, B.L., Wilson, S.K. 2017. Global warming and recurrent mass bleaching of corals. *Nature* 543, 373–377. doi:10.1038/nature21707
- Jacobovitz, M.R., Rupp, S., Voss, P.A., Gornik, S.G., Guse, A. 2019. Dinoflagellate symbionts escape vomocytosis by host cell immune suppression. bioRxiv, <https://doi.org/10.1101/864579>.
- Jernigan, K.K., Bordenstein, S.R. 2014. Ankyrin domains across the Tree of Life. *PeerJ* 2, e264.
- Kappler, C., Meister, M., Lagueux, M., Gateff, E. Hoffmann, J.A., Reichart, J.M. 1993. Insect immunity. Two 17 bp repeats nesting a kappa B-related sequence confer inducibility to the dipterin gene and bind a polypeptide in bacteria-challenged *Drosophila*. *EMBO Journal* 12, 1561–1568.
- Kawai, T., Akira, S. 2007. Signaling to NF- κ B by Toll-like receptors. *Trends in Molecular Medicine* 13, 460–469. doi:10.1016/j.molmed.2007.09.002
- Kieran, M., Blank, V., Logeat, F., Vandekerckhove, J., Lottspeich, F., Le Bail, O., Urban, M.B., Kourilsky, P., Baeuerle, P.A., Israël, A. 1990. The DNA binding subunit of NF- κ B is identical to factor KBF1 and homologous to the *rel* oncogene product. *Cell* 62, 1007–1018.
- King, N., 2005. Choanoflagellates. *Current Biology* 15, R113-R114.
- King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., Marr, M., Pincus, D., Putnam, N., Rokas, A., Wright, K.J., Zuzow, R., Dirks, W., Good, M., Goodstein, D., Lemons, D., Li, W., Lyons, J.B., Morris, A., Nichols, S., Richter, D.J., Salamov, A., JGI Sequencing, Bork, P., Lim, W.A., Manning, G., Miller, W.T., McGinnis, W., Shapiro, H., Tjian, R., Grigorie, I.V., Rokhsar, D. 2008. The genome of the

- choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451, 783–788.
- Kushmaro, A., Loya, Y., Fine, M., Rosenberg, E. 1996. Bacterial infection and coral bleaching. *Nature* 380, 396. doi:10.1038/380396a0
- Larroux, C., Fahey, B., Liubicich, D., Hinman, V.F., Gauthier, M., Gongora, M., Green, K., Wörheide, G., Leys, S.P., Degnan, B.M. 2006. Developmental expression of transcription factor genes in a demosponge: insights into the origin of metazoan multicellularity. *Evolution and Development* 8, 150–173. <https://doi.org/10.1111/j.1525-142X.2006.00086.x>
- Lin, L., DeMartino, G.N., Greene, W.C. 1998. Cotranslational biogenesis of NF- κ B p50 by the 26S proteasome. *Cell* 92:819–828.
- Lowell C. A. 2011. Src-family and Syk kinases in activating and inhibitory pathways in innate immune cells: signaling cross talk. *Cold Spring Harbor Perspectives in Biology*, 3, a002352. <https://doi.org/10.1101/cshperspect.a002352>
- Mansfield, K.M., Carter, N.M., Nguyen, L., Cleves, P.A., Alshanbayeva, A., Williams, L.M., Penvose, A.R., Crowder, C., Finnerty, J.R., Siggers, T.W., Weis, V.M., Gilmore, T.D. 2017. Transcription factor NF- κ B is modulated by symbiotic status in a sea anemone model of cnidarian bleaching. *Scientific Reports* 7, 16025.
- Mansfield, K.M., Cleves, P.A., Vlack, E.V., Kriefall, N.G., Benson, B.E., Camacho, D.J., Hemond, O., Pedroza, M., Siggers, T., Pringle, J.R., Davies, S.W., Gilmore, T.D. 2019. Varied effects of algal symbionts on transcription factor NF- κ B in a sea anemone and a coral: possible roles in symbiosis and thermotolerance. *bioRxiv*, <https://doi.org/10.1101/640177>.
- Mansfield, K.M., Gilmore, T.D. 2019. Innate immunity and cnidarian-Symbiodiniaceae mutualism. *Developmental and Comparative Immunology* 90, 199–209.
