Evolution of deuterostome immune systems

by

Michael G. Tassia

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of Philosophy

Auburn, Alabama
December 11th, 2021

Keywords: Hemichordata, Deuterostomia, innate immunity, pattern-recognition receptors, differential gene expression, molecular evolution

Approved by:
Kenneth M. Halanych, Chair, Professor of Biological Sciences
Elizabeth H. Schwartz, Associate Professor of Biological Sciences
Ryan C. Range, Assistant Professor of Biological Sciences
Jamie R. Oaks, Assistant Professor of Biological Sciences

Abstract

Contemporary theory on the evolution of metazoan immune systems is primarily derived of work on a few historically established model species (including *Drosophila,* human, and mouse). Though these species have provided invaluable insight into the molecular and cellular components of immunity, they nonetheless cannot alone provide sufficient phylogenetic contrast to accurately capture the nuances of immunity evolution. In the modern sequencing era of comparative biology, studies leveraging transcriptomic and genomic datasets have provided valuable insight into the ancestry of immunity molecular toolkits within Deuterostomia – a superphylum comprised of Hemichordata (acorn worms and pterobranchs), Echinodermata (urchins, sea stars, and their allies), and Chordates (tunicates, lancelets, and vertebrates). In my PhD thesis, I focus on remedying a gap in knowledge by improving representation of hemichordates in the context of immunity and place these findings in the framework of deuterostome evolution.

This dissertation contains one literature review and three research chapters (**Chapters 2-5**). **Chapter 2** is zoological account of the hemichordate clade, Enteropneusta (commonly called acorn worms). In addition to overviewing the anatomy, physiology, and diversity within the clade, this publication also highlights the sparsity of immunology within the phylum. **Chapter 3** focuses on the Toll-like receptor (TLR) pathway, a central component of pathogen recognition and innate immunity signaling. This chapter was the first published work of my dissertation, and its findings provided the groundwork for the subsequent chapters. **Chapter 4** concerns the molecular toolkit involved in antiviral immunity in the acorn worm hemichordate, *Saccoglossus kowalevskii*. The capacity for *S. kowalevskii* to recognize and transcriptionally react to viral stimulus is described. This work is a critical step forward for hemichordate immunology and towards inferring the ancestry of deuterostome immunity evolution. The final data chapter (**Chapter 5**) expands upon the work of **Chapter 3** by investigating the evolution of key innate immunity proteins across Metazoa, with focus on deuterostome taxa. For this work, I developed a bioinformatic tool called TIAMMAt (*Taxon-Informed Adjustment of Markov Model Attributes*). This study highlights the value of improving the representation of non-model species in comparative evolutionary studies using immunity as a case study.

Acknowledgements

Science does not often provide its denizens the occasion to be unrelentingly thankful to the people and institutions that make this career possible. I'd like to enthusiastically take this opportunity to say thank you to those who have made a lasting impression on me and my career.

First, I'd like to thank my advisor Ken Halanych for entertaining my wild ideas and endorsing my "adventures down the scientific rabbit hole." For the many occasions that he encouraged diving *even* deeper into a project, there have been just as many times where he has had to remind me that I sometimes (often) tend to reach a little too deep into the conceptual cookie-jar – sometimes (often) losing sight of my initial goals. During my PhD, Ken has helped me to attend countless scientific meetings, spend countless hours digging for worms in the mud, and even spend months at sea to visit Antarctica. I am profusely thankful for his support and advice for my scientific career and in life.

Thank you to the members of the Halanych lab that have tolerated me all these years. I feel truly fortunate to have worked in a lab where I not only consider its members my colleagues, but also good friends. I want to directly thank all of whom have overlapped with and helped me stay sane (past and present): Viktoria Bogantes, Damien Waits, Kyle David, Nathan Whelan, Caitlin Redak, Candace Grimes, Matt Galaska, Elisa Acosta, Oluchi Oyekwe, Yu Sun, Yuanning Li, Susan Rashid, and Pam Brannock. I would also like to thank my PhD committee members: Beth Schwartz, Ryan Range, and Jamie Oaks for their expertise and motivation.

Thank you to Chris Lowe for hosting me at MBL over the years and teaching me how to work with *Saccoglossus*. Thank you to Billie Swalla for introducing me to hemichordates in the first place! Thank you to Patricia Morse for being so persistently supportive during my scientific career. I'm grateful to my collaborators with whom I've had the privilege to publish with and enjoy many late-night conversations about the wacky world of marine invertebrates: Kevin Kocot, James Townsend, Joie Cannon, and Alejandro Damian-Serrano. A huge shoutout to the science teams and crews of the *RV Hugh R. Sharp* and the *RVIB Nathanial B. Palmer*. I'd like to thank the funding sources and institutions that have supported me during my PhD: Auburn University, Marine

Biological Laboratories, Society of Developmental Biology, Society of Integrative and Comparative Biology, and The National Science Foundation

To my mom and dad who have always supported and encouraged me to pursue the career that would make me happiest, I am endlessly grateful to have parents like you. To my brother, Jono, who (despite moving halfway across the planet in the middle of my PhD) still makes time to play video games with me just like when we were kids. To my other three "brothers", Anthony, Ben, and Chris, thank you for being a constant source of joy and absurdity in my life. A sincere thank you to my dearest friends, Damien and Viktoria. And thank you to my Rocket League teammates, Jono, Anthony, and Damien, for letting me carry you all these years.

Finally, I'd like to express my most heartfelt thanks to my wife, Kat, and our three "underbaked" cats, Ella, Emma, and Olivia. My dissertation is dedicated to you.

Table of Contents

List of Tables

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Abbreviations

BIC	Bayesian Information Criterion
DEG	Differentially Expressed Gene
DGE	Differential Gene Expression
HMM	Hidden Markov Model
IRF	Interferon Regulatory Factor
LRT	Likelihood Ratio Test
PAMP	Pathogen-Associated Molecular Pattern
PP	Posterior Probability
PRR	Pattern Recognition Receptor
NLR	NOD-Like Receptor
RLR	RIG-I-Like Receptor
TIAMMAt	Taxon-Informed Adjustment of Markov Model Attributes
TLR	Toll-Like Receptor

Chapter 1: Introduction

1.1 General Introduction and Background

Innate immunity represents a robust line of defense against pathogens. Interestingly, some components of innate immune systems possess homology dating back to the origins of Metazoa (Gauthier et al., 2010). In contrast, adaptive immune systems have evolved several times, with the gnathostome immunoglobulin-based system often regarded as the most evolutionarily complex (Litman et al., 2010). Whereas there is a large body of work delineating molecular determinants and cellular interfaces of innate and adaptive immunity of vertebrates (Boehm, 2012), far less is known about the immune systems of invertebrates.

Innate immune response pathways are initiated and regulated by a cohort of conserved pattern recognition receptors (PRRs) (Takeuchi and Akira, 2010). PRRs include, among others, Toll-Like Receptors (TLR), NOD-Like Receptors (NLR) and RIG-I-Like Receptors (RLR) (**Fig. 2**). Ligation of PRRs by specific pathogen associated molecular patterns (PAMPs) elicits a range of signaling pathways often resulting in pro-inflammatory cytokine production (Akira and Takeda, 2004). Common PAMPs include lipopolysaccharide (Gram-negative bacteria), peptidoglycan (Gram-positive bacteria), dsRNA (viruses), and zymosan-A (fungi) – all of which are often intrinsically vital to the pathogen that may attack the animal.

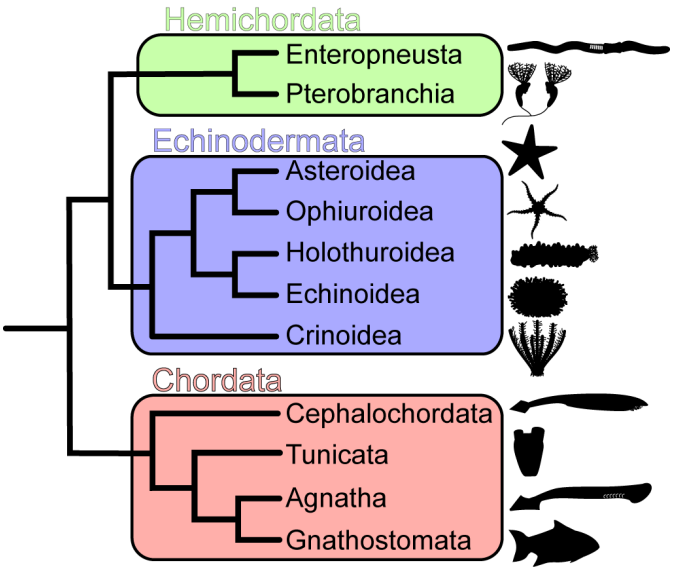


Figure 1: Phylogeny of Deuterostomia – comprised of Hemichordata, Echinodermata, and Chordata. From Tassia et al. 2017.

Although many conserved innate immune pathways exist across Metazoa, the TLR signaling pathway is among the most-well studied. Unlike receptors of adaptive immune systems, TLRs are not modified within a single individual by gene rearrangements (Litman et al., 2010). Interestingly, the number of TLRs varies across taxa, suggesting a possible mechanism for variation in the breadth of pathogen recognition controlled by expansions/reductions in the number

of unique TLR proteins encoded in the genome. Whereas vertebrate lineages typically possess 10-19 TLR homologs, cephalochordate genomes code up to 72 TLRs, tunicates only 3, and urchins over 200 (Buckley and Rast, 2015; Tassia et al., 2017). Additionally, signaling adaptors vary among specific TLR orthologs – sometimes inducing specialized cytokine production, such as antiviral interferons (**Fig. 2**) (Akira and Takeda, 2004; Medzhitov, 2001). The large variation of

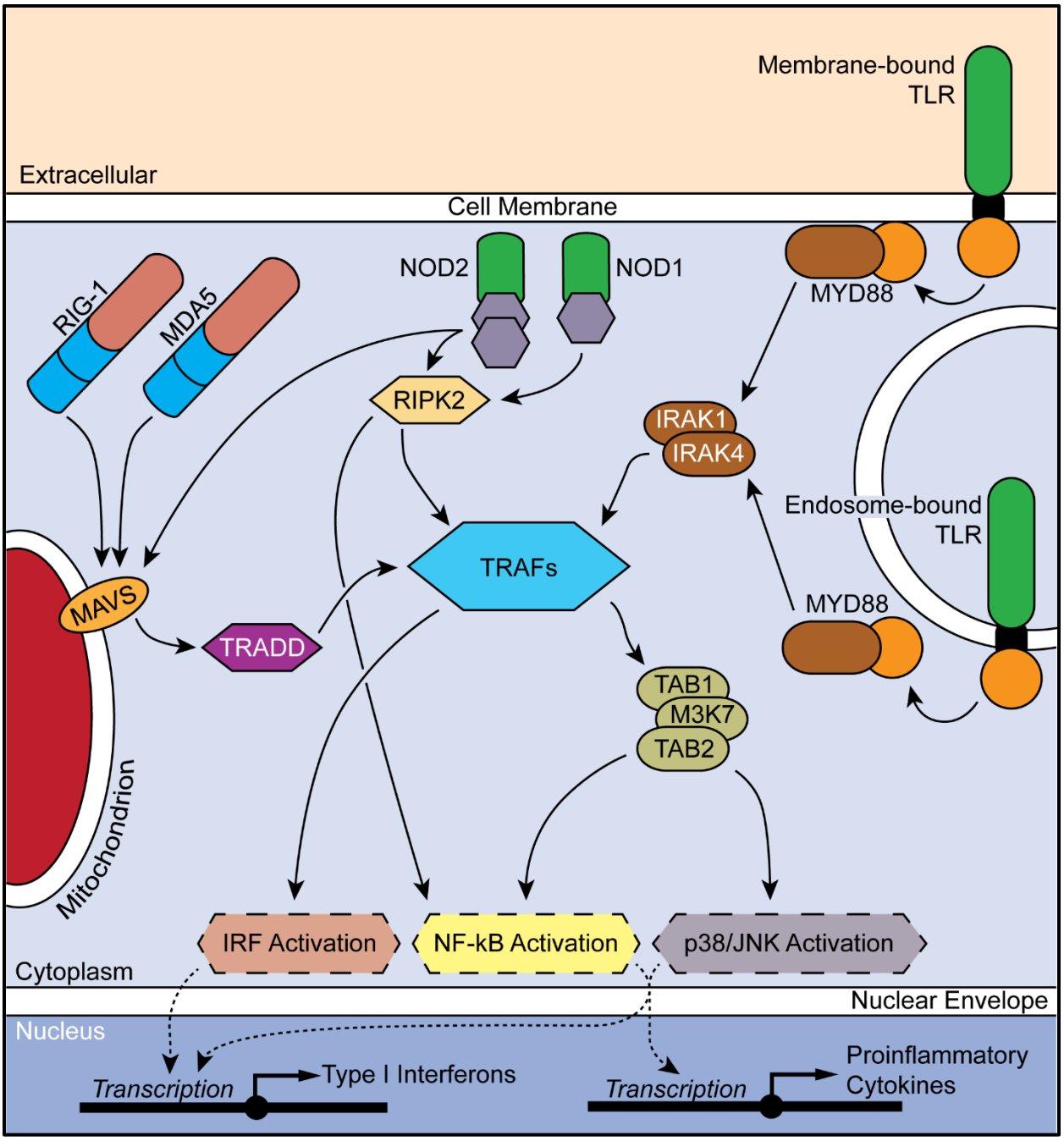


Figure 1: Major PRR signaling pathways including Toll-like receptor (TLR) signaling (right), NOD-like receptor (NLR) signaling (center), and RIG-1-like receptor (RLR) signaling (left).

TLR gene number, combined with a lack of functional data (aside from vertebrates and echinoderm larvae), creates a gap in knowledge about the ancestral characteristics of deuterostome innate immune pathways.

Alongside TLRs, NLRs and RLRs similarly rely on the recognition and binding of PAMPs to initiate molecular immunity signal transduction which, like TLRs, leads to activation of IRF- and/or NFkB-driven cytokine expression (**Fig. 2**) (Loo and Gale, 2011). However, unlike TLRs, which are type-I transmembrane proteins, NLRs and RLRs act cytoplasmically (Takeuchi and Akira, 2010) and are comparatively understudied. Whereas individual TLRs and/or NLRs recognize a broad suite of PAMPs indicative of bacterial, fungal, or viral infection, RLRs exclusively recognize viral nucleic acid moieties (Loo and Gale, 2011). Akin to TLRs and their associated signaling pathways, NLRs and RLRs have also been shown to play roles in the immune response of both vertebrates and invertebrates (Lange et al., 2011; Zhang et al., 2014), indicating these PRR-driven pathways also represent an evolutionarily ancient branch of metazoan innate immunity present well before the origin of vertebrates.

Recent availability of genomic data (Cannon et al., 2014; Simakov et al., 2015) combined with a strongly supported phylogeny for all major deuterostome groups (Cannon et al., 2014; Near et al., 2012; Telford et al., 2014) provides an unprecedented opportunity to explore evolution of deuterostome innate immunity. Together, hemichordates (pterobranchs and enteropneusts, or acorn worms) and echinoderms form Ambulacraria, the sister clade to chordates (**Fig. 1**) (Cannon et al., 2014). Modern hemichordates, unlike extant echinoderms, possess several morphological (e.g., pharyngeal gill slits) and molecular traits (e.g., anteroposterior axis-patterning gene boundaries within the adult nervous system) often considered chordate synapomorphies which can be more accurately attributed to the deuterostome last common ancestor (Brown et al., 2008; Lowe et al., 2015; Peterson and Eernisse, 2001).

Over the course of my doctoral dissertation, I employ a balance of bioinformatic and molecular methods to investigate immunity in hemichordates, a subject overlooked within scientific literature since the 1980's (Millar and Ratcliffe, 1987a, 1987b; Rhodes and Ratcliffe, 1983). The works described below fill a major taxonomic gap in immunity research among deuterostome taxa, ultimately providing contrast essential to accurately inferring the ancestral state of immunity in the last common ancestor to deuterostomes.

1.2 References:

Akira S, Takeda K. 2004. Toll-like receptor signalling. *Nat. Rev. Immunol.* 4: 499–511.

Boehm T. 2012. Evolution of vertebrate immunity. *Curr. Biol.* 22: R722–R732.

Brown FD, Prendergast A, Swalla BJ. 2008. Man is but a worm: chordate origins. *Genesis* 46: 605–613.

Buckley KM, Rast JP. 2015. Diversity of animal immune receptors and the origins of recognition complexity in the deuterostomes. *Dev. Comp. Immunol.* 49: 179–189.

Cannon JT, Kocot KM, Waits DS, Weese DA, Swalla BJ, Santos SR, Halanych KM. 2014. Phylogenomic Resolution of the Hemichordate and Echinoderm Clade. *Curr. Biol.* 24: 2827–2832.

Gauthier ME, Du Pasquier L, Degnan BM. 2010. The genome of the sponge *Amphimedon queenslandica* provides new perspectives into the origin of Toll-like and interleukin 1 receptor pathways. *Evol. Dev.* 12: 519–533.

Lange C, Hemmrich G, Klostermeier UC, López-Quintero JA, Miller DJ, Rahn T, Weiss Y, Bosch TCG, Rosenstiel P. 2011. Defining the origins of the NOD-like receptor system at the base of animal evolution. *Mol. Biol. Evol.* 28: 1687–1702.

Litman GW, Rast JP, Fugmann SD. 2010. The origins of vertebrate adaptive immunity. *Nat. Rev. Immunol.* 10: 543–53.

Loo YM, Gale M. 2011. Immune Signaling by RIG-I-like Receptors. *Immunity* 34: 680–692.

Lowe CJ, Clarke DN, Medeiros DM, Rokhsar DS, Gerhart J. 2015. The deuterostome context of chordate origins. *Nature* 520: 456–465.

Medzhitov R. 2001. Toll-like receptors and innate immunity. *Nat. Rev. Immunol.* 1: 135–145.

Millar D, Ratcliffe N. 1987a. The antibacterial activity of the hemichordate *Saccoglossus ruber* (Enteropneusta). *J. Invertebr. Pathol.* 50: 191–200.

Millar D, Ratcliffe N. 1987b. Activity and preliminary characterisation of a hemagglutinin from the hemichordate *Saccolgossus ruber*. *Dev. Comp. Immunol*. 11: 309–320.

Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. *Proc. Natl. Acad. Sci.* 109: 13698–13703.

Peterson KJ, Eernisse DJ. 2001. Animal phylogeny and the ancestry of bilaterians: Inferences from morphology and 18S rDNA gene sequences. *Evol. Dev.* 3: 170–205.

Rhodes CP, Ratcliffe NA. 1983. Coelomocytes and defense reactions of the primitive chordates, *Branchiostoma lanceolatum* and *Saccoglossus horsti. Dev. Comp. Immunol.* 7: 695–698.

Simakov O, Kawashima T, Marlétaz F, Jenkins J, Koyanagi R, Mitros T, Hisata K, Bredeson J, Shoguchi E, Gyoja F, et al. 2015. Hemichordate genomes and deuterostome origins. *Nature* 527: 1–19.

Takeuchi O, Akira S, 2010. Pattern Recognition Receptors and Inflammation. *Cell* 140: 805–820.

Tassia MG, Whelan NV, Halanych KM. 2017. Toll-like receptor pathway evolution in deuterostomes. *Proc. Natl. Acad. Sci.* 114: 7055–7060.

Telford MJ, Lowe CJ, Cameron CB, Ortega-Martinez O, Aronowicz J, Oliveri P, Copley RR. 2014. Phylogenomic analysis of echinoderm class relationships supports Asterozoa. *Proc. R. Soc. B Biol. Sci.* 281: 20140479

Zhang Y, Yu F, Li J, Tong Y, Zhang Y, Yu Z. 2014. The first invertebrate RIG-I-like receptor (RLR) homolog gene in the pacific oyster *Crassostrea gigas*. *Fish Shellfish Immunol*. 40: 466–471.

Chapter 2: Enteropneust Morphology*

2.1 Introduction

Enteropneusts, also called acorn worms, are free-living, vermiform deuterostomes found often associated with soft, subtidal sediments or associated with the epibenthos of the deep-sea (**Figure 1**) (Tassia et al. 2016). First collected in the Marshall Islands of the Indo-Pacific by Eschscholtz in 1825, *Ptychodera flava* was initially considered to be an undescribed species of sea cucumber. Eschscholtz's work went unnoticed, however, and most early works on acorn worms was centered around the description of *Balanoglossus clavigerus* by Delle Chiaje in 1829. Though several enteropneust taxa would be described in the subsequent decades, it would not be until 1870 when Gegenbaur would erect the name Enteropneusti (Greek – εντερο: "entero," *intestine*; πνευμον: "pneumon," *lung*) following Kowalevsky (1866) when enteropneust gill slits were first described. The name Hemichordata was not erected until 1885 when Bateson allied enteropneusts within Chordata under the premise of homologous developmental schemes. Whereas the name Hemichordata remains in use today, the original placement of Hemichordata within Chordata was rebutted in 1893 by Spengel and ultimately ejected from Chordata by Van der Horst (1939) – substantiated by Metschnikoff's (1881) assertion to unify Enteropneusta and Echinodermata under the name Ambulacraria based on larval morphology. However, the alliance of these phyla and their phylogenetic placement as sister to Chordata would not be substantialized and formalized until Halanych (1995) using 18S rDNA sequence data (see section **2.5 Phylogeny**). To date, 108 enteropneust species have been described (Tassia et al. 2016) and are organized into four major groups: Harrimaniidae, Spengelidae, Ptychoderidae, and Torquaratoridae (**Figure 1**) (Cannon et al. 2014).

*This chapter has been published as: Tassia MG, Cannon JT, Halanych KM. 2018. "9. Enteropneusta". *Miscellaneous Invertebrates,* edited by A. Schmidt-Rhaesa. Berlin, Boston: De Gruyter, 299-326.

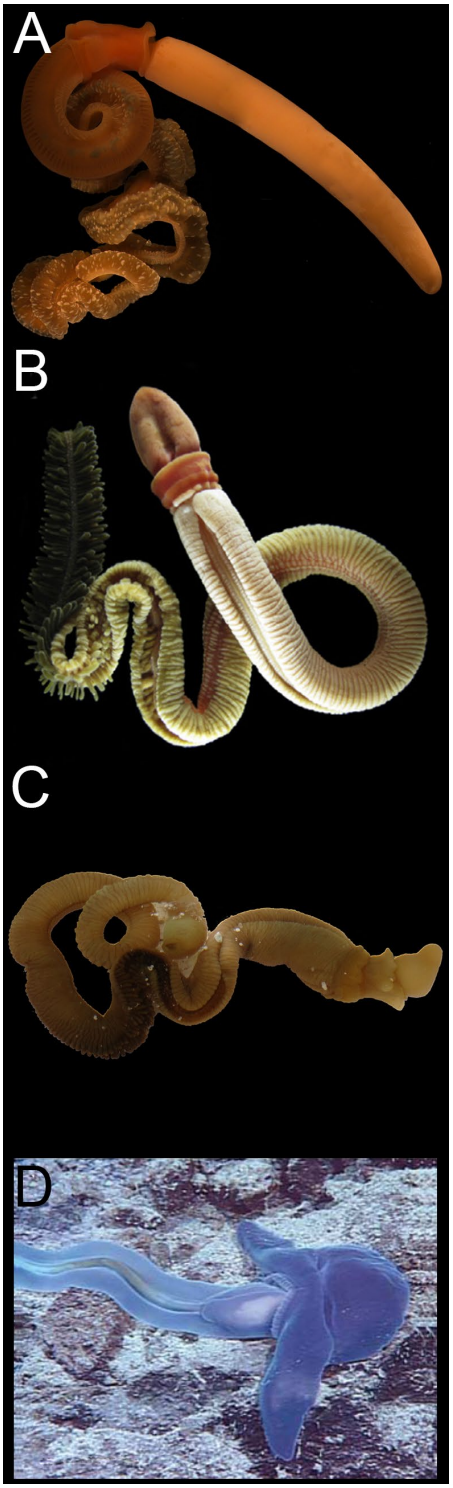


Figure 1: The four major enteropneust groups. A) Harrimaniidae, *Saccoglossus kowalevskii*; B) Spengelidae, *Schizocardium californicum* (adapted from Gonzalez et al. 2017); C) Ptychoderidae, *Ptychodera flava* (from Röttinger & Lowe 2012); D) Torquaratoridae, *Yoda purpurata* (from Holland et al. 2005).

2.2 Morphology

2.2.1 General Anatomy

Size among acorn worm species ranges from meiofaunal, such as in *Meioglossus psammophilus* (611µm; Worsaae et al. 2012), to the especially large *Balanoglossus gigas* (Muller 1898) described from the Brazilian intertidal – exceeding 1.5m in length as an adult. Body coloration varies widely among genera, but often falls into the fleshy tan to washed orange hues for infaunal species. Strikingly, coloration in epibenthic deep-sea taxa varies more broadly, including ranges of translucencies (e.g., *Allapasus aurantiacus*), purple (*Yoda purpurata*), and deep red (*Tergivelum cinnebarium*). Enteropneusts are particularly soft bodied and lack any gross sclerotized structures; as such, they are prone to fragmentation during collection. Osborn et al. (2011) even remark, "Most torquaratorid [enteropneusts] are poorly muscularized, gelatinous and fragile in the extreme." The acorn worm body plan is partitioned into three major body regions, from anterior to posterior: proboscis, collar, and trunk (also referred to as protosome, mesosome, and metasome, respectively).

Proboscis morphology varies from species to species in a gradient from short and "acorn-shaped" (e.g., *Ptychodera flava*) to long and vermiform when relaxed (e.g., *Saccoglossus kowalevskii*). In the most commonly studied infaunal genera (e.g., *Ptychodera* and *Saccoglossus*), the proboscis' primary roles are in locomotion (specifically burrowing) and probing within the environment in association with feeding (Ritter 1902). Preceding the mouth, on the ventral surface of the proboscis, a pre-oral ciliary organ has been strongly implicated in the feeding and chemosensation (Knight-Jones 1952, Gonzalez & Cameron 2009). In epibenthic deep-sea torquaratorids, however, the role of the proboscis is unclear. Torquaratorids uniquely possess, "[A] proboscis and collar each conspicuously broader from side-to-side than in their other dimensions (anteroposterior and dorsoventral) …" (Holland et al. 2005). Among the remaining families, however, the proboscis is rounded, conical, or vaguely cylindrical with a single coelom, and is well-muscularized with circular and longitudinal muscle. In some species, a dorsal groove may extend anteriorly from the proboscis stalk (Deland et al. 2010).

Posteriorly, the proboscis narrows to a slender stalk and terminates connecting to the dorsal interior of the collar (**Figure 2A**). In some species, one or two proboscis pore(s) may be present on the dorsal surface of the proboscis stalk associated with the glomerulus (Cameron 2002b) (see

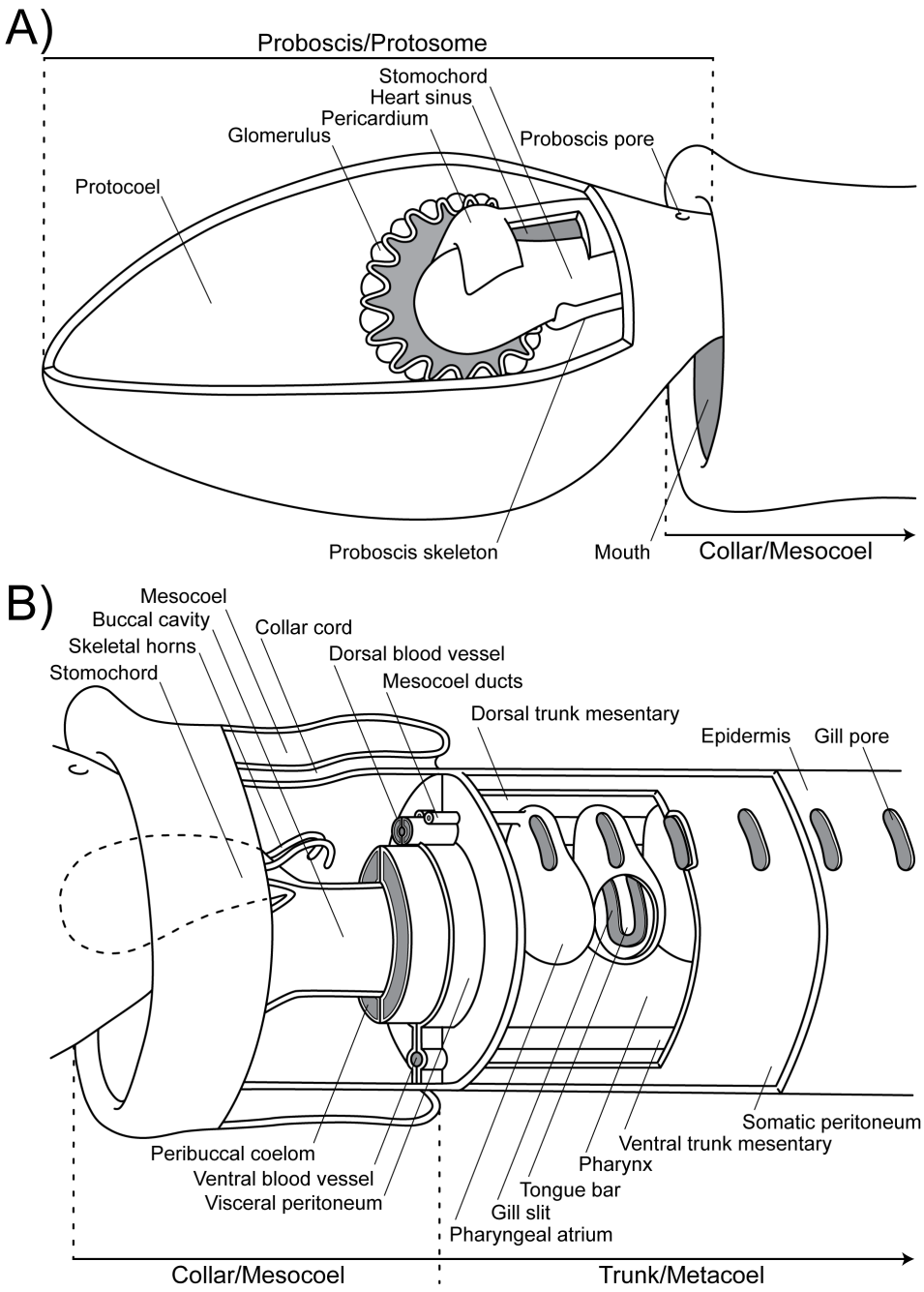


Figure 2: Generalized enteropneust anatomy. A) Proboscis; B) Collar and trunk. Diagrams inspired by Bullock 1945, Benito & Pardos 1997, Cameron 2005).

section **2.2.7 Excretory and Circulatory System**). Internally, the proboscis contains a blind buccal diverticulum called the stomochord (**Figure 2A**) (Bateson 1885). Although early scholars hypothesized the stomochord as homologous to the chordate notochord (see Annona et al. 2015), some even suggesting it as a bona fide notochord (Bateson 1885), canonical notochord-specifying

transcription factors are absent in the developing hemichordate stomochord (Satoh et al. 2014). The stomochord appears to function as a scaffold for several essential organ systems including the dorsal blood-vessel sinus, pericardium, and glomerulus – collectively referred to as the heart-kidney, or proboscis, complex. Interestingly, the heart-kidney complex is reportedly absent from torquaratorids, with the exception of the genus *Allapasus* (Priede et al. 2012), though this may be in part due to the diffuculty of preservation when collecting deep-sea taxa. Ventral to the stomochord is a cartilaginous Y-shaped proboscis skeletal element derived of stomochord basement membrane (**Figure 3**) (Burdon-Jones 1956). The proboscis skeleton bifurcates posteriorly as two horns and terminates within the collar coelom. Additionally, species belonging to the enteropneust family Spengelidae possess proboscis horns which extend the full length of the collar – a diagnostic synapomorphy for the family Spengelidae (Cameron & Perez 2012). This contrasts with horns of *Ptychodera flava* (Ptychoderidae) or *Saccoglossus kowalevskii* (Harrimaniidae) which terminate partially through the collar (Hyman 1959). The proboscis skeleton is also reduced to a medial plate or altogether absent in Torquaratoridae.

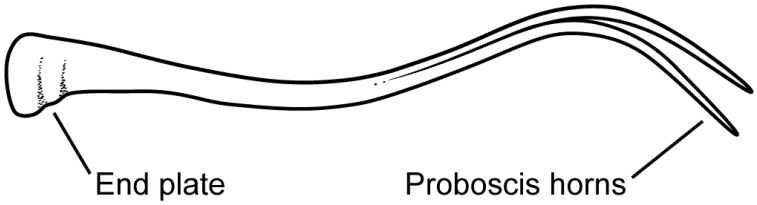


Figure 3: *Protoglossus koehleri* proboscis skeleton, lateral view. Anterior terminus on left, posterior termini on right. Overall length is approximately 1mm.

The enteropneust collar is a cylindrical feature intermediately situated between the proboscis and trunk (**Figure 2**). Collar tissue may partially overlap proboscis and mouth (called a collarette when present), while to a lesser degree overlapping with the trunk posteriorly. At the ventral boundary between collar and proboscis is the mouth (**Figure 2A**). Notably, without assistance from the proboscis, enteropneusts are unable to extrinsically close their mouth (Knight-Jones 1953). Collar external morphology may differ among taxa by circumferential depressions or smooth-to-ruffled collarettes (**Figure 1**) (Hyman 1959). Akin to proboscis morphology, deep-sea torquaratorids display the most morphologically striking collars among extant enteropneust groups – sometimes possessing broad lateral lips which likely assist in funneling epibenthic sediment to the mouth (**Figure 1D**) (Priede et al. 2012).

The third and most posterior body region is the trunk. In a generalized enteropneust model (**Figure 2B**), the trunk can be partitioned into three subsections (listed from anterior to posterior): 1) Branchiogenital region containing reproductive structures and pharyngeal gut, 2) Hepatic region containing digestive gut (also called intestine) and hepatic sacs, and 3) Post-hepatic/Caudal region, sometimes terminating a post-anal tail in juvenile harrimaniids (Colwin & Colwin 1953). In ptychoderids and torquaratorids, the boundaries between all three trunk sections are readily visible and boundary demarcations can often be pronounced (Holland et al. 2005). In contrast, Harrimaniids lack hepatic saccules; as a result, the boundary between pharyngeal and digestive gut can often only be distinguished by the termination of gill slit iteration (Hyman 1959). Spengelids may or may not possess hepatic sacs, and do not possess genital wings, a reproductive structure discussed below (Cameron & Perez 2012).

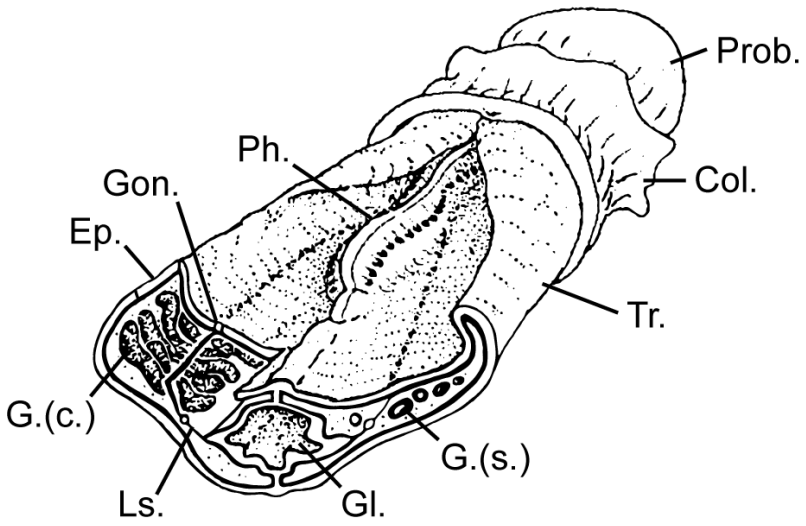


Figure 4: Diagrammatic reconstruction of *Balanoglossus* highlighting genital wings (adapted from Packard 1968). *Col.,* Collar; *Ep.*, Epidermis; *G.(s.)*, Gonads (section); *G.(c.),* Gonads (cutaway); *Gl.*, Gut lumen; *Gon.*, Gonopore; *Ls.*, Lateral septum; *Ph.*, Pharynx; *Prob.*, Proboscis; *Tr.*, Trunk.

Both the relative length and morphology of the branchiogenital region varies between groups. Ptychoderids and torquaratorids possess flat, lateral extensions which curl dorsally to form the genital wings where gonads are stored (**Figure 4**). Whereas the genital wings do not persist past the branchiogenital region in ptychoderids, the lateral wings of torquaratorids can continue along the entire length of the trunk (however, gonads may only present in the most anterior portions (Priede et al. 2012). *Stereobalanus* (Harrimaniidae) uniquely possesses four short gonads immediately posterior to the collar (see Reinhard 1942). External morphology may be entirely insufficient for distinguishing between trunk regions in some species. The number of enteropneust

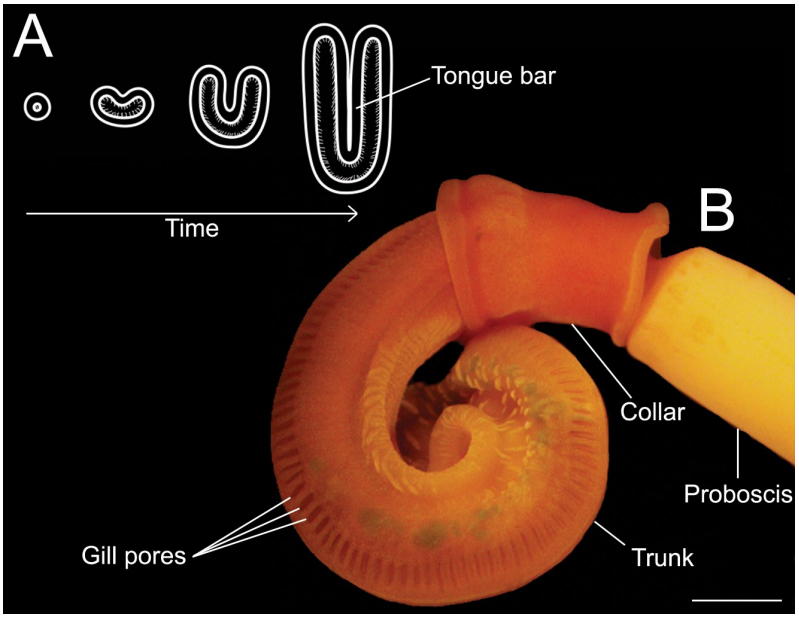


Figure 5: Pharyngeal morphology during development (**A**) and in adult *Saccoglossus kowalevskii* (**B**). **A**) Developmental scheme of gill slit morphogenesis highlighting ventral tongue bar extension (figure adapted from Hyman 1959). **B**) Adult *S. kowalevskii* highlighting external pharynx morphology. Scale bar ≈ 500µm.

gill slits can vary from one (*Meioglossus psammophilus,* Worsaae et al. 2012) to indeterminate (*Balanoglossus aurantiacus*, Harmer 1910) – with additional pairs added with age. Gill slits are found in an iterative, paired series of punctate-to-vertical slits flanking the mid-dorsum of pharynx epithelium (**Figure 2B, 5A**). Homologous to those familiar in chordates, additional gill slits arise through a coordination between gut epithelium (i.e. gill slit formation) and the epidermis (i.e. gill pore formation) (Gillis et al. 2012). During development, gill slit pairs form on either side of the pharynx's dorsolateral surface and are often ellipsoid in shape. Over time, gill pore apices will begin to extend dorsally, separated by a medial tissue extension called secondary gill bars (or tongue bars), finally forming their mature U-shape (**Figure 5B**). Between individual gill slits, primary gill bars (or septa) provide structural support for the gill apparatus and, similar to secondary gill bars, septa are primarily composed of fibrillar collagen (Benito & Pardos 1997, Rychel et al. 2005). In ptychoderids and some spengelids, lateral skeletal elements (referred to as synapticles) join primary and secondary gill bars as lateral "bridges"; this feature is absent in members of Harrimaniidae and Torquaratoridae (Cameron 2005, Holland et al. 2009, Deland et al. 2010, Priede et al. 2012, Osborn et al. 2013). Pharyngeal gill slits do not join directly to the epidermal gill pores in enteropneusts; instead, each gill slit opens first to a single branchial pouch,

or atrium (**Figure 2B**) (Hyman 1959). In most species, individual gill slits possess a single atrium such as in the ptychoderid *Balanoglossus misakiensis* (Hyman 1959). Contrasting most other enteropneusts, the harrimaniid *Stereobalanus canadensis* possesses a single fused atrium with a single consolidated gill pore shared by the entire pharyngeal apparatus (Reinhard 1942).

Posterior to the branchiogenital trunk region resides the hepatic region, or digestive gut– often distinguishable by the initiation of externally visible paired hepatic sacculations, termination of gill slit iteration, and/or tapering and termination of genital/lateral wings. Across enteropneust families, however, some of these features may or may not be present. In ptychoderids, the hepatic/digestive gut is readily visible by the tapering of the genital wings and conspicuous hepatic caeca along the mid-dorsum. Harrimaniids lack genital wings and external hepatic caeca; thus, digestive gut is most easily recognized by the termination of gill slit iteration. Spengelids and torquaratorids possess a mixture of these hepatic region features. *Schizocardium californicum* (Spengelidae) possesses both genital wings and hepatic caeca, whereas *Glandiceps hacksi* (Spengelidae) lacks exteriorly visible hepatic caeca (Cameron & Perez 2012). In torquaratorids, genital wings may be well-defined and restricted to a region coinciding with the pharynx (i.e., a traditional branchiogenetic region as seen in *Yoda purpurata*) or persist the entire length of the trunk (e.g. *Allapasus spp.*) (Holland et al. 2012, Priede et al. 2012), making it an unreliable feature for demarcating branchiogenital from digestive gut in some torquaratorid species. Hepatic sacculations are present but may be occluded visually by other trunk morphologies and are not as pronounced as in ptychoderids; when visible, hepatic intestine is most easily recognized by a region of dark coloration following the branchiogenital trunk (Priede et al. 2012). Finally, the posthepatic trunk is comparably featureless relative to the previously mentioned trunk regions (Hyman 1959). A ventral, vacuolated cord referred to as the pygochord (Willey 1899a) has been reported in the posthepatic trunk of *Saxipendium coronatum* (Woodwick & Sesenbaugh 1985), *Ptychodera flava* (Eschscholtz 1825), and *Glossobalanus berkleyi* (Willey 1931) to be present in the posterior posthepatic gut. However, its presence in histological preparations has been proposed to be an artifact of tissue collapse during fixation (Cameron 2005) drawing skepticism to it's biological validity.

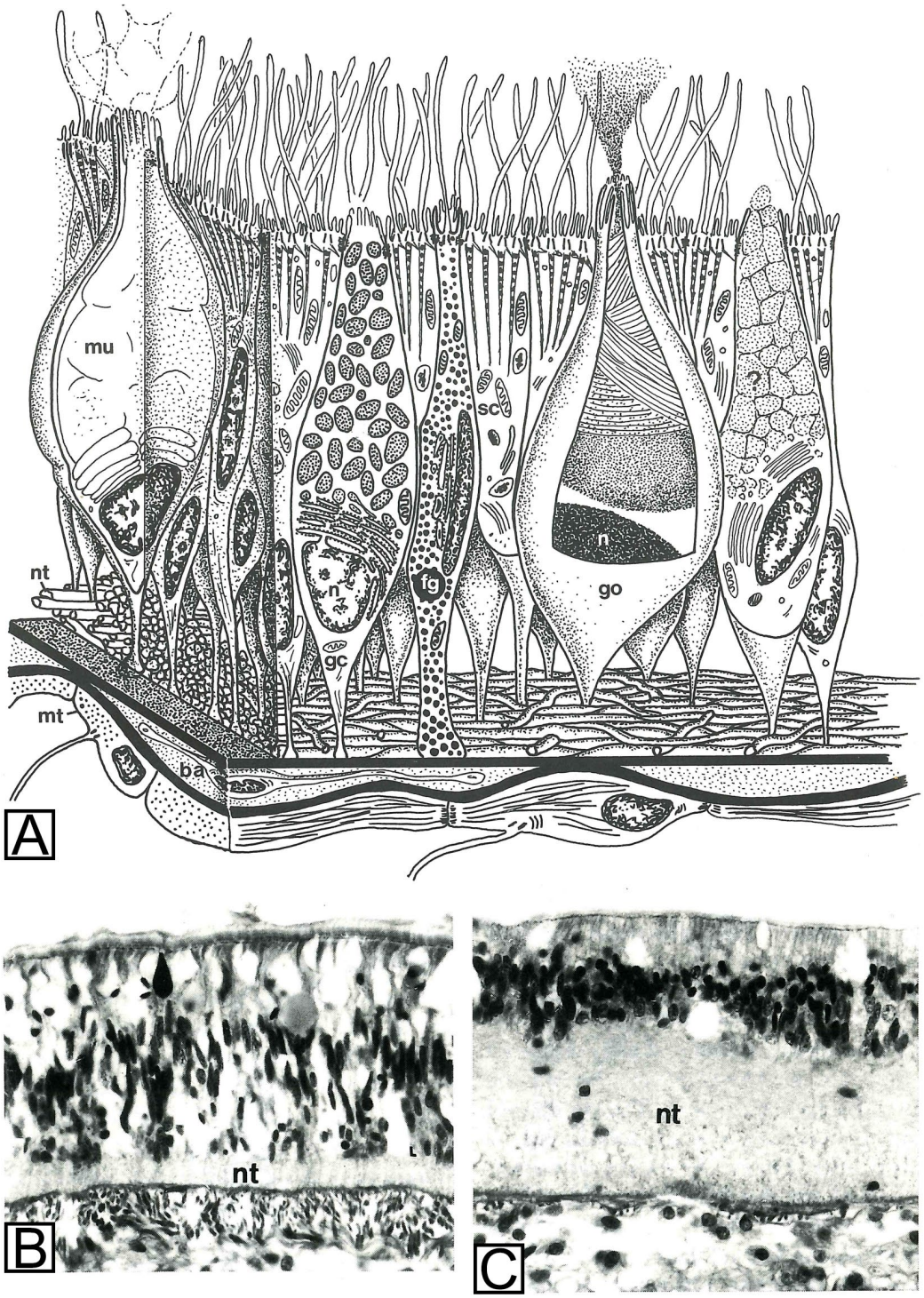


Figure 6: Enteropneust integument. Figure recreated from Benito & Pardos (1997). A) Diagrammatic representation; B) Epidermis of the proboscis (magnified x460); C) Middorsal epidermis of the proboscis stalk (magnified x460). *ba,* blood amoebocyte; *fg*, fine grain cell; *gc,* coarse grain cell; *go,* goblet cell; *mt,* mesothelium; *mu,* mucous cell; *n*, nucleus; *nt*, nerve net; *sc,* supporting cell; *?*, undetermined gland cell.

2.2.2 Integument

The integument is richly punctuated with glandular cells which secrete a thick mucus involved in feeding (Barrington 1940, Knight-Jones 1953, Cameron 2002a), locomotion (Knight-Jones 1952; Hyman 1959), and defense (Millar & Ratcliffe 1987). Mucus secretions have also been implicated in tube-building for both extant Antarctic acorn worms (Halanych et al. 2013) and fossil members (Caron et al. 2013); though this trait of building tubes (temporary or permanent) is not shared by most extant enteropneusts.

The enteropneust integument generally consists of a dense pseudostratified epithelium of ciliated and glandular cells situated above an intraepithelial nerve plexus apical to the basal lamina (**Figure 6**) (Saita et al. 1978, Welsch 1984, Benito & Pardos 1997). Both glandular and ciliated epithelial cells possess a thin, stalk-like structure which traverses through the intraepithelial nerve layer to connect epidermal cells to the basal membrane. Multiple glandular cell morphologies have been categorized and detailed (e.g., mucous cells, mulberry cells, goblet cells, fine granular cells); for more detail on these cell morphologies and reported functions, see Benito & Pardos, 1997 (**Figure 6A**). Schneider (1902) and Hyman (1959) both suggest a thin, "reticulated membrane" may separate the epithelium and nervous layer. This tissue, however, has not been confirmed by electron microscopy (Benito & Pardos 1997). Ciliated-to-glandular compositional ratios, epidermal thickness, and nerve layer thickness can vary between, and within, the three major body regions of acorn worms.

Proboscis integument is comprised of a uniform mixture of columnar ciliated cells and a variety of glandular cells (**Figure 6B**). The nerve layer becomes thickest at the proboscis stalk where it spatially dominates the integument (**Figure 6C**) (Benito & Pardos 1997). In stark contrast, the collar possesses considerable variation in integument histological composition, alternating between visually distinct transverse regions of densely-to-intermediately packed glandular cells (diffused by columnar ciliated cells). The number of these epidermal annulations in the collar integument varies between species, ranging from three in Ptychoderids (Spengel 1893) to five in the harrimaniid *Protoglossus* (Burdon-Jones 1956). Finally, the trunk integument is also densely packed with glandular epithelia (Barrington 1940) and is thinnest at the boundaries of individual hepatic sacculae and epidermal annulations (when present) (Schneider 1902, Hyman 1959, Barrington 1965, Benito & Pardos 1997). The intraepidermal nerve layer also becomes thickest in

the dorsal- and ventral-apices of the trunk – forming the dorsal and ventral trunk nerve cords, respectively (Bullock 1940, Silén 1950). Though, it should be noted that these are structurally intraepidermal, in contrast to the bona fide collar nerve cord (see section **2.2.5 Nervous System**).

2.2.3 Body Cavities

As adults, enteropneusts generally possess five major body coeloms: an unpaired proboscis coelom (protocoel), paired collar coeloms (mesocoels), and paired trunk coeloms (metacoels) (Harmer 1910); each possessing a peritoneum and are derived of enterocoely during development (Hyman 1959, Kaul-Strehlow & Rottinger 2015). A sixth smaller coelomic cavity, which will later differentiate into the pericardial/heart vessel, emerges shortly after the other five major coelomic cavities via schizocoely (Kaul-Strehlow & Stach 2011). Coelomocytes (i.e., free-wandering nonepithelial cells) can be often be found migrating within coelomic cavities. According to Benito and Pardos (1997), coelomocytes possess the capacity for amoeboid movement and likely share a progenitor cell lineage with blood cells. Experimental evidence also suggest enteropneust coelomocytes are involved in immune reactions (Rhodes & Ratcliffe 1983). In addition to coelomocytes, enteropneust coeloms are filled with a transparent fluid that is poorly characterized.

In most acorn worm taxa, the relative distribution of protocoel within the proboscis is often reduced compared to the space occupied by muscle and connective tissue. In most taxa, the protocoel cavity is restricted to the posterior-most region of the proboscis, often only surrounding the heart-kidney complex (Hyman 1959). In contrast, the protocoel of *Protoglossus koehleri* (Caullery & Mesnil 1900) extends to the tip of the proboscis (Burdon-Jones 1956). Dorsal and ventral mesenteries/septa may be present in the proboscis coelom near the proboscis stalk – nearest to the heart-kidney complex – dividing the protocoel into left and right halves (Hyman 1959). However, septa morphology, presence/absence, and length vary from species to species within and between enteropneust families. In *Saccoglossus kowalevskii* (Harrimaniidae), the heart vesicle functions as the dorsal septum, separating the protocoel into left and right halves towards the proboscis base (Hyman 1959), and entirely lacks a ventral proboscis septum (Cameron et al. 2010). In contrast, bona fide dorsal and ventral mesenteries are present in *Protoglossus koehleri* (Burdon-Jones 1956), another harrimaniid. Posteriorly, the protocoel may terminate blindly at the boundary of the proboscis and collar or open to the environment via 1-2 proboscis pores (**Figure 2A**)

associated with the posterior partitions of the protocoel. Although the number and orientation of the proboscis pore(s) varies between species, the majority of described species possess a single proboscis pore associated with the left, and often larger, protocoel partition (Hyman 1959, Benito & Pardos 1997, Cameron et al. 2010, Deland et al. 2010, Cameron & Perez 2012). Proboscis pores appear to be absent altogether in torquaratorids (Holland et al. 2005, Holland et al. 2009, Holland et al. 2012). Proboscis pores and their association with the enteropneust excretory system will be discussed in section **2.2.7 Excretory System** below.

Although the mesocoel develops as a paired organ, a defined septum dividing the left and right compartments is entirely lost in most adult enteropneust species – though this partition is still maintained in members of the harrimaniid genus *Protoglossus* (Burdon-Jones 1956, Deland et al. 2010). In taxa which possess incomplete, or partially lack, mesocoel mesenteries, often the dorsal mesentery is still present as a fragment between the dorsal collar epidermis and the collar cord (see section **2.2.5 Nervous System**), which is exemplified in the ptychoderid genus *Ptychodera* (Hyman 1959, Luttrell et al. 2012). Anteriorly, the collar coelom extends two projections, on either side of the buccal cavity, into the protocoel of the proboscis stalk where it secretes a semi-ridged material called chondroid tissue which is associated with proboscis skeleton (Hyman 1959). This serves as a cartilage-like scaffolding element for the proboscis skeleton and is especially well-developed in harrimaniids and spengelids (Hyman 1959, Deland et al. 2010). In the posterior regions of the mesocoel, two coelomic ducts, located on either side of the buccal cavity at the level of the gill slits, project from the collar coelom into the first pair of gill slits (**Figure 2B**). The mesocoel ducts ultimately connect the collar coelom to the external environment by way of the branchial system (Spengel 1893, Hyman 1959, Benito & Pardos 1997).

The trunk coelom, or metacoel, like the collar coelom, is developmentally paired. Whereas the metacoel's ventral mesentery is often complete in adult acorn worms, the dorsal mesentery is frequently incomplete – extending only as far as the dorsal blood vessel from the dorsum. The lack of a complete septum ultimately yeilds a contiguous cavity between the left and right halves of the trunk coelom. In most described taxa, the transverse septum isolating metacoel from mesocoel does not directly correspond to the external collar-trunk boundary – again, with the exception of *Protoglossus* (Burdon-Jones 1956). Instead, the transverse septum separating the two body cavities often invades as evaginations from the trunk coelom into the collar coelom as two disparate

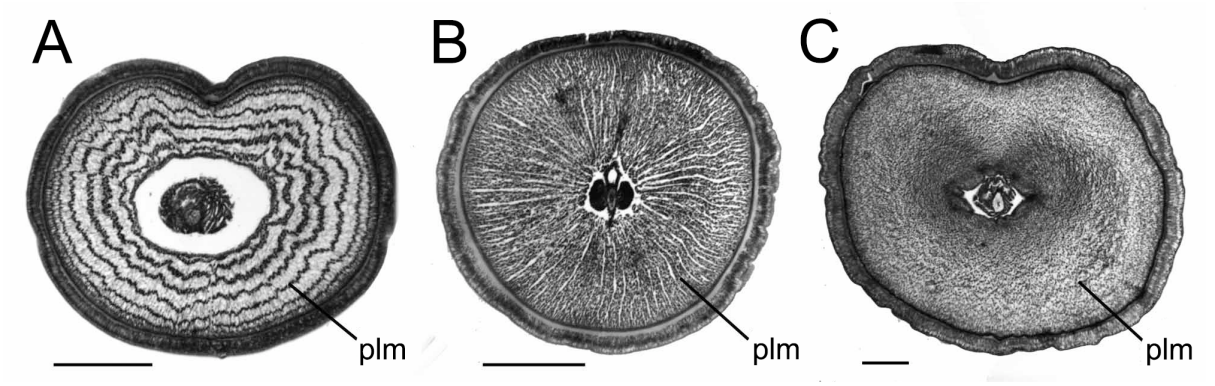


Figure 7: Light micrographs of transverse sections of the proboscis from Deland et al. (2010). **A**) *Saccoglossus pusillus* showing the arrangement of proboscis longitudinal musculature in concentric rings. **B**) *Protoglossus mackiei* showing the arrangement of proboscis longitudinal musculature in radial plates. **C**) *Mesoglossus macginitiei* showing the diffuse arrangement of the proboscis longitudinal musculature. *plm*, proboscis longitudinal muscles. Scale bar = 500μm.

features (**Figure 2B**) (Hyman 1959). First, peribuccal coelomic diverticulae may project anteriorly from the metacoel surrounding the buccal cavity and often extends further dorsally than ventrally. This trait is exhibited by ptychoderids, some harrimaniids (e.g., variably present across *Saccoglossus spp.* and notably lacking in *Protoglossus*), some spengelids (e.g., present in *Schizocardium* and absent in *Glandiceps*), and it is unclear whether peribuccal diverticulae are present in Torquaratoridae (Cameron 2005, Cameron et al. 2010, Deland et al. 2010, Cameron & Perez 2012). The second coelomic projection from the trunk coelom is a paired perihaemal cavity surrounding the dorsal vessel. The paired perihaemal cavities may project as far as the proboscis skeleton in the proboscis stalk (Harmer 1910, Hyman 1959); when fully developed, these cavities are often filled with longitudinal muscle and may or may not coalesce into a single unpaired cavity in the anterior collar still surrounding the dorsal blood vessel (Hyman 1959).

Alongside the five coeloms comprising the major body cavities in the proboscis, collar, and trunk, the pericardium is an unpaired coelomic organ situated dorsally upon the stomochord (**Figure 2A**) (Spengel 1893, Schepotieff 1907, Van der Horst 1939, Benito & Pardos 1997). The function of the pericardium as the central motor to the enteropneust circulatory system will be covered in the section **2.2.7 Circulatory & Excretory System** below. Between the pericardium and the dorsal surface of the stomochord is the blood-vessel sinus, a cavity (sensu lato) on which

the pericardium's contractile nature acts to motivate blood circulation across the enteropneust body.

2.2.4 Musculature

Beneath the proboscis-integument's basal lamina lies a uniform layer of circular muscle which varies in thickness from relatively thin in ptychoderids (Hyman 1959, Cameron & Ostiguy 2013) to conspicuously thick in spengelids (Cameron & Perez 2012). Beneath the circular muscle layer, proboscis longitudinal muscles have been reported to possess several forms of organization (**Figure 7**). Harrimaniids have the most reported variation in longitudinal muscle arrangements among enteropneust families, organized as concentric rings (diagnostic of the genus *Saccoglossus**;*** **Figure 7A**), radial plates (present in the genera *Stereobalanus, Harrimania, Horstia,* and *Protoglossus;* **Figure 7B**), or diffuse (present in the genera *Ritteria, Saxipendium,* and *Mesoglossus**;*** **Figure 7C**) (Deland et al. 2010). The arrangement of longitudinal muscle fibers in ptychoderids is as radial plates (Spengel 1893, Deland et al. 2010), diffuse in Spengelids (Cameron & Perez 2012), and diffuse or radially arranged in Torquaratorids (Holland et al. 2012, Priede et al. 2012). Longitudinal muscle fibers in the proboscis may be physically partitioned by the dorsal protocoel mesentery. Proboscis musculature is instrumental to acorn worm locomotion, both in the more common sediment-dwelling species (e.g. *Saccoglossus spp.;* Knight-Jones 1952, Hyman 1959) and epibenthic animals (e.g., *Allapasus isidis;* Jones et al. 2013).

Collar musculature is most well-documented within Ptychoderidae by Spengel (1893), Van der Horst (1939), and Hyman (1959). A layer of circular muscle lies directly beneath the basal lamina of the collarette proximal to the proboscis stalk (**Figure 8**). Opposite to this, anchoring to the exterior tissues of the collarette, bundles of muscle separate basal lamina and the circular muscle layer, ultimately traversing diagonally through the collar coelom and attaching to the collar-trunk septum proximal to the buccal cavity. In Harrimaniids and Spengelids, the longitudinal muscles in the collar are present as two bundles, each flanking the buccal cavity on the left or right (Deland et al. 2010, Cameron & Perez 2012). Internally, a layer of circular muscle surrounds the buccal epithelium. Interestingly, conspicuously dense muscle fibers are present on the left and right sides of the buccal cavity in *Tergivelum* (Holland et al. 2009), a synapomorphy

for the genus (Priede et al. 2012). The perihaemal coeloms are also filled with longitudinal muscle (Hyman 1959).

Similar to the proboscis and collar, trunk muscle organization varies between major enteropneust groups (Hyman 1959). In members of Ptychoderidae, circular muscle fibers are exterior to the longitudinal muscle (Spengel 1893, Cameron & Ostiguy 2013). In contrast, the circular muscle layer in Spengelids lies beneath the longitudinal muscles of the trunk (Cameron & Perez 2012). Harrimaniids lack trunk circular muscle fibers, which constitutes a diagnostic feature for the family (Deland et al. 2010). In harrimaniids, however, the longitudinal muscles, which dominantly occupy either side of the trunk's ventral septum, may form two dense longitudinal bands (Hyman 1959, Deland et al. 2010). There is little reported evidence for muscle organization in torquaratorid trunks.

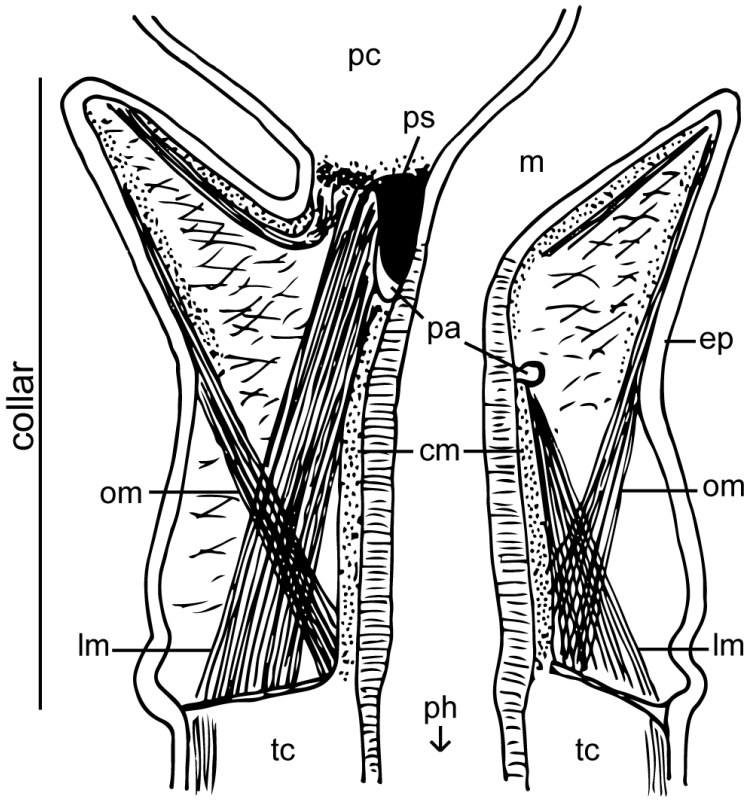


Figure 8: Sagittal section diagram through a generalized ptychoderid collar adapted from Van der Horst (1939). Figure is oriented as ventrum on the right and anterior pointed upward. *cm,* circular muscle filling peribuccal coelom of buccal cavity; *ep*, epidermis; *lm,* longitudinal muscle; *m*, mouth; *om*, oblique muscle; *pa,* peribuccal artery (subsection of the lateral arteries, see section **2.2.7 Circulatory System**); *pc,* proboscis coelom; *ph;* pharynx; *ps,* proboscis skeleton; *tc,* trunk coelom.