- Margolis, S.R., Dietzen, P.A., Hayes, B.M., Wilson, S.C., Remick, B.C., Chou, S., Vance, R.E. 2021. The STING ligand 2'3'-cGAMP induces an NF- κ B-dependent anti-bacterial innate immune response in the starlet sea anemone *Nematostella vectensis*. *bioRxiv*, <https://doi.org/10.1101/2021.05.13.443009>
- Melillo, D., Marino, R., Italiani, P., Boraschi, D. 2018. Innate immune memory in invertebrate metazoans: a critical appraisal. *Frontiers in Immunology*, 9, 1915. <https://doi.org/10.3389/fimmu.2018.01915>
- Miller, D.J., Hemmrich, G., Ball, E.E., Hayward, D.C., Khalturin, K., Funayama, N., Agata, K., Bosch, T.C.G. 2007. The innate immune repertoire in cnidaria--

- ancestral complexity and stochastic gene loss. *Genome Biology* 8, R59.
doi:10.1186/gb-2007-8-4-r59
- Minakhina, S., Steward, R. 2006. Nuclear factor-kappa B pathways in *Drosophila*.
Oncogene 25, 6749–6757. doi:10.1038/sj.onc.1209940
- Mócsai, A., Ruland, J., Tybulewicz, V. (2010). The SYK tyrosine kinase: a crucial player
in diverse biological functions. *Nature Reviews. Immunology* 10, 387–402:
<https://doi.org/10.1038/nri2765>
- Müller, W.E.G., Müller, I.M. 2003. Origin of the metazoan immune system:
identification of the molecules and their functions in sponges. *Integrative and
Comparative Biology* 43, 281–292. <https://doi.org/10.1093/icb/43.2.281>
- Musser, J.M., Schippers, K.J., Nickel, M., Mizzon, G., Kohn, A.B., Pape, C., Hammel,
J.U., Wolf, F., Liang, C., Hernández-Plaza, A., Achim, K., Schieber, N.L.,
Francis, W.R., Sergio Vargas, R., Kling, S., Renkert, M., Feuda, R., Gaspar, I.,
Burkhardt, P., Bork, P., Beck, M., Kreshuk, A., Wörheide, G., Huerta-Cepas, J.,
Schwab, Y., Moroz, L.L., Arendt, D. 2019. Profiling cellular diversity in sponges
informs animal cell type and nervous system evolution. *bioRxiv*,
<https://doi.org/10.1101/758276>.
- Neubauer, E.F., Poole, A.Z., Weis, V.M., Davy, S.K., 2016. The scavenger receptor
repertoire in six cnidarian species and its putative role in cnidarian-dinoflagellate
symbiosis. *PeerJ* 4, e2692.
- Nonaka, M., Kimura, A. 2006. Genomic view of the evolution of the complement
system. *Immunogenetics*, 58, 701–713. <https://doi.org/10.1007/s00251-006-0142-1>
- Ocampo, I.D., Zarate-Potes, A., Pizarro, V., Rojas, C.A., Vera, N.E., Cadavid, L.F.,
2015. The immunotranscriptome of the Caribbean reef-building coral
Pseudodiploria strigosa. *Immunogenetics* 67, 515–530.
- O'Neill, L.A.J., Dunne, A., Edjeback, M., Gray, P., Jefferies, C., Wietek, C. 2003. Mal
and MyD88: adapter proteins involved in signal transduction by Toll-like
receptors. *Journal of Endotoxin Research* 9, 55–59.
doi:10.1179/096805103125001351
- Parra-Acero, H., Ros-Rocher, N., Perez-Posada, A., Kozyczkowska, A., Sánchez-Pons,
N., Nakata, A., Suga, H., Najle, S.R., Ruiz-Trillo, I. 2018. Transfection of
Capsaspora owczarzaki, a close unicellular relative of animals. *Development* 145,
dev162107.

- Pinzón, J.H., Kamel, B., Burge, C.A., Harvell, C.D., Medina, M., Weil, E., Mydlarz, L.D. 2015. Whole transcriptome analysis reveals changes in expression of immune-related genes during and after bleaching in a reef-building coral. *Royal Society Open Science* 2(4), 140214. <https://doi.org/10.1098/rsos.140214>
- Pita, L., Hoepfner, M., Ribes, M., Hentschel, U. 2018. Differential expression of immune receptors in two marine sponges upon exposure to microbial-associated molecular patterns. *Scientific Reports* 8,16081.