The pericardial epithelium can be regionally subdivided into an outer protocolemic epithelium and an inner pericardial epithelium (**Figure 9**) (Spengel 1893, Schepotieff 1907, Van der Horst 1939, Benito & Pardos 1997). These two regions are richly muscularized as myoepithelia and their patterns are most well-described in *Saccoglossus* (Harrimaniidae; Balser & Ruppert 1990) and *Glossobalanus* (Ptychoderidae; Benito & Pardos 1997). The muscle arrangement of the two epithelial surfaces are perpendicularly arranged relative to one another, with longitudinal muscles in the protocoelomic epithelium and circular muscles in the pericardial epithelium – an arrangement hypothesized to optimize contractile pressure (Benito & Pardos 1997). In the spengelid genus *Schizocardium*, the pericardium bifurcates anteriorly – a diagnostic character for the genus (Cameron & Perez 2012). Van der Horst (1939) additionally describes two species of

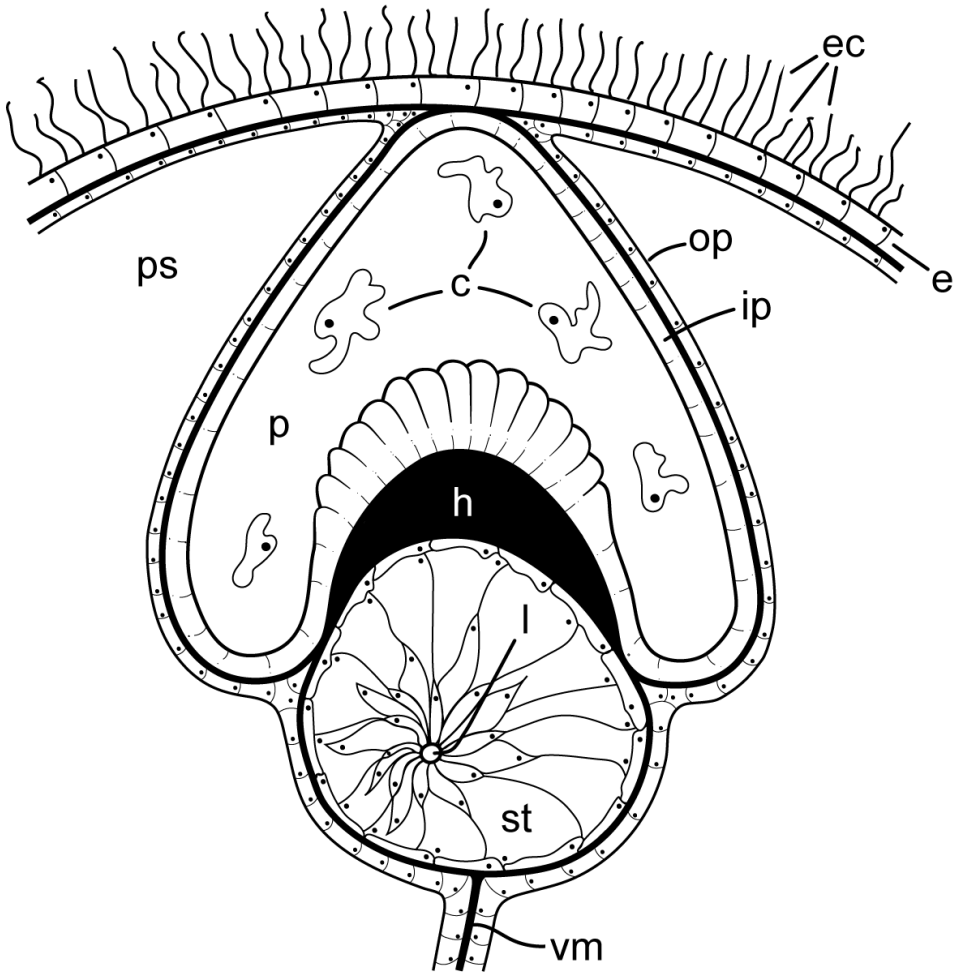


Figure 9: Diagrammatic cross section of enteropneust heart-kidney complex. *c,* coelomocytes; *e*, epidermis; *ec*, epidermal cilia; *h*, heart; *ip*, pericardial epithelium; *l,* stomochord lumen; *op,* protocoel epithelium; *p,* pericardium; *ps*, protocoel; *st,* stomochord epithelium.

Balanoglossus which possess a pericardium which completely encloses the blood-vessel sinus – superficially resembling the closed anatomy of the vertebrate heart (Hyman 1959).

2.2.5 Nervous System & Sensory Structures

Enteropneust nervous systems have been extensively documented in the classic works of Bullock (1940, 1945), Silén (1950) (considering *Glossobalanus marginatus*), and Knight-Jones (1952) (considering *Saccoglossus ruber,* previously *S. cambrensis*), as well as being discussed in the broader anatomical surveys of Van der Horst (1939), Dawydoff (1948), and Hyman (1959).

Acorn worm nervous systems take form dominantly as an intraepidermal (or basiepidermal) nerve plexus – filling space between epidermal cell stalks and the basal lamina, permeating anteroposteriorly, circularly, and obliquely (**Figure 6B & C**) (Bullock 1945, Hyman 1959). Within the proboscis, the nerve plexus is often homogenous and disperse from anterior to posterior, with a particular exception among some species in the genus *Saccoglossus*, where nerve fibers condense mid-dorsally (**Figure 10**) (Bullock 1945, Knight-Jones 1952, Cameron et al. 2010). Between species, the relative thickness of the nerve plexus layer within the epidermis also varies. Just anterior to the proboscis stalk, nerve fibers aggregate and tend to radiate circularly resembling a nerve ring (Knight-Jones 1952, Hyman 1959). In the proboscis *Saccoglossus,* the longitudinal nerve bundles originating at the nerve ring radiate anteriorly towards the tip of tip of the proboscis (Knight-Jones 1952). Nerve fibers condense dorsolaterally within the proboscis stalk (resembling a fan-shape) and are weakly structured ventrally, nearly absent below the proboscis skeleton within the stalk (Bullock 1945, Knight-Jones 1952). The nerve plexus of the collar is organized irregularly anteriorly, and nerves become increasingly longitudinally oriented in the posterior collar epidermis, ultimately joining the prebranchial nerve rings within the posterior collar (**Figure 10C**). Within the trunk, the intraepithelial nerve plexus persists constitutively while also condensing both dorsally and ventrally to form nerve cords, with the ventral nerve cord conspicuously denser than its dorsal counterpart in most enteropneust taxa (Spengel 1893, Van der Horst 1939, Bullock 1945, Silén 1950, Knight-Jones 1952), with a possible exception among torquaratorids (Priede et al. 2012). Generally, the trunk's nerve plexus shows greater organization in ventral and anterior portions than the dorsal and posterior portions, respectively (Knight-Jones 1952).

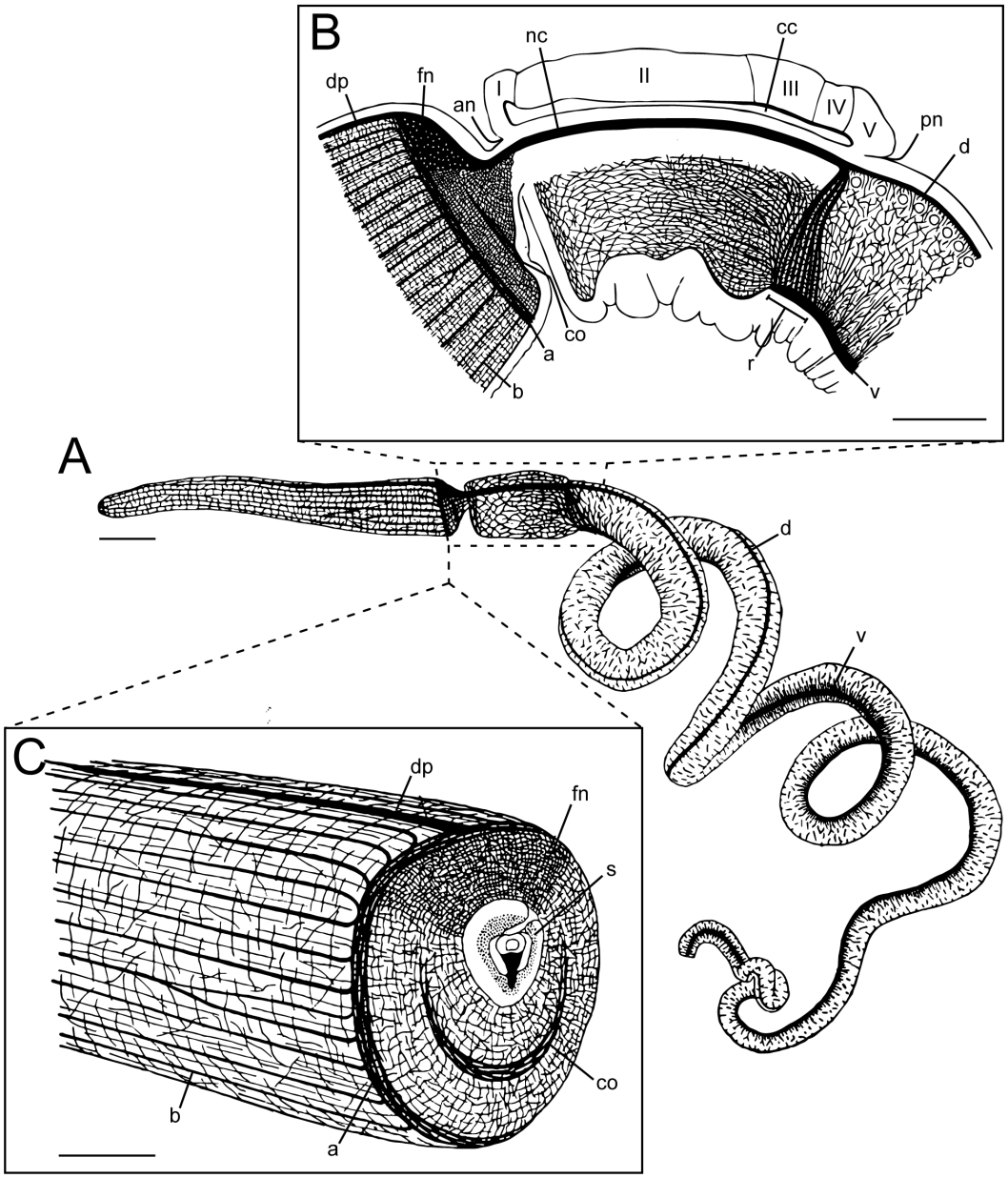


Figure 10: Diagrammatic representation of nervous system organization in *Saccoglossus cambrensis* (Adapted from Knight-Jones 1952). In all images, anterior is left and posterior is right. **A**) Gross organization of nervous system across body regions. Scale bar = 2mm. **B**) Sagittal section of posterior proboscis, collar, and anterior trunk found within the box highlighted in **A**. Scale bar = 0.5mm. **C**) Transverse section of posterior proboscis at the level of the left-most dotted line in **A**. Scale bar = 0.5mm. *a*, anterior nerve ring; *an*, anterior neuropore; b, *longitudinal nerve bundles*; *cc*, collar coelom; *co;* crescentic nerve bundle of the pre-oral ciliary organ; *d*, dorsal nerve bundle of the trunk; *dp*, dorsal nerve bundle of the proboscis; *fn*, fan-shaped thickening of the dorsal proboscis nerve-layer; *nc*, neurocord; *pn,* posterior neuropore; *r*, prebranchial nerve ring(s); *v*, ventral nerve bundle of the trunk; *I-V,* various histological regions of the dorsal collar epidermis.

In addition to the nerve plexus, enteropneusts possess only a single truly internalized nerve cord (often referred to as the collar cord) held within the dorsum of the collar which is contiguous with the dorsal intraepidermal nerve plexus both anteriorly towards the proboscis and posteriorly towards the trunk. The collar cord possesses nerve fibers primarily oriented in a longitudinal manner and the cord may contain a continuous lumen (e.g., *Ptychodera bahamensis*), lack any internal cavity, (e.g., *Schizocardium peruvianum*), or possess small, scattered crescentic spaces called lacunae (Bullock 1945, Silén 1950, Knight-Jones 1952, Hyman 1959, Deland et al. 2010, Cameron & Perez 2012, and Cameron & Ostiguy 2013). The collar cord cavity may also open to the environment via anterior and/or posterior neuropores, with the former being absent in Ptychoderidae (Cameron 2005, Knight-Jones 1952).

The collar cord has classically received much attention as the hypothesized nervous center in enteropneusts – primarily made based on its gross morphological similarity to the dorsal hollow nerve cord of chordates. Unlike the traditional dorsal hollow nerve cord of chordates, the collar cord is restricted to a relatively small region of the enteropneust, appears to be exclusively involved in signal conduction and not information processing (Dilly et al. 1970, Cameron & Mackie 1996), and is not a site of concentration for efferent nerves (Bullock 1945, Silén 1950, Knight-Jones 1952, Hyman 1959, Ruppert 2005). Despite these discrepancies, the collar cord of hemichordates does arise developmentally through coordinated tissue movements reminiscent of chordate neurulation (Nomaksteinsky et al. 2009). Interesting still, Lowe et al. (2003) have shown the ectodermal nerve plexus of *Saccoglossus kowalevskii* expresses orthologs to 22 chordate central nervous system patterning genes in an identical anteroposterior pattern as their counterparts in the chordate central nervous system – further complicating central nervous system homology between enteropneust and chordate nervous systems. With the advent of molecular tools addressing homology of tissues, the homology of the hemichordate collar cord and chordate dorsal hollow nerve cord has been the subject of intense experimental scrutiny and remains a topic of great interest (see Luttrell et al. 2012, Miyamoto & Wada 2013, and Kaul-Strehlow et al. 2017).

Sensory cells are dispersed across most epidermal regions, though especially concentrated on the surface of the proboscis. Notably, on the ventral surface of the proboscis anterior to the mouth, a U-shaped preoral ciliary organ is present and likely functions in chemoreception (Bullock 1940, Brambell & Goodhart 1941, Bullock 1945, Knight-Jones 1952). The preoral ciliary organ

has also been hypothesized to be homologous to the cephalochordate wheel-organ and vertebrate hypophysis (Brambell & Cole 1939, Gonzalez & Cameron 2009), though this hypothesis has yet to receive experimental support. With regard to photoreception, the enteropneust tornaria larvae possess apical eye-spots (see section **2.3 Reproduction & Development**), and adult worms are known to be negatively phototaxic, possessing light-sensitive tissues concentrated in the collar and proboscis (Hess 1936, Hess 1938). The availability of genetic datasets for a wide range of hemichordate taxa (Cannon et al. 2014), in addition to full genome sequences on *Ptychodera flava* and *Saccoglossus kowalevskii* (Simakov et al. 2015), have facilitated the identification of opsin genes among hemichordates and other ambulacrarians (D'Aniello et al. 2015), ultimately laying a foundation for further study on photoreception among hemichordates as well as many invertebrate deuterostomes.

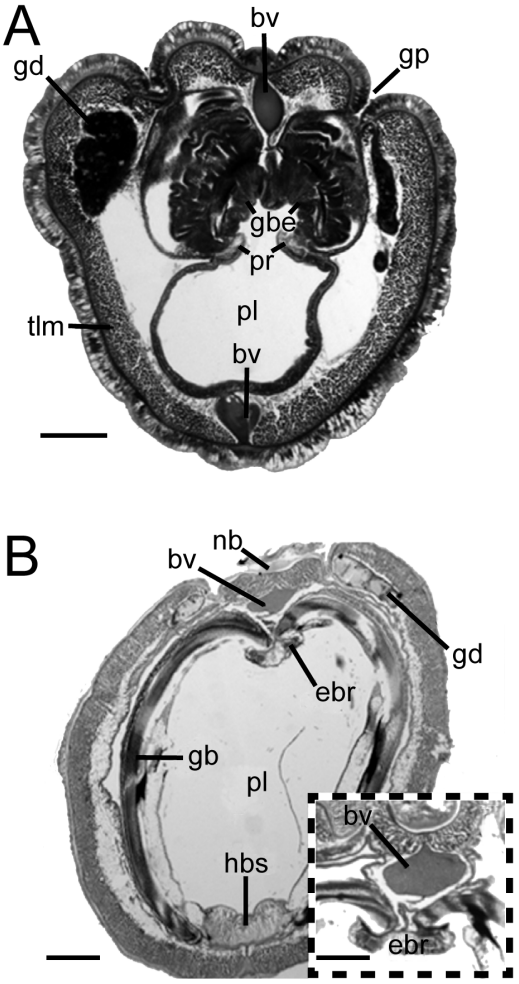


Figure 11: Light micrograph cross sections of enteropneust pharyngeal gut morphology. **A**) *Glossobalanus hartmanae* (Ptychoderidae) from Cameron & Ostiguy (2013). Scale bar = 0.5mm. **B**) *Schizocardium californicum* (Spengelidae) from Cameron & Perez (2012). Inset focuses on dorsal blood vessel. Scale bar = 0.5mm (0.3mm for inset). bv, blood vessel; *ebr*, epibranchial ridge; *gb*, gill bar; *gbe*, gill bar epithelium; *gd*, gonad; *gp*, gill pore; *hbs*, hypobranchial strip; *nb*, nerve bundle; *pl*, pharyngeal lumen; *pr,* parabranchial ridge; *tlm*, trunk longitudenal muscle.

2.2.6 Digestive & Branchial Systems

2.2.6.1 Digestive System

The enteropneust digestive system is organized as an alimentary canal of four major partitions: buccal tube, pharynx, esophagus, and intestine. As mentioned above (see section **2.2.1 General Anatomy**), histological and/or external regionalization of these partitions may or may not be readily apparent by gross anatomy in a taxon-specific manner.

The mouth, which opens ventrally between the proboscis and ventral collarette, is followed by the buccal tube, a digestive cavity lined with a ciliated and vacuolated columnar epithelium borne exclusively within the collar (Van der Horst 1939, Knight-Jones 1953). Anteriorly, a buccal diverticulum (commonly called the stomochord) extends into the proboscis and acts primarily as a scaffolding element for the heart-kidney complex. It should be mentioned the homology of the stomochord to the chordate notochord has been hotly debated since Bateson (1886) (for further reading, see Newell 1952, Silén 1954, and Annona et al. 2015) but will not be discussed further in that context here. Diagnostic of spengelid taxa, the stomochord additionally possesses a long anterior process extending into the proboscis muscle matrix (Hyman 1959). Internally, the stomochord is histologically similar to the buccal tube and has been reviewed extensively in Benito & Pardos (1997). The buccal tube may also possess lateral constrictions caused by the pressing of the two horns of the proboscis skeleton (**Figure 3**) into the lateral surfaces of the tube (Hyman 1959).

Posterior to the buccal tube following the collar-trunk boundary, the digestive tract progresses into the pharynx (**Figure 2B**). Dorsoventral differentiation between digestive (ventral) and branchial (dorsal) pharynx is often present (**Figure 11**), though varies between taxa. For example, the ratio between digestive and branchial pharynx in the genus *Protoglossus* is more or less equal (Deland et al. 2010). Similarly, in *Ptychodera flava* the dorsoventral bias is slight, though distinguishable by the presence of a lateral constriction separating the two (**Figure 11A & B**) (Hyman 1959). Notably, this same constriction is also present among torquaratorids (Priede et al. 2012). In contrast, the genus *Schizocardium* is diagnosable by the presence of gills extending nearly the whole circumference of pharyngeal gut, relieving only a small hypobranchial strip for digestive pharynx (**Figure 11C & D**) (Cameron & Perez 2012). Regarding the digestive portion of the pharynx, it is composed primarily of glandular and ciliated cell types (Knight-Jones 1953). Posteriorly, the branchial components (see section **2.2.6.2 Branchial System**) of the pharynx diminish gradually, giving way to the esophagus.

The esophagus immediately follows the termination of gill slit posterior of the pharynx. Within the esophagus, the gut epithelium may become increasingly undulatory in transverse section (as is the case for most harrimaniid taxa and Spengelidae) or remain consistent along its length (as is the case for Ptychoderidae and the harrimaniid genus *Protoglossus*) (Hyman 1959).

In *Glandiceps* (Spengelidae), the esophagus becomes dorsoventrally differentiated as it progresses – with increased furrow depths localized to the ventral epithelium (Spengel 1907). In the esophagi of some spengelid and harrimaniid taxa, dorsal pores (hypothesized to be degenerate gill slits by Hyman 1959) may be present (see Deland et al. 2010, Cameron et al. 2010, and Cameron & Perez 2012 for taxon-specific details), allowing for communication between the esophageal lumen and the environment.

The fourth and final section of the gut is intestine. Anteriorly, the intestine may possess readily visible dorsal hepatic sacculations arranged in bilaterally-organized pairs (present in ptychoderids and *Schizocardium*) and are involved in digestion when present (Bridges & Woodwick 1994). In the absence of these organs, the dorsal epithelium of the intestine may still be distinguishable from the rest of the intestinal epithelium through the presence of brown-green pigment and is intensely glandular (Hyman 1959) – likely performing digestive functions similar to the hepatic sacculations described above. In Ptychoderid taxa, a longitudinally oriented dorsolateral groove may also be present on one or both sides of the intestinal lumen (**Figure 11A & B**) – often diminishing and disappearing posteriorly (Cameron & Ostiguy 2013). Following the hepatic region, digestive epithelium remains primarily ciliated and is referred to as posthepatic intestine. Whereas in most enteropneust taxa the posthepatic intestine is relatively featureless (terminating at the anus), ptychoderids feature a longitudinally oriented subepithelial band ventrally situated to the intestinal lumen (Willey 1899a). Although its function is unclear, the pygochord has also been posited to be homologous to the chordate notochord (Willey 1899b), though with little success (Annona et al. 2015).

2.2.6.2 Branchial System

The enteropneust branchial system is present on the dorsolateral surfaces of the pharyngeal gut and commonly conforms to a similar gross morphology across species (**Figure 2B**). A notable exception to this pattern is in *Stereobalanus*, where extensive fusion of branchiogenital tissues has occurred (Reinhard 1942). The number of gill slit pairs can be a valuable diagnostic feature (Van der Horst 1939) and can vary anywhere between a single pair in the meiofaunal taxon *Meioglossus psammophilus* (Worsaae et al. 2012) to indeterminate in *Balanoglossus aurantiacus* (Harmer 1910). The number of gill slits present in an individual also increases ontogenically, with new pairs

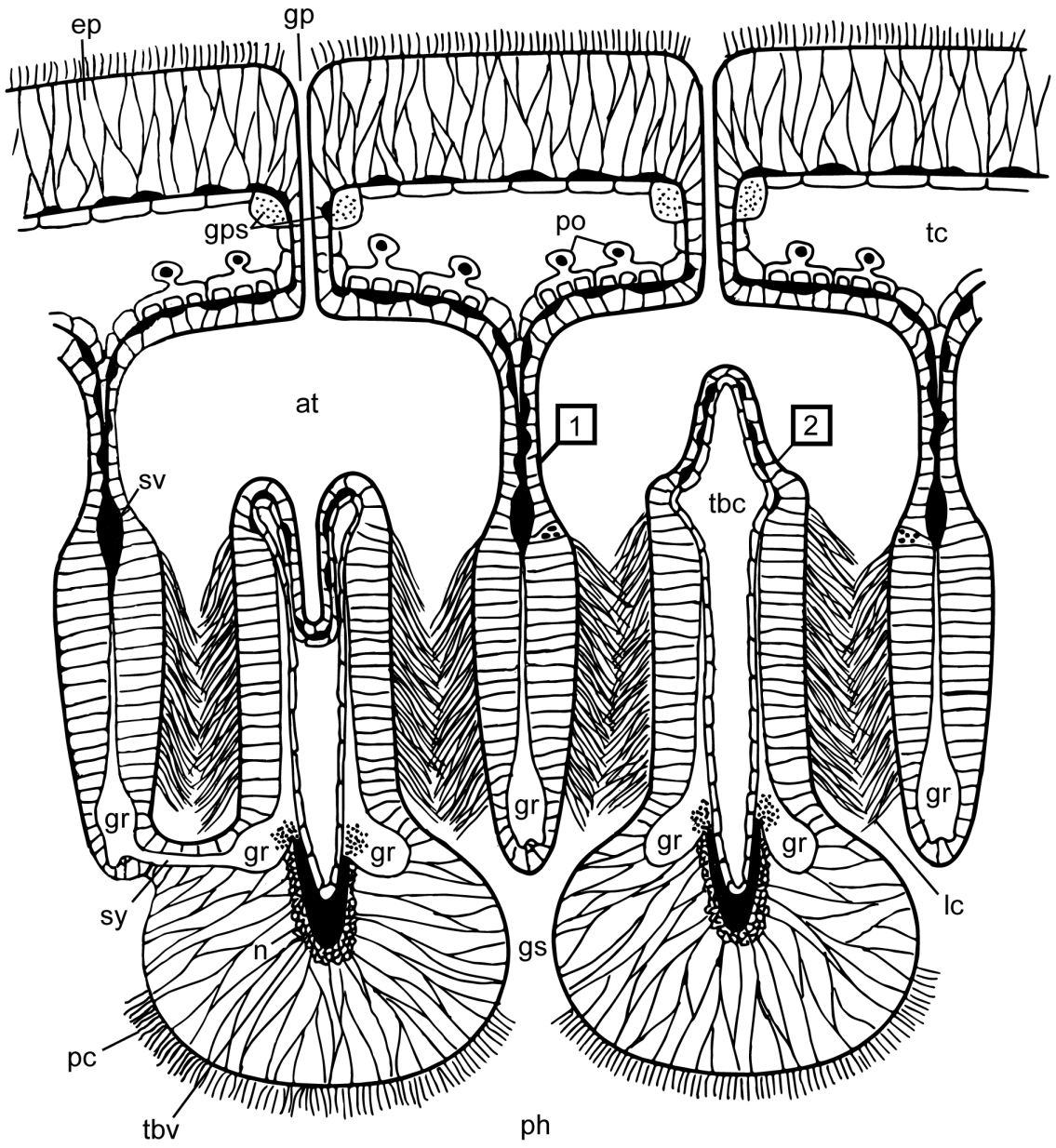


Figure 12: Parasagittal section through gill slits and associated structures adapted from Benito & Pardos (1997). External environment oriented towards the top of the image. *1*, primary gill bar (or septum); *2*, secondary gill bar (or tongue bar); *at*, atrium; *ep*, epidermis; *gp*, gill pore; *gps*, gill pore sphincter; *gr*, gill skeletal rod; *gs*, gill slit; *lc*, lateral cilia; *n*, nerve net; *pc*, pharyngeal cilia; *ph*, pharynx lumen; *po*, podocytes; *sv*, septal blood vessel; *sy,* synapticle; *tbc*, tongue bar coelom; *tbv*, tongue bar vessel.

being added on the posterior boundaries of the pharynx with age (Hyman 1959). In an evolutionary context, molecular and phylogenetic studies investigating the homology between hemichordate and chordate gill slits have supported a single common ancestry, placing the origin of pharyngeal gill slits at the last common ancestor of deuterostomes (Halanych 1995, Gillis et al. 2012).

Internally, the enteropneust branchial system initially develops as bilaterally symmetric pores distending from the dorsolateral surfaces of the pharyngeal epithelium and fusing with ectodermal epithelium of the trunk (**Figure 2B**) (Bateson 1885). Each gill slit remains separated from its sister via a dorsal strip of gut epithelium called the epibranchial ridge (**Figure 11C & D**). Beginning ellipsoidal in shape, the ventromedial surface of the gill slit's epithelium extends dorsally while the gill silt itself extends ventrally on either side, ultimately forming the gill pore's mature U-shape (**Figure 5**). The tissue intermediate to the two "prongs" of each U-shaped slit is referred to as a tongue bar (or secondary bar) and contains coelomic space internally; in contrast, the tissue separating each U-shaped gill aperture is called a septum (or primary bar) and is solid in construction (Harmer 1910, Hyman 1959). The lateral surface of each gill slit is richly ciliated (**Figure 12**), likely to facilitate water flow for gas exchange (Benito & Pardos 1997). As such, the branchial system is well vascularized (see section **2.2.7 Excretory & Circulatory Systems**) (van der Horst 1939, Hyman 1959).

Primary and secondary gill bars are structurally fortified by the inclusion of hairpin-shaped skeletal elements comprised of acellular cartilage (Cole & Hall 2004, Rychel et al. 2005, Rychel & Swalla 2007). Each skeletal element is approximately U-shaped, but inverted relative to the gill slit ultimately allowing each skeletal element to extend one arm into the primary gill bar (i.e., the septum) and the second arm into the secondary gill bar (i.e., tongue bar; **Figure 13**) (Spengel 1893, Van der Horst 1939). Whereas each primary and secondary gill bar contains two rods extended by two individual skeletal elements (with the exception of the first and last gill slit which are only supported by a single skeletal element), the rods extending into each tongue bar remain separated whereas the rods within each septum become fused (**Figure 13**) (Burdon-Jones 1956, Hyman 1959). Successive skeletal elements may also contain lateral processes, collectively called synapticulae, connecting skeletal rods of the tongue bars to those of the septa. The presence/absence of synapticulae varies extensively among hemichordate taxa – present in

Ptychoderidae, *Spengelia* (Spengelidae), and *Schizocardium* (Spengelidae), and absent in Torquaratoridae, Harrimaniidae, and *Willeyia* (Spengelidae) (Cameron 2005, Deland et al. 2010).

Externally, individual gill slits communicate to the environment by first opening into an atrium (or branchial sac) – a cavity intermediate between the gill slit and the circular gill pore situated on the dorsolateral surface of the epidermis (**Figure 2B & 12**) (Dawydoff 1948, Hyman 1959). For most enteropneusts, each gill slit is paired with a single branchial sac, although the first few gill slits may share a common atrium and gill pore (Van der Horst 1939, Hyman 1959). In contrast, Dawydoff (1948) additionally described the gill slits of *Balanoglossus misakiensis* (Ptychoderidae), *Willyia bisulcate* (Spengelidae), and *Glossobalanus ruficollis* (Ptychoderidae) are often partitioned in such a way that multiple gill slits are grouped together to share single atria and their associated gill pores. More discrepant still, each row of the branchial apparatus in *Stereobalanus canadensis* is associated with a single deep longitudinal slit on the pharynx – unique among enteropneusts (Reinhard 1942).

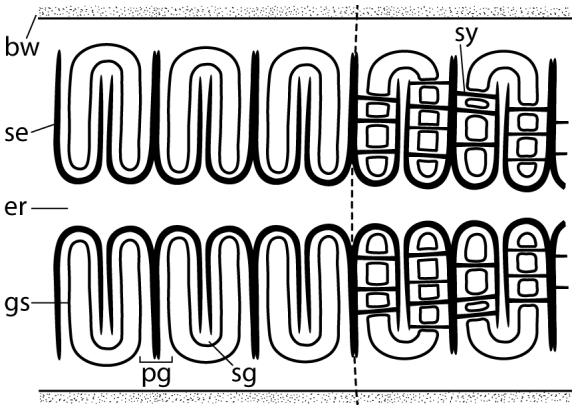


Figure 13: Gill skeletal structures as illustrated from a dorsal cutaway of the pharyngeal gut epithelium. Arms of two adjacent skeletal elements fuse within primary gill bars, whereas arms extending into tongue bars remain separate entities. Dotted line represents hypothetical separation between skeletal elements without synapticulae (left) and skeletal elements possessing synapticulae (right). *bw*, body wall; *er*, epibranchial ridge; *gs*, gill slit; *pg*, primary gill bar (septum); *se*, skeletal element; *sg*, secondary gill bar (tongue bar); *sy*, synapticulae

2.2.7 Excretory & Circulatory Systems

The enteropneust circulatory system is regionally specialized and comprised of an anterior and posterior provision relative to the contractile pericardium and blood-vessel sinus (or heart), both of which are situated dorsally upon the stomochord (**Figures 2A, 9, & 14**) (Spengel 1893, Schepotieff 1907, Van der Horst 1939). The anatomy of the pericardium itself is covered in section **2.2.3 Body Cavities** above. Following contraction of the pericardium, arterial blood passes anteriorly into the glomerulus sinus – the primary excretory organ of the enteropneust. The glomerulus is composed of many small blood sinuses lined with monocellular podocytes essential

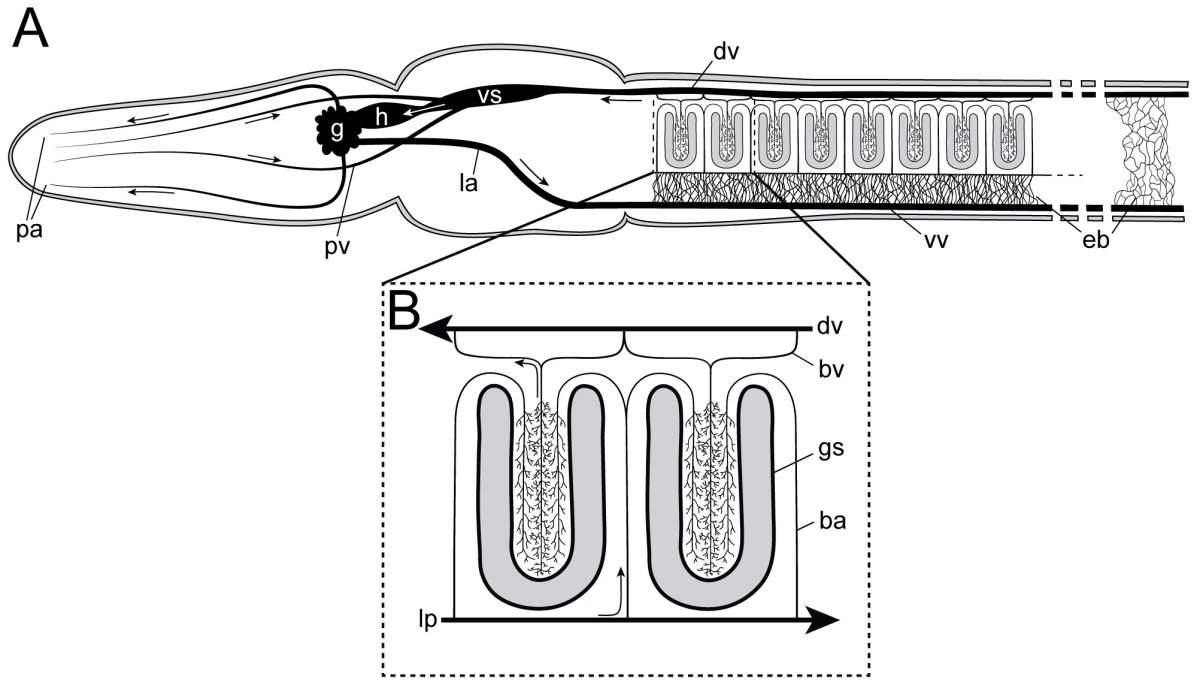


Figure 14: Diagram of vascular system and blood flow across enteropneust anatomy. Arrows denote the direction of blood flow for their associated structures. **A**) Generalized schematic of blood flow and various vascular structures. **B**) Magnification of branchial vascular system. *ba*, afferent branchial artery; *bv*, efferent branchial vein; *dv*, dorsal vessel; *eb*, epidermal blood plexus; *g*, glomerulus; *gs*, gill slit; *h*, heart; *la,* lateral (peribuccal) arteries; *lp*, lateral pharyngeal vessel; *pa,* proboscis arteries; *pv*, proboscis veins; *vs*, venous sinus; *vv,* ventral vessel.

to the filtering blood of metabolic waste which is ultimately converted into urine and excreted via the proboscis pore (Balser & Ruppert 1990, Benito & Pardos 1997).

Following filtration in the glomerulus, the circulatory system functionally bifurcates. Anteriorly, a dorsal and ventral artery supply blood to the proboscis (**Figure 14A**). These arteries reside within the dorsal and ventral proboscis mesenteries, respectively, until reaching the proboscis wall where they both branch considerably (Van der Horst 1939, Hyman 1959). Posteriorly, blood leaves the glomerulus via two lateral arteries flanking either side of the proboscis complex, penetrates into the collar, and finally wraps ventrally – coalescing as the ventral blood vessel (**Figure 14A**). The collar itself is supplied blood by a ventral collar artery – an anterior projection of the ventral blood vessel – imbedded in the ventral collar septum when present (Hyman 1959). The collar is additionally supplied blood by two lacunar networks occupying the coelomic surface beneath the epidermis and buccal tube epithelium.

The ventral vessel's primary function is as the main artery supplying blood to the trunk (**Figure 14**). The circulatory system's organization differs between the pharyngeal and post-pharyngeal gut. Associated with both the left and right halves of the pharynx, a lateral vessel supplies blood to the enteropneust's branchial apparatus (**Figure 14**). Each individual gill slit receives two arteries which permeate the lateral sides of a tongue bar (see section **2.2.6.2 Branchial System**) and branch into an arteriole plexus (**Figure 14B**). From here, blood enters venous extensions of dorsal blood vessel and is ultimately transported anteriorly to the heart sinus via the dorsal blood vessel, completing the cycle from the perspective of the branchial system (Pardos & Benito 1984). With respect to blood supply in non-pharyngeal trunk tissues, the ventral blood vessel diverts blood into the subepithelial plexus of the epidermis and subepithelial lacunae of the gut where it ultimately rejoins with the dorsal blood vessel, completing the circulatory cycle from the perspective of non-pharyngeal trunk tissues (Pardos & Benito 1990).

Enteropneust blood is comprised of two major components: blood pigment and blood cells (Benito & Pardos 1997). Blood pigment is extracellular and ranges from 15-25nm – data interpolated primarily from *Saccoglossus* (Dilly 1969, Balser & Ruppert 1990) and *Glossobalanus* (Wilke 1972, Pardos & Benito 1990). Blood cells (also called amoebocytes or amoeboid cells) are poorly characterized but are known to extend pseudopodial processes which facilitated pinocytosis of blood pigment particles (Pardos & Benito 1990).

2.3 Reproduction & Development

2.3.1 Reproductive System

Enteropneusts are dioecious, though generally indistinguishable based on external morphology with the exception of some tissue translucency permitting visibility to the differentially pigmented gonads of males and females. The gonads themselves are often internalized within the genital ridges of the branchiogenital trunk or genital wings (present in Ptychoderidae and Torquaratoridae). Notably, however, *Allapasus aurantiacus* females possess primary ova which are carried externally in epidermal pouches (Holland et al. 2012). Gonads are round-to-ellipsoid in cross section and held within the dorsolateral peritoneum of the trunk coelom (**Figures 4 & 11**). Each gonad's associated gonopore is often associated with a single gill slit.

Enteropneusts naturally free-spawn and fertilization is external. Spawn timing in vivo is largely undocumented across taxa and is primarily anecdotal when available (Burdon-Jones 1951, Urata et al. 2012). Female egg masses are often imbedded deeply within diffuse mucus cords, and the number of eggs can range from under 100 to several thousands (Burdon-Jones 1951). Ova are relatively large for invertebrates (300-1000μm in diameter), and sperm follow the conventional animal arrangement (Colwin & Colwin 1962, Benito & Pardos 1997). Among harrimaniids, which are direct developers, eggs are particularly yolky (Kaul-Strehlow & Rottinger 2015).

2.3.2 Development

Enteropneusts undergo either direct development (Harrimaniidae; **Figure 15A-F**) or indirect development by way of a tornaria larva (Ptychoderidae, Spengelidae; **Figure 15G-L**). The developmental mode of Torquaratorids is currently undocumented, though brooding has been described in the torquaratorid taxon *Coleodesmium* (Osborn et al. 2013). Developmental biology among direct developers has been most extensively studied in the genus *Saccoglossus* including *S. kowalevskii* (Bateson 1885, Bateson 1886, Colwin & Colwin 1953, Kaul-strehlow & Stach 2013), *S. pusillus* (Davis 1908), and *S. horsti* (Burdon-Jones 1952), whereas indirect development has been sampled across genera including the spengelids *Schizocardium californicum* (Gonzalez et al. 2017) and *Glandiceps hacksi* (Urata et al. 2014), as well as in the ptychoderid genera *Balanoglossus* (Morgan 1891, Morgan 1894, Stiasny 1914a, Stiasny 1914b, Urata & Yamaguchi 2004, Miyamoto & Saito 2007) and *Ptychodera* (Tagawa et al. 1998, Nakajima et al. 2004, Nielsen & Hay-Schmidt 2007). In the context of preparation of embryos for cell and molecular biology, studies have focused on *Saccoglossus kowalevskii, Ptychodera flava*, and *Schizocardium californicum* (Lowe et al. 2004, Gonzalez et al. 2017).

Regardless of developmental mode, early cleavage events following fertilization and through to gastrulation are consistent across Enteropneusta (Kaul-Strehlow & Rottinger 2015). Upon fertilization, the vitelline envelope lifts from the zygote's cell membrane – the degree of separation varies between taxa (Urata & Yamaguchi 2004, Kaul-strehlow & Stach 2013). Enteropneust development follows the canonical holoblastic radial cleavage pattern; as such, the first two cleavages (zygote → 2-cell → 4-cell) occur in the meridional plane and are oriented perpendicularly with respect to one another. The third cleavage (4-cell → 8-cell) occurs in the longitudinal plane, producing two layers of cells (*i.e.,* animal and vegetal poles). Fourth cleavage

produces a single layer of 8 cells in the animal pole (future anterior ectoderm), whereas the four vegetal pole cells cleave longitudinally to yield a large vegetal tier (future middle/posterior ectoderm) and small vegetal tier (future endomesoderm) (Darras et al. 2011). Prior to gastrulation, the embryo flattens out along the anteroposterior axis, and the archenteron begins invaginating at the vegetal pole.

Following gastrulation in direct developers, such as *Saccoglossus kowalevskii*, the blastopore closes off, and the embryo begins to elongate along the anteroposterior axis (Colwin & Colwin 1953). Transversely in the posterior region of the embryo, a ciliated band forms (opisthotroch, **Figure 15A**), and the embryo can be seen spinning around within the vitelline envelope as a result of ciliary beating. As the ciliated embryo elongates, the external demarcations separating proboscis, collar, and trunk begin to form. First, a single groove develops circumferentially about the anterior regions of the embryo – demarcating the differentiating proboscis. This event coincides internally with the separation of the protocoel from the anterior archenteron while the paired meso- and metacoels begin to evaginate from the middle and posterior lateral-archenteron, respectively, via enterocoely (Kaul-strehlow & Stach 2013). At this point, pale pigment spots appear across the embryo (Colwin & Colwin 1953). Next, a second groove forms posterior to the first groove (though to a shallower depth) and delimits the boundary between collar and trunk (**Figure 15B**). This groove's appearance occurs coincidentally with the full separation of the meso- and metacoels from the archenteron (Kaul-Strehlow & Stach 2013). The sixth coelomic cavity, the pericardium, also differentiates at this point; in contrast to the other coeloms, however, the pericardium forms through schizocoely of the dorsal epidermis near the dorsoposterior regions of the proboscis cavity (Kaul-Strehlow & Stach 2011). Once the three major body regions are demarcated, the embryo elongates along its anteroposterior axis and flexes ventrally (commonly referred to as the early-kink phase, **Figure 15C**). A groove is sometimes visible along the dorsal midline as elongation and flexion is initiated (Colwin & Colwin 1953; **Figure 15D**). Prior to hatching, a single pair of gill pores perforate the dorsolateral surface of the anterior trunk; all following gill slits develop post-hatching and settlement (**Figure 15E**) (Colwin & Colwin 1953, Kaul-strehlow & Stach 2013). For *Saccoglossus kowalevskii*, developmental timing has been well characterized and is consistent – requiring approximately 5 days from fertilization to juvenile hatching at 22℃ (Colwin & Colwin 1953, Lowe et al. 2004, Kaul-Strehlow

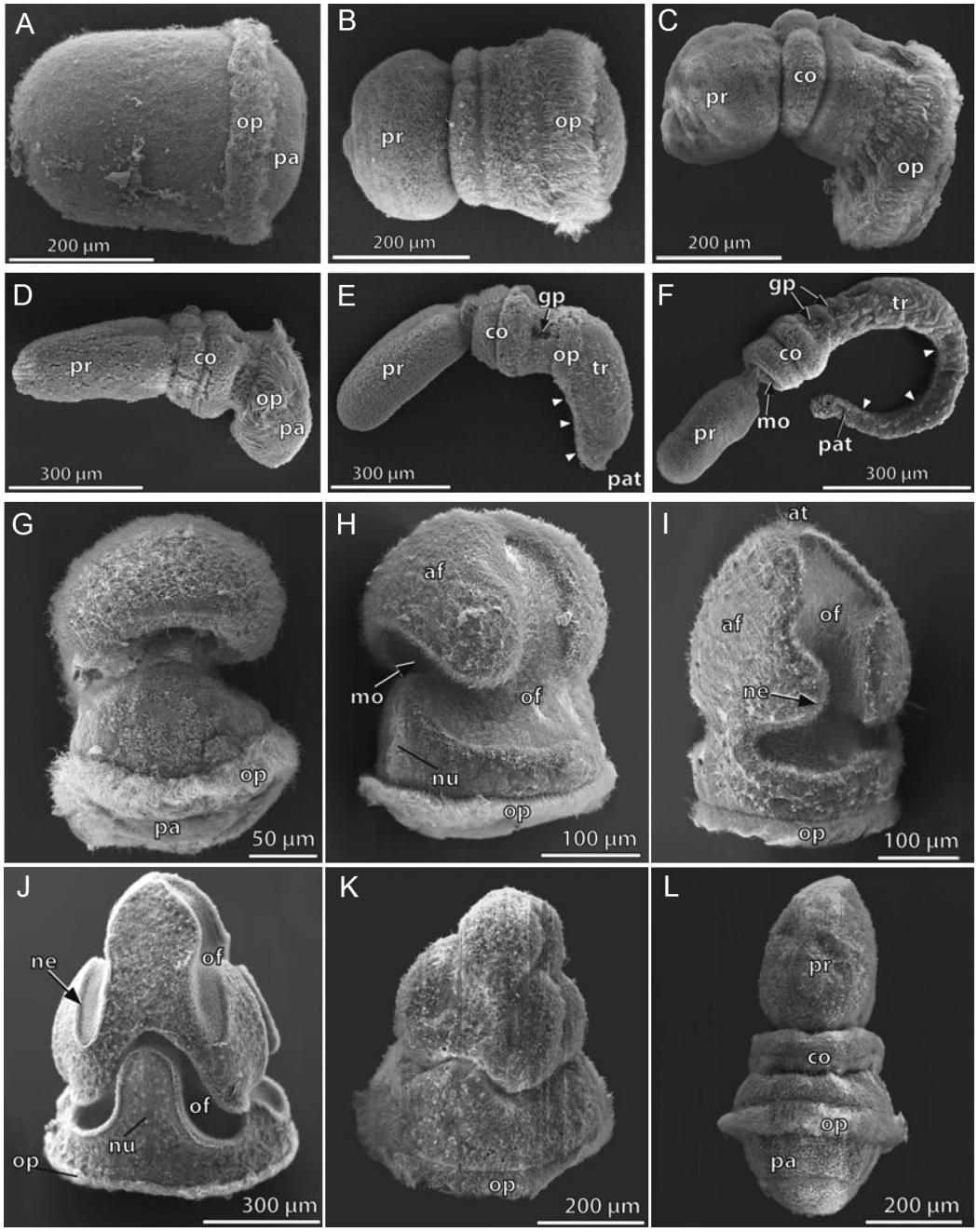


Figure 15. Late developmental schemes of *Saccoglossus kowalevskii* (direct developer, **A-F**) and *Balanoglossus misakiensis* (indirect developer, **G-L**) as portrayed in Kaul-Strehlow & Rottinger (2015). **A**) Late gastrula; **B**) Early kink; **C**) Dorsal flexure; **D**) One-gill-slit hatchling; **E**) Early settling juvaline. Arrowheads denote ventrally-biased elongation; **F**) Three-gill-slit juvenile; **G**) Early hatched Heider-stage tornaria; **H**) Late Heider-stage tornaria; **I**) Early Metschnikoff-stage tornaria; **J**) Late Metschnikoff-stage tornaria; **H**) Spengel-stage tornaria; **I**) Agassiz-stage tornaria. *af,* aboral field; *at,* apical tuft; *co,* collar; *gs*, gill slit; *gp*, gill pore; *mo,* mouth opening; *ne*, neotroch; *nu*, neurotroch; *of*, oral field; *op*, opisthotroch; *pa*, perianal feed; *pat*, postanal tail; *pr*, proboscis; *tr,* trunk.

& Rottinger 2015). Following hatching, the juvenile enteropneust utilizes the opisthotroch to swim, followed by settlement in sediment.

Indirect developing enteropneusts possess a larval intermediate called a tornaria larva (Bateson 1885, Tagawa et al. 1998). Following blastopore closure at the end of gastrulation, the internal protocoel separates from the archenteron, elongates dorsally, and fuses with the ectoderm to form the larval hydropore (a future proboscis pore) while the archenteron simultaneously extends towards the ventral ectoderm. Finally, the protocoel extends a thin process the anterior tip of the embryo underneath the future apical tuft while the archenteron bends ventrally in preparation for the development of the mouth post-hatching (Tagawa et al. 1998). At this point, the early tornaria larval stage hatches from the fertilization envelope as a Müller-stage tornaria (Kaul-Strehlow & Rottinger 2015). Once hatched, the early tornaria larva's apical tuft grows longer, the archenteron fuses with the ventral ectoderm to form the mouth, a perianal ciliary band emerges (telotroch), and the gut differentiates into the three gastrointestinal segments (esophagus, stomach, and intestine from anterior to posterior) while the overall size of the larva increases and elongates longitudinally (Tagawa et al. 1998, Nielsen & Hay-Schmidt 2007). This morphological stage is called the Heider-stage tornaria (**Figure 15G & H**). After gut segments have differentiated, the characteristic neotroch primary lobes of a dipleurula larvae develop and separates oral- and aboral-field – ultimately facilitating feeding in the Metchnikoff-stage tornaria (**Figure 15I & J**) (Bateson 1885, Nielsen & Hay-Schmidt 2007). Eye spots may be visible proximal to the apical tuft (Brandenburger et al. 1973). A ventral ciliated band forming the neurotroch is also present at this stage (Kaul-Strehlow & Rottinger 2015). As growth continues, the perianal monociliated cells develop compound cilia and will eventually become the post-oral opisthotroch in the late-tornaria larva (Nielsen & Hay-Schmidt 2007). The larvae will develop secondary lobes (sometimes called tentacles) associated with the neotroch of the oral-field until a fully mature Krohn-stage tornaria is formed (Hadfield 2002) – at which point it can persist in the water column for an undetermined length of time.

Transition from tornaria larva into a juvenile acorn worm is achieved through a regressive metamorphosis where several features become reduced. The first regressive metamorphic state is the Spengel-stage (**Figure 15K**), denoted externally by a reduction in overall larval size and the retrogression of the neotroch on the future proboscis (Hadfield 2002, Nielsen & Hay-Schmidt

2007, Kaul-Strehlow & Rottinger 2015). In coordination with the external modifications, the protocoel begins to enlarge within the preoral regions of the larva and the meso- and metacoels emerge via enterocoely (Kaul-Strehlow & Rottinger 2015). Next, the neotroch is lost entirely, and the adult body regions become distinct and elongate (Agassiz-stage, **Figure 15L**). A deep constriction demarcates the external boundary between proboscis and collar, the stomochord and its associated heart-kidney complex has developed internally, meso- and metacoels have expanded into their mature orientations, and one-to-few gill slits may be present on the dorsolateral surfaces of the trunk (Kaul-Strehlow & Rottinger 2015). Finally, at the time of settlement and metamorphosis, the larva will lose its opisthotroch and begins to resemble the adult morphology.

2.4 Distribution & Ecology

Enteropneusts are found throughout the world's oceans, from the intertidal to >8,000 meters in depth. Tassia et al. (2016) review the global distribution of recognized species, although the authors caution that much hemichordate biodiversity remains to be described, so any such report will be an underestimate. Enteropneusts are benthic, living in burrows, under stones, among algal holdfasts, or on the surface of deep sea sediments (Hyman 1959, Osborn et al. 2011). *Saccoglossus* and *Balanoglossus* are particularly known for their burrowing behavior. Some enteropneusts ingest sediment, extract organic material during digestion, and then produce a characteristic fecal casting on the benthic surface. Not all enteropneusts produce these castings, however. There have been few studies making use of time lapse imaging to study the fecal castings of deep sea torquaratorid enteropneusts (Smith et al. 2005, Anderson et al. 2011, Jones et al. 2013).

Many enteropneusts give off a characteristic odor, presumably from the bromo-organics they produce, which have been hypothesized to be unpalatable defensive compounds (Thomas 1972, Fielman and Targett 1995, King et al. 1995, Kicklighter et al. 2003). Kicklighter et al. (2004) found that *Saccoglossus kowalevskii* was unpalatable to two sympatric species of fish due to its high concentrations of 2,3,4-tribromopyrrole, although the worm was still readily consumed by a sympatric crab.

There are few studies on acorn worms' species interactions. Some species of enteropneusts have been reported to live inside the burrows of other species, for example, *Glossobalanus ruficollis* is found inside *Balanoglossus carnosus* burrows (Willey 1899c, Okuda 1939, Hyman 1959). Protozoan parasites, particularly gregarines, have been reported in the hepatic sacs of

Balanoglossus clavigerus, Harrimaniia kupferri (Hyman 1959), and *Glossobalanus minutus* (Wakeman et al. 2014). A parasitic copepod in *Ptycodera flava* was described in Tung et al. (2014). Several species of ptychoderid have been observed to be bioluminescent, although the in vivo mechanism and the function of this response remain unknown (reviewed in Oba et al. 2017).

While ecological studies on adult enteropneusts are scarce, research on larval ecology is even more rare. Little is known about larval feeding or predation in tornaria larvae. One recent study found that growth rates in the tornaria larvae of *Schizocardium* sp. from Texas were adversely affected by exposure to chemically dispersed crude oil (Almeda et al. 2014). Ecological interactions of both enteropneust adults and larvae will benefit greatly from increased scientific scrutiny in the future.

2.5 Phylogeny

Since the advent of molecular phylogenetics, there has been a consensus on relationships between the three major deuterostome phyla, Chordata, Echinodermata, and Hemichordata (see discussion in Halanych 2004, Swalla and Smith 2008, Kocot et al. 2010). Phylogenomics has further upheld Ambulacraria (Bourlat et al. 2006, Dunn et al. 2008, Lartillot and Philippe 2008, Philippe et al. 2009, Cannon et al. 2014, Simakov et al. 2015), and there is additional support for this clade from morphology (Cameron 2005, Ruppert 2005), and a combined analysis of ribosomal, mitochondrial and nuclear protein-coding genes (Bourlat et al. 2008). The Ambulacraria hypothesis suggests that shared morphological features between hemichordates and chordates, such as pharyngeal gill slits and the post-anal tail, were likely to have been present in the deuterostome ancestor.

Given general agreement on the position of hemichordates within deuterostomes, focus has turned to resolving the relationships within Hemichordata. Historically, there were few explicit phylogenetic hypotheses, although traditional wisdom assumed pterobranchs and enteropneusts were reciprocally monophyletic taxa based on their disparate morphology and life history (Dawydoff 1948, Hyman 1959). Beginning with Halanych (1995), studies using 18S rDNA sequence data have recovered Enteropneusta as paraphyletic, with pterobranchs originating within the acorn worm lineage sister to Harrimaniidae (Cameron et al. 2000, Bourlat et al. 2003, Cannon

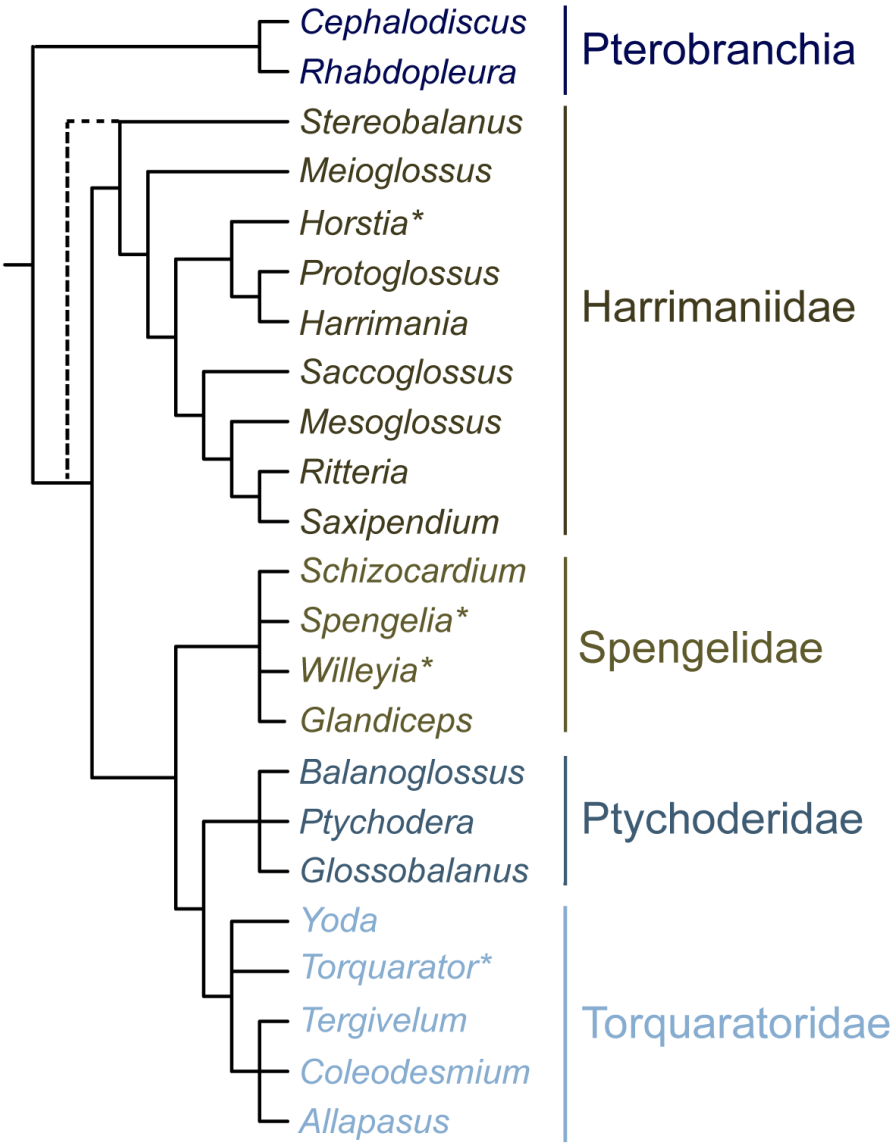


Figure 16. Phylogenetic hypothesis for the described genera of Enteropneusta. Tree summarizes molecular phylogenetic results from Cannon et al. 2009, Osborn et al. 2011, Worsaae et al. 2012, Cannon et al. 2013, Osborn et al. 2013, and Cannon et al. 2014, with the addition of morphological data from Holland et al. 2005 and Deland et al. 2010. Genus names followed by an asterisk are represented by morphological data only.

et al. 2009, Worsaae et al. 2012). In their 28S rDNA analyses, however, Winchell et al. (2002) found pterobranchs as sister to a monophyletic Enteropneusta, a result that has also been recovered by Cameron's (2005) morphological cladistic analysis, and phylogenomics (Cannon et al. 2014).

Within Enteropneusta, phylogenetic hypotheses have been lacking, although Hyman (1959) noted that harrimaniids were the least complex of the enteropneusts, followed by spengelids and ptychoderids. Aside from this type of subjective assessment, the earliest objective studies addressing hemichordate relationships included Halanych (1995), with 18S rDNA data from three hemichordate taxa, and Winchell et al. (2002), who included six hemichordates in a broader study on deuterostome relationships using 18S and 28S rDNA data. Cameron (2005) conducted a morphological cladistic analysis including all known hemichordate genera at the time, although resolution within Enteropneusta was poor.

Cannon et al. (2009) specifically addressed hemichordate relationships using 18S rDNA and two mitochondrial markers; several subsequent studies have built upon the Cannon et al. (2009) dataset (Osborn et al. 2011, Worsaee et al. 2012, Cannon et al. 2013, Cedhagen and Hansson, 2013, Holland et al. 2013). Cannon et al. (2014) included 14 hemichordate species from all recognized families in a phylogenomic analysis of datasets consisting of 162-299 genes. Generally, these studies have recovered traditional monophyletic enteropneust families with the exception of placing the unusual harrimaniid *Stereobalanus canadensis* as sister to all other enteropneusts (Cannon et al. 2014, Li et al. under review). *Saxipendium,* which had been considered to constitute its own family, was also found to be nested within Harrimaniidae (Cannon et al. 2009, 2014, Deland et al. 2010). Additionally, ribosomal, phylogenomic, and mitochondrial datasets have suggested that Torquaratoridae is within Ptychoderidae (Cannon et al. 2013, 2014, Li et al. under review). Although represented by no more than two species in any molecular phylogenetic analysis to date, Spengelidae has been consistently recovered as sister to Torquaratoridae + Ptychoderidae. Harrimaniidae is sister to the remaining enteropneust families. Molecular phylogenetic analyses published to date have not contained sufficient species coverage to thoroughly test monophyly of most genera, nor the relationships of genera within families. A phylogenetic hypothesis for the relationships of described extant hemichordate genera is given in **Figure 16**.

In addition to the described species of enteropneusts, there are several undescribed taxa that have been characterized molecularly or documented photographically (Cannon et al. 2009, Anderson et al. 2011, Osborn et al. 2011, Cannon et al. 2013). Cannon et al. (2009) include sequence data from a tornaria larva that is not phylogenetically close to any other species, falling

sister to the clade of Torquaratoridae + Ptychoderidae, but not within Spengelidae (Cannon et al. 2009, Cannon et al. 2013, Osborn et al. 2013). In addition, there are at least six undescribed species of Torquaratoridae with published molecular sequence data (Cannon et al. 2009, Osborn et al. 2011, Cannon et al. 2013), and many that have been photographed but not collected (Holland et al. 2005, Smith et al. 2005, Andersen et al. 2011, Osborn et al. 2011). Cannon et al. (2013) utilize molecular data to characterize a novel group of cold water harrimaniids that is split into four clades, as well as a distinct Icelandic harrimaniid that does not phylogenetically align with any other genus with available sequence data.

2.6 Diversity

In 2010, Deland et al. presented a revised taxonomy of Harrimaniidae based on historical collections, started over 100 years prior, from Ritter, Bullock and Rao. These collections, comprised of complete specimens, sections, notes, and micrographs, had been passed from Bullock to Burdon-Jones, who deposited the material at the Smithsonian Institution upon his retirement. Bullock contacted Cameron to complete the work, and after Bullock's death in 2005, several papers have now been published posthumously based on this material (Cameron et al. 2010, Deland et al. 2010, Cameron & Perez 2012, Cameron and Ostiguy 2013), substantially updating (and consolidating) enteropneust taxonomy and improving representation of morphological diversity within select acorn worm groups.