- Pita, L., Fraune, S., Hentschel, U. 2016. Emerging sponge models of animal-microbe symbioses. *Frontiers in Microbiology* 7, 2102. <https://doi.org/10.3389/fmicb.2016.02102>
- Pomerantz, J.L., Baltimore, D. 1999. NF-kappaB activation by a signaling complex containing TRAF2, TANK and TBK1, a novel IKK-related kinase. *EMBO Journal* 18, 6694–6704. <https://doi.org/10.1093/emboj/18.23.6694>
- Poole, A.Z., Weis, V.M. 2014. TIR-domain-containing protein repertoire of nine anthozoan species reveals coral-specific expansions and uncharacterized proteins. *Developmental and Comparative Immunology* 46, 480–488. doi:10.1016/j.dci.2014.06.002
- Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V.V., Jurka, J., Genikhovich, G., Grigoriev, I.V., Lucas, S.M., Steele, R.E., Finnerty, J.R., Technau, U., Martindale, M.Q., Rokhsar, D.S. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317, 86-94.
- Reichhart, J.M., Georgel, P., Meister, M., Lemaitre, B., Kappler, C., Hoffmann, J.A. 1993. Expression and nuclear translocation of the rel/NF- κ B-related morphogen dorsal during the immune response of *Drosophila*. *Comptes Rendus de l'Académie des Sciences- Series III* 316, 1218–1224.
- Richter, D.J., Fozouni, P., Eisen, M.B., King, N. 2018. Gene family innovation, conservation and loss on the animal stem lineage. *eLife* 7, e34226.
- Riesgo, A., Farrar, N., Windsor, P.J., Giribet, G., Leys, S.P. 2014. The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. *Molecular Biology and Evolution* 31, 1102–1120.
- Roth, S., Stein, D., Nüsslein-Volhard, C. 1989. A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the *Drosophila* embryo. *Cell* 59, 1189–1202.

- Roy, A., Srivastava, M., Saqib, U., Liu, D., Faisal, S.M., Sugathan, S., Bishnoi, S., Baig, M.S. 2016. Potential therapeutic targets for inflammation in Toll-like receptor 4 (TLR4)-mediated signaling pathways. *International Immunopharmacology* 40, 79-99.
- Ryzhakov, G., Teixeira, A., Saliba, D., Blazek, K., Muta, T., Ragoussis, J., Udalova, I.A. 2013. Cross- species analysis reveals evolving and conserved features of the Nuclear factor κ B (NF- κ B) proteins. *Journal of Biological Chemistry* 288, 11546–11554. <https://doi.org/10.1074/jbc.M113.451153>
- Santodomingo, N., Vanhoorne, B., Kelly, M., Hooper, J.N.A., 2012. Global diversity of sponges (Porifera). *PLoS One* 7, e35105. <https://doi.org/10.1371/journal.pone.0035105>
- Savinova, O.V., Hoffmann, A., Ghosh, G. 2009. The Nfkb1 and Nfkb2 proteins p105 and p100 function as the core of high-molecular-weight heterogeneous complexes. *Molecular Cell* 34, 591–602. <https://doi.org/10.1016/j.molcel.2009.04.033>
- Schmitz, J.F., Zimmer, F., Bornberg-Bauer, E. 2016. Mechanisms of transcription factor evolution in Metazoa. *Nucleic Acids Research* 44, 6287–6297. <https://doi.org/10.1093/nar/gkw492>
- Sebé-Pedrós, A., Irimia, M., Del Campo, J., Parra-Acero, H., Russ, C., Nusbaum, C., Biencowe, B.J., Ruiz-Trillo, I. 2013. Regulated aggregative multicellularity in a close unicellular relative of Metazoa. *eLife* 2, e01287.
- Sebé-Pedrós, A., Ballaré, C., Parra-Acero, H., Chiva, C., Tena, J.J., Sabidó, E., Gómez-Skarmeta, J.L., Di Croce, L., Ruiz-Trillo, I. 2016. The dynamic regulatory genome of *Capsaspora* and the origin of animal multicellularity. *Cell* 165, 1224–1237.
- Sebé-Pedrós, A., de Mendoza, A., Lang, B.F., Degnan, B.M., Ruiz-Trillo, I. 2011. Unexpected repertoire of metazoan transcription factors in the unicellular holozoan *Capsaspora owczarzaki*. *Molecular Biology and Evolution* 28, 1241–1254. <https://doi.org/10.1093/molbev/msq309>
- Senftleben, U., Cao, Y., Xiao, G., Greten, F.R., Krähn, G., Bonizzi, G., Chen, Y., Hu, Y., Fong, A., Sun, S.C., Karin, M. 2001. Activation by IKK α of a second, evolutionary conserved, NF- κ B signaling pathway. *Science* 293, 1495–1499. <https://doi.org/10.1126/science.1062677>
- Shinzato, C., Shoguchi, E., Kawashima, T., Hamada, M., Hisata, K., Tanaka, M., Fujie, M., Fujiwara, M., Koyanagi, R., Ikuta, T., Fujiyama, A., Miller, D.J., Satoh, N. 2011. Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* 476, 320–323.

- Siboni, N., Abrego, D., Motti, C.A., Tebben, J., Harder, T. 2014. Gene expression patterns during the early stages of chemically induced larval metamorphosis and settlement of the coral *Acropora millepora*. PLoS One 9, e91082.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7, 539.
- Siggers, T., Chang, A.B., Teixeira, A., Wong, D., Williams, K.J., Ahmed, B., Ragoussis, J., Udalova, I.A., Smale, S.T., Bulyk, M.L. 2011. Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF- κ B family DNA binding. Nature Immunology 13, 95–102. doi:10.1038/ni.2151
- Silverman, N., Maniatis, T. 2001. NF- κ B signaling pathways in mammalian and insect innate immunity. Genes & Development 15, 2321–2342. doi:10.1101/gad.909001
- Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D.J., Di Franco, A., Roure, B., Satoh, N., Quéinnec, É., Ereskovsky, A., Lapébie, P., Corre, E., Delsuc, F., King, N., Wörheide, G., Manuel, M. 2017. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. Current Biology 27, 958–967. <https://doi.org/10.1016/j.cub.2017.02.031>
- Simpson, A.G.B., Slamovits, C.H. Archibald, J.M. 2017. “Protist Diversity and Eukaryote Phylogeny” in *Handbook of the Protists*, J.M. Archibald, *et al.*, Eds. (Springer International Publishing, 2017), pp. 1–21.
- Sogabe, S., Hatleberg, W.L., Kocot, K.M., Say, T.E., Stoupin, D., Roper, K.E., Fernandez-Valverde, S.L., Degnan, S.M., Degnan, B.M. 2019. Pluripotency and the origin of animal multicellularity. Nature 570, 519–522. <https://doi.org/10.1038/s41586-019-1290-4>
- Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E.A., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N.H., Stanke, M., Adamska, M., Darling, A., Degnan, S.M., Oakley, T.H., Plachetzki, D.C., Zhai, Y., Adamski, M., Calcino, A., Cummins, S.F., Goodstein, D.M., Harris, C., Jackson, D.J., Leys, S.P., Shu, S., Woodcroft, B.J., Vervoort, M., Kosik, K.S., Manning, G., Degnan, B.M., Rokhsar, D.S. 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature 466, 720–726.
- Starczynowski, D.T., Trautmann, H., Pott, C., Harder, L., Arnold, N., Leeman, J.R., Siebert, R., Gilmore, T.D. 2007. Mutation of an IKK phosphorylation site within the transactivation domain of REL in two patients with B-cell lymphoma

- enhances REL's in vitro transforming activity. *Oncogene* 26, 2685–2694. doi:10.1038/sj.onc.1210089
- Steele, R.E., David, C.N., Technau, U. 2011. A genomic view of 500 million years of cnidarian evolution. *Trends in Genetics* 27, 7–13. doi:10.1016/j.tig.2010.10.002
- Stefanik, D.J., Lubinski, T.J., Granger, B.R., Byrd, A.L., Reitzel, A.M., DeFilippo, L., Lorenc, A., Finnerty, J.R. 2014. Production of a reference transcriptome and transcriptomic database (EdwardsiellaBase) for the lined sea anemone, *Edwardsiella lineata*, a parasitic cnidarian. *BMC Genomics* 15, 71.
- Steward, R. 1987. *Dorsal*, an embryonic polarity gene in *Drosophila*, is homologous to the vertebrate proto-oncogene, *c-rel*. *Science* 238, 692–694.