At the time of this publication, the 108 described enteropneust species are classified in four families, with no ordinal level classifications. Diversity within the group is underestimated, with some authors estimating species numbers closer to 500-1,000 (Appeltans et al. 2012). The four currently recognized enteropneust families are Ptychoderidae Spengel 1893, Spengelidae Willey 1898, Harrimaniidae Spengel 1901, and Torquaratoridae Holland et al. 2005. In addition to these four families, Planctosphaeroidea is a rarely collected monotypic group known only as large, modified tornaria larvae of *Planctosphaera pelagica*. Whereas Planctosphaeroidea was once considered a separate class of hemichordate (Van der Horst 1936, Hyman 1959), it is now commonly considered to be the larva of an unknown enteropneust (Hadfield and Young 1983). These planktonic organisms can measure up to 25mm. They have been collected in the Atlantic and the Pacific, from depths of 75 to ~1000m, but only about 30 individuals have been collected

since their discovery in 1910 (Hart et al. 1994). Due to their uncertain affinities and taxonomic status, we have not included them in the classifications herein.

In the following section, we briefly describe the four recognized families in the order they were described, and the genera that comprise them, also in order of year of description. This discussion is not intended to be a complete list of synonyms and collection sites. For further details, see Supplementary Table 1 from Tassia et al. (2016). Many significant identifying characteristics of enteropneust species are internal morphology that must be observed via histological sectioning. In the following section, external anatomy has been emphasized, with the intention that this discussion may prove useful for non-specialists to identify acorn worms in the field. References to more thorough internal anatomical descriptions are given under each genus. For a taxonomic key to the families of Enteropneusta see Deland et al. (2010), which also contains a key to the genera of Harrimaniidae. Cameron and Perez (2012) provides a key to the genera of Spengelidae; Cameron and Ostiguy (2013) provides a key to the genera of Ptychoderidae. Information for family and genus descriptions is from these sources unless otherwise indicated.

Ptychoderidae Spengel 1893

Ptychoderidae includes 44 species in three recognized genera, *Ptychodera, Glossobalanus,* and *Balanoglossus*. Ptychoderids have been described as the most derived enteropneusts, due to structures such as synapticles in the gill slits, pronounced regionalization of the trunk, and lateral septa (Hyman 1959). Ptychoderidae is defined by the absence of abdominal pores, and the presence of lateral septa in the trunk. The primary and secondary gill bars are connected via synapticles. The pericardium is simple, and there are dorsal nerve roots in the collar. Hepatic caeca are usually present, and the horns of the proboscis skeleton rarely extend past the anterior half of the collar (Cameron and Ostiguy 2013).

Ptychoderids develop indirectly via tornaria larvae. *Ptychodera flava* has also been the subject of the majority of studies to date on indirect development among enteropneusts (e.g., Tagawa et al. 1998, Henry 2001, Nakajima et al. 2004, Lin et al. 2016). More recently, *Balanoglossus misakiensis* and *Balanoglossus simodensis* have been proposed as models for hemichordate indirect development, in part due to the ease with which these species can be reared

in the lab (Urata and Yamaguchi 2004, Miyamoto and Saito 2007, Ikuta et al. 2009, Miyamoto et al. 2010, Miyamoto and Saito 2010).

In molecular systematics studies, the relationships between the three ptychoderid genera are extremely variable. *Glossobalanus* is non-monophyletic in several studies (Cannon et al. 2013, Osborn et al. 2013), suggesting some sequences may have been misidentified, or this genus requires systematic revision. *Ptychodera flava*, the first hemichordate to be described, is now one of two enteropneust species to have a complete published genome (Simakov et al. 2015).

Ptychodera Eschscholtz 1825

In *Ptychodera* species, the atrium opens via long slits that are exposed in their entirety to the outside. The genital wings are well developed.

Ptychodera flava had the highest reported species distribution in Tassia et al. (2016), although the population and species boundaries within *Ptychodera* are in need of further assessment. Urata (2015) analyzed 18S rDNA and mitochondrial 16S rDNA from *Ptychodera flava* specimens in Japan and found that these sequences were identical to sequences from *Ptychodera flava* collected in Moorea (Cannon et al. 2009), over 10,000 kilometers away, suggesting that this species may indeed have an extensive geographic distribution.

Balanoglossus Delle Chiaje 1829

As in *Ptychodera*, the genital wings in *Balanoglossus* are well developed. The primary distinction between *Ptychodera* and *Balanoglossus* in terms of external anatomy is that *Balanoglossus* has small pores opening to the atrium, rather than fully exposed long gill slits.

Glossobalanus Spengel 1893

In *Glossobalanus*, there are genital ridges, rather than the genital wings found in other ptychoderid genera. The genital ridges usually extend into the hepatic region of the trunk. In most

species, the hepatic sacs are consistent in shape and size, and ordered in two rows. Gill slits are small in *Glossobalanus*.

Spengelidae Willey 1899

Schizocardium, Willeyia, *Glandiceps*, and *Spengelia* comprise Spengelidae (20 species), which is defined by the presence of a digit-like projection, or "vermiform process", at the anterior end of the stomochord. Spengelids additionally have a layer of circular muscle inside the layer of longitudinal muscle in the trunk, and the horns of the proboscis skeleton are long, generally extending through the entirety of the collar. They typically do not have dorsal nerve roots arising from the collar cord. Spengelids are described as having a combination of ptychoderid and harrimaniid features. For example, *Spengelia* and *Schizocardium* possess gill slit synapticles, whereas *Glandiceps* and *Willeyia* do not (Hyman 1959). In species where development is known, spengelids develop via tornaria larvae (e.g., *Glandiceps* – Rao 1953, Urata et al. 2014; S*chizocardium californicum* – Gonzalez et al. 2017). In general, this group contains many rare and poorly-studied species, thus there are few studies on spengelids. Illustrating this, a specimen of *Glandiceps abyssicola*, not seen since its discovery on the Challenger expedition in 1873, was recently rediscovered (Holland et al. 2013). However, there has been progress on developmental biology in *Schizocardium californicum* (Gonzalez et al. 2017) and has gained attention as an emerging model species for indirect developing enteropneusts.

Schizocardium Spengel 1893

The digit-like projection at the anterior end of the stomochord is very long in *Schizocardium*. The gill slits are also very long, extending almost the entire circumference of the pharyngeal trunk, leaving only a narrow hypobranchial strip of digestive pharynx. Synapticles are present, there are no dorsal gonads in the branchial region, and the hepatic sacs are pronounced. Additionally, the pericardium bifurcates anteriorly into long tubes, and these are surrounded by a paired glomerulus.

Glandiceps Spengel 1891

Synapticles, nerve roots, and hepatic sacs are absent. The ventral part of the pharynx is well developed. *Glandiceps* has an unusual swimming behavior, and swarms of swimming acorn worms have been described near Java (Spengel 1909) and in the Seto Inland Sea of Japan (Ikeda 1908, Yoshimatu and Nishikawa 1999). Urata et al. (2012) observed swimming behavior in cultured specimens of *Glandiceps hacksi* and found that the worms could readily leave their burrows and swim by contorting their proboscis into a pear shape, flattening their trunk dorsoventrally, and wiggling their trunks. Worms swam on average 71 seconds, with the longest observed swimming time of 165 seconds (Urata et al. 2012).

Spengelia Willey 1898

Synapticles are present, hepatic sacs are absent, and nerve roots may be present or absent. Pericardial diverticula are short, and there are dermal pits in the genital region.

Willeyia Punnet 1903

The collar in *Willeyia* is longer than it is broad. Synapticles, nerve roots, and hepatic sacs are all absent. The pericardial diverticula are short, and dorsal gonads are absent.

Harrimaniidae Spengel, 1901

Harrimaniidae (40 species) is comprised of members of the genera *Saccoglossus, Harrimania, Stereobalanus, Protoglossus, Mesoglossus, Ritteria, Saxipendium, Horstia, Meioglossus,* and *Xenopleura*. Harrimaniids are defined by the absence of many features; including circular muscles in the trunk, lateral septa, gill slit synapticles, and hepatic caeca (Deland et al. 2010).

Harrimaniids have direct development, with a non-descript larval stage bearing little resemblance to tornaria. *Saccoglossus kowalevskii* has become a well-known study organism for developmental work (Colwin and Colwin 1949, 1953, 1962, Stach and Kaul 2011, Kaul-Strehlow and Stach 2013) and developmental gene expression work (Lowe et al. 2003, 2006, Aronowicz

and Lowe 2006, Lowe 2008, Darras et al. 2011, Green et al. 2013). These studies have yielded important insights by comparing hemichordate and chordate development. The *Saccoglossus kowalevskii* genome is now available (Simakov et al. 2015).

Saccoglossus Schimkewitsch 1892

The genus *Saccoglossus* is most recognizable for the extremely long proboscis found in most of its members. Typically at the proboscis length is at least twice the width, but it may be much longer, and the proboscis has a middorsal longitudinal groove. The collar is generally very short relative to the proboscis. The longitudinal muscle fibers of the proboscis are arranged in concentric rings (**Figure 7A**).

Harrimania Ritter 1900

Harrimania's proboscis is cone-shaped, and is slightly longer than it is broad, the longitudinal muscles of the proboscis are arranged in radial plates (**Figure 7B**), and there are often two proboscis pores. The collar is broader than it is long, and the horns of the proboscis skeleton are very long, even reaching into the trunk in some species (Cameron 2002, Deland et al. 2010). The gonads form rows of simple sacs dorsally and laterally (Deland et al. 2010).

Stereobalanus Spengel 1901

Stereobalanus is an unusual genus with four genital regions (two dorsolateral and two ventrolateral) immediately posterior to the collar. Broad gill openings with externally visible gill tongues are found between these paired gonad regions, opening to a single fused atrium (Reinhard 1942). The longitudinal musculature of the proboscis is arranged in radial plates (**Figure 7B**).

Xenopleura Gilgrist 1925

This genus is described from a single specimen from the southern tip of Africa and is reportedly viviparous. *Xenopleura* is characterized by medullary folds in the dorsal trunk. Longitudinal muscles in the proboscis are diffuse (**Figure 7C**), and there is a single proboscis pore.

Protoglossus van der Horst 1935

There is a deep dorsal groove at the posterior end of the short and conical proboscis in *Protoglossus.* Additionally, the pre-oral ciliary organ is prominent and horseshoe-shaped. The proboscis longitudinal musculature is radial (**Figure 7B**), and there is a single proboscis pore on the left side. The horns of the proboscis skeleton extend to the posterior of the collar, which creates ridges on each side of the buccal cavity. The anterior edge of the collar is ruffled (Deland et al. 2010).

Saxipendium Woodwick & Sensenbaugh 1985

Commonly called "the spaghetti worm", *Saxipendium coronatum* is a member of deep sea hydrothermal vent communities. The proboscis skeleton is crown-shaped in cross section, with long horns that curve backwards. The longitudinal muscles in the proboscis are diffuse (**Figure 7C**).

Deland et al. (2010) formally synonymized Saxipendiidae Woodwick and Sensenbaugh (1985) into Harrimaniidae after molecular phylogenetic analyses demonstrated that *Saxipendium coronatum* belonged in this group (Cannon et al. 2009). A second species, *Saxipendium implicatum*, was described in Holland et al. (2012) from seamounts in the deep sea off the central coast of California.

Horstia Deland *et al.* 2010

In *Horstia*, the proboscis is quite short and rounded, with longitudinal musculature arranged in radial plates (**Figure 7B**). The gonads project from the trunk surface as modules. The

gill pores are found on an elevated ridge, and the trunk itself narrows posteriorly from the collar (see **Figure 2A**, Deland et al. 2010).

Mesoglossus Deland *et al.* 2010

Based primarily on Ritter's unpublished manuscript written ca. 1900, Deland et al. (2010) erected the genus *Mesoglossus* to contain those species formerly of the genus *Saccoglossus* where the longitudinal musculature of the proboscis is arranged diffusely (**Figure 7C**), rather than in concentric rings (which is a defining character of *Saccoglossus*) or radial bundles. The proboscis is additionally approximately twice as long as it is wide, without a dorsal groove. Only lateral gonads are present, no dorsal gonads. There is one proboscis pore, generally on the left.

Ritteria Deland *et al.* 2010

Ritteria has a short proboscis, and the longitudinal muscles of the proboscis are dispersed (**Figure 7C**). There is a single well-developed proboscis pore on the left side of the proboscis stalk, which is very reduced. The collar is wider than it is long. There are two pairs of dorsolateral genital ridges, and the branchial pores are found in a groove between these ridges.

Meioglossus Worsaae *et al.* 2012

A genus of meiofaunal harrimaniid acorn worm, *Meioglossus*, has been described from material found in Bermuda and Belize (Worsaae et al. 2012). The genus is characterized as microscopic, interstitial, and having a completely ciliated body. The proboscis is elongated and has a pair of proboscis pores at the base. The collar region is more than half the length of the proboscis. The sole described species, *Meioglossus psammophilus*, reaches a maximum body length of 0.6mm, making it the smallest known enteropneust (Worsaae et al. 2012).

Torquaratoridae Holland et al. 2005

Torquaratoridae was described in 2005 based on morphological analysis of a few specimens collected in the deep northeastern Pacific (Holland et al. 2005). This family now

contains seven species in five genera, *Torquarator, Tergivelum, Allapasus, Yoda*, and *Coleodesmium*. To date, all torquaratorids have been found at 350 meters or greater. Broad-collared acorn worms have been photographed in the deep sea since the 1960's (Bourne and Heezen 1965, Ewing and Davis 1967), but due to the extreme fragility of these animals, collecting intact specimens for species description has not been possible until recent advances in deep sea remote-operated vehicles (ROVs) (Osborn et al. 2011). Molecular phylogenetic results have indicated that Torquaratoridae is sister to Ptychoderidae (Cannon et al. 2009, Holland et al. 2009, Osborn et al. 2011, Worsaae et al. 2012). Torquaratorid life history is still poorly understood, although some species demonstrate unusual characteristics relative to their shallower cohorts. For example, videos taken by ROV have shown *Allapasus aurantiacus* both burrowing and drifting above the benthos, indicating a possible benthopelagic life history (Holland et al. 2012).

Torquaratorids are characterized by a broad proboscis and collar, and the proboscis skeleton is either absent or greatly reduced to a small plate. Synapticles are absent, and there are prominent hepatic caeca. The adult stomochord is separated from the buccal cavity of the collar or is otherwise absent (Holland et al. 2005, Osborn et al. 2011).

Torquarator Holland et al. 2005

As the first described genus of Torquaratoridae, many of the characters used in the diagnosis of *Torquarator bullocki*, such as the relative breath of the collar, the absence of synapticles, etc., have subsequently been shown to be diagnostic of the whole family. Uniquely, *Torquarator bullocki* is described as having a proboscis skeleton with very short anterior and posterior horns, which makes this the sole example of Torquaratoridae where the proboscis skeleton is at all elaborated beyond a simple plate. Unlike other described species in this family, there is no molecular sequence data for *Torquarator bullocki*.

Tergivelum Holland et al. 2009

There are currently two described species of *Tergivelum*, *T. baldwinae* and *T. cinnabarinum.* The primary external diagnostic character of *Tergivelum* is paired "black veils" that

run along 30-50% of the length of the trunk immediately following the collar. These structures were lost in collected specimens of *T. cinnabarinum* but were evident in images taken prior to collection. Internally, this genus is characterized by prominent buccal muscles on the left and right side of the mouth, which are strikingly robust in contrast to the remaining musculature. The two species are morphologically very similar, and are distinguished by color, with *T. baldwinae* being dark brown and *T. cinnabarinum* named for its red, cinnabar coloration.

Allapasus Holland *et al.* 2012

The proboscis complex in this genus is located in the proboscis stalk, more posteriorly than in other enteropneust families, and contains a plate-like proboscis skeleton without skeletal horns and no pericardial sac. In the female holotype specimen of *A. aurantiacus*, numerous ovaries are located in lateral wings that run the length of the trunk. Each ovary holds a single primary oocyte and is contained in a pouch of epidermal tissue that is connected to the lateral wing tissue by a thin stalk. *A. aurantiacus* and *A. isidis* are very similar, with primary differences in the descriptions relating to the fact that the holotype of *A. aurantiacus* is female, while the *A. isidis* holotype is male. The two species differ in color, however, with *A. aurantiacus* being orange and *A. isidis* yellow.

There are no observations of either of the two species of *Allapasus* leaving a fecal trail. Behavioral observations are limited to *Allapasus aurantiacus*, which has been observed partially burrowed in sea floor sediments. This burrowing may be facilitatied by the musculature in the proboscis and collar, which are both more robust than in other torquaratorids. Prior to collection, the *A. aurantiacus* holotype was recorded drifting a few centimeters above the sediment, apparently after voiding its gut contents, with no evidence of muscular undulations.

Yoda Priede et al. 2012

The name *Yoda* comes from the shape of elongated lateral lips that account for two thirds of the width of the collar, and taper to points, resembling the ears of the Star Wars character, Yoda. The sole described species of this genus, *Yoda purpurata*, is dark reddish purple in color.

Specimens have been observed fully exposed on the seafloor surface, creating an irregularly meandering fecal trail.

Coleodesmium Osborn et al. 2013

The sole described species of *Coleodesmium*, *C. karaensis* has a unique tubular sheath in the proboscis skeleton surrounding the collar nerve cord. In contrast to *Allapasus*, which is otherwise similar in general outward shape, the musculature in the proboscis and collar is weakly developed. The single collected specimen is a brooding female with embryos at differing stages of early development located on the surface of the pharyngeal region. This species was collected in the Russian Arctic at a depth of about 350 meters, marking the shallowest reported collection of a torquaratorid enteropneust.

2.7 References

Almeda R., Bona S., Foster C.R., & Buskey E.J. (2014): Dispersant Corexit 9500A and chemically dispersed crude oil decreases the growth rates of meroplanktonic barnacle nauplii (*Amphibalanus improvisus*) and tornaria larvae (*Schizocardium sp.*). *Marine Environmental Research*, 99: 212–217.

Anderson T.J., Przeslawski R., & Tran M. (2011): Distribution, abundance and trail characteristics of acorn worms at Australian continental margins. *Deep-Sea Research Part II-Topical Studies in Oceanography*, 58(7-8): 970–978.

Annona G., Holland N.D., & D'Aniello S. (2015): Evolution of the notochord. *EvoDevo*, 6(1): 30.

Appeltans W., Ahyong S.T., Anderson, G., et al. (2012): The magnitude of global marine species diversity. *Current Biology*, 22(23): 2189–2202.

Aronowicz, J., & Lowe, C.J. (2006): Hox gene expression in the hemichordate *Saccoglossus kowalevskii* and the evolution of deuterostome nervous systems. *Integrative and Comparative Biology*, 46(6): 890–901.

Balser E.J. & Ruppert E.E. (1990): Structure, ultrastructure, and function of the preoral heart-

kidney in *Saccoglossus kowalevskii* (Hemichordata, Etneropneusta) including new data on the stomochord. *Acta Zoologica*, 71(4): 235–249.

Barrington E.J.W. (1940): Observations on feeding and digestion in *Glossobalanus minutus*. *Quarterly Journal of Microscopical Science*, 82: 227–260.

Barrington E.J.W. (1965): *The biology of Hemichordata and Protochordata*, San Francisco: W.H. Freeman.

Bateson W. (1885): The later stages in the development of *Balanoglossus kowalevskyi*, with a suggestion as to the affinities of the Enteropneusta. *Quarterly Journal of Microscopical Science*, 25: 81–122.

Bateson W. (1886): The early stages in the development of *Balanoglossus*. *Studies from the Morphological laboratory in the University of Cambridge*, 2: 131–160.

Benito J. & Pardos F. (1997): Hemichordata. In: Harrison F.W. & Ruppert E.E. (eds.) *Microscopic Anatomy of Invertebrates*. Wiley-Liss, New York: 15–101.

Bourlat S., Nielsen C., Lockyer A., Littlewood D., & Telford M. (2003): *Xenoturbella* is a deuterostome that eats molluscs. *Nature*, 424(6951): 925–928.

Bourlat S.J., Juliusdottir T., Lowe C.J., et al. (2006): Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature*, 444(7115): 85–88.

Bourlat S.J., Nielsen C., Economou A.D., & Telford M.J. (2008): Testing the new animal phylogeny: A phylum level molecular analysis of the animal kingdom. *Molecular Phylogenetics and Evolution*, 49(1): 23–31.

Bourne D. & Heezen B. (1965): A wandering enteropneust from abyssal Pacific and distribution of spiral tracks on the sea floor. *Science,* 150(3692): 60–63.

Brambell F.W.R. & Cole H.A. (1939): The preoral ciliary organ of the Enteropneusta: Its occurence, structure, and possible phylogenetic significance. *Journal of Zoology*, B109(2): 181–193.

Brambell F.W.R. & Goodhart C.B. (1941): *Saccoglossus horsti*, sp. n., an enteropneust occurring in the Solent. *Journal of the Marine Biology Association*, xxv: 283–301.

Brandenburger J.L., Woolacott R.M., & Eakin R.M. (1973): Fine structure of eyespots in tornarian larvae (Phylum: Hemichordata). *Zeitschrift für Zellforschung und Mikroskopische Anatomie*, 142(1): 89–102.

Bridges T.S. & Woodwick K.H. (1994): Comparative morphology and function of hepatic caeca in four enteropneusts. *Acta Zoologica*, 75(4): 371–378.

Bullock T.H. (1940): The functional organization of the nervous system of Enteropneusta. *Biological Bulletin*, 79: 91–113.

Bullock T.H. (1945): The anatomical organization of the nervous system of Enteropneusta. *The Quarterly Journal of Microscopical Science*, 86: 55–111.

Burdon-Jones C. (1951): Observations on the spawning behaviour of *Saccoglossus horsti* Brambell & Goodhart, and of other Enteropneusta. *Journal of the Marine Biological Association of the United Kingdom*, 29(3): 625–638.

Burdon-Jones C. (1952): Development and biology of the larva of *Saccoglossus horsti* (Enteropneusta). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 236(639): 553–590.

Burdon-Jones C. (1956): Observations on the enteropneust, *Protoglossus koehleri* (Caullery & Mesnil). *Proceedings of the Zoological Society of London*, 127: 35–59.

Cameron C.B. & Mackie G.O. (1996): Conduction pathways in the nervous system of *Saccoglossus* sp. (Enteropneusta). *Canadian Journal of Zoology*, 74: 15–19.

Cameron C.B., Garey J., & Swalla, B.J. (2000): Evolution of the chordate body plan: New insights from phylogenetic analyses of deuterostome phyla. *Proceedings of the National Academy of Sciences of the United States of America*, 97(9): 4469–4474.

Cameron C.B. (2002a): Particle retention and flow in the pharynx of the enteropneust worm *Harrimania planktophilus*: The filter-feeding pharynx may have evolved before the chordates. *Biological Bulletin*, 202(2): 192–200.

Cameron C.B. (2002b): The anatomy, life habits, and later development of a new species of enteropneust, *Harrimania planktophilus* (Hemichordata : Harrimaniidae) from Barkley

Sound. *Biological Bulletin*, 202(2): 182–191.

Cameron C.B. (2005): A phylogeny of the hemichordates based on morphological characters. *Canadian Journal of Zoology*, 83(1): 196–215.

Cameron C.B., Deland C., & Bullock T. (2010): A revision of the genus *Saccoglossus* (Hemichordata: Enteropneusta: Harrimaniidae) with taxonomic descriptions of five new species from the Eastern Pacific. *Zootaxa*, 2483: 1–22.

Cameron C.B. & Perez, M. (2012): Spengelidae (Hemichordata: Enteropneusta) from the Eastern Pacific including a new species, *Schizocardium californicum*, from California. *Zootaxa*, 3569: 79–88.

Cameron C.B. & Ostiguy A. (2013): Three new species of *Glossobalanus* (Hemichordata: Enteropneusta: Ptychoderidae) from western North America. *Zootaxa*, 3630(1): 143–154.

Cannon J.T., Rychel A.L., Eccleston H., Halanych K.M., & Swalla B.J. (2009): Molecular phylogeny of Hemichordata, with updated status of deep-sea enteropneusts. *Molecular Phylogenetics and Evolution*, 52(1): 17–24.

Cannon J.T., Swalla B.J., & Halanych K.M. (2013): Hemichordate molecular phylogeny reveals a novel cold-water clade of harrimaniid acorn worms. *The Biological Bulletin,* 225(3): 194–204.

Cannon J.T., Kocot K.M., Waits D.S., Weese D.A., Swalla B.J., Santos S.R., & Halanych K.M. (2014): Phylogenomic resolution of the hemichordate and echinoderm clade. *Current Biology*, 24(23): 2827–2832.

Caron J.B.B., Morris S.C., & Cameron C.B. (2013): Tubicolous enteropneusts from the Cambrian period. *Nature*, 495(7442): 503–506.

Caullery M. & Mesnil F. (1900): Sur une nouvelle espece del *Balanoglossus* (*B. koehleri*) habitant les cotes de la Manche. *Comptes rendus des seances de la Societe de biologie et de ses filiales*, LII: 256–259.

Cedhagen T. & Hansson H.G. (2013): Biology and distribution of hemichordates (Enteropneusta) with emphasis on Harrimaniidae and description *Protoglossus bocki* sp. nov. from Scandinavia. *Helgoland Marine Research*, 67(2): 251–265.

Cole A.G. & Hall B.K. (2004): The nature and significance of invertebrate cartilages revisited: Distribution and histology of cartilage and cartilage-like tissues within the Metazoa. *Zoology*, 107(4): 261–273.

Colwin L.H. & Colwin A.L. (1949): The fertilization reaction in the egg of *Saccolgossus* (*Dolichoglossus*) *kowalevskii. Biological Bulletin,* 97(2): 237–237.

Colwin A.L. & Colwin L.H. (1953): The normal embryology of *Saccoglossus kowalevskii* (Enteropneusta). *Journal of Morphology*, 92(3).

Colwin L.H. & Colwin A.L. (1962): Induction of spawning in *Saccoglossus kowalevskii* (Enteropneusta) at Woods Hole. *Biological Bulletin*, 123(2): 461–520.

D'Aniello S., Delroisse J., Valero-Gracia A., et al. (2015): Opsin evolution in the Ambulacraria. *Marine Genomics*, 24(2): 177-183.

Darras S., Gerhart J., Terasaki M., Kirschner M., & Lowe C.J. (2011): Beta-Catenin specifies the endomesoderm and defines the posterior organizer of the hemichordate *Saccoglossus kowalevskii*. *Development*, 138(5): 959–970.

Davis B.M. (1908): The early life-history of *Dolichoglossus pusillus* Ritter. *University of California Publications in Zoology*, 4: 187–226.

Dawydoff C. (1948): Classe des Enteropneustes. In: Grasse P.P. (ed.) *Traite de Zoologie, Vol. XI*. Masson et Cie, Paris: 369–453.

Deland C., Cameron C.B., Rao K.P., Ritter W.E., & Bullock T.H. (2010): A taxonomic revision of the family Harrimaniidae (Hemichordata: Enteropneusta) with descriptions of seven species from the Eastern Pacific. *Zootaxa*, 2408: 1–30.

Delle Chiaje S. (1829): Memorie sulla storia e notomia degli animali senza vertebre del Regno di Neapel. *Napoli,* 4: 1–72.

Dilly P.N. (1969): The nerve fibres in the basement membrane and related structures in

Saccoglossus horsti (Enteropneusta). *Zeitschrift fur Zellforschung und Mikroskopische Anatomie*, 97(1): 69–83.

Dilly P.N., Welsch U., & Storch V. (1970): The structure of the nerve fibre layer and neurocord in enteropneusts. *Zeitschrift für Zellforschung und Mikroskopische Anatomie*, 103(1): 129–148.

Dunn C.W., Hejnol A., Matus D.Q., et al. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. Nature, 452(7188): 745–749.

Eschscholtz F. (1825): *Berichtüber die zoologische Ausbeute während der Reise von Kronstadt bis St. Peter-und Paul.* Oken's Isis.

Ewing M. & Davis R.A. (1967): Lebensspuren photographed on the ocean floor. *The John Hopkins Oceanographic Studies*, 3.

Fielman K., & Targettt N. (1995): Variation of the 2,3,4-Tribromopyrrole and its sodium sulfamate salt in the hemichordate *Saccoglossus kowalevskii*. Marine Ecology-Progress Series, 116(1-3): 125-136.

Gillis J.A., Fritzenwanker J.H. & Lowe C.J. (2012): A stem-deuterostome origin of the vertebrate pharyngeal transcriptional network. *Proceedings of the Royal Society of London B: Biological Sciences*, 279(1727): 237–246.

Gonzalez P. & Cameron C.B. (2009): The gill slits and pre-oral ciliary organ of *Protoglossus* (Hemichordata: Enteropneusta) are filter-feeding structures. *Biological Journal of the Linnean Society*, 98(4): 898–906.

Gonzalez P., Uhlinger K.R., & Lowe C.J. (2017): The adult body plan of indirect developing hemichordates develops by adding a Hox-patterned trunk to an anterior larval territory. *Current Biology*, 27(1): 87–95.

Green S.A., Norris R.P., Terasaki M., & Lowe C.J. (2013): FGF signaling induces mesoderm in the hemichordate *Saccoglossus kowalevskii. Development*, 140(5): 1024–1033.

Hadfield M. & Young R. (1983): *Planctosphaera* (Hemichordata, Enteropneusta) in the Pacific Ocean. *Marine Biology*, 73(2): 151–153.

Hadfield M.G. (2002) Phylum Hemichordata. In: Young C.M., Sewell M.A., & Rice M.E. (eds.)

Atlas of Marine Invertebrate Larvae. Academic Press, San Diego: 553–564.

Halanych K.M. (1995): The phylogenetic position of the pterobranch hemichordates based on 18S rDNA sequence data. *Molecular Phylogenetics and Evolution*, 4(1): 72–76.

Halanych K.M. (2004): The new view of animal phylogeny. *Annual Review of Ecology Evolution and Systematics*, 35: 229–256.

Halanych K.M., Cannon J.T., Mahon A.R., Swalla B.J., & Smith C.S. (2013): Modern Antarctic acorn worms form tubes. *Nature Communications*, 4: 1–4.

Harmer S.F. (1910): Hemichordata. In: Harmer S.F. & Shipley A.E. (eds.) *The Cambridge Natural History*. MacMillan and Co., Limited, London.

Hart M., Miller R., & Madin L. (1994): Form and feeding mechanism of a living *Planctosphaera pelagica* (Phylum Hemichordata). *Marine Biology,* 120(4): 521–533.

Hess W.N. (1936): Reaction to light in *Ptychodera bahamensis*. *Pap. Tortugas Lab.*, 475: 77–86.

Hess W.N. (1938): Reactions to light and the photoreceptors of *Dolichoglossus kowalesvkyi. Journal of Experimental Zoology*, 79(1).

Henry J., Tagawa K., & Martindale M. (2001): Deuterostome evolution: early development in the enteropneust hemichordate, *Ptychodera flava. Evolution & Development*, 3(6): 375–390.

Holland N.D., Clague D.A., Gordon D.P., Gebruk A., Pawson D.L., & Vecchione M. (2005): "Lophenteropneust" hypothesis refuted by collection and photos of new deep-sea hemichordates. *Nature*, 434(7031): 374–376.

Holland N.D., Jones W.J., Ellena J., Ruhl H.A., & Smith Jr. K.L. (2009): A new deep-sea species of epibenthic acorn worm (Hemichordata, Enteropneusta). *Zoosystema*, 31(2): 333–346.

Holland N.D., Kuhnz L.A., & Osborn, K.J. (2012): Morphology of a new deep-sea acorn worm (class Enteropneusta, phylum Hemichordata): A part-time demersal drifter with externalized ovaries. *Journal of Morphology*, 273(7): 661–671.

Holland N.D., Osborn K.J., Gebruk A.V., & Rogacheva A. (2013): Rediscovery and augmented description of the HMS "Challenger" acorn worm (Hemichordata, Enteropneusta),

Glandiceps abyssicola, in the equatorial Atlantic abyss. *Journal of the Marine Biological Association of the United Kingdom*, 93(8): 2197–2205.

Holland P.W.H., Hacker A.M., & Williams N.A. (1991): A molecular analysis of the phylogenetic affinities of *Saccoglossus cambrensis* Brambell & Cole (Hemichordata). *Proceedings of the Royal Society B: Biological Sciences.* 332(1264): 185-189.

Hyman L.H. (1959): Phylum Hemichordata. In: *The Invertebrates*. McGraw-Hill, New York: 72–207.

Ikeda I. (1908): On the swimming habit of a Japanese enteropneust, *Glandiceps hacksii* Marion. *Annotationes Zoologicae Japonenses*, 6: 255–257.

Ikuta T., Miyamoto N., Saito Y., Wada H., Satoh N., & Saiga H. (2009): Ambulacrarian prototypical Hox and ParaHox gene complements of the indirect-developing hemichordate *Balanoglossus simodensis*. *Development Genes and Evolution*, 219(7): 383–389.

Jones D.O.B., Alt C.H.S., Priede I.G., et al. (2013): Deep-sea surface-dwelling enteropneusts from the Mid-Atlantic Ridge: Their ecology, distribution and mode of life. *Deep-Sea Research II*, 98: 374–387.

Kaul-Strehlow S. & Stach T. (2011): The pericardium in the deuterostome *Saccoglossus kowalevskii* (Enteropneusta) develops from the ectoderm via schizocoely. *Zoomorphology*, 130(2): 107–120.

Kaul-Strehlow S. & Stach T. (2013): A detailed description of the development of the hemichordate *Saccoglossus kowalevskii* using SEM. *Frontiers in Zoology*, 1053(1): 1–32.

Kaul-Strehlow S. & Rottinger E. (2015): Hemichordata. In: Wanninger A. (ed.) *Evolutionary Developmental Biology of Invertebrates 6: Deuterostomia*. Springer-Verlag Wien, Vienna: 59–89.

Kaul-Strehlow S., Urata M., Praher D., & Wanniger A. (2017): Neuronal patterning of the tubular collar cord is highly conserved among enteropneusts but dissimilar to the chordate neural tube. *Scientific Reports*, 7(1): 7003.

Kicklighter C., Kubanek J., Barsby T., & Hay M. (2003): Palatability and defense of some tropical infaunal worms: Alkylpyrrole sulfamates as deterrents to fish feeding. *Marine Ecology-Progress Series*, 263: 299–306.

Kicklighter C., Kubanek J., & Hay M. (2004): Do brominated natural products defend marine worms from consumers? Some do, most don't. *Limnology and Oceanography*, 49(2): 430–441.

King G., Giray C., & Kornfield I. (1995): Biogeographical, biochemical and genetic differentiation among North American saccoglossids (Hemichordata; Enteropneusta; Harrimaniidae). *Marine Biology*, 123(2): 369–377.

Knight-Jones E.W. (1952.): On the nervous system of *Saccoglossus cambrensis* (Enteropneusta). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 236(634): 315–354.

Knight-Jones E.W. (1953): Feeding in *Saccoglossus* (Enteropneusta). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 123(3): 637–654.

Kocot K.M., Cannon J.T., & Halanych K.M. (2010): Elucidating animal phylogeny. In: DeSalle R. & Schierwater B. (eds.), *Key Transitions in Animal Evolution*. Science Publishers, Ensfield: 15-33.

Kowalevsky A. (1866): Anatomie des *Balanoglossus*. *Mem. Acad. Imper. Sci. St. Petersbourg*, 10(7).

Lartillot N. & Philippe H. (2008): Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1496): 1463–1472.

Lin C.Y., Tung C.H., Yu J.K., & Su Y.H. (2016): Reproductive periodicity, spawning induction, and larval metamorphosis of the hemichordate acorn worm *Ptychodera flava. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 326(1): 47–60.

Lowe C.J., Wu M., Salic A., et al. (2003): Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. *Cell*, 113(7): 853–865.

Lowe C.J., Tagawa K., Humphreys T., Kirschner M., & Gerhart J. (2004): Hemichordate embryos:

Procurement, culture, and basic methods. *Methods in Cell Biology*, 74: 171-194.

Lowe C.J., Terasaki M., Wu M., et al. (2006): Dorsoventral patterning in hemichordates: Insights into early chordate evolution. *PLoS Biology*, 4(9): 1603–1619.

Lowe C.J. (2008): Molecular genetic insights into deuterostome evolution from the direct-developing hemichordate *Saccoglossus kowalevskii*. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 363(1496): 1569–1578.

Luttrell S., Konikoff C., Byrne A., Bengtsson B., & Swalla B. (2012): Ptychoderid hemichordate neurulation without a notochord. *Integrative and Comparative Biology*, 52(6): 829–834.

Maisey J.G. (1986): Heads and Tails: A chordate phylogeny. *Cladistics*, 2(4): 201–256.

Metschnikoff, V. (1881): Uber die systematishe Stellung von *Balanoglossus*. *Zoologischer Anzeiger,* 4: 139-143.

Millar D. & Ratcliffe N. (1987): The antibacterial activity of the hemichordate *Saccoglossus ruber* (Enteropneusta). *Journal of Invertebrate Pathology*, 50(3): 191–200.

Miyamoto M. & Saito Y. (2007): Morphology and development of a new species of *Balanoglossus* (Hemichordata: Enteropneusta: Ptychoderidae) from Shimoda, Japan. *Zoological Science*, 24(12): 1278–1285.

Miyamoto N., Nakajima Y., Wada H., & Saito Y. (2010): Development of the nervous system in the acorn worm *Balanoglossus simodensis*: Insights into nervous system evolution. *Evolution & Development*, 12(4): 416–424.

Miyamoto N. & Saito Y. (2010): Morphological characterization of the asexual reproduction in the acorn worm *Balanoglossus simodensis*. *Development Growth & Differentiation*, 52(7): 615–627.

Miyamoto N. & Wada H. (2013): Hemichordate neurulation and the origin of the neural tube. *Nature communications*, 4: 2713.

Morgan T.H. (1891): The growth and metamorphosis of tornaria. *Journal of Morphology*, 5: 407–458.

Morgan T.H. (1894): The development of Balanoglossus. *Journal of Morphology*, 9: 1–86.

Muller F. (1898): Observações sobre a fauna marinha. *Rev. Mus. Paulista*, 3: 35.

Nakajima Y., Humphreys T., Kaneko H., & Tagewa K. (2004): Development and neural organization of the tornaria larva of the Hawaiian hemichordate, *Ptychodera flava*. *Zoological Science*, 21(1): 69–78.

Newell G. (1952): The homology of the stomochord of the Enteropneusta. *Journal of Zoology,* 121(4): 741–746.

Nielsen C. & Hay-Schmidt A. (2007): Development of the enteropneust *Ptychodera flava*: Ciliary bands and nervous system. *Journal of Morphology*, 268: 551–570.

Nomaksteinsky M., Röttinger E., Dufour H.D., Chettough Z., Lowe C.J., Martindale M.Q., & Brunet J.F. (2009): Centralization of the deuterostome nervous system predates chordates. *Current Biology*, 19(15): 1264–1269.

Oba Y., Stevani C.V., Oliveira A.G., Tsarkova A.S., Chepurnykh T.V., & Yampolsky I.V. (2017): Selected least studied but not forgotten bioluminescent systems. *Photochemistry and Photobiology*, 93(2): 405–415.

Okuda S. (1939): The Enteropneusta from the Palau Islands. *Journal of the Faculty of Science, Hokkaido University: Zoology*, 7: 17–25.

Osborn K.J., Kuhnz L.A., Priede I.G., Urata M., Gebruk A.V., & Holland N.D. (2011): Diversification of acorn worms (Hemichordata, Enteropneusta) revealed in the deep sea. *Proceedings of the Royal Society B: Biological Sciences*, 279(1733): 1646–1654.

Osborn K.J., Gebruk A.V., Rogacheva A., & Holland N.D. (2013): An externally brooding acorn worm (Hemichordata, Enteropneusta, Torquaratoridae) from the Russian Arctic. *Biological Bulletin*, 225(2): 113–123.

Packard A. (1968): Asexual reproduction in *Balanoglossus* (Stomochordata). *Proceedings of the Royal Society B: Biological Sciences*, 171: 261–272.

Pardos F. & Benito J. (1984): Blood circulatory system in the pharynx of an enteropneust: *Glossobalanus minutus* (Ptychoderidae). *Rivista Di Biologia*, 77: 69–85.

Pardos F. & Benito J. (1990): The main trunk vessels and blood components of *Glossobalanus*

minutus (Enteropneusta). *European Archives of Biology*, 101: 455–468.

Philippe H., Derelle R., Lopez P., et al. (2009): Phylogenomics revives traditional views on deep animal relationships. *Current Biology*, 19(8): 706–712.

Priede I.G., Osborn K.J., Gebruk A.V., Jones D., Shale D., Rogacheva A., Holland N.D. (2012): Observations on torquaratorid acorn worms (Hemichordata, Enteropneusta) from the North Atlantic with descriptions of a new genus and three new species. *Invertebrate Biology*, 131(3): 244–257.

Rao K. (1953): The development of *Glandiceps* (Enteropneusta, Spengelidae). *Journal of Morphology*, 93(1): 1–17.

Reinhard E. (1942): *Stereobalanus canadensis*. *Journal of the Washington Academy of Sciences*, 32.

Rhodes C.P. & Ratcliffe N.A. (1983): Coelomocytes and defence reactions of the primitive chordates, *Branchiostoma lanceolatum* and *Saccoglossus horsti*. *Developmental & Comparative Immunology*, 7(c): 695–698.

Ritter W.E. (1902): The movements of the Enteropneusta and the mechanisms by which they are accomplished. *The Biological Bulletin*, 3: 255–261.

Romer A.S. (1967): Major steps in vertebrate evolution. *Science*, 158(3809): 1629–1637.

Röttinger E. & Lowe C.J. (2012): Evolutionary crossroads in developmental biology: Hemichordates. *Development*, 139: 2463–2475.

Ruppert E.E. (2005): Key characters uniting hemichordates and chordates: Homologies or homoplasies? *Canadian Journal of Zoology*, 83(1): 8–23.

Rychel A.L., Smith S.E., Shimamoto H.T., & Swalla B.J. (2005): Evolution and development of the chordates: Collagen and pharyngeal cartilage. *Molecular Biology and Evolution*, 23(3): 541–549.

Rychel A.L. & Swalla B.J. (2007): Development and evolution of chordate cartilage. *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution*, 308: 325–335.

Saita A., Castellani L.C., & Tripepi S. (1978): The integument of *Glossobalanus minutus*

Kowalevsky (Enteropneusta Ptychoderidae) ultrastructural analysis. *Italian Journal of Zoology*, 12(2–3): 115–179.

Satoh N., Tagawa K., Lowe C.J., et al. (2014): On a possible evolutionary link of the stomochord of hemichordates to pharyngeal organs of chordates. *Genesis*, 52(12): 925–934.

Schaeffer B. (1987): Deuterostome monophyly and phylogeny. *Evolutionary Biology*, 21: 179–235.

Schepotieff A. (1907): Die Pterobranchier. 1. Teil. 1. Abschnitt. Die Anatomie von *Rhabdopleura*. *Zool Jahrb Abt Anat D Ont Tiere Bd*, 23: 463–534.

Schneider K.C. (1902): Lehrbuch der Wergleichenden Histologie der Tiere. *Jena:* 672–692.

Schram F.R. (1991): Cladistic analysis of metazoan phyla and the placement of fossil problematica. In: Simonetta A.M. & Conway-Morris S. (eds.), *The Early Evolution of Metazoa and the Significance of Problematic Taxa*. Cambridge University Press, Cambridge.

Silén L. (1950): On the nervous system of *Glossobalanus marginatus* meek (Enteropneusta). *Acta Zoologica*, 31: 149–175.

Silén, L. (1954): Reflections concerning the "stomochord" of the Enteropneusta. *Journal of Zoology*, 124(1): 63–67.

Simakov O., Kawashima T., Marlétaz F., et al. (2015): Hemichordate genomes and deuterostome origins. *Nature*, 527: 1–19.

Smith K., Holland N., & Ruhl H. (2005): Enteropneust production of spiral fecal trails on the deep-sea floor observed with time-lapse photography. *Deep-Sea Research Part I-Oceanographic Research Papers*, 52(7): 1228–1240.

Spengel J.W. (1891) Über die Gattungen der Enteropneusten. *Verhandlungen der Deutschen Zoologischen Gesellschaft 1*: 47–48

Spengel J.W. (1893): Die Enteropneusten des Golfes von Neapel. *Fauna und Flora des Golfes von Neapel und der Angrenzenden Meeres-Abschnitte*. Friedlander & Sons, Berlin.

Spengel J.W. (1907): Studien über die Enteropneusten der Siboga-Expedition nebst Beobachtungen an verwandten Arten. *Monographie 26, Siboga-Expiditie, uitkomsten op*

zoologisch, botanisch, oceanographisch en geologisch gebied verzameld in Nederlandisch Oost-Indië 1899–1900. Brill, Leiden.

Spengel J.W. (1909): Pelagisches Vorkommen von Enteropneusten. *Zoologischer Anzeiger*, 34: 54–59.

Stach T. & Kaul S. (2011): The postanal tail of the enteropneust *Saccoglossus kowalevksii* is a ciliary creaping organ without distinct similarities to the chordate tail. *Acta Zoologica,* 92(2): 150-160.

Stiasny G. (1914a): Studium uber die Entwicklung des *Balanoglossus clavigerus* Delle Chiaje. I. Die Entwicklung der Tornaria. *Mitt Zool Stat Neapel*, 22: 22–75.

Stiasny G. (1914b): Studium uber die Entwicklung des *Balanoglossus clavigerus* Delle Chiaje. II. Darstellung der weiteren Entwicklung bis zur Metamorphose. *Mitt Zool Stat Neapel*, 22: 255–290.

Swalla B.J. & Smith A.B. (2008): Deciphering deuterostome phylogeny: molecular, morphological and palaeontological perspectives. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1496): 1557–1568.

Tagawa K., Nishino A., Humphreys T., & Satoh N. (1998): The spawning and early development of the Hawaiian acorn worm (Hemichordate), *Ptychodera flava*. *Zoological Science*, 15(1): 85–91.

Tassia M.G., Cannon J.T., Konikoff C.E., Shenkar N., Halanych K.M., & Swalla B.J. (2016): The global diversity of Hemichordata. *Plos One*, 11(10): e0162564.

Thomas I. (1972): Action of the gut in *Saccoglossus otagoensis* (Hemichordata: Enteropneusta). *New Zealand Journal of Marine and Freshwater Research*, 6(4): 560–569.

Tung C.H., Cheng Y.R., Lin C.Y., Ho J.S., Kuo C.H., Yu J.K., & Su Y.H. (2014): A new copepod with transformed body plan and unique phylogenetic position parasitic in the acorn worm *Ptychodera flava. The Biological Bulletin*, 226(1): 69–80.

Urata M. & Yamaguchi M. (2004): The development of the enteropneust hemichordate *Balanoglossus misakiensis* kuwano. *Zoological Science*, 21(5): 533–540.

Urata M., Iwasaki S., & Ohtsuka S. (2012): Biology of the swimming acorn worm *Glandiceps hacksi* from the Seto Inland Sea of Japan. *Zoological Science*, 29(5): 305–310.

Urata M., Iwasaki S., Ohtsuka S., & Yamaguchi M. (2014): Development of the swimming acorn worm *Glandiceps hacksi*: Similarity to holothuroids. *Evolution and Development*, 16(3): 149–154.

Urata M. (2015): Molecular identification of *Ptychodera flava* (Hemichordata: Enteropneusta): Reconsideration in light of nucleotide polymorphism in the 18S ribosomal RNA gene. *Zoological Science*, 32(3): 307–313.

Van der Horst C.J. (1936): Planktosphaera and Tornaria. *Quarterly Journal of Microscopical Sciences*, 78: 605-613.

Van der Horst C.J. (1939): Hemichordata. In: Bronn H.G. (ed). *Klassen und Ordnungen des Tierreichs.* Vol. 4: Abt. 4, Buch 2, Teil 2: 1–739

Wakeman K.C., Reimer J.D., Jenke-Kodama H., & Leander B.S. (2014): Molecular phylogeny and ultrastructure of *Caliculium glossobalani* n. gen. et sp. (Apicomplexa) from a Pacific *Glossobalanus minutus* (Hemichordata) confounds the relationships between marine and terrestrial gregarines. *Journal of Eukaryotic Microbiology*, 61(4): 343–353.

Welsch U. (1984): Hemichordata. In: Bereiter-Hahn J., Matoltsy A.G., & Richards K.S. (eds.) *Biology of the Integument 1: Invertebrates*. Springer-Verlag, Berlin: 790–799.

Wilke U. (1972): Die Feinstruktur des Glomerulus von *Glossobalanus minutus* Kowalewsky (Enteropneusta). *Cytobiologie*, 5: 439–447.

Willey A. (1899a): Enteropneusta from the South Pacific, with notes on the West Indian species. *Willey's Zoological Results*, III: 32–335.

Willey A. (1899b): Remarks on some recent work on the Protochorda, with a condensed account of some fresh observations on the Enteropneusta. *Quarterly Journal of Microscopical Sciences*, 42: 233–244.

Willey A. (1899c): Zoological results based on material from New Britain, New Guinea, Loyalty Islands and elsewhere, collected during the years 1895, 1896 and 1897. *University Press*.

Willey A. (1931): *Glossobalanus berkeleyi*, a new enteropneust from the West Coast. *Transactions of the Royal Society of Canada*, 5: 19–28.

Winchell C., Sullivan J., Cameron C., Swalla B., & Mallatt J. (2002): Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. *Molecular Biology and Evolution*, 19(5): 762–776.

Woodwick K.H. & Sesenbaugh T. (1985): *Saxipendium coronatum*, new genus, new species (Hemichordata: Enteropneusta): the unusual spaghetti worms of the Galápagos Rift hydrothermal vents. *Proceedings of the Biological Society of Washington*, 98: 351–365.

Worsaae K., Sterrer W., Kaul-Strehlow S., Hay-Schmidt A., & Giribet G. (2012): An anatomical description of a miniaturized acorn worm (Hemichordata, Enteropneusta) with asexual reproduction by paratomy. *Plos One*, 7(11): e48529.

Yoshimatu S. & Nishikawa T. (1999): Swimming swarms of a usually benthic enteropneust *Glandiceps sp.* in the Seto Inland Sea, Japan, found in 1998. *Zoological Science*, 16: Supplement: 39.

Chapter 3: Toll-like Receptor Pathway Evolution in Deuterostomes[†]

3.1 Abstract

Animals have evolved an array of pattern-recognition receptor families essential for recognizing conserved molecular motifs characteristic to pathogenic microbes. One such family is the Toll-Like receptors (TLRs). Upon pathogen binding, TLRs initiate specialized cytokine signaling catered to the class of invading pathogen. This signaling is pivotal for activating adaptive immunity in vertebrates, suggesting a close evolutionary relationship between innate and adaptive immune systems. Despite significant advances towards understanding TLR-facilitated immunity in vertebrates, knowledge of TLR-pathway evolution in other deuterostomes is limited. By analyzing genomes and transcriptomes across 37 deuterostome taxa, we shed light on the evolution and diversity of TLR pathway signaling elements. Here, we show that the deuterostome ancestor possessed a molecular toolkit homologous to that which drives canonical MYD88-dependent TLR signaling in contemporary mammalian lineages. We also provide evidence that TLR3-facilitated antiviral signaling predates the origin of its TCAM1-dependence recognized in the vertebrates. SARM1, a negative regulator of TCAM1-dependent pathways in vertebrates, was also found to be present across all major deuterostome lineages despite the apparent absence of TCAM1 in invertebrate deuterostomes. Whether the presence of SARM1 is due to its role in immunity regulation, neuron physiology, or a function of both is unclear. Additionally, Bayesian phylogenetic analyses corroborate several lineage-specific TLR gene expansions in urchins and cephalochordates. Importantly, our results underscore the need to sample across taxonomic groups to understand evolutionary patterns of the innate immunity foundation on which complex immunological novelties arose.

3.2 Significance statement

Innate immunity provides critical defense against pathogen invasion, and mutations in its cellular mechanisms have been implicated in autoimmunity, immune suppression, and other diseases. However, knowledge of innate immunity pathways is largely biased towards model

[†]This chapter has been published as: Tassia MG, Whelan NV, Halanych KM. 2017. Toll-like receptor pathway evolution in deuterostomes. *Proceedings of the National Academy of Sciences* 114, 7055-7060

species. As such, evolutionary interpretations suffer from large taxonomic gaps ultimately weakening the strength of evolutionary inference. Our phylogenetic approach shows the molecular machinery of the canonical TLR pathway was present in the last deuterostome ancestor prior to the rise of chordate lineages. Thus, TLR pathways with multiple gene-gene interactions have been conserved for over 500 million years within vertebrates. Moreover, we provide evidence suggesting TLR3 may represent an ancient, evolutionarily conserved molecular interface for viral immune-stimulation present across Deuterostomia.

3.3 Introduction

Innate immunity provides vital cellular and molecular defense against invading pathogens (1). Unlike immunological memory facilitated by jawed vertebrate immunoglobulins (2) and lamprey variable lymphocyte receptors (3), molecular components of innate immunity do not recombine to diversify the breadth of defensive molecules (4). Thus to provide substantial defense against a diversity of

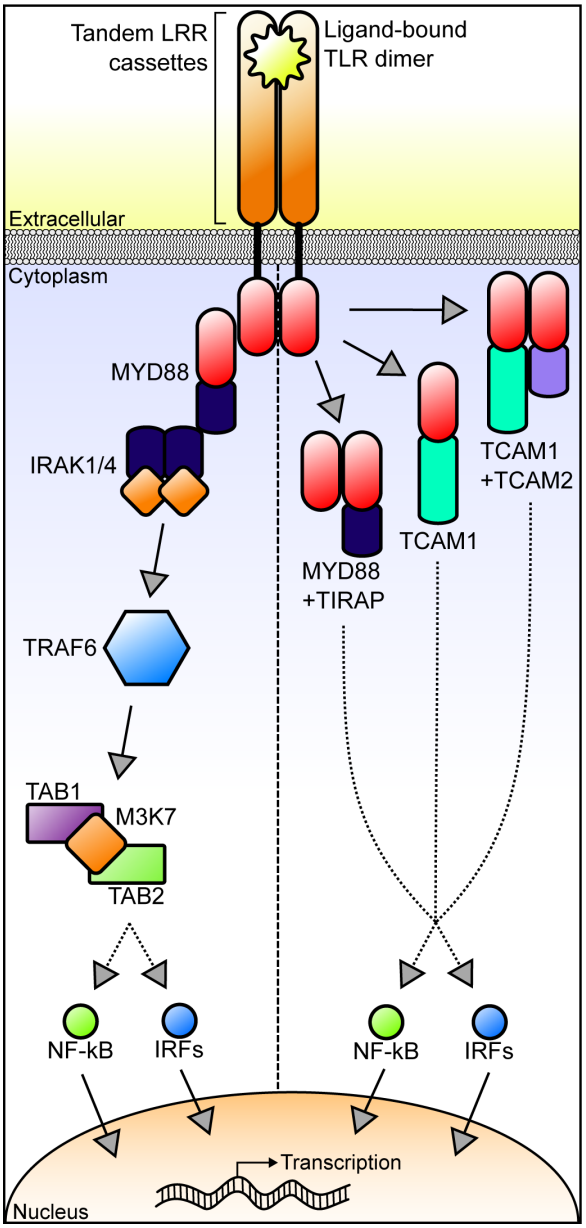


Figure 1: Diagram of major TLR pathways. Upon ligand binding and receptor dimerization, TLRs interact with a TIR-domain-containing adaptor protein. Canonically, TLR signal transduction occurs through the MYD88-dependent pathway (left). In some cases, such as with TLR3 and TLR4, TLRs require other TIR-domain-containing adaptors to successfully signal for cytokine expression (right). SARM1, a TIR-domain-containing negative regulator for TCAM1-dependent signaling pathways, is not shown. Red ellipses denote conserved TIR domains.

infectious agents with limited resources, the innate immune system exploits evolutionarily conserved pathogen-associated molecular patterns (PAMPs) (1, 4). PAMPs, such as Gram-negative lipopolysaccharide or viral double-stranded RNA (dsRNA), often serve fundamental biological roles (4). Such structures are typically conserved over evolutionary time, providing targets for animal pattern-recognition receptors (5). Although there are many well-recognized families of innate immunity pattern-recognition receptors, Toll-like receptors (TLRs) evolved early in animals and have been extensively studied in model systems (6).

TLRs, named after the Toll protein in *Drosophila melanogaster* (7), are a group of type-I transmembrane glycoproteins localized to plasma membranes and endosomes (8). All TLRs possess three major regions: extracellular domain of tandem leucine-rich repeats (LRRs), a transmembrane helix, and a cytoplasmic Toll/interleukin-1 receptor (TIR) domain. The breadth of TLR-facilitated immunity is determined by the ectodomain structure of LRRs and their associated glycosylated superstructure (8). Upon binding to PAMPs, TLRs dimerize, and a signal is transduced cytoplasmically via the TIR domain. Receptor dimers subsequently interact with cytoplasmic TIR-domain containing adaptor proteins (i.e. MYD88, TIRAP, TCAM1, and/or TCAM2) (9). Canonical signaling is mediated by MYD88 (10) (**Fig. 1**). This MYD88-dependent pathway proceeds through IRAK1/4, TRAF6, TAB1/2, and M3K7, terminating in activation of NF-kB for translocation to the nucleus where it acts as a transcription factor for a host of pro-inflammatory cytokines (10). This pathway rapidly provokes an inflammatory response and recruits additional phagocytic cells to confine and neutralize invading pathogens (11)**.** Once thought to possess limited immune potential, research in jawed vertebrates has revealed that several TLRs possess pathogen-specific signaling pathways, and are vital for activating adaptive immunity pathways (12).

TLRs are functionally partitioned into two categories: those localized to host cell membranes and primarily recognize microbial cell membrane components (TLR1, 2, 4, 5, 6, and 10) and those localized to endosomes and recognize nucleic acids (TLR3, 7, 8, and 9) (10). TLR3, a vertebrate ortholog responsible for recognizing viral dsRNA, stimulates downstream signaling exclusively through the TIR-domain containing adaptor TCAM1 (9, 10). Independent of MYD88, TLR3 not only initiates downstream NF-kB activation, but it also initiates type-I interferon signaling that is fundamental to antiviral immunity (10). Concurrent with TLR3 activation and

interaction with TCAM1, another TIR-domain-containing protein, SARM1, increases in concentration and subsequently acts as a negative regulator for the TCAM1-dependent pathway (9). This negative feedback loop is vital for efficient TLR-signaling regulation where overstimulation would be harmful to the host (10). The remaining two TIR-domain containing adaptors, TIRAP and TCAM2, are individually insufficient for TLR signal transduction. Instead, these proteins function as "sorting" proteins – with TIRAP promoting MYD88-dependent pathways and TCAM2 promoting TCAM1-dependent signaling pathways (9). In contrast to early viewpoints which suggested innate immunity acted merely as a molecular bridge to adaptive immunity (13), the presence of pathogen-specialized TLR signaling pathways and their involvement in signaling immune responses indicates innate immunity itself acts as a barrier to microbe pathogenesis.

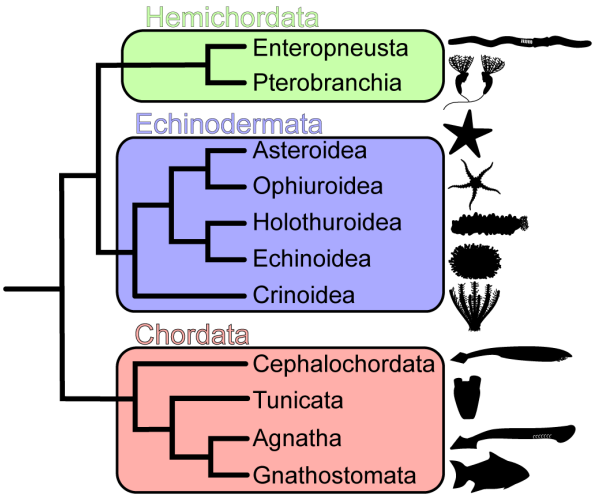


Figure 2. Deuterostome relationships as reported by recent phylogenomic studies (14). Echinoderms and hemichordates form the superphylum Ambulacraria – the sister group to Chordata.

Jawed vertebrates possess approximately 10 TLRs that have been functionally characterized. Far less is known of TLR diversity among other deuterostome groups. In addition to vertebrates, Deuterostomia consists of echinoderms (e.g. sea stars and urchins), hemichordates (acorns worms and pterobranchs), cephalochordates (lancelets), and tunicates (e.g., sea squirts) (**Fig. 2**). Genome surveys have revealed the purple urchin *Strongylocentrotus purpuratus* and the lancelet *Branchiostoma floridae* have expanded repertoires of 253 and 72 TLRs, respectively (15–18). In these species, the majority of TLRs appear to be the result of lineage-specific expansions as opposed to gene loss in jawed vertebrates (17, 18). How these TLR expansions affect breadth of pathogen recognition has yet to be determined. In contrast, the tunicate *Ciona intestinalis* appears to possess only three TLRs, though *C. intestinalis* TLRs have broader PAMP recognition than those known in mammalian systems (19). *Saccoglossus kowalevskii,* an acorn worm hemichordate, has been reported to possess eight TLRs (18).

In addition to TLRs, several immunity-related features appear to be evolutionarily and functionally conserved across deuterostome lineages including: pathogen-responsive phagocytic cell types, regulation of canonical cytokine homologs upon immune challenge, and differential regulation of TLR orthologs upon microbial challenge (**Table 1**). TLRs in the lancelet *Branchiostoma belcheri* have been shown to undergo obligatory MYD88 interactions to activate downstream NF-kB, consistent with observations in mammals (24). In the urchin *S. purpuratus*, gut epithelia have been shown to undergo stereotypical inflammatory responses in the presence of bacterial agents (21). This inflammatory response elicits specialization and migration of phagocytic cell-types to regions of infection – mediated in part by TNFs and IL-17 signaling homologs (21). As a group, these conserved immune mechanisms suggest the ancestor of all

Table 1. Functional conservation of immunity elements among invertebrate deuterostomes

	Chordata		**Ambulacraria**	
	Tunicata	**Cephalochordata**	**Echinodermata**	**Hemichordata**
Phagocytes/ Coelomocytes	Present (22)	Present (20)	Present (21)	Present (20)
Cytokines and/or TFs* expressed in infection	NF-kB, IL-1, TNFα (19, 22)	NF-kB, IL-17, IRFs (24-26)	NF-kB, TNF, IL-17s (21, 23)	Unknown
TLRs				
Subcellular expression	Cell membrane + endosomes (19)	Cell membrane and/or endosomes (24)	Unknown	Unknown
Cell/Tissue expression	Pharynx and gut (19)	Epidermis, pharynx, and gut (24)	Coelomocytes and gut epithelium (18)	Unknown
PAMP-dependent regulation	Present (19)	Present (5)	Present (18)	Unknown
Molecular interactions	Activates NF-kB; induces TNFα expression (19)	Activates NF-kB via MYD88 (24)	Unknown	Unknown

*TFs= Transcription Factors

deuterostome possessed a common innate immunity toolkit with evolutionarily conserved function.