- Stibbs H.H., Owczarzak, A., Bayne, C.J., DeWan, P. 1979. Schistosome sporocyst-killing amoebae isolated from *Biomphalaria glabrata*. *Journal of Invertebrate Pathology* 33, 159–170.
- Stöven, S., Silverman, N., Junell, A., Hedengren-Olcott, M., Erturk, D., Engström, Y., Maniatis, T., Hultmark, D. 2003. Caspase-mediated processing of the *Drosophila* NF- κ B factor Relish. *Proceedings of the National Academy of Sciences of the United States of America* 100, 5991–5996.
- Stroud, J.C., Chen, L. 2003. Structure of NFAT bound to DNA as a monomer. *Journal of Molecular Biology* 334, 1009–1022.
- Suga, H., Chen, Z., de Mendoza, A., Seb  -Pedr  s, A., Brown, M.W., Kramer, E., Carr, M., Kerner, P., Vervoort, M., S  nchez-Pons, N., Torruella, G., Derelle, R., Manning, G., Lang, B.F., Russ, C., Haas, B.J., Roger, A.J., Nusbaum, C., Ruiz-Trillo, I. 2013. The *Capsaspora* genome reveals a complex unicellular prehistory of animals. *Nature Communications* 4, 1–9.
- Sullivan, J.C., Kalaitzidis, D., Gilmore, T.D., Finnerty, J.R. 2007. Rel homology domain-containing transcription factors in the cnidarian *Nematostella vectensis*. *Development, Genes and Evolution* 217, 63–72.
- Sullivan, J.C., Wolenski, F.S., Reitzel, A.M., French, C.E., Traylor-Knowles, N., Gilmore, T.D., Finnerty, J.R. 2009. Two alleles of NF- κ B in the sea anemone *Nematostella vectensis* are widely dispersed in nature and encode proteins with distinct activities. *PLoS One* 4, e7311.
- Sun, S.-C. 2011. Non-canonical NF- κ B signaling pathway. *Cell Research*, 21, 71–85.
- Swofford, D.L. 2001. PAUP*: Phylogenetic Analysis Using Parsimony (and other methods) 4.0.b5.

- Tincu, J.A., Taylor, S.W. 2004. Antimicrobial peptides from marine invertebrates. *Antimicrobial Agents and Chemotherapy* 48, 3645–3654. <https://doi.org/10.1128/AAC.48.10.3645-3654.2004>
- Tompkins-Macdonald, G.J., Gallin, W.J., Sakarya, O., Degnan, B. Leys, S.P., Boland, L.M. 2009. Expression of a poriferan potassium channel: insights into the evolution of ion channels in metazoans. *Journal of Experimental Biology* 212, 761–767. <https://doi.org/10.1242/jeb.026971>
- Traylor-Knowles, N., Rose, N.H., Sheets, E.A., Palumbi, S.R. 2017. Early transcriptional responses during heat stress in the coral *Acropora hyacinthus*. *Biological Bulletin* 232, 91–100.
- Tulin, S., Aguiar, D., Istrail, S., Smith, J. 2013. A quantitative reference transcriptome for *Nematostella vectensis* early embryonic development: a pipeline for *de novo* assembly in emerging model systems. *EvoDevo* 4, 16.
- Valanne, S., Wang, J.-H., Rämet, M. 2011. The *Drosophila* Toll signaling pathway. *Journal of Immunology* 186, 649–656. doi:10.4049/jimmunol.1002302
- Van Soest, R.W.M., Boury-Esnault, N., Vacelet, J., Dohrmann, M., Erpenbeck, D., De Voogd, N.J., Santodomingo, N., Vanhoorne, B., Kelly, M., Hooper, J.N.A. 2012. Global diversity of sponges (Porifera). *PLoS One* 7, e35105. doi:10.1371/journal.pone.0035105.
- Vargas, S., Schuster, A., Sacher, K., Büttner, G., Schätzle, S., Läubli, B., Hall, K., Hooper, J.N.A., Erpenbeck, D., Wörheide, G. 2012. Barcoding sponges: an overview based on comprehensive sampling. *PLoS One* 7, 39345. <https://doi.org/10.1371/journal.pone.0039345>
- Vasselon, T., Detmers, P.A. 2002. Toll receptors: a central element in innate immune responses. *Infection and Immunity* 70, 1033–1041. doi:10.1128/IAI.70.3.1033-1041.2002
- Voolstra, C.R., Li, Y., Liew, Y.J., Baumgarten, S., Zoccola, D., Flot, J.-F., Tambutté, S., Allemand, D., Aranda, M. 2017. Comparative analysis of the genomes of *Stylophora pistillata* and *Acropora digitifera* provides evidence for extensive differences between species of corals. *Scientific Reports* 7, 1–14.