Although origins and ancestral function of TLR signaling among animals are currently unclear, MYD88-facilitated TLR signaling is known to have been present in the bilaterian ancestor (6, 28). In contrast, virus-targeted TCAM1-dependent TLR signaling is only known from studies in select vertebrate taxa (e.g., mouse, human, and zebrafish) (9). As such, available evidence suggests TCAM1-facilitated TLR signaling evolved in the vertebrate lineage at a similar time as the emergence of adaptive immunities (29). This hypothesis has been supported by the reported absence of a TCAM1 homolog among invertebrate model systems (9, 16, 17). However, past comparisons between established vertebrate and invertebrate models provide limited phylogenetic resolution required for an accurate understanding of TLR pathway evolution. In this study, we seek to illuminate the complement of TLR pathway components possessed in early deuterostomes and inform subsequent molecular innovation among contemporary lineages.

3.4 Results and Discussion

3.4.1 TLR signaling adaptors and their associated pathways

Employing bioinformatics tools (**Fig. S1**) to analyze genomic and transcriptomic data from 37 invertebrate deuterostome taxa, including humans which was used as a genomic control (**Table S1**), our findings suggest the deuterostome ancestor possessed homologs to all canonical MYD88-dependent TLR signaling components (**Table 2**). *Th*e presence of downstream TLR signaling elements across all major deuterostome lineages may indicate conserved function of the pathway. The notion of conservation and a shared immunological ancestry among deuterostomes is supported by studies that show functional similarity between invertebrate deuterostomes and vertebrate lineages (**Table 1**). Mentioned previously, TLRs require MYD88 mediation to activate NF-kB in the lancelet (24) and the TLR-mediated and NF-kB activation has also been shown in the tunicate *C. intestinalis* (19). Homologs to several TLR-pathway induced cytokines are expressed in immune challenged invertebrates, and have been implicated in inflammatory and immune responses for sea urchin embryos (21).

Although all other MYD88-dependent signaling mediators were identified by conserved domain architectures, invertebrate deuterostomes appear to lack a typical TAB2 signaling mediator (**Table 2 & Fig. S2**). TAB2, which facilitates coupling TRAF6 to the TAB1/M3K7 complex and

Table 2. Presence of TLR pathway signaling homologs in deuterostomes*.

TIR-domain-containing Adaptors

	MYD88	*TIRAP*	*TCAM1*	*TCAM2*	*SARM1*
Hemi.	+	+		+	+
Echi.	+				+
Ceph.	+	+		+	+
Tuni.	+				+
Vert.	+	+	+	+	+

Signaling Mediators

	IRAK1/4	*TRAF6*	*TAB1*	*TAB2*	*M3K7*
Hemi.	+	+	+	+	+
Echi.	+	+†	+	+	+
Ceph.	+	+	+	+	+
Tuni.	+	+†	+	+	+
Vert.	+	+	+	+	+

*See **Fig. S2** for species-specific homolog numbers.
†Identified only in targeted molecular studies (17,30).

activating NF-kB (**Fig. 1**), was frequently identifiable by primary sequence homology but unable to be corroborated by assessment of typical domain architecture (except in the human control dataset). Specifically, each putative homolog identified lacked the N-terminal CUE domain essential in vertebrates for substantial activation of NF-kB (31). When placed in a phylogenetic framework, including both known TABs and those identified in this study, TAB2 orthology was supported with high confidence (bootstrap = 100%; **Fig. S3**) suggesting TAB2's CUE domain is a vertebrate-specific novelty. As TAB2 homologs without the CUE domain were independently identified in 22 different datasets, the lack of the CUE domain is assumed to be correct and not due to sequence assembly error. Interestingly, a functional TAB2 homolog that contains a CUE domain has been also reported in *Drosophila melanogaster* (32). This finding suggests convergence between vertebrate and *D. melanogaster* TAB2s regarding a CUE domain – perhaps

selected for by molecular kinetics. Under a binary parsimony framework (i.e., presence vs. absence), a single deuterostome loss-and-regain event is equally parsimonious. Provided the CUE domain's potentially ancient eukaryotic origins (independent of TAB2 homology (33)), understanding evolution of this domain will require much deeper taxon sampling within Metazoa.

The second major pathway-defining protein, TCAM1, could not be identified in any genomic or transcriptomic dataset sampled herein aside from humans (**Table 2 & Fig. S2**). In mammals, TCAM1 is necessary for TLR3-facilitated antiviral cytokine signaling via interferons, as well as sufficient for TLR4 signal transduction (9, 10). Despite absence of TLR3's obligatory signaling adaptor, we found strong phylogenetic support (99% posterior probability) for TLR3 orthologs among invertebrate deuterostomes (**Figs 3, S4,** & **S5**). The sole exception was within tunicates, but the absence of TLR3 orthologs in tunicates is unsurprising given their apparent TLR reduction (34). If function of these TLR3 orthologs is conserved across taxa, the implication is that antiviral TLR3 signaling predates the origin of the TCAM1-dependent pathway. Molecular evidence suggests the use of interferons (the primary antiviral cytokine family activated in TLR3-facilitated signaling) and their upstream transcriptional regulators (interferon regulatory factors; IRFs) for antiviral function are a vertebrate innovation (35, 36). Strikingly, IRFs from the lancelet *Branchiostoma belcheri* have been shown to effectively recognize promoter regions of several human interferons (26), and are tightly regulated in response to dsRNA infection (24). These observations are consistent with the hypothesis that antiviral TLR3 signaling predates the emergence of TCAM1. Further molecular investigation will be required to determine the binding specificity of invertebrate TLR3 orthologs and their downstream signaling components. Such studies will be invaluable for understanding the ancestry and functional evolution of antiviral TLR signaling as well as identifying subsequent vertebrate-specific molecular innovations.

The third TIR-domain-containing adaptor, SARM1, acts in a negative feedback-loop for TCAM1-dependent pathways (e.g. TLR3 and TLR4), providing robust regulation where overexpression may yield deleterious effects. We found SARM1 to be present among all deuterostome phyla, despite the apparent absence of TCAM1 (**Table 2 & Fig. S2**). Thus, SARM1's function in TLR signaling regulation may have existed prior to the origin of the TCAM1-dependent pathway in mammals. This conclusion is supported by research on the lancelet *B. belcheri* where SARM1 was found to play an inhibitory role for MYD88-dependent signaling

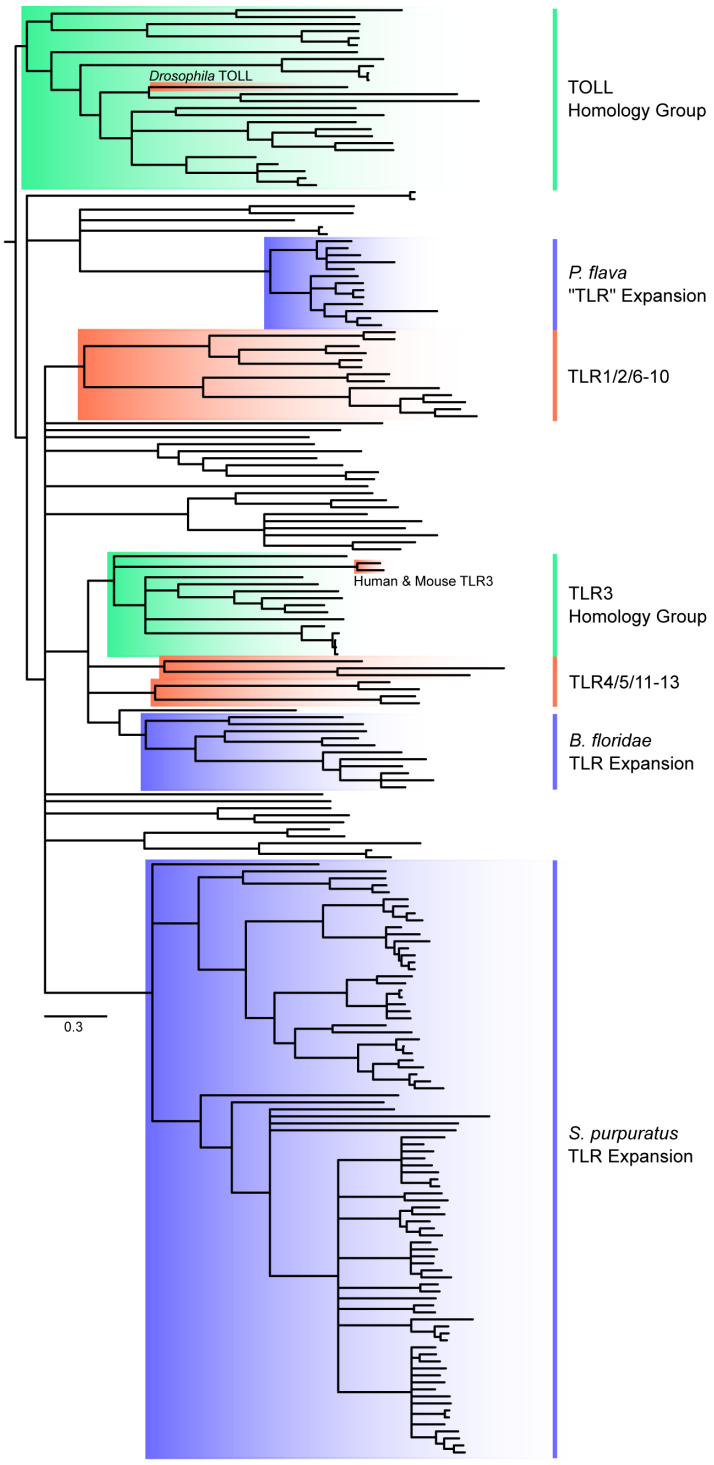


Figure 3. Deuterostome TLR tree generated using full TLR protein sequences in ExaBayes. Human and mouse TLRs, as well as *Drosophila* Toll, are included as positive controls and have been highlighted in red for orientation to known orthology groups. Tips have been removed for ease of interpretation; see **Figs S4 & S5** for more detail**.** All nodes have ≥95% posterior probability.

rather than TCAM1-dependent signaling (37). SARM1 has also been shown to be central for a variety of neuronal processes (e.g. maintenance, behavior, and development (38)) and is a key player in injury-induced axon apoptosis in vertebrates (39). Additionally, SARM1 is implicated in embryological neuron development prior to its function in TLR signaling regulation during *B. belcheri* ontogenesis (37). Thus, our detection of SARM1 homologs among invertebrate deuterostomes may be independent of TLR signaling, but rather a function of is role in neuron physiology/injury. Whether the original function of SARM1 was neuronal physiology/apoptosis, immunity regulation, or a coordination of these functions is still unclear.

The last two TIR-domain-containing adaptors TCAM2 and TIRAP are individually insufficient for TLR signal transduction. Rather, TCAM2 and TIRAP facilitate recruitment of TCAM1 and MYD88, respectively, providing downstream signaling specialization contingent upon the PAMP-bound TLR ortholog. Although our bioinformatics pipeline identified putative TCAM1 and TIRAP homologs among hemichordates and cephalochordates, orthology could not be supported when placed in a phylogenetic framework. These proteins may interact directly with TLRs and/or other TIR-domain-containing adaptors; however, functional characterization will be required to elucidate their molecular signaling roles.

3.4.2 Lineage specific TLR expansions

Evidence for TLR expansions, as reported in echinoderms and lancelets (17, 18), could not conclusively be detected in hemichordates **(Figs. 3, S4,** & **S5)**. A possible exception appears in the acorn worm *Ptychodera flava*, whose genome possesses 27 TLR homologs. However, a large group of *P. flava*'s TLRs (n=13) possess multiple cysteine-rich LRR clusters (atypical even when compared to *Drosophila*-like multi-cysteine cluster TLRs) and large regions that do not match any characterized protein domain (40, 41). When placed in a phylogenetic framework (**Fig. 3, S4,** & **S5**), these TLRs form a monophyletic clade (100% posterior probability) closely related to homologs of more typical TLR structure. Considering the structural divergence from all other TLRs sampled in this study, these TLRs may possess functional divergence specific to *P. flava*. In contrast, *Saccoglossus kowalevskii*, for which a genome is also available (**Table S1**), possesses 13 TLRs.

Using genomic gene-model datasets, coupled with transcriptomic evidence, we were only able to identify 104 TLRs in the urchin *S. purpuratus* and 19 in the lancelet *B. floridae* which could be confidently corroborated by domain architecture. Our estimates of TLR diversity in these two species are fewer than previously reported, possibly suggesting overly-stringent homolog detection standards (see below). However, our methods identified all TLRs, signaling mediators, and gene variants present in the human gene-model dataset, a control for our bioinformatics pipeline. In an attempt to identify sources of error, we employed the bioinformatics pipeline used in Buckley & Rast, 2012 (18) on the most recent versions of the *S. purpuratus* and *B. floridae* genomes available at the time of this study (**Table S1**), and we were still unable to replicate previously reported numbers of TLRs (253 and 72, respectively). Even when modifying detection methods in such a way that would likely result in high false-discovery rates, we detected at most 185 and 39 TLRs in *S. purpuratus* and *B. floridae*, respectively. A large clade of *S. purpuratus* TLRs contained consistent overlapping domain signals from the cytoplasmic TIR domain and what appear to be cytoplasmic LRRs (***Figure S5***). Although this may merely be a domain-prediction artifact, the *S. purpuratus* genome was the only dataset to consistently show this predicted structure. The function of expanded TLR repertoires in immune function or coopted for other roles remains unclear without further functional investigation.

With regard to *S. purpuratus*, the discrepancy between our results and previous reports (16, 18) can be attributed to two differences in the approach of bioinformatics workflows. First, our bioinformatics pipeline identifies unique peptide sequences that cannot be locally clustered by 100% identity, thus removing any translational redundancies and/or fragments of longer contigs (see **TLR Pathway Homolog Identification** below). This greatly reduces chances of including variants of the same TLR polypeptide that only differs by an N-terminal and/or C-terminal extension and thus cannot be confidently concluded to be unique TLRs. Second, extrinsic pseudogene prediction (16, 18) was not included in our analysis. If one combines the number of TLR pseudogenes identified in Buckley and Rast, 2012 (18) with the number of TLRs identified through our overestimation pipeline, results are consistent with previous estimates. Notably, however, we were only able to reach this result when deliberately overestimating the number of TLRs encoded in the *S. purpuratus* genome. We suspect that similar reasons led to discrepancies between our results and those previously reported for *B. belcheri* (18).

3.5 Conclusions

The last common deuterostome ancestor inherited a molecular toolkit sufficient for MYD88-dependent TLR signaling. However, unlike pathways in select model vertebrate taxa and *Drosophila*, conditions for complete signal transduction in the deuterostome ancestor likely lacked a typical TAB2 signaling mediator. TLR3 orthologs were recovered with strong phylogenetic support across invertebrate deuterostomes despite the apparent absence of its obligate signaling adaptor, TCAM1, which is currently known only from vertebrates. Given our findings, coupled with functional evidence for evolutionarily, and functionally conserved antiviral signaling mechanisms (26), we hypothesize that TLR3-facilitated antiviral cytokine signaling predates the origin of the TCAM1-dependent pathway. Considering the ubiquity of marine viruses (42) and their obligate virus-host interactions, extant deuterostome lineages may have inherited a TLR3-mediated antiviral defense from their most recent common ancestor, which was a marine organism. TLR3-mediated antiviral defense would then have been subsequently modified in vertebrates.

3.6 Materials and Methods

3.6.1 Data Acquisition and Assembly

Datasets used for analyses, identified homologs, and their respective accession numbers are available in **Table S1**. cDNA downloaded as raw Illumina RNA-seq reads were digitally normalized via khmer (43) and assembled with Trinity using default parameters (44). Data generated on 454 sequencing platforms were assembled via 454's NEWBLER. With regard to genomic data, only predicted and confirmed gene models were used for homolog identification. A total of 37 taxa were analyzed. Notably, transcriptomic datasets are only evaluated for presence of homologs and no conclusions were made on the absence of any particular homolog. Similarly, several transcriptomic datasets (such as the pterobranch hemichordates *Cephalodiscus nigrescens* and *Rhabdopleura normani*) had relatively low numbers of unique contigs and yielded no detectable homologs. Taxa included in this study were selected to represent phylogenetic depth and distribution representative of all major deuterostome groups while maintaining biocomputational feasibility by limiting vertebrate sampling.

3.6.2 TLR Pathway Homolog Identification

Transdecoder (version 2.0.1) (45) was used on nucleotide sequences to identify putative open-reading frames (ORFs) and their associated protein sequences. Following TransDecoder's log-likelihood scoring metric, high-scoring, small ORFs encapsulated by larger ORFs of the same reading frame were consolidated to avoid redundancy. Extracted amino-acid sequences were additionally clustered by 100% identity using CD-HIT (version 4.6.1) (46) to control for translation redundancies and fragments. Following translation, amino-acid sequences were queried against the SwissProt database (downloaded May 2015) (47) using blastp (version 2.2.29+) (48) with an e-value cutoff of 1E-4. Sequences with best-hits to target proteins were annotated for protein domain architecture using the SMART (40) and Pfam (41) databases included in InterProScan **(**version 5.17-56.0) (49). Domain architectures were cross-referenced with the putative homology for each sequence to control for non-target or fragmented proteins. Absence of stop-codons was not used as a criterion for removal as putative TLR genes because of the potential for extensive data-loss; additionally, this type of filtering merely controls for C-terminal fragments. TMHMM (version 2.0c) (50) was used to predict TLR transmembrane helices; however, absence of a transmembrane helix was not used as a criterion for TLR structural homology as TLRs with known transmembrane helices (i.e., human and mouse TLR9) lacked confident transmembrane predictions with this software. The complete homolog identification bioinformatic pipeline can be found at *https://github.com/mtassia/Homolog_identification* and is diagramed in **Fig. S1**.

Leucine-rich repeats were additionally annotated with LRRfinder (51) to filter for particularly short fragments (See **Table S2** for TLR domain architecture criteria) or TLR products of erroneous gene-prediction. As LRRfinder's prediction confidence statistics were unreliable and appeared to prioritize over prediction rather than under prediction given its performance on human TLRs, we did not remove TLRs with predicted LRR coordinates lying within cytoplasmic regions (as was the case for a group of *S. purpuratus* TLRs). Novel hemichordate TLRs identified from *Saccoglossus kowalevskii* and *Ptychodera flava* genomes were mapped back to their respective genomes using BWA-MEM (version 0.7.12) (52) to confirm gene-models and check for overlap.

3.6.3 Phylogenetic Analyses

To evaluate orthology of recovered sequences, human and mouse TLR pathway protein sequences (in addition to *Drosophila* TOLL) were obtained from SwissProt database and used for phylogenetic inference. We used the E-INS-I alignment algorithm in MAFFT (version 7.215) (53), which optimizes alignments for sequences of multiple domains separated by hypervariable regions, to align complete TLR amino acid sequences. ProtTest (version 3.4) (54) was used to test and select best-fit amino acid substitution rates employing the Bayesian Inference Criterion (BIC). A gene tree of TLRs was inferred with ExaBayes (version 1.4.2) (55). ExaBayes was executed with two parallel runs of four Metropolis-coupled chains analyzed for $1x10^6$ generations (sampled every 500 generations) using a gamma-distributed rate-heterogeneity, empirical amino acid state frequencies, and a fixed amino acid substitution model (WAG). All chains appeared to converge and convergence statistics are available in **Table S3**. Partitioning TLRs by cytoplasmic, transmembrane, and extracellular regions was tested using PartitionFinder (version 1.1.1) (56); however, partitioned analyses did not improve phylogenetic resolution. A majority rule consensus tree was generated after discarding the first 25% sampled MCMC generations (250,000) as burn-in and visualized with Mesquite (version 3.10) (57). All nodes with posterior probabilities less than 95% were collapsed. Two additional TLR phylogenetic analyses were inferred using only TIR domains to compare with previous results of TLR orthology inference (58, 59). Methods, results, and topologies for TIR-domain-only trees are available via "**Constrained phylogenetic inference of TLRs***"* sections of the supplement and **Fig. S6**, respectively. Aligning matrices of full TLR contigs provided greater nodal support and increased resolution relative to the TIR-domain only tree (**Figs. 3 & S6**). Non-TLR gene trees were inferred using RAxML's (version 8.0.23) (60) rapid bootstrap analysis with subsequent best-fit tree inference. Amino acid substitution matrices for non-TLR gene trees were also inferred using ProtTest and phylogenetic analyses were run with gamma-distributed rate-heterogeneity and empirical amino acid state frequencies. RAxML was preferred over ExaBayes for small gene trees due to its accessibility when working with small datasets and short computation time.

3.6.4 TLR Overestimation Pipeline

Genome assemblies of *Branchiostoma floridae* and *Strongylocentrotus purpuratus* (**Table S1**) were analyzed in an attempt to detect TLR diversifications similar to previous findings (17, 18). Whole genome scaffolds were translated in all open reading frames using EMBOSS' *getorf* (61) command with a minimum peptide length of 75 amino acids. Sequences were annotated for Pfam domains using HMMER with default detection thresholds. Polypeptides possessing TIR and LRRs were considered 'complete' and subsequently clustered by 100% identity by cd-hit to remove any sequences that may have been fragments of larger sequences in datasets. Domain architecture was not used to validate TLR homology in this reanalysis to provide highest possible 'TLR' counts. Although this pipeline was assembled in an attempt to maximize overestimation of TLRs in a given genome, we were still unable to identify the magnitude of TLR diversity identified in previous studies (17, 18).

3.7 Acknowledgements

The Auburn University Cellular and Molecular Biosciences graduate research apprenticeship provided funding to M.G.T. and the Alabama Supercomputer Authority. The findings and conclusions in the article are those of the authors and do not necessarily represent the findings and conclusions of the U.S. Fish and Wildlife Service. This is Molette Biology Laboratory contribution 64 and Auburn University Marine Biology Program contribution 158.

3.8 References

1. Beutler B (2004) Innate immunity: An overview. *Mol Immunol* 40(12):845–859.
2. Rast JR, Litman GW (1998) Towards understanding the evolutionary origins and early diversification of rearranging antigen receptors. *Immunol Rev* 116:79–86.
3. Pancer Z, et al. (2004) Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* 430:174–180.
4. Akira S, Uematsu S, Takeuchi O (2006) Pathogen recognition and innate immunity. *Cell* 124:783–801.

5. Dzik JM (2010) The ancestry and cumulative evolution of immune reactions. *Acta Biochim Pol* 57(4):443–466.

6. Gauthier MEA, Du Pasquier L, Degnan BM (2010) The genome of the sponge *Amphimedon queenslandica* provides new perspectives into the origin of Toll-like and interleukin 1 receptor pathways. *Evol Dev* 12(5):519–533.

7. Lemaitre B, Nicolas E, Michaut L, Reichhart J-M, Hoffmann JA (1996) The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in *Drosophila* adults. *Cell* 86:973–983.

8. Botos I, Segal DM, Davies DR (2011) The structural biology of Toll-like receptors. *Structure* 19:447–459.

9. O'Neill LAJ, Bowie AG (2007) The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat Rev Immunol* 7:353–364.

10. Kawai T, Akira S (2010) The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol* 11(5):373–384.

11. Sanjuan MA, et al. (2007) Toll-like receptor signalling in macrophages links the autophagy pathway to phagocytosis. *Nature* 450:1253–1257.

12. Pasare C, Medzhitov R (2004) Toll-like receptors: linking innate and adaptive immunity. *Microbes Infect* 6:1382–1387.

13. O'Neill LA, Golenbock D, Bowie AG (2013) The history of Toll-like receptors - redefining innate immunity. *Nat Rev Immunol* 13:453–460.

14. Cannon JT, et al. (2014) Phylogenomic resolution of the hemichordate and echinoderm clade. *Curr Biol* 24(23):2827–2832.

15. Buckley KM, Rast JP (2015) Diversity of animal immune receptors and the origins of recognition complexity in the deuterostomes. *Dev Comp Immunol* 49:179–189.

16. Hibino T, et al. (2006) The immune gene repertoire encoded in the purple sea urchin genome. *Dev Biol* 300:349–365.

17. Huang S, et al. (2008) Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome Res* 18:1112–1126.

18. Buckley KM, Rast JP (2012) Dynamic evolution of toll-like receptor multigene families in echinoderms. *Front Immunol* 3(136):1-14.

19. Sasaki N, Ogasawara M, Sekiguchi T, Kusumoto S, Satake H (2009) Toll-like receptors of the ascidian *Ciona intestinalis*: prototypes with hybrid functionalities of vertebrate Toll-like receptors. *J Biol Chem* 284(40):27336–27343.

20. Rhodes CP, Ratcliffe NA (1983) Coelomocytes and defence reactions of the primitive chordates, *Branchiostoma lanceolatum* and *Saccoglossus horsti*. *Dev Comp Immunol* 7:695–698.

21. Ho ECH, et al. (2016) Perturbation of gut bacteria induces a coordinated cellular immune response in the purple sea urchin larva. *Immunol Cell Biol* 94:861–874.

22. Cima F, Franchi N, Ballarin L (2016) Origin and functions of tunicate hemocytes. *Evolution of the Immune System,* ed Malagoli D (Academic Press, London), pp 29–49.

23. Smith LC, et al. (2006) The sea urchin immune system. *Invertebr Surviv J* 3:25–39.

24. Yuan S, Ruan J, Huang S, Chen S, Xu A (2015) Amphioxus as a model for investigating evolution of the vertebrate immune system. *Dev Comp Immunol* 48:297–305.

25. Wu B, Jin M, Gong J, Du X, Bai Z (2011) Dynamic evolution of CIKS (TRAF3IP2/Act1) in metazoans. *Dev Comp Immunol* 35:1186–1192.

26. Yuan S, et al. (2015) Characterization of amphioxus IFN regulatory factor family reveals an archaic signaling framework for innate immune response. *J Immunol* 195(12):5657-5666.

27. Clow AL, Raftos AD, Gross SP, Smith LC (2004) The sea urchin complement homologue, SpC3, functions as an opsonin. *J Exp Biol* 207:2147–2155.

28. Zhang Y, et al. (2013) Characteristic and functional analysis of Toll-like receptors (TLRs) in the lophotrocozoan, *Crassostrea gigas*, reveals ancient origin of TLR-mediated innate immunity. *PLoS One* 8(10):1–15.

29. Leulier F, Lemaitre B (2008) Toll-like receptors — taking an evolutionary approach. *Nat Rev Genet* 9:165–178.

30. Terajima D, et al. (2003) Identification of candidate genes encoding the core components of the cell death machinery in the *Ciona intestinalis* genome. *Cell Death Differ* 10:749–753.

31. Kanayama A, et al. (2004) TAB2 and TAB3 activate the NF-kB pathway through binding to polyubiquitin chains. *Mol Cell* 15:535–548.

32. Zhuang ZH, et al. (2006) *Drosophila* TAB2 is required for the immune activation of JNK and NF-kappaB. *Cell Signal* 18:964–970.

33. Ponting CP (2002) Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res* 30(17):3643–3652.

34. Yu C, et al. (2005) Genes "waiting" for recruitment by the adaptive immune system: the insights from amphioxus. *J Immunol* 174:3493–3500.

35. Tamura T, Yanai H, Savitsky D, Taniguchi T (2008) The IRF family transcription factors in immunity and oncogenesis. *Annu Rev Immunol* 26:535–584.

36. Nehyba J, Hrdličková R, Bose HR (2009) Dynamic Evolution of Immune System Regulators: The History of the Interferon Regulatory Factor Family. *Mol Biol Evol* 26(11):2539–2550.

37. Yuan S, et al. (2010) Amphioxus SARM involved in neural development may function as a suppressor of TLR signaling. *J Immunol* 184(12):6874–6881.

38. Lin C-W, Chen C-Y, Cheng S-J, Hu H-T, Hsueh Y-P (2014) Sarm1 deficiency impairs synaptic function and leads to behavioral deficits, which can be ameliorated by an mGluR allosteric modulator. *Front Cell Neurosci* 8(87):1-10.

39. Gerdts J, Summers DW, Milbrandt J, DiAntonio A (2016) Axon self-destruction: new links among SARM1, MAPKs, and NAD+ metabolism. *Neuron* 89:449–460.

40. Letunic I, Doerks T, Bork P (2015) SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res* 43:257–260.

41. Finn RD, et al. (2015) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44:279–285.

42. Suttle CA (2007) Marine viruses – major players in the global ecosystem. *Nat Rev Microbiol* 5:801–812.

43. Brown CT, Howe A, Zhang Q, Pyrkosz AB, Brom TH (2012) A reference-free algorithm for computational normalization of shotgun sequencing data. *arXiv* 1203.4802v2:1–18.

44. Grabherr MG, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29(7):644–652.

45. Haas BJ, et al. (2014) De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat Protoc* 8(8):1–43.

46. Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23):3150–3152.

47. The Uniprot Consortium (2014) UniProt: a hub for protein information. *Nucleic Acids Res* 43:204–212.

48. Camacho C, et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10(421):1-9.

49. Jones P, et al. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236–1240.

50. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580.

51. Offord V, Coffey TJ, Werling D (2010) LRRfinder: A web application for the identification of leucine-rich repeats and an integrative Toll-like receptor database. *Dev Comp Immunol* 34:1035–1041.

52. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*:1303.3997v2:1-3.

53. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772–780.

54. Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest-HPC: fast selection of best-fit models of protein evolution. *Bioinformatics* 27(8):1164–1165.

55. Aberer AJ, Kobert K, Stamatakis A (2014) ExaBayes: massively parallel bayesian tree inference for the whole-genome era. *Mol Biol Evol* 31(10):2553–2556.

56. Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol* 29(6):1695–1701.

57. Maddison W, Maddison D (2016) Mesquite: a modular system for evolutionary analysis. Version 3.2 http://mesquiteproject.org.

58. Roach JC, et al. (2005) The evolution of vertebrate Toll-like receptors. *Proc Natl Acad Sci USA* 102(27):9577–9582.

59. Roach JM, Racioppi L, Jones CD, Masci AM (2013) Phylogeny of Toll-Like receptor signaling: adapting the innate response. *PLoS One* 8(1):1–9.

60. Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.

61. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European molecular biology open software suite. *Trends Genet* 16(6):276–277.

62. Simakov O, et al. (2015) Hemichordate genomes and deuterostome origins. *Nature* 527:459-465.

3.9 Supporting Information

Constrained phylogenetic inference of TLRs. Utilizing the same base dataset, TIR domains amino acid sequences were extracted from TLRs, aligned, and placed in both a likelihood and Bayesian phylogenetic framework using the bioinformatic workflow described in the **Materials and Methods** of the main text. Although both resulting likelihood and Bayesian TIR-

domain-only topologies (**Fig. S6**) were consistent with the topology generated using full TLR sequences (**Fig. 3**), support values in the TIR domain only tree were low, and the resulting topology lacked the resolution yielded by using whole TLR alignment matrices. Despite being allowed to run for nearly double the number of generations as the full TLR Bayesian analysis, MCMC convergences statistics for the TIR-domain-only matrix were reflective of less stationary sampling of parameter space relative to each MCMC chain when compared to the full TLR analysis (*asdsf* = 1.11% and 0.60%, respectively; **Table S3**). According to ExaBayes' recommended convergence diagnostics, an *asdsf* value of 1.11% still falls within the "acceptable convergence" range, whereas ≤ 1.00% is regarded as "excellent convergence" (55). We expect orthology inference across 207 sequences using an amino acid matrix of only 246 sites is insufficient for phylogenetic inference at our target depth. Furthermore, by inclusion of a broad distribution of deuterostome taxa – more accurately representing deep deuterostome phylogeny – we are able to more reliably align complete TLR sequences which would be impossible when restricted to comparisons between model species (e.g., *S. purpuratus* vs. *C. intestinalis* vs. mouse).

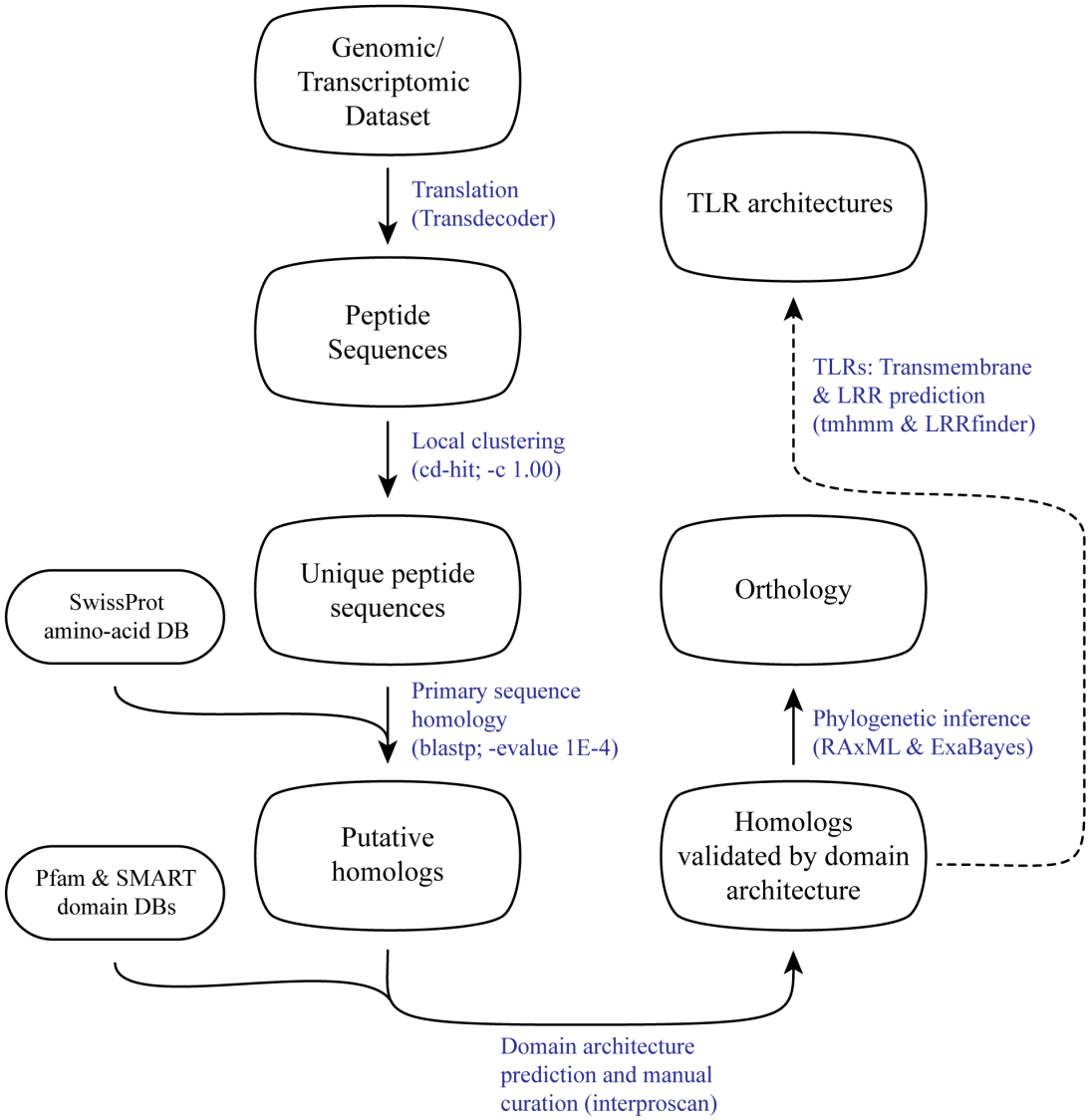


Figure S1. Diagram of bioinformatic pipeline for homology identification. See details in **Materials and Methods** in main text.

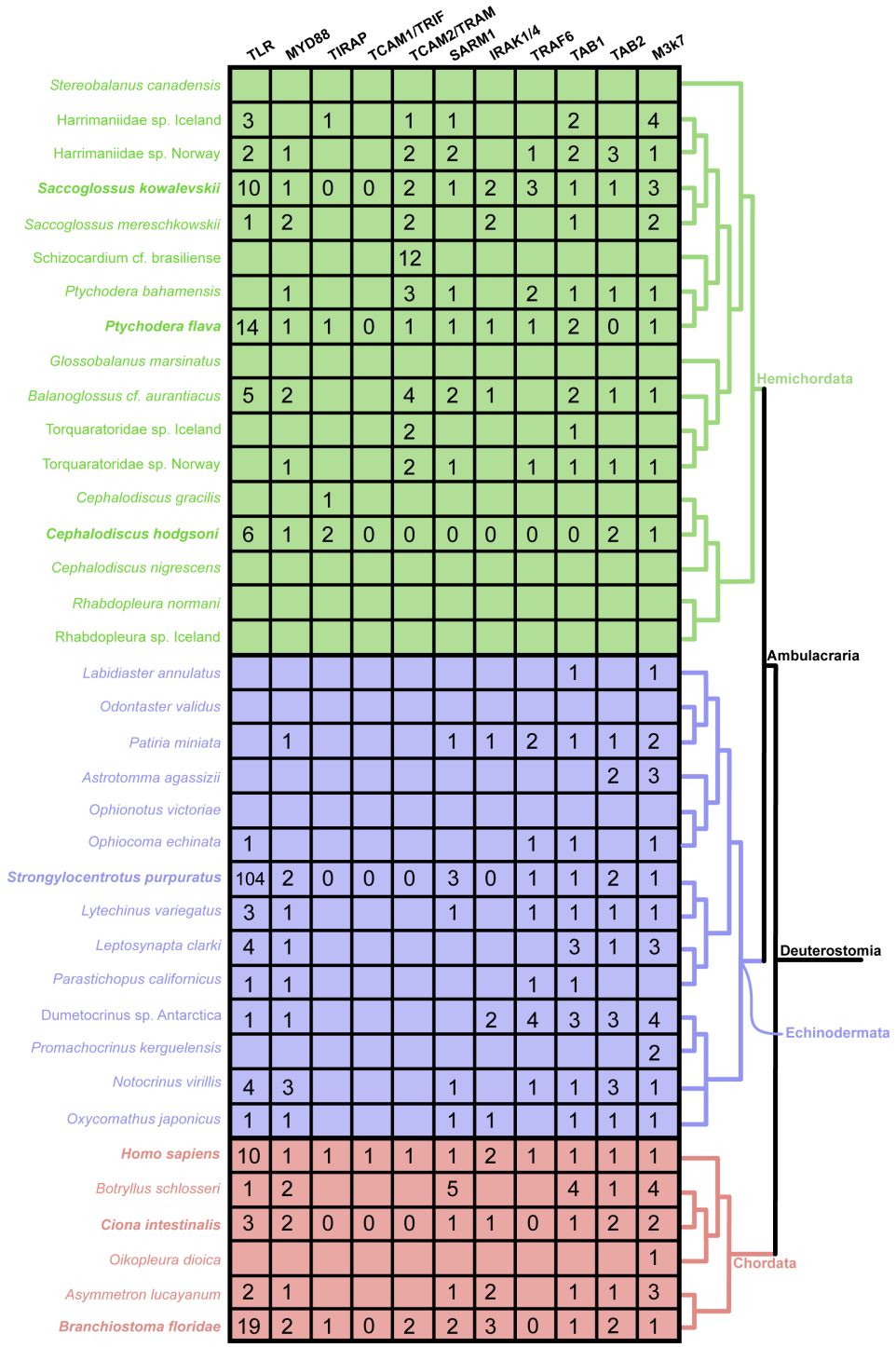


Figure S2. Matrix containing number of homologs identified per taxon. Taxa informed by both transcriptomic and genomic data are indicated in bold. Cells lacking values represent missing data for taxa informed solely by transcriptomic evidence. In contrast, gene absence is denoted with "0" only in taxa for which we analyzed genomic data. A cladogram depicting relationships among taxa is to the right of the matrix. Blocks/clades are colored by phylum: Hemichordata = green, Echinodermata = blue, and Chordata = red.

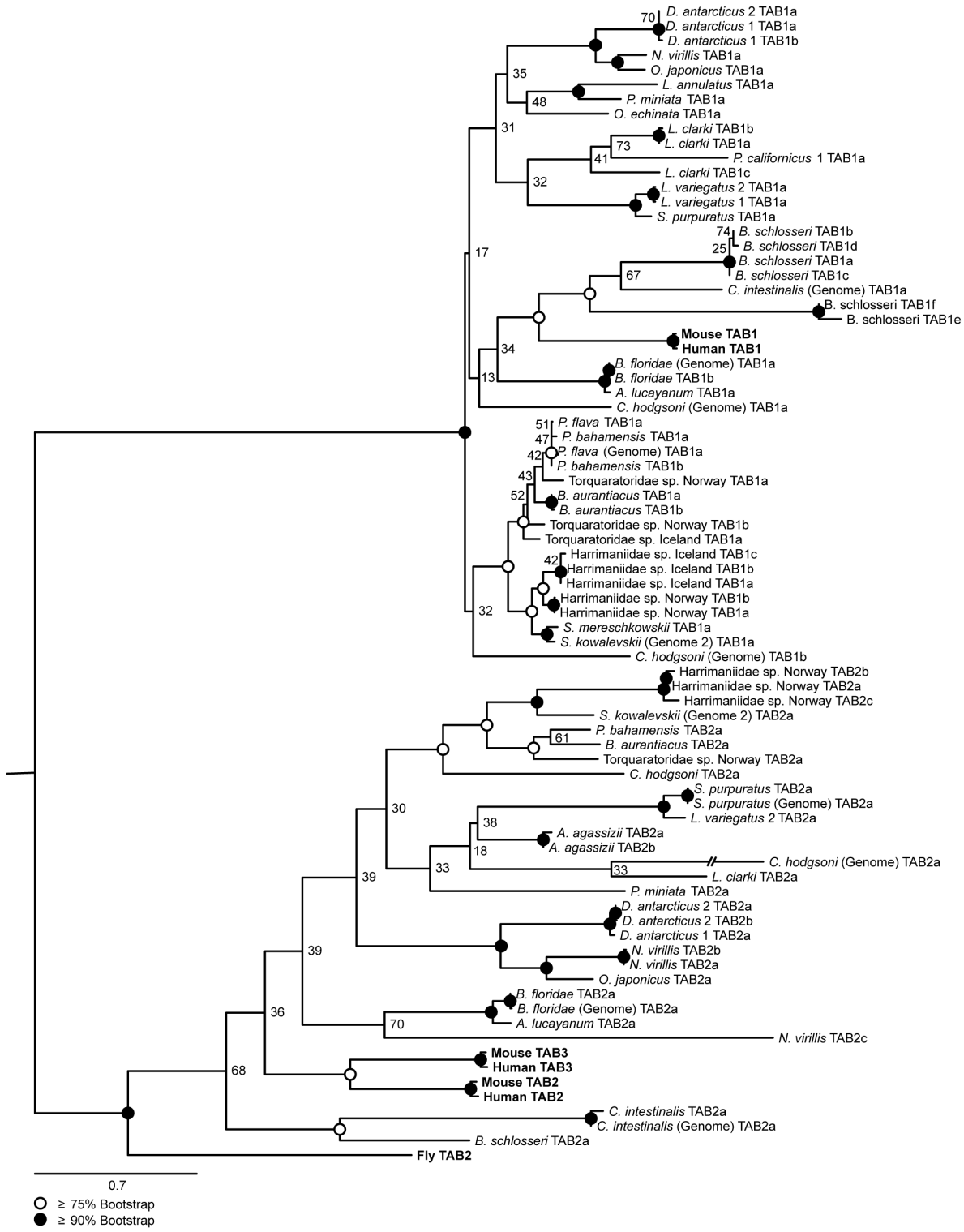


Figure S3. TAB1/2/3 gene-tree built using RAxML (60) rapid bootstrap analysis (1000 replicates) and subsequent inference of best-fitting tree topology. Truncated TAB2's identified among invertebrate deuterostomes ally with human, mouse, and *Drosophila* TAB2/3s. Bold names reflect sequences downloaded from public data repositories (**Table S1**).

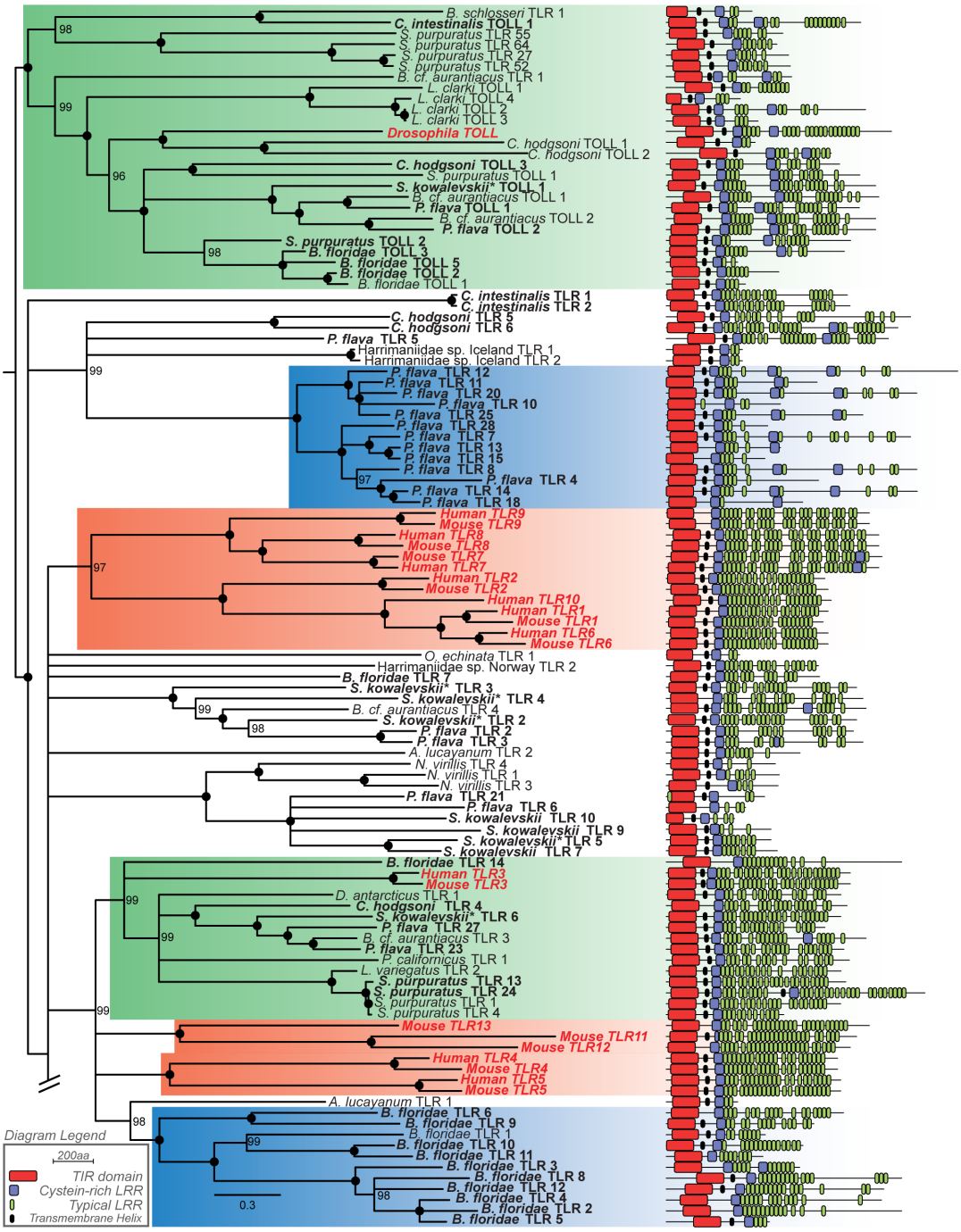


Figure S4. Detailed Bayesian gene-tree of deuterostome TLRs. Reference sequences from human, mouse, and *Drosophila* are highlighted in bold red font whereas deuterostome sequences derived from genomic data are bold. Sub-tree contains Toll homology group, *P. flava* "TLR" expansion, reference TLR groups, TLR3 homology group, and *Branchiostoma floridae* TLR expansion. Clade color highlights are the same as those presented in **Fig. 3.** Black circles denote nodes with 100% posterior probability. Clades supported by less than 95% posterior probability have been collapsed. *Saccoglossus kowalevskii* TLRs with asterisks (*) denote TLRs which were obtained from the revised *S. kowalevskii* genome (62).

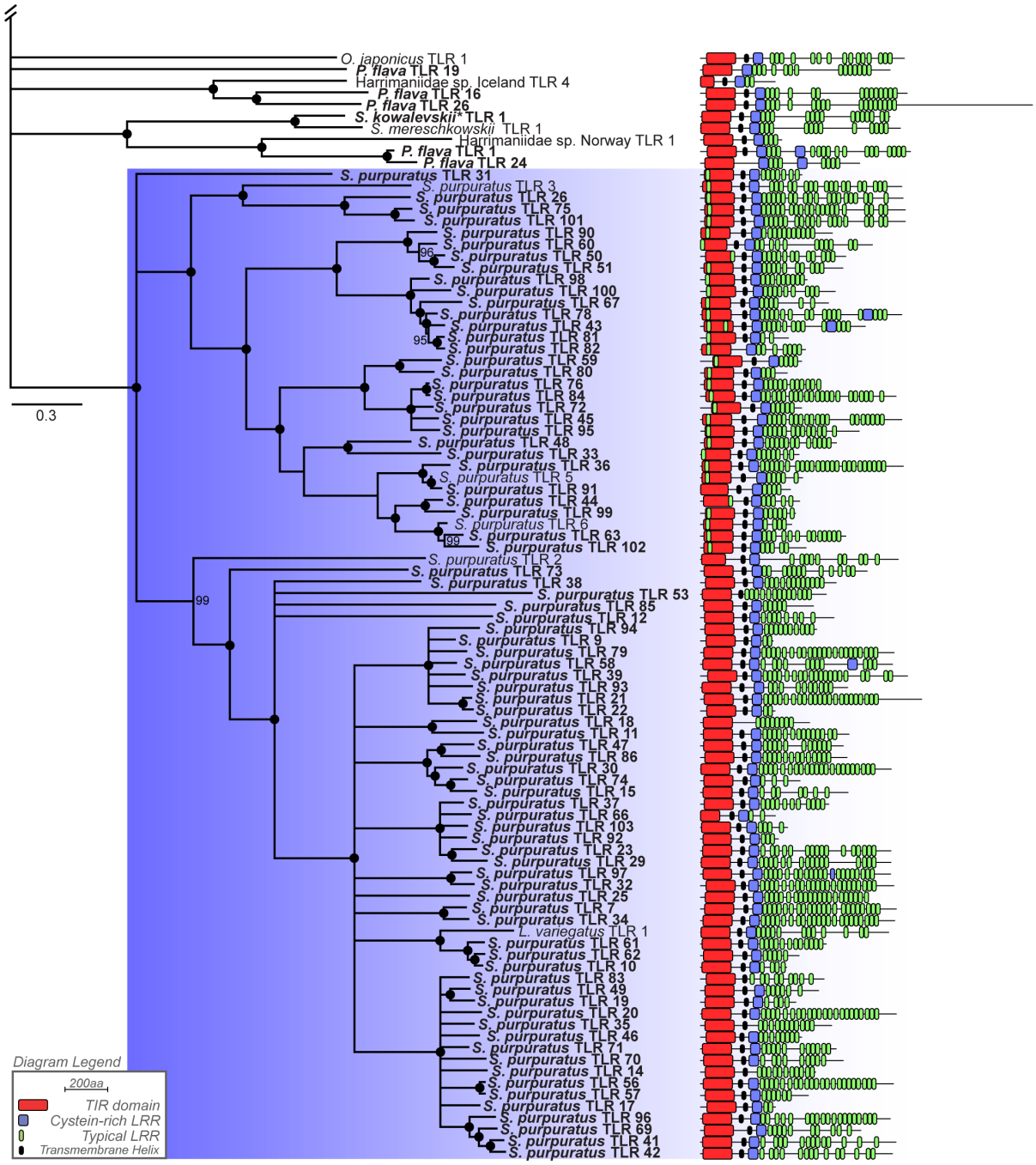


Figure S5. Detailed Bayesian gene-tree of deuterostome TLRs. Reference sequences from human, mouse, and *Drosophila* are highlighted in bold red font whereas deuterostome sequences derived from genomic data are bold. Sub-tree contains *Strongylocentrotus purpuratus* TLR expansion. Clade color highlights are the same as those presented in **Fig. 3.** Black circles denote nodes with 100% posterior probability. Clades supported by less than 95% posterior probability have been collapsed. *Saccoglossus kowalevskii* TLRs with asterisks (*) denote TLRs which were obtained from the revised *S. kowalevskii* genome (62).

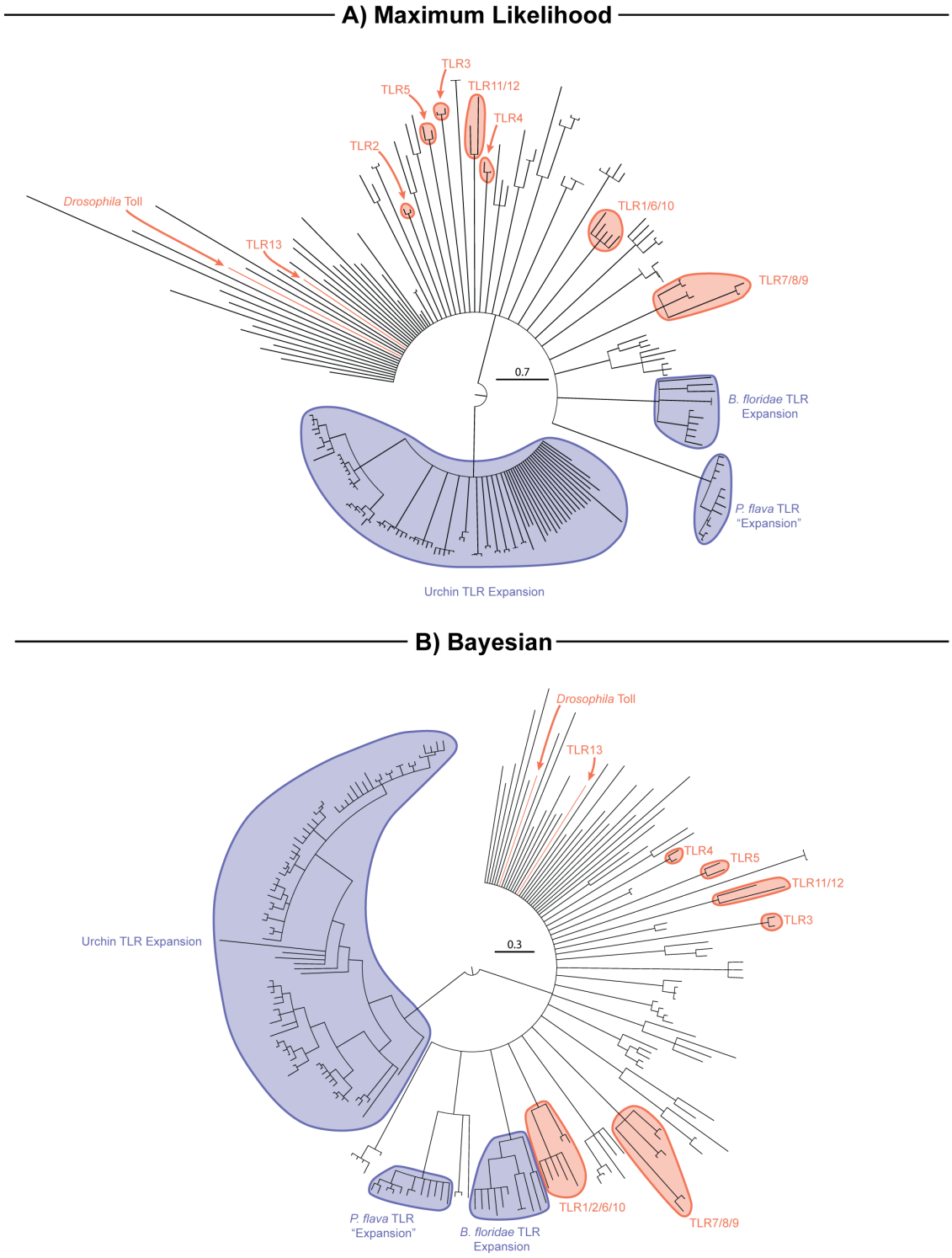


Figure S6. TIR-domain-only TLR amino acid trees. Topologies are consistent with those displayed in **Figs. 3**, **S4** and **S5**, though with considerably lower resolution and node support values. **A)** Tree produced using RAxML; all nodes have ≥75% bootstrap support. **B)** Tree produced using ExaBayes; all nodes have ≥95% posterior probability.

Table S1: Accession numbers associated with molecular datasets and TLR pathway homologs identified per taxon. SRAs and genome assemblies are publicly available via NCBI (www.ncbi.nlm.nih.gov). TLR pathway homologs are available via figshare (10.6084/m9.figshare.4987169).

Datasets Analyzed

Taxon	**Clade**	**Data Type**	**# Unique proteins**	**Accession #**
Asymmetron lucayanum	Chordata: Cephalochordata	T	48578	SRR1138336
Branchiostoma floridae 1	Chordata: Cephalochordata	T	41091	SRR923751
Branchiostoma floridae 2	Chordata: Cephalochordata	GM	27622	GCA_000003815.1 Version 2 CDS
Branchiostoma floridae 3	Chordata: Cephalochordata	GA	N/A	GCF_000003815.1
Botryllus schlosseri	Chordata: Tunicata	T	71778	SRR1604859
Ciona intestinalis 1	Chordata: Tunicata	T	10356	SRR1324870
Ciona intestinalis 2	Chordata: Tunicata	GM	15290	GCF_000224145.2 CDS
Oikoleura dioica	Chordata: Tunicata	T	25709	SRR1693766
Homo sapiens	Chordata: Vertebrata	GM	92248	GCF_000001405.30 CDS
Astrotomma agassizii	Echinodermata: Asteroidea	T	26264	SRR1695485
Labidiaster annulatus	Echinodermata: Asteroidea	T	16573	SRR1695480
Odontaster validus	Echinodermata: Asteroidea	T	160	SRR1695479
Patiria miniata	Echinodermata: Asteroidea	T	29570	SRR1138705
Dumetocrinus antarcticus 2	Echinodermata: Crinoidea	T	39854	SRR5564112
Dumetocrinus sp Antarctica 1	Echinodermata: Crinoidea	T	28635	SRR1695483
Notocrinus virillis	Echinodermata: Crinoidea	T	43536	SRR5564113
Oxycomathus japonicus	Echinodermata: Crinoidea	T	18972	SRR1138706
Promachocrinus kerguelensis 1	Echinodermata: Crinoidea	T	546	SRR1695482
Promachocrinus kerguelensis 2	Echinodermata: Crinoidea	T	20474	SRR5564111
Lytechinus variegatus 1	Echinodermata: Echinoidea	T	8362	PRJNA62467
Lytechinus variegatus 2	Echinodermata: Echinoidea	T	26739	SRR1139214

Strongylocentrotus purpuratus 1	Echinodermata: Echinoidea	T	25384	PRJNA81157
Strongylocentrotus purpuratus 2	Echinodermata: Echinoidea	GM	32868	GCF_000002235.4 CDS
Strongylocentrotus purpuratus 3	Echinodermata: Echinoidea	GA	N/A	GCF_000002235.4
Leptosynapta clarki	Echinodermata: Holothuroidea	T	33707	SRR1695478
Parastichopus californicus 1	Echinodermata: Holothuroidea	T	23923	SRR1695477
Parastichopus californicus 2	Echinodermata: Holothuroidea	T	12080	SRR1139198
Ophiocoma echinata	Echinodermata: Ophiuroidea	T	25243	SRR1138707
Ophionotus victoriae	Echinodermata: Ophiuroidea	T	377	SRR1695484
Balanoglossus cf. aurantiacus	Hemichordata: Enteropneusta	T	35437	SRR1695460
Glossobalanus marsinatus	Hemichordata: Enteropneusta	T	3735	SRR1695459
Harrimaniidae sp. Iceland	Hemichordata: Enteropneusta	T	49117	SRR1695462
Harrimaniidae sp. Norway	Hemichordata: Enteropneusta	T	55596	SRR1695463
Ptychodera bahamensis	Hemichordata: Enteropneusta	T	30780	SRR1695458
Ptychodera flava 1	Hemichordata: Enteropneusta	T	9674	SRR1029584
Ptychodera flava 2	Hemichordata: Enteropneusta	GM	30537	GCA_001465055.1
Saccoglossus kowalevskii 1	Hemichordata: Enteropneusta	T	5147	SRR071719
Saccoglossus kowalevskii 2	Hemichordata: Enteropneusta	GM	22059	GCF_000003605.2
Saccoglossus kowalevskii 3	Hemichordata: Enteropneusta	GM	29411	[See Simakov et al., 2016 (62)]
Saccoglossus mereschkowskii	Hemichordata: Enteropneusta	T	41268	SRR1695461
Schizocardium cf. brasiliense	Hemichordata: Enteropneusta	T	31767	SRR1695467
Stereobalanus canadensis 1	Hemichordata: Enteropneusta	T	1890	SRR1697122
Stereobalanus canadensis 2	Hemichordata: Enteropneusta	T	2224	SRR1695465
Torquaratorid sp. Iceland	Hemichordata: Enteropneusta	T	25570	SRR1695468
Torquaratorid sp. Norway	Hemichordata: Enteropneusta	T	33277	SRR1695469
Cephalodiscus gracilis	Hemichordata: Pterobranchia	T	13366	SRR1695473

Cephalodiscus hodgsoni 1	Hemichordata: Pterobranchia	T	54201	SRR1695470
Cephalodiscus hodgsoni 2	Hemichordata: Pterobranchia	GM	40310	See [10.6084/m9.figshare.4987169]
Cephalodiscus nigrescens	Hemichordata: Pterobranchia	T	1782	SRR1695472
Rhabdopleura normani	Hemichordata: Pterobranchia	T	361	SRR1695475
Rhabdopleura sp. Iceland	Hemichordata: Pterobranchia	T	558	SRR1695476

TLR Pathway Homologs Identified

Species	**Homolog**	**Sequence Header**
Astrotomma agassizii	MAP3K7	AstagM3K7_1
Astrotomma agassizii	MAP3K7	AstagM3K7_2
Astrotomma agassizii	MAP3K7	AstagM3K7_3
Astrotomma agassizii	TAB2	AstAgTAB2_1
Astrotomma agassizii	TAB2	AstAgTAB2_2
Astrotomma agassizii	IRAK1	AsyluIRAK1_1
Asymmetron lucayanum	IRAK4	AsyluIRAK4_1
Asymmetron lucayanum	MAP3K7	AsyluM3K7_1
Asymmetron lucayanum	MAP3K7	AsyluM3K7_2
Asymmetron lucayanum	MAP3K7	AsyluM3K7_3
Asymmetron lucayanum	MYD88	AsyluMYD88_1
Asymmetron lucayanum	SARM1	AsyluSARM1_1
Asymmetron lucayanum	TAB1	AsyluTAB1_1
Asymmetron lucayanum	TAB2	AsyluTAB2_1
Asymmetron lucayanum	TLR	AsyluTLR_1
Asymmetron lucayanum	TLR	AsyluTLR_2
Balanoglossus aurantiacus	IRAK1	BalauIRAK4_1
Balanoglossus aurantiacus	MAP3K7	BalauM3K7_1
Balanoglossus aurantiacus	MYD88	BalauMYD88_1
Balanoglossus aurantiacus	MYD88	BalauMYD88_2
Balanoglossus aurantiacus	SARM1	BalauSARM1_1
Balanoglossus aurantiacus	SARM1	BalauSARM1_2
Balanoglossus aurantiacus	TAB1	BalauTAB1_1
Balanoglossus aurantiacus	TAB1	BalauTAB1_2
Balanoglossus aurantiacus	TAB2	BalauTAB2_1
Balanoglossus aurantiacus	TICAM2	BalauTCAM2_1
Balanoglossus aurantiacus	TICAM2	BalauTCAM2_2
Balanoglossus aurantiacus	TICAM2	BalauTCAM2_3
Balanoglossus aurantiacus	TICAM2	BalauTCAM2_4
Balanoglossus aurantiacus	TLR	BalauTLR_1

Balanoglossus aurantiacus	TLR	BalauTLR_2
Balanoglossus aurantiacus	TLR	BalauTLR_3
Balanoglossus aurantiacus	TLR	BalauTLR_4
Balanoglossus aurantiacus	TLR	BalauTOLL_1
Balanoglossus aurantiacus	TLR	BalauTOLL_2
Botryllus schlosseri	MAP3K7	BotscM3K7_1
Botryllus schlosseri	MAP3K7	BotscM3K7_2
Botryllus schlosseri	MAP3K7	BotscM3K7_3
Botryllus schlosseri	MAP3K7	BotscM3K7_4
Botryllus schlosseri	MYD88	BotscMYD88_1
Botryllus schlosseri	MYD88	BotscMYD88_2
Botryllus schlosseri	SARM1	BotscSARM1_1
Botryllus schlosseri	SARM1	BotscSARM1_2
Botryllus schlosseri	SARM1	BotscSARM1_3
Botryllus schlosseri	SARM1	BotscSARM1_4
Botryllus schlosseri	SARM1	BotscSARM1_5
Botryllus schlosseri	TAB1	BotscTAB1_1
Botryllus schlosseri	TAB1	BotscTAB1_2
Botryllus schlosseri	TAB1	BotscTAB1_3
Botryllus schlosseri	TAB1	BotscTAB1_4
Botryllus schlosseri	TAB1	BotscTAB1_5
Botryllus schlosseri	TAB1	BotscTAB1_6
Botryllus schlosseri	TAB2	BotscTAB2_1
Botryllus schlosseri	TLR	BotscTLR_1
Branchiostoma floridae	IRAK1	BraflIRAK4_1
Branchiostoma floridae	IRAK1	BraflIRAK4_2
Branchiostoma floridae	IRAK1	BraflIRAK4_3
Branchiostoma floridae	MAP3K7	BraflM3K7_1
Branchiostoma floridae	MYD88	BraflMYD88_1
Branchiostoma floridae	MYD88	BraflMYD88_2
Branchiostoma floridae	SARM1	BraflSARM1_1
Branchiostoma floridae	SARM1	BraflSARM1_2
Branchiostoma floridae	TAB1	BraflTAB1_1
Branchiostoma floridae	TAB1	BraflTAB1_2
Branchiostoma floridae	TAB2	BraflTAB2_1
Branchiostoma floridae	TAB2	BraflTAB2_2
Branchiostoma floridae	TICAM2	BraflTCAM2_1
Branchiostoma floridae	TICAM2	BraflTCAM2_2
Branchiostoma floridae	TIRAP	BraflTIRAP_1
Branchiostoma floridae	TLR	BraflTLR_1
Branchiostoma floridae	TLR	BraflTLR_10
Branchiostoma floridae	TLR	BraflTLR_11
Branchiostoma floridae	TLR	BraflTLR_12

Branchiostoma floridae	TLR	BraflTLR_13
Branchiostoma floridae	TLR	BraflTLR_14
Branchiostoma floridae	TLR	BraflTLR_2
Branchiostoma floridae	TLR	BraflTLR_3
Branchiostoma floridae	TLR	BraflTLR_4
Branchiostoma floridae	TLR	BraflTLR_5
Branchiostoma floridae	TLR	BraflTLR_6
Branchiostoma floridae	TLR	BraflTLR_7
Branchiostoma floridae	TLR	BraflTLR_8
Branchiostoma floridae	TLR	BraflTLR_9
Branchiostoma floridae	TLR	BraflTOLL_1
Branchiostoma floridae	TLR	BraflTOLL_2
Branchiostoma floridae	TLR	BraflTOLL_3
Branchiostoma floridae	TLR	BraflTOLL_4
Branchiostoma floridae	TLR	BraflTOLL_5
Cephalodiscus gracilis	TIRAP	CepgrTIRAP_1
Cephalodiscus hodgsoni	MAP3K7	CephoM3K7_1
Cephalodiscus hodgsoni	MYD88	CephoMYD88_1
Cephalodiscus hodgsoni	TAB1	CephoTAB1_1
Cephalodiscus hodgsoni	TAB1	CephoTAB1_2
Cephalodiscus hodgsoni	TAB2	CephoTAB2_1
Cephalodiscus hodgsoni	TAB2	CephoTAB2_2
Cephalodiscus hodgsoni	TIRAP	CephoTIRAP_1
Cephalodiscus hodgsoni	TIRAP	CephoTIRAP_1
Cephalodiscus hodgsoni	TLR	CephoTLR_1
Cephalodiscus hodgsoni	TLR	CephoTLR_2
Cephalodiscus hodgsoni	TLR	CephoTLR_3
Cephalodiscus hodgsoni	TLR	CephoTLR_4
Cephalodiscus hodgsoni	TLR	CephoTLR_5
Cephalodiscus hodgsoni	TLR	CephoTLR_6
Cephalodiscus hodgsoni	TLR	CephoTOLL_1
Cephalodiscus hodgsoni	TLR	CephoTOLL_2
Cephalodiscus hodgsoni	TLR	CephoTOLL_3
Ciona intestinalis	IRAK1	CioinIRAK4_1
Ciona intestinalis	MAP3K7	CioinM3K7_1
Ciona intestinalis	MAP3K7	CioinM3K7_2
Ciona intestinalis	MYD88	CioinMYD88_1
Ciona intestinalis	MYD88	CioinMYD88_2
Ciona intestinalis	SARM1	CioinSARM1_1
Ciona intestinalis	TAB1	CioinTAB1_1
Ciona intestinalis	TAB2	CioinTAB2_1
Ciona intestinalis	TAB2	CioinTAB2_2
Ciona intestinalis	TLR	CioinTLR_1

Ciona intestinalis	TLR	CioinTLR_2
Ciona intestinalis	TLR	CioinTOLL_1
Dumetocrinus antarcticus	IRAK1	DumanIRAK1_1
Dumetocrinus antarcticus	IRAK1	DumanIRAK4_1
Dumetocrinus antarcticus	MAP3K7	DumanM3K7_1
Dumetocrinus antarcticus	MAP3K7	DumanM3K7_1
Dumetocrinus antarcticus	MAP3K7	DumanM3K7_2
Dumetocrinus antarcticus	MAP3K7	DumanM3K7_2
Dumetocrinus antarcticus	MYD88	DumanMYD88_4
Dumetocrinus antarcticus	TAB1	DumanTAB1_1
Dumetocrinus antarcticus	TAB1	DumanTAB1_2
Dumetocrinus antarcticus	TAB1	DumanTAB1_2
Dumetocrinus antarcticus	TAB2	DumanTAB2_1
Dumetocrinus antarcticus	TAB2	DumanTAB2_2
Dumetocrinus antarcticus	TAB2	DumanTAB2_3
Dumetocrinus antarcticus	TLR	DumanTLR_1
Dumetocrinus antarcticus	TRAF6	DumanTRAF6_1
Dumetocrinus antarcticus	TRAF6	DumanTRAF6_2
Dumetocrinus antarcticus	TRAF6	DumanTRAF6_3
Dumetocrinus antarcticus	TRAF6	DumanTRAF6_4
Harrimaniid sp. Iceland	MAP3K7	HarIcM3K7_1
Harrimaniid sp. Iceland	MAP3K7	HarIcM3K7_2
Harrimaniid sp. Iceland	MAP3K7	HarIcM3K7_3
Harrimaniid sp. Iceland	MAP3K7	HarIcM3K7_4
Harrimaniid sp. Iceland	TAB1	HarIcTAB1_1
Harrimaniid sp. Iceland	TAB1	HarIcTAB1_2
Harrimaniid sp. Iceland	TAB1	HarIcTAB1_3
Harrimaniid sp. Iceland	TICAM2	HarIcTCAM2_1
Harrimaniid sp. Iceland	TIRAP	HarIcTIRAP_1
Harrimaniid sp. Iceland	TLR	HarIcTLR_1
Harrimaniid sp. Iceland	TLR	HarIcTLR_2
Harrimaniid sp. Iceland	TLR	HarIcTLR_3
Harrimaniid sp. Iceland	TLR	HarIcTLR_4
Harrimaniid sp. Norway	MAP3K7	HarNoM3K7_1
Harrimaniid sp. Norway	MAP3K7	HarNoM3K7_2
Harrimaniid sp. Norway	MYD88	HarNoMYD88_1
Harrimaniid sp. Norway	SARM1	HarNoSARM1_1
Harrimaniid sp. Norway	SARM1	HarNoSARM1_2
Harrimaniid sp. Norway	TAB1	HarNoTAB1_1
Harrimaniid sp. Norway	TAB1	HarNoTAB1_2
Harrimaniid sp. Norway	TAB2	HarNoTAB2_1
Harrimaniid sp. Norway	TAB2	HarNoTAB2_2
Harrimaniid sp. Norway	TAB2	HarNoTAB2_3

Harrimaniid sp. Norway	TICAM2	HarNoTCAM2_1
Harrimaniid sp. Norway	TICAM2	HarNoTCAM2_2
Harrimaniid sp. Norway	TLR	HarNoTLR_1
Harrimaniid sp. Norway	TLR	HarNoTLR_2
Harrimaniid sp. Norway	TRAF6	HarNoTRAF6_1
Labidiaster annulatus	MAP3K7	LabanM3K7_1
Labidiaster annulatus	TAB1	LabanTAB1_1
Leptosynapta clarki	MAP3K7	LepclM3K7_1
Leptosynapta clarki	MAP3K7	LepclM3K7_2
Leptosynapta clarki	MAP3K7	LepclM3K7_3
Leptosynapta clarki	MYD88	LepclMYD88_1
Leptosynapta clarki	TAB1	LepclTAB1_1
Leptosynapta clarki	TAB1	LepclTAB1_2
Leptosynapta clarki	TAB1	LepclTAB1_3
Leptosynapta clarki	TAB2	LepclTAB2_1
Leptosynapta clarki	TLR	LepclTOLL_1
Leptosynapta clarki	TLR	LepclTOLL_2
Leptosynapta clarki	TLR	LepclTOLL_3
Leptosynapta clarki	TLR	LepclTOLL_4
Lytechinus variegatus	MAP3K7	LytvaM3K7_1
Lytechinus variegatus	MYD88	LytvaMYD88_1
Lytechinus variegatus	SARM1	LytvaSARM1_1
Lytechinus variegatus	TAB1	LytvaTAB1_1
Lytechinus variegatus	TAB1	LytvaTAB1_2
Lytechinus variegatus	TAB2	LytvaTAB2_1
Lytechinus variegatus	TLR	LytvaTLR_1
Lytechinus variegatus	TLR	LytvaTLR_2
Lytechinus variegatus	TLR	LytvaTOLL_1
Lytechinus variegatus	TRAF6	LytvaTRAF6_1
Notocrinus virillis	MAP3K7	NotviM3K7_1
Notocrinus virillis	MAP3K7	NotviM3K7_2
Notocrinus virillis	MAP3K7	NotviM3K7_3
Notocrinus virillis	MYD88	NotviMYD88_1
Notocrinus virillis	MYD88	NotviMYD88_2
Notocrinus virillis	MYD88	NotviMYD88_3
Notocrinus virillis	SARM1	NotviSARM1_1
Notocrinus virillis	TAB1	NotviTAB1_1
Notocrinus virillis	TAB2	NotviTAB2_1
Notocrinus virillis	TAB2	NotviTAB2_2
Notocrinus virillis	TAB2	NotviTAB2_3
Notocrinus virillis	TLR	NotviTLR_1
Notocrinus virillis	TLR	NotviTLR_2
Notocrinus virillis	TLR	NotviTLR_3

Notocrinus virillis	TLR	NotviTLR_4
Notocrinus virillis	TRAF6	NotviTRAF6_1
Oikiopleura dioica	MAP3K7	OikdiM3K7_1
Ophiocoma echinata	MAP3K7	OphecM3K7_1
Ophiocoma echinata	TAB1	OphecTAB1_1
Ophiocoma echinata	TLR	OphecTLR_1
Ophiocoma echinata	TRAF6	OphecTRAF6_1
Oxycomathus japonicus	IRAK1	OxyjaIRAK4_1
Oxycomathus japonicus	MAP3K7	OxyjaM3K7_1
Oxycomathus japonicus	MYD88	OxyjaMYD88_1
Oxycomathus japonicus	SARM1	OxyjaSARM1_1
Oxycomathus japonicus	TAB1	OxyjaTAB1_1
Oxycomathus japonicus	TAB2	OxyjaTAB2_1
Oxycomathus japonicus	TLR	OxyjaTLR_1
Parastichopus californicus	MYD88	ParcaMYD88_1
Parastichopus californicus	TAB1	ParcaTAB1_1
Parastichopus californicus	TLR	ParcaTLR_1
Parastichopus californicus	TRAF6	ParcaTRAF6_1
Patiria miniata	IRAK1	PatmiIRAK4_1
Patiria miniata	MAP3K7	PatmiM3K7_1
Patiria miniata	MAP3K7	PatmiM3K7_2
Patiria miniata	MYD88	PatmiMYD88_1
Patiria miniata	SARM1	PatmiSARM1_1
Patiria miniata	TAB1	PatmiTAB1_1
Patiria miniata	TAB2	PatmiTAB2_1
Patiria miniata	TRAF6	PatmiTRAF6_1
Patiria miniata	TRAF6	PatmiTRAF6_2
Promachocrinus kerguelensis	MAP3K7	ProkeM3K7_1
Promachocrinus kerguelensis	MAP3K7	ProkeM3K7_2
Ptychodera bahamansis	MAP3K7	PtybaM3K7_1
Ptychodera bahamansis	MYD88	PtybaMYD88_1
Ptychodera bahamansis	SARM1	PtybaSARM1_1
Ptychodera bahamansis	TAB1	PtybaTAB1_1
Ptychodera bahamansis	TAB1	PtybaTAB1_2
Ptychodera bahamansis	TAB2	PtybaTAB2_1
Ptychodera bahamansis	TICAM2	PtybaTCAM2_1
Ptychodera bahamansis	TICAM2	PtybaTCAM2_2
Ptychodera bahamansis	TICAM2	PtybaTCAM2_3
Ptychodera bahamansis	TLR	PtybaTOLL_1
Ptychodera bahamansis	TRAF6	PtybaTRAF6_1
Ptychodera bahamansis	TRAF6	PtybaTRAF6_1
Ptychodera flava	IRAK1	PtyflIRAK1_1
Ptychodera flava	MAP3K7	PtyflM3K7_1

Ptychodera flava	MYD88	PtyflMYD88_1
Ptychodera flava	SARM1	PtyflSARM1_1
Ptychodera flava	TAB1	PtyflTAB1_1
Ptychodera flava	TAB1	PtyflTAB1_2
Ptychodera flava	TICAM2	PtyflTCAM2_1
Ptychodera flava	TIRAP	PtyflTIRAP_1
Ptychodera flava	TLR	PtyflTLR_1
Ptychodera flava	TLR	PtyflTLR_10
Ptychodera flava	TLR	PtyflTLR_11
Ptychodera flava	TLR	PtyflTLR_12
Ptychodera flava	TLR	PtyflTLR_13
Ptychodera flava	TLR	PtyflTLR_14
Ptychodera flava	TLR	PtyflTLR_15
Ptychodera flava	TLR	PtyflTLR_16
Ptychodera flava	TLR	PtyflTLR_17
Ptychodera flava	TLR	PtyflTLR_18
Ptychodera flava	TLR	PtyflTLR_19
Ptychodera flava	TLR	PtyflTLR_2
Ptychodera flava	TLR	PtyflTLR_20
Ptychodera flava	TLR	PtyflTLR_21
Ptychodera flava	TLR	PtyflTLR_22
Ptychodera flava	TLR	PtyflTLR_23
Ptychodera flava	TLR	PtyflTLR_24
Ptychodera flava	TLR	PtyflTLR_25
Ptychodera flava	TLR	PtyflTLR_26
Ptychodera flava	TLR	PtyflTLR_27
Ptychodera flava	TLR	PtyflTLR_28
Ptychodera flava	TLR	PtyflTLR_3
Ptychodera flava	TLR	PtyflTLR_4
Ptychodera flava	TLR	PtyflTLR_5
Ptychodera flava	TLR	PtyflTLR_6
Ptychodera flava	TLR	PtyflTLR_7
Ptychodera flava	TLR	PtyflTLR_8
Ptychodera flava	TLR	PtyflTLR_9
Ptychodera flava	TLR	PtyflTOLL_1
Ptychodera flava	TLR	PtyflTOLL_2
Ptychodera flava	TRAF6	PtyflTRAF6_1
Saccoglossus kowalevskii	IRAK1	SackoIRAK1_1
Saccoglossus kowalevskii	IRAK1	SackoIRAK4_1
Saccoglossus kowalevskii	MAP3K7	SackoM3K7_1
Saccoglossus kowalevskii	MAP3K7	SackoM3K7_2
Saccoglossus kowalevskii	MAP3K7	SackoM3K7_3
Saccoglossus kowalevskii	MYD88	SackoMYD88_1

Saccoglossus kowalevskii	SARM1	SackoSARM1_2
Saccoglossus kowalevskii	TAB1	SackoTAB1_1
Saccoglossus kowalevskii	TAB2	SackoTAB2_1
Saccoglossus kowalevskii	TICAM2	SackoTCAM2_1
Saccoglossus kowalevskii	TICAM2	SackoTCAM2_1
Saccoglossus kowalevskii	TLR	SackoTLR_1
Saccoglossus kowalevskii	TLR	SackoTLR_10
Saccoglossus kowalevskii	TLR	SackoTLR_11
Saccoglossus kowalevskii	TLR	SackoTLR_2
Saccoglossus kowalevskii	TLR	SackoTLR_3
Saccoglossus kowalevskii	TLR	SackoTLR_4
Saccoglossus kowalevskii	TLR	SackoTLR_5
Saccoglossus kowalevskii	TLR	SackoTLR_6
Saccoglossus kowalevskii	TLR	SackoTLR_7
Saccoglossus kowalevskii	TLR	SackoTLR_8
Saccoglossus kowalevskii	TLR	SackoTLR_9
Saccoglossus kowalevskii	TLR	SackoTOLL_1
Saccoglossus kowalevskii	TRAF6	SackoTRAF6_1
Saccoglossus kowalevskii	TRAF6	SackoTRAF6_2
Saccoglossus kowalevskii	TRAF6	SackoTRAF6_3
Saccoglossus mereschkowskii	IRAK1	SacmeIRAK4_1
Saccoglossus mereschkowskii	IRAK1	SacmeIRAK4_2
Saccoglossus mereschkowskii	MAP3K7	SacmeM3K7_1
Saccoglossus mereschkowskii	MAP3K7	SacmeM3K7_2
Saccoglossus mereschkowskii	MYD88	SacmeMYD88_1
Saccoglossus mereschkowskii	MYD88	SacmeMYD88_2
Saccoglossus mereschkowskii	TAB1	SacmeTAB1_1
Saccoglossus mereschkowskii	TICAM2	SacmeTCAM2_1
Saccoglossus mereschkowskii	TICAM2	SacmeTCAM2_2
Saccoglossus mereschkowskii	TLR	SacmeTLR_1
Schizocardium brasiliense	TICAM2	SchbrTCAM2_1
Schizocardium brasiliense	TICAM2	SchbrTCAM2_10
Schizocardium brasiliense	TICAM2	SchbrTCAM2_11
Schizocardium brasiliense	TICAM2	SchbrTCAM2_12
Schizocardium brasiliense	TICAM2	SchbrTCAM2_2
Schizocardium brasiliense	TICAM2	SchbrTCAM2_3
Schizocardium brasiliense	TICAM2	SchbrTCAM2_4
Schizocardium brasiliense	TICAM2	SchbrTCAM2_5
Schizocardium brasiliense	TICAM2	SchbrTCAM2_6
Schizocardium brasiliense	TICAM2	SchbrTCAM2_7
Schizocardium brasiliense	TICAM2	SchbrTCAM2_8
Schizocardium brasiliense	TICAM2	SchbrTCAM2_9
Strongylocentrotus purpuratus	MAP3K7	StrpuM3K7_1

Strongylocentrotus purpuratus	MYD88	StrpuMYD88_1
Strongylocentrotus purpuratus	MYD88	StrpuMYD88_2
Strongylocentrotus purpuratus	SARM1	StrpuSARM1_1
Strongylocentrotus purpuratus	SARM1	StrpuSARM1_2
Strongylocentrotus purpuratus	SARM1	StrpuSARM1_3
Strongylocentrotus purpuratus	TAB1	StrpuTAB1_1
Strongylocentrotus purpuratus	TAB2	StrpuTAB2_1
Strongylocentrotus purpuratus	TAB2	StrpuTAB2_2
Strongylocentrotus purpuratus	TLR	StrpuTLR_1
Strongylocentrotus purpuratus	TLR	StrpuTLR_10
Strongylocentrotus purpuratus	TLR	StrpuTLR_100
Strongylocentrotus purpuratus	TLR	StrpuTLR_101
Strongylocentrotus purpuratus	TLR	StrpuTLR_102
Strongylocentrotus purpuratus	TLR	StrpuTLR_103
Strongylocentrotus purpuratus	TLR	StrpuTLR_11
Strongylocentrotus purpuratus	TLR	StrpuTLR_12
Strongylocentrotus purpuratus	TLR	StrpuTLR_13
Strongylocentrotus purpuratus	TLR	StrpuTLR_14
Strongylocentrotus purpuratus	TLR	StrpuTLR_15
Strongylocentrotus purpuratus	TLR	StrpuTLR_16
Strongylocentrotus purpuratus	TLR	StrpuTLR_17
Strongylocentrotus purpuratus	TLR	StrpuTLR_18
Strongylocentrotus purpuratus	TLR	StrpuTLR_19
Strongylocentrotus purpuratus	TLR	StrpuTLR_2
Strongylocentrotus purpuratus	TLR	StrpuTLR_20
Strongylocentrotus purpuratus	TLR	StrpuTLR_21
Strongylocentrotus purpuratus	TLR	StrpuTLR_22
Strongylocentrotus purpuratus	TLR	StrpuTLR_23
Strongylocentrotus purpuratus	TLR	StrpuTLR_24
Strongylocentrotus purpuratus	TLR	StrpuTLR_25
Strongylocentrotus purpuratus	TLR	StrpuTLR_26
Strongylocentrotus purpuratus	TLR	StrpuTLR_27
Strongylocentrotus purpuratus	TLR	StrpuTLR_28
Strongylocentrotus purpuratus	TLR	StrpuTLR_29
Strongylocentrotus purpuratus	TLR	StrpuTLR_3
Strongylocentrotus purpuratus	TLR	StrpuTLR_30
Strongylocentrotus purpuratus	TLR	StrpuTLR_31
Strongylocentrotus purpuratus	TLR	StrpuTLR_32
Strongylocentrotus purpuratus	TLR	StrpuTLR_33
Strongylocentrotus purpuratus	TLR	StrpuTLR_34
Strongylocentrotus purpuratus	TLR	StrpuTLR_35
Strongylocentrotus purpuratus	TLR	StrpuTLR_36
Strongylocentrotus purpuratus	TLR	StrpuTLR_37

Strongylocentrotus purpuratus	TLR	StrpuTLR_38
Strongylocentrotus purpuratus	TLR	StrpuTLR_39
Strongylocentrotus purpuratus	TLR	StrpuTLR_4
Strongylocentrotus purpuratus	TLR	StrpuTLR_40
Strongylocentrotus purpuratus	TLR	StrpuTLR_41
Strongylocentrotus purpuratus	TLR	StrpuTLR_42
Strongylocentrotus purpuratus	TLR	StrpuTLR_43
Strongylocentrotus purpuratus	TLR	StrpuTLR_44
Strongylocentrotus purpuratus	TLR	StrpuTLR_45
Strongylocentrotus purpuratus	TLR	StrpuTLR_46
Strongylocentrotus purpuratus	TLR	StrpuTLR_47
Strongylocentrotus purpuratus	TLR	StrpuTLR_48
Strongylocentrotus purpuratus	TLR	StrpuTLR_49
Strongylocentrotus purpuratus	TLR	StrpuTLR_5
Strongylocentrotus purpuratus	TLR	StrpuTLR_50
Strongylocentrotus purpuratus	TLR	StrpuTLR_51
Strongylocentrotus purpuratus	TLR	StrpuTLR_52
Strongylocentrotus purpuratus	TLR	StrpuTLR_53
Strongylocentrotus purpuratus	TLR	StrpuTLR_54
Strongylocentrotus purpuratus	TLR	StrpuTLR_55
Strongylocentrotus purpuratus	TLR	StrpuTLR_56
Strongylocentrotus purpuratus	TLR	StrpuTLR_57
Strongylocentrotus purpuratus	TLR	StrpuTLR_58
Strongylocentrotus purpuratus	TLR	StrpuTLR_59
Strongylocentrotus purpuratus	TLR	StrpuTLR_6
Strongylocentrotus purpuratus	TLR	StrpuTLR_60
Strongylocentrotus purpuratus	TLR	StrpuTLR_61
Strongylocentrotus purpuratus	TLR	StrpuTLR_62
Strongylocentrotus purpuratus	TLR	StrpuTLR_63
Strongylocentrotus purpuratus	TLR	StrpuTLR_64
Strongylocentrotus purpuratus	TLR	StrpuTLR_65
Strongylocentrotus purpuratus	TLR	StrpuTLR_66
Strongylocentrotus purpuratus	TLR	StrpuTLR_67
Strongylocentrotus purpuratus	TLR	StrpuTLR_68
Strongylocentrotus purpuratus	TLR	StrpuTLR_69
Strongylocentrotus purpuratus	TLR	StrpuTLR_7
Strongylocentrotus purpuratus	TLR	StrpuTLR_70
Strongylocentrotus purpuratus	TLR	StrpuTLR_71
Strongylocentrotus purpuratus	TLR	StrpuTLR_72
Strongylocentrotus purpuratus	TLR	StrpuTLR_73
Strongylocentrotus purpuratus	TLR	StrpuTLR_74
Strongylocentrotus purpuratus	TLR	StrpuTLR_75
Strongylocentrotus purpuratus	TLR	StrpuTLR_76

Strongylocentrotus purpuratus	TLR	StrpuTLR_77
Strongylocentrotus purpuratus	TLR	StrpuTLR_78
Strongylocentrotus purpuratus	TLR	StrpuTLR_79
Strongylocentrotus purpuratus	TLR	StrpuTLR_8
Strongylocentrotus purpuratus	TLR	StrpuTLR_80
Strongylocentrotus purpuratus	TLR	StrpuTLR_81
Strongylocentrotus purpuratus	TLR	StrpuTLR_82
Strongylocentrotus purpuratus	TLR	StrpuTLR_83
Strongylocentrotus purpuratus	TLR	StrpuTLR_84
Strongylocentrotus purpuratus	TLR	StrpuTLR_85
Strongylocentrotus purpuratus	TLR	StrpuTLR_86
Strongylocentrotus purpuratus	TLR	StrpuTLR_87
Strongylocentrotus purpuratus	TLR	StrpuTLR_88
Strongylocentrotus purpuratus	TLR	StrpuTLR_89
Strongylocentrotus purpuratus	TLR	StrpuTLR_9
Strongylocentrotus purpuratus	TLR	StrpuTLR_90
Strongylocentrotus purpuratus	TLR	StrpuTLR_91
Strongylocentrotus purpuratus	TLR	StrpuTLR_92
Strongylocentrotus purpuratus	TLR	StrpuTLR_93
Strongylocentrotus purpuratus	TLR	StrpuTLR_94
Strongylocentrotus purpuratus	TLR	StrpuTLR_95
Strongylocentrotus purpuratus	TLR	StrpuTLR_96
Strongylocentrotus purpuratus	TLR	StrpuTLR_97
Strongylocentrotus purpuratus	TLR	StrpuTLR_98
Strongylocentrotus purpuratus	TLR	StrpuTLR_99
Strongylocentrotus purpuratus	TLR	StrpuTOLL_1
Strongylocentrotus purpuratus	TLR	StrpuTOLL_2
Strongylocentrotus purpuratus	TRAF6	StrpuTRAF6_1
Torquaratorid sp. Iceland	TAB1	TorIcTAB1_1
Torquaratorid sp. Iceland	TICAM2	TorIcTCAM2_1
Torquaratorid sp. Iceland	TICAM2	TorIcTCAM2_2
Torquaratorid sp. Norway	MAP3K7	TorNoM3K7_1
Torquaratorid sp. Norway	MYD88	TorNoMYD88_1
Torquaratorid sp. Norway	SARM1	TorNoSARM1_1
Torquaratorid sp. Norway	TAB1	TorNoTAB1_1
Torquaratorid sp. Norway	TAB1	TorNoTAB1_2
Torquaratorid sp. Norway	TAB2	TorNoTAB2_1
Torquaratorid sp. Norway	TICAM2	TorNoTCAM2_1
Torquaratorid sp. Norway	TICAM2	TorNoTCAM2_2
Torquaratorid sp. Norway	TLR	TorNoTLR_1
Torquaratorid sp. Norway	TLR	TorNoTLR_2
Torquaratorid sp. Norway	TLR	TorNoTLR_3
Torquaratorid sp. Norway	TRAF6	TorNoTRAF6_1

Footnotes: Data type key: T = Transcriptome, GM = Genome gene models, GS = Genome Assembly

Table S2: SwissProt proteins (47) and their associated domain architectures

SwissProt Accession	*Entry name*	*Associated Domain Architectures*
P51617	IRAK1_HUMAN	Death + Protein Kinase
Q62406	IRAK1_MOUSE	Death + Protein Kinase
Q9NWZ3	IRAK4_HUMAN	Death + Protein Kinase
Q8R4K2	IRAK4_MOUSE	Death + Protein Kinase
O43318	M3K7_HUMAN	Protein Kinase
Q62073	M3K7_MOUSE	Protein Kinase
Q99836	MYD88_HUMAN	TIR + Death
P22366	MYD88_MOUSE	TIR + Death
Q6SZW1	SARM1_HUMAN	TIR + SAM + SAM
Q6PDS3	SARM1_MOUSE	TIR + SAM + SAM
Q15750	TAB1_HUMAN	PPM-type Phosphatase
Q8CF89	TAB1_MOUSE	PPM-type Phosphatase
Q9NYJ8	TAB2_HUMAN	CUE + RanBP2-type Zinc-finger
Q99K90	TAB2_MOUSE	CUE + RanBP2-type Zinc-finger
Q8IUC6	TCAM1_HUMAN	TIR + RHIM
Q80UF7	TCAM1_MOUSE	TIR + RHIM
Q86XR7	TCAM2_HUMAN	TIR
Q8BJQ4	TCAM2_MOUSE	TIR
P58753	TIRAP_HUMAN	TIR
Q99JY1	TIRAP_MOUSE	TIR
Q15399	TLR1_HUMAN	TIR + (> 3 LRRs)
Q9EPQ1	TLR1_MOUSE	TIR + (> 3 LRRs)
O60603	TLR2_HUMAN	TIR + (> 3 LRRs)
Q9QUN7	TLR2_MOUSE	TIR + (> 3 LRRs)
O15455	TLR3_HUMAN	TIR + (> 3 LRRs)
Q99MB1	TLR3_MOUSE	TIR + (> 3 LRRs)
O00206	TLR4_HUMAN	TIR + (> 3 LRRs)
Q9QUK6	TLR4_MOUSE	TIR + (> 3 LRRs)
O60602	TLR5_HUMAN	TIR + (> 3 LRRs)
Q9JLF7	TLR5_MOUSE	TIR + (> 3 LRRs)
Q9Y2C9	TLR6_HUMAN	TIR + (> 3 LRRs)
Q9EPW9	TLR6_MOUSE	TIR + (> 3 LRRs)
Q9NYK1	TLR7_HUMAN	TIR + (> 3 LRRs)
P58681	TLR7_MOUSE	TIR + (> 3 LRRs)
Q9NR97	TLR8_HUMAN	TIR + (> 3 LRRs)
P58682	TLR8_MOUSE	TIR + (> 3 LRRs)
Q9NR96	TLR9_HUMAN	TIR + (> 3 LRRs)
Q9EQU3	TLR9_MOUSE	TIR + (> 3 LRRs)
Q9BXR5	TLR10_HUMAN	TIR + (> 3 LRRs)

Q6R5P0	TLR11_MOUSE	TIR + (> 3 LRRs)
Q6QNU9	TLR12_MOUSE	TIR + (> 3 LRRs)
Q6R5N8	TLR13_MOUSE	TIR + (> 3 LRRs)
P08953	TOLL_DROME	TIR + (> 3 LRRs)
Q9Y4K3	TRAF6_HUMAN	MATH + Ring-type Zinc-finger + Traf-type Zinc Finger
P70196	TRAF6_MOUSE	MATH + Ring-type Zinc-finger + Traf-type Zinc Finger

Table S3: Convergence statistics from Bayesian phylogenetic inference of deuterostome TLRs as computed by ExaBayes (version 1.4.2) (55).

	# Gen.	*Statistic**	*Value*	*Min*	*Max*	*Mean*	*Median*
Full TLRs	10.0E6	*Parameter ESS*		204	7360	1390	1280
		Parameter PSRF		1.00	1.00	1.00	1.00
		ASDSF	0.60%				
		MSDSF	4.42%				
TIR-Only	19.6E6	*Parameter ESS*		135	4030	1190	1080
		Parameter PSRF		1.00	1.01	1.00	1.00
		ASDSF	1.11%				
		MSDSF	8.70%				

*ESS = Effective sample size, PSRF = Potential scale reduction factor, ASDSF = Average standard deviation of split frequencies, MSDSF = Maximum standard deviation of split frequencies.

Chapter 4: Induced Immune Reaction in the Acorn Worm, *Saccoglossus kowalevskii,* has Implications for Understanding the Evolution of Antiviral Immunity

4.1 Abstract

Comparative genomics and evolutionary analyses leveraging species have provided strong evidence that the molecular toolkit required for a robust immune response was present in the last common ancestor to Deuterostomes – a superphylum of animals comprised of hemichordates, echinoderms, and chordates. Although these genetic surveys provide essential groundwork for inferring the evolutionary history of immune gene families within the superphylum, there are very few studies which experimentally isolate genes directly involved in invertebrate deuterostome immunity. In this study, we perform a differential gene expression analysis on the hemichordate deuterostome, *Saccoglossus kowalevskii*, to identify genes directly implicated in the immune response against viral dsRNA. By comparing individuals injected with poly(I:C), a synthetic dsRNA analog and potent ligand of several pattern recognition receptors, to those injected with a buffer, we identify 455 genes which possess statistically significant changes (Wald's adjusted p-values < 0.05) in their mean expression following dsRNA exposure. Moreover, by performing these experiments in a time-dependent schema, we show that expressional variation for 368 genes can be explained by an interaction between time and treatment (LRT adjusted p-values < 0.05) – suggesting viral exposure elicits a dynamic transcriptomic response in *S. kowalevskii*. Although some differentially expressed genes possessed annotations congruent with canonical immune reactions known in vertebrates, many genes identified in *S. kowalevskii* could not be annotated and remain uncharacterized. Together, the transcriptional response and DEGs described in *S. kowalevskii* represent a substantial advance in establishing a phylogenetically informative and comparative framework necessary for elucidating the evolution of deuterostome immune systems.

4.2 Significance Statement

Understanding the conserved genetic elements underlying molecular immunity is necessary to inform the evolutionary history which preceded the adaptive immune systems of vertebrates (**Figure 1A**). By investigating the transcriptomic response to viral stimulus in the invertebrate deuterostome, *Saccoglossus kowalevskii*, we provide vital phylogenetic contrast and

insight into the molecular mechanisms behind viral immunity. Despite possessing a similar innate immunity genetic repertoire to vertebrates, *S. kowalevskii* leverages many genes during its antiviral immune response which could not be annotated via database-dependent methods. Congruent with similarly structured studies in echinoderms, our findings suggest that vertebrates may have lost genetic complexity with regard to their innate immunity system following their divergence from the last common ancestor of deuterostomes.

4.3 Introduction

Evolutionary analyses of -omic scale data have shown that the molecular machinery responsible for pathogen recognition and transcriptional regulation of immunity effectors is broadly conserved across Deuterostomia (**Figure 1A**; see **Chapters 3 & 5**), implicating an already established and robust genetic toolkit sufficient for rapid host-defense in the LCA of deuterostomes. Although comparative molecular evolution studies provide invaluable evidence on the ancestry of immune gene/protein diversity, only a few invertebrate deuterostomes have been the subject of functional studies that characterize their genetic toolkit involved in immunity (Guo & Li 2021). Lack of taxonomic representation not only constrains inferences concerning ancestral vertebrate immunity but may also misrepresent the diversity of immune systems leveraged across deuterostomes. These effects become further exacerbated when considering the evolution and diversity of specialized immune responses such as antiviral immunity.

Experiments leveraging RNA-seq for differential gene expression (DGE) analyses have provided valuable insight into the deployment of antiviral immune factors in echinoderms (Fuess et al. 2015; Ruiz-Ramos et al. 2020; Wu et al. 2020), the clade of invertebrate deuterostomes to have afforded the most attention with respect to viral disease ecology and immunology (Guo & Li 2021, Hewson et al. 2014, Miner et al. 2018). DGE studies provide direct evidence identifying genes involved in antiviral immune reactions and can be placed in an evolutionary context to inform the ancestry of deuterostome immunity. Viruses have been shown to be the infectious agent behind sea star wasting disease, a particularly nefarious condition which has led to mass mortality events of sea stars across the Eastern Pacific (including the keystone species, *Pisaster ochraceous* (Schiebelhut et al. 2018), and several fatal diseases in the sea cucumber, *Apostichopus japonicus* (a Chinese aquaculture species with an economic value reportedly exceeding 5 billion USD in

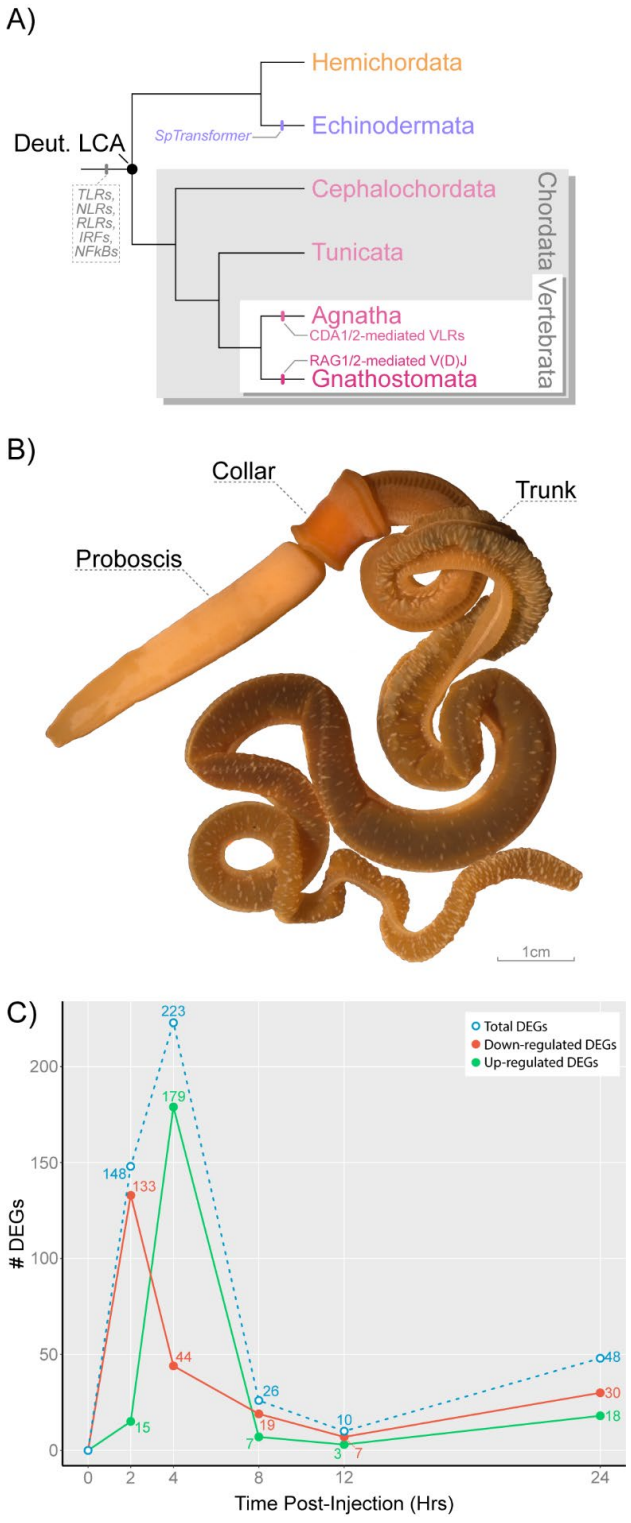


Figure 1: A) Deuterostome phylogeny annotated for the evolution of key immunity pathways. Deut. = Deuterostome. B) Morphology of the hemichordate acorn worm, *Saccoglossus kowalevskii*. C) Plot of the number differentially expressed genes (DEGs) in treatment *S. kowalevskii* individuals per timepoint relative to control conditions (Wald's adj. $p < 0.05$).

2012; Zhang et al. 2015, Guo & Li 2021). As such, genetic investigations into echinoderm immunity provide context not only for deuterostome immunity evolution, but also as important groundwork for future conservation efforts. Strikingly, DGE analyses have revealed that echinoderm immunity invokes many genes which cannot be sufficiently annotated – with only 1183 (31.7% of total) differentially expressed genes (DEGs) in *P. ochraceous* receiving high-confidence SwissProt annotations (Fuess et al. 2015), and 1180 (67.5% of total) DEGs in *Holothuria leucospilota* able to be annotated with at least a single GO (gene ontology) term (Wu et al. 2020). These findings highlight the importance of taxonomic representation in comparative immunology and genomic studies (David et al. 2019), as further phylogenetic comparisons are necessary to determine whether annotation inefficiency among echinoderm antiviral genes is due to novelty (and/or sequence divergence) or, instead, representative of an evolutionarily ancient antiviral toolkit which has been lost, or at least undocumented, in vertebrates.

As sister phylum to the echinoderms (Halanych 1995, Cannon et al. 2014, Laumer et al. 2019), hemichordates play an instrumental role in elucidating the ancestry of deuterostome traits, particularly in the field of evolutionary developmental biology (Lowe 2021). However, very little is known about hemichordate immunity. Hemichordate immunity received some attention in the 1980s when the acorn worm hemichordate, *Saccoglossus horsti,* was shown to process phagocytes within the collar and trunk coelomic cavities (**Figure 1B**) that are responsive to *in vitro* bacterial infection (Rhodes & Ratcliffe 1983), and whole-body homogenate of the congeneric species, *Saccoglossus ruber*, exhibited strong antibacterial properties (Millar & Ratcliffe 1983). More recently, comparative genomics/transcriptomics have revealed that hemichordates inherited a molecular toolkit containing several major innate immunity protein families, including Toll-like receptors (TLRs) and their canonical pathway (Tassia et al. 2017), NOD-like receptors (NLRs), RIG-I-like receptors (RLRs), interferon regulatory factors (IRFs), and nuclear factor κB (NF-κB) family members (**Chapters 3 & 5**). Phylogenetic reconstruction of deuterostome TLRs has also provided strong support for hemichordates possessing a TLR3 ortholog (Tassia et al. 2017). Interestingly, type-I interferons, a group of secreted signaling molecules central to antiviral immunity (i.e., antiviral cytokines), are hypothesized to be a vertebrate novelty, as no clear homologs have been identified in hemichordates or any other invertebrate phylum (Majzoub et al. 2019). Considering the molecular components canonically upstream of interferon transcription are conserved in hemichordates (i.e., TLR3, RLRs, and IRFs), the signaling molecules involved in

hemichordate antiviral immunity remains an open question. Nonetheless, the evolutionary conservation of TLR3, RLRs, and IRFs within Hemichordata fortifies the hypothesis that extant deuterostomes possess a shared antiviral immunity toolkit.

Here, we describe an antiviral transcriptional immune response in the acorn worm, *Saccoglossus kowalevskii* (**Figure 1B**), and place our findings in a comparative and evolutionary framework. This experiment was designed to identify genes involved in *S. kowalevskii*'s defense against RNA viruses and provide insight into the time-dependent transcriptional effects following viral stimulus. To accomplish these objectives, we simulated an acute viral infection within the proboscis cavity (**Figure 1B**) by injecting *S. kowalevskii* individuals with the synthetic viral double-stranded RNA (dsRNA) analog, poly(I:C), a potent ligand of TLR3 and RLRs known to elicit an antiviral immune response within vertebrates (Kaur et al. 2019). Transcriptomes of individuals injected with poly(I:C) were compared against those from individuals which received a control injection containing only buffer (see **Materials & Methods** below). For each experimental replicate (n = 4), worms were sacrificed at 0, 2, 4, 8, 12, and 24hrs post-injection (hpi) and individual sequencing libraries were independently constructed for three body regions (i.e., proboscis, collar, and trunk) from each worm. This design allowed us to control for tissue-specific effects (**Figure S1**) and instead focus on systemic transcriptional effects in response to viral infection. In total, 144 transcriptomes were generated for our study (4 experimental replicates, 2 treatment conditions per replicate, 6 timepoints per condition, 3 tissues per worm; **Table S1**).

4.4 Results

S. kowalevskii possesses an immune system capable of recognizing and transcriptionally reacting to dsRNA (**Figure 1C, Figure 2**). For each timepoint comparison, the number of DEGs reported and discussed henceforth are those which possessed statistically significant differences in mean expression between treatment individuals relative to control conditions at 0hpi (Wald's adjusted p-value < 0.05 and FDR cutoff = 0.05). Consolidated across all time points, 455 genes were differentially expressed due to poly(I:C) exposure (**Table 1**). The largest quantifiable transcriptional effects occurred at 2hpi (148 DEGs) and 4hpi (223 DEGs), with gene expression primarily down-regulated at 2hpi (133/148 DEGs) and up-regulated at 4hpi (178/223 DEGs). At timepoints following 4hpi, transcriptional effects were comparatively more subtle (**Figure 1C**). A

Figure 2: Annotation overview of DEGs per timepoint relative to control conditions (Wald's adj. $p < 0.05$) with focus on genes down-regulated 2hpi and up-regulated 4hpi. Circles depict Log2FC for a single gene in treatment individuals relative to control conditions at 0hpi, and bars extending from each circle represent Log2FC standard error. Genes are colored if their Log2FC > 1 (green) or Log2FC < 1 (red). PANTHER pathway annotations are shown DEGs up-regulated 4 hours post-injection (top) and down-regulated 2 hours post-injection (bottom).

Table 1: Summary statistics of treatment vs. control DEGs

	Differentially Expressed Genes (Wald's adjusted p < 0.05)			# DEGs with SwissProt IDs (% of DEG Category with SwissProt IDs)		
Timepoint (vs. 0hrs)	Total DEGs	Up-reg. DEGs	Down-reg. DEGs	Total Annotated	Up-reg. DEGs Annotated	Down-reg. DEGs Annotated
2hrs	148	15	133	91 (61%)	12 (80%)	79 (59%)
4hrs	223	179	44	145 (81%)	108 (60%)	37 (84%)
8hrs	26	7	19	18 (69%)	4 (57%)	14 (74%)
12hrs	10	3	7	8 (80%)	3 (100%)	5 (71%)
24hrs	48	18	30	21 (44%)	5 (28%)	16 (53%)
Total	455	222	233	283 (62%)	132 (59%)	151 (65%)

total of 117 DEGs at 2hpi, 88 DEGs at 4hpi, 19 DEGs at 8hpi, 9 DEGs at 12hpi, and 45 DEGs at 24hpi exhibit at least a 2-fold change in expression. A likelihood ratio test (LRT adjusted p-value < 0.05 and FDR cutoff = 0.05) showed that a significant amount of the count variation for 368 genes was explained by the interaction between condition and time.

To provide functional context to the transcriptional effects observed at each timepoint, DEGs were assigned function via the PANTHER classification system (Thomas et al. 2003, Mi et al. 2010, Mi et al. 2019). Because PANTHER's functional classification relies on the assumption that orthologous proteins likely share similar functions (i.e., the ortholog conjecture; David et al. 2020), DEGs at each timepoint were first assigned identifications according to their best-hit sequence within the SwissProt database, a subset of the UniProt Knowledgebase which contains protein sequences of known function (Uniprot Consortium 2021). Notably, only 283 DEGs (62.2% of all DEGs) could be confidently assigned a SwissProt ID by sequence similarity (DIAMOND blastx e-value > 0.001) (**Table 1**). Given the known roles of noncoding RNAs (ncRNAs) during mammalian immunity regulation (Turner et al. 2014), we additionally attempted to identify ncRNAs among DEGs without SwissProt IDs. Using the RNA search tool Infernal (Nawrocki & Eddy 2013), only seven genes across all timepoints could be partially annotated as ncRNAs from the Rfam database (Kalvari et al. 2021).

S. kowalevskii's transcriptional immune response contains many factors acting in pathways unrecognized from vertebrates (**Figures S2–5)**. Only 46 DEGs (10.1% of all DEGs and 16.3% of those with assigned SwissProt IDs) could be correlated with PANTHER pathway data, and 106 DEGs (23.3% of all DEGs and 37.5% of those with assigned SwissProt IDs) could be correlated

with a PANTHER protein class. Among SwissProt IDs which possess PANTHER pathway annotations (**Table S2**), several notable immunity-associated pathways were present, including those associated with chemokine/cytokine-associated inflammatory responses (2- and 4hpi), apoptosis (2-, 4-, and 24hpi), interferon-gamma signaling (2hpi), and interleukin signaling (2- and 4hpi). DEGs down-regulated at 2hpi and up-regulated at 4hpi were associated with greatest number of pathways (21 and 18 pathways, respectively) (**Figure 2**). In contrast, none of the DEGs at 12hpi could be associated with any of the 177 protein pathways within PANTHER.

Protein class annotations were also greatest in the pool of down-regulated DEGs at 2hpi and up-regulated DEGs at 4hpi (16 and 32 protein classes, respectively). Within these groups, protein classes could be confidently assigned to 58 DEGs at 2hpi (39.1% of all DEGs at 2hpi and 63.7% of those with assigned SwissProt IDs at 2hpi) and 81 DEGs at 4hpi (36.3% of all DEGs at 4hpi and 55.9% of those assigned SwissProt IDs at 4hpi) (**Table S2**). Notably, 47.9% of all DEGs assigned protein classes, among those downregulated at 2hpi, (**Table S2, Figures S4 & S5)** were extracellular matrix proteins, suggesting a concerted change in cellular/tissue structure(s) in response to viral presence. In contrast, the greatest single proportion of protein class annotations up-regulated at 4hpi (i.e., ATP-binding cassette transporters) comprised only 10.3% of the protein classes annotated, suggesting the genes up-regulated at 4hpi confer a diverse array of biochemical roles.

Gene Ontology (GO) term (Ashburner et al. 2000, Gene Ontology Consortium 2020) annotation followed a similar trend as PANTHER classification (**Table S3**), with only 106 DEGs assigned molecular function (23.3% of all DEGs and 37.5% of those assigned SwissProt IDs), 116 assigned biological processes (26.1% of all DEGs and 41.0% of those assigned SwissProt IDs), and 140 assigned cellular components (30.8% of all DEGs and 49.5% of those assigned SwissProt IDs). Akin to PANTHER annotations, the greatest number of GO terms were assigned to 4hpi and 2hpi (**Table S3**).

Translated DEGs were also assigned KEGG Ontologies (KO; Aramaki et al. 2020) and Pfam domain annotation (Mistry et al. 2020). Unlike PANTHER which relies on sequence similarity metrics, KOs and Pfam domain annotations relied on databases of hidden markov models to infer homology between query sequences and database entries. Whereas 142 translated

DEGs could be assigned KOs (31.2% of all DEGs), 371 proteins could be annotated with at least one Pfam domain (81.5% of all DEGs).

4.5 Discussion

The acorn worm, *Saccoglossus kowalevskii* (**Figure 2**) showed a marked transcriptional response to injection of poly(I:C), a synthetic analog for viral dsRNA. Many of the DEGs observed in this response are orthologs to known viral immunity response genes in vertebrates (Tables S2 - 5) but numerous genes could not be annotated, perhaps indicating unrecognized mechanisms used during viral immunity among deuterostomes. Consistent with transcriptome-based differential gene expression studies in echinoderms (Fuess et al. 2015, Wu et al. 2020), *S. kowalevskii'*s immune response to dsRNA involves many genes which cannot be identified via translated similarity to SwissProt proteins (n = 172; 37.8% of all DEGs) or KEGG ontologies (n = 142; 31.2% of all DEGs), substantially hindering broad functional inference for these genes. Whereas functional inference of the full protein may be difficult, broad annotation of proteins with Pfam domains (n = 371; 81.5% of all DEGs) does provide some predictive strength. Very few (n = 7) could be assigned even partial homology to known ncRNA families within Rfam, an important consideration that is necessitated by previous data showing ncRNAs play a role in mammalian immunity (Turner et al. 2014).

By directly accounting for tissue-specific transcriptional effects in our experimental design, we are able to identify DEGs that represent *S. kowalevskii*'s systemic response to acute viral infection. This distinction is important, as discrete hematopoietic and immunocyte sources are yet to be identified among hemichordates (**Chapter 2**). Given work in urchin showing that immune cell recruitment following infection is similar to that known from vertebrates (Ho et al. 2016), careful consideration was taken during the design of our experiment to account for immunity-dependent transcriptional effects which may occur in body territories distant from the injection site. Furthermore, nearly 90% of expressional variance across sequence libraries generated in this study can be attributed to tissue-specific transcriptomic differences – suggesting a high degree of tissue specialization between proboscis, collar, and trunk.

In vertebrates, dsRNA viruses recognized by dendritic cells and macrophages promote the activation of NF-κB and IRFs, transcription factors responsible for the induction of pro-inflammatory and/or antiviral cytokine expression (Mogensen & Paludan 2001). Whereas these transcription factors are known to be present across Metazoa (**Chapter 5**), their targets and the hypothetical cytokine/chemokine complement outside of mammals is poorly understood. Although the ideal experiment to determine the direct targets of NF-κB or IRF activation would utilize targeted expression-vectors, our study provides some insight into the downstream effectors of immune system activation (and potentially NF-kB and/or IRF activation). Importantly, functional annotation at each timepoint indicates *S. kowalevskii*'s defense against dsRNA viruses involves immune pathways that act in tightly regulated circuits where, in some cases, the same pathway may be simultaneously up- and down-regulated (e.g., apoptosis signaling is both up- and down-regulated at 2hpi) or consecutively regulated in opposing directions between timepoints (e.g., inflammation mediated by cytokine/chemokine signaling down-regulated at 2hpi, and up-regulated at 4hpi) (**Figures S2 & S3**). In mammals, such effects are attributed to the avoidance of immunity overactivation, where inflammatory or apoptotic pathways can be re-routed towards deleterious conditions such as severe autoimmunity disorders (Blach-Olszewska & Leszek 2007). Further investigation into the regulatory machinery, and direct experimental evidence connecting transcription factor to their target genes, will be invaluable for determining the evolutionary relationship between regulatory effects in *S. kowalevskii* and other deuterostomes.

In our study, we provide evidence that *S. kowalevskii* can recognize and respond to dsRNA by transcriptionally regulating genes homologous to those from biomedical model species that have prescribed canonical roles in immunity (**Figure 2, S2-5**); furthermore, these findings are congruent with findings among echinoderms (Fuess et al. 2015, Wu et al. 2020). Strikingly, however, we also identified many genes involved in *S. kowalevskii*'s antiviral defense which could not be sufficiently annotated (**Tables S2-4**). Given comparable lack of annotation in echinoderms, uncharacterized immune genes identified in *S. kowalevskii* can be hypothesized to represent any combination of the following evolutionary scenarios: A) They belong to a plesiomorphic deuterostome antiviral immunity toolkit which was subsequently lost, diverged, or coopted for other functions in vertebrates, B) The genes represent lineage-specific novelties in the clade comprised of hemichordates and echinoderms, the Ambulacraria (**Figure 1A**), or C) These genes

emerged after hemichordates and echinoderms diverged. Further phylogenetic analyses will be necessary to discern among these scenarios.

Given the adaptive benefits that have been attributed the acquisition of secondary immune systems (Müller et al. 2018), such as those presented by jawed-vertebrate immunoglobulins and jawless-vertebrate variable lymphocyte receptors (Flajnik 2018), characterizing the evolutionarily conserved elements of immunity is vital for understanding the conditions on which adaptive immunities evolved. The work presented here fills a significant gap in knowledge necessary for making phylogenetically informed hypotheses/conclusions on the ancestry of deuterostome antiviral immunity.

4.6 Materials & Methods

4.6.1 Animal Handling

Worms were obtained intertidally from Waquoit Bay, MA in September 2018 and 2019. Detailed information on *S. kowalevskii* collection and general upkeep can be found in Lowe et al. 2004. All worms were female and had spawned at least 4 days prior to treatment. Immediately prior to injection, individual worms were placed into a single well of a 6-well plate containing filtered sea water (see Lowe et al. 2004), and each plate was placed onto ice to relax worms prior to injection. During injection, individuals were temporarily submerged into a 3.75% solution of magnesium chloride (dissolved in filtered sea water) to inhibit muscle contraction. Following injection, worms were replaced to their 6-well plates and placed into a flow-through sea water table until their incubation time was complete.

4.6.2 Experimental Design

Two experimental replicates were performed in both September 2018 and September 2019. Individual worms were injected with either A) an agonist solution containing 10 μg/mL high- and low-molecular weight Poly(I:C) (tlrl-pic and tlrl-picw, respectively; InvivoGen) dissolved in an injection buffer containing 1X phosphate-buffered saline (10mM sodium phosphate, 0.15mM sodium chloride) and 5mg/mL Calceine fluorescent dye, or B) only injection buffer. Six timepoints

were used for our study: 0, 2, 4, 8, 12, and 24 hours post-injection. After incubation, individuals were dissected into proboscis, collar, and trunk before being flash-frozen in liquid nitrogen and stored at -80°C. For each experimental replicate, 36 samples were generated (2 conditions, 6 timepoints, and 3 body regions). In total, 144 samples were collected and prepared for sequencing (**Table S1**).

Needles were pulled from Sutter Instruments Co.'s thin wall borosilicate glass with filament (BF100-78-10) on a Sutter Instruments Co.'s Model P-97 (parameters: P = 200, Heat = 520, Pull = 60, Vel = 100, Time = 175). Needles were loaded with 1µL of injection solution (see above) and the injection was propelled by nitrogen gas using a MPPI-3 Pressure Injector (Applied Scientific Instrumentation, Inc). Worms were injected into the posterior proboscis directly anterior to the proboscis neck.

4.6.3 Library Preparation and Sequencing

RNA extractions were performed using the Qiagen RNeasy Mini Kit (Qiagen) and on-column RNase-Free DNase Set (Qiagen). RNA concentration and purity were assessed on a TapeStation 2200 (Agilent Technologies, Inc.). Sequencing libraries were constructed with the QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen) in conjunction with the i5 6 nt Dual Indexing Add-on Kit (Lexogen). Optimum amplification cycles per library was assessed using the PCR Add-on Kit for Illumina (Lexogen). Libraries were sent to the Hudson Alpha Institute for Biotechnology (Huntsvillle, AL, USA) where they were sequenced on a NovaSeq (single-end, 100bp; Illumina, Inc.). Sequencing statistics are reported in **Table S1**.

4.6.4 Sequence Pre-processing

All 144 sequence read libraries were cleaned using *fastp* (version 0.19.7; Chen et al. 2018) to remove adapter sequences (AGATCGGAAGAGCACACGTCTGAACTCCAGTCA), 3' polyX tails, and perform per-base quality trimming using a "cut right" sliding window with default parameters.

Each library was mapped to the *Saccoglossus kowalevskii* reference genome (version Skow_1.1; downloaded from the NCBI genome repository in December 2019) using STAR (version 2.7.0; Dobin et al. 2013). STAR parameters follow Lexogen's recommendations for the QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina; mapping statistics can be seen in **Table S1**. Following read alignment, count matrices were obtained using HTSeq-count (version 0.11.2; Anders et al. 2015) with "intersection-nonempty" defined for handling reads which overlap multiple features, strandedness set to sense orientation, and both feature type and ID attribute set to "gene" for quantification. Read map quantification was not normalized prior to differential gene expression analyses.

4.6.5 Differential Gene Expression Analysis

DESeq2 (version 1.30.1; Love et al. 2014) was used to perform differential gene expression analyses with the following design formula:

$$\sim Condition + Time\ Post\ Injection\ (TPI) + Tissue + Condition{:}TPI$$

For Wald Test's, Log2 fold changes (LFCs) per gene between treatment and control groups at each timepoint were estimated using a linear contrast of *Condition+Condition:TPI* where the reference levels were control and 0hrs post-injection for *Condition* and *TPI*, respectively, and a false discovery rate (adjusted p) cutoff of 0.05. All design variables were treated as categorial factors. The *TPI* and *Tissue* factors were included in the design formula as a measure to explain variation caused by tissue-specific gene expression (**Figure S1**) and time-dependent effects that can be isolated from the treatment contrast (such as wound healing or biorhythms). Importantly, the overall experimental design of our study controls for several biological factors (i.e., sex and sexual maturity) that could otherwise influence the number of genes (and subsequent conclusions) associated with *S. kowalevskii*'s antiviral immune response. To account for low mean-count genes, LFC shrinkage was performed using the ashr package (Stephens 2016) in DESeq2. Additionally, a Likelihood Ratio Test was also performed using the following reduced formula:

$$\sim Condition + Time\ Post\ Injection\ (TPI) + Tissue$$

Data were visualized using ggplot2 (version 3.3.3; Wickham 2016). DESeq2 and ggplot2 were conducted in RStudio (version 1.2.5001; RStudio Team 2020) (version 4.0.4; R Core Team 2021).