- Warner, J.F., Guerlais, V., Amiel, A.R., Johnston, H., Nedoncelle, K., Röttinger, E. 2018. NvERTx: a gene expression database to compare embryogenesis and regeneration in the sea anemone *Nematostella vectensis*. *Development* 145, dev162867.

- Wehrl, M., Steinert, M., Hentschel, U. 2007. Bacterial uptake by the marine sponge *Aplysina aerophoba*. *Microbial Ecology* 53, 355–365.
<https://doi.org/10.1007/s00248-006-9090-4>
- Weis, V.M. 2008. Cellular mechanisms of cnidarian bleaching: stress causes the collapse of symbiosis. *Journal of Experimental Biology* 211, 3059–3066.
- Wenger, Y., Buzgariu, W., Reiter, S., Galliot, B. 2014. Injury-induced immune responses in *Hydra*. *Seminars in Immunology* 26, 277–294.
- Wiens, M., Korzhev, M., Krasko, A., Thakur, N.L., Perović-Ottstadt, S., Breter, H.J., Ushijima, H., Diehl-Seifert, B., Müller, I.M., Müller, W.E.G. 2005. Innate immune defense of the sponge *Suberites domuncula* against bacteria involves a MyD88-dependent signaling pathway: induction of a perforin-like molecule. *Journal of Biological Chemistry* 280, 27949–27959.
<https://doi.org/10.1074/jbc.M504049200>
- Wiens, M., Korzhev, M., Perović-Ottstadt, S., Luthringer, B., Brandt, D., Klein, S., Müller, W.E.G. 2007. Toll-like receptors are part of the innate immune defense system of sponges (Demospongiae: Porifera). *Molecular Biology and Evolution* 24, 792–804. doi: 10.1093/molbev/msl208.
- Williams, L.M., Fuess, L.E., Brennan, J.J., Mansfield, K.M., Salas-Rodriguez, E., Welsh, J., Awtry, J., Banic, S., Chacko, C., Chezian, A., Dowers, D., Estrada, F., Hsieh, Y.-H., Kang, J., Li, W., Malchiodi, Z., Malinowski, J., Matuszak, S., McTigue IV, T., Mueller, D., Nguyen, B., Nguyen, M., Nguyen, P., Nguyen, S., Njoku, N., Patel, K., Pellegrini, W., Pliakas, T., Qadir, D., Ryan, E., Schiffer, A., Thiel, A., Yunes, S.A., Spiliotis, K.E., Pinzón C, J.H., Mydlarz, L.D., Gilmore, T.D. 2018. A conserved Toll-like receptor-to-NF- κ B signaling pathway in the endangered coral *Orbicella faveolata*. *Developmental and Comparative Immunology* 79, 128–136.
- Williams, L.M., Inge, M.M., Mansfield, K.M., Rasmussen, A., Afghani, J., Agrba, M., Albert, C., Andersson, C., Babaei, M., Babaei, M., Bagdasaryants, A., Bonilla, A., Browne, A., Carpenter, S., Chen, T., Christie, B., Cyr, A., Dam, K., Dulock, N., Erdene, G., Esau, L., Esonwune, S., Hanchate, A., Huang, X., Jennings, T., Kasabwala, A., Kehoe, L., Kobayashi, R., Lee, M., LeVan, A., Liu, Y., Murphy, E., Nambiar, A., Olive, M., Patel, D., Pavesi, F., Petty, C.A., Samofalova, Y., Sanchez, S., Stejskal, C., Tang, Y., Yap, A., Cleary Jr., J.P., Yunes, S.A., Siggers, T., Gilmore, T.D. 2020. Transcription factor NF- κ B in a basal metazoan, the sponge, has conserved and unique sequences, activities, and regulation. *Developmental and Comparative Immunology* 104, 103559.
- Williams, L.M., Gilmore, T.D. 2020. Looking down on NF- κ B. *Molecular and Cellular Biology* 40, e00104-20.