4.6.6 Functional Annotation

PANTHER (Thomas et al. 2003) and GO-term (Ashburner et al. 2000, Gene Ontology Consortium 2020) classification per differentially expressed gene (DEG) was performed by first assigning each DEG its single best-hit SwissProt identification (Accessed July 2021; The Uniprot Consortium 2021) using DIAMOND's translated search mode (Buchfink 2021), where best hit was defined as the subject with the highest bit score, lowest expect value, and highest percent of identical matches. SwissProt IDs were then input to the PANTHER (database version 16.0; Mi et al. 2019) Gene List Analysis tool (accessed June 2021; Mi et al. 2010) to obtain PANTHER pathway, PANTHER protein class, and GO-term data using all organisms as reference. Rfam (version 14.6; Kalvari et al. 2021) annotation was performed using Infernal's *cmscan* command with default parameters (version 1.1; Nawrocki & Eddy 2013). Translated DEGs (using Transdecoder version 3.0.1; Haas et al. 2013) were assigned with KEGG orthologies with the KofamKOALA web server using an E-value threshold of 0.01 (release version 99.0, accessed July 2021; Aramaki et al. 2020), and Pfam (Mistry et al. 2020) annotation was performed with HMMER's *hmmscan* using default parameters (Eddy 2009) and filtered for best-fit domains using the *Best_fit_domains.py* tool from TIAMMAt (https://github.com/mtassia/TIAMMAt; see **Chapter 5**).

4.7 Acknowledgments

This work was supported by The National Science Foundation (grant number IOS – 1755377 to KMH, Rita Graze and Elizabeth Hiltbold Schwartz). Computational resources were

made available by the Alabama Supercomputer Authority and the Auburn University Hopper Cluster.

4.8 References

S. Anders, P.T. Pyl, W. Huber, HTSeq – a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166-169 (2015).

T. Aramaki, *et al.,* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251-2252 (2020)

M. Ashburner, *et al.,* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25-29 (2000)

Z. Blach-Olszewska, J. Leszek, Mechanisms of over-activated innate immune system regulation in autoimmune and neurodegenerative disorders. *Neuropsychiatr. Dis. Treat.* **3**: 365-372 (2007)

B. Buchfink, K. Reuter, H.G. Drost, Sensitive protein alignments at tree-of-lie scale using DIAMOND. *Nat. Methods* **18**: 366-368.

J.T. Cannon, *et al.*, Phylogenomic resolution of the hemichordate and echinoderm clade. *Curr. Biol.* **24**, 2827-2832 (2014)

S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34,** i884-i890 (2018).

K.T. David, A.E. Wilson, K.M. Halanych, Sequencing disparity in the genomic era. *Mol. Biol. Evol.* **36**, 1624-1627 (2019).

K.T. David, J.R. Oaks, K.M. Halanych, Patterns of gene evolution following duplications and speciations in vertebrates. *PeerJ* **8**: e8813 (2020)

A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21 (2013).

S.R. Eddy, A new generation of homology search tools based on probabilistic inference. *Genome Informatics* **23**, 205-211 (2009)

M.F. Flajnik, A cold-blooded view of adaptive immunity. *Nat. Rev. Immunol.* **18**, 438-453 (2018)

L.E. Fuess, *et al.,* Up in Arms: Immune and Nervous System Response to Sea Star Wasting Disease. *PLoS ONE* **10** (2015).

The Gene Ontology Consortium, The Gene Ontology resource: enriching a Gold mine. *Nucleic Acids Res.* **49**, D325-334 (2020)

M. Guo, C. Li, Current progress on identification of virus pathogens and the antiviral effectors in echinoderms. *Dev. Comp. Immunol.* **116**, 103912 (2021).

B.J. Haas, *et al.,* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494-1512 (2013)

K.M. Halanych, The phylogenetic position of the pterobranch hemichordates based on 18S rDNA sequence data. *Mol. Phylogenet. Evol.* **4**, 72-76 (1995).

I. Hewson, *et al.,* Densovirus associated with sea-star wasting disease and mass mortality. *Proc. Nat. Acad. Sci.* **111**, 17278-17283 (2014).

E.C.H. Ho, *et al.* Perturbation of gut bacteria induces a coordinated cellular immune response in the purple sea urchin larva. *Immunol. Cell Biol.* **94**: 861-874 (2016)

I. Kalvari, *et al.*, Rfam 14: expanded coverage of metagenomic, viral, and microRNA families. *Nucleic Acids Research* **49**, D192-D200 (2021).

D. Kaur, *et al.,* PRRDB 2.0: a comprehensive database of pattern-recognition receptors and their ligands. *Database* **2019** (2019).

C.E. Laumer, *et al.,* Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proc. R. Soc. B* **286** (2019).

M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15:** 550 (2014)

C.J. Lowe, K. Tagawa, T. Humphreys T., M. Kirschner, J. Gerhart, Hemichordate embryos: procurement, culture, and basic methods. *Methods in cell biology* **74**, 171-194 (2004).

C.J. Lowe, Molecular insights into deuterostome evolution from hemichordate developmental biology. *Curr. Top. Dev. Biol.* **141**, 75-117 (2021).

K. Majzoub, F. Wrensch, T.F. Maumert, The innate antiviral response in animals: an evolutionary perspective from flagellates to humans. *Viruses* **11** (2019)

H. Mi *et al.,* PANTHER version 7: improved phylogenetic trees, orthologs, and collaboration with the Gene Ontology Consortium. *Nucl. Acids Res.* **38**: D204-D210. (2010)

H. Mi et al. Protocol Update for Large-scale genome and gene function analysis with PANTHER Classification System (v.14.0). *Nat Protoc.* **14**: 703-721 (2019).

D.A. Millar, N.A. Ratcliffe, The Antibacterial Activity of the Hemichordate *Saccoglossus ruber* (Enteropneusta). *J. Invertebrate Pathol.* **50**, 191-200 (1987).

C.M. Miner, *et al.,* Large-scale impacts of sea star wasting disease (SSWD) on intertidal sea stars and implications for recovery. *PLoS ONE* **13** (2018).

J. Mistry, *et al.,* Pfam: The proteins families database in 2021. *Nucleic Acids Res.* **49**, D412-419 (2020)

T.H. Mogensen, S.R. Paludan, Molecular pathways in virus-induced cytokine production. *Microbiol. Mol. Biol. Rev.* **65**: 131-150 (2001)

V. Müller, R.J. de Boer, S. Bonhoeffer, E. Szathmáry, An evolutionary perspective on the systems of adaptive immunity. *Bio. Rev.* **93**: 505-528 (2018)

E.P. Nawrocki, S.R. Eddy, Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933-2935 (2013).

C.P. Rhodes, N.A. Ratcliffe, Coelomocytes and defense reactions of the primitive chordates, *Branchiostoma lanceolatum* and *Saccoglossus horsti*. *Dev. Comp. Immunol.* **7**: 695-698 (1983).

R Core Team, R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org/ (2021)

RStudio Team, RStudio: Integrated Development for R. *RStudio, PBC, Boston, MA*, http://www.rstudio.com/ (2020)

D.V. Ruiz-Ramos, L.M. Schiebelhut, K.J. Hutt, J.P. Wares, M.N. Dawson, An initial comparative genomic autopsy of wasting disease in sea stars. *Mol. Ecol.* **29**: 1087-1102 (2020).

L.M. Schiebelhut, J.B. Puritz, M.N. Dawson, Decimation by sea star wasting disease and rapid genetic changes in a keystone species, *Pisaster ochraceous*. *Proc. Nat. Acad. Sci.* **115**, 7069-7074 (2018).

M. Stephens, False discovery rates: a new deal. *Biostatistics* **18**: 275-294 (2017).

M.G. Tassia, N.V. Whelan, K.M. Halanych, Toll-like receptor pathway evolution in deuterostomes. *Proc. Nat. Acad. Sci.* **114**, 7055-7060 (2017)

P.D. Thomas, *et al.*, PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13**: 2129-2141. (2003)

M. Turner, A. Galloway, E. Vigorito, Noncoding RNA and its associated proteins as regulatory elements of the immune system. *Nat. Immunol.* **15**, 484-481 (2014).

The UniProt Consortium, UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**: D1 (2021).

H. Wickham, ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York* (2016).

X. Wu, *et al.*, Transcriptomic analysis of sea cucumber (*Holothuria leucospilota*) coelomocytes revealed the echinoderm cytokine response during immune challenge. *BMC Genomics* **21** (2020).

L. Zhang, X. Song, J. Hamel, A. Mercier, "Chapter 16 – Aquaculture, Stock Enhancement, and Restocking" in Developments in Aquaculture and Fisheries Science, H. Yang, J Hamel, A. Mercier, Eds. (Elsevier, 2015), pp. 289-322.

4.9 Supporting Information

Table S1: Sample metadata, sequencing statistics, and mapping rates

	Metadata				Sequencing Reads			Mapping	
Sample	Rep	Condition	Tissue	TPI	Raw	Cleaned	% Ret*	Unique	Multi
2C0C	2	Control	Collar	0	6513256	6167465	94.69%	4370017 (70.86%)	1640080 (26.59%)
2C0P	2	Control	Proboscis	0	6619666	6369347	96.22%	5952043 (93.45%)	262007 (4.11%)
2C0T	2	Control	Trunk	0	5337846	5206737	97.54%	4787075 (91.94%)	223568 (4.29%)
2T0C	2	Treatment	Collar	0	8259324	7954854	96.31%	6655264 (83.66%)	1021746 (12.84%)
2T0P	2	Treatment	Proboscis	0	5586058	5420449	97.04%	5108482 (94.24%)	191129 (3.53%)
2T0T	2	Treatment	Trunk	0	6619882	6329306	95.61%	5482857 (86.63%)	554899 (8.77%)
2C2C	2	Control	Collar	2	6408972	6081601	94.89%	5220510 (85.84%)	668559 (10.99%)
2C2P	2	Control	Proboscis	2	9861000	9147991	92.77%	7418897 (81.10%)	1504439 (16.45%)
2C2T	2	Control	Trunk	2	5850429	5586214	95.48%	4651997 (83.28%)	774876 (13.87%)
2T2C	2	Treatment	Collar	2	6154648	5869524	95.37%	3495613 (59.56%)	2213756 (37.72%)
2T2P	2	Treatment	Proboscis	2	8358600	8114891	97.08%	7467084 (92.02%)	479738 (5.91%)
2T2T	2	Treatment	Trunk	2	4743187	4593878	96.85%	4151569 (90.37%)	230594 (5.02%)
2C4C	2	Control	Collar	4	7776629	7514611	96.63%	6190966 (82.39%)	1127023 (15.00%)
2C4P	2	Control	Proboscis	4	7958525	7770086	97.63%	7077782 (91.09%)	530606 (6.83%)
2C4T	2	Control	Trunk	4	7856773	7721049	98.27%	7168792 (92.85%)	268139 (3.47%)
2T4C	2	Treatment	Collar	4	6575393	6293802	95.72%	5221827 (82.97%)	848670 (13.48%)
2T4P	2	Treatment	Proboscis	4	6344244	5945192	93.71%	5119037 (86.10%)	675931 (11.37%)
2T4T	2	Treatment	Trunk	4	7273025	7016244	96.47%	6259260 (89.21%)	418213 (5.96%)
2C8C	2	Control	Collar	8	5580484	5249739	94.07%	4064753 (77.43%)	1048030 (19.96%)
2C8P	2	Control	Proboscis	8	4651049	4249479	91.37%	3810908 (89.68%)	329442 (7.75%)
2C8T	2	Control	Trunk	8	6246604	5962549	95.45%	5118344 (85.84%)	582091 (9.76%)
2T8C	2	Treatment	Collar	8	5453671	5216084	95.64%	4537574 (86.99%)	538579 (10.33%)
2T8P	2	Treatment	Proboscis	8	6134518	5950218	97.00%	5625481 (94.54%)	180688 (3.04%)
2T8T	2	Treatment	Trunk	8	6433870	6136305	95.38%	5140862 (83.78%)	775107 (12.63%)
2C12C	2	Control	Collar	12	6913987	6548496	94.71%	3515271 (53.68%)	2841104 (43.39%)
2C12P	2	Control	Proboscis	12	9159487	8927213	97.46%	7496607 (83.97%)	1237093 (13.86%)
2C12T	2	Control	Trunk	12	5714963	5451287	95.39%	4810941 (88.25%)	437102 (8.02%)
2T12C	2	Treatment	Collar	12	6096237	5924673	97.19%	5458664 (92.13%)	291897 (4.93%)
2T12P	2	Treatment	Proboscis	12	5806881	5644299	97.20%	5185295 (91.87%)	323119 (5.72%)
2T12T	2	Treatment	Trunk	12	6958901	6636202	95.36%	5631918 (84.87%)	751803 (11.33%)
2C24C	2	Control	Collar	24	6444470	6019677	93.41%	4259508 (70.76%)	1602224 (26.62%)
2C24P	2	Control	Proboscis	24	6821810	6251237	91.64%	5510949 (88.16%)	584799 (9.35%)
2C24T	2	Control	Trunk	24	6825255	6563990	96.17%	5974563 (91.02%)	346034 (5.27%)
2T24C	2	Treatment	Collar	24	2354917	2236345	94.96%	1575813 (70.46%)	603378 (26.98%)
2T24P	2	Treatment	Proboscis	24	7679506	7079732	92.19%	6164599 (87.07%)	754846 (10.66%)
2T24T	2	Treatment	Trunk	24	6260388	5951618	95.07%	4918331 (82.64%)	818862 (13.76%)
3C0C	3	Control	Collar	0	7959684	7336677	92.17%	5361362 (73.08%)	1775294 (24.20%)
3C0P	3	Control	Proboscis	0	7907310	7326804	92.66%	6577277 (89.77%)	531793 (7.26%)
3C0T	3	Control	Trunk	0	7181077	6922738	96.40%	5093543 (73.58%)	1633244 (23.59%)
3T0C	3	Treatment	Collar	0	5690772	5408477	95.04%	3920033 (72.48%)	1321926 (24.44%)
3T0P	3	Treatment	Proboscis	0	7239449	6662025	92.02%	5722962 (85.90%)	781662 (11.73%)
3T0T	3	Treatment	Trunk	0	7130503	6907257	96.87%	5940785 (86.01%)	693213 (10.04%)
3C2C	3	Control	Collar	2	9274808	8952379	96.52%	4896091 (54.69%)	3810433 (42.56%)
3C2P	3	Control	Proboscis	2	6566317	6273093	95.53%	5781454 (92.16%)	347806 (5.54%)
3C2T	3	Control	Trunk	2	5837990	5668087	97.09%	5191878 (91.60%)	293999 (5.19%)
3T2C	3	Treatment	Collar	2	8028109	7521804	93.69%	5521721 (73.41%)	1781839 (23.69%)
3T2P	3	Treatment	Proboscis	2	7061698	6463452	91.53%	5687865 (88.00%)	612457 (9.48%)
3T2T	3	Treatment	Trunk	2	6926214	6734503	97.23%	6168230 (91.59%)	308365 (4.58%)
3C4C	3	Control	Collar	4	6390858	6143853	96.14%	3943033 (64.18%)	1973703 (32.12%)
3C4P	3	Control	Proboscis	4	5993608	5598079	93.40%	4923400 (87.95%)	516585 (9.23%)
3C4T	3	Control	Trunk	4	6564535	6347748	96.70%	5718327 (90.08%)	387765 (6.11%)
3T4C	3	Treatment	Collar	4	8556850	7888025	92.18%	6797949 (86.18%)	749843 (9.51%)
3T4P	3	Treatment	Proboscis	4	5968893	5770785	96.68%	5383439 (93.29%)	254049 (4.40%)
3T4T	3	Treatment	Trunk	4	5730831	5479068	95.61%	4812387 (87.83%)	399161 (7.29%)
3C8C	3	Control	Collar	8	5154655	4958781	96.20%	3341584 (67.39%)	1490293 (30.05%)

3C8P	3	Control	Proboscis	8	5807005	5342217	92.00%	4494142 (84.13%)	724794 (13.57%)
3C8T	3	Control	Trunk	8	8420164	8002595	95.04%	6343650 (79.27%)	1443964 (18.04%)
3T8C	3	Treatment	Collar	8	6632491	6297382	94.95%	4601590 (73.07%)	1455465 (23.11%)
3T8P	3	Treatment	Proboscis	8	10327251	9883725	95.71%	8784407 (88.88%)	895743 (9.06%)
3T8T	3	Treatment	Trunk	8	7457274	7228023	96.93%	6268530 (86.73%)	745845 (10.32%)
3C12C	3	Control	Collar	12	6148670	5932636	96.49%	4787838 (80.70%)	984910 (16.60%)
3C12P	3	Control	Proboscis	12	5768557	5606878	97.20%	5322743 (94.93%)	160805 (2.87%)
3C12T	3	Control	Trunk	12	5069516	4990290	98.44%	4645079 (93.08%)	156866 (3.14%)
3T12C	3	Treatment	Collar	12	7036687	6690400	95.08%	5370563 (80.27%)	1034916 (15.47%)
3T12P	3	Treatment	Proboscis	12	6351778	5999193	94.45%	5417998 (90.31%)	429705 (7.16%)
3T12T	3	Treatment	Trunk	12	6001892	5817068	96.92%	5332684 (91.67%)	260297 (4.47%)
3C24C	3	Control	Collar	24	6081225	5790525	95.22%	4158333 (71.81%)	1429633 (24.69%)
3C24P	3	Control	Proboscis	24	6628564	6432038	97.04%	6071478 (94.39%)	202682 (3.15%)
3C24T	3	Control	Trunk	24	7309285	7055457	96.53%	6333170 (89.76%)	486257 (6.89%)
3T24C	3	Treatment	Collar	24	5778315	5546276	95.98%	4210220 (75.91%)	1186209 (21.39%)
3T24P	3	Treatment	Proboscis	24	5308798	5157396	97.15%	4885209 (94.72%)	157901 (3.06%)
3T24T	3	Treatment	Trunk	24	4980744	4875991	97.90%	4533492 (92.98%)	144669 (2.97%)
D4C0C	4	Control	Collar	0	6868670	6477512	94.31%	4368309 (67.44%)	1862472 (28.75%)
D4C0P	4	Control	Proboscis	0	5620586	5233406	93.11%	4399361 (84.06%)	710442 (13.58%)
D4C0T	4	Control	Trunk	0	8706931	8337416	95.76%	6627191 (79.49%)	1242860 (14.91%)
D4T0C	4	Treatment	Collar	0	7988559	7544671	94.44%	5292331 (70.15%)	1979860 (26.24%)
D4T0P	4	Treatment	Proboscis	0	9063671	8407548	92.76%	6007608 (71.45%)	2201754 (26.19%)
D4T0T	4	Treatment	Trunk	0	5393736	5257154	97.47%	4768113 (90.70%)	289965 (5.52%)
D4C2C	4	Control	Collar	2	5713786	5331557	93.31%	3972782 (74.51%)	1187987 (22.28%)
D4C2P	4	Control	Proboscis	2	5290255	5135578	97.08%	4634383 (90.24%)	390185 (7.60%)
D4C2T	4	Control	Trunk	2	7247470	6992856	96.49%	6215177 (88.88%)	400884 (5.73%)
D4T2C	4	Treatment	Collar	2	8493454	8088639	95.23%	6581724 (81.37%)	1204269 (14.89%)
D4T2P	4	Treatment	Proboscis	2	8395474	8152251	97.10%	7012419 (86.02%)	984123 (12.07%)
D4T2T	4	Treatment	Trunk	2	4542746	4405781	96.98%	4050761 (91.94%)	170774 (3.88%)
D4C4C	4	Control	Collar	4	6006543	5707679	95.02%	4609339 (80.76%)	937726 (16.43%)
D4C4P	4	Control	Proboscis	4	8431385	8005446	94.95%	6851010 (85.58%)	970288 (12.12%)
D4C4T	4	Control	Trunk	4	4051456	3912466	96.57%	3538592 (90.44%)	230190 (5.88%)
D4T4C	4	Treatment	Collar	4	9133602	8626345	94.45%	5485546 (63.59%)	2772143 (32.14%)
D4T4P	4	Treatment	Proboscis	4	7456306	6869497	92.13%	5926302 (86.27%)	760706 (11.07%)
D4T4T	4	Treatment	Trunk	4	6438671	6171842	95.86%	5365641 (86.94%)	467423 (7.57%)
D4C8C	4	Control	Collar	8	8421559	8158439	96.88%	4931495 (60.45%)	2969969 (36.40%)
D4C8P	4	Control	Proboscis	8	7671829	7160221	93.33%	6372755 (89.00%)	567492 (7.93%)
D4C8T	4	Control	Trunk	8	6935579	6577132	94.83%	5226100 (79.46%)	894208 (13.60%)
D4T8C	4	Treatment	Collar	8	6460220	6274663	97.13%	2364013 (37.68%)	3798819 (60.54%)
D4T8P	4	Treatment	Proboscis	8	5362233	5195874	96.90%	4697446 (90.41%)	386600 (7.44%)
D4T8T	4	Treatment	Trunk	8	10294873	9945730	96.61%	8346686 (83.92%)	1198400 (12.05%)
D4C12C	4	Control	Collar	12	6538333	6121249	93.62%	5046539 (82.44%)	824174 (13.46%)
D4C12P	4	Control	Proboscis	12	6489481	5815248	89.61%	4782410 (82.24%)	857157 (14.74%)
D4C12T	4	Control	Trunk	12	7882213	7601848	96.44%	6752692 (88.83%)	476475 (6.27%)
D4T12C	4	Treatment	Collar	12	7615314	7291568	95.75%	5119583 (70.21%)	1947820 (26.71%)
D4T12P	4	Treatment	Proboscis	12	9172740	8495553	92.62%	7492987 (88.20%)	770470 (9.07%)
D4T12T	4	Treatment	Trunk	12	9285462	8862010	95.44%	7293881 (82.31%)	1250778 (14.11%)
D4C24C	4	Control	Collar	24	6441980	6161584	95.65%	5048131 (81.93%)	948399 (15.39%)
D4C24P	4	Control	Proboscis	24	6792552	6482955	95.44%	5964327 (92.00%)	382157 (5.89%)
D4C24T	4	Control	Trunk	24	7499029	7226370	96.36%	6258770 (86.61%)	682655 (9.45%)
D4T24C	4	Treatment	Collar	24	6888815	6738061	97.81%	6279929 (93.20%)	271243 (4.03%)
D4T24P	4	Treatment	Proboscis	24	6323963	6073715	96.04%	5698589 (93.82%)	208274 (3.43%)
D4T24T	4	Treatment	Trunk	24	9091657	8850005	97.34%	8127416 (91.84%)	373747 (4.22%)
D5C0C	5	Control	Collar	0	6741291	6428587	95.36%	4857687 (75.56%)	1382323 (21.50%)
D5C0P	5	Control	Proboscis	0	6936474	6738707	97.15%	6240966 (92.61%)	355332 (5.27%)
D5C0T	5	Control	Trunk	0	11939362	11603196	97.18%	10301594 (88.78%)	882578 (7.61%)
D5T0C	5	Treatment	Collar	0	9440731	8957886	94.89%	7005343 (78.20%)	1703204 (19.01%)
D5T0P	5	Treatment	Proboscis	0	6672742	6515341	97.64%	6131493 (94.11%)	255979 (3.93%)

D5T0T	5	Treatment	Trunk	0	5619606	5393752	95.98%	4907263 (90.98%)	263109 (4.88%)
D5C2C	5	Control	Collar	2	8345630	8205253	98.32%	2598378 (31.67%)	5456905 (66.51%)
D5C2P	5	Control	Proboscis	2	6717209	6541916	97.39%	6085430 (93.02%)	315147 (4.82%)
D5C2T	5	Control	Trunk	2	8313949	7961581	95.76%	6636493 (83.36%)	914837 (11.49%)
D5T2C	5	Treatment	Collar	2	9281689	8885460	95.73%	7740369 (87.11%)	930236 (10.47%)
D5T2P	5	Treatment	Proboscis	2	6935958	6741867	97.20%	6038532 (89.57%)	580623 (8.61%)
D5T2T	5	Treatment	Trunk	2	6629357	6366757	96.04%	5719072 (89.83%)	390444 (6.13%)
D5C4C	5	Control	Collar	4	8968741	8579678	95.66%	6736223 (78.51%)	1619027 (18.87%)
D5C4P	5	Control	Proboscis	4	8959151	8616720	96.18%	8042917 (93.34%)	396014 (4.60%)
D5C4T	5	Control	Trunk	4	9526634	9217766	96.76%	7887907 (85.57%)	1005780 (10.91%)
D5T4C	5	Treatment	Collar	4	9546943	8995323	94.22%	7015907 (78.00%)	1693093 (18.82%)
D5T4P	5	Treatment	Proboscis	4	7152826	6493425	90.78%	5442802 (83.82%)	868433 (13.37%)
D5T4T	5	Treatment	Trunk	4	9116102	8760066	96.09%	6859052 (78.30%)	1625778 (18.56%)
D5C8C	5	Control	Collar	8	9917822	9350048	94.28%	7684994 (82.19%)	1367868 (14.63%)
D5C8P	5	Control	Proboscis	8	7299196	6797345	93.12%	5616280 (82.62%)	994830 (14.64%)
D5C8T	5	Control	Trunk	8	6884499	6666754	96.84%	5948638 (89.23%)	455711 (6.84%)
D5T8C	5	Treatment	Collar	8	8450871	8153545	96.48%	5675398 (69.61%)	2240049 (27.47%)
D5T8P	5	Treatment	Proboscis	8	8510535	8141322	95.66%	7553406 (92.78%)	397634 (4.88%)
D5T8T	5	Treatment	Trunk	8	8964101	8532158	95.18%	7295918 (85.51%)	823004 (9.65%)
D5C12C	5	Control	Collar	12	7972181	7772655	97.50%	6954625 (89.48%)	613612 (7.89%)
D5C12P	5	Control	Proboscis	12	10554566	10195779	96.60%	9216872 (90.40%)	751906 (7.37%)
D5C12T	5	Control	Trunk	12	8686478	8157965	93.92%	6530827 (80.05%)	1301587 (15.95%)
D5T12C	5	Treatment	Collar	12	7033179	6669278	94.83%	5886725 (88.27%)	552411 (8.28%)
D5T12P	5	Treatment	Proboscis	12	4848823	4701432	96.96%	4219772 (89.76%)	373781 (7.95%)
D5T12T	5	Treatment	Trunk	12	8402106	8110750	96.53%	6858394 (84.56%)	912999 (11.26%)
D5C24C	5	Control	Collar	24	6943749	6701190	96.51%	4105884 (61.27%)	2360975 (35.23%)
D5C24P	5	Control	Proboscis	24	6734130	6058386	89.97%	5014630 (82.77%)	881669 (14.55%)
D5C24T	5	Control	Trunk	24	7812330	7412088	94.88%	5146666 (69.44%)	1579097 (21.30%)
D5T24C	5	Treatment	Collar	24	6701966	6352876	94.79%	5469281 (86.09%)	690669 (10.87%)
D5T24P	5	Treatment	Proboscis	24	7322777	6894277	94.15%	6199253 (89.92%)	500174 (7.25%)
D5T24T	5	Treatment	Trunk	24	6656478	6321918	94.97%	4799281 (75.91%)	1245145 (19.70%)

*Percentage of reads which retained after *fastp* quality filter

Table S2: Summary of PANTHER annotation efficiency

	Fraction of SwissProt IDs with PANTHER Pathway Annotation		Fraction of SwissProt IDs with PANTHER Protein Class Annotation	
Timepoint (vs. 0hrs)	Up-reg. DEGs	Down-reg. DEGs	Up-reg. DEGs	Down-reg. DEGs
2hrs	4/12 (33%)	13/79 (16%)	10/12 (83%)	48/79 (61%)
4hrs	14/108 (13%)	4/37 (11%)	58/108 (54%)	23/37 (62%)
8hrs	1/4 (25%)	3/14 (21%)	3/4 (75%)	8/14 (57%)
12hrs	0/3 (0%)	0/5 (0%)	0/3 (0%)	2/5 (40%)
24hrs	1/5 (20%)	6/16 (38%)	4/5 (80%)	4/16 (25%)
Total	20/132 (15%)	26/151 (17%)	75/132 (57%)	85/151 (56%)

Table S3: Summary of GO annotation efficiency

	Fraction of SwissProt IDs with Molecular Function Annotation		Fraction of SwissProt IDs with Biological Process Annotation		Fraction of SwissProt IDs with Cellular Component Annotation	
Timepoint (vs. 0hrs)	Up-reg. DEGs	Down-reg. DEGs	Up-reg. DEGs	Down-reg. DEGs	Up-reg. DEGs	Down-reg. DEGs
2hrs	4/12 (33%)	16/79 (20%)	5/12 (42%)	13/79 (16%)	5/12 (42%)	35/79 (44%)
4hrs	51/108 (47%)	14/37 (38%)	56/108 (52%)	21/37 (57%)	55/108 (51%)	25/37 (68%)
8hrs	3/4 (75%)	8/14 (57%)	1/4 (25%)	8/14 (57%)	1/4 (25%)	7/14 (50%)
12hrs	0/3 (0%)	2/5 (40%)	0/3 (0%)	3/5 (60%)	0/3 (0%)	3/5 (60%)
24hrs	1/5 (20%)	7/16 (44%)	1/5 (20%)	8/16 (50%)	2/5 (40%)	7/16 (44%)
Total	59/132 (45%)	47/151 (31%)	63/132 (48%)	53/151 (35%)	63/132 (48%)	77/151 (51%)

Table S4: Summary of SwissProt, KEGG, and Pfam annotation efficiency

	Fraction of DEGs with SwissProt Annotation		Fraction of DEGs with KEGG Annotation		Fraction of DEGs with Pfam Annotation	
Timepoint (vs. 0hrs)	Up-reg. DEGs	Down-reg. DEGs	Up-reg. DEGs	Down-reg. DEGs	Up-reg. DEGs	Down-reg. DEGs
2hrs	12/15 (80%)	79/133 (59%)	6/15 (40%)	18/133 (14%)	13/15 (87%)	106/133 (80%)
4hrs	108/179 (60%)	37/44 (84%)	65/179 (36%)	23/44 (52%)	140/179 (78%)	39/44 (89%)
8hrs	4/7 (57%)	14/19 (74%)	3/7 (43%)	9/19 (47%)	7/7 (100%)	19/19 (100%)
12hrs	3/3 (100%)	5/7 (71%)	2/3 (67%)	3/7 (43%)	2/3 (67%)	6/7 (86%)
24hrs	5/18 (28%)	16/30 (53%)	2/18 (11%)	11/30 (37%)	11/18 (61%)	28/30 (93%)
Total	132/222 (59%)	151/233 (65%)	78/222 (35%)	64/233 (27%)	173/222 (78%)	198/233 (85%)

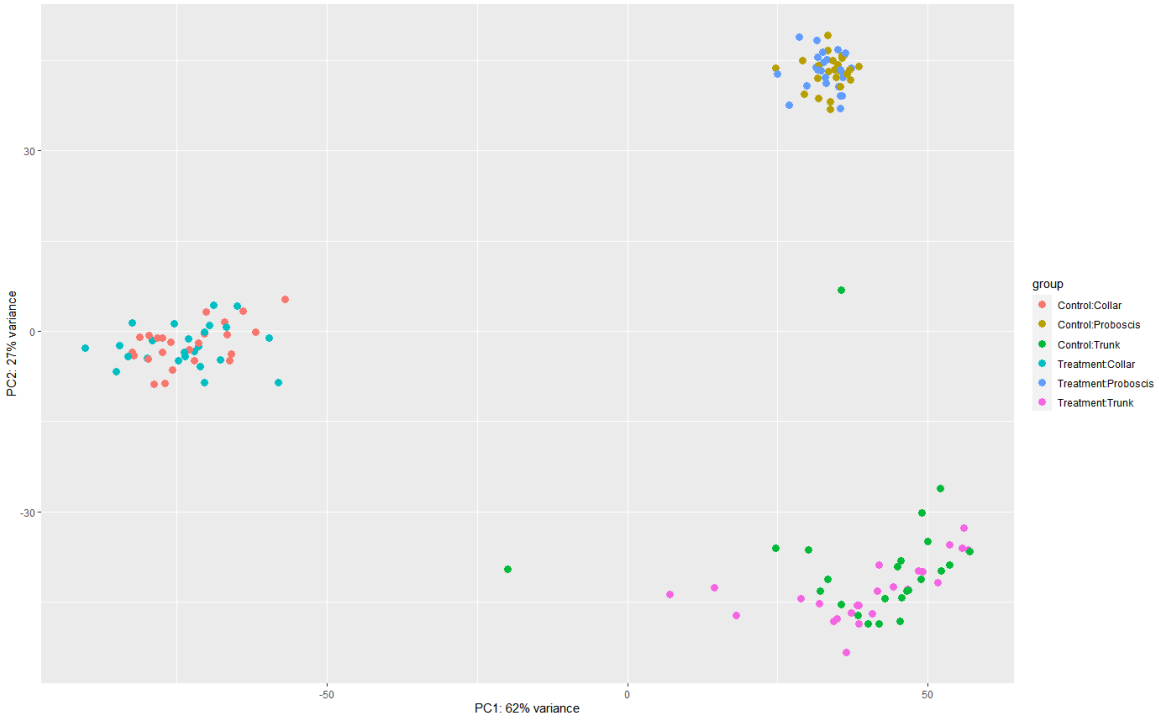


Figure S1. Principle component analysis of mean gene expression variance per sequencing library (after correcting for size factors). Samples are colored by condition (treatment or control) and tissue (proboscis, collar, or trunk). Approximately 89% of variance can be attributed to differences in gene expression among the three tissue types.

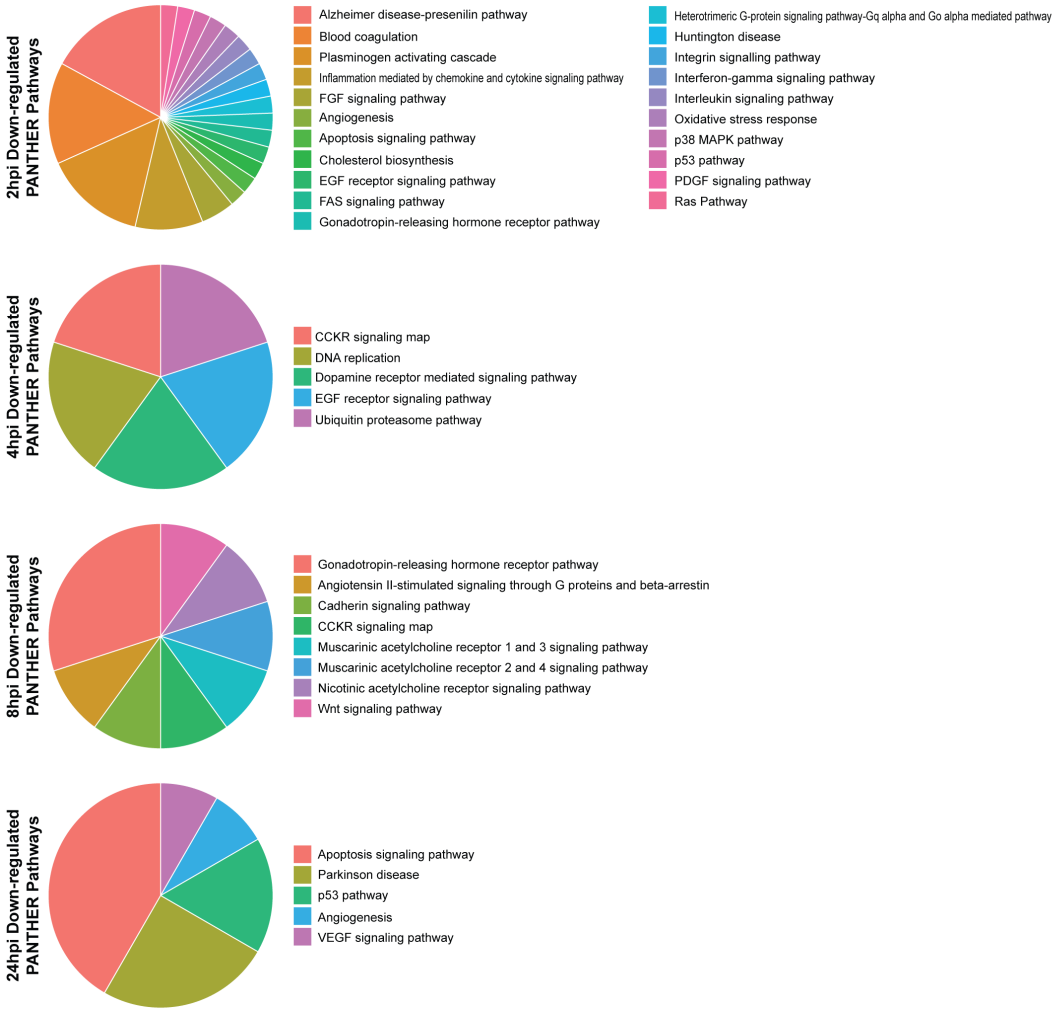


Figure S2. Pie plots showing the relative composition of PANTHER pathways for genes down-regulated at each timepoint. 12hpi is not shown as it did not possess any DEGs that could be annotated for a PANTHER pathway.

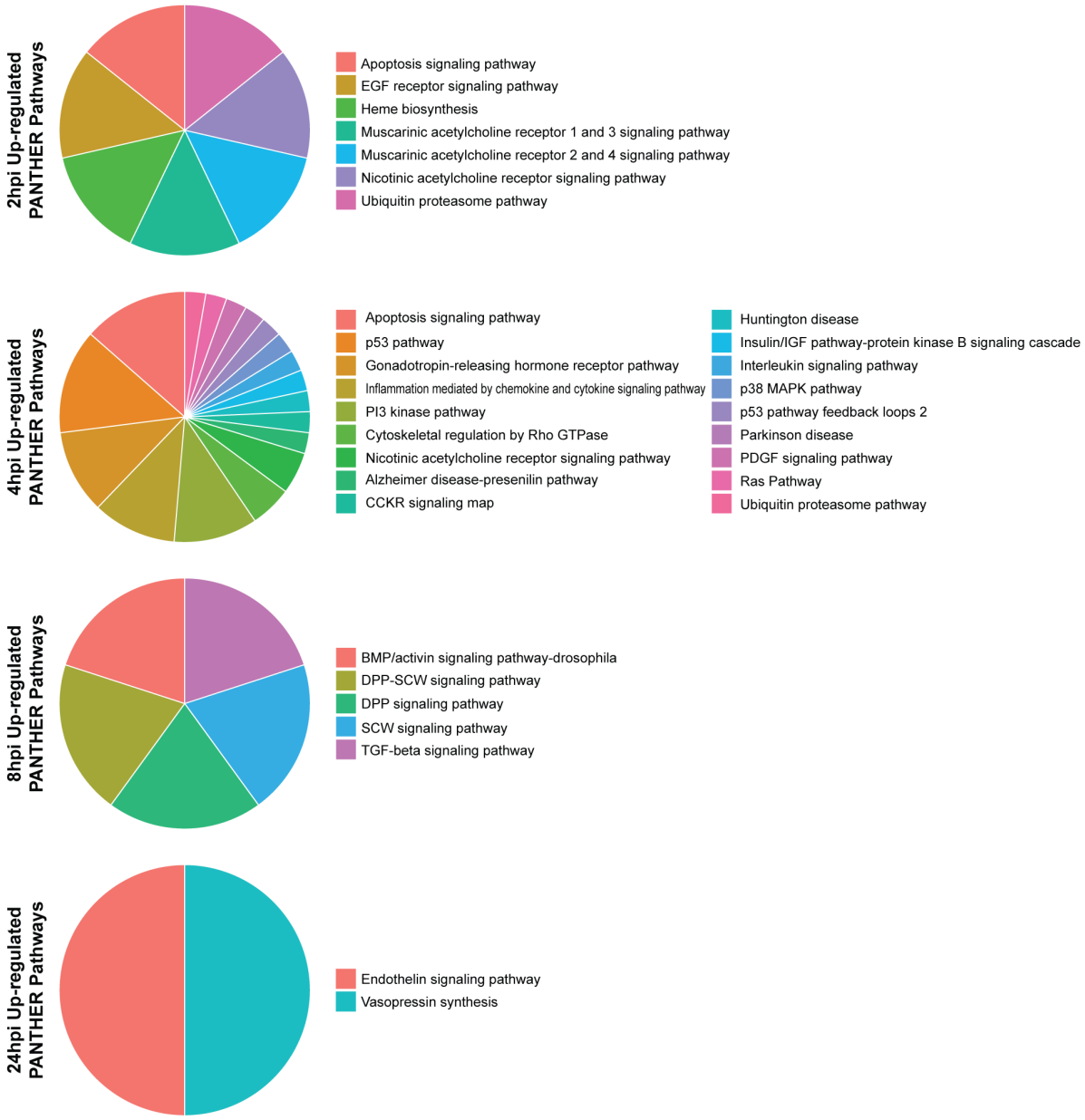


Figure S3. Pie plots showing the relative composition of PANTHER pathways for genes up-regulated at each timepoint. 12hpi is not shown as it did not possess any DEGs that could be annotated for a PANTHER pathway.

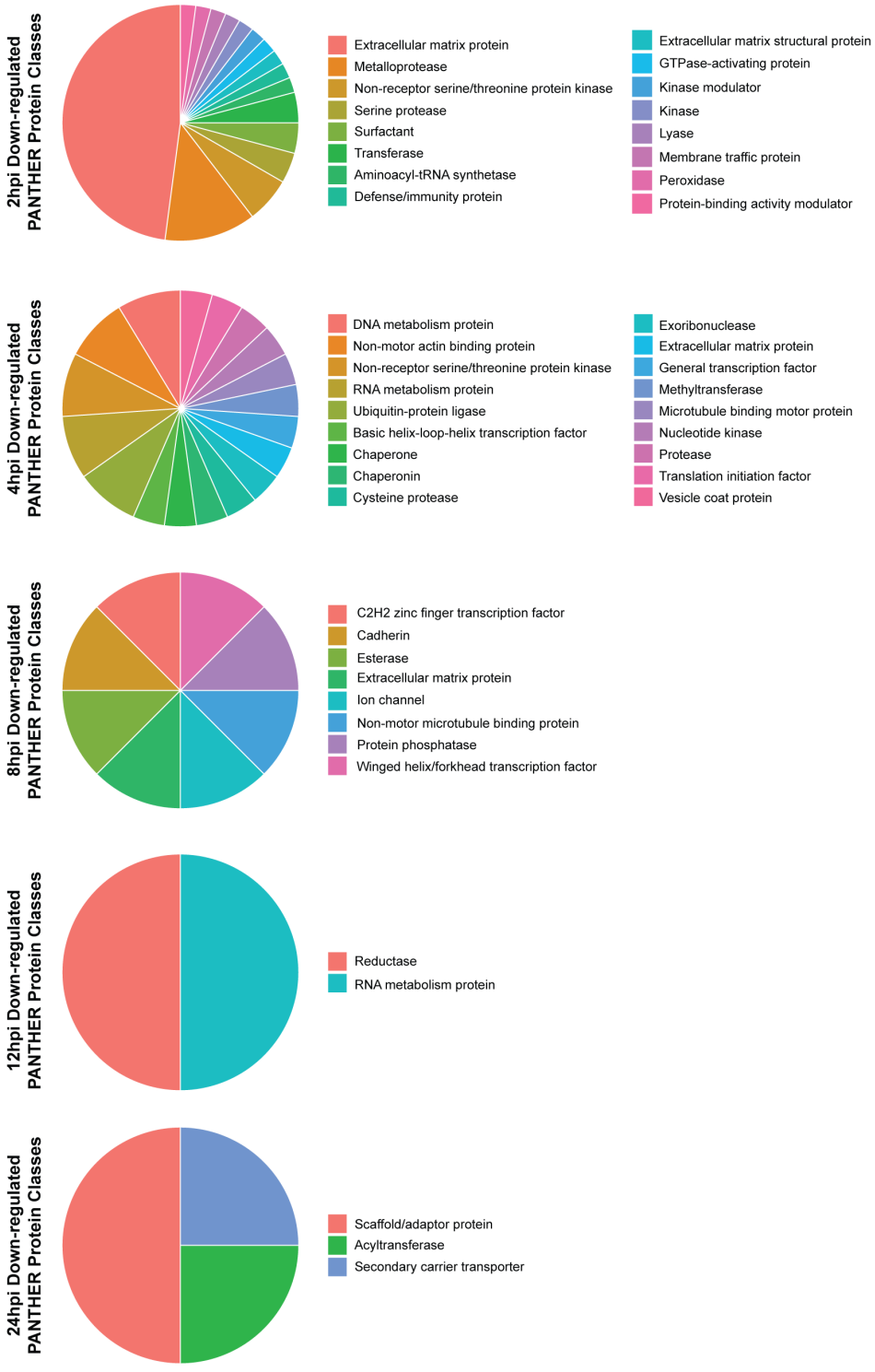


Figure S4. Pie plots showing the relative composition of PANTHER protein classes for genes down-regulated at each timepoint.

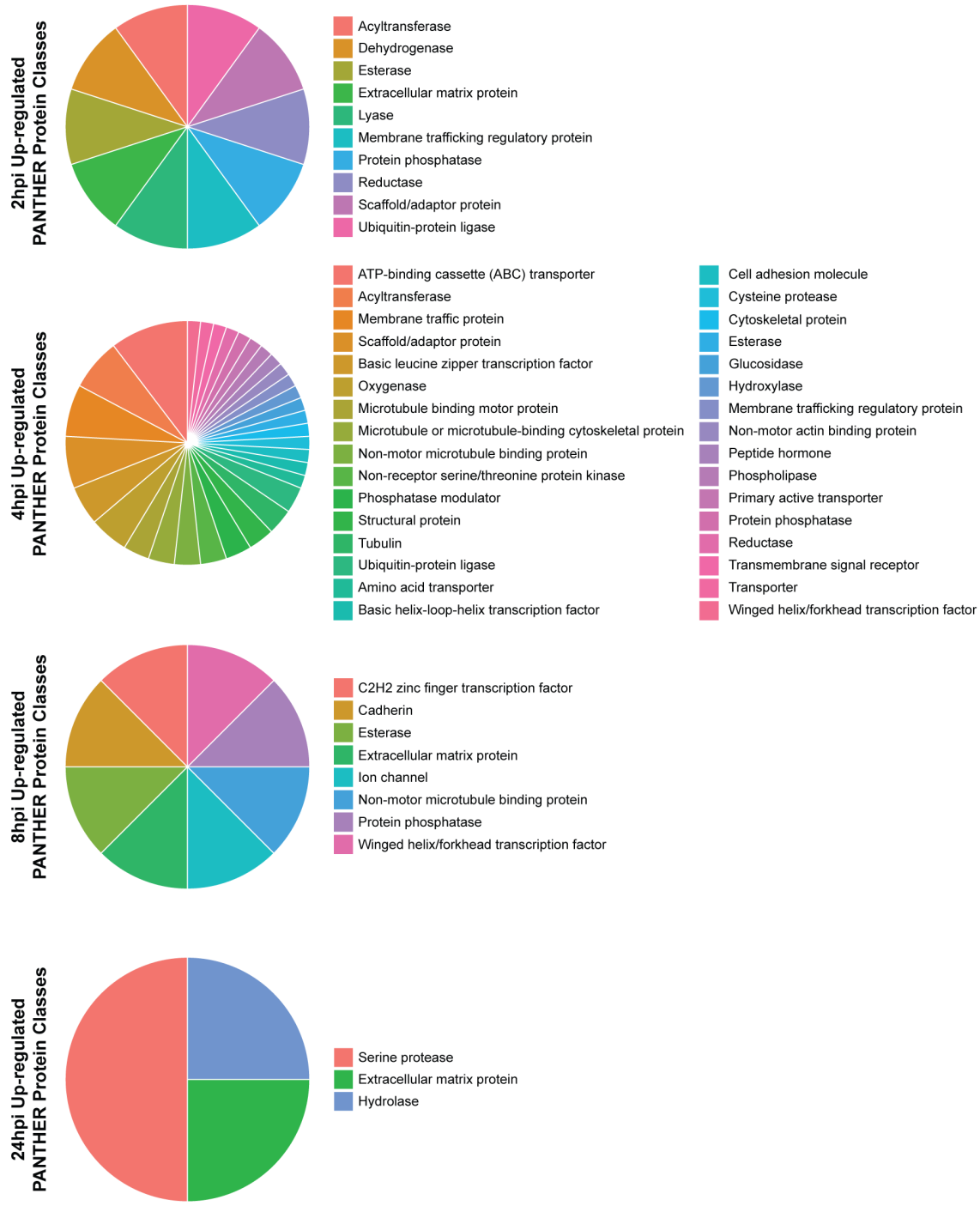


Figure S5. Pie plots showing the relative composition of PANTHER protein classes for genes up-regulated at each timepoint. 12hpi is not shown as it did not possess any DEGs that could be annotated for a PANTHER protein class.

Chapter 5: TIAMMAt: Leveraging Biodiversity to Revise Protein Domain Models, Evidence from Innate Immunity[‡]

5.1 Abstract

Sequence annotation is fundamental for studying the evolution of protein families, particularly when working with non-model species. Given the rapid, ever-increasing number of species receiving high-quality genome sequencing, accurate domain modeling that is representative of species diversity is crucial for understanding protein family sequence evolution and their inferred function(s). Here, we describe a bioinformatic tool called TIAMMAt (*Taxon-Informed Adjustment of Markov Model Attributes*) which revises domain profile hidden Markov models (HMMs) by incorporating homologous domain sequences from underrepresented and non-model species. Using innate immunity pathways as a case study, we show that revising profile HMM parameters to directly account for variation in homologs among underrepresented species provides valuable insight into the evolution of protein families. Following adjustment by TIAMMAt, domain profile HMMs exhibit changes in their per-site amino acid state emission probabilities and insertion/deletion probabilities while maintaining the overall structure of the consensus sequence. Our results show that domain revision can heavily impact evolutionary interpretations for some families (i.e., NLR's NACHT domain), whereas impact on other domains (e.g., rel homology domain and interferon regulatory factor domains) is minimal due to high levels of sequence conservation across the sampled phylogenetic depth (i.e., Metazoa). Importantly, TIAMMAt revises target domain models to reflect homologous sequence variation using the taxonomic distribution under consideration by the user. TIAMMAt's flexibility to revise any subset of the Pfam database using a user-defined taxonomic pool will make it a valuable tool for future protein evolution studies, particularly when incorporating (or focusing on) non-model species.

[‡] This chapter has been submitted as: Tassia MG, David KT, Townsend JP, Halanych KM. Submitted 2021. TIAMMAt: Leveraging biodiversity to revise protein domain models, evidence from innate immunity. *Molecular Biology and Evolution*

5.2 Introduction

Accurate assignment of protein identity is a fundamental component of molecular studies involving non-model species. Such studies often begin by tethering an uncharacterized protein's identity to a homolog of known function to infer, for example, residue-specific selective pressures (Buckley & Rast 2012), protein-protein interaction networks (Szklarczyk et al. 2014), or evolutionary divergence (Tassia et al. 2017). Errors in these assessments can be costly. In the field of evolutionary and developmental biology, for example, over- or underestimating the full complement of protein family members in a non-model species can compromise the design of genetic reporter constructs (Cavalieri & Spinelli 2014) or CRISPR/Cas9 targets (Connahs et al. 2019). These errors cost researchers time, financial resources, and can negatively impact the accuracy of scientific conclusions.

Comparative molecular studies employing non-model species (Buckley & Rast 2015; Brennan & Gilmore 2018) often utilize a common bioinformatic approach when assigning evolutionary affinity and putative function to uncharacterized proteins (Loewenstein et al. 2009). Initially, protein identity is typically labeled using primary sequence similarity, which measures the number of pairwise matches between two sequences. Although similarity metrics aid protein identification (prematurely extrapolated to indicate orthology in some cases; Chen et al. 2007), similarity alone is insufficient to infer function in an evolutionary context (Liu et al. 2018). Given the pitfalls when relying on similarity alone, uncharacterized protein sequences are also placed in a phylogenetic context to verify homology (Tassia et al. 2017) and further annotated with domains – amino acid sequence patterns which can be used to assign function to discrete territories within a full amino acid sequence (Wojcik & Schächter 2001; Zhao et al. 2008). When used in concert, phylogenetic methods and domain annotation can reinforce hypotheses on protein family evolution and their functional variation across deep evolutionary timescales (Buckley & Rast 2012; Costa-Paiva et al. 2017; Tassia et al. 2017; Gerdol et al. 2017; Costa-Paiva et al. 2018). For example, mammalian inflammatory and apoptotic caspases invariably possess a carboxy-terminal protease effector domain and paralogs within the family can be categorized by their amino-terminus CARD or DED domain(s) (Man & Kanneganti 2016). These same rules remain consistent when applied to categorizing caspases in *Hydra,* a freshwater cnidarian (Lasi et al. 2010). Importantly,

annotation of an uncharacterized protein with domain structure requires a database of known protein domains.

The Pfam database contains a well-curated and frequently updated catalogue of domain models placed in an evolutionary context for protein studies across the tree of life (Sonnhammer et al. 1997; Finn et al. 2016; El-Gebali et al. 2018). Each Pfam domain entry is created as follows: 1) a seed alignment is generated from representative sequences containing a conserved pattern that has been characterized in at least one of the sampled species; 2) the seed is then used to build a domain profile hidden Markov model (HMM) using the open source HMMER software package (Eddy 2009); lastly, 3) the new profile HMM is searched against Pfam's proteomic sequence database as quality control and to provide evolutionary context (Sonnhammer et al. 1997; Eddy 2009; El-Gebali et al. 2018; Mistry et al. 2020). Encoding Pfam domains as profile HMMs, in turn, allows protein domain searches to adopt the robust statistical framework underlying HMMs and information entropy (Hernando et al. 2005), along with the benefit that domain profile HMMs are rapidly searchable (Eddy 2009). Although variation encoded within the model is designed to capture homologs from species outside those represented directly within the seed alignment (El-Gebali et al. 2018), many domain profiles are derived of only a few species, reducing the model's capacity to identify homologous domain sequences in phylogenetically distant taxa. Currently, domain seed alignments are dominated by sequences from a few biomedical model taxa (**Fig. 1**), or closely related taxa, and the trend in sequencing bias towards these model systems is becoming increasingly exacerbated (David et al. 2019).

Using innate immunity proteins as a case study, we show that revising domain profile seed alignments to directly account for underrepresented protein diversity aids homolog identification in non-model animal species. Innate immunity signaling relies on pattern recognition receptors (PRRs) which recognize broad categories of microbes (such as RNA viruses or Gram-positive bacteria) by binding specific pathogen-associated moieties (Beutler 2004). Unlike adaptive immunities which evolved independently in both jawed- and jawless vertebrates (Flajnik & Kasahara 2010), PRRs were likely present in the last common ancestor to all animal lineages (Bosch 2013) and some innate immunity protein families have undergone several notable lineage-specific diversifications (Buckley & Rast 2012; Gerdol et al 2017). Important for the context of our study, PRRs rely on domain-domain interactions for activation and signal transduction

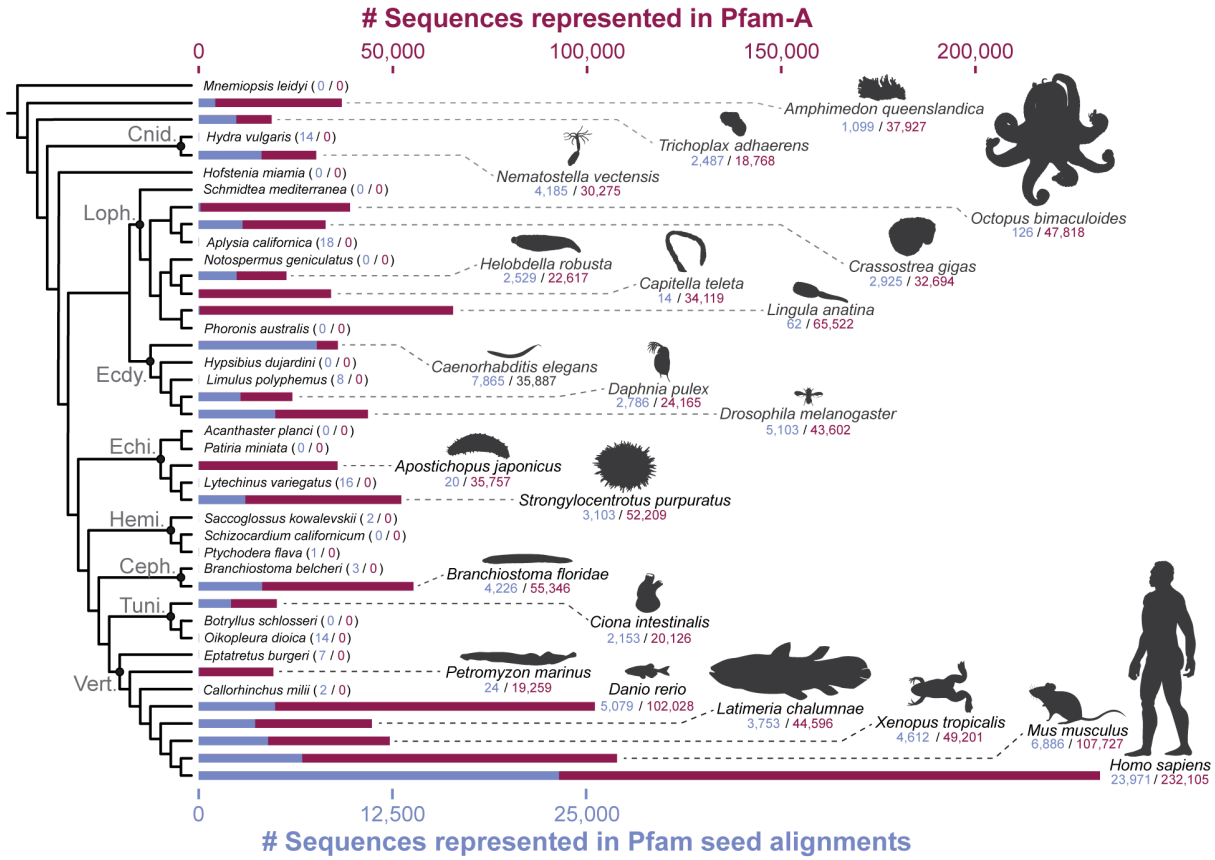


Figure 1. Taxon representation within Pfam database. In blue (left values following species names) are the total number of occurrences a species appears across all Pfam seed alignments. In red (right values following species names) are the total number of sequences within a species' reference proteome captured by all Pfam domain profiles. Blue and red bars duplicate the numeric values next to species names and the blue bar is superimposed on the red bar, each with its own independent scale displayed at the top (Red) and bottom (Blue) of the plot. Cladogram depicts consensus phylogenetic relationships derived from Laumer et al. 2019. Abbreviations: Cnid. – Cnidaria, Loph. – Lophotrochozoa, Ecdy. – Ecdysozoa, Echi. – Echinodermata, Hemi. – Hemichordata, Ceph. – Cephalochordata, Tuni. – Tunicata, Vert. – Vertebrata.

(O'Neill & Bowie 2007), possess defined domain architectures (Akira & Takeda 2004; Kowalinski et al. 2011; Lechtenberg et al. 2014), and have dominantly been studied in biomedical model species (Leulier & Lemaitre 2008). Among the most well-described PRRs are NOD-like receptors (NLRs; Lechtenberg et al. 2014), Toll-like receptors (TLRs; Akira & Takeda 2004), and RIG-I-like receptors (RLRs; Kowalinski et al. 2011). Although these three PRR families differ from one another in their domain architectures and signal transduction partners (**Supplementary Fig. 1**), all three converge on the activation of nuclear factor κB (NF-κB) and/or interferon regulatory factors (IRFs). These transcription factors promote expression of pro-inflammatory cytokines (e.g.,

interleukins and tumor-necrosis factors), antimicrobial-, and/or antiviral peptides (Hiscott 2007; Zhang et al. 2017). The current perspective on PRR signaling is intimately tied to domain architecture, emphasizing the importance of protein annotation as a fundamental prerequisite when placing PRRs in a comparative and evolutionary framework.

Here, we show that revising profile seed alignments aids identification of domain homologs in non-model species and can provide insight into protein family evolution. The value of phylogenetically representative domain models cannot be overstated as identifying protein homologs across deep evolutionary timescales is a challenge that continues to grow as genomes become more accessible, particularly for those of historically underrepresented species (Buckley & Rast 2012; Costa-Paiva et al. 2017; Tassia et al. 2017; Gerdol et al. 2017; Costa-Paiva et al. 2018). To this end, we explore the effects of revising domains which are essential for animal innate immunity signaling pathways, a group of evolutionarily ancient protein families within Metazoa that rely on domain-domain interactions and show considerable variation between taxa. Below, we describe a domain revision protocol called TIAMMAt (*Taxon-Informed Adjustment of Markov Model Attributes;* pronounced "TEE-a-mat" or "TEE-a-maht") and apply it to the domains at the core of PRR signaling to reveal the effects of narrow phylogenetic representation within domain seed alignments on domain homolog detection in non-model species.

5.3 New Approaches

TIAMMAt provides an automated and reproducible method for revising Pfam domain profile HMMs to capture homologous sequence diversity contingent upon a user-defined taxonomic distribution. TIAMMAt fundamentally relies on HMMER's suite of profile-to-sequence comparison tools and their direct association with Pfam domain database entries. Although TIAMMAt is utilized in the context of metazoan innate immunity for our study, the program can revise any domain profile(s) within Pfam based on a user-defined taxonomic pool. For example, TIAMMAt can be applied to investigate the evolution of the death domain superfamily in all eukaryotes just as it can be used to revise and identify globin domains within arachnids. For each domain revised by TIAMMAt, the program will produce 1) a domain profile HMM which directly accounts for homologous sequence variation within the queried taxon/taxa, and 2) the subset of proteins from each taxon which possess the domain of interest (before and

after revision) (**Fig. 2**). Importantly, TIAMMAt is a versatile tool for protein evolution studies that can be catered to the investigator's subject of research.

TIAMMAt executes the following steps for each target domain profile (see **Materials & Methods, Fig. 2, Supplementary Fig. 2,** and **Supplementary Table 1**). First, each supplied proteome (defined here as the whole collection of protein sequences derived of an organism's genome) is searched for occurrences of the target domain where the target domain must meet two conditions: A) The per-target/per-domain expectation values (e-values) do not exceed HMMER's reporting or inclusion thresholds (default per-domain/-target reporting threshold ≤ 10.0 and per-domain/-target inclusion threshold ≤ 0.01, respectively), and B) the target domain has the lowest per-domain e-value within the sequence envelope in which it is identified (relative to every other domain in the Pfam database). Conceptually, coordination of these two filtering conditions constrains the domain revision process to not only incorporate amino acid sequences labeled as "true homologs" to the target domain (as defined by HMMER's profile-to-sequence comparison pipeline; Eddy et al. 2009), but also that the target domain has the lowest probability of being a false positive within the

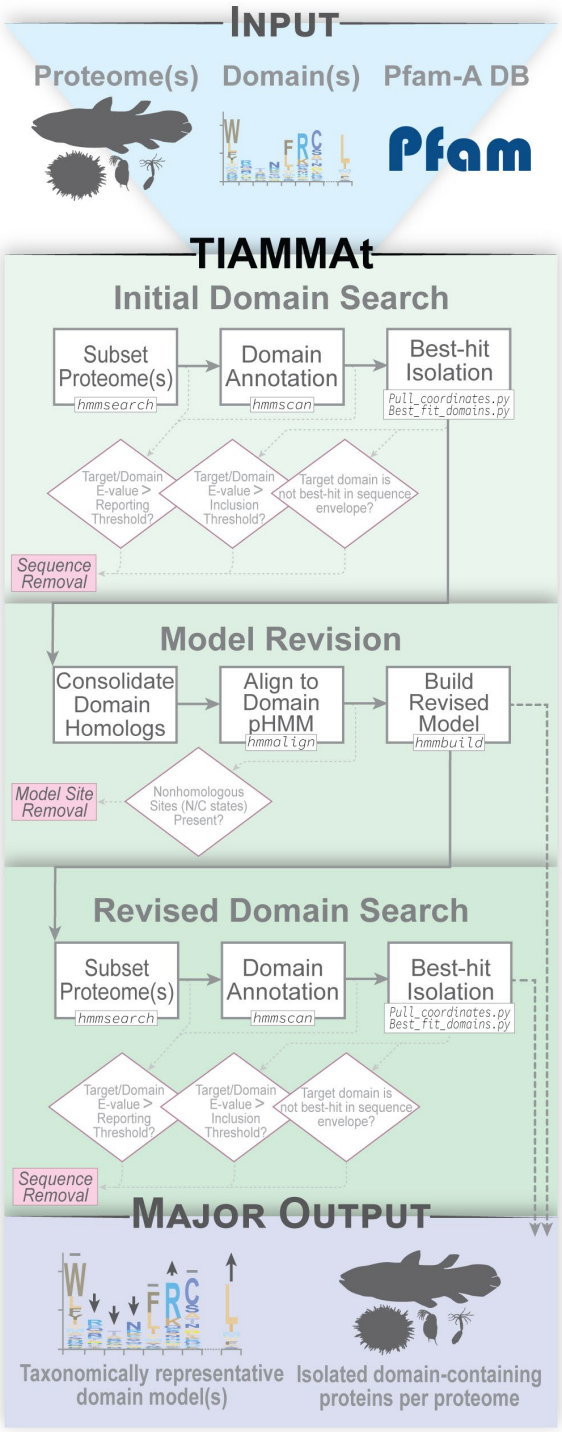


Figure 2. Bioinformatic operations by TIAMMAt. Boxes and solid arrows symbolize major workflow of TIAMMAt. Dashed lines and diamonds show filtering criteria used by TIAMMAt. For further detail, see **Supplementary Fig. 2**.

sequence envelope it was identified (relative to all other domains). These constraints are specifically designed to avoid false positives which could otherwise be introduced by incorporating similar, but non-target, motifs into the domain seed alignment (e.g., CARD and DED are similar and related, but exhibit distinct interactive properties; Jiang et al. 2012). Importantly, TIAMMAt can also be run with user-specified per-target/per-domain e-value thresholds. Such flexibility may be useful if, for example, investigators intend to experimentally test a domain's function in a pharmacological context using a protein derived from non-model species (Agrawal et al. 2016). In this scenario, increasing the stringency of these thresholds permits investigators to identify sequence structures strictly similar to those captured in the original domain model, while also accounting for evolutionary distance between their subject species and those used to build the original seed alignment.

Following annotation filtering, TIAMMAt extracts all best-hit domain sequences and aligns them to the profile HMM along with the original Pfam seed sequences. This revised seed alignment, which is a direct derivation of the original alignment (and constrained *a priori* to align to original domain profile HMM), is then reconstructed into a single new revised domain profile HMM. Finally, all proteomes are searched once again for the target domain(s) using all the original and revised models generated during the run (**Fig. 2**). For the purposes of our study, we revised domains integral to the functions of TLRs, RLRs, NLRs, NF-κBs, and IRFs (**Supplementary Tables 1 & 2, Supplementary Figs. 3 & 4**) using 39 publicly available proteomic datasets representative of taxa across the metazoan phylogeny (**Supplementary Table 3**), focusing particularly on species which have been labeled within scientific literature as emerging or non-model (e.g., Simakov et al. 2013; Simakov et al. 2015; Hall et al. 2017; Gehrke et al. 2019).

5.4 Results & Discussion

5.4.1 Trends in Model Revision

Given that the model revision process employed by TIAMMAt is dependent upon the original model, the revised model's properties are constrained in a few key ways (**Fig. 3**): 1) Overall sequence structure remains consistent between base and revised models, where low information content regions retain low impact on sequence hit probabilities and high information

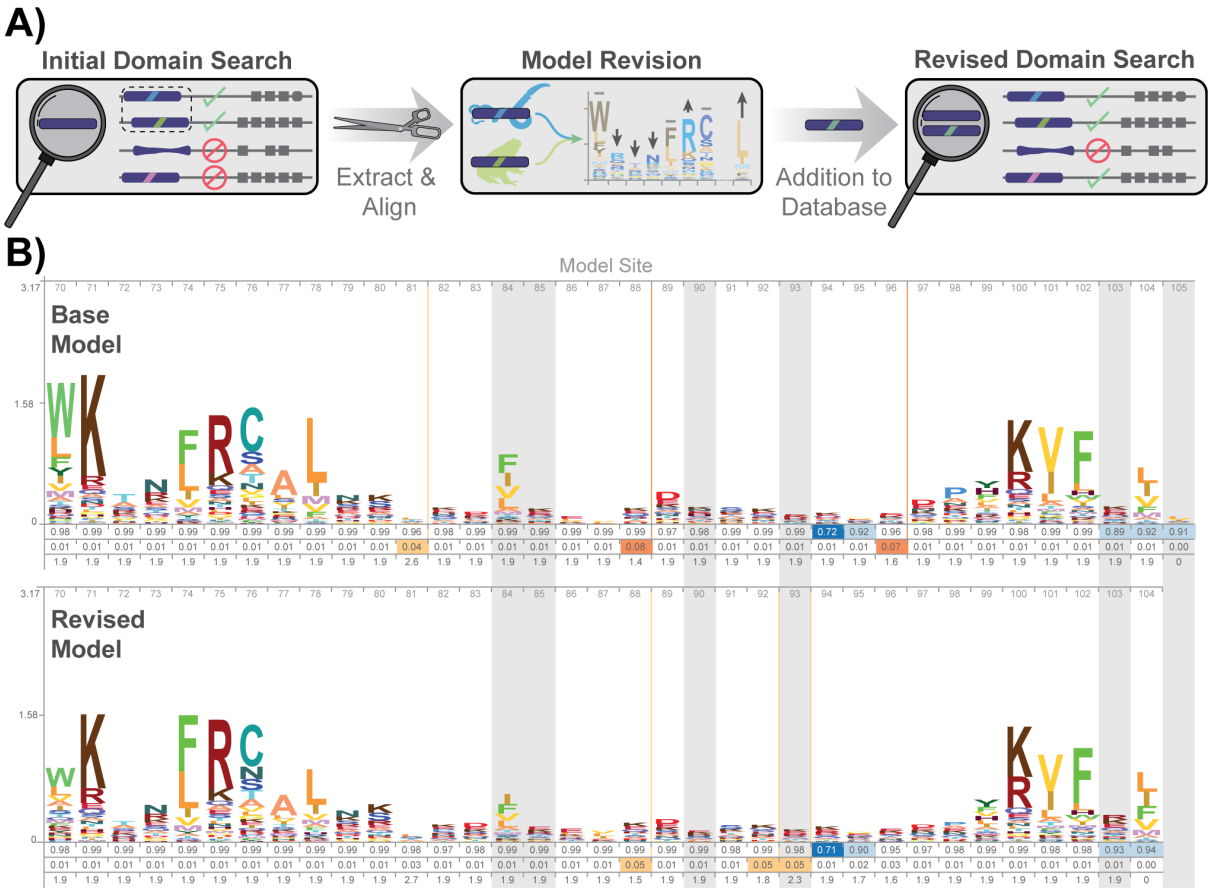


Figure 3. Domain revision by TIAMMAt. A) Schematic overview of the three major operations performed by TIAMMAt (see **Materials & Methods** for details). First, target domains are searched for among input proteomes. These domains are extracted and aligned to the associated domain profile HMM. Second, the alignment is recompiled into a revised domain profile HMM. Lastly, revised domains are appended to a local installation of Pfam and used to re-annotate all sequences which possess either the base or revised model. B) Visual alignment of IRF domain (PF00605) C-terminus Skylign graphs (Wheeler et al. 2014) showing common parameter adjustments after domain revision, including changes in most probable amino acid state emission per site (grey columns), non-consensus state trimming (last column), and overall adjustments in information content (bit score) per site (Y-axis value per site). X-axes below each diagram are as follows (from top to bottom): occupancy probability, probability of insertion following site, and length of insertion following site. Vertical bars mark sites where the insertion probability > 0.01. Relative height of amino acid symbols reflects their emission probability relative to all other amino acid states at that site.

content regions retain their overall structure and comparatively high statistical weight; 2) the revised model's length is partially constrained to the base model's length by trimming nonhomologous residues near ends of the alignment (via *hmmalign'*s *--trim* flag), avoiding overparameterization which could be produced as the new seed alignment incorporates new sequences; 3) most changes are adjustments in the emission probabilities per residue per site of the domain profile and changes in insertion probabilities, but not changes in overall consensus sequence structure.

For all analyses, we included the human proteome as a positive control for domain revision. All target domain-containing sequences identified before revision in human were also identified after revision. Additionally, the human sequences found to possess a target domain only after revision (**Supplementary Table 2**) met one of two conditions: A) following phylogenetic analysis, the newly identified protein clustered with sequences known to possess the original (pre-revision) domain, suggesting model revision produced a profile which describes sequence variation absent in the original model; or B) the sequence represented a poorly understood protein and the revised domain was assigned to a sequence envelope where no other domain met HMMER's inclusion threshold (i.e., the revised domain fell within an unannotated sequence envelope). Importantly, the magnitude of change for individual amino acid emission probabilities per-site was not equivalent across all model revisions, suggesting TIAMMAt is sensitive to the degree of sequence variation within the input domain(s). In turn, revision of some domains showed limited change. For example, domain revision yielded 4 and 8 additional IRF and NF-κB family members, respectively (**Supplementary Table 2**). When all IRF and NF-κB proteins were placed in a phylogenetic context, resolution of deeper nodes was poor (**Supplementary Figs. 5 & 6**), and lack of domain architecture diversity among these transcription factor family members further exacerbated the challenge of interpreting novel IRF and NF-κB members in an evolutionary context. In contrast, revision of domains central to NLR, TLR, and RLR signaling pathways produced more dramatic changes that could be interpreted in an evolutionary framework.

TIAMMAt revises domain profile HMMs to capture homologous sequence variation represented within the proteomes provided; as such, taxon selection (and dataset quality) directly influences the evolutionary context for revised domains and their potential use in subsequent searches. Revision of immunity-associated domains using, for instance, a single genus of

crustaceans may not produce revised domains appropriate for studies at the scale of Metazoa. However, revising a domain using a single clade of organisms would yield interesting and valuable results if that clade of organisms is already known to possess divergent proteins, particularly for domains directly implicated in protein-protein interactions (as is the case for the TIR and NACHT domains). Under these circumstances, model revision using narrow taxon sampling would facilitate identification of lineage-specific domain structures. Importantly, because each domain profile HMM describes the variation observed for a single Pfam domain, the original and revised models are not mutually exclusive.

Below, we report and discuss the effects of domain model revision on three key innate immunity PRR families: NLRs, TLRs, and RLRs. Direct comparisons between studies investigating PRR diversity across Metazoa can be difficult due to differences in bioinformatics and the definitions used to define each family (e.g., Nehyba et al. 2009; Buckley & Rast 2012; Yuen et al. 2013; Tassia et al. 2017). Nevertheless, the number of domain-containing sequences identified by TIAMMAt prior to domain revision are reported in **Supplementary Tables 4-7** and have been categorized to reflect conservative estimates of the number of PRR family members recognized in the literature (Buckley & Rast 2015; Pugh et al. 2016; Gerdol et al. 2018). Although TIAMMAt's domain filtering conditions are explicitly designed to avoid false positives, phylogenetic methods serve as a valuable framework to support evolutionary relationships between protein sequences identified before and after model revision. In each tree generated during our analyses, all domain-containing proteins exclusively identified after revision fell within orthology groups comprised of proteins identified prior to revision.

5.4.2 NOD-like Receptors

NACHT domain revision yielded the greatest increase in the number of domain-containing sequences in our analyses (**Fig. 4, Supplementary Tables 4-6**). We defined NLRs as proteins possessing both a NACHT domain and a terminal series of LRRs, consistent with literature on the structural perspectives of NLR signaling kinetics and previous NLR surveys (Laroui et al. 2011; Mo et al. 2012; Meunier & Broz 2017). Following NACHT revision, we identified novel NLRs in the sea snail, *Aplysia californica* (n=1 additional), seastars *Acanthaster planci* (n=1) and *Patiria miniata* (n=5), the sea cucumber, *Apostichopus japonicus* (n=1), acorn worms *Ptychodera flava*

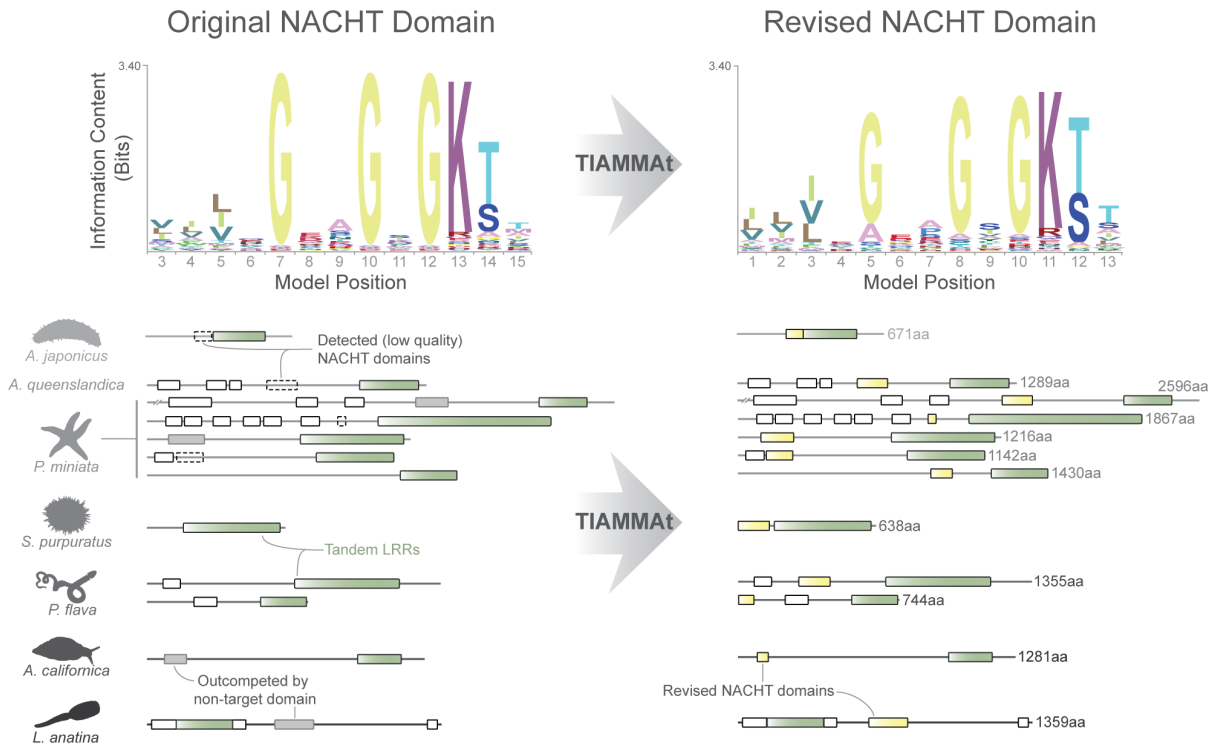


Figure 4. NACHT domain model revision identifies previously unrecognized NLRs. Top: Skylign (Wheeler et al. 2014) graph for positions 3-15 of the original NACHT domain model (PF05729; left) and positions 1-13 of the revised domain (right). Relative height of amino acid symbols reflects their emission probability relative to all other amino acid states at that site. Bottom: Domain diagrams of NLR structures before (left) and after (right) domain revision, highlighting the utility of incorporating taxonomic diversity into the NACHT domain seed alignment when working with underrepresented taxa. Regarding domain diagrams, green bars represent tandem leucine-rich repeats, yellow bars represent NACHT domains, dotted-outline empty bars represent NACHT domains below inclusion threshold, grey bars represent domains which outcompete NACHT for best-fit domain within the sequence envelope, and white bars represent other NLR accessory domains.

(n=2) and *Schizocardium californicum* (n=1), and the purple urchin, *Strongylocentrotus purpuratus* (n=1; **Supplementary Table 6**). Aside from novel CARD-containing NLRs (i.e., NLRC subfamily) identified in *P. flava* and *S. californicum*, all other NLRs identified after revision by TIAMMAt could not be classified into the four canonical NLR subfamilies (Kanneganti et al. 2007; Meunier & Broz, 2017) based solely on domain architecture (**Supplementary Fig. 7; Supplementary Table 6**). This result is consistent with previous findings showing NLRs exhibit more variety in their N-terminal domains among invertebrate taxa than within Vertebrata (Lange et al. 2011; Hamada et al. 2013; Yuen et al. 2013). Moreover, PYD-containing NLRs (i.e., NLRP subfamily) appear to be exclusive to euteleosts in our dataset (i.e., *Latimeria*, zebrafish, and

human), even after domain revision. Coincidentally, PYD, independent of NACHT, could only be identified in euteleost taxa (data not shown). Unlike the NLRCs which can directly elicit cell-death behaviors through homotypic CARD interactions, NLRPs (which possess an N-terminal PYD in place of a CARD) require ASC as a signaling intermediate (**Supplementary Fig. 1**), a short adaptor protein containing both a PYD and CARD (Lamkanfi & Dixit, 2012), before signaling for cell-death.

Our evolutionary analysis supports previous studies (Messier-Solek 2010; Hamada et al. 2013; Yuen et al. 2013; Gerdol et al. 2018) which suggest vertebrate-defined NLR subfamilies (i.e., NLRAs, NLRBs, NLRCs, and NLRPs) are insufficient for classifying NLRs outside Vertebrata (**Supplementary Table 6, Supplementary Fig. 8)**. Noncanonical NLRs identified in our study include a collection N-terminal death domain (or juxtaposed death and CARD domains) NLRs present in cephalochordates (*Branchiostoma belcheri* and *B. floridae*) and echinoderms (*Acanthaster planci, Patiria miniata, Apostichopus japonicus, Strongylocentrotus purpuratus, Lytechinus variegatus*) (**Supplementary Table 6**). Assuming the overall domain structures of metazoan NLRs retain their functional regionalization (i.e., C-terminal LRRs operate as ligand-binding, NACHT domains promote oligomerization, and the N-terminal domains are responsible for protein-protein interaction and signal transduction), the presence of noncanonical death domain superfamily members among NLRs may indicate a degree of evolutionary flexibility connecting pathogen recognition to the various death domain superfamily-associated signaling effects such as inflammation, apoptosis, cytokine/chemokine expression, and transcriptional regulation (Park et al. 2007; Kwon et al. 2012).

Outside of the death domain superfamily, NLRs identified in nine invertebrate taxa possess a higher eukaryotes and prokaryotes nucleotide-binding (HEPN) domain at their N-terminus. In a previous survey of HEPN domain sequence evolution across the tree of life (Anantharaman et al. 2013), HEPN proteins were predicted to act as either RNA sensors or catabolic RNases associated with RNA-dependent host-defense and stress response. Although we can loosely predict HEPN-NLRs may function as a cytoplasmic sensor for some category of RNAs, broader taxon sampling among underrepresented animal phyla and targeted molecular studies will be required to validate these proteins' hypothetical role in immunity.

The N-terminal domain of NLRs is far more diverse than what has traditionally been represented within vertebrates. The non-canonical NLRs identified in this study represent an underappreciated subset of the NLR protein family, perhaps indicative of more diverse functional roles for the family over the course of animal evolution. Moreover, because the search protocol employed by TIAMMAt isolates all proteins containing a target domain (which meet TIAMMAt's statistical prerequisites), several NACHT domain-containing proteins with undocumented affinity for NLR signaling pathways were identified before and after revision (**Supplementary Fig. 7, Supplementary Table 6**). Given their role in facilitating protein-protein interactions between two or more NACHT-containing proteins (Lamkanfi & Dixit, 2012), these unclassified NACHT domain-containing proteins warrant further investigation for their potential role in NLR signaling regulation across Metazoa.

5.4.3 Toll-like Receptors

Following TIR domain revision (PF01582 and PF13676), additional Toll-like receptor (TLR) proteins were identified in the tunicates *Ciona intestinalis* (n = 2 additional) and *Botryllus schlosseri* (n = 1), the stalked brachiopod, *Lingula anatina* (n = 1), and the lancelet chordate, *Branchiostoma belcheri* (n = 1) (**Fig. 5**). Whereas novel TLRs identified in *L. anatina* and *B. belcheri* occur in a background of >20 and >40 TLRs, respectively (Halanych & Kocot 2014; Huang et al. 2015; Gerdol et al. 2018), proteins detected in tunicates after revision are proportionally more substantial, doubling the number of reported TLRs in *B. schlosseri* from 1 to 2 (Tassia et al. 2017; Franchi et al. 2019) and in *C. intestinalis* from 3 to 5 (Buckley & Rast 2015; Tassia et al. 2017). For all novel TLRs identified, the revised TIR domain was exclusively predicted in previously unannotated space downstream of tandem LRR cassettes, not within a territory where it statistically outcompeted another high confidence, but unrelated, domain annotation. Thus, TIAMMAt's results yielded a domain architecture fitting the canonical schema for TLRs (Akira & Takeda 2004). A TIR domain was omitted in the original annotations of these TLRs for two different reasons. For *Lingula*'s and *Branchiostoma*'s novel TLRs, a TIR domain met HMMER's default reporting threshold prior to revision (per-domain and per-sequence e-values < 10.0). However, the domain did not meet the inclusion threshold requirement (per-sequence e-value < 0.01) to confidently be labeled as a statistically significant homolog. In

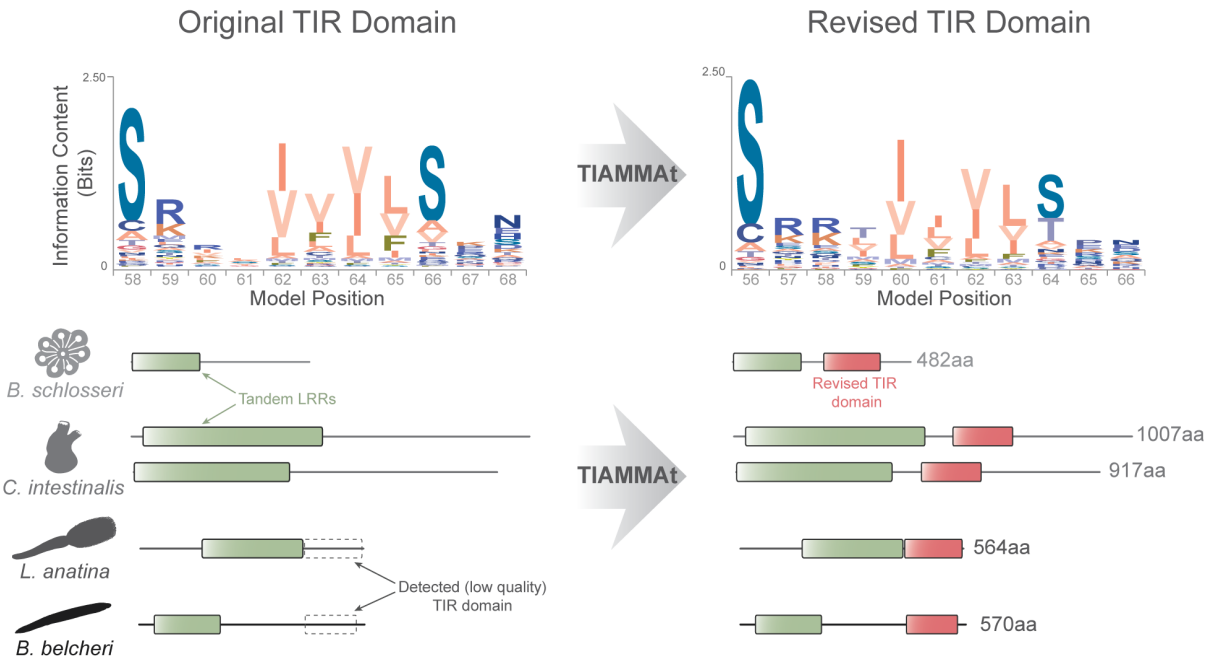


Figure 5. TIR domain model revision identifies previously unrecognized TLRs. Top: Skylign (Wheeler et al. 2014) graph for positions 58-68 of the original TIR domain model (PF01582; left) and after revision (right). Relative height of amino acid symbols reflects their emission probability relative to all other amino acid states at that site. Bottom: Domain diagrams of TLR structures before (left) and after (right) domain revision, highlighting the utility of incorporating taxonomic diversity into the TIR domain seed alignment when working with underrepresented taxa. Regarding domain diagrams, green bars represent tandem leucine-rich repeats, red bars represent TIR domains, dotted-outline empty bars represent TIR domains below inclusion threshold.

contrast, the novel tunicate TLRs lacked any reportable TIR domain prior to revision (**Fig. 5**), suggesting the newly identified tunicate proteins contain divergent TIR domains relative to sequences in the original seed alignment. Prior analyses have shown both of *Ciona*'s previously described TLRs act as a functional blend of several vertebrate homologs (Sasaki et al. 2009; Satake & Sekiguchi 2012). Notably, the divergent tunicate TLR may be causally tied to tunicate's rapid rate of molecular evolution relative to their sister phylum, Vertebrata (Berná & Alvarez-Valin 2014).

TIR domain revision also supported previous data (Gerdol et al. 2017) suggesting TIR-domain-containing (TIR-DC) proteins have experienced a high degree of evolutionary change across Metazoa. Several TIR-DC families possess notable taxonomic distributions with implications for TLR pathway evolution (**Supplementary Table 4; Supplementary Fig. 7**).

Stimulator of interferon genes (STING), an evolutionarily ancient facilitator of innate immunity responses against exogenous RNA and dsDNA (Wu et al. 2014), was reported to uniquely possess a TIR domain in several lophotrochozoan lineages (Gerdol et al. 2017), implicating an intersection between TLR- and STING-facilitated immunity. Our results corroborate these findings, reporting an additional TIR-DC STING protein in the nemertean, *Notospermus geniculatus*, and two more copies in the oyster, *Crassostrea virginica,* following TIR domain revision (**Supplementary Table 4**). Furthermore, whereas homologs to MYD88 and SARM1 (canonical TIR-DC adaptor proteins responsible for signal transduction and regulation of TLRs, respectively; O'Neill & Bowie 2007) possess ancestry predating the emergence of Vertebrata (Tassia et al. 2017; Toschachev & Neuwald 2020), many evolutionarily conserved TIR-DC proteins (defined in Gerdol et al. 2017) identified here lack any homologs within Vertebrata (**Supplementary Table 4**). Even when including proteomes from non-mammalian vertebrate lineages (i.e., hagfish, lamprey, and *Latimeria*) when running TIAMMAt, vertebrate TIR-DC proteins appear to be restricted to TLRs, IL-1Rs, and the five traditional TLR adaptors (O'Neill & Bowie 2007). Although there may be some relationship between the emergence of adaptive immunity and the limited number of TIR-DC protein structures within vertebrates, the non-canonical TIR-DC proteins identified across metazoan taxa may also represent a more flexible role for TIR domains outside the confines of the TLR pathway.

5.4.4 RIG-I-like Receptors

Revision of the RLR C-terminal domain (CTD), which is unique to three canonical RLR family members (retinoic acid-inducible gene, RIG-I; melanoma differentiation antigen 5, MDA5; and laboratory of genetic and physiology 2, LGP2; **Supplementary Fig. 1**; Esser-Nobis et al. 2020), revealed novel RLR proteins in the cnidarian, *Hydra vulgaris* (n=1 additional), and the sea star, *Patiria miniata* (n=2; **Supplementary Table 5**). Unlike canonical RLRs, novel proteins identified in *H. vulgaris* and *P. miniata* have atypical and individually distinct domain organizations. The novel protein identified in *Hydra* has a reversed architecture (with an N-terminal RLR "C-terminal domain"), an incomplete central helicase, and lacks CARD domains, similar in structure to the vertebrate LGP2 protein. In contrast, *Patiria'*s novel proteins both possess appropriately positioned C-terminal CTDs. However, one of the two newly identified

Patiria RLRs lacks a central helicase, the second possesses a duplicated CTD, and both possess a single N-terminal death effector domain (DED). Moreover, the novel domain architectures described above are not unique to the post-domain-revision dataset as several non-canonical RLR-related domain architectures (defined by the presence of the RLR-specific C-terminal domain) were detected across Metazoa even before domain revision. For example, *Hydra* possesses a second reversed RLR protein and *Hofstenia miamia* (a member of the clade, Xenacoelamorpha) possesses two reverse RLR proteins which, together with *Hydra'*s proteins, comprise a well-supported monophyletic orthology group (>90% posterior probability; **Supplementary Fig. 9**). Given that all canonical RLRs (i.e., RIG-I, MDA5, and LGP2) share a central DExD/H-box helicase and a CTD, which together give RLR's their RNA recognition capacities (Pippig et al. 2009; Jiang et al. 2011; Luo et al. 2011; Reikine et al. 2014), the proteins with incomplete helicases described in this paragraph provide an interesting opportunity to investigate the function of the RLR CTD independent of a proximal helicase.

We placed all RLRs identified in our study into a Bayesian phylogenetic framework to compare with previous phylogenetic hypotheses on RLR evolution and to expand RLR sampling to include the less conventional RLR family members described above (**Supplementary Fig. 9**). Concordant with previous studies (Mukherjee et al. 2013; Pugh et al. 2016), we resolve RIG-I and MDA5/LGP2 orthology groups with deep representation of deuterostome taxa, except tunicates which possess their own RLR orthogroup. Interestingly, an orthology group comprised of RLRs with N-terminal DED domains (including the two novel *Patiria* sequences described above) was recovered with maximal support. DED, like CARD, is a member of the death domain superfamily (Park et al. 2007). Independent of RLR signaling, DED operates through homotypic domain-domain interactions and is vital for the regulation of cell death, including interactions mediated by caspase-8 and -10 (Valmiki & Ramos, 2009; Riley et al. 2015; Man & Kanneganti 2016). Although they belong to the same superfamily, functional evidence has shown the CARDs of RLRs and the DED of caspase-8 are not functionally equivalent (Jiang et al. 2012), suggesting DED-containing RLRs present among invertebrates may function independently of the canonical RLR signaling pathway. Given the ancient origins of cell death regulation through DED-DED interactions among animals (Sakamaki et al. 2015; Man & Kanneganti 2016), the ubiquitous threat of viral infection (Forterre 2006), and the potential coupling of DED-dependent signaling to the dsRNA recognition

via RLRs containing an N-terminal DED, we hypothesize that RLRs possess additional family members among invertebrates which act through rapid DED-dependent apoptotic pathways.

5.4.5 Future Prospects of TIAMMAt

In our application of TIAMMAt on innate immunity protein families, we demonstrated the value of improving representation of non-model species in Pfam domain seed alignments. Strikingly, for each of the PRR signaling families we considered, protein domain architecture diversity appears to be underestimated across Metazoa even independent of domain revision (**Supplementary Fig. 7**), and the effect becomes more severe when directly accounting for homologous sequence variation in domains among non-model species. These findings are consistent with previous studies that highlight the value of leveraging underrepresented species to investigate protein family evolution (Zhang et al. 2012; Yuen et al. 2013; Gerdol et al 2017).

The design of TIAMMAt was stimulated through a combination of uncovering a lack of even phylogenetic representation within domain profile seed alignments (**Fig. 1**) and the inferential value of annotating uncharacterized proteins with Pfam domains (e.g, Hibino et al. 2006; Costa-Paiva et al. 2017; Gerdol et al. 2017; Tassia et al. 2017). As such, TIAMMAt is designed to be compatible with any taxonomic distribution and collection of Pfam domains of interest to the user. Given enough computational and proteomic resources, TIAMMAt could be applied, for example, to domain/protein evolution studies among all opisthokonts, eukaryotes, or even all organisms. In contrast, TIAMMAt can also be used to revise a domain based on a single genus, yielding a revised domain profile where the per-site amino acid emission probabilities are narrowed to be a strict representation of that genus' domain sequence variance. For example, where one study could revise all globin domains using a broadly sampled metazoan dataset to help identify and investigate oxygen transport protein evolution, another study could revise the same domains using only hydrothermal vent and/or cold seep species where oxygen-transport has become highly specialized (Hourdez & Lallier 2007). Though these two scenarios may focus on the same Pfam domains, the revised domains produced by TIAMMAt reflect different taxonomic assemblages and address two discrete evolutionary questions.

TIAMMAt can be flexibly inserted into bioinformatic pipelines where domain annotation plays a role in inferring function of uncharacterized proteins. As non-model and underrepresented species continue to be sequenced at an accelerated rate (David et al. 2017), TIAMMAt provides bioinformaticians the option to account for phylogenetic distance during protein domain annotation – being particularly valuable when data are generated with the direct intent of addressing protein family/pathway evolution. In studies which concentrate on protein families instead of newly sequenced species, TIAMMAt can be used in a fashion that is similar to our case study on innate immunity, shifting focus from the traditional biomedical model species to a broader comparative scope by leveraging biodiversity already available in public data repositories.

5.5 Materials & Methods

5.5.1 Input Dataset Acquisition

Protein sequence accessions for the 39 metazoan taxa used in this study are available in **Supplementary Table 3**. Species were chosen to represent a broad phylogenetic distribution across Metazoa with compensation for representation bias within the Pfam database (**Fig. 1**). Regarding the two species where protein sequence datasets were not directly downloadable at the time of acquisition (i.e., *Hofstenia miamia* and *Schmidtea mediterranea*), scaffolded genomes and accompanying protein models were used to generate a protein sequence dataset using *gffread* (https://github.com/gpertea/gffread). In the context of our study, we do not discriminate between protein sequences derived of direct protein sequencing (reviewed in Callahan et al. 2020) and those inferred through bioinformatic translation of nucleotide datasets. Similarly, we recognize each species' proteome is not reflective of the same degree of sequencing revision or protein annotation (David et al. 2019). As a result, proteomes belonging to deeply sequenced species, such as humans, encode a high number of isoforms per protein when compared with more enigmatic taxa (Uhlén et al. 2015). To compensate for uneven annotations across taxa, we make the assumption that all protein isoform predictions possess equal probability to be expressed and are functional. Importantly, because we employ proteomic datasets derived primarily of genome sequencing projects, assessments made in our study are at the level of unique protein species encoded within the genome (accounting for all modeled isoforms of a single gene), not the measure of genes present.

The domain profile HMMs and seeds associated with key innate immunity proteins are summarized in **Supplementary Table 1.** Particularly, we chose domains traditionally associated with TLRs (i.e., TIR & TIR_2 domains; Tassia et al. 2017), RLRs (i.e., RIG-I_C-RD & CARD domains; Liu et al. 2017), NLRs (i.e., NACHT & CARD domains; Elinav et al. 2011), IRFs (i.e., IRF & IRF3 domains; Nehyba et al. 2009), and NFkB (i.e., RHD domain; Hayden & Ghosh 2011). All domain models and their seeds were obtained from Pfam version 32.0 (El-Gebali et al. 2018). Phylogenetic breadth represented within seed alignments before and after revision are shown in **Supplementary Figs. 3 & 4**. Additional LRR annotation was supplemented with Interproscan's (version 5.26-65.0) Gene3D annotation (version 4.1.0; Lees et al. 2014) due to HMMER's difficulty for positively annotating boundaries between individual repeat cassettes (Pellegrini 2015; Mistry et al. 2020).

5.5.2 Database Bias

Pfam domain profile seed alignments were downloaded from the Pfam 32.0 FTP server on April 7th, 2020. The Pfam-A database, which is generated from HMMs constructed from the seed alignments, was also downloaded. Species codes were then extracted and aggregated from both Pfam-A and the seed alignments to get a count estimate of species representation in the seeds themselves, as well as how those seeds may contribute to representation (or lack thereof) in the full database (**Fig. 1**).

5.5.3 Domain Profile HMM Revision

TIAMMAt automates revision of Pfam domain models to capture homologous sequence diversity based upon taxonomic distribution provided by the user (**Fig. 2**, **Supplementary Fig. 2**). The program is written using open-source software packages and is publicly available via GitHub (https://github.com/mtassia/TIAMMAt.git). Looping through the individual domain profile HMMs compiled above, TIAMMAt begins by searching proteomes for a single domain signature using HMMER's *hmmsearch* (version 3.1b2; Eddy 2009) under either default reporting/inclusion thresholds (used in this study) or user-defined thresholds (**Supplementary Table 7**). For each target sequence reporting a hit to the target domain, the target sequence is isolated from its parent

proteome and scanned for all Pfam domains using *hmmscan*, again with default thresholds (used here) or user-defined values. TIAMMAt then parses *hmmscan*'s domain table output to identify the best-fit domain architecture per sequence. Specifically, TIAMMAt first omits any hits which do not meet the per-sequence and per-domain (both conditional and independent) E-value inclusion thresholds of 0.01 (or the value specified by the user). The remaining hits are then ranked in ascending order of per-domain conditional E-values (with a lower bound of zero) and filtered of overlapping annotations, always maintaining the better-scoring domain hit over an overlapping weaker-scoring hit. This annotation parsing schema produces a non-overlapping list of highest-confidence domains per sequence which must at least meet the per-domain E-value inclusion threshold. Notably, some sequences which reported a potential hit to the target domain during the *hmmsearch* step may not report the same target domain after filtering due to conditional statistics after including all other domains in the Pfam database. Such annotations are considered noise from the perspective of the program and are omitted from the following steps due to lack of statistical substantiation, avoiding incorporation of false positives during revision.

Following domain annotation and identification of sequences with a best-fitting target domain, TIAMMAt extracts all best-fitting domain targets from their parent sequences (e.g., all TIR domains found within the *Saccoglossus kowalevskii* proteome). All isolated domains and the domain's seed sequences are aligned to the relevant domain profile HMM using *hmmalign* with the optional *--trim* argument to trim nonhomologous residues – particularly those which may accumulate at the termini of the model. Next, TIAMMAt runs *hmmbuild* to generate a revised domain profile HMM from *hmmalign*'s output Stockholm alignment. After the domain model has been revised, TIAMMAt loops once more through *hmmsearch* and *hmmscan*, this time isolating sequences which possess either the base or revised domain model hits (which meet the same threshold requirements specified above).

Once all domains have been revised, TIAMMAt executes a final *hmmscan* using a Pfam database appended with all revised domain models from the current TIAMMAt run – permitting each sequence to be annotated with base or revised domains of all those considered which, until this point, had all been revised in isolation of one another. This step is particularly important if the domains being revised are, in combination, descriptive of a single protein family (e.g., NACHT and CARD domain revisions as they relate to NOD-like receptors). Post-revision datasets from

our immunity study (both sequences and markov models) are also available via the TIAMMAt github repository (https://github.com/mtassia/TIAMMAt.git).

We recommend using TIAMMAt only after careful consideration of input proteome dataset quality and completeness, such as using protein datasets derived of published genomes where such effects have been considered and explicitly controlled or performing genome quality assessments like BUSCO (Waterhouse et al. 2018) or BlobTools (Laetsch & Blaxter 2017).

5.5.4 Phylogenetic Methods

Each protein family was aligned using MAFFT version 7's L-INS-I protocol (Katoh & Standley 2013). Phylogenetic reconstruction was performed using IQ-TREE version 1.6.12 (Nguyen et al. 2015). We employ IQ-TREE's ModelFinder subprogram (Kalyaanamoorthy et al. 2017) to infer best-fit substitution models and the ultrafast bootstrap approximation method for node support (10,000 generations; Minh et al. 2013). Phylogenetic trees were initially visualized using the iTOL web server (Letunic & Bork, 2019) and all nodes with ultrafast bootstrap support less than 95% are collapsed and considered unsupported per IQ-TREE's statistical guidelines. Anchoring sequences were downloaded from the UniProt SwissProt database (The UniProt Consortium, 2019) in Fall 2020.

Bayesian phylogenetic reconstruction of RLR protein relationships was performed using ExaBayes version 1.5 (Aberer et al. 2014). Two independent runs of four Metropolis-coupled chains each were executed in parallel for 1 x 10^7 generations, sampled every 100 generations, using a γ-distributed rate heterogeneity, empirical amino acid state frequencies, and a fixed substitution model of VT, which was determined to be the best-fit amino acid substitution matrix via BIC by ModelFinder (Kalyaanamoorthy et al. 2017). Chain convergence was confirmed by the presence of average standard deviation of split frequencies < 0.01 and effective sample size per parameter ≥ 100. A majority-rule consensus tree was generated after discarding the first 25% of sampled Markov Chain Monte Carlo (MCMC) generations as burn-in and visualized using the iTOL web server (Letunic & Bork, 2019). Unedited tree files for both likelihood and Bayesian phylogenetic inferences from this study are available via the TIAMMAt github repository (https://github.com/mtassia/TIAMMAt.git).

5.6 Acknowledgements

This work was supported by The National Science Foundation (grant number IOS – 1755377 to KMH, Rita Graze and Elizabeth Hiltbold Schwartz), and KTD was supported by The National Science Foundation's Graduate Research Fellowship Program. Thank you to Elizabeth H. Schwartz, Rita Graze, Ryan Range, Jamie Oaks, and Nathan Whelan for discussing topics on individual protein families and code development. We would also like to thank all testers of TIAMMAt during development and helpful feedback from reviewers and editors. Computational resources were made available by the Alabama Supercomputer Authority and the Auburn University Hopper Cluster.

5.7 Data Availability Statement

TIAMMAt and results of this manuscript are publicly available and maintained through the TIAMMAt GitHub repository (https://github.com/mtassia/TIAMMAt.git). Regarding data analyzed in our study, taxonomic accessions and their source repositories are available via **Supplementary Table 3**, and the Pfam models are available via **Supplementary Table 1**.

5.8 References

Aberer AJ, Kobert K, Stamatakis A. 2014. ExaBayes: Massively parallel Bayesian tree inference for the whole-genome era. *Mol Biol Evol.* 31(10): 2553-2556.

Akira S, Takeda K. 2004. Toll-like receptor signaling. *Nat Rev Immunol* 4: 499-511.

Agrawal S, Adholeya A, Deshmukh SK. 2016. The pharmacological potential of non-ribosomal peptides from marine sponge and tunicates. *Front Immunol.* 25: doi: 10.3389/fphar.2016.00333

Anantharaman V, Makarova KS, Burroughs AM, Koonin EV, Aravind L. 2013. Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis, and RNA processing. *Biol Direct* 8(15)

Beutler B. 2004. Innate immunity: an overview. *Mol Immunol.* 40: 845-859.

Berná L, Alvarez-Valin F. 2014. Evolutionary genomics of fast evolving tunicates. *Genome Biol. Evol.* 6(7): 1724-1738.

Bosch TCG. 2013. Cnidarian-microbe interactions and the origin of innate immunity in metazoans. *Annu Rev Microbiol.* 67: 499-518.

Brennan JJ, Gilmore TD. 2018. Evolutionary origins of Toll-like receptor signaling. *Mol Biol Evol.* 35(7):1576-1587.

Buckley KM, Rast JP. 2012. Dynamic evolution of toll-like receptor multigene families in echinoderms. *Front Immunol.* 3(136) doi: 10.3389/fimmu.2012.00136

Buckley KM, Rast JP. 2015. Diversity of animal immune receptors and the origins of recognition complexity in the deuterostomes. *Dev Comp Immunol.* 49, 179-189.

Callahan N, Tullman J, Kelman Z, Marino M. 2020. Strategies for development of a next-generation protein sequencing platform. *Trends in biochemical sciences* 45(1): 76-89.

Cavalieri V, Spinelli G. 2014. Early asymmetric cues triggering the dorsal/ventral gene regulatory network of the sea urchin embryo. *eLife* 3: doi:10.7554/eLife.04664

Chen F, Mackey AJ, Vermunt JK, Roos DS. 2007. Assessing performance of orthology detection strategies applied to eukaryotic genomes. *PLoS ONE* 2(4):e383

Connahs H, Tlili S, van Creij J, Loo TYJ, Banerjee TD, Saunders TE, Monteiro A. 2019. Activation of butterfly eyespots by Distal-less is consistent with a reaction-diffusion process. *Development* 146: doi:10.1242/dev.169367

Costa-Paiva EM, Shrago CG, Halanych KM. 2017. Broad phylogenetic occurrence of oxygen-binding hemerythrins in bilaterians. *Genome Biol Evol.* 9: 2580-2591.

Costa-Paiva EM, Shrago CG, Coates CJ, Halanych KM. 2018. Discovery of novel hemocyanin genes in metazoans. *Biol Bull.* 235: 134-151

David KT, Wilson AE, Halanych KM. 2019. Sequencing disparity in the genomic era. *Mol Biol Evol.* 36(8): 1624-1627.

Eddy SR. 2009. A new generation of homology search tools based on probabilistic inference. *Genome Informatics* 23: 205-211.

Eddy SR. 2011. Accelerated Profile HMM Searches. *PLoS Comput Biol.* 7(10): e1002195

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A. 2018. The Pfam protein families database in 2019. *Nucleic Acids Res.* doi: 10.1093/nar/gky995.

Elinav E, Strowig T, Henao-Mejia J, Flavell RA. 2011. Regulation of the antimicrobial responses by NLR proteins. *Immunity* 34: 665-679.

Esser-Nobis K, Hatfield LD, Gale Jr M. 2020. Spatiotemporal dynamics of innate immune signaling via RIG-I-like receptors. *PNAS.* doi: 10.1073/pnas.1921861117.

Flajnik MF, Kasahara M. 2010. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat Rev Genet.* 11: 47-59.

Forterre P. 2006. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* 117, 5-16

Franchi N, Ballarin L, Peronato A, Cima F, Grimabaldi A, Girardello R, de Eguileor M. 2019. Functional amyloidogenesis in immunocytes from the colonial ascidian *Botryllus schlosseri*: Evolutionary perspective. *Dev Comp Immunol.* 90: 108-120.

Gehrke AR, Neverett E, Luo Y, Brandt A, Ricci L, Hulett RE, Gompers A, Ruby RG, Rokhsar DS, Reddien PW, Srivastava M. 2019. Acoel genome reveals the regulatory landscape of whole-body regeneration. *Science* 363, eauu6173.

Gerdol M, Venier P, Edomi P, Pallavicini A. 2017. Diversity and evolution of TIR-domain-containing proteins in bivalves and Metazoa: new insights from comparative genomics. *Dev Comp Immunol.* 70: 145-164.

Gerdol M, Luo Y, Satoh N, Pallavicini A. 2018. Genetic and molecular basis of the immune system in the brachiopod *Lingula anatina. Dev Comp Immunol.* 82: 7-30.

Halanych KM, Kocot KM. 2014. Repurposed transcriptomic data facilitate discover of innate immunity Toll-like receptor (TLR) genes across Lophotrochozoa. *Biol Bull.* 227: 201-209.

Hall MR, Kocot KM, Baughman KW, Fernandez-Valverde SL, Gauthier MEA, Hatleberg WL, Krishnan A, McDougall C, Motti CA, Shoguchi E, et al. 2017. The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature* 544, 231-234.

Hamada M, Shoguchi E, Shinzato C, Kawashima T, Miller DJ, Satoh N. 2013. The complex NOD-like receptor repertoire of the coral *Acropora digitifera* includes novel domain combinations. *Mol Biol Evol.* 30(1): 167-176.

Hayden MS, Ghosh S. 2011. NF-κB in immunobiology. *Cell Research* 21: 223-244.

Hernando D, Crespi V, Cybenko G. 2005. Efficient computation of the hidden Markov model entropy for a given observation sequence. *IEEE Transactions on Information Theory* 51(7): 2681-2685.

Hibino T, Loza-Coll M, Messier C, Majeske AJ, Cohen AH, Terwilliger DP, Buckley KM, Brockton V, Nair SV, Berney K, et al. 2006. The immune gene repertoire encoded in the purple sea urchin genome. *Dev Biol.* 300: 349-365.

Hiscott J. 2007. Convergence of the NF-κB and IRF pathways in the regulation of innate antiviral response. *Cytokine & Growth Factor Reviews* 18: 483-490.

Hourdez S, Lallier FH. 2007. Adaptations to hypoxia in hydrothermal-vent and cold-seep invertebrates. *Rev Environ Sci Biotechnol.* 6: 143-157

Huang S, Yan S, Guo L, Yanhong Y, Li J, Wu T, Liu T, Yang M, Wu K, Liu H, et al. 2015. Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome Res.* 18: 1112-1126.

Jiang F, Ramanathan A, Miller MT, Tang G, Gale Jr. M, Patel SS, Marcotrigiano J. 2011. Structural basis of RNA recognition and activation by innate immune receptor RIG-I. *Nature* 479: 423-429.

Jiang X, Kinch LN, Brautigam CA, Chen X, Du F, Grishin NV, Chen ZJ. 2012. Ubiquitin-induced oligomerization of the RNA sensors RIG-I and MDA5 activates antiviral innate immune response. *Immunity* 36, 959-973.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14, 587-589.

Kanneganti T, Lamkanfi M, Núñez G. 2007. Intracellular NOD-like receptors in host defense and disease. *Immunity* 27: 549-559.

Katoh K, Standly DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol.* 30(4): 772-780.

Kowalinski E, Lunardi T, McCarthy AA, Louber J, Brunel J, Grigorov B, Gerlier D, Cusack S. 2011. Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* 147: 423-435.

Kwon D, Yoon JH, Shin S, Jang T, Kim H, So I, Jeon J, Park HH. 2012. A comprehensive manually curated protein-protein interaction database for the death domain superfamily. *Nucleic Acids Res.* 40: D331-336

Laetsch DR, Blaxter ML. 2017. BlobTools: Interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. *F1000Research* 6: 1287

Lange C, Hemmrich G, Klostermeier UC, López-Quintero JA, Miller DJ, Rahn T, Weiss Y, Bosch TCG, Rosenstiel P. 2011. Defining the origins of the NOD-like receptor system at the base of animal evolution. *Mol Biol Evol.* 28(5): 1687-1702.

Lamkanfi M, Dixit VM. 2012. Inflammasomes and their role in health and disease. *Annu Rev Cell Dev Biol.* 28:137-161

Laroui H, Yan Y, Narui Y, Ingersoll SA, Ayyadurai S, Charania MA, Zhou F, Wang B, Salaita K, Sitaraman SV, et al. 2011. L-Ala-γ-D-Glu-meso-diaminopimelic acid (DAP) interacts directly with leucin-rich region domain of nucleotide-binding oligomerization domain 1, increasing phosphorylation activity of receptor-interacting serine/threonine-protein kinase 2 and its interaction with nucleotide-binding oligomerization domain 1. *J Biol Chem.* 286(35): 31003-31013.

Lasi M, David CN, Böttger A. 2010. Apoptosis in pre-Bilaterians: *Hydra* as a model. *Apoptosis* 15, 269-278.

Laumer CE, Fernández R, Lemer S, Combosch D, Kocot KM, Resgo A, Andrade SCS, STerrer W, Sørensen MV, Giribet G. 2019. Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proc R Soc B.* 286: doi:10.1098/rspb.2019.0831

Lechtenberg BC, Mace PD, Riedl SJ. 2014. Structural mechanisms in NLR inflammasome signaling. *Current Opinion in Structural Biol.* 29: 17-25.

Lees JG, Lee D, Studer RA, Dawson NL, Sillitoe I, Das S, Yeats C, Dessailly BH, Rentzsch R, Orengo CA. 2014. Gene3D: Multi-domain annotations for protein sequence and comparative genome analysis. *Nucleic Acids Res.* D1, D240-D245.

Letunic I, Bork P. 2019. Interactive Tree of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47(W1): W256-259

Leulier F, Lemaitre B. 2008. Toll-like receptors – taking an evolutionary approach. *Nat Rev Genet.* 9: 165-178.

Liu M, Liao W, Buckley KM, Yang SY, Rast JP, Fugmann SD. 2018. AID/APOBEC-like cytidine deaminases are ancient immune mediators in invertebrates. *Nat Commun.* 9, doi:10/1038/s41467-018-04273-x

Liu Y, Olagnier D, Lin R. 2017. Host and viral modulation of RIG-I-mediated antiviral immunity. *Front Immuol.* 7(662), doi: 10.3389/Immu.2016.00662.

Loewenstein Y, Raimondo D, Redfern OC, Watson J, Frishman D, Linial M, Oregno C, Thornston J, Tramantano A. 2009. Protein function annotation by homology-based inference. *Genome Biol.* 10(207), doi:10.1186/gb-2009-10-2-207.

Luo D, Ding SSC, Vela A, Kohlway A, Lindenbach BD, Pyle AM. 2011. Structural insights into RNA recognition by RIG-I. *Cell* 147: *409-422.*

Man SM, Kanneganti T. 2016. Converging roles of caspases in inflammasome activation, cell death, and innate immunity. *Nat Rev Immunol.* 16, 7-21.

Messier-Solek C, Buckley KM, Rast JP. 2010. Highly diversified innate receptor systems and new forms of animal immunity. *Semin Immunol.* 22(1): 39-47.

Meunier E, Broz P. 2017. Evolutionary convergence and divergence in NLR function and structure. *Trends Immunol.* 38(10): 744-757

Minh BQ, Nguyen MAT, Haeseler V. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol.* 30(5): 1188-1195

Mistry J, Cihuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. 2020. Pfam: The proteins families database in 2021. *Nucleic Acids Res.* doi:10.1093/narlgkaa913

Mo J, Boyle JP, Howard CB, Monie TP, Davis BK, Duncan JA. 2012. Pathogen sensing by nucleotide-binding oligomerization domain-containing protein 2 (NOD2) is mediated by direct binding to muramyl dipeptide and ATP. *J Biol Chem.* 287(27): 23057-23067.

Mukherjee K, Korithoski B, Kolaczkowski B. 2013. Ancient origins of vertebrate-specific innate antiviral immunity. *Mol Biol Evol.* 31(1):140-153.

Nehyba J, Hrdličková R, Bose HR. 2009. Dynamic evolution of immune system regulators: the history of interferon regulatory factor family. *Mol Biol Evol.* 26(11): 2539-2550.

Nguyen L, Schmidt HA, Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1): 268-274.

O'Neill LAJ, Bowie AD. 2007. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat Rev Immunol.* 7: 353-364.

Park HH, Lo Y, Lin S, Wang L, Yang JK, Wu H. 2007. The death domain superfamily in intracellular signaling of apoptosis and inflammation. *Annu Rev Immunol.* 25: 561-586.

Pellegrini M. 2015. Tandem repeats in proteins: prediction algorithms and biological role. *Front Bioeng Biotechnol.* 3(143): doi:10.3389/fbioe.2015.00143

Pippig DA, Hellmuth JC, Cui S, Kirchhofer A, Lammens K, Lammens A, Schmidt A, Rothenfusser S, Hopfner K. 2009. The regulatory domain of the RIG-I family ATPase LGP2 senses double-stranded RNA. *Nuc Acids Res.* 37(6): 2014-2025.

Pugh C, Kolaczkowski O, Manny A, Korinthoski B, Kolaczkowski B. 2016. Resurrecting ancestral structural dynamics of an antiviral immune receptor: adaptive binding pocket reorganization

repeatedly shifts RNA preference. *BMC Evol Biol.* 16(241). doi:10.1186/s12862-016-0818-6.

Reikine S, Nguyen JB, Modis Y. 2014. Pattern recognition and signaling mechanisms of RIG-I and MDA5. *Front Immunol.* 5(342), doi:10.3389/fimmu.2014.00342

Riley JS, Malik A, Holohan C, Longley DB. 2015. DED or alive: assembly and regulation of death effect domain complexes. *Cell Death and Disease* 6. doi: 10.1038/cddis.2015.213

Sakamaki K, Imai K, Tomii K, Miller DJ. 2015. Evolutionary analysis of caspase-8 and its paralogs: Deep origins of the apoptotic signaling pathways. *Bioessays* 37: 767-776.

Sasaki N, Ogasawara M, Sekiguchi T, Kusumoto S, Satake H. 2009. Toll-like receptors of the ascidian *Ciona intestinalis* prototypes with hybrid functionalities of vertebrate Toll-like receptors. *J Biol Chem.* 284: 27336-27343.

Satake H, Sakeiguchi T. 2012. Toll-like receptors of deuterostome invertebrates. *Front Immunol.* 3(34). doi:10.3389/fimmu.2012.00034

Simakov O, Marletaz F, Cho S, Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo D, Larsson T, Lv J, Arendt D, et al. 2013. Insights into bilaterian evolution from three spiralian genomes. *Nature* 493, 526-531.

Simakov O, Kwashima T, Marlétaz F, Jenkins J, Koyanagi R, Mitros T, Hisata K, Bredeson J, Shoguchi E, Gyoja F, et al. 2015. Hemichordate genomes and deuterostome origins. *Nature* 527, 459-465.

Sonnhammer ELL, Eddy SR, Durbin R. 1997. Pfam: A comprehensive database of protein domain families based on seed alignments. *Proteins: Structure, Function, and Bioinformatics* 28(3), 405-420.

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al. 2014. STRING v10: protein-protein interaction networks integrated over the tree of life. *Nucleic Acids Res.* 43: D447-452

Tassia MG, Whelan NV, Halanych KM. 2017. Toll-like receptor pathway evolution in deuterostomes. *PNAS* 114(27): 7055-7060.

The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506-D515.

Toshchakov VY, Neuwald AF. 2020. A survey of TIR domain sequence and structure divergence. *Immunogenetics:* 1-23

Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, et al. 2015. Tissue-based map of the human proteome. *Science* 347(6220): 1260419.

Valmiki MG, Ramos JW. 2009. Death effector domain-containing proteins. *Cell Mol Life Sci.* (66): 814-830.

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol.* 35(3), 543-548

Wheeler TJ, Clements J, Finn RD. 2014. Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics* 15, 7.

Wojcik J, Schächter V. 2001. Protein-protein interaction map inference using interacting domain profile pairs. *Bioinformatics* 17 (1), S296-S305.

Wu X, Wu F, Wang X, Wang L, Siedow JN, Zhang W, Pei Z. 2014. Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. *Nucleic Acids Res.* 42: 8243-8257.

Yuen B, Bayes JM, Degnan SM. 2013. The characterization of sponge NLRs provides insight into the origin and evolution of this innate immune gene family in animals. *Mol Biol Evol.* 31(1):106-120.

Zhang Q, Lenardo MJ, Baltimore D. 2017. 30 years of NF-κB: a blossoming of relevance to human pathobiology. *Cell* 168: 37-57.

Zhang X, Wang Z, Zhang X, Le MH, Sun J, Xu D, Cheng J, Stacey G. 2012. Evolutionary dynamics of protein domain architecture in plants. *BMC Evol Biol.* 12(6)

Zhao X, Wang Y, Chen L, Aihara K. 2008. Protein domain annotation with integration of heterogeneous information sources. *Proteins* 72: 461-473.

5.9 Supplementary Information

Supplementary Table 1. Domains examined with TIAMMAt and selected model statistics per-domain. All models obtained from Pfam-A version 32.0.

	Pfam Acc.	# Seqs in Seed	Model Len.	Effective Num. Seqs	Relative entropy	Information entropy	pRel E	Kulber-Liebach distance (from null)
Pre-Revision:								
CARD	PF00619	49	85	8.12	0.67	0.69	0.61	0.08
IRF	PF00605	88	105	1.76	0.59	0.52	0.52	0.07
IRF-3	PF10401	29	179	2.42	0.59	0.57	0.52	0.03
NACHT	PF05729	25	166	3.99	0.59	0.61	0.5	0.02
RHD_DNA_bind	PF00554	21	169	1.56	0.59	0.57	0.51	0.04
RIG-I_C-RD	PF11648	74	117	2.30	0.59	0.49	0.52	0.08
TIR	PF01582	23	176	3.10	0.59	0.57	0.5	0.05
TIR_2	PF13676	528	118	9.26	0.59	0.58	0.39	0.03
Post-revision:								
CARD	N/A	2010	79	14.66	0.72	0.74	0.62	0.08
IRF	N/A	360	104	2.5	0.59	0.52	0.52	0.08
IRF-3	N/A	239	175	4.48	0.59	0.55	0.49	0.03
NACHT	N/A	2756	157	10.07	0.59	0.59	0.47	0.03
RHD_DNA_bind	N/A	347	168	3.96	0.59	0.57	0.49	0.05
RIG-I_C-RD	N/A	263	110	3.37	0.59	0.49	0.51	0.09
TIR	N/A	1141	141	5.68	0.59	0.55	0.48	0.06
TIR_2	N/A	2189	107	8.33	0.59	0.56	0.41	0.04

Supplementary Table 2. Number of best-fit domains identified before revision (left value) and after revision (right value).