- Williams, L.M. Sridhar, S., Samaroo, J., Adindo, E.K., Addanki, A., BB522 Molecular Biology Laboratory, Aguirre Carrión, P.J., DiRusso, C.J., Rodriguez-Sastre, N., Siggers, T., Gilmore, T.D. 2021. Diversification of transcription factor NF- κ B in protists. bioRxiv, <https://biorxiv.org/cgi/content/short/2021.03.15.435342v1>
- Wolenski, F.S., Bradham, C.A., Finnerty, J.R., Gilmore, T.D. 2013. NF- κ B is required for cnidocyte development in the sea anemone *Nematostella vectensis*. *Developmental Biology* 373, 205–215.
- Wolenski, F.S., Chandani, S., Stefanik, D.J., Jiang, N., Chu, E., Finnerty, J.R., Gilmore, T.D. 2011. Two polymorphic residues account for the differences in DNA binding and transcriptional activation by NF- κ B proteins encoded by naturally occurring alleles in *Nematostella vectensis*. *Journal of Molecular Evolution* 73, 325–336.
- Wolenski, F.S., Garbati, M.R., Lubinski, T.J., Traylor-Knowles, N., Dresselhaus, E., Stefanik, D.J., Goucher, H., Finnerty, J.R., Gilmore, T.D. 2011. Characterization of the core elements of the NF- κ B signaling pathway of the sea anemone *Nematostella vectensis*. *Molecular and Cellular Biology* 31, 1076–1087.
- Wolenski, F.S. 2012. The NF-kappaB signaling pathway of the sea anemone *Nematostella vectensis*: functional characterization of core elements and two naturally occurring polymorphisms. Boston University, PhD Dissertation. <https://open.bu.edu/handle/2144/12677>
- Wolenski, F.S., Layden, M.J., Martindale, M.Q., Gilmore, T.D., Finnerty, J.R. 2013. Characterizing the spatiotemporal expression of RNAs and proteins in the starlet sea anemone, *Nematostella vectensis*. *Nature Protocols* 8, 900–915. doi:10.1038/nprot.2013.014
- Wolfowicz, I., Baumgarten, S., Voss, P.A., Hambleton, E.A., Voolstra, C.R., Hatta, M., Guse, A. 2016. *Aiptasia* sp. larvae as a model to reveal mechanisms of symbiont selection in cnidarians. *Scientific Reports* 6, 1–12.
- Woznica, A., Gerdt, J.P., Hulett, R.E., Clardy, J., King, N. 2017. Mating in the closest living relatives of animals is induced by a bacterial chondroitinase. *Cell* 170, 1175–1183.e11.
- Yin, Z., Zhu, M., Davidson, E.H., Bottjer, D.J., Zhao, F., Tafforeau, P., 2015. Sponge grade body fossil with cellular resolution dating 60 Myr before the Cambrian. *Proceedings of the National Academy of Sciences of the United States of America* 112, E1453–E1460. <https://doi.org/10.1073/pnas.1414577112>
- Ying, H., Hayward, D.C., Cooke, I., Wang, W., Moya, A., Siemering, K.R., Sprungala, S., Ball, E.E., Forêt, S., Miller, D.J. 2019. The whole-genome sequence of the coral *Acropora millepora*. *Genome Biology and Evolution* 11, 1374–1379.

- Zárate-Potes, A., Ocampo, I.D., Cadavid, L.F. 2019. The putative immune recognition repertoire of the model cnidarian *Hydractinia symbiolongicarpus* is large and diverse. *Gene* 684, 104–117.
- Zhang, Y., Liu, Jianjun, Yao, S., Li, F., Xin, L., Lai, M., Bracchi-Ricard, V., Xu, H., Yen, W., Meng, W., Liu, S., Yang, L., Karmally, S., Liu, Jin, Zhu, H., Gordon, J., Khalili, K., Srinivasan, S., Bethea, J.R., Mo, X., Hu, W. 2012. Nuclear factor kappa B signaling initiates early differentiation of neural stem cells. *Stem Cells* 30, 510–524. <https://doi.org/10.1002/stem.1006>
- Zhou, Z., Wu, Y., Zhang, C., Li, C., Chen, G., Yu, X., Shi, X., Xu, Y., Wang, L., Huang, B. 2017. Suppression of NF- κ B signal pathway by NLRC3-like protein in stony coral *Acropora aculeus* under heat stress. *Fish and Shellfish Immunology* 67, 322-330. doi:10.1016/j.fsi.2017.06.027

CURRICULUM VITAE