	CARD	IRF	IRF-3	NACHT	RHD_DNA_Bind	RIG-I_C-RD	TIR	TIR_2
"Early Diverging"								
Ctenophora								
Mnemiopsis leidyi	0/0	2/2	1/3	5/7	0/0	0/0	0/0	16/19
Porifera								
Amphimedon queenslandica	32/121	3/3	3/3	55/55	2/2	3/3	0/1	5/9
Placozoa								
Trichoplax adhaerens	42/67	2/2	0/0	10/17	0/0	0/0	0/2	14/16
Cnidaria								
Hydra vulgaris	11/11	6/6	3/3	65/83	3/3	1/2	0/5	19/19
Nematostella vectensis	11/11	5/5	1/1	23/28	3/3	3/3	2/7	15/17
Xenacoelamorpha								
Acoelomorpha								
Hofstenia miamia	8/13	1/1	0/0	20/22	3/5	0/2	0/1	37/37
Lophotrochozoa								
Platyhelminthes								
Schmidtea mediterranea	8/10	1/1	1/1	3/16	5/5	29/29	0/6	31/39
Mollusca								
Aplysia californica	26/31	7/7	5/5	11/23	5/5	5/5	28/62	57/70
Crassostrea virginica	195/231	10/10	1/1	50/67	7/9	10/10	131/242	178/247
Octopus vulgaris	34/34	8/8	2/2	9/9	4/4	1/1	20/26	21/26
Annelida								
Capitella teleta	179/192	3/3	0/0	128/140	3/3	3/3	43/76	50/69
Helobdella robusta	10/16	5/5	0/0	5/9	4/4	2/2	4/12	10/16
Nemertea								
Notospermus geniculatus	23/32	12/12	14/15	34/43	4/4	7/7	10/35	53/58
Brachiopoda								
Lingula anatina	165/187	4/4	3/3	31/40	3/3	8/8	40/84	78/86
Phoronida								
Phoronis australis	53/55	4/4	1/1	22/30	4/4	2/2	21/48	54/62
Ecdysozoa								
Nematoda								
Caenorhabditis elegans	3/3	0/0	0/0	1/3	0/0	3/3	0/1	9/9
Tardigrada								
Hypsibius dujardini	2/3	0/0	0/0	0/0	4/4	0/0	0/1	3/3
Arthropoda								
Limulus polyphemus	15/15	15/15	0/0	17/34	27/28	0/0	17/20	9/14
Daphnia pulex	1/1	0/0	0/0	24/24	4/4	0/0	4/8	6/7
Drosophila melanogaster	0/1	0/0	0/0	3/17	21/21	0/0	13/19	14/17
Deuterostomia								
Echinodermata								
Acanthaster planci	23/25	6/6	8/8	174/185	10/10	16/16	17/53	71/78
Patiria miniata	12/13	4/4	3/4	112/125	7/7	6/8	5/16	37/38
Apostichopus japonicus	16/18	3/3	3/3	249/294	4/4	2/2	8/13	22/23
Strongylocentrotus purpuratus	28/32	10/10	6/6	295/323	9/9	19/19	185/208	97/118
Lytechinus variegatus	14/17	3/3	6/6	62/79	3/3	3/3	2/9	51/55
Hemichordata								
Ptychodera flava	26/31	3/3	2/2	63/84	4/5	8/8	69/105	90/116
Saccoglossus kowalevskii	13/14	2/2	2/2	25/35	6/6	4/4	17/24	33/38
Schizocardium californicum	21/28	3/3	2/2	128/138	7/7	9/9	135/188	104/135
Cephalochordata								
Branchiostoma belcheri	135/140	16/16	11/11	88/107	7/7	6/6	61/102	90/107
Branchiostoma floridae	108/117	10/10	4/4	70/82	2/2	5/5	43/70	67/87
Tunicata								
Ciona intestinalis	7/8	15/15	13/13	37/44	7/7	5/5	15/20	12/14
Botryllus schlosseri	7/8	8/8	12/12	207/226	4/5	8/8	10/21	15/25
Oikopleura dioica	0/0	2/2	1/1	2/2	2/2	0/0	1/1	1/2
Vertebrata								
Eptratretus burgeri	36/39	8/8	8/8	31/31	27/28	4/4	16/24	8/9
Petromyzon marinus	38/46	8/8	6/6	6/8	8/10	1/1	18/26	12/12
Callorhinchus milii	45/45	16/16	12/12	81/83	16/16	3/3	19/24	8/9
Latemeria chalumnae	44/46	15/15	10/10	59/64	24/24	4/4	27/33	10/11
Danio rerio	131/131	23/23	16/16	406/416	29/29	5/5	47/52	7/8
Homo sapiens	162/167	50/50	53/53	175/187	82/82	4/4	133/137	8/14

Supplementary Table 3. Taxonomic dataset representation.

	Dataset Version	Source	Proteome Accession
"Early Diverging"			
Ctenophora			
Mnemiopsis leidyi	[Accessed June 2019]	NIH NHGRI	ML2.2.aa
Porifera			
Amphimedon queenslandica	v1.0	NCBI	GCF_000090795.1_v1.0_protein
Placozoa			
Trichoplax adhaerens	v1.0	NCBI	GCF_000150275.1_v1.0_protein
Cnidaria			
Hydra vulgaris	Hydra_RP_1.0	NCBI	GCF_000004095.1_Hydra_RP_1.0_protein
Nematostella vectensis	ASM20922v1	NCBI	GCF_000209225.1_ASM20922v1_protein
Xenacoelamorpha			
Acoelomorpha			
Hofstenia miamia	[Accessed June 2019]	Srivastava Lab	[hmi_genome + hmi_gene_annotation + *gffread -y -S*]
Lophotrochozoa			
Platyhelminthes			
Schmidtea mediterranea	[Accessed June 2019]	PlanMine	[dd_Smes_g4 + smes_v2_hconf_SMESG + *gffread -y -S*]
Mollusca			
Aplysia californica	AplCal3.0	NCBI	GCF_000002075.1_AplCal3.0_protein
Crassostrea virginica	C_virginica-3.0	NCBI	GCF_002022765.2_C_virginica-3.0_protein
Octopus vulgaris	ASM634580v1	NCBI	GCF_006345805.1_ASM634580v1_protein
Annelida			
Capitella teleta	Capca1	NCBI	GCA_000328365.1_Capca1_protein
Helobdella robusta	Helobdella robusta v1.0	NCBI	GCF_000326865.1_Helobdella_robusta_v1.0_protein
Nemertea			
Notospermus geniculatus	v2.0	OIST MGU	nge_genome_v2.0_prot
Brachiopoda			
Lingula anatina	v2.0	OIST MGU	lan_genome_v2.0_prot
Phoronida			
Phoronis australis	v2.0	OIST MGU	pau_genome_v2.0_prot
Ecdysozoa			
Nematoda			
Caenorhabditis elegans	WBcel235	NCBI	GCF_000002985.6_WBcel235_protein
Tardigrada			
Hypsibius dujardini	nHd_3.1	NCBI	GCA_002082055.1_nHd_3.1_protein
Arthropoda			
Limulus polyphemus	Limulus_polyphemus-2.1.2	NCBI	GCF_000517525.1_Limulus_polyphemus-2.1.2_protein
Daphnia pulex	v1.0	NCBI	GCA_000187875.1_V1.0_protein
Drosophila melanogaster	Release 6 plus ISO1 MT	NCBI	GCF_000001215.4_Release_6_plus_ISO1_MT_protein
Deuterostomia			
Echinodermata			
Acanthaster planci	OKI-Apl_1.0	NCBI	GCF_001949145.1_OKI-Apl_1.0_protein
Patiria miniata	V2.0	EchinoBase	Pmin_proteins_v2.0
Apostichopus japonicus	ASM275485v1	NCBI	GCA_002754855.1_ASM275485v1_protein
Strongylocentrotus purpuratus	Spur_5.0	NCBI	GCF_000002235.5_Spur_5.0_protein
Lytechinus variegatus	Version 2.2	EchinoBase	Lvar2.2-protein
Hemichordata			
Ptychodera flava	1.014	OIST MGU	Pfl_public_ver1.0.prot
Saccoglossus kowalevskii	Skow_1.1	NCBI	GCA000003605.1_Skow_1.1_protein
Schizocardium californicum	[Unreleased]	[Unreleased]	[Unreleased]
Cephalochordata			
Branchiostoma belcheri	Haploidv18h27	NCBI	GCF_001625305.1_Haploidv18h27_protein
Branchiostoma floridae	Version 2	NCBI	GCF_000003815.1_Version_2_protein
Tunicata			
Ciona intestinalis	KH	NCBI	GCF_000224145.3_KH_protein
Botryllus schlosseri	[Accessed June 2019]	ANISEED	botryllus_protein_fasta
Oikopleura dioica	ASM20953v1	NCBI	GCA_000209535.1_ASM20953v1_protein
Vertebrata			
Eptatretus burgeri	v3.2	ENSEMBL	Eptatretus_burgeri.Eburgeri_3.2.pep.all
Petromyzon marinus	v3.1	SIMRBASE	PMZ_v3.1_proteins
Callorhinchus milii	Callorhinchus_milii-6.1.3	NCBI	GCF_000165045.1_Callorhinchus_milii-6.1.3_protein
Latemeria chalumnae	LatCha1	NCBI	GCF_000225785.1_LatCha1_protein
Danio rerio	GRCz11	NCBI	GCF_000002035.6_GRCz11_protein
Homo sapiens	GRCh38.p12	NCBI	GCF_000001405.38_GRCh38.p12_protein

Supplementary Table 4. TIR-domain-containing protein family members identified before and after model revision. *

	TLR	MYD88	SARM1	IL-1R	STING	TRIF	ecTIR-DC 1/2/4	ecTIR-DC 3	ecTIR-DC 5	ecTIR-DC 6	ecTIR-DC 7	ecTIR-DC 9	ecTIR-DC 11	ecTIR-DC 12	ecTIR-DC 13	ecTIR-DC 14	Unclassified
"Early Diverging"																	
Ctenophora																	
Mnemiopsis leidyi		2					6						1		1		9
Porifera																	
Amphimedon queenslandica		2		2			1						1				3
Placozoa																	
Trichoplax adhaerens							5		1				2		2	1	6
Cnidaria																	
Hydra vulgaris							13										3
Nematostella vectensis	1			2			6		2			1			1		3
Xenacoelamorpha																	
Acoelomorpha																	
Hofstenia miamia	1						18		1			2	3		2		10
Lophotrochozoa																	
Platyhelminthes																	
Schmidtea mediterranea		3		6			14								1	4	6
Mollusca																	
Aplysia californica	49	1	1				16		1			1	5		2	3	8
Crassostrea virginica	163	2	2	3	17/19		64		1			1	3	3	11	10	36
Octopus vulgaris	17						7				4	1	4		1		6
Annelida																	
Capitella teleta	9	3	1		2		67						1		1	2	12
Helobdella robusta	6		1				8/9										1
Nemertea																	
Notospermus geniculatus	17	2		5	0/1		21		1			1	4		3	11	12
Brachiopoda																	
Lingula anatina	28/29	1	1	2	2		45/47		1			1	3	4	2	6	28
Phoronida																	
Phoronis australis	30	1	1	1			22	1	5		7	1			2	1	11
Ecdysozoa																	
Nematoda																	
Caenorhabditis elegans	1		8														0
Tardigrada																	
Hypsibius dujardini	1		1												1		0
Arthropoda																	
Limulus polyphemus	17	6					3								1		4
Daphnia pulex	6		1				4										0
Drosophila melanogaster	16		8				3										0
Deuterostomia																	
Echinodermata																	
Acanthaster planci	7	2	3	17			12/13		1		15/18	2	9				20
Patiria miniata	1	4	1	6							2	1			2		16
Apostichopus japonicus	4	1	1	4			7	1							3		10
Strongylocentrotus purpuratus	162	10	5	9/11			38/39	4	1		0/1	3	7		4		47
Lytechinus variegatus	1	2	2				25/26	2							10		16
Hemichordata																	
Ptychodera flava	35		1	10			57/60				3				6		56/58
Saccoglossus kowalevskii	7	3	1	1			16					1	1		1		19
Schizocardium californicum	87	3		14			58/59				4		1			1	79
Cephalochordata																	
Branchiostoma belcheri	48/49	4/6	2	2			35/36		1		5	1				1	56
Branchiostoma floridae	16	4/5	2				37/39		1		4		2			1	49
Tunicata																	
Ciona intestinalis	3/5	1		5			9		4								1
Botryllus schlosseri	1/2		1	1			24/30		1								0
Oikopleura dioica	1						1										0
Vertebrata																	
Eptatretus burgeri	9	5		7			3										0
Petromyzon marinus	11	1		1			18										3
Callorhinchus milii	12	1	1	7		3	3										0
Latemeria chalumnae	14	2	1	12		1	4										3
Danio rerio	33	1	1	16			2										1
Homo sapiens	55	3/5	1	72		1	8										1

*All ecTIR-DC Families are defined in Gerdol et al. 2017. Families 8, 10, and 15 removed as no members are identifiable exclusively by Pfam domain architecture. Values shown as [# Before]/[# After]; single values indicate no new sequences identified after domain revision.

Supplementary Table 5. RLR CTD-containing protein family members identified before and after model revision.*

	Orphan RIG-I CTD	Incomplete Helicase	RIG-I (DDX58) / MDA5 (IFIH1)	LGP2 (DHX58)	Single CARD RLR	DEATH-RLR	DED-RLR	Unclassified
"Early Diverging"								
Ctenophora								
Mnemiopsis leidyi								0
Porifera								
Amphimedon queenslandica		1				2		0
Placozoa								
Trichoplax adhaerens								0
Cnidaria								
Hydra vulgaris		1/2						0
Nematostella vectensis		1		2				0
Xenacoelamorpha								
Acoelomorpha								
Hofstenia miamia		2						0
Lophotrochozoa								
Platyhelminthes								
Schmidtea mediterranea	3	7		11	3			4
Mollusca								
Aplysia californica	1	3		1				0
Crassostrea virginica			8	2				0
Octopus vulgaris			1					0
Annelida								
Capitella teleta			1	2				0
Helobdella robusta				2				0
Nemertea								
Notospermus geniculatus	1	2			2			2
Brachiopoda								
Lingula anatina			2	2			1	3
Phoronida								
Phoronis australis				1			1	0
Ecdysozoa								
Nematoda								
Caenorhabditis elegans								0
Tardigrada								
Hypsibius dujardini								0
Arthropoda								
Limulus polyphemus								0
Daphnia pulex								0
Drosophila melanogaster								0
Deuterostomia								
Echinodermata								
Acanthaster planci			7	6	1		2	0
Patiria miniata		1/3		1	1			3
Apostichopus japonicus	1	1						0
Strongylocentrotus purpuratus	1	4	1	7	1	2	3	0
Lytechinus variegatus	1	2						0
Hemichordata								
Ptychodera flava		2		1	1			4
Saccoglossus kowalevskii			1		2			1
Schizocardium californicum		4	2		1		1	1
Cephalochordata								
Branchiostoma belcheri				3			1	1
Branchiostoma floridae				3			1	1
Tunicata								
Ciona intestinalis			4					1
Botryllus schlosseri	1	4						3
Oikopleura dioica								0
Vertebrata								
Eptratretus burgeri				4				0
Petromyzon marinus								1
Callorhinchus milii			2	1				0
Latemeria chalumnae		1	3					0
Danio rerio			2	2				1
Homo sapiens			2	2				0

*Values shown as [# Before]/[# After]; single values indicate no new sequences identified after domain revision.

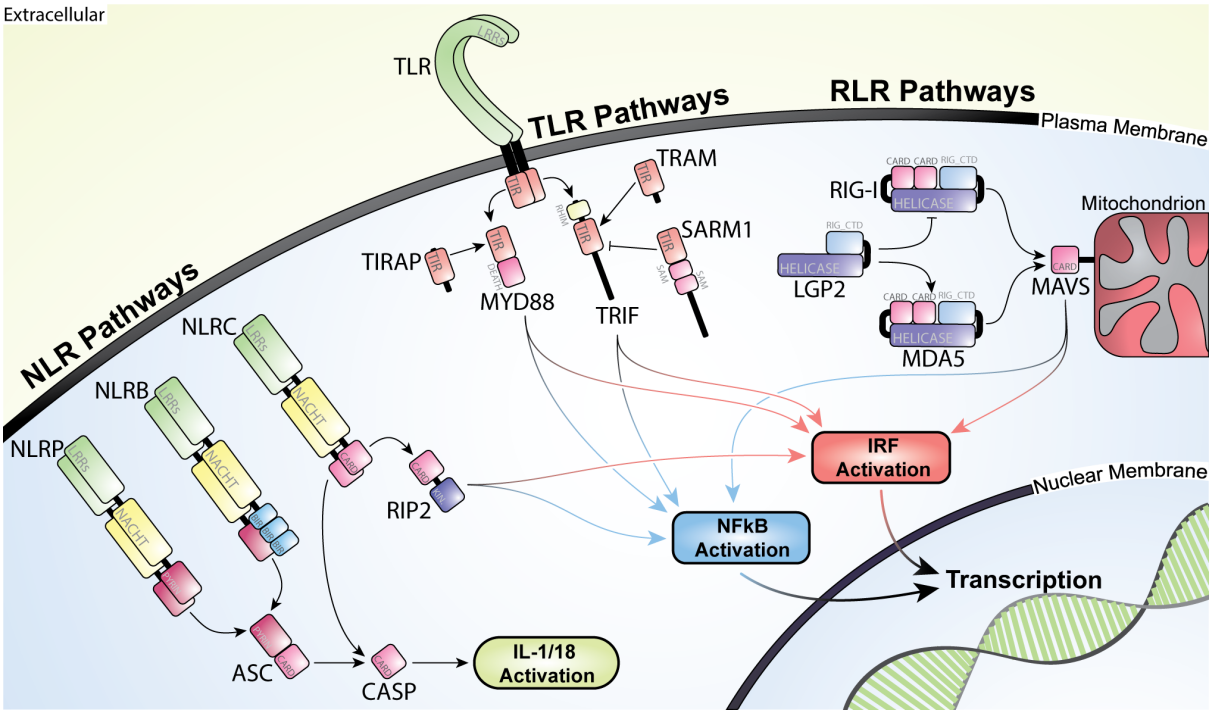
Supplementary Table 6. NACHT-containing protein family members identified before and after model revision.*

	Orphan NACHT	NLRB	NLRC	Tunicate_NLRC	zNLRC	NLRP	NLR_C-term	Unclassified NLR	TEP1	ANKRD50	TANC1	NWD1	Death Family NACHT	Unclassified NACHT-DC
"Early Diverging"														
Ctenophora														
Mnemiopsis leidyi	1/2								1					3/4
Porifera														
Amphimedon queenslandica	20			9			9		2			2	6	7
Placozoa														
Trichoplax adhaerens	9/10								0/1					1/6
Cnidaria														
Hydra vulgaris	29/36								0/1				29/31	7/14
Nematostella vectensis								4	1				1	17/22
Xenacoelamorpha														
Acoelomorpha														
Hofstenia miamia	5							4	1	1		1		8/10
Lophotrochozoa														
Platyhelminthes														
Schmidtea mediterranea	0/3							2		0/2				1/9
Mollusca														
Aplysia californica	1/3						0/1	1	4	0/1		0/1		5/12
Crassostrea virginica								7	2	0/2		1		40/55
Octopus vulgaris	1						1		1					6
Annelida														
Capitella teleta	28/33		44				17	3	1			2	20	13/20
Helobdella robusta	2/5													3/4
Nemertea														
Notospermus geniculatus	8/11						1	1	2	1		0/1	1	20/25
Brachiopoda														
Lingula anatina	4/5						1	5	1	1		2/3	2/3	15/21
Phoronida														
Phoronis australis	3/5							1		1		1	4	11/17
Ecdysozoa														
Nematoda														
Caenorhabditis elegans									1			0/2		1
Tardigrada														
Hypsibius dujardini														0
Arthropoda														
Limulus polyphemus	4/6						8		1	3/13				1/6
Daphnia pulex	19						2			3				0
Drosophila melanogaster	2									0/9		0/4		1/2
Deuterostomia														
Echinodermata														
Acanthaster planci	10/12			48			28	33/34	1	0/1	0/4	7	11	36/39
Patiria miniata	13/16			24/25			29/31	11/13		0/2	0/1	3	6	26/28
Apostichopus japonicus	184/203			1			5/6	2		0/1		0/1	0/1	70/79
Strongylocentrotus purpuratus	135/137			28			46/47	3	1	0/16		1	42/43	39/47
Lytechinus variegatus	21/34			4			6			0/1			15	16/19
Hemichordata														
Ptychodera flava	20/25		0/1				6/7	6			0/1	6	8/9	17/29
Saccoglossus kowalevskii	6									0/1	0/1	1		18/26
Schizocardium californicum	14/16		0/1				13	43	2	0/1		1	35	20/25
Cephalochordata														
Branchiostoma belcheri	8/9		1	27			11	13	1	0/8		0/1	5/7	22/29
Branchiostoma floridae	6/8		1	11			7	12	1			1/2	3/6	28/34
Tunicata														
Ciona intestinalis	20/26						13	1	0/1*					3
Botryllus schlosseri	196/212						11		0/1					0/2
Oikopleura dioica	1													1
Vertebrata														
Eptatretus burgeri	11		6				7						2	5
Petromyzon marinus	2		1				1						1	1/3
Callorhinchus milii	44/45		3				28	1		0/1				5
Latemeria chalumnae	6	1	3			10	25	1		1	0/3	6		6/8
Danio rerio	104/106		7		125	28	71	3	2		0/6	1	48	5/7
Homo sapiens	15	3	39			56	42		4	0/2	0/10	10		6

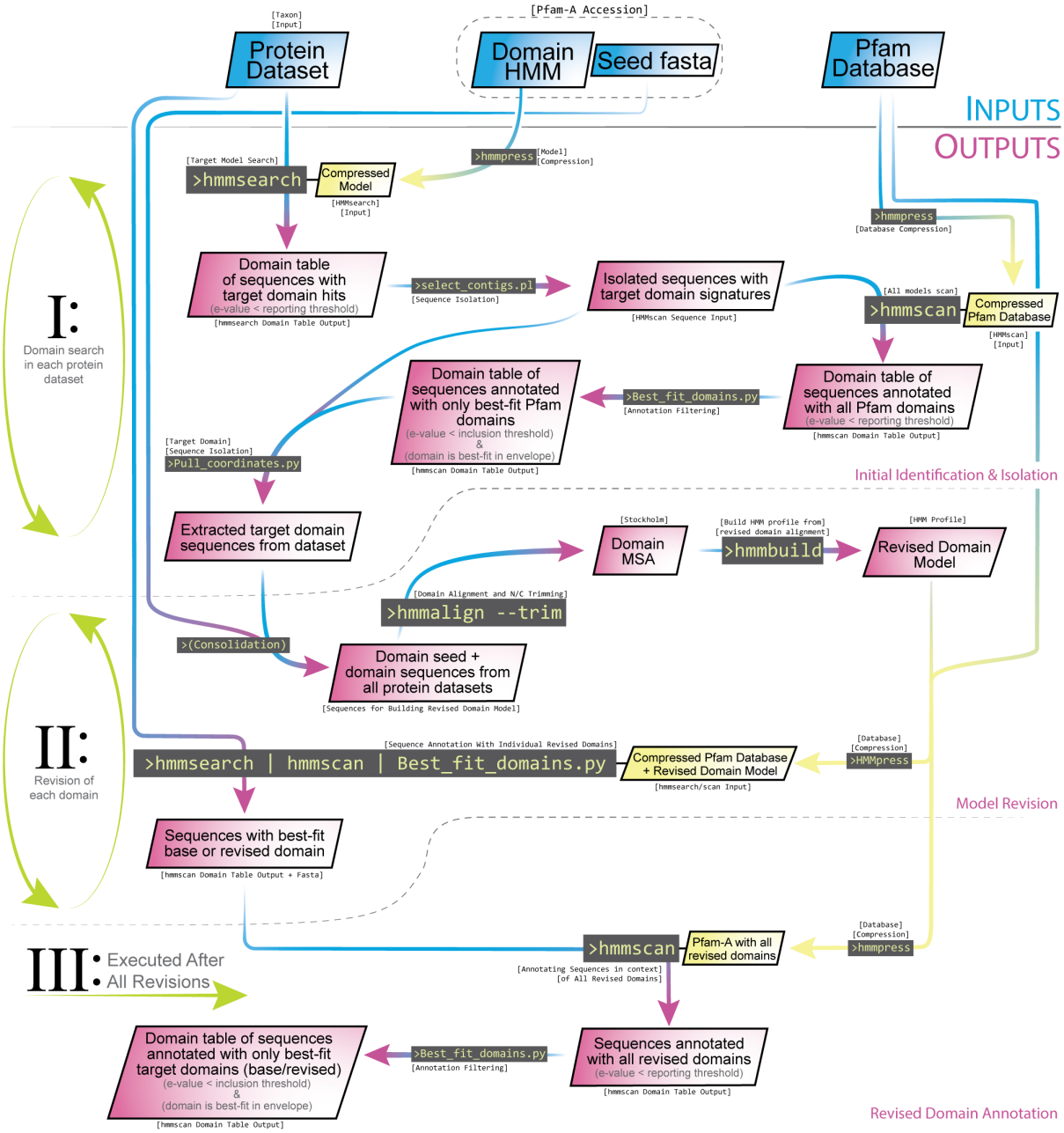
*NLRAs are omitted from count as they cannot be defined by Pfam domains alone. Values shown as [# Before]/[# After]; single values indicate no new sequences identified after domain revision.

Supplementary Table 7. Number of domains reported by *hmmsearch* identified before revision (left value) and after revision (right value).

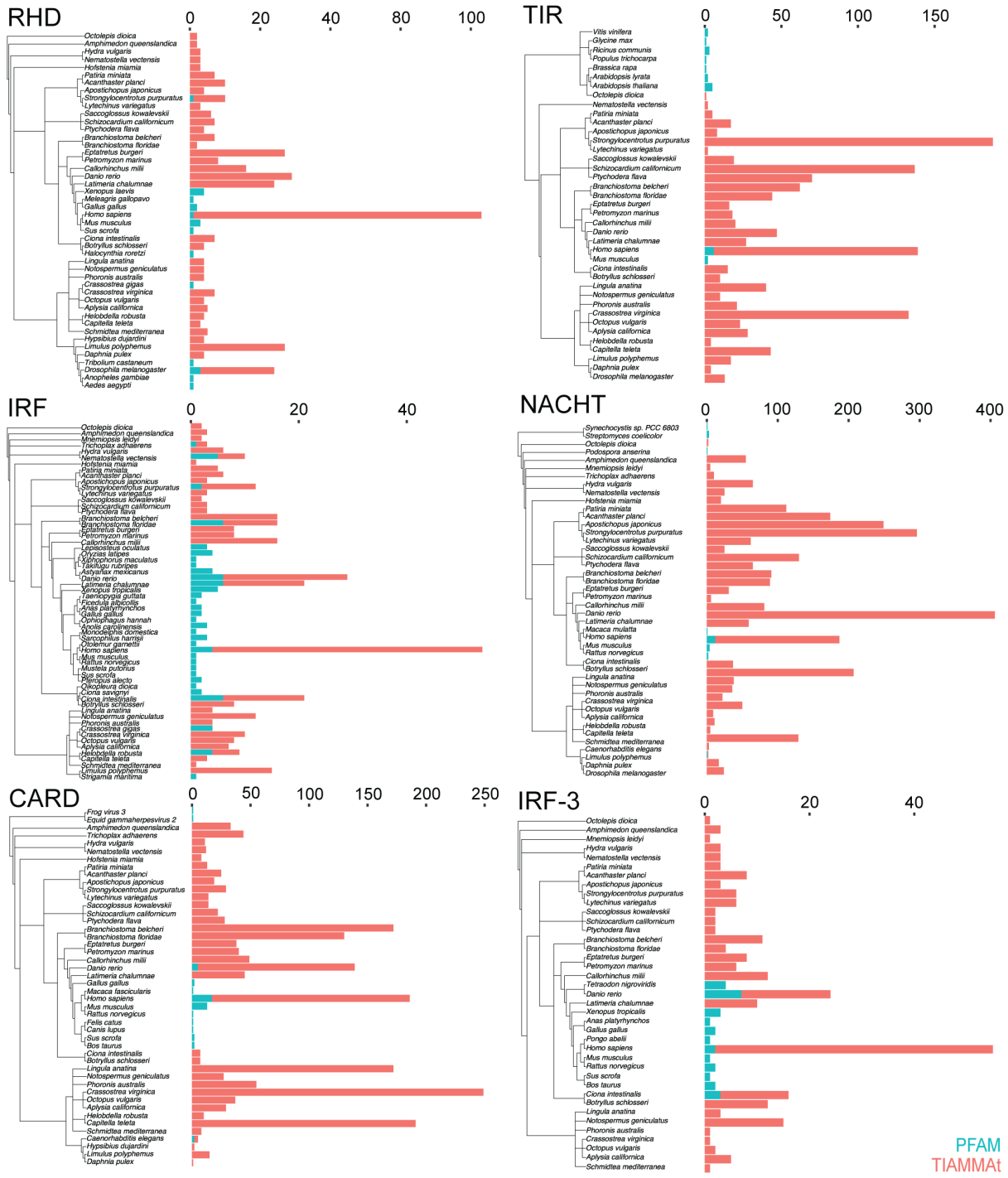
	CARD	IRF	IRF-3	NACHT	RHD_DNA_Bind	RIG-I_C-RD	TIR	TIR_2
"Early Diverging"								
Ctenophora								
Mnemiopsis leidyi	2/7	3/3	7/8	88/134	0/0	7/9	12/15	18/21
Porifera								
Amphimedon queenslandica	96/242	5/5	8/13	219/318	2/2	5/8	6/8	11/15
Placozoa								
Trichoplax adhaerens	48/71	2/3	1/3	104/175	1/1	2/4	12/17	19/27
Cnidaria								
Hydra vulgaris	14/17	6/8	7/7	158/230	4/4	2/4	11/14	26/31
Nematostella vectensis	16/19	6/8	3/4	148/204	3/3	5/7	13/15	21/23
Xenacoelamorpha								
Acoelomorpha								
Hofstenia miamia	14/21	1/2	3/3	118/183	5/7	7/10	28/32	38/44
Lophotrochozoa								
Platyhelminthes								
Schmidtea mediterranea	16/24	1/2	5/5	192/312	5/5	33/36	27/35	35/44
Mollusca								
Aplysia californica	36/44	9/9	7/7	170/266	6/7	8/8	79/90	91/94
Crassostrea virginica	230/269	18/20	22/29	581/763	10/12	21/23	288/304	335/347
Octopus vulgaris	36/48	8/11	5/6	112/197	4/4	8/11	34/45	42/43
Annelida								
Capitella teleta	195/206	4/4	3/4	237/298	3/4	4/5	85/100	99/107
Helobdella robusta	16/20	5/6	2/3	84/138	5/5	5/9	14/19	17/19
Nemertea								
Notospermus geniculatus	40/50	13/13	16/17	184/279	4/4	13/15	50/75	70/85
Brachiopoda								
Lingula anatina	186/214	6/7	8/9	172/250	5/5	11/15	102/121	130/131
Phoronida								
Phoronis australis	61/66	4/5	6/7	121/177	5/5	3/4	63/77	77/83
Ecdysozoa								
Nematoda								
Caenorhabditis elegans	4/5	0/0	1/7	90/149	0/0	6/9	3/9	11/14
Tardigrada								
Hypsibius dujardini	2/3	0/0	7/12	64/110	4/5	2/8	2/3	6/6
Arthropoda								
Limulus polyphemus	16/30	15/16	15/21	166/305	27/30	0/0	24/32	44/51
Daphnia pulex	1/7	1/2	1/3	130/167	5/5	3/5	7/10	12/14
Drosophila melanogaster	1/6	6/6	6/10	147/259	21/22	7/9	19/26	27/27
Deuterostomia								
Echinodermata								
Acanthaster planci	31/54	7/12	20/20	383/481	10/10	25/25	74/87	96/103
Patiria miniata	14/17	5/6	8/9	258/325	7/7	15/20	28/32	49/49
Apostichopus japonicus	20/25	4/5	8/8	360/438	5/5	5/10	25/30	33/38
Strongylocentrotus purpuratus	35/43	11/11	16/18	493/627	9/9	21/24	254/272	287/299
Lytechinus variegatus	14/30	5/6	7/7	259/360	3/5	4/4	32/50	56/59
Hemichordata								
Ptychodera flava	39/52	3/4	5/8	167/233	5/7	10/11	123/159	166/180
Saccoglossus kowalevskii	21/26	2/3	9/10	125/184	7/7	5/11	39/44	52/54
Schizocardium californicum	39/51	3/5	6/9	242/329	9/9	11/13	203/230	251/258
Cephalochordata								
Branchiostoma belcheri	154/181	16/16	24/26	250/337	8/8	6/8	136/158	162/180
Branchiostoma floridae	122/136	10/10	6/7	177/231	2/3	6/17	88/100	128/140
Tunicata								
Ciona intestinalis	12/20	15/15	18/18	127/211	8/9	10/12	23/26	30/33
Botryllus schlosseri	11/15	14/15	15/16	306/378	5/7	10/11	24/37	37/43
Oikopleura dioica	1/2	3/5	4/7	65/117	2/2	2/3	2/2	6/6
Vertebrata								
Eptatretus burgeri	41/42	9/9	18/18	148/223	28/29	9/13	23/24	29/36
Petromyzon marinus	48/55	10/10	7/10	67/97	9/11	6/8	25/30	34/36
Callorhinchus milii	54/71	17/18	19/24	196/279	16/19	10/13	23/27	32/32
Latemeria chalumnae	57/64	18/20	26/36	205/318	25/62	17/18	36/38	52/60
Danio rerio	138/166	23/26	34/43	651/824	29/29	13/21	55/59	60/67
Homo sapiens	177/197	58/62	105/116	688/1015	82/83	25/29	141/146	168/168



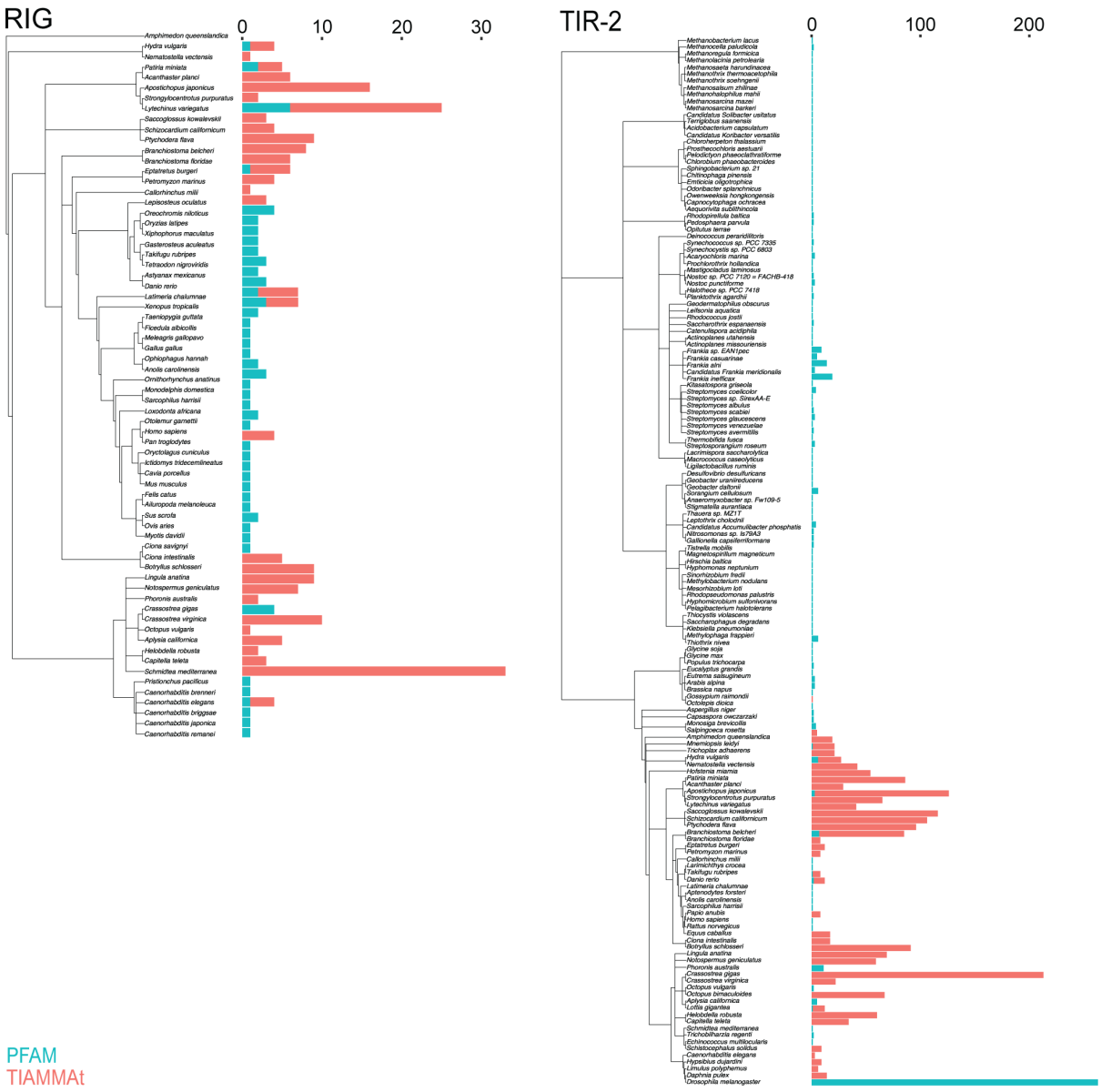
Supplementary Figure 1. Diagram of NLR, TLR, and RLR signaling pathways with all domains labeled. (Left) NLRs are cytoplasmically localized and possess a C-terminal series of leucine-rich repeats responsible for ligand binding, and a central NACHT domain involved in oligomerization and activation (Lechtenberg et al. 2014). NLR subfamilies differ in their N-terminal domain(s) which promote transcription factor activation or inflammasome assembly (Meunier & Broz, 2017). (Middle) TLRs are type-I transmembrane proteins localized to cell or endosomal membranes. Their N-terminal leucine-rich repeats bind pathogen-associated moieties, and the C-terminal TIR domain undergoes homotypic TIR domain interactions with one of five TIR-domain-containing adaptor proteins (Akira & Takeda 2004). (Right) RLRs are cytoplasmically localized and are exclusively involved in nucleic acid sensing. The central helicase and C-terminal regulatory domain are involved in ligand binding and autoregulation, whereas the N-terminal CARD domains are involved in signal transduction (Reikine et al. 2014). All three pathways converge on the activation of NF-κB and IRF activation, transcription factors which promote the expression of host-defense compounds like pro-inflammatory cytokines and antiviral peptides, respectively.



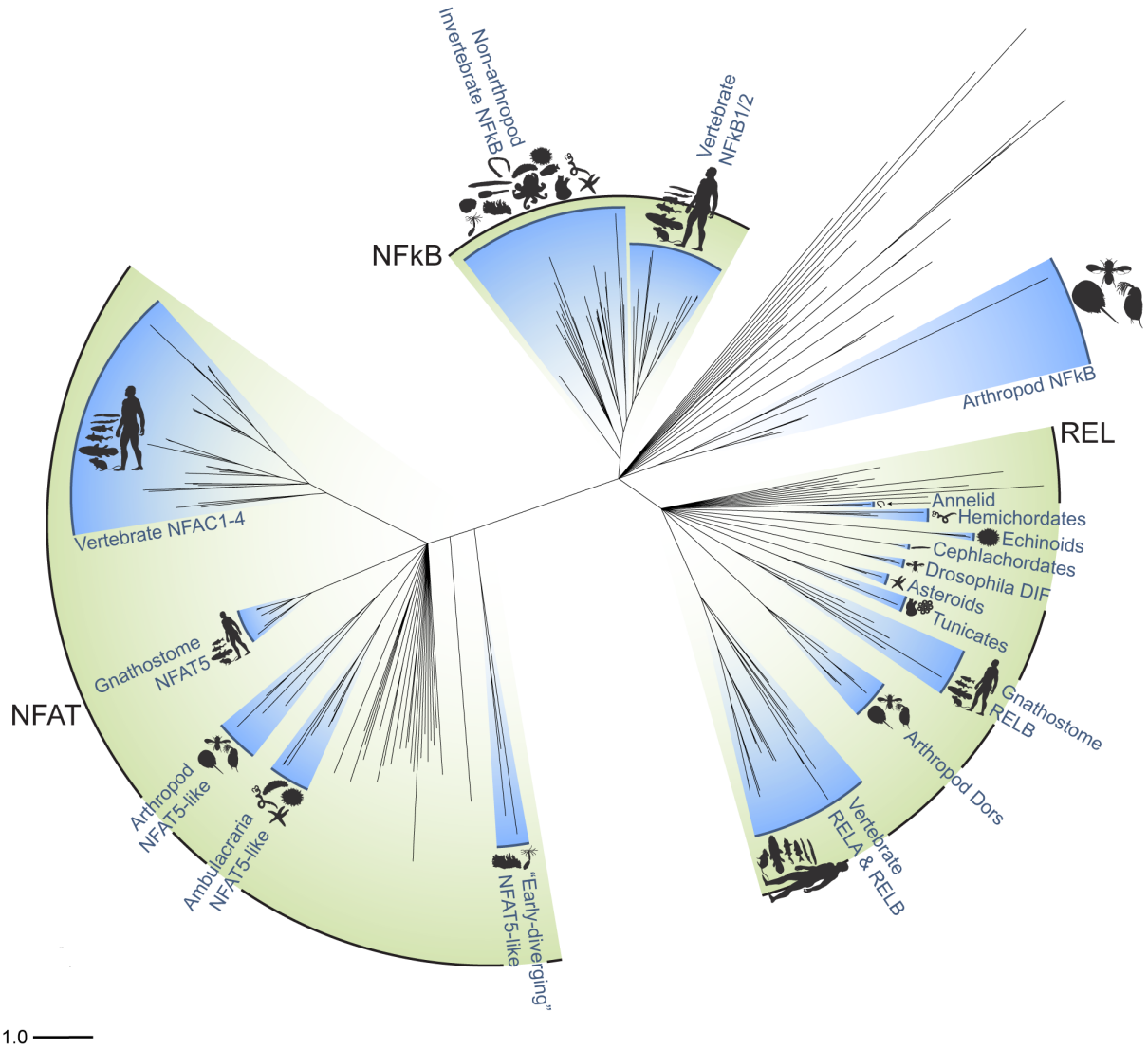
Supplementary Figure 2. Detailed schematic of commands and data analysis performed by TIAMMAt (see **Materials & Methods** for further detail).



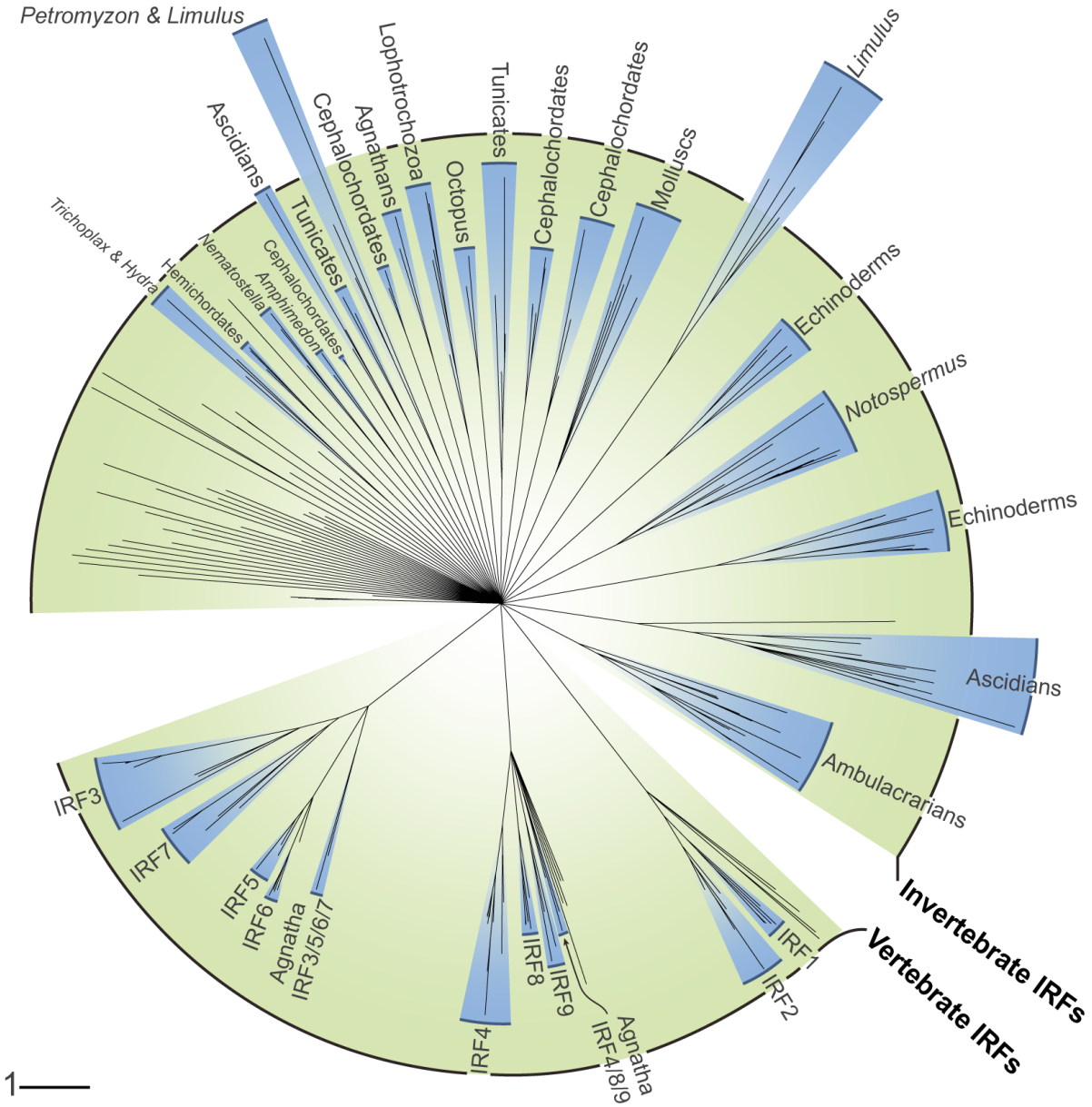
Supplementary Figure 3. Bar plots showing the number of sequences per species represented in a domain seeds of RHD, TIR, IRF, NACHT, CARD, and IRF-3 before (blue) and after (red) domain revision by TIAMMAt. Blue and red bars are superimposed, and their scales are displayed above each individual plot. Phylogenetic relationships of species represented in each domain seed were derived of their NCBI taxonomy.



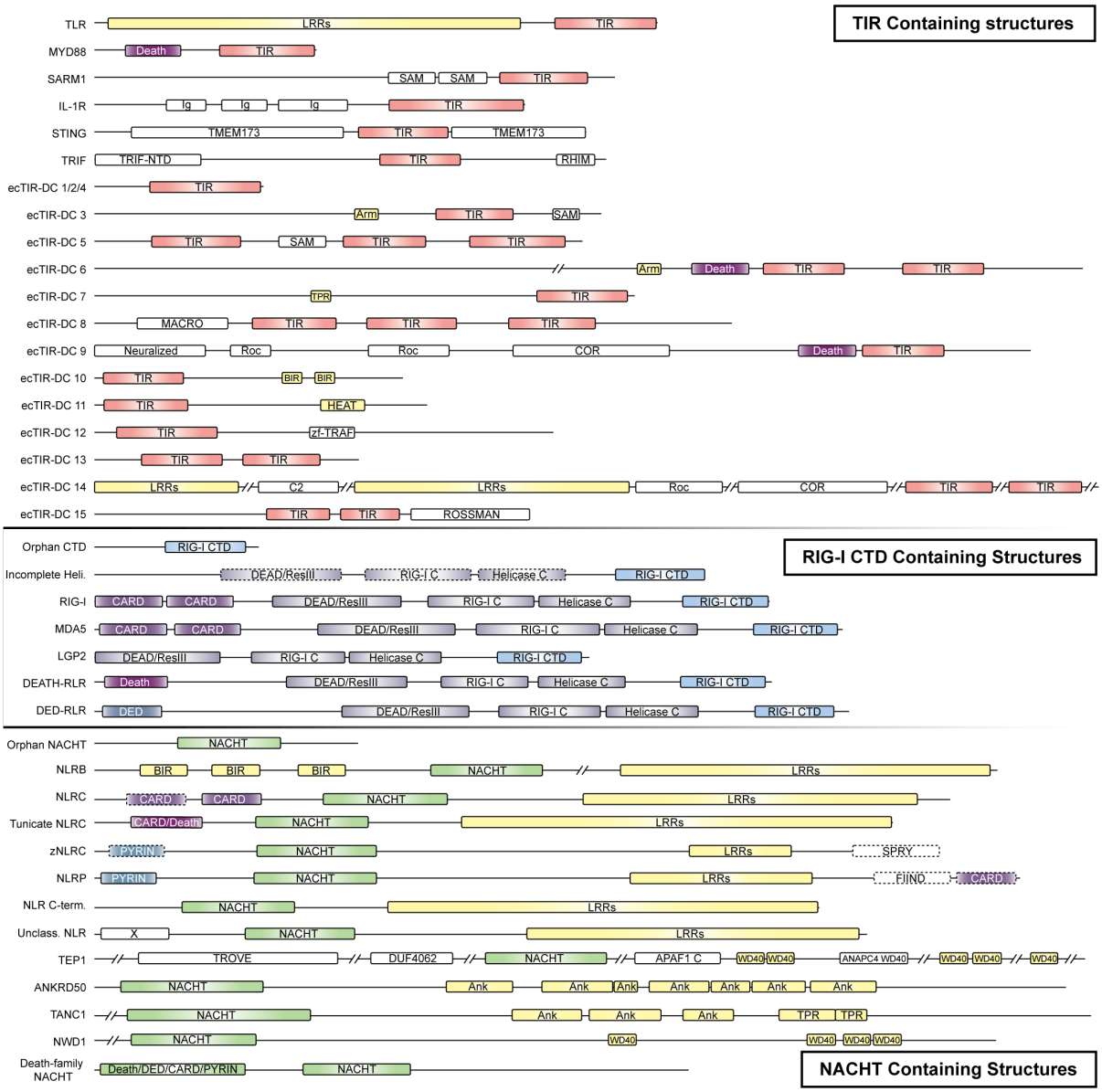
Supplementary Figure 4. Bar plots showing the number of sequences per species represented in a domain seeds of RIG and TIR-2 before (blue) and after (red) domain revision by TIAMMAt. Blue and red bars are superimposed, and their scales are displayed above each individual plot. Phylogenetic relationships of species represented in each domain seed were derived of their NCBI taxonomy.



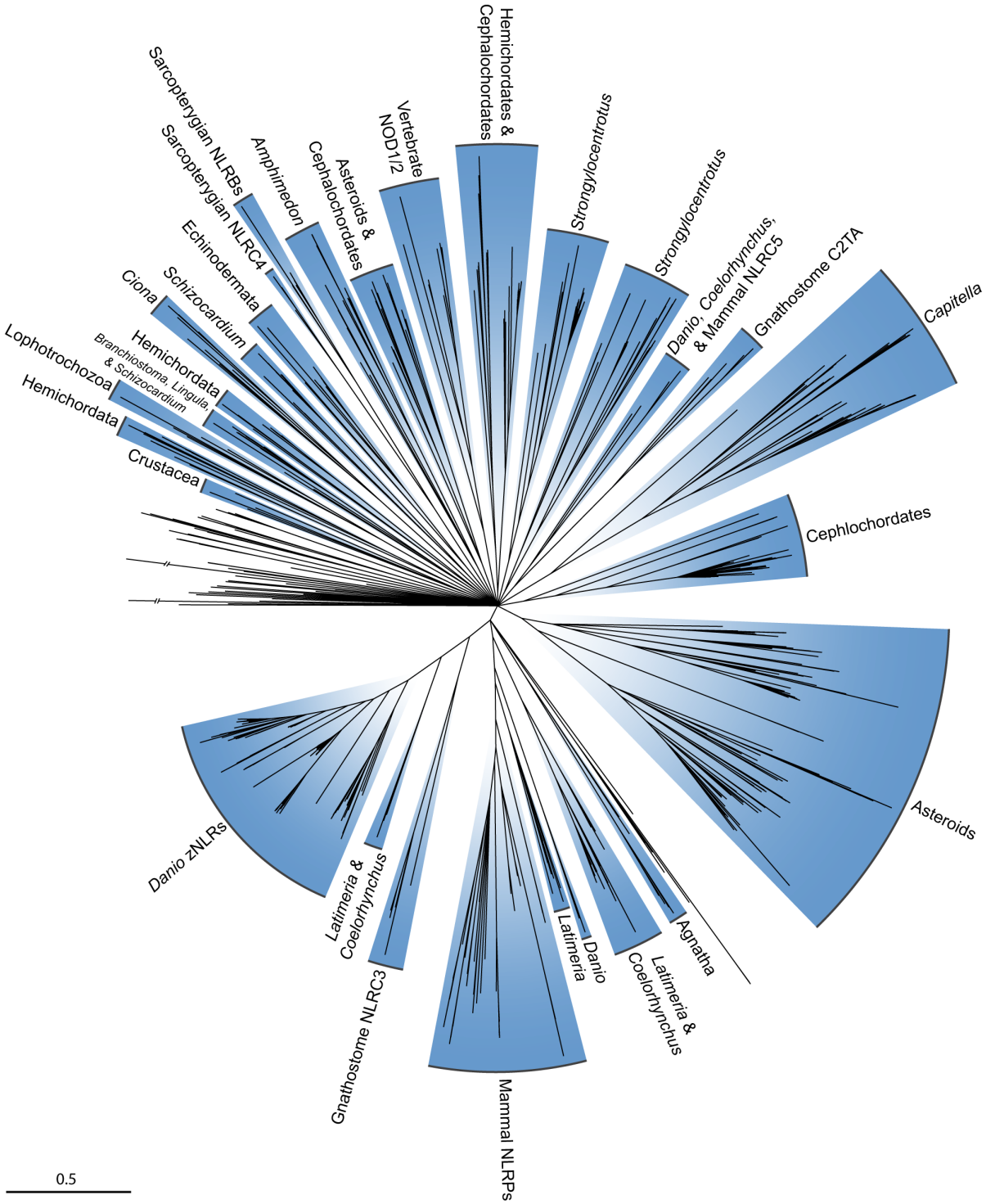
Supplementary Figure 5. Maximum-likelihood phylogenetic reconstruction of NF-κB family proteins using IQ-TREE. All nodes possess ultrafast-bootstrap support ≥95%. Best-fit model by BIC: VT+F+R6. Scale bar in number of substitutions per site.



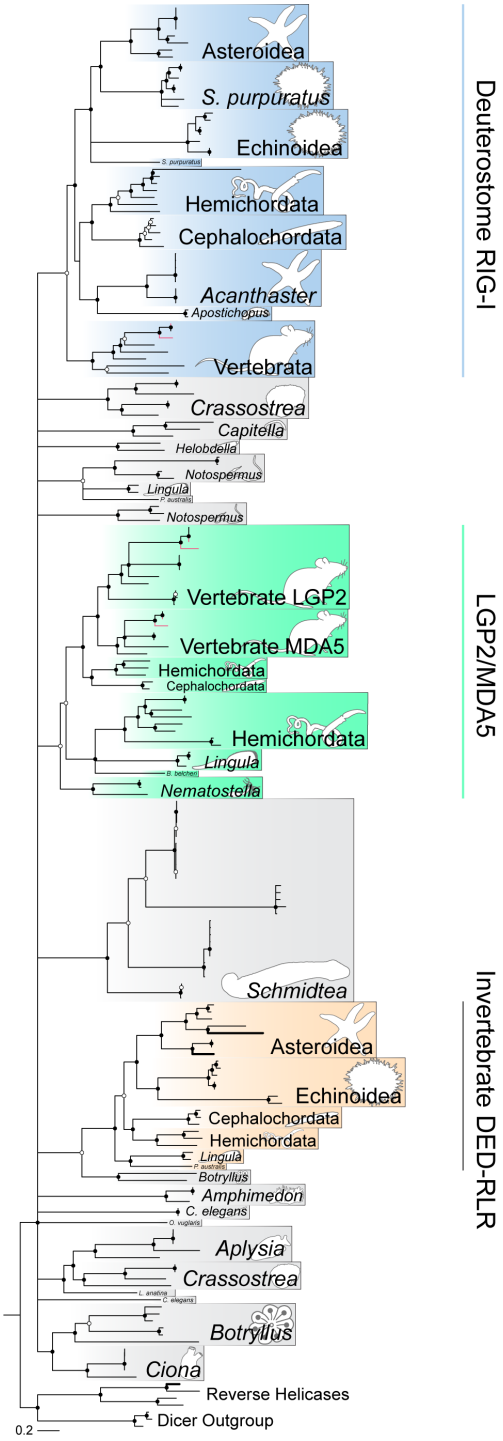
Supplementary Figure 6. Maximum-likelihood phylogenetic reconstruction of IRFs using IQ-TREE. All nodes possess ultrafast-bootstrap support ≥95%. Best-fit model by BIC: VT+R6. Scale bar in number of substitutions per site.



Supplementary Figure 7. Domain architectures associated TIR, RLR C-terminal domain, and NACHT domain revision. Dotted outlines denote domains which are not necessary for protein classification, though may be present.



Supplementary Figure 8. Summarized topology of maximum-likelihood phylogenetic reconstruction of all proteins containing both NACHT domains and LRRs using IQ-TREE. All nodes possess ultrafast-bootstrap support ≥95%. Best-fit substitution model by BIC: VT+R10. Scale bar in number of substitutions per site.



Supplementary Figure 9. Summarized topology of Bayesian RLR phylogenetic reconstruction. Bold branches mark RLRs identified only after domain revision. Nodes with posterior probabilities (PP) <90% are collapsed. Nodes 90≤PP<100% are marked with white circles. Nodes with 100% PP are marked with black circles. Best-fit substitution model by BIC: VT+F+G. Scale bar in number of substitutions per site.

Chapter 6: Conclusions

6.1 Closing Remarks

Contemporary perspectives on the evolution of immunity are primarily derived from work on a select distribution of animal species which established a historical role in biomedical research, including *Drosophila*, *C. elegans*, mouse, and human (Beutler 2004). As such, the evolution of immunity, both its genetic determinants and their functional derivatives, is influenced by a lack of sufficient taxon sampling required for inferring broad evolutionary trends across Metazoa. Although improvements to both the quality and cost of sequencing have yielded considerable advancements towards informing the evolution of immunity-associated gene/protein families, there nonetheless remain many gaps in knowledge which hinder accurate inference of immunity ancestry. The primary goal of my doctoral research was to investigate and describe the genetic components underlying hemichordate immunity in a phylogenetically comparative context adequate for informing deuterostome immunity evolution.

Within Deuterostomia – the superphylum comprised of Hemichordata, Echinodermata, and Chordata – hypotheses on immunity evolution have largely been composed in a comparative context which overlook hemichordates. Whereas improved genomic accessibility has permitted several publications investigating the immune gene complement among several invertebrate deuterostome taxa, including urchins, tunicates, and lancelets (Buckley & Rast 2015), little was known about hemichordate's full immune gene repertoire. Following several publications describing the Toll-like receptor (TLR) repertoire in echinoderms (e.g., Buckley & Rast 2012, Buckley & Rast 2015), I sought to improve resolution of deuterostome TLR pathway evolution by sampling heavily among hemichordates. To this end, in **Chapter 3,** I developed a bioinformatic pipeline to validate homology assignments (to TLR pathway components) by comparing domain architecture of functionally uncharacterized proteins to those among biomedical models where function is known. Ultimately, this study showed that the canonical TLR pathway, which is vital for pathogen recognition by the innate immune system, is conserved across deuterostome lineages. Moreover, by leveraging broad taxon sampling representative of a deeply sampled deuterostome phylogeny, Bayesian phylogenetic inference of deuterostome TLRs showed TLR3, the ortholog responsible for recognizing viral dsRNA in vertebrates, is conserved across deuterostome phyla (with the exception of tunicates). Conservation of TLR3 among invertebrate deuterostomes

lineages suggests the molecular toolkits underlying antiviral immunity may have retained their functional roles over the course of evolution. The findings of this chapter raised additional questions that predicate the subsequent research of my dissertation: A) Can hemichordates recognize and respond to viral stimulus, and B) Do other immune gene families exhibit comparable evolutionary patterns as TLRs?

In **Chapter 4,** I sought to experimentally identify genes directly involved in the antiviral immune response of the acorn worm hemichordate, *Saccoglossus kowalevskii*, and place these results in a comparative context to inform antiviral immunity evolution within Deuterostomia. For this study, my colleagues and I developed an experimental framework where gene expression would be quantifiably compared between *S. kowalevskii* individuals injected with poly(I:C), a synthetic analog of viral dsRNA and potent agonist of dsRNA-dependent immune reactions, and individuals injected only with the solvent buffer. Furthermore, we resolve the temporal effects of gene regulation following injection (with 2-, 4-, 8-, 12-, and 24-hour post-injection timepoints). This study identified 455 genes which exhibit statistically significant differences in mean expression (Wald's adjusted p-value < 0.05) in individuals injected with poly(I:C) when compared with those which received buffer injections. Although some differentially expressed genes (DEGs) could be functionally categorized into conventional immune response categories (e.g., inflammation, apoptosis, or interleukin signaling), a large proportion of DEGs could not be confidently assigned function via sequence similarity to those of known function. Prior to this work, only a small number of similar studies have been performed among invertebrate deuterostomes (let alone among all invertebrate metazoans). Strikingly, among the few comparable studies in echinoderms (the sister phylum to hemichordates), a large proportion of genes also remain uncharacterized by traditional functional annotation methods (Fuess et al. 2015; Wu et al. 2020), suggesting invertebrate deuterostomes possess antiviral genetic toolkits which are not (at least currently) recognized in vertebrates.

In the final data chapter of my dissertation (**Chapter 5**), I investigated the evolution of several protein families central to pathogen recognition and innate immunity. Though this chapter primarily focuses on software I developed to remedy the shortcomings of taxonomic bias within the Pfam database, this study also reports domain architecture diversity and phylogenetic relationships for TLRs, NLRs, RLRs, NFκBs, and IRFs across Metazoa (with deepest sampling

within Deuterostomia). The results of this research show protein architecture diversity among TLRs, NLRs, and RLRs is underestimated when viewed from the lens of biomedical model species. Furthermore, the identification of these protein family members among non-model species is measurably hindered by sequence models which only directly capture variation present within model species. Similar to the findings of **Chapter 3**, hemichordates possess NLRs, RLRs, NFκBs, and IRFs sufficient to host a robust innate immune response following infection – a hypothesis supported by the findings of **Chapter 4**.

In consolidation, the chapters of my dissertation represent a critical step forward for understanding hemichordate immunity and informing the evolutionary history of deuterostome immunity.

6.2 References

Beutler B (2004) Innate immunity: An overview. *Mol. Immunol.* **40**:845–859.

Buckley KM, Rast JP (2012) Dynamic evolution of Toll-like receptor multigene families in echinoderms. *Front. Immunol.* **3**

Buckley KM, Rast JP (2015) Diversity of animal immune receptors and the origins of recognition complexity in the deuterostomes. *Dev. Comp. Immunol.* **49**:179–189

Fuess LE, et al. 2015. Up in Arms: Immune and Nervous System Response to Sea Star Wasting Disease. *PLoS ONE* **10**.

Wu X, et al. 2020. Transcriptomic analysis of sea cucumber (*Holothuria leucospilota*) coelomocytes revealed the echinoderm cytokine response during immune challenge. *BMC Genomics* **21**.